Accepted Manuscript

Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter
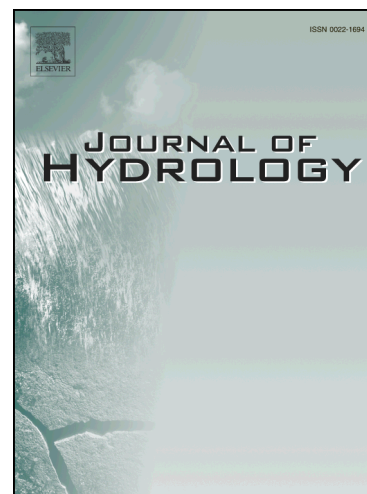
Francesco Zovi, Matteo Camporese, Harrie-Jan Hendricks Franssen, Johan Alexander Huisman, Paolo Salandin

Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter

Francesco Zovi[a,*], Matteo Camporese[a,**], Harrie-Jan Hendricks Franssen[b], Johan Alexander Huisman[b], Paolo Salandin[a]

[a]*Department of Civil, Environmental and Architectural Engineering, University of Padova, Padova, Italy*
[b]*Agrosphere (IBG3), Forschungszentrum Jülich, Jülich, Germany*

Abstract

Alluvial aquifers are often characterized by the presence of braided high-permeable paleo-riverbeds, which constitute an interconnected preferential flow network whose localization is of fundamental importance to predict flow and transport dynamics. Classic geostatistical approaches based on two-point correlation (i.e., the variogram) cannot describe such particular shapes. In contrast, multiple point geostatistics can describe almost any kind of shape using the empirical probability distribution derived from a training image. However, even with a correct training image the exact positions of the channels are uncertain. State information like groundwater levels can constrain the channel positions using inverse modeling or data assimilation, but the method should be able to handle non-Gaussianity of the parameter distribution. Here the normal score ensemble Kalman filter (NS-EnKF) was chosen as the inverse conditioning algorithm to tackle this issue. Multiple point geostatistics and NS-EnKF have already been tested in synthetic examples, but in this study they are used for the first time in a real-world case study. The test site is an alluvial unconfined aquifer in northeastern Italy with an extension of approximately 3 km^2. A satellite training image showing the braid shapes of the nearby river and electrical resistivity

---
[*]now at Lloyds Banking Group General Insurance, Leeds, UK
[**]Corresponding author
Email address: `matteo.camporese@unipd.it` (Matteo Camporese)

tomography (ERT) images were used as conditioning data to provide information on channel shape, size, and position. Measured groundwater levels were assimilated with the NS-EnKF to update the spatially distributed groundwater parameters (hydraulic conductivity and storage coefficients). Results from the study show that the inversion based on multiple point geostatistics does not outperform the one with a multiGaussian model and that the information from the ERT images did not improve site characterization. These results were further evaluated with a synthetic study that mimics the experimental site. The synthetic results showed that only for a much larger number of conditioning piezometric heads, multiple point geostatistics and ERT could improve aquifer characterization. This shows that state of the art stochastic methods need to be supported by abundant and high-quality subsurface data.

Keywords: multiple point geostatistics, normal score transform, ensemble Kalman filter, groundwater, modeling

1. Introduction

Hydrogeological modeling plays a fundamental role for a large number of earth sciences and engineering problems, such as groundwater management, aquifer remediation, and underground waste disposal and management. In or-
5 der to be capable of reliable predictions, models require a detailed knowledge of the aquifer geological structure (e.g., extent and thickness of hydrogeological units, boundary conditions for flow and transport) and flow and transport parameters of the aquifer (e.g., hydraulic conductivity and porosity, dispersivity). In practice, only limited information on the spatial variation of these parameters
10 is available. Therefore, we must often deal with highly uncertain hydrogeologic models, especially with reference to the hydraulic conductivity. Current practice in stochastic hydrogeologic modeling consists of assuming a multivariate Gaussian distribution for log-transformed hydraulic conductivity (De Marsily, 1986). However, cases in nature where this assumption is not valid are abundant, e.g.,
15 when highly permeable preferential flowpaths such as alluvial paleo-channels

2

lie within a statistically homogeneous hydraulic conductivity field. Traditional geostatistical simulators based on two-points correlation models (i.e. the variogram) cannot describe such shapes and can yield inappropriate stochastic realizations failing to reproduce the true probability distribution of the analyzed property (Gómez-Hernández and Wen, 1998). Therefore, more advanced approaches that can tackle these limitations are increasingly being used.

By using copula functions (e.g. Bárdossy, 2006; Bárdossy and Li, 2008) or multiple point geostatistics (MPG) (Caers and Zhang, 2005; Hu and Chugunova, 2008), it is possible to generate any kind of morphological structure, such as curvilinear facies and interconnected channels that often characterize fluvial deposits. Copulas describe the dependence structure between random variables without information on the marginal distributions by representing the dependence between the random variables over the range of quantiles. MPG algorithms do not require the definition of a variogram model, but instead rely on a training image from which the empirical probability distribution of the shapes and structures that need to be mimicked is obtained. MPG algorithms were successfully applied in groundwater flow and transport problems to describe multimodal spatially heterogeneous parameter fields (Feyen and Caers, 2005; Huysmans and Dassargues, 2009). However, in real-world applications, an adequate training image may not always be available. Moreover, the use of an inappropriate training image can lead to unrealistic stochastic realizations, drastically affecting the modeling results (Jafarpour and McLaughlin, 2009).

Among many inversion modeling frameworks, one that is particularly appealing when dealing with groundwater problems with heterogeneous parameters is sequential data assimilation. The primary objective of data assimilation is to provide an optimal estimate of the system state and parameters given a set of measurements and a dynamical model with known uncertainties. The ensemble Kalman filter (EnKF) (Evensen, 1994; Burgers et al., 1998), in particular, has become popular in various scientific fields, such as meteorology, hydrogeology and petroleum engineering (e.g., Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008; Crestani et al., 2013; Houtekamer and Mitchell,

3

1998; Aanonsen et al., 2009; Camporese et al., 2011). The main advantages of EnKF over other data assimilation methods are: (i) the ability to handle different sources of uncertainty in a relatively straightforward manner, thanks to its

50 Monte Carlo-based approach; (ii) the computational efficiency (e.g., Hendricks Franssen and Kinzelbach, 2009); (iii) the possibility to estimate a posterior probability density function (pdf) rather than a single optimal solution; and (iv) the ability to include new observation data from on-line sensors in real-time.

One of the main limitations of the ensemble Kalman filter is that it provides

55 an optimal solution (in a least-squares sense) only if the state variables follow a Gaussian distribution, a condition that is rarely met in practical problems. Therefore, different approaches were developed to account for non-Gaussian distributions. These approaches include: (i) particle filters (Rings et al., 2010; Pasetto et al., 2012), which do not rely on any assumption on the pdf of the

60 states and parameters, but typically require larger ensemble sizes compared to EnKF and can therefore be computationally unaffordable; (ii) Gaussian mixture model methods, which approximate the non-Gaussian probability density functions using a probabilistic model with a finite number of Gaussian pdfs (Chen and Liu, 2000; Apte et al., 2007; Sun et al., 2009; Liu et al., 2016); (iii)

65 transformed re-parametrization, where non-Gaussian system variables are replaced with alternative variables more suitable to be approximated by a Gaussian distribution (Chen et al., 2009; Chang et al., 2010); (iv) iterative EnKF, which addresses the issue of non-Gaussianity by repeating the updating process many times at each update step until the required match between observations

70 and predicted state variables is obtained (Gu and Oliver, 2007; Li and Reynolds, 2007; Reynolds et al., 2006; Sakov et al., 2012; Hendricks Franssen and Kinzelbach, 2008); and (v) the normal score approach, where the non-Gaussian distribution of each state variable is transformed into a Gaussian distribution before each update step. After the update, the normal score transformed variables are

75 back-transformed to their original distribution. Applications of the latter approach in the field of groundwater hydrology can be found in Zhou et al. (2011), Schöniger et al. (2012), and Crestani et al. (2013).

4

Although the normal score ensemble Kalman filter (NS-EnKF) was already combined with multiple point geostatistics to develop a data assimilation frame-

80 work for heterogeneous aquifers with bimodal hydraulic conductivity distributions, to date it was tested only in synthetic studies (Zhou et al., 2011; Li et al., 2012; Schöniger et al., 2012; Zhou et al., 2012; Xu et al., 2013; Xu and Gómez-Hernández, 2016). Only a few case studies can be found in the literature where EnKF-based data assimilation frameworks were tested in real field experiments for pa-

85 rameter estimation in groundwater hydrology (Hendricks Franssen et al., 2011; Panzeri et al., 2015; Crestani et al., 2015). However, such evaluations are essential because many assumptions, such as the adopted geostatistical model, remain highly speculative, due to the inherently limited amount of information that is available to characterize the subsurface.

90 The main objective of this paper is to evaluate the performance of a data assimilation framework for the estimation of heterogeneous subsurface parameters based on EnKF and MPG in a real-world test case. The study area is the Settolo experimental site, an alluvial phreatic aquifer with an extension of approximately 3 km^2 located in Northeastern Italy. The aquifer is character-

95 ized by the presence of paleo-riverbeds acting as preferential flowpaths. We use satellite data to obtain a training image that captures the plausible characteristic shapes of the paleo-riverbeds. Piezometric head data collected in the field are then assimilated with NS-EnKF into a two-dimensional groundwater flow model to update states and aquifer parameters. An additional novel as-

100 pect of this study is the use of electrical resistivity tomography images in the MPG simulator to condition the channel positions in the prior realizations of the hydraulic conductivity fields. Along with the application to the field case study, a synthetic case study that mimics the Settolo site is developed, in order to give further insights on the potential and drawbacks of the proposed

105 assimilation/inversion framework.

5

2. Inversion approach

2.1. Groundwater model

Two-dimensional horizontal flow in an unconfined aquifer is described by the nonlinear Dupuit-Boussinesq equation (De Marsily, 1986):

$$\frac{\partial}{\partial x}\left[K_x\left(h-z_0\right)\frac{\partial h}{\partial x}\right]+\frac{\partial}{\partial y}\left[K_y\left(h-z_0\right)\frac{\partial h}{\partial y}\right]=\left[S_s\left(h-z_0\right)+S_y\right]\frac{\partial h}{\partial t}+q_a, \quad (1)$$

where K_x and K_y (m/s) are the hydraulic conductivities along the x and y spatial coordinates, respectively, h (m) is the water table elevation, z_0 (m) is the elevation of the aquifer bottom (assumed impermeable), S_s (m^{-1}) is the specific storage, S_y is the specific yield (m^3 m^{-3}), and q_a (m/s) represents a

110

115 generic nodal source or sink term. Equation (1) is integrated in space by means of the linear finite element method. An unstructured mesh can be used, allowing variable element size according to different spatial discretization needs. For integration over time, an adaptive time stepping with backward Euler scheme is adopted. The model is completed by appropriate initial and boundary con-

120 ditions that will be described in detail later.

We opted for a two-dimensional model after a preliminary comparison with a fully three-dimensional Richards equation solver (Zovi, 2014). The comparison highlighted that the groundwater dynamics in the Settolo aquifer is mainly horizontal and using a three-dimensional model would only imply a much larger

125 computational effort without evident benefits in terms of solution accuracy.

2.2. Ensemble Kalman Filter with normal score transform

The normal score ensemble Kalman filter (NS-EnKF) is a variant of the EnKF (Burgers et al., 1998; Evensen, 1994) that addresses the non-Gaussianity of the system state variables by transforming the components of the state vectors

130 for all ensemble members, such that each of these components honors a standard Gaussian distribution (Zhou et al., 2011; Li et al., 2012). In this study, we integrate the NS-EnKF with the previously described groundwater model to update the system state and parameters through the assimilation of observed

6

groundwater levels. While Schöniger et al. (2012) used the NS-EnKF for subsur-

135 face parameter estimation only, based on the parameter space EnKF proposed by Nowak (2009), parameter updating is here achieved here by considering an augmented system state vector $\mathbf{x}$

$$\mathbf{x} = \left[\mathbf{h}^T \ \mathbf{Y}^T \ \mathbf{S_y}^T \ \mathbf{S_s}^T \right], \tag{2}$$

where $\mathbf{h}$ is a N_n-dimensional vector containing the water table elevations in

140 each node of the numerical mesh, $\mathbf{Y}$ consists of the hydraulic log-conductivity values ($Y = logK$) in the N_e elements, while $\mathbf{S_y}$ and $\mathbf{S_s}$ are one-dimensional, as we assume that specific yield and specific storage are spatially homogeneous. A total of NMC realizations of the vector $\mathbf{x}$ are collected in a matrix $\mathbf{X}$ of dimensions $N_x \times NMC$, where $N_x = N_n + N_e + 2$.

145 The proposed inversion approach can be summarized as follows.

1. Generation of the prior ensemble. NMC equally likely stochastic realizations of Y, S_y and S_s are generated, assuming they are the only sources of uncertainty. To test different assumptions and capture all the relevant sources of uncertainty, the initial ensemble of Y is generated by differ-150 ent combinations of the Single Normal Equation Simulation (SNESIM) algorithm (Strebelle, 2002; Liu, 2006; Remy et al., 2009), for generating facies distribution, and Sequential Gaussian Simulation (SGSIM) algorithm (Gómez-Hernández and Journel, 1993; Deutsch and Journel, 1998; Remy et al., 2009), for populating the different facies. Available condi-155 tioning information consists of electrical resistivity tomography data and satellite images. The initial ensemble of S_y and S_s is generated by drawing random numbers from uniform distributions with pre-set minimum and maximum values. The initial ensemble of groundwater levels (i.e., initial conditions) is generated with a steady-state run for each realization 160 of the parameters. This generation step will be described in more detail in Section 4.2.

2. Forward model evaluation. The groundwater model is run from the previous assimilation time t_{i-1} to the current assimilation time t_i for each

7

realization of the ensemble in order to compute the distribution of ground-
165 water levels h. This step is the most computationally intensive, because
NMC model evaluations are needed. However, these model evaluations
can be easily run in parallel, thereby reducing the computational time.

3. Normal score transformation. All entries of the augmented state vector
(h, Y, S_y, and S_s values), as well as the observed heads z, are normal
170 score transformed:

$$\hat{h} = \Phi_h\left(h\right), \quad \hat{Y} = \Phi_Y\left(Y\right), \quad \hat{S_y} = \Phi_{S_y}\left(S_y\right), \quad \hat{S_s} = \Phi_{S_s}\left(S_s\right), \quad (3)$$

where $\hat{h}$, $\hat{Y}$, $\hat{S_y}$, and $\hat{S_s}$ are the normal score transformed variables and for
each entry of the state vector a different transformation function (Φ_h, Φ_Y,
Φ_{S_y}, and Φ_{S_s}) is generated. For Y, we always use the initial function Φ_Y
175 computed at the first assimilation time, which remains unchanged dur-
ing the iterations, in order to preserve the prior model structure and the
statistical information included in the initial ensemble (Zhou et al., 2011;
Li et al., 2012). For h, S_y, and S_s, we use a new transformation function
at each time step. The observed heads z are transformed into $\hat{z}$ using a
180 transformation function derived from the simulated heads in the nodes
corresponding to the measurement locations, following Schöniger et al.
(2012). For all the transformations, anamorphosis functions are built that
link the original variable to its transformed equivalent. One of the draw-
backs of this approach is that the relationship between untransformed and
185 transformed variables is only known in a finite number of discrete points,
so all the other values must be obtained by linear interpolation. To pre-
vent the generation of unrealistic values, we make sure that extrapolation
never occurs by adding reasonable and realistic minimum and maximum
extreme values to the reference dataset, whose respective Gaussian cumu-
190 lative distribution function (CDF) values are set to 0.000001 and 0.999999,
respectively.

4. Assimilation. An augmented state vector $\hat{\mathbf{X}}$ is built by assembling $\hat{\mathbf{h}}$, $\hat{\mathbf{Y}}$,
$\hat{\mathbf{S_y}}$, and $\hat{\mathbf{S_s}}$, and then updated by means of the measurements available in

8

N_{obs} observation points, according to the following equations:

$$\hat{\mathbf{X}}^u = \hat{\mathbf{X}} + \boldsymbol{\alpha}\hat{\mathbf{G}}\left(\hat{\mathbf{z}} - \mathbf{H}\hat{\mathbf{X}}\right), \tag{4}$$

$$\hat{\mathbf{G}} = \hat{\mathbf{P}}\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\hat{\mathbf{P}}\mathbf{H}^{\mathbf{T}} + \hat{\mathbf{R}}\right)^{-1}, \tag{5}$$

$$\hat{\mathbf{P}} = \frac{1}{NMC}\left(\hat{\mathbf{X}} - \overline{X}e^T\right)\left(\hat{\mathbf{X}} - \overline{X}e^T\right)^T, \tag{6}$$

$$\overline{X_i} = \frac{1}{NMC}\sum_{j=1}^{NMC}\hat{X}_{i,j}, \ i = 1,\ldots,N_x, \tag{7}$$

where $\hat{\mathbf{X}}^u$ is the updated transformed state vector of size $[N_x \times NMC]$; $\boldsymbol{\alpha}$ is a dampening diagonal matrix $[N_x \times N_x]$ that contains dampening factors (ranging between 0 and 1) for each component of the augmented state vector (Hendricks Franssen and Kinzelbach, 2008); $\hat{\mathbf{G}}$ is the Kalman gain $[N_x \times N_{obs}]$; $\hat{\mathbf{z}}$ is the ensemble of transformed observations $[N_{obs} \times NMC]$, obtained by perturbing the N_{obs} transformed observation values with a random noise characterized by mean equal to zero and standard deviation ϵ (Burgers et al., 1998); $\hat{\mathbf{R}}$ is the transformed observation error covariance matrix $[N_{obs} \times N_{obs}]$), which here is assumed to be diagonal, i.e., errors at different measurement locations are uncorrelated; $\mathbf{H}$ is the observation matrix $[N_{obs} \times N_x]$ that maps the measurements to the augmented state and in our case is composed only by 0's and 1's, as measurement locations correspond exactly to mesh nodes; $\hat{\mathbf{P}}$ is the augmented state covariance matrix $[N_x \times N_x]$, with $\overline{X}$ being the ensemble mean $[N_x \times 1]$ and $e^T = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}$ $[1 \times NMC]$. If $N_{obs} \ll N_x$, as in most practical applications, the Kalman gain can be computed in an efficient way without the need to calculate directly the whole ensemble covariance matrix $\hat{\mathbf{P}}$, saving memory and CPU operations (Evensen, 2003). Covariance localization and inflation, as described by Kurtz et al. (2012, 2014) and Xu et al. (2013), are also implemented in the inversion approach.

5. Normal score back-transformation. After the update, the augmented state vector is back-transformed to its original distribution using the inverse of

9

the transformation functions:

$$h = \Phi_h^{-1}\left(\hat{h}\right), \quad Y = \Phi_Y^{-1}\left(\hat{Y}\right), \quad S_y = \Phi_{S_y}^{-1}\left(\hat{S}_y\right), \quad S_s = \Phi_{S_s}^{-1}\left(\hat{S}_s\right), \tag{8}$$

The inverse of the transformation function is not explicitly computed, but is approximated by an interpolation of the constructed anamorphosis function. The updated h distribution is the initial condition for the next time step, and points 3 to 5 are repeated until the end of the simulation period.

The ensemble Kalman filter can be affected by filter inbreeding, which is an undesired effect whereby ensemble variance is increasingly underestimated with time. This is due to spurious covariances between pairs of grid cells where no correlation would be expected, but that arise due to the finite ensemble size (Evensen, 2009). Updating on the basis of these spurious covariances causes a strong variance reduction, as observations are interpreted as informative by the filter whereas in reality they are not. Filter inbreeding can be reduced by using a large number of realizations (e.g., Chen and Zhang, 2006) or techniques such as dampening (Hendricks Franssen and Kinzelbach, 2008) and covariance inflation or localization (Evensen, 2009). Localization aims to remove spurious correlations from the covariance matrix (Nan and Wu, 2011). Here we adopt an isotropic fifth-order correlation function (Hamill et al., 2001) that sets to zero the covariance for pairs of points that are separated by a distance larger than a pre-defined length scale. Covariance inflation aims at limiting filter inbreeding by empirically inflating the covariance matrix. Here, the time-dependent inflation algorithm proposed by Wang and Bishop (2003) is used.

3. Study area

The research site is located along the left bank of the Piave River, in the Northern Treviso province (North-East Italy), close to the city of Valdobbiadene (Figure 1). The area lies in a piedmont region of great relevance for aquifer

10

recharge, as the river leaves its mountain catchment and feeds the underlying unconfined aquifer. The site has an extension of 2.7 km^2 and elevation ranging

₂₅₀ between 155 to 165 m a.s.l. The aquifer is heavily exploited for irrigation and drinking water supply. The main drivers of groundwater flow are infiltration from the Piave River, which determines the southern and western boundary of the groundwater flow domain, and water extraction from a number of drinking water wells, indicated with p1, p2, and p3 in Figure 1.

₂₅₅ River stage and groundwater level are continuously measured in twenty locations as shown in Figure 1. Four of these locations are used to define the time-variable boundary conditions of the groundwater model domain (I1, I2, w7 and wS). The remaining data from sixteen boreholes are used for data assimilation. Pumping rates at the production wells and rainfall rate are also

₂₆₀ continuously measured (Zovi, 2014). More details about the experimental site can be found at http://settolo.dicea.unipd.it/index.php.

Additional available data include five electrical resistivity tomography (ERT) profiles (Figure 2), whose position is shown in Figure 1. The resistivity images were obtained using the 2-D resistivity, 3-D current ERT code R2 by A. Binley,

₂₆₅ Lancaster University (http://www.es.lancs.ac.uk/people/amb/Freeware/freeware.htm), to invert field data collected with 10 m resolution Schlumberger arrays and the roll-along technique to concatenate consecutive transects. The ERT profiles highlight a pronounced zonal heterogeneity mainly due to the presence of highly permeable paleo-riverbeds (Sartor, 2006). Therefore, it is assumed that a bi-

₂₇₀ modal hydraulic conductivity distribution would be needed to reproduce the heterogeneity of the study area.

In-situ pumping and slug tests resulted in hydraulic conductivity values that range from about 5×10^{-4} to 5×10^{-2} m/s, specific yield values from 0.16 to 0.29, and specific storage values from 1.5×10^{-2} m^{-1} to 2.5×10^{-2} m^{-1}.

₂₇₅ A preliminary analysis of the pumping test data with the method proposed by Schneider and Attinger (2008) resulted in an estimate of the horizontal integral scale of 84 m. Note that all these in-situ tests were performed in a small area around well p1 (Figure 1) and therefore cannot be considered as representative

11

for the whole aquifer.

4. Model and simulation setup

4.1. Model setup

The Settolo aquifer was discretized with an unstructured finite element triangular mesh consisting of 11939 elements and 5866 nodes. To be able to better reproduce the local groundwater table variations, finer triangles (2 m size) were used around the pumping wells, while coarser triangles of 20 m size were used away from the wells. The latter size corresponds to almost 1/4 of the minimum expected integral scale (84 m). The locations of observation and pumping wells correspond exactly to nodes of the mesh.

Boundary conditions for the groundwater model are shown in Figure 1. The Piave river to the western and southern boundary of the flow domain constitutes a known variable head boundary condition that is measured by the gauges I1 and I2 (Figure 1) and linearly interpolated along the A-B segment. At the northern boundary, a fluvio-glacial alluvial terrace constitutes an almost impermeable boundary (C-D and E-A segments), except where the Funer valley intersects the test site (D-E segment). Here, the groundwater level measured in well w7 is used in order to account for the recharge contribution from the upstream basin. To the south-east, the boundary condition for the B-C segment is obtained from the groundwater level measurement in well wS. Pumping well extraction rates are imposed as sink terms in the corresponding nodes.

A preliminary calibration of the groundwater flow model with homogeneous parameters using measured groundwater levels gave a value of $1.4 \cdot 10^{-2}$ m/s for the hydraulic conductivity, which is one order of magnitude higher than the values obtained with pumping and slug tests. Based on the ERT transects, we adopt a conceptual model dividing the aquifer in two different zones: channels, with higher hydraulic conductivity, in a background with lower conductivity. We assume that channel segments (indicated as red transparent rectangles in Figure 2) correspond to portions of the subsurface where the 400 Ωm iso-resistivity line

12

extends below a depth ranging from an average of 30 m in the North to 50 m in the South. This depth is indicated by red lines in Figure 2 and represents the
310 assumed aquifer thickness in the groundwater model.

4.2. Real-world experiments

We simulated a time period of 365 days (from 1 February 2011 to 31 January 2012) and selected a daily interval for data assimilation. The selected time period includes both flood events and a long water production period, in order
315 to consider the aquifer response to different forcing conditions (Figure 3). In our data assimilation experiments we used an ensemble size (NMC) of 2000, which is often considered to be large enough for accurate estimation and avoiding filter inbreeding (Hendricks Franssen and Kinzelbach, 2008; Camporese et al., 2015).

A binary indicator field identifying the facies distribution (i.e., channel and
320 matrix) was generated with the SNESIM algorithm (Strebelle, 2002; Liu, 2006; Remy et al., 2009). Next, both facies were populated with hydraulic conductivity values generated by the SGSIM algorithm (Gómez-Hernández and Journel, 1993; Deutsch and Journel, 1998). The parameters for both algorithms were assigned according to the following steps:

325 1. *Training image.* A crucial point for multiple point geostatistics, as pointed out by Jafarpour and McLaughlin (2009), is the need for a representative training image, because its choice highly affects the shape of the generated facies. We derived a training image from a satellite image of the current floodplain of the Piave River (Figure 4). By doing so, it is assumed that
330 the geomorphological phenomena responsible for the deposition of fluvial sediments in past ages are also responsible for the generation of present braided river shapes. This assumption might be considered questionable, because past morphodynamic conditions, i.e., river discharge and sediment size and type, were not the same as nowadays, due to tectonic activity,
335 climate change and anthropogenic activities such as the construction of fluvial dams and dikes (Leopold et al., 1957; Leopold, 1995). Furthermore,

13

the derivation of a training image from a satellite image is subject to interpretation, for example in the choice of the channel widths. Nevertheless, we considered the use of satellite photographs the only feasible choice in this case.

To partly address uncertainty in the selection of the training image, we added uncertainty to the MPG generation by randomly varying several parameters of the SNESIM algorithm. The varied parameters were: P_x and P_y, i.e., the training image scaling factors for the generated channels in x and y directions, respectively, the planar rotation α_{TI} (i.e., the inclination on the $x - y$ plane of the generated channels), and the areal fraction of the two facies $\%_{ch}$. For each realization, these four parameters were randomly drawn from a uniform distribution with minimum and maximum values as defined in Table 1. This approach is similar to the one used by Jafarpour and Tarrahi (2011).

2. *Conditioning to geophysical data.* The facies distribution obtained from the ERT transects was considered as conditioning data for the facies generation with SNESIM (Figure 4). As the distinction between channels and background in the ERT transects is also subject to interpretation, the geophysical data should not be considered as hard data during conditioning. Therefore, uncertainty was added to this conditioning step by imposing a 60% probability that a particular conditioning point obtained from the geophysical data is honored by SNESIM. In addition, 20% of the realizations were not conditioned to geophysical data in the initial ensemble, to allow for diversity and the possibility that our geophysical interpretation is not accurate.

3. *Population of the facies with Gaussian random fields.* The two zones obtained from SNESIM are populated with hydraulic conductivity values extracted from a Gaussian random field (GRF) using the SGSIM algorithm (Gómez-Hernández and Journel, 1993; Deutsch and Journel, 1998; Remy et al., 2009). An anisotropic exponential correlation function was adopted. The main uncertainty lies in the definition of the mean ($\langle Y \rangle$),

14

variance (σ_Y^2), integral scales in the x and y direction (I_x and I_y), and anisotropy angle (α_{xy}) of the correlation function for both the high and
370 low conductivity zones. From available field data (Section 3), plausible ranges for these parameters were obtained (Table 1). As for the SNESIM parameters, a set of values was obtained using a uniform random draw between the predetermined minimum and maximum values, independently for the two facies. In this way, the initial ensemble contains realizations
375 generated with a range of geostatistical parameters, which is expected to lead to a better representation of the uncertainty in the structure of the $\log K$ field (Jafarpour and Tarrahi, 2011). It is then left to the EnKF to condition the realizations to the available groundwater measurements.

To summarize, each realization is generated as follows: (1) extract random val-
380 ues for SNESIM parameters (P_x, P_y, α_{TI}, $\%_{ch}$); (2) generate facies distributions with SNESIM from the training image, conditioning 80% of the realizations to geophysical data; (3) extract random values for SGSIM parameters ($\langle Y \rangle$, σ_Y^2, I_x, I_y, α_{xy}), independently for each zone; (4) populate both zones with random fields generated with SGSIM. This procedure was used to generate a first
385 ensemble of hydraulic conductivity fields, hereafter denoted as "Ensemble 1". Panels a), b), and d) of Figure 5 show three examples of realizations from this ensemble.

A second ensemble, denoted as "Ensemble 2", was generated by using the same procedure as described above, with the exception that the facies distri-
390 bution was not conditioned to our interpretation of the ERT transects. One realization sampled from this ensemble is shown in Figure 5c) as an example.

To further elaborate on the advantages of considering MPG, the results of the simulations with the first two ensembles were compared with those obtained with a more standard approach to generate the initial ensemble. Here, an initial
395 set of Gaussian random fields without zonation was generated with SGSIM (Gómez-Hernández and Journel, 1993; Deutsch and Journel, 1998; Remy et al., 2009). Again, uncertainty in the definition of the geostatistical parameters for

15

SGSIM was considered following the approach outlined above. One realization of this ensemble, denoted as "Ensemble 3", is provided as an example in Figure 6.

⁴⁰⁰ Two other groundwater parameters considered uncertain are specific yield and specific storage (S_y and S_s, respectively). Their influence on the groundwater dynamics in our site, however, is less important than that of Y. This was confirmed by a preliminary sensitivity analysis of the groundwater model. For this reason, it was assumed that these parameters are homogeneous in space.

⁴⁰⁵ For every realization, S_s and S_y were sampled randomly from a uniform distribution in the ranges $0.1 \leq S_y \leq 0.35$ and $10^{-5} \leq S_s \leq 10^{-1}$.

The proposed approaches were evaluated using six scenarios, as shown in Table 2. Each of the three initial ensembles were conditioned by assimilation of groundwater level data with both the NS-EnKF and the standard EnKF

⁴¹⁰ (without normal score transformation). The parameter ensemble resulting from each scenario was then assessed by running the groundwater model without data assimilation for an additional 90-day verification period (from 1 December 2010 to 28 February 2011).

4.3. Synthetic inversion experiments

⁴¹⁵ To better investigate the capabilities and limitations of the proposed inversion approach for our experimental site, we also conducted a synthetic test, which mimics the real-world case. More specifically, this synthetic test case was carried out for three main purposes. First, due to the limited number of observation wells available in the real-world case, it is important to assess to what

⁴²⁰ extent the results of the real-world case are a consequence of the limited amount of information available. Tests with different data sets allow us to evaluate the improvement in aquifer characterization and model predictions that could be achieved if more data were available. Second, a synthetic test case can give additional insights on the performance of the normal score transformation and

⁴²⁵ the conditioning to geophysical data and further support the conclusions drawn from the real world case study. Third, for the synthetic test case, covariance localization and inflation were also tested, to explore whether the use of such

16

techniques can significantly improve the data assimilation results.

Using the same approach as described in section 4.2, a $\log K$ field was gen-
430 erated that was considered as the reference field. The forward problem was
solved to generate synthetic groundwater levels from which measurement data
were taken for the subsequent data assimilation scenarios. The same boundary
conditions, time window, mesh discretization, groundwater model setup and
NS-EnKF parameters were used as in the real-world application. Thus, the
435 only important difference was that we knew a priori the true $\log K$ distribution,
so that the estimation error could be directly estimated.

The reference field shown in Figure 7 was generated with the same training
image used in the real-world application, and was conditioned to geophysical
data, as in section 4.2. Synthetic numerical experiments were then conducted
440 using 14 scenarios, which are summarized in Table 3. We used three different
sets of observation points: 16 observations, distributed exactly like the real-
world case; 65 observations, representing a refined measurement network, and a
much refined case with 250 observations. Each of the three sets was tested with
and without normal score transformation and with and without conditioning
445 on geophysical data. Covariance localization and inflation were tested only for
the case with 65 observations, normal score transformation and conditioning to
geophysical data.

4.4. Performance assessment

To assess the performance of the proposed modeling framework in the various
450 scenarios, a number of measures were used. At each assimilation time t the Y
field is updated by EnKF (with or without normal score transform, depending
on the scenario), and for each grid element i the ensemble mean is calculated
as follows:

$$EM_Y\left(i,t\right) = \frac{1}{NMC} \sum_{j=1}^{NMC} Y_{i,j,t};$$ (9)

455 while the ensemble variance is given by

$$EV_Y\left(i,t\right) = \frac{1}{NMC} \sum_{j=1}^{NMC} \left(Y_{i,j,t} - EM_Y\left(i,t\right)\right)^2;$$ (10)

17

and the ensemble spread by:

$$ES_Y(t) = \frac{1}{A_{tot}} \sum_{i=1}^{N_e} A_i EV_Y(i,t) \tag{11}$$

where $Y_{i,j,t}$ is the $\log K$ of the ith element, j^{th} realization at time step t, A_{tot}
is the total area of the mesh and A_i is the area of the ith element. The EM_Y
map after the last update (365th day) represents the best possible estimate of
the true Y field that can be obtained from the available data, while the EV_Y
map represents the local degree of uncertainty in the estimation.

To evaluate the performance of the best estimate of the $\log K$ field as well as
the whole final ensemble of realizations for each scenario, the root mean square
error of the groundwater levels simulated in the 90-day verification period was
calculated in two ways:

$$RMSE_{wl}^{ens} = \sqrt{\frac{1}{N_{obs} \cdot N_t \cdot NMC} \sum_{j=1}^{NMC} \sum_{t=1}^{N_t} \sum_{p=1}^{N_{obs}} \left(h_{p,j}(t) - h_p^{meas}(t) \right)^2}, \tag{12}$$

and

$$RMSE_{wl} = \sqrt{\frac{1}{N_{obs} \cdot N_t} \sum_{t=1}^{N_t} \sum_{p=1}^{N_{obs}} \left(\overline{h_p}(t) - h_p^{meas}(t) \right)^2}, \tag{13}$$

where $\overline{h_p}(t)$ is the simulated water level at the pth observation point and t^{th}
time step obtained using the ensemble mean distribution EM_Y, $h_p^{meas}(t)$ is
the measured groundwater level, and $h_{p,j}(t)$ is the simulated water level using
the jth realization of the ensemble at the pth observation point and t^{th} time
step. To validate the model, a time discretization suitable to fully catch the
groundwater dynamics was chosen, i.e., one observed data every 3 hours (8 time
steps each day). $RMSE_{wl}^{ens}$ represents the capability of the whole ensemble to
reproduce the piezometric data. $RMSE_{wl}$, instead, represents the capability of
the ensemble mean to reproduce the real data. The ensemble mean represents
the "best estimate", but does not contain all the information about uncertainty
that the filter can provide. Moreover, the hydraulic head distribution resulting
from the best Y field estimate is not necessarily the best head field due to

18

nonlinearity. Therefore, for completeness, we included both measures in our performance analysis.

485   The data assimilation results are also compared with an open-loop run, i.e., $RMSE_{wl}^{ens}$ and $RMSE_{wl}$ were also computed with the initial ensemble without any update by EnKF or NS-EnKF. The open-loop run represents the initial level of uncertainty and is the estimate without data assimilaton. Such a comparison highlights the benefits of using EnKF or NS-EnKF for updating the log$K$ field.

490   Table 2 summarizes the six scenarios with the relevant parameters.

In the synthetic test case, the true hydraulic conductivity distribution is known and we can therefore evaluate directly the error of the Y estimates. Here, the mean absolute error (MAE) weighted with the element area was used, instead of the $RMSE$, as it is less sensitive to the occasional large error

495   introduced by the possible misrepresentation of different local zones. As for the $RMSE$, also MAE can be computed in two ways:

$$MAE_Y^{ens}(t) = \sqrt{\frac{1}{NMC \cdot A_{tot}} \sum_{i=1}^{N_e} \sum_{j=1}^{NMC} A_i |Y_{i,j}(t) - Y_i^{true}|}, \qquad (14)$$

and

$$MAE_Y(t) = \sqrt{\frac{1}{A_{tot}} \sum_{i=1}^{N_e} A_i |EM_Y(i,t) - Y_i^{true}|}, \qquad (15)$$

500   where Y_i^{true} is the reference log$K$ value of the ith element. In addition, a facies matching ratio ($FMR_\%$) was calculated in the synthetic test case by converting a continuous log$K$ field into a binary indicator field. In our experiments, we used a threshold value of -4.6 log(m/s), which corresponds to the centre of mass of the histogram of the reference log$K$ field shown in Figure 7. The $FMR_\%$ is

505   the percentage of area that was correctly classified as channel or background material.

19

5. Results and discussion

5.1. Real-world test case

The wall-clock time required to complete the data assimilation simulations was around 30 hours on a 4-core desktop workstation. However, in a second round of runs it was possible to reduce such time to 25-30 minutes on a cluster with 256 allocated processors.

Table 2 reports the performance statistics of the six scenarios for the real-world application, while Figures 8 and 9 show the six retrieved EM_Y maps for all scenarios and the corresponding EV_Y maps. The obtained $\log K$ distributions of all the scenarios show similarities in the zone where the measurements are most dense (Figure 8), while major differences arise elsewhere. In all six scenarios no clear curvilinear channel shapes seem to emerge. Nevertheless, the $RMSE_{wl}^{ens}$ and $RMSE_{wl}$ for groundwater levels (Table 2) show an improvement with respect to the open loop runs for all scenarios. The best performing scenario is the one with the initial ensemble generated with a GRF and without normal score transformation (scenario 3), while initial ensembles generated with MPG exhibit a poorer performance. This is probably due to the fact that the measurements are mostly concentrated around the central part of the domain and cannot provide enough information to identify the channels. On the other hand, scenario 3, which is not constrained by MPG and ERT data, has more degrees of freedom and can adapt better to the observations.

Figures 8 and 9 also show that scenarios with normal score transformation always result in a final ensemble variance that is larger compared to the corresponding runs without normal score. It has been demonstrated that transforming the variables can introduce noise into the EnKF procedure, due to a significant reduction of the correlation between state variables and parameters (Schöniger et al., 2012; Crestani et al., 2013). As a result, in the scenarios with normal score transform, the updates cannot propagate in space as they would normally do without transformation and therefore the hydraulic head values far from the measurement locations have more freedom to differentiate and main-

20

tain a higher variance.

Figures 10 and 11 show measured and simulated groundwater levels for the 90-day verification period in different observation wells and for scenarios 540 3 (multiGaussian random fields) and 4 (MPG with conditioning on ERT data and normal score transform), respectively. The results highlight that, although on average the performance of scenario 3 is better, scenario 4 exhibits a larger ensemble spread that always envelops the real data, even if the ensemble mean is far from the true response. On the other hand, scenario 3 has a smaller en-545 semble spread, thus all the realizations tend to collapse to the ensemble mean, leading to some significant local mismatches, as in borehole w42 (Figure 10). This is further confirmed by the values of ES_Y in Table 2. Therefore, scenario 4 (ensemble 1 with normal score transform) seems to better represent the uncertainty of the estimation and it can be argued that $RMSE$ alone is not sufficient 550 to capture the overall performance of the inversion results, as it does not provide a meaningful measure of the uncertainty.

It is difficult to determine which approach provides the best estimates of the actual $\log K$ distribution, because of two reasons: (i) the observation wells used for data assimilation were used also for performance evaluation, and this can 555 lead to an unrealistic validation; (ii) the use of only 16 observation wells seems to be inadequate to obtain good inversion results. This latter issue also cleary emerges from the ensemble variance maps in Figures 8 and 9.

5.2. Synthetic test case

Table 3 summarizes the performance of the fourteen synthetic test case sce-560 narios. As in the real-world case, the results of the synthetic test case also show an improvement for all the scenarios with respect to the open loop run. Figures 12 and 13 show that no channels seem to appear in the inversion results obtained with only 16 observations. This is fully consistent with the results of the real world case study. In contrast, the inversion results obtained with 65 and 565 250 observations show much clearer channels that are correctly retrieved. Increasing the number of observation points clearly leads to a better performance,

21

regardless of which measure of performance is considered (Table 3).

The performance statistics in terms of MAE_Y^{ens} (Table 3) show that conditioning initial realizations to ERT transects results in an improvement when the
570 number of observation wells is increased. Compared to the open loop, MAE_Y^{ens}
for scenarios 1, 5, and 11 (i.e., 16, 65, and 250 observations) is 10%, 24%, and
28% lower, respectively. Scenarios where normal score transformation was used
also resulted in a better delineation of the channels (Figure 13), with sharper
boundaries. Scenarios without normal score transform exhibit a smooth tran-
575 sition between the two facies, resulting in some $\log K$ values (cyan elements in
Figure 12) that are not represented in the reference $\log K$ field (Figure 7).

Analyzing the results in Table 3, the best scenario for predicting Y (based
on MAE_Y^{ens}) for the case of 16 observation points (as for the real-world appli-
cation) is without normal score transform and with geophysical conditioning,
580 while for the cases with more observations (65 and 250), the lowest MAE_Y^{ens} is
obtained for the cases with normal score transform and geophysical condition-
ing. Compared to the same scenarios without normal score transformation, the
cases with 65 or 250 observations result in values of MAE_Y and MAE_Y^{ens} that
are 2-5% smaller if normal score transformation is used, while geophysical con-
585 ditioning can reduce by up to 12% the error if compared to the same scenarios
without geophysical conditioning. A ranking of the simulation runs based on
groundwater levels ($RMSE_{WL}^{ens}$) leads to quite different results. For instance,
in all the cases with 16, 65 and 250 observation points, the best performing
scenario in terms of $RMSE_{WL}^{ens}$ is the one without normal score transform and
590 geophysical conditioning, while different rankings are obtained for MAE_Y and
$RMSE_{WL}$. The latter gave the best scores in the cases with 65 and 250 observa-
tions if conditioning to geophysical information was performed, without normal
score transformation. Nevertheless, normal score transformation resulted in just
slightly larger errors. This suggests that the validation obtained by the $RMSE$
595 of groundwater levels computed at the assimilation nodes may not be the best
indicator of model performance and can lead to misleading interpretations. Val-
idation based on facies matching ratio ($FMR_\%$ in Table 3) is in agreement with

22

$RMSE_{WL}$, except for the case with 16 observations, and also gives the best results if geophysical information was used for conditioning. Again, results were
₆₀₀ slightly worse if in addition a normal score transformation was performed.

In general, covariance localization and inflation did not improve the assimilation results (compare scenarios 9 and 10 in Table 3 with scenario 7). We attribute this to the fact that uncertainty was already well-represented in the scenarios and therefore there was no need to correct for variance underestima-
₆₀₅ tion. Moreover, localization based on isotropic distance is probably not appropriate for the case considered here, because it excludes all the $\log K$ values that fall outside a certain distance from the measurement points from the update. As the Settolo aquifer has a main flow direction from north-west to south-east, the piezometric behavior in one point highly depends on the K of all the up-
₆₁₀ stream locations. The localization scheme adopted here, with a 600 m exclusion range, probably excluded some important information that the NS-EnKF needs to suitably update the system states.

5.3. Discussion

Overall, the results for the real-world case study illustrate the problems
₆₁₅ we can expect to encounter in practical applications of MPG-NS-EnKF. Even though geological evidence suggests the presence of channelized structures that cannot be reproduced by a multiGaussian model, it cannot be demonstrated that an MPG-based model outperforms a multiGaussian one. The results are not clearly improved by conditioning to geophysical information either, or by us-
₆₂₀ ing NS-EnKF instead of EnKF. This may be related to the limited information available: (i) imprecise information about the geological structure so that the training image (or in general the geostatistical model) is uncertain, and (ii) too few hydraulic conductivity and piezometric head data to constrain the spatial distribution of hydraulic conductivity. In line with our results, Schöniger et al.
₆₂₅ (2015) found that hydrogeological inversions based on geostatistics are significantly affected by prior geological knowledge and incorporating wrong information may lead to poor performance. In our study, the prescription of the channel

structure does not seem accurate enough and the hydrogeophysical information from ERT is either too uncertain or too little for the inversion to actually bene-

630 fit from this hard constraint, while in Schöniger et al. (2015) a relatively simple geological zonation based on accurate prior information clearly outperformed Gaussian random fields, because sharp contrasts in $\log K$ could not be resolved by the multiGaussian approach.

The synthetic test case confirmed the results of the real-world case study,

635 as for 16 observations no advantage was observed by including geophysical information and multiple point geostatistics with normal score transformation. However, the synthetic scenarios also showed that for a larger number of head data a clear positive impact of geophysical conditioning may be expected. On the other hand, even for many head data, MPG in combination with NS-EnKF

640 did not show a clear advantage over a conventional approach. This could be partly explained also by the effect of boundary conditions, which in our test case are more realistic than those typically used in previous synthetic studies. For example, Zhou et al. (2011), Xu et al. (2013), and Schöniger et al. (2012), who showed a better performance of NS-EnKF compared to EnKF, worked with

645 prescribed flux boundary conditions for at least part of the boundaries or included pumping tests that stressed most of the modelling domain. This is likely to have a significant impact on the pressure head distribution and makes the groundwater level measurements more informative with respect to the hydraulic conductivity distribution, helping to identify also non-multi-Gaussian hydraulic

650 conductivity fields.

In summary, our results suggest that in practical problems it is difficult to obtain an improved characterization with the help of advanced stochastic approaches, which probably require a large number of data. Additional data types (e.g., temperature, electrical conductivity), combined with hydraulic head

655 data, might help to better delineate the heterogeneous hydraulic conductivity distribution and states in complex aquifers such as the one investigated here.

24

6. Summary and conclusions

We applied an inversion approach based on the normal score transform ensemble Kalman filter to a real-world test case in Northern Italy, where continuous measurements of water table level in an unconfined aquifer were assimilated. Using multiple point geostatistics with a training image derived from satellite data and conditioning the initial realizations to electrical resistivity tomography transects, the proposed approach succeeded to estimate a posterior probability distribution function of the log-transformed hydraulic conductivity (K) field that minimized the error in the groundwater level predictions, and also provided an estimate of the related uncertainty. However, the proposed inversion approach could not reconstruct the characteristic paleo-channel shapes, due to the relatively small number of observation points with respect to the extension of the study area.

This was confirmed in a subsequent synthetic case study, which showed that the $\log K$ distribution could be retrieved more accurately with more observation points (65 and 250, compared to 16). This suggests that there are practical limitations on the use of advanced stochastic methods like those used in this study. Although there is evidence for the presence of channelized structures in the aquifer at this site, not enough information is available to constrain the prior multiple point geostatistical model and the position of the paleo-channels. Therefore, the inversions relying on multiple point geostatistics did not outperform inversions with the simpler multiGaussian model, even though information from several geophysical investigations was included in the inverse conditioning.

Similar issues might be encountered in many other hydrogeological sites and our results highlight that advanced stochastic methods can still be very useful, but they should be carefully chosen according to the level of residual uncertainty that practitioners are willing to accept. Such uncertainty can be drastically reduced only if abundant and high-quality subsurface data are available, which obviously imply costly experimental investigations. Whether or not it is worth the effort of course depends on the specific problem at hand, its scale, and the

25

final objective, but, in general, hydrogeologists should stress to stakeholders that uncertainty must always be dealt with and correct deterministic solutions are not possible in practice.

Acknowledgments

References

Aanonsen, S., Nævdal, G., Oliver, D., Reynolds, A., Vallès, B., 2009. The ensemble Kalman filter in reservoir engineering–a review. SPE Journal 14. doi:10.2118/117274-PA.

Apte, A., Hairer, M., Stuart, A.M., Voss, J., 2007. Sampling the posterior: An approach to non-gaussian data assimilation. Physica D: Nonlinear Phenomena 230, 50–64.

Bárdossy, A., 2006. Copula-based geostatistical models for groundwater quality parameters. Water Resources Research 42, W11416. doi:10.1029/2005WR004754. w11416.

Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copulas. Water Resources Research 44, W07412. doi:10.1029/2007WR006115. w07412.

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

Caers, J., Zhang, T., 2005. Multiple-point geostatistics: A quantitative vehicle for integrating geologic analogs into multiple reservoir models. AAPG Memoir , 383–394.

Camporese, M., Cassiani, G., Deiana, R., Salandin, P., 2011. Assessment of local hydraulic properties from electrical resistivity tomography monitoring of a three-dimensional synthetic tracer test experiment. Water Resources Research 47. doi:10.1029/2011WR010528.

Camporese, M., Cassiani, G., Deiana, R., Salandin, P., Binley, A., 2015. Coupled and uncoupled hydrogeophysical inversions using ensemble Kalman filter assimilation of ERT-monitored tracer test data. Water Resources Research 51, 3277–3291. doi:10.1002/2014WR016017.

Chang, H., Zhang, D., Lu, Z., 2010. History matching of facies distribution with the EnKF and level set parameterization. Journal of Computational Physics 229, 8011–8030.

Chen, R., Liu, J.S., 2000. Mixture Kalman filters. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62, 493–508.

Chen, Y., Oliver, D.S., Zhang, D., 2009. Data assimilation for nonlinear problems by ensemble Kalman filter with reparameterization. Journal of Petroleum Science and Engineering 66, 1–14. doi:10.1016/j.petrol.2008.12.002.

Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Advances in Water Resources 29, 1107–1122. doi:10.1016/j.advwatres.2005.09.007.

Crestani, E., Camporese, M., Baú, D., Salandin, P., 2013. Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. Hydrology and Earth System Sciences 17, 1517–1531. doi:10.5194/hess-17-1517-2013.

Crestani, E., Camporese, M., Salandin, P., 2015. Assessment of hydraulic conductivity distributions through assimilation of travel time data from ERT-monitored tracer tests. Advances in Water Resources 84, 23–36. doi:10.1016/j.advwatres.2015.07.022.

27

De Marsily, G., 1986. Quantitative Hydrogeology: Groundwater Hydrology for Engineers. 1 ed., Academic Press.

Deutsch, C., Journel, A., 1998. Geostatistical software library and users guide (GSLIB) .

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans 99, 10143—10162.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 53, 343–367. doi:10.1007/s10236-003-0036-9.

Evensen, G., 2009. The ensemble Kalman filter for combined state and parameter estimation. IEEE Contr. Syst. Mag. 29, 83–104. doi:10.1109/MCS.2009.932223.

Feyen, L., Caers, J., 2005. Multiple-point geostatistics: a powerful tool to improve groundwater flow and transport predictions in multi-modal formations, in: Geostatistics for Environmental Applications. Springer, p. 197–208.

Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of multiGaussian fields, in: Geostatistics Troia'92. Springer, 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, 47–61. doi:10.1016/S0309-1708(96)00031-0.

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

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review 129, 2776—2790. doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

28

Hendricks Franssen, H.J., Kaiser, H., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., Kinzelbach, W., 2011. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-

770   aquifer interaction and parameter updating. Water Resources Research 47. doi:10.1029/2010WR009480.

Hendricks Franssen, H.J., 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.

775 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, 261–274.

Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review 126, 796–811.

780   doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

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

Huysmans, M., Dassargues, A., 2009. Application of multiple-point

785   geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). Hydrogeology Journal 17, 1901–1911. doi:10.1007/s10040-009-0495-2.

Jafarpour, B., McLaughlin, D., 2009. Estimating channelized-reservoir permeabilities with the ensemble Kalman filter: The importance of ensemble design.

790   SPE Journal 14. doi:10.2118/108941-PA.

Jafarpour, B., Tarrahi, M., 2011. Assessing the performance of the ensemble Kalman filter for subsurface flow data integration under variogram uncertainty. Water Resources Research 47.

29

Kurtz, W., Hendricks Franssen, H.J., Kaiser, H.P., Vereecken, H., 2014. Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. Water Resources Research 50, 1665–1688. doi:10.1002/2013WR014823.

Kurtz, W., Hendricks Franssen, H.J., Vereecken, H., 2012. Identification of time-variant river bed properties with the ensemble Kalman filter. Water Resources Research 48. doi:10.1029/2011WR011743.

Leopold, L.B., 1995. Fluvial processes in geomorphology. Courier Dover Publications.

Leopold, L.B., Wolman, M.G., Wolman, M.G., Wolman, M.G., 1957. River channel patterns: braided, meandering, and straight. US Government Printing Office Washington, DC.

Li, G., Reynolds, A., 2007. An iterative ensemble Kalman filter for data assimilation, in: SPE Annual Technical Conference and Exhibition.

Li, L., Zhou, H., Franssen, H.J.H., Gómez-Hernández, J.J., 2012. Groundwater flow inverse modeling in non-MultiGaussian media: performance assessment of the normal-score ensemble Kalman filter. Hydrology and Earth System Sciences 16, 573–590. doi:10.5194/hess-16-573-2012.

Liu, B., Gharamti, M., Hoteit, I., 2016. Assessing clustering strategies for Gaussian mixture filtering a subsurface contaminant model. Journal of Hydrology 535, 1–21. doi:10.1016/j.jhydrol.2016.01.048.

Liu, Y., 2006. Using the SNESIM program for multiple-point statistical simulation. Computers and Geosciences 32, 1544–1563.

Nan, T., Wu, J., 2011. Groundwater parameter estimation using the ensemble Kalman filter with localization. Hydrogeology Journal 19, 547—561. doi:10.1007/s10040-010-0679-9.

30

Nowak, W., 2009. Best unbiased ensemble linearization and the quasi-linear kalman ensemble generator. Water Resources Research 45, n/a–n/a. doi:10.1029/2008WR007328. w04431.

Panzeri, M., Riva, M., Guadagnini, A., Neuman, S., 2015. EnKF coupled with groundwater flow moment equations applied to Lauswiesen aquifer, Germany. Journal of Hydrology 521, 205–216. doi:10.1016/j.jhydrol.2014.11.057.

Pasetto, D., Camporese, M., Putti, M., 2012. Ensemble kalman filter versus particle filter for a physically-based coupled surface–subsurface model. Advances in Water Resources 47, 1 – 13. doi:10.1016/j.advwatres.2012.06.009.

Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics with SGeMS: A User's Guide. Cambridge University Press.

Reynolds, A.C., Zafari, M., Li, G., 2006. Iterative forms of the ensemble Kalman filter, in: 10th European Conference on the Mathematics of Oil Recovery.

Rings, J., Huisman, J.A., Vereecken, H., 2010. Coupled hydrogeophysical parameter estimation using a sequential bayesian approach. Hydrology and Earth System Sciences 14, 545–556. doi:10.5194/hess-14-545-2010.

Sakov, P., Oliver, D.S., Bertino, L., 2012. An iterative EnKF for strongly nonlinear systems. Monthly Weather Review 140, 1988—2004.

Sartor, L., 2006. Studio idrogeologico propedeutico al piano di caratterizzazione del sito ex discarica Saltarel. Relazione geologica. Amministrazione Comunale di Valdobbiadene. Valdobbiadene, TV.

Schneider, C.L., Attinger, S., 2008. Beyond thiem: A new method for interpreting large scale pumping tests in heterogeneous aquifers. Water Resources Research 44. doi:10.1029/2007WR005898.

Schöniger, A., Illman, W.A., Wöhling, T., Nowak, W., 2015. Finding the right balance between groundwater model complexity and experimental effort via bayesian model selection. Journal of Hydrology 531, Part 1, 96 – 110. doi:10.1016/j.jhydrol.2015.07.047.

31

Schöniger, A., Nowak, W., Hendricks Franssen, H.J., 2012. Parameter estima-
tion by ensemble Kalman filters with transformed data: Approach and ap-
plication to hydraulic tomography. Water Resources Research 48, n/a–n/a.
doi:10.1029/2011WR010462.

Strebelle, S., 2002. Conditional simulation of complex geological struc-
tures using multiple-point statistics. Mathematical Geology 34, 1–21.
doi:10.1023/A:1014009426274.

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

Wang, X., Bishop, C.H., 2003. A comparison of breed-
ing and ensemble transform Kalman filter ensemble forecast
schemes. Journal of the Atmospheric Sciences 60, 1140—1158.
doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

Xu, T., Gómez-Hernández, J.J., 2016. Characterization of non-Gaussian
conductivities and porosities with hydraulic heads, solute concentrations,
and water temperatures. Water Resources Research 52, 6111–6136.
doi:10.1002/2016WR019011.

Xu, T., Gómez-Hernández, J.J., Zhou, H., Li, L., 2013. 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 hy-
draulic conductivity field. Advances in Water Resources 54, 100–118.

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, 844–864.

Zhou, H., Li, L., Gómez-Hernández, J.J., 2012. Characterizing curvilinear fea-
tures using the localized Normal-Score ensemble Kalman filter. Abstract and
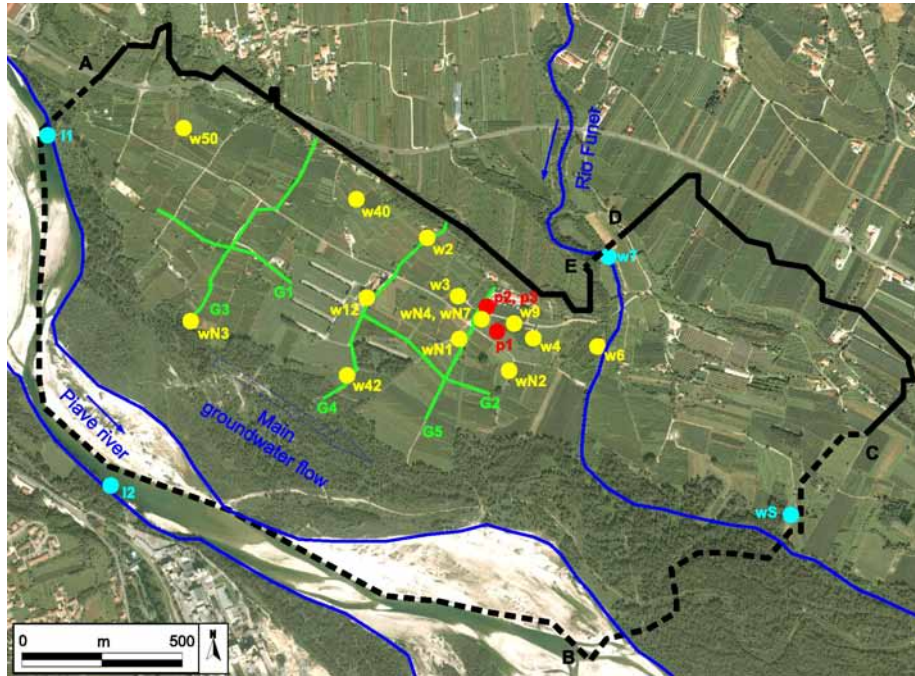Applied Analysis 2012, 1–18. doi:10.1155/2012/805707.

32

Figure 1: Map of the Settolo study area. The black lines denote the contour of the modeled domain and the boundary conditions: solid for no flow (Neumann) and dashed for prescribed head (Dirichlet). Prefixes "w", "p", "I", and "G" denote observation wells, pumping wells, river stage stations, and electrical resistivity tomographies, respectively.

Zovi, F., 2014. Assessment of heterogeneous hydraulic properties in natural aquifers at the intermediate scale. Ph.D. thesis. University of Padova. Padova, Italy.

Table 1: Geostatistical parameters used to generate the initial ensembles of $\log K$ random fields. For each realization, parameters were randomly extracted from a uniform distribution whose support is defined by the "min" and "max" values. Angles are indicated with respect to the main flow direction.

	min	max
Ensembles 1 and 2: multiple point geostatistics + Gaussian random field		
SNESIM parameters for facies generation		
P_x and P_y	0.85	1.15
α_{TI} [°]	-15	+15
$\%_{ch}$	40%	60%
SGSIM parameters for zone 1 (channels)		
$\langle Y \rangle$ [log(m/s)]	$\log(2 \times 10^{-2})$	$\log(5 \times 10^{-2})$
σ_Y^2 [log(m/s)^2)]	0.1	0.2
I_x [m]	85	160
I_y [m]	45	85
α_{xy} [°]	-10	+10
SGSIM parameters for zone 2 (matrix)		
$\langle Y \rangle$ [log(m/s)]	$\log(5 \times 10^{-4})$	$\log(2 \times 10^{-3})$
σ_Y^2 [log(m/s)^2)]	0.2	0.5
I_x [m]	85	160
I_y [m]	45	85
α_{xy} [°]	-10	+10
Ensemble 3: Gaussian random field		
SGSIM parameters		
$\langle Y \rangle$ [log(m/s)]	$\log(2 \times 10^{-3})$	$\log(2 \times 10^{-2})$
σ_Y^2 [log(m/s)^2)]	0.25	0.75
I_x [m]	85	160
I_y [m]	45	85
α_{xy} [°]	-10	+10
Ensembles 1, 2, and 3		
S_y [-]	0.10	0.35
S_s [m^{-1}]	10^{-5}	10^{-1}

Table 2: Description and performance statistics of the six scenarios for the real test case. "Ens 1" indicates the initial ensemble generated with multiple point geostatistics (MPG) and geophysical conditioning (GC). "Ens 2" indicates the initial ensemble generated with MPG but no GC. "Ens 3" indicates the initial ensemble generated with sequential Gaussian simulation only (no MPG). "OL" indicates open loop runs (without data assimilation). The values of ensemble spread, mean K, S_y, and S_s reported in the table refer to the last update. Bold numbers indicate the best performing experiment.

Scenario	Initial ensemble	Normal score	$RMSE_{wl}$ [m]	$RMSE_{wl}^{ens}$ [m]	ES_Y [log(m/s)^2]	Avg. K [m/s]	S_y [-]	S_s [m^{-1}]
OL	Ens 1	N/A	0.400	0.720	1.73	7.5×10^{-3}	0.28	5.0×10^{-3}
OL	Ens 2	N/A	0.407	0.691	1.78	8.0×10^{-3}	0.28	5.0×10^{-3}
OL	Ens 3	N/A	0.396	0.654	0.96	6.3×10^{-3}	0.28	5.0×10^{-3}
1	Ens 1	N	0.238	0.256	1.53	1.0×10^{-2}	0.26	7.6×10^{-3}
2	Ens 2	N	0.244	0.306	1.59	1.4×10^{-2}	0.38	9.0×10^{-3}
3	Ens 3	N	**0.155**	**0.205**	0.55	9.5×10^{-3}	0.20	4.1×10^{-3}
4	Ens 1	Y	0.203	0.310	1.57	5.2×10^{-3}	0.10	1.7×10^{-3}
5	Ens 2	Y	0.179	0.269	1.62	5.3×10^{-3}	0.13	1.2×10^{-3}
6	Ens 3	Y	0.302	0.496	0.56	4.5×10^{-3}	0.21	3.2×10^{-3}

Table 3: Description and performance statistics of the fourteen scenarios for the synthetic test case. Bold numbers indicate the best performance for each group of numerical experiments. The values of mean absolute errors, facies matching ratio, and ensemble spread reported in the table refer to the last update.

Scenario	Observations	Normal score	Geoph. cond	MAE_Y^{ens} [log(m/s)]	MAE_Y [log(m/s)]	$RMSE_{WL}^{ens}$ [m]	$RMSE_{WL}$ [m]	$FMR_\%$	ES_Y [log(m/s)^2]
OL	N/A	N/A	Y	1.59	1.31	0.498	0.196	66	1.48
OL	N/A	N/A	N	1.71	1.44	0.582	0.292	50	1.55
1	16	N	Y	**1.43**	1.30	0.378	0.118	**59**	0.87
2	16	N	N	1.51	1.41	**0.310**	**0.093**	54	0.83
3	16	Y	Y	1.49	**1.25**	0.366	0.217	58	1.34
4	16	Y	N	1.64	1.39	0.362	0.194	52	1.44
5	65	N	Y	1.21	1.20	0.183	**0.008**	**69**	0.22
6	65	N	N	1.38	1.37	**0.177**	0.009	62	0.21
7	65	Y	Y	**1.19**	1.18	0.183	0.011	67	0.29
8	65	Y	N	1.34	1.34	0.187	0.012	59	0.27
9	65 w/loc.	Y	Y	1.34	**1.07**	0.215	0.241	64	1.29
10	65 w/infl.	Y	Y	1.29	1.28	0.182	0.016	63	0.34
11	250	N	Y	1.14	1.11	0.106	**0.036**	**74**	0.36
12	250	N	N	1.17	1.13	**0.093**	0.098	71	0.35
13	250	Y	Y	**1.09**	**1.06**	0.113	0.056	71	0.52
14	250	Y	N	1.14	1.10	0.103	0.085	68	0.58
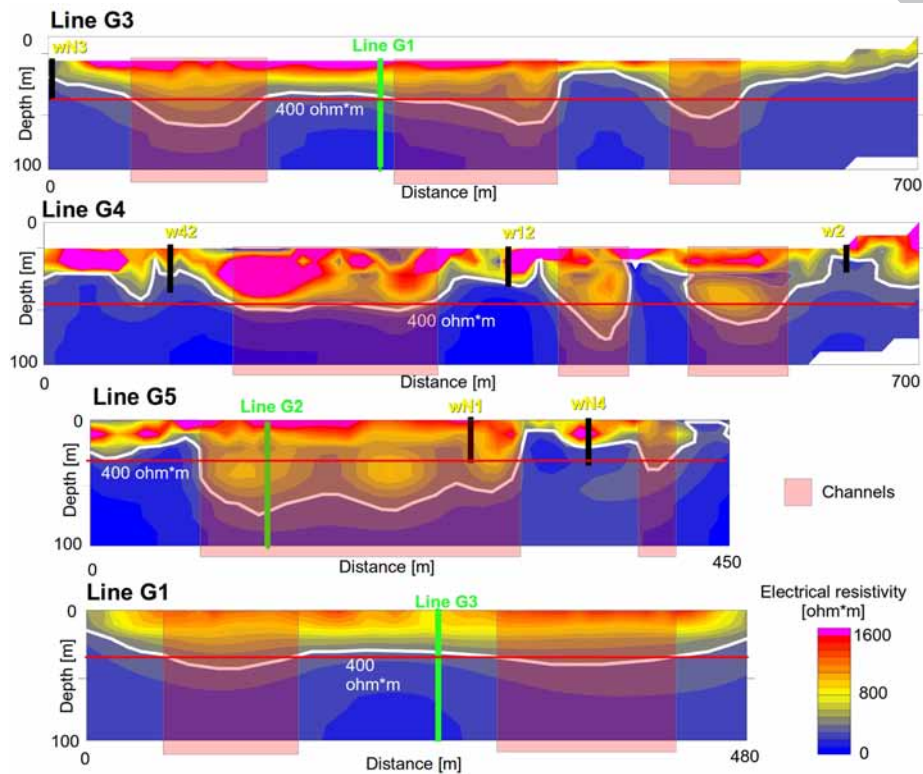
Figure 2: Electrical resistivity tomography transects. The location of the transects is reported in Figure 1. Orientation is from southwest to northeast for lines G3, G4, and G5, from northwest to southeast for line G1. Intersections with other tomographies are shown in green, while black segments represent wells. The 400 Ωm white line indicates the assumed impermeable bedrock.
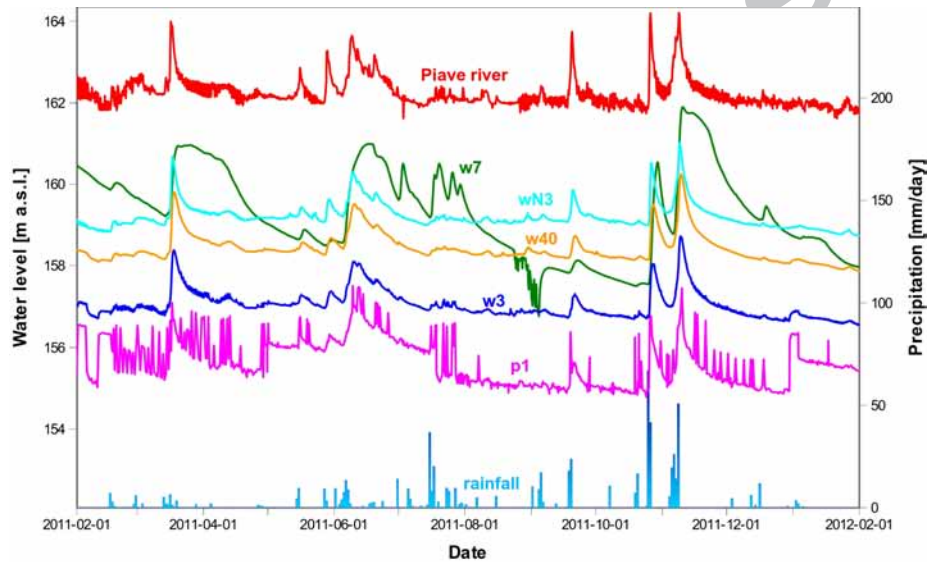
Figure 3: Example of groundwater level data assimilated with EnKF in the real-world inversion runs, together with the main hydrological drivers of groundwater dynamics, i.e., the Piave River stage and rainfall rate. For clarity, only data from a selected number of observation wells is shown.
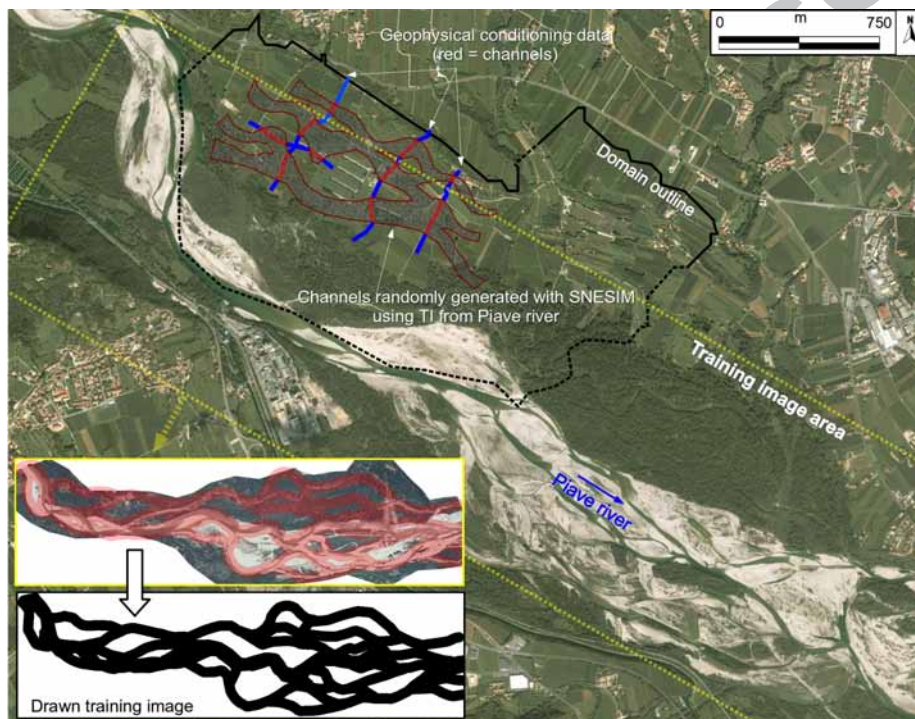
Figure 4: Training image extracted from satellite data and used as input to the SNESIM algorithm for the generation of random facies distributions honoring geophysical data.
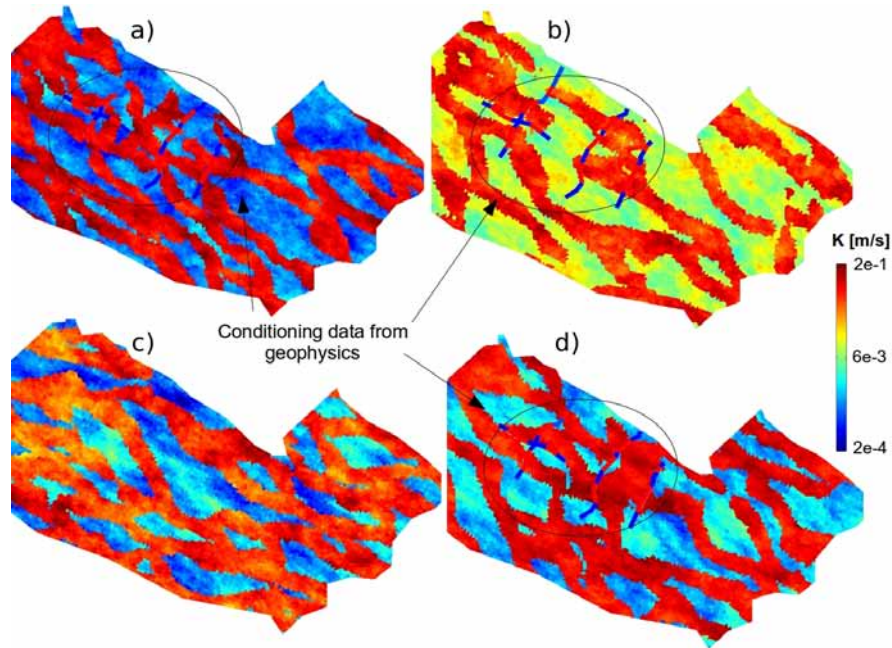
Figure 5: Example of four realizations of $\log K$ fields generated with MPG and conditioned to geophysical data (ensemble 1), except for the lower left field, which was generated without conditioning (ensemble 2). The four realizations were generated with different sets of geostatistical parameters.
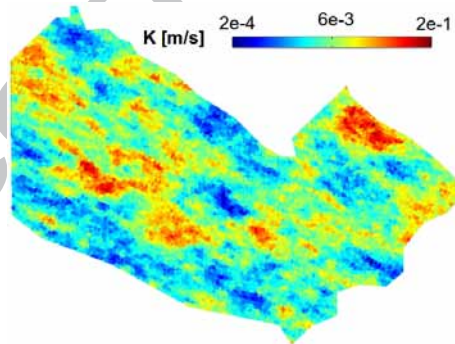
Figure 6: Realization of a $\log K$ gaussian random field (ensemble 3) generated with SGSIM.

40
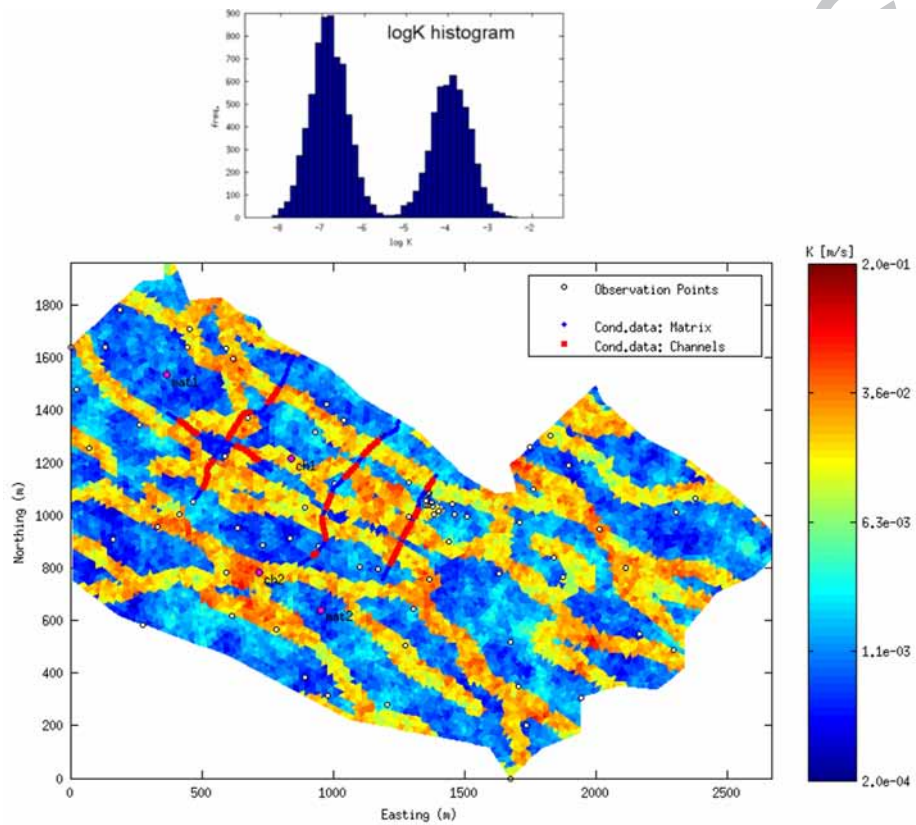
Figure 7: Reference log$K$ field for the synthetic test case, generated with the same parameters as ensemble 1 (Table 1), and corresponding probability distribution function.
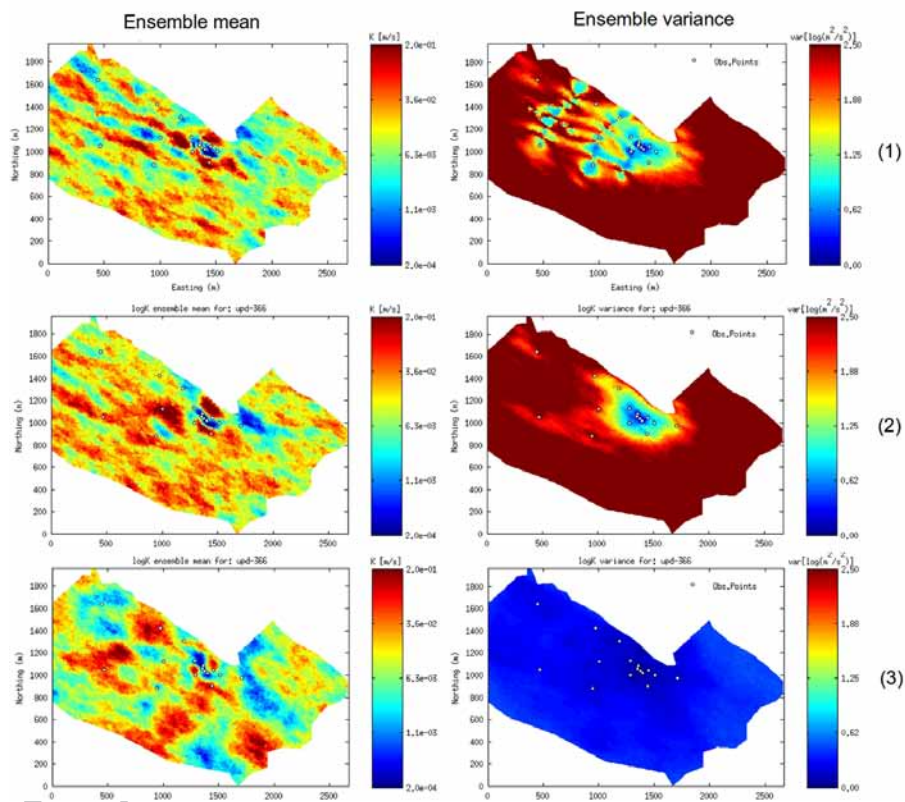
Figure 8: Ensemble mean (EM_Y) and variance (EV_Y) of the log$K$ distribution as computed at the last update for scenarios 1 (top), 2 (middle), and 3 (without normal score transform, bottom) of the real-world test case (Table 2).
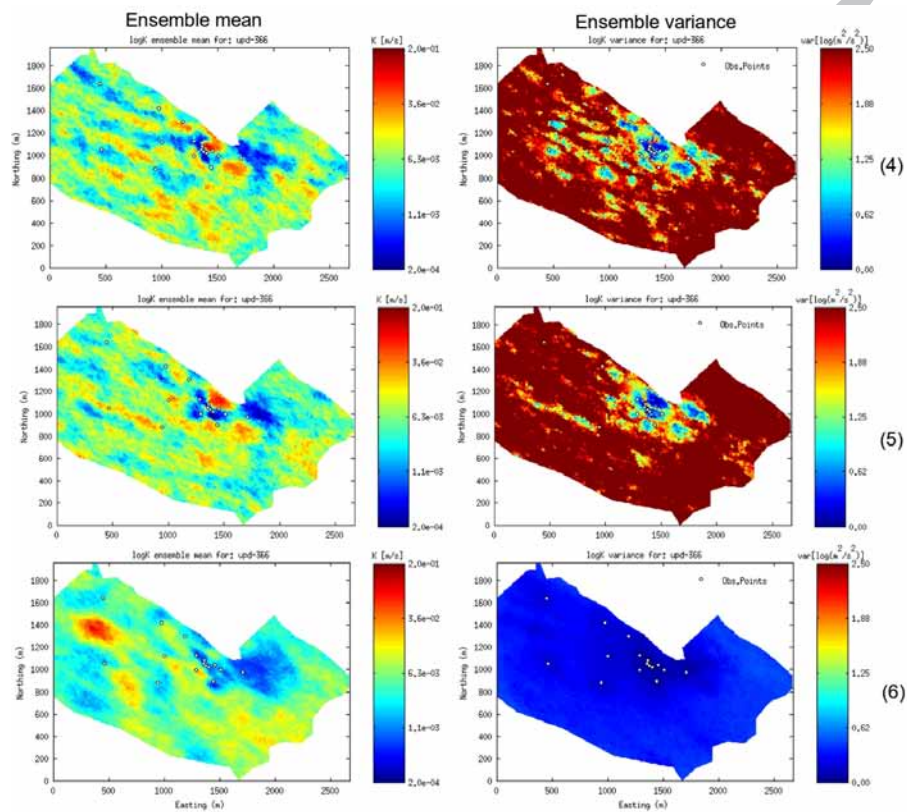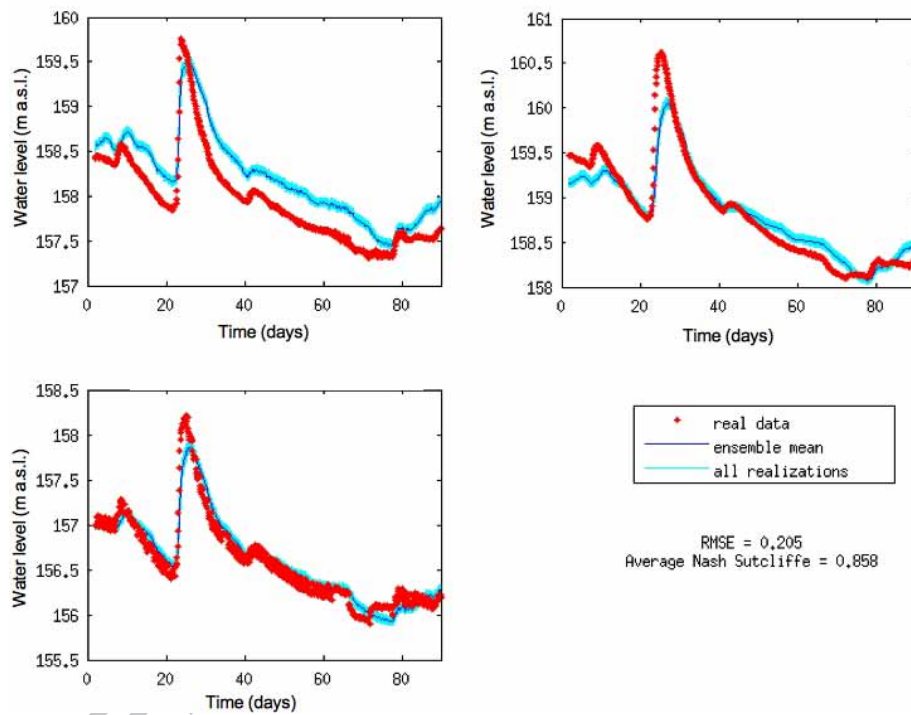
Figure 9: Ensemble mean (EM_Y) and variance (EV_Y) of the log$K$ distribution as computed at the last update for scenarios 4 (top), 5 (middle), and 6 (with normal score transform, bottom) of the real-world test case (Table 2).

Figure 10: Comparison between measured and simulated groundwater levels in three observation wells (w42, w40, and wN2, clockwise from top left, respectively) for the verification period. The simulations were run using the ensemble of K distributions derived from scenario 3 of the real-world test case (see Table 2).
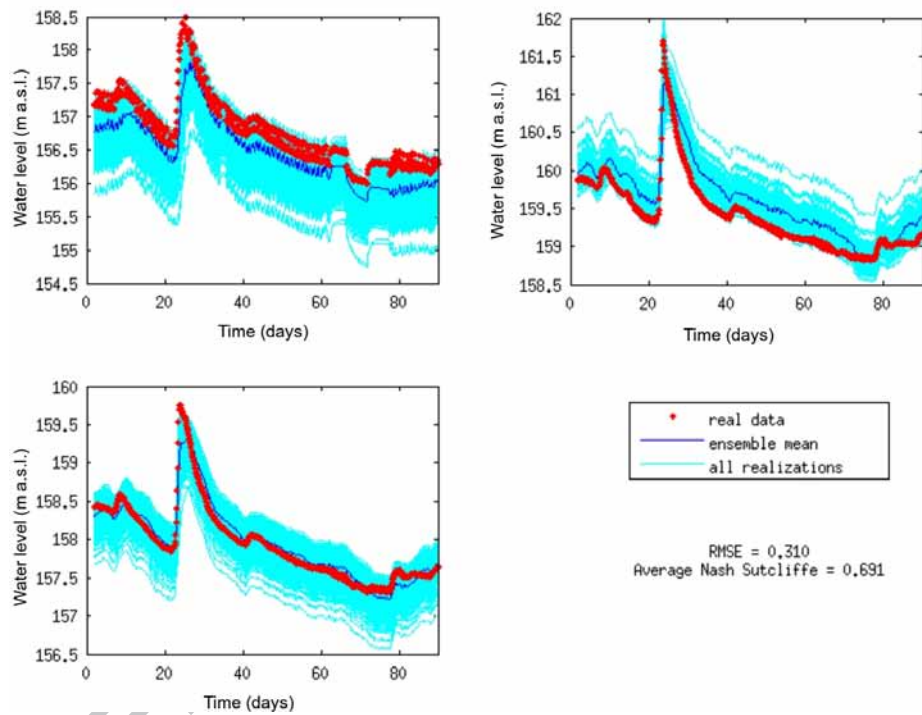
Figure 11: Comparison between measured and simulated groundwater levels in three observation wells (w9, wN3, and w42, clockwise from top left, respectively) for the verification period. The simulations were run using the ensemble of K distributions derived from scenario 4 of the real-world test case (see Table 2).
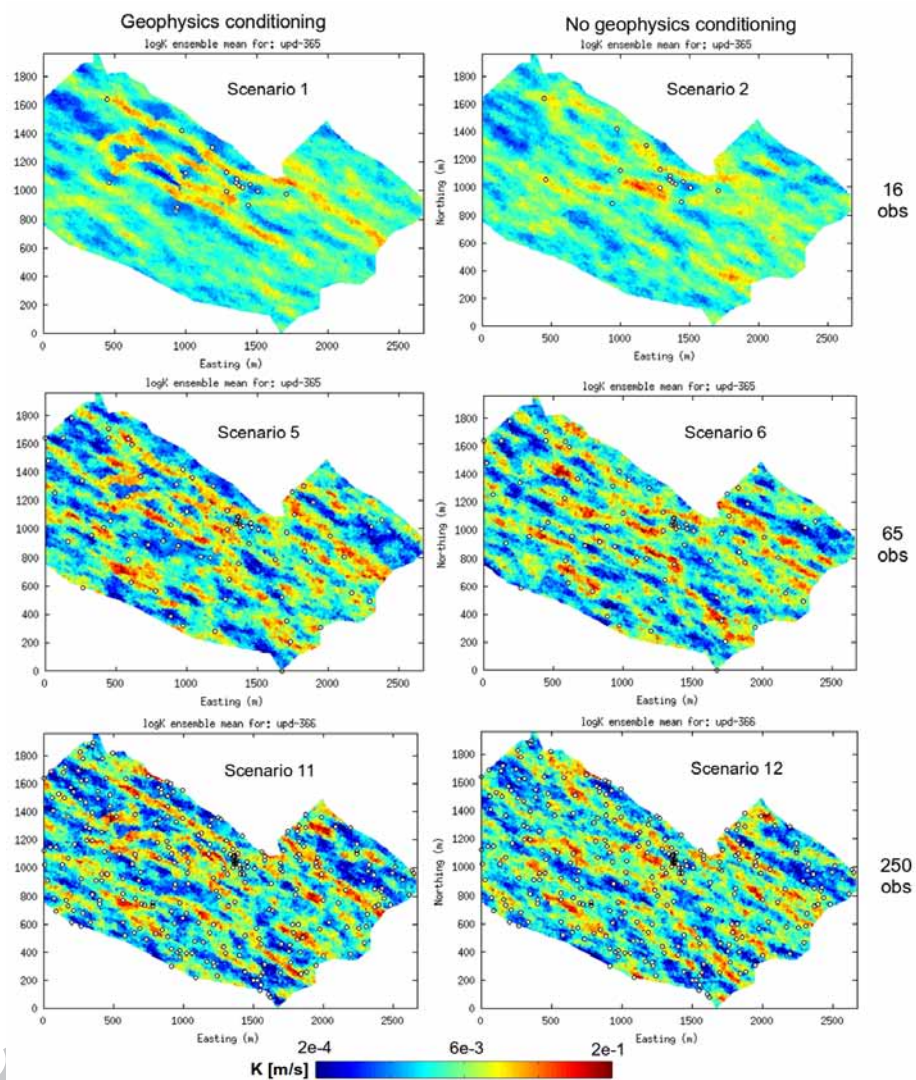
45

Figure 12: Ensemble mean of the $\log K$ distributions resulting from the last update of the synthetic test scenarios without normal score transform, for both the cases with and without conditioning on geophysical data and for different sets of observations (white dots in the figure).
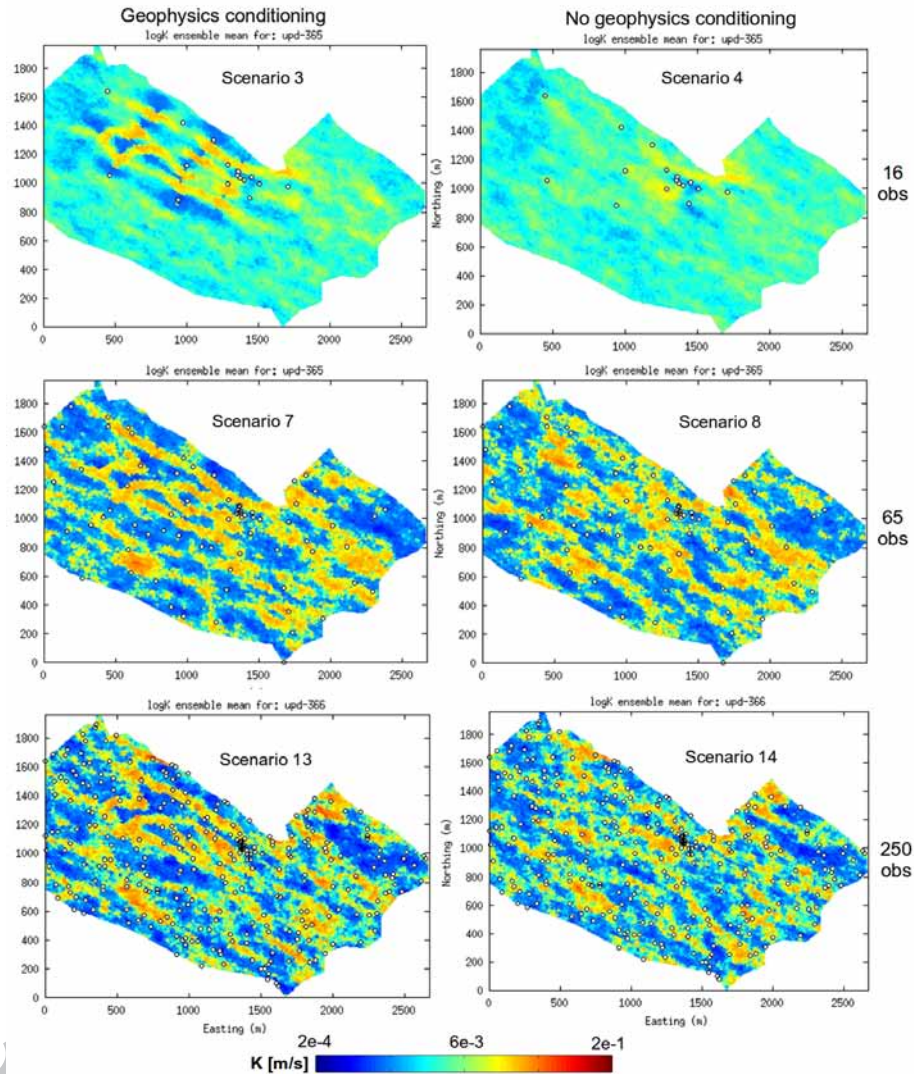
46

Figure 13: Ensemble mean of the log$K$ distributions resulting from the last update of the synthetic test scenarios with normal score transform, for both the cases with and without conditioning on geophysical data and for different sets of observations (white dots in the figure).

47

- Hydraulic head data are assimilated in a real aquifer with ensemble Kalman filter
- Multiple point geostatistics (MPG) and normal score transform used with ERT data
- Data assimilation provides good estimate of heterogeneous hydraulic conductivity
- Multiple point geostatistics and conditioning to ERT data do not improve results
- An additional synthetic study shows more data are needed for MPG to be effective