



Geofísica Internacional

ISSN: 0016-7169

arostan@geofisica.unam.mx

Universidad Nacional Autónoma de
México
México

Vázquez-Guillén, F.; Auvinet, Guichard
Hydraulic conductivities identification via Ensemble Kalman Filtering with transformed
data considering the risk of systematic bias
Geofísica Internacional, vol. 56, núm. 4, octubre-noviembre, 2017, pp. 317-333
Universidad Nacional Autónoma de México
Distrito Federal, México

Available in: <http://www.redalyc.org/articulo.oa?id=56853159001>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

Hydraulic conductivities identification via Ensemble Kalman Filtering with transformed data considering the risk of systematic bias

F. Vázquez-Guillén*, Guichard Auvinet

Received: February 02, 2016; accepted: September 12, 2017; published on line: October 01, 2017

Resumen

En hidrología subterránea, el conjunto de filtros Kalman (EnKF) se acopla con modelos del flujo y transporte de agua subterránea para resolver el problema inverso. Se han propuesto varias extensiones del EnKF para mejorar su desempeño al tratar con campos aleatorios no multi-gaussianos de la conductividad hidráulica. Una de esas variantes es el EnKF con transformación de datos (tEnKF), el cual utiliza la anamorfosis gaussiana dentro de una etapa de condicionamiento. Aunque esta transformación se ha utilizado en el pasado para identificar conductividades hidráulicas, estudios previos han ignorado el riesgo de introducir un sesgo sistemático en la evolución espaciotemporal del campo de la carga hidráulica durante las etapas de pronóstico que las etapas de condicionamiento podrían no corregir conforme transcurre el tiempo. Este artículo propone que la aplicación del tEnKF en medios porosos aleatorios generados sintéticamente debe tener en cuenta este riesgo incorporando en el conocimiento *a priori* una estructura de correlación multi-gaussiana para las conductividades y adoptando un campo de referencia con estructura de correlación asimétrica. Como un ejemplo de esta aplicación, en este artículo se identifican conductividades hidráulicas utilizando el tEnKF resolviendo un problema de flujo monofásico, unidimensional, en un medio poroso aleatorio continuo. Se utilizan conceptos comunes en geoestadística para explicar las hipótesis en las que se basan el EnKF y el tEnKF y también para establecer un vínculo claro entre el tEnKF y la simulación estocástica de campos aleatorios condicionales.

Key words: Simulación estocástica, campos aleatorios condicionales, anamorfosis gaussiana, problema inverso, campos aleatorios no multi-gaussianos.

Abstract

In subsurface hydrology, Ensemble Kalman Filtering (EnKF) has been coupled with groundwater flow and transport models to solve the inverse problem. Several extensions of the EnKF have been proposed to improve its performance when dealing with non-multi-Gaussian random field models of the hydraulic conductivity. One such variant is the EnKF with transformed data (tEnKF), which uses Gaussian anamorphosis within a conditioning step. Although this transformation has been used in the past to identify hydraulic conductivities, previous studies have ignored the risk of introducing a systematic bias in the spatiotemporal evolution of the hydraulic head field during the forecast steps that the update steps may not correct over time. This paper proposes that in order to evaluate the performance of tEnKFs, applications in synthetically generated random porous media should take into account this risk by incorporating prior knowledge with a multi-Gaussian conductivity correlation structure, and by adopting a reference field with asymmetric correlation structure. As an example of this application, hydraulic conductivities using the tEnKF were identified by solving a one-dimensional, single phase flow problem in a continuous random porous medium. Common concepts in Geostatistics are used to explain the hypothesis underlying both EnKF and tEnKF and to establish a clear link between the tEnKF and the stochastic simulation of conditional random fields.

Palabras clave: Stochastic simulation, conditional random fields, Gaussian anamorphosis, inverse problem, Non multi-Gaussian random fields.

F. Vázquez-Guillén*

G. Auvinet

Coordinación de Geotecnia y Estructuras

Instituto de Ingeniería

Universidad Nacional Autónoma de México

Ciudad Universitaria

Delegación Coyoacán 04510

México CDMX, México.

*Corresponding author: fvazquezg@exii.unam.mx

Introduction

Groundwater modeling as a tool for sustainable development and utilization of groundwater resources in land subsidence studies requires knowledge of the hydrogeologic properties of the medium. A number of studies have shown that parameters such as hydraulic conductivity and porosity of porous media exhibit spatial variations. Most of the times, these variabilities can only be determined at few and sparse locations. Thus, there is uncertainty in the characterization of the spatial variability of those properties. Owing to this uncertainty, predictions conducted with groundwater flow models are also uncertain. To make plausible predictions and reduce the associated uncertainty, stochastic inverse modeling techniques are frequently applied. These techniques allow the identification of parameters and the quantification of parameter and prediction uncertainty consistent with the measured data and the groundwater flow model.

In subsurface hydrology, hydraulic conductivity fields are often characterized assuming that its spatial variability can be interpreted as a realization of a random field model (Dagan, 1989; Gelhar, 1993). The parameterization of the random field is obtained through the observations of the realization itself assuming the field is ergodic (Deutsch and Journel, 1992; Chilés and Delfiner, 1999). Then, inferences at unobserved locations and the uncertainty of these inferences are determined on the basis of the direct observations by either estimation or simulation techniques (e.g. Journel and Huijbregts, 1978; Chilés and Delfiner, 1999). While estimation techniques provide one single “best” estimate of the hydraulic conductivity field, simulation techniques yield multiple realizations of that field (Journel and Huijbregts, 1978). In order to reduce the uncertainty in the inference of the hydraulic conductivity field, indirect observations of it, e.g., of hydraulic heads, are also taken into account in these tasks by solving the typical inverse problem of hydrogeology (Chilés and Delfiner, 1999). Other informative variables such as flow rates and species concentrations may also be incorporated in the inference process to further constrain the spatial fluctuations of the conductivities. Although solutions to the inverse problem lead to hydraulic conductivity fields which are compatible with the measured hydraulic head data, it is recognized that there are an infinite number of other conductivity fields which may also match the same hydraulic head data (RamaRao *et al.*, 1995). The inverse problem

is therefore ill-posed and a unique, exact solution is generally not available. Instead, a solution which coincides with the observations is commonly sought (Tarantola, 2005).

The inverse problem can be solved under steady conditions or under transient conditions using estimation or simulation techniques (Gómez-Hernández and Wen, 1994; Chilés and Delfiner, 1999). A comprehensive review of the evolution of several methods for solving the stochastic inverse problem in hydrogeology has been presented elsewhere (Zhou *et al.*, 2014). Among these approaches, simulation-based inversion techniques are often preferred over estimation-based inversion techniques because it has been proven that the single “best” estimate provided by the latter does not capture the range of variability of real fields of conductivities. As a result, flow and transport predictions conducted in these fields are very poor (Gómez-Hernández and Wen, 1994). The most widely accepted simulation-based inversion techniques in groundwater modeling use the Monte Carlo (MC) method. Within the MC framework, the available observations of the state variables are integrated into a prior random field of conductivities through an iterative procedure to obtain a posterior random field of conductivities expressed as a set of conditional simulations. Examples of such kinds of approaches are: the self-calibration method (Sahuquillo *et al.*, 1992; Gómez-Hernández *et al.*, 1997), the pilot-point method (RamaRao *et al.*, 1995), the Markov-Chain MC method (Oliver *et al.*, 1997), the gradual deformation method (Hu, 2000; Capilla and Llopis-Albert, 2009; Hu *et al.*, 2013) and the random mixing method (Bárdossy and Hörning, 2016). One common characteristic to all of these MC-based inversion methods is that they are formulated as optimization problems where the unknown parameter field is represented by the nodes of a mesh. Thus, the use of dimensional reduction techniques is indispensable in large dimensional problems. The main difference among the approaches is the way the optimization problem is solved.

An alternative MC-based inversion technique that can be used to integrate available observations of the hydraulic head into a prior random field model of conductivities is the Ensemble Kalman Filter (EnKF). In this approach, the observations are integrated sequentially in time using the groundwater model itself to evolve the hydraulic head field in a physically plausible manner (Katzfuss *et al.*, 2016). The EnKF was developed by Evensen (1994) as an extension of the Kalman Filter (KF) (Kalman, 1960) to deal with non-linear

systems. The main difference between EnKF and KF is that the former uses an ensemble representation for the state variables from which any statistical moment can be calculated whenever it is needed. Instead of using the MC method, earlier extensions of the KF used perturbation theory to handle non-linear dynamics. This is the case of the Extended KF (EKF) (Evensen, 1992; Leng and Yeh, 2003). However, it has been found that the EnKF outperformed the EKF, especially in highly non-linear systems (Miller *et al.*, 1999; Reichle *et al.*, 2002). The EnKF differs from other MC-based filters, such as particle filters, in the process by which the observations are integrated into the prior random field model.

The EnKF procedure consists of two main steps. The first is the forecast step, which uses MC sampling to propagate the uncertainty in the hydraulic conductivity field through the groundwater model to approximate the spatiotemporal evolution of the hydraulic head field at the time the observations are available. The second is the update step that integrates the measured hydraulic head data into the prior random field model of conductivities by conditioning each prior realization to the available observations. The conditioning process is performed only on the basis of the available data at the time of analysis using a linear estimation technique. Thus, only the mean, auto-covariance and cross-covariance functions are used when computing the posterior random field of conductivities. This strategy lends the scheme computational efficiency and makes it suitable for large dimensional problems, yet the forecast step may still be highly demanding in terms of computation. Since the conditional simulations thus obtained are consistent with the system dynamics, predictions of response variables and an investigation of the uncertainty of these predictions can also be conducted simultaneously.

The EnKF only converges to an optimal solution when the random fields involved are multi-Gaussian and when the functional relationship between state and parameter variables is linear. Due to the non-linearity of the groundwater equations, it seems reasonable to expect that state fields will be non multi-Gaussian even in multi-Gaussian parameter fields. Thus, assuming joint multi-Gaussian distributions between conductivities and heads is often not correct in practice. Moreover, in multi-Gaussian random fields the spatial correlation structure is symmetric; i.e., the values at opposite percentiles with respect to the mean present exactly the same spatial

correlation structure (Journel and Deutsch, 1993; Journel and Zhang, 2006). The highest continuity is observed at mean values and the extreme high/low values appear as isolated clusters. As a result, connected paths of extreme values do not occur in multi-Gaussian random fields (Gómez-Hernández and Wen, 1998; Knudby and Carrera, 2005). On the contrary, field evidence suggests that the patterns of spatial variability in natural soil formations differ significantly from such multi-Gaussian dependence characteristics. Journel and Alabert (1989) found stronger spatial correlation structures at low values than at high values in field measurements of air permeability taken on a vertical slab of Berea sandstone; this stronger correlation was even stronger than that imposed by a multi-Gaussian random field model. Asymmetric correlation structures have also been found in conductivity fields with relatively small heterogeneity (Haslauer *et al.*, 2012). Thus, natural porous media seem to be non multi-Gaussian with respect to their spatial distribution of conductivities independently of their degree of heterogeneity.

Several extensions of the EnKF have been examined in the last decade in order to achieve a more versatile tool capable of handling non multi-Gaussian random field models in parameter identification problems. Sun *et al.* (2009) reformulated the update step of the EnKF using Gaussian mixture models and clustering techniques. Sarma and Chen (2009) developed a generalization of the EnKF based on kernel principal component analysis. Emerick (2017) presented an investigation of the performance of different principal component analysis-based approaches. Bertino *et al.* (2003) modified the update step of the EnKF by applying Gaussian transformations. The advantages of this last extension compared to the classical EnKF method were confirmed in several studies (Zhou *et al.* 2011; Li *et al.* 2012; Schöniger *et al.* 2012; Erdal *et al.* 2015). One common characteristic to all of these EnKF-based methods is the use of auto-covariance and cross-covariance functions as unique descriptors of the spatial dependence at each update step. Unlike these approaches, Zhou *et al.* (2012) developed the so-called Ensemble PATtern matching method (EnPAT) based on multiple-point geostatistical simulation techniques which use multiple-point covariance functions rather than traditional two-point covariance functions to determine the spatial correlations. Extensions of this approach were developed later (Li *et al.* 2013, 2014, 2015). Although the EnPAT method outperforms the classical EnKF and eliminates the multi-Gaussian assumption implicit in the update

step of the latter (Li *et al.* 2015), it strongly relies on the concept of training image; i.e. a conceptual model of the geological structure of the formation under study, the construction of which can be problematic, especially in 3D applications, and the benefit of which is more evident in fluvial deposits than in other type of geologic formations.

The EnKF with Gaussian transformations (tEnKF) has been particularly well accepted in several geophysical areas (Bertino *et al.* 2003; Simon and Bertino, 2009, 2012; Béal *et al.* 2010; Zhou *et al.* 2011; Li *et al.* 2012; Schöniger *et al.* 2012; Xu *et al.* 2013; Erdal *et al.* 2015; Zovi *et al.* 2017). In subsurface hydrology, it has been applied to the identification of the hydraulic conductivity of synthetically generated channelized aquifers that display the same asymmetry as the prior random field model (Zhou *et al.* 2011; Li *et al.* 2012; Xu *et al.* 2013). Thus, the domain under study has the same spatial variability characteristics as those of the set where the spatiotemporal evolution of the hydraulic head field is sought. As a result, the performance of tEnKF is promising because the deviations from the reference field are then corrected, up to a certain extent, with the update steps. However, these studies do not take into account the risk of introducing a systematic bias in the spatiotemporal evolution of the hydraulic head field during the forecast steps that the update steps may not correct over time. While they attribute the promising performance of tEnKF to Gaussian transformations, they disregard the implications of the information incorporated in the prior conductivity fields. This information should be taken into account to explain the performance of the tEnKF because the update step remains suboptimal and is performed under the multi-Gaussian model which is unable to capture channelized structures with the covariance function as the sole descriptor of an evolving spatial dependence. The absence of knowledge about higher order moments of the parameter field inevitably implies a risk of introducing such systematic bias.

This paper proposes that in order to evaluate the performance of tEnKFs, applications in synthetically generated random porous media should take into account the risk of systematic bias by incorporating prior knowledge with a multi-Gaussian conductivity correlation structure, and by considering the spatial distribution of the hydraulic conductivity of a reference field as having an asymmetric correlation structure. This view of the problem follows the idea of Kerrou *et al.* (2008), yet

differs significantly from its mathematical framework and scope. Our experimental setting is closer to a common situation found in practice where one has access to a rough approximation to the mean, variance and auto-covariance function of the real field, but the asymmetry of the spatial correlation structure of that field is unknown. As an example of this application, this paper identifies hydraulic conductivities using the tEnKF by solving a one-dimensional, single phase flow problem in a continuous random porous medium. To explain the hypothesis underlying both EnKF and tEnKF and to establish a clear link between the tEnKF and the stochastic simulation of conditional random fields, common concepts in Geostatistics are used. Ultimately, the aim of this paper is to motivate further discussions about the benefit of incorporating transient hydraulic head responses in the identification of hydraulic conductivity fields subject to this kind of constraint and about further potential improvements to the update step of the EnKF to overcome the multi-Gaussian assumption.

Groundwater flow equations

In this section, the dynamic model describing single-phase fluid flow in a one-dimensional, vertical, fully saturated porous medium with spatially variable hydraulic conductivity is analyzed:

$$\frac{\partial}{\partial x} \left[K_s(\mathbf{x}) \frac{\partial H}{\partial x} \right] = S_s \frac{\partial H}{\partial t} \quad (1)$$

subject to initial and boundary conditions

$$H_{t=0} = h_0, H_{\Gamma_D} = h_1 \quad (2)$$

where H is the hydraulic head [L] in the domain Ω , \mathbf{x} is the spatial coordinate ($\mathbf{x} = x_3$ [L], where x_3 represents the vertical coordinate which is positive upward), $K_s(\mathbf{x})$ is the saturated hydraulic conductivity [L/T], S_s is the specific storage [L⁻¹], h_0 represents the initial head and h_1 the prescribed head at Dirichlet boundary Γ_D .

In the present example, specific storage as well as initial and boundary conditions are treated as deterministic constants. An error-free dynamic model is also assumed. Hence, the model prediction is only affected by the uncertainty in $K_s(\mathbf{x})$. To model this uncertainty, a stationary random field model is adopted. In such an approach, $Y(\mathbf{x})$ defines the collection of n continuous scalar random variables of the natural logarithm of the saturated hydraulic conductivity, i.e. $Y(\mathbf{x}) = \ln(K_s(\mathbf{x}))$ indexed

at the spatial locations \mathbf{x} in the domain Ω_x with $\Omega_x \in \mathfrak{R}^1$. Since $K_s(\mathbf{x})$ is a random field, equation (1) becomes a stochastic differential equation and the flow response H becomes a spatiotemporal random field. $K_s(\chi, t)$ is defined as the collection of N continuous scalar random variables of the hydraulic head indexed at the spatial locations χ in the domain Ω_χ with $\Omega_\chi \in \mathfrak{R}^1$ and times $t \in \{0, 1, 2, \dots\}$. Although $Y(\mathbf{x})$ is assumed as stationary, $H(\chi, t)$ will be non-stationary in space and time because the domain is bounded (Zhang, 2002). In the following formulations, an ensemble interpretation of both random fields is applied.

The EnKF method

In this section, an alternative formulation for parameter estimation of the EnKF based on common concepts in Geostatistics is presented.

Forecast step

The EnKF uses MC sampling to approximate $H(\chi, t)$ assuming that it evolves like a first order Markov process (Evensen, 2003), i.e. $P[H(\chi, t) | H(\chi, t-1), H(\chi, t-2), \dots] = P[H(\chi, t) | H(\chi, t-1)]$. Hence, only the most recent past determines the multivariate conditional Cumulative Distribution Function (CDF) of $H(\chi, t)$ given the whole past. This simplified evolution of $H(\chi, t)$ can be written as:

$$H(\chi, t) = \mathfrak{F}(H(\chi, t-1); \theta) \quad (3)$$

where $\mathfrak{F}(\cdot)$ is a forecast operator representing the dynamic model, i.e. the behavior of the state process as time evolves, and θ is a vector of parameters involved in that description. For example, to determine $H(\chi, t=1)$, it is necessary to specify initial and boundary conditions, a prior ensemble of saturated hydraulic conductivity fields, and the specific storage coefficient. The information incorporated in the prior conductivity fields is thus a key issue in the performance of EnKFs. Finally, note that in equation (3) there are no assumptions about the type of CDF of the parameters or about the linearity of the considered dynamic model.

Update step

The update step of the EnKF approximates the univariate conditional CDFs of $Y(\mathbf{x})$ given that N_h observations of $H(\chi, t=1)$ are known, i.e. $F_{Y_i | H_i, \dots, H_{N_h}}(y_i; \mathbf{x}_i | D) = P[Y(\mathbf{x}_i) \leq y_i | H(\chi_\alpha, t=1) = h_\alpha, t=1, \alpha=1, \dots, N_h]$

by conditioning each realization of $Y(\mathbf{x})$ under the hypothesis that the joint multivariate distributions of $Y(\mathbf{x})$ and $H(\chi, t=1)$, as well as the multivariate distributions of $H(\chi, t=1)$ and $Y(\mathbf{x})$, are Gaussian. Thus, the problem is reduced to the stochastic simulation of one conditional random field assuming a multi-Gaussian model. This update step is performed according to:

$$\mathbf{U}_t^u = \mathbf{U}_0 + \mathbf{K}_t[\mathbf{Z}_t - \mathbf{H}_t^f] \quad (4)$$

where $\mathbf{U}_t^u = [\hat{y}(\mathbf{x}_1), \hat{y}(\mathbf{x}_2), \dots, \hat{y}(\mathbf{x}_{N_h})]^T$ is an n -dimensional vector of updated realizations of the log-conductivity at $t=1$, $\mathbf{U}_0 = [y(\mathbf{x}_1), y(\mathbf{x}_2), \dots, y(\mathbf{x}_{N_h})]^T$ is an n -dimensional vector of simulated realizations of the log-conductivity (realizations *a priori*), $\mathbf{Z}_t = [h_{1,t=1}, h_{2,t=1}, \dots, h_{N_h,t=1}]^T$ is the vector of observations with dimension N_h , $\mathbf{H}_t^f = [h(\chi_1, t=1), h(\chi_2, t=2), \dots, h(\chi_{N_h}, t=1)]^T$ is a reduced vector of forecasted states (realizations of the hydraulic head at the locations of the observations) of dimension N_h , and $\mathbf{K}_t = [\lambda_1(\mathbf{x}_1), \lambda_1(\mathbf{x}_2), \dots, \lambda_1(\mathbf{x}_{N_h}); \dots, \lambda_{N_h}(\mathbf{x}_1), \lambda_{N_h}(\mathbf{x}_2), \dots, \lambda_{N_h}(\mathbf{x}_{N_h})]$ is a matrix of dimension $n \times N_h$ called "Kalman gain", with $\lambda_1(\mathbf{x}_\alpha)$ representing the relative importance of the observations in estimating the value of $Y(\mathbf{x}_i)$. These weighting functions are solutions of the following systems of linear equations:

$$\sum_{i=1}^n \sum_{\alpha=1}^{N_h} \lambda_i(\chi_\alpha) C_H(s, \tau) = C_{YH}(s, \tau) \quad (5)$$

where $C_H(s, \tau)$ represents the spatiotemporal auto-covariance functions between hydraulic heads with $s=(\chi_\alpha, \chi_j)$ for $j=1, \dots, N$ and $\tau=(t=1, t=1)$, and $C_{YH}(s, \tau)$ represents the spatiotemporal cross-covariance functions between log-conductivities and hydraulic heads at $s=(\mathbf{x}, \chi_j)$ with $\tau=(t=0, t=1)$. In the EnKF, both covariance functions are determined statistically over the ensemble of realizations of $Y(\mathbf{x})$ and $H(\chi, t=1)$. These covariance functions are therefore empirical, but on average over several realizations, they can be expected to lead to positive definite matrices and may be used directly without modeling. On the other hand, given their statistical origin, they lead to the "filter inbreeding effect", i.e. the underestimation of variance over time after several update steps (Hendricks and Kinzelbach, 2008). As a result, the final updated realizations may look almost identical to each other and are virtually equal to the ensemble mean (Zhou et al., 2011). Xu et al. (2013) showed, through hydrogeology applications, that filter inbreeding can be reduced with

covariance localization and covariance inflation techniques. Alternative strategies for reducing this problem were presented earlier by Hendricks and Kinzelbach (2008). The implementation of these techniques in the present application is beyond the scope of this paper.

In most applications of the EnKF method, the observations are affected by random errors which are characterized by a normal distribution with zero mean and a diagonal covariance function which represents that, at different measurement locations, these errors are also independent. Adding random errors to the observations serves the purpose of increasing the ensemble variance over time (Burgers *et al.*, 1998). Hence, considering noisy observations contributes somehow to the reduction of the filter inbreeding effect. Some authors also find it useful to add random errors to the observations to stabilize the inversion of the covariance functions and avoid small singular values dominating the solution. Other authors address this last problem formally by means of the Tikhonov regularization functional (e.g. Johns and Mandel, 2008; Elsheikh *et al.* 2013). In the present paper, extensive numerical experiments performed by the authors showed that the Gaussian transformations presented in the following section contribute significantly to the stabilization of the inversion of the covariance functions. Thus, adding random errors to the observations for this purpose is not necessary within the Gaussian space. Moreover, systematic errors affecting the observations could also be considered in the update step because instruments may induce biases which introduce fictitious correlations between variables. Furthermore, errors might not be Gaussian. Therefore, error-free measurements were considered in this study. In future investigations, it will be important to address the evaluation of the effect of both types of errors on the spatial distribution of the uncertainty and its interpretation, as well as the quantitative impact of adding white noise to the observations on the filter inbreeding effect.

The EnKF method with transformed data (tEnKF)

There are two additional steps and one modified update step that have to be implemented in the EnKF method with transformed data (tEnKF). These three steps are described in this section using common concepts in Geostatistics. Then, some useful formulas for post-processing the results and recommendations for the numerical implementation of the tEnKF are presented.

Gaussian transformation step

Under the multi-Gaussian hypothesis implicit in the classical EnKF procedure, the univariate conditional CDF of $Y(\mathbf{x})$, given that N_h observations of $H(\chi_1, t=1)$ are known, will be Gaussian with conditional expectation and conditional variance given by the simple Kriging (cokriging) estimates (Journel and Huijbregts, 1978). However, for the Kriging variance to be an unbiased estimate, both $Y(\mathbf{x})$ and $H(\chi_1, t=1)$ have to be zero-mean Gaussian random fields (Shinozuka and Zhang, 1996). Bertino *et al.* (2003) realized this situation and therefore proposed to apply Gaussian transformations, at least locally, to both random fields. The Gaussian transformation maps non-Gaussian distributed random variables into Gaussian random variables (with zero mean and unity variance) according to:

$$\tilde{Y}(\mathbf{x}_i) = \Phi^{-1}(F_Y(y; \mathbf{x}_i)) \quad (6)$$

$$\tilde{H}(\mathbf{x}_i, t=1) = \Phi^{-1}(F_H(h; \mathbf{x}_i; t=1)) \quad (7)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the univariate Gaussian distribution function, and $F_Y(y; \mathbf{x}_i)$ and $F_H(h; \mathbf{x}_i; t=1)$ are the local distribution functions of $Y(\mathbf{x})$ and $H(\chi_1, t=1)$, respectively. In the tEnKF the empirical versions of these distribution functions are used. The empirical distribution function is defined by $F_n(c) = 1/n_r \sum_{j=1}^{n_r} I\{C_j \leq c\}$, where $I\{\cdot\}$ stands for an indicator random variable that takes a value equal to 1 whenever its argument is true, and 0 otherwise with n_r equal to the number of log-conductivity fields. It should be recalled that the empirical distribution function will converge to its theoretical counterpart as the number of realizations in the ensemble tends to infinity (Billingsley, 1995). The variables with the tilde in equations 6-7 symbolize the transformed random variables. In addition to the transformations expressed by these equations, a standardization of the observations at each location have to be applied. The function relating y to \tilde{y} or h to \tilde{h} in the x - y Cartesian plane, is called the Gaussian anamorphosis function and was introduced by Rosenblatt (1952) (Figure 1).

Note that since only marginal transformations are applied, the joint multivariate distributions of $Y(\mathbf{x})$ and $H(\chi_1, t=1)$, as well as the multivariate distributions of $H(\chi_1, t=1)$ and $Y(\mathbf{x})$, are not modified.

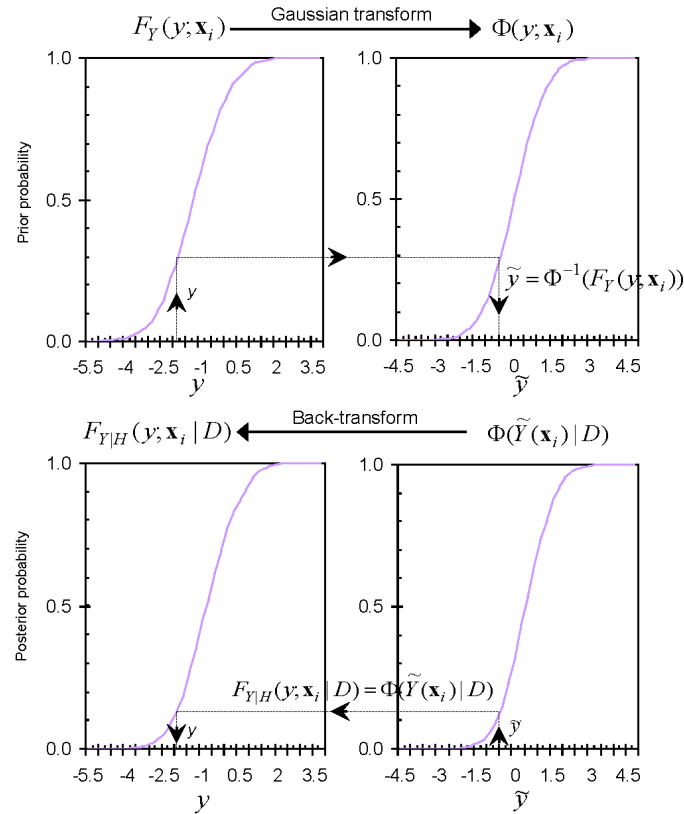


Figure 1. Transformation step of the EnKF method: a) Gaussian transformation process, b) Back-transformation process.

Modified update step

Once the transformations described in the previous section have been applied, the update step of the tEnKF method can now be written in terms of the anamorphosed variables:

$$\tilde{U}_t^u = \tilde{U}_0 + [\tilde{K}_t \quad \tilde{Z}_t - \tilde{H}_t^f] \quad (8)$$

where the weighting functions $\lambda_i(\mathbf{x}_a)$ of matrix $\tilde{\mathbf{K}}_t$ are obtained from the following system of linear equations which is written in terms of the auto-covariances and cross-covariances of the anamorphosed variables:

$$\sum_{l=1}^n \sum_{\alpha=1}^{N_h} \lambda_l(\chi_\alpha) C_{\tilde{H}}(s, \tau) = C_{\tilde{Y}\tilde{H}}(s, \tau) \quad (9)$$

Back-transformation step

After applying the modified update step, it is necessary to return to the original space for interpretative purposes. The mapping of the conditioned anamorphosed random variables into the original non-Gaussian distributed variables is made with:

$$(Y(\mathbf{x}_i)|D) = F_{\tilde{Y}|\tilde{H}}^{-1}(\Phi(\tilde{Y}(\mathbf{x}_i)|D)) \quad (10)$$

This means that the conditional CDF value of the original variable is identified with the conditional CDF value at its corresponding Gaussian transform value (Goovaerts, 1997) (Figure 1). The inverse of the univariate conditional CDF, i.e. $F_{\tilde{Y}|\tilde{H}}^{-1}(\cdot)$, is in fact a pseudoinverse function of its theoretical counterpart, but when the latter is strictly increasing, the pseudoinverse equals the usual inverse (Nelsen, 1997).

Computation of conditional moments

The conditional mean $\mu_{Y|H}(\mathbf{x})$ and conditional variance $\sigma_{Y|H}^2(\mathbf{x})$ of $Y(\mathbf{x})$, given that N_h observations of $H(\chi_1, t=1)$ are known, may be computed from:

$$\mu_{Y|H}(\mathbf{x}) = E\{Y^c(\mathbf{x})\} \quad (11)$$

$$\sigma_{Y|H}^2(\mathbf{x}) = E\{(Y^c(\mathbf{x}) - E\{Y^c(\mathbf{x})\})(Y^c(\mathbf{x}) - E\{Y^c(\mathbf{x})\})\} \quad (12)$$

where $Y(\mathbf{x})|D = Y^c(\mathbf{x})$.

Numerical implementation

The mathematical model described in the previous sections is coded in FORTRAN programming language on the LINUX platform and run in the HPC cluster "Tonatiuh" at the Institute of Engineering, UNAM. A block diagram describing its numerical implementation is depicted in Figure 2. Note that the conditioning process is repeated at the next time at which observations are available, but the new prior log-conductivity random field becomes the posterior one at time $t=1$.

Application of the tEnKF to the identification of conductivities

Multivariate spatiotemporal random fields have been used in a variety of geophysical applications. For example, Bodas-Salcedo *et al.* (2003) combined spatiotemporal random fields with the Kalman filter method to predict solar radiation in the earth-atmosphere system; Suciú (2014) used a diffusion model to predict solutes transport in groundwater under

uncertainty about spatiotemporal evolution of velocity fields; a similar approach was used by Suciú *et al.* (2016) to model reactive transport; Sanchez *et al.* (2016) developed a spatiotemporal dynamic model based on the classical EnKF for Bayesian inference of rainfall; and Liang *et al.* (2016) used a stochastic groundwater flow model to analyze the effect of uncertainty in recharge and transmissivity on the spatiotemporal variations of groundwater level in an unconfined aquifer. Finally, Moslehi and de Barros (2017) investigated the impact of uncertainty in spatial variability of soil hydraulic conductivity on several environmental performance metrics that are relevant for environmental risk assessments, such as species concentrations and arrival times, using a stochastic advection-dispersion model to represent the spatiotemporal evolution of the concentration field.

In this paper, the tEnKF is applied to infer the hydraulic conductivity field in a confined porous medium on the basis of the observed spatiotemporal variations of the hydraulic head field. For the sake of simplicity, groundwater flow through a one-dimensional, vertical, fully saturated random porous medium is considered (equations 1–2). It is assumed that hydraulic head at the lower boundary of the porous medium diminishes at a constant rate known from historical records and that this decrement is associated with groundwater withdrawal from wells located farther away. This simplified setting allows useful preliminary evaluations of the tEnKF. However, it is recognized that these two hypotheses can be avoided by 3D modeling of the system including the wells, and by considering that the spatial variability of the hydraulic conductivity in porous media is in fact 3D. The coupling of this dynamic model within the tEnKF is being developed and the results will be presented in further publications.

The porous medium is 40m deep and is discretized into 80 finite elements each with a length of 0.5m. Each finite element e_i for $i=1, \dots, 80$ is assigned a log-conductivity value $y_i^{ref}(\mathbf{x}_i)$ according to the following procedure. First, a multi-Gaussian field $g(\mathbf{x})$ with exponential auto-covariance function and correlation scale $a=2.5\text{m}$ is simulated using a modified version of the SGSIM random field generator (Deutsch and Journel, 1992) (Figure 3a). Second, the V-transform is applied in the following manner (Bárdossy and Li, 2008):

$$v(\mathbf{x}) = \begin{cases} k(g(\mathbf{x})m) & \text{if } g(\mathbf{x}) \geq m \\ k(g(\mathbf{x})m) & \text{if } g(\mathbf{x}) < m \end{cases} \quad (13)$$

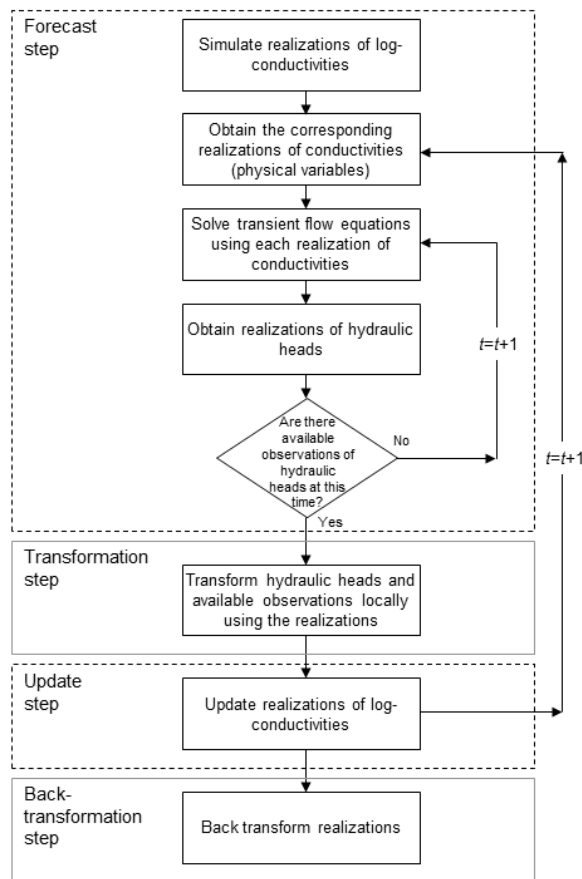


Figure 2. Block diagram of the EnKF with transformed data (tEnKF).

with arbitrarily chosen parameters $m=0$ and $k=1$, to the previously generated $g(\mathbf{x})$ field to obtain the transformed $v(\mathbf{x})$ field (Figure 3b). Third, a Gaussian distribution is imposed to the $v(\mathbf{x})$ field as $y'=\Phi^{-1}[F_v(v)]$ where $F_v(v)$ is the empirical CDF of the $v(\mathbf{x})$ field. Finally, this $y'(\mathbf{x})$ field is scaled to a normally distributed $y^{ref}(\mathbf{x})$ field with mean value $m_y=-1.654$ and variance $s_y^2=0.997$ as: $y^{ref}(\mathbf{x})=m_y+y'(\mathbf{x})s_y$. Each one of these values is assumed to be constant within its finite element e_i . This log-conductivity field, which is displayed in Figure 3c, is called the reference field and is considered the "true state of nature".

Several interesting properties of the V -transformation should be mentioned. First, the symmetric distribution function of the Gaussian field $g(\mathbf{x})$ is transformed into an asymmetrical distribution function through parameters m and k . Second, the empirical auto-covariance function of $g(\mathbf{x})$ is not preserved in $v(\mathbf{x})$ because the V -transformation is non-monotonous (Figure 4). Third, the spatial correlation of $v(\mathbf{x})$ is stronger for the values above the median than for the values below the median, i.e. the spatial correlation of $v(\mathbf{x})$ is asymmetric. This characteristic of the field holds after imposing the Gaussian (normal) distribution function onto it because the Gaussian transformation is monotonous (Deutsch and Journel, 1992).

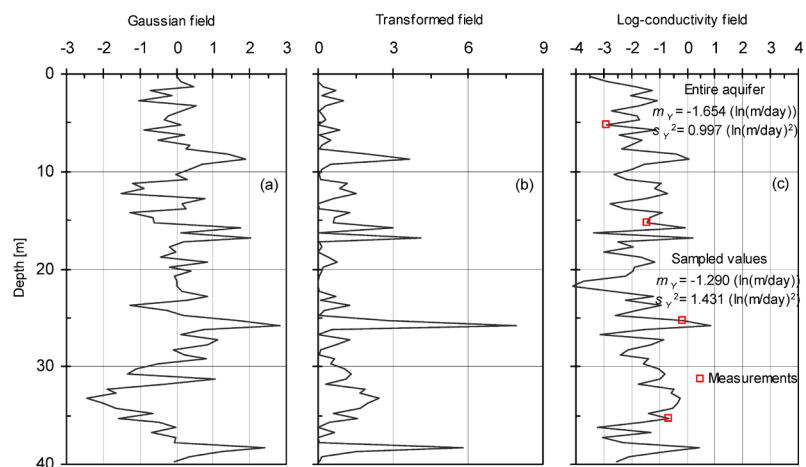
In the present example, since $y^{ref}(\mathbf{x})$ is normally distributed, $k_s(\mathbf{x})=\exp(y^{ref}(\mathbf{x}))$ is lognormally distributed with expected value $\mu_{Ks}=0.315$ m/day and variation coefficient $CV_{Ks}=1.31$. Conductivity fields with lognormal univariate distributions and asymmetric spatial correlation are considered to be more representative of natural porous media (Gómez-Hernández and Wen, 1998; Journel

and Zhang, 2006). Specific storage coefficient is assumed to be equal to 0.001m^{-1} throughout the porous medium.

Using the reference field of conductivities, groundwater head responses are generated by solving a transient flow condition with finite elements (Smith and Griffiths, 2004). At $t=0$ days, the initial distribution of heads is hydrostatic. At $t>0$ days, hydraulic head decreases with time at a rate of 0.15 m/day during 150 days at the lower boundary. For the purpose of the present numerical example, the distribution of heads at $t=90$ days is assumed to be the initial condition (denoted as $t=0$ days in Figure 5 and henceforth). It is further assumed that groundwater head responses are available at times $t=3$, $t=18$ and $t=60$ days at the two locations indicated in Figure 5. Thus, two histories with three hydraulic head values are generated. These indirect, informative variables of the hydraulic conductivity of the reference field are considered available transient piezometric observations. At each one of these three times, the update step of the tEnKF scheme is performed.

The reference field is sampled at four locations indicated in Figure 3c and the values are considered direct log-conductivity measurements. The mean and variance of the set of sampled values are reported in the same figure. Observe that these statistics overestimate the mean and variance of the reference field. Observe also in Figure 4 that an exponential auto-covariance function with correlation scale $\alpha=2.5\text{m}$ overestimates the correlation scale of the reference field as well. The sampled mean and variance, as well as the auto-covariance function mentioned above, are used to simulate two thousand unconditional multi-Gaussian log-conductivity

Figure 3. One dimensional fields: a) Initial Gaussian field, b) Field after applying the V -transform with parameters $m=0$ and $k=1$ to the initial Gaussian field, c) Final log-conductivity field after imposing a marginal normal distribution with expected value $m_y=-1.654$ and variance $s_y^2=0.997$ to the V -transformed field. Statistics of sampled values (empty squares) are also reported.



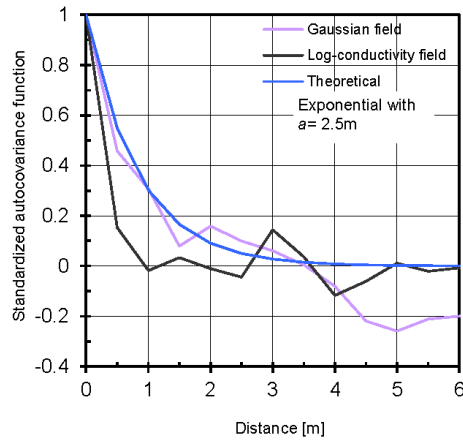


Figure 4. Standardized auto-covariance functions of the initial Gaussian field and reference log-conductivity field. An exponential function is also shown for comparison.

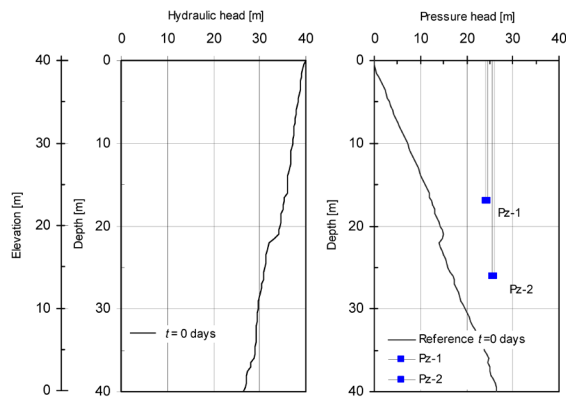


Figure 5. Profile of hydraulic heads in the reference field at $t=0$ days. The depths of the tips of two piezometers (Pz-1 and Pz-2) are indicated with filled squares.

realizations. Since the realizations are multi-Gaussian, they do not exhibit the asymmetric spatial correlation structure of the reference field (Journal and Deutsch, 1993). This set of log-conductivity fields attempts to model a situation in which the mean value, variance and auto-correlation function of the reference field are only roughly estimated *a priori* and the asymmetry of this field is ignored. It represents, in fact, the prior uncertainty about spatial variability of the conductivity in the porous medium.

The chosen number of simulated realizations ensures the stability of the following two error measures, according to preliminary computations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y^*(\mathbf{x}_i) - y^{ref}(\mathbf{x}_i))^2} \quad (14)$$

where n is the number of log-conductivities in the flow domain, $y^*(\mathbf{x}_i)$ is the estimated mean log-conductivity at location \mathbf{x}_i , and $y^{ref}(\mathbf{x}_i)$ is the reference log-conductivity also at location \mathbf{x}_i .

The SPREAD is computed as:

$$SPREAD = \sqrt{\frac{1}{n} \sum_{i=1}^n s_{en}^2(\mathbf{x}_i)} \quad (15)$$

where $s_{en}^2(\mathbf{x}_i)$ is the variance of the estimation of the log-conductivity at location \mathbf{x}_i computed statistically over the ensemble of realizations.

It is worth mentioning that RMSE is a measure of the difference of the means of the estimated and reference fields and SPREAD is a measure of the dispersion of the estimated field around the reference field. Therefore, they can be viewed as measures of accuracy and precision of the estimations, respectively.

Results and discussion

Effect of the Gaussian transformation

Figure 6 illustrates the Gaussian transformation process of the hydraulic head at an arbitrarily selected node before the first update step. In general, the shape of the local distributions depends on the location of the node in the flow domain and on the boundary conditions of the problem at hand. In all cases, the local distribution functions can be transformed into Gaussian distributions by building local Gaussian anamorphosis functions numerically, as explained earlier. As can be seen in Figure 6(a), although the original values exhibit a skewed distribution, the transformed variable becomes symmetric around the mean showing the well-known bell-shape of the Gaussian anamorphosis (Figures 6b and 6c).

Figure 7 represents the relationship between log-conductivities and heads at arbitrarily selected locations, before (Figure 7a) and after (Figure 7b) applying the respective Gaussian transformations. Given that the Gaussian transformation is monotonous, the bivariate characteristics of the dependence, such as the correlation structure at different percentiles, are not modified (Deutsch and Journal, 1992; Chilès and Delfiner, 1999). However, the linear

correlation coefficient, which depends on the kind of marginal distributions of the random variables, might be different before and after transformations. In the particular case of the variables at the locations indicated in Figure 7 (at the upper right corner of each figure), this coefficient presents nearly the same value before and after transformations. Therefore, the implicit pseudo-linearization effect associated to the Gaussian anamorphosis reported by Shöniger *et al.* (2012) should be considered application dependent.

Effects of conditioning on log-conductivities alone

The impact of conditioning realizations of log-conductivities to direct measurements only is analyzed. Figure 8 displays comparisons between the reference field of log-conductivities and the mean of the conditional realizations of log-conductivities. Contrasting both fields, it is observed that at the measurement locations the conditional value is the measured value.

Figure 9 reproduces profiles of standard deviations (uncertainty) computed with the conditional realizations of log-conductivities. Observe that the effect of conditioning is to reduce, overall, the prior uncertainty and to collapse it to zero at the locations of the measurements.

Effects of conditioning on log-conductivities and transient heads

In what follows, the additional impact of conditioning the realizations of log-conductivities to transient heads responses is examined. Comparisons of the reference field of log-conductivities with the mean of the conditional realizations of log-conductivities of the tEnKF at times $t=3$, $t=18$ and $t=60$ days are shown in Figures 10a, b and c, respectively.

Contrasting the RMSE and SPREAD values shown at the bottom right corner of Figure 10a with the same two error measures shown in Figure 8, it can be observed that the estimation

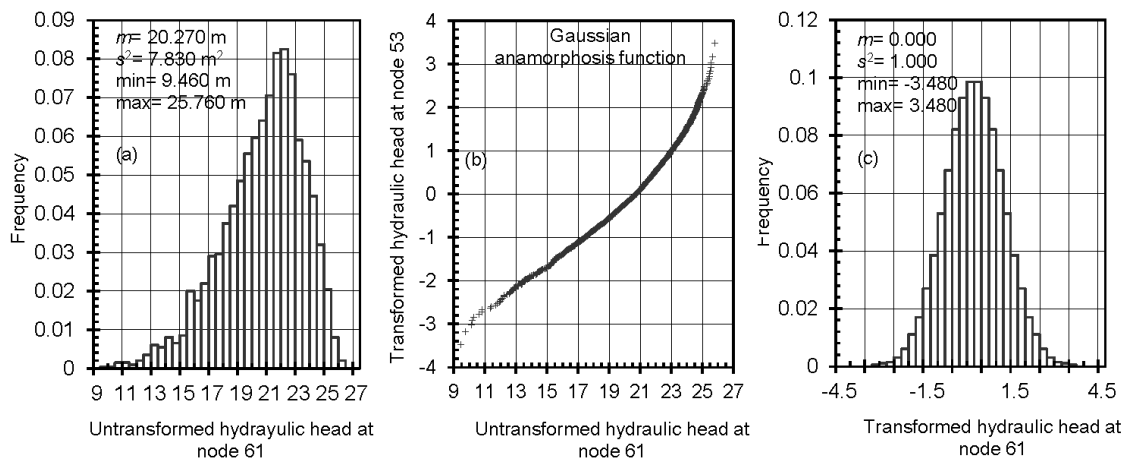


Figure 6. Gaussian transformation of hydraulic heads at node 61: a) Histogram of untransformed hydraulic heads, b) Gaussian anamorphosis function (with zero mean and unity variance), c) Histogram of hydraulic heads after the Gaussian anamorphosis.

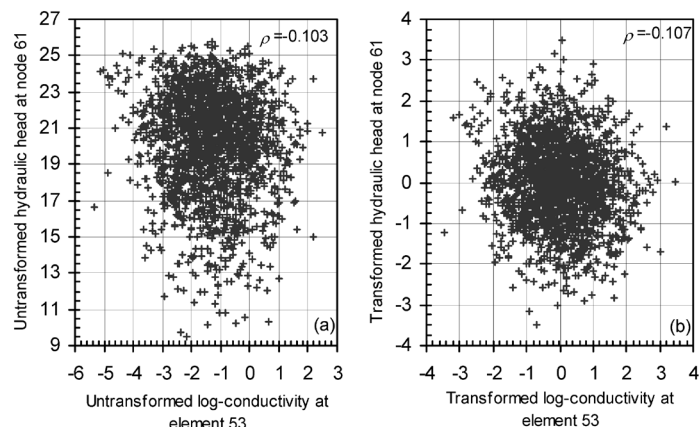


Figure 7. Relationship between log-conductivity and hydraulic head at two arbitrary selected locations (ρ is the Pearson correlation coefficient): a) Before the Gaussian transformation of both variables, b) After the Gaussian transformation of both variables.

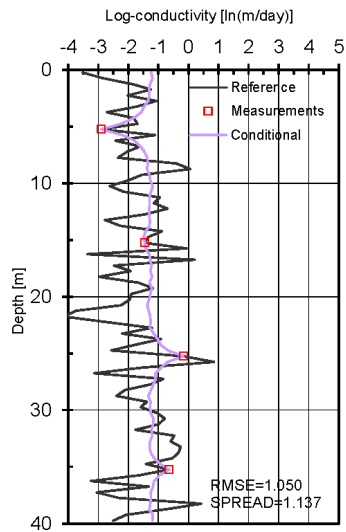


Figure 8. Log-conductivity fields conditional to log-conductivities alone. The reference field is also shown.

of the hydraulic head field becomes more accurate and precise after assimilating the first pair of measured hydraulic heads with the tEnKF. Further improvements in the estimation of the reference field are observed as more measured hydraulic head data are assimilated. For instance, the RMSE and SPREAD values for the conditional mean log-conductivity field of the tEnKF at time $t=60$ days are 0.964 and 1.080, whereas before the filtering process they were equal to 1.050 and 1.137, respectively. The accuracy of the estimation with the tEnKF can be attributed to the fluctuations that occur between measurements which follow the variability of the reference medium more closely. However, it should be recalled that the RMSE and SPREAD values measure the quality of the local estimation only, i.e. they do not indicate anything about the quality of the multivariate estimation. Bivariate empirical copulas can be used to perform this evaluation (Bárdossy and Li, 2008).

The standard deviation profiles calculated with the realizations of log-conductivities of the tEnKF are reported in Figure 11. The overall uncertainty decreases as longer groundwater head records are incorporated into the update step, except at the locations of direct measurements where uncertainty is zero at all times. The Figure shows that uncertainty is smaller around depths of 17 m and 26 m (where the tips of the two piezometers are located) than at other depths.

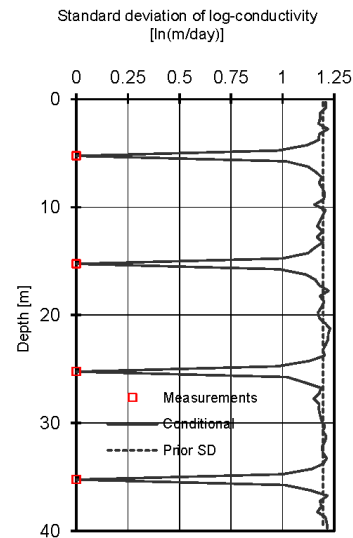


Figure 9. Profiles of conditional standard deviations of log-conductivities with respect to depth (empty squares indicate the locations of known values).

Figure 12 displays the frequency distributions of log-conductivities of the reference field and of the prior and posterior ensembles at the end of the conditioning process. Recall that the mean value of the conductivity of the reference field was overestimated by the prior random field; hence, the distribution function of this field is located to the right of the field's frequency function. Looking at the distributions of the posterior fields, it is observed that they exhibit some features of the reference frequencies (like some of the "peaks" of both branches) and that they are slightly displaced toward the left of the frequencies of the prior random field. This result illustrates the attempt of the filter to lead the prior frequencies toward the reference frequencies when a small number of instruments and short records of measured hydraulic head responses ($t=3$, $t=18$ and $t=60$ days) are used in the conditioning process and when the initial representation of the asymmetric correlation structure of the true field is ignored by the prior random field model. Although there is a benefit, the degree of improvement of the initial approximation should be further investigated because the depth and distance between instruments, as well as the times at which observations are available, and the length of the histories of hydraulic head responses may also have an effect on the results.

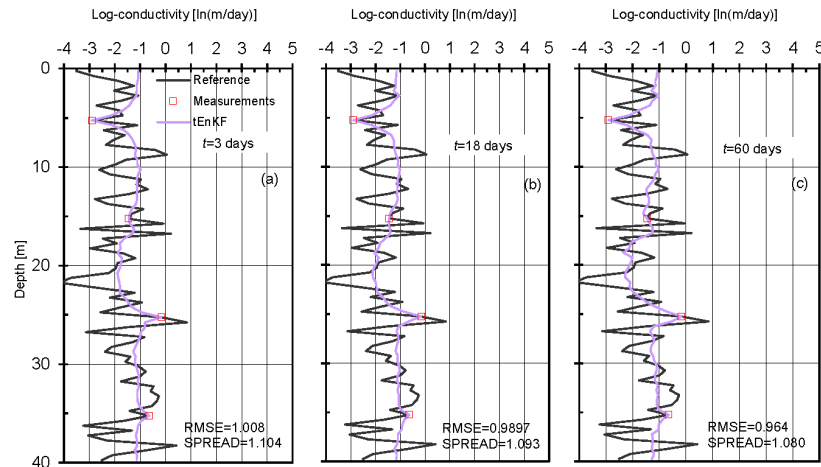


Figure 10. Log-conductivity fields conditional to histories of hydraulic heads with the tEnKF method. The reference field is also shown: a) At $t=3$ days, b) At $t=18$ days, c) At $t=60$ days.

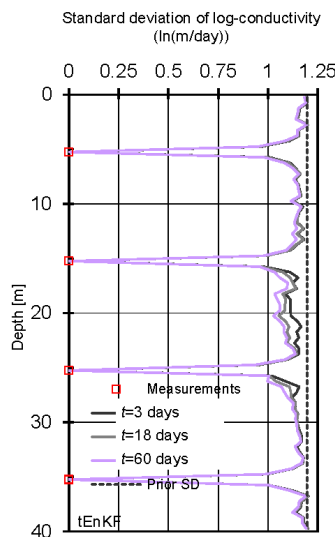
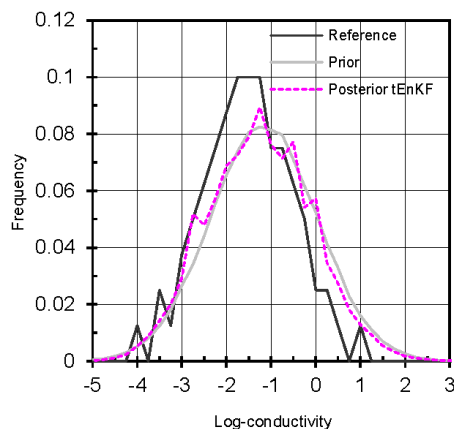


Figure 11. Profiles of conditional standard deviation of log-conductivities with respect to depth at different times (empty squares indicate the locations of known values).



Conclusions

This paper proposed that in order to evaluate the performance of tEnKFs in synthetically generated fields of the hydraulic conductivity, it is necessary to take into account the risk of introducing a systematic bias in the spatiotemporal evolution of the hydraulic head field by incorporating prior knowledge with a multi-Gaussian conductivity correlation structure, and by adopting a reference field with asymmetric correlation structure. This setting aims to offer a truer representation of common situations in practice in which the first two moments and the auto-covariance function of the real field are roughly known, but the asymmetry of the spatial correlation structure of that field is unknown.

As an example of the proposed approach, hydraulic conductivities were identified using the tEnKF by solving a one-dimensional, single phase flow problem in a continuous random porous medium. Three effects on the reproduction of the reference field were evaluated: the effect of the Gaussian transformation process, the effect of incorporating only measured hydraulic conductivities and the effect of incorporating measured hydraulic conductivities and hydraulic heads data. The results of this example indicate that when the asymmetry of the spatial correlation structure of the reference field was unknown *a priori*, the tEnKF was still capable of improving the initial approximation. However, no definitive conclusions can be drawn from

Figure 12. Frequency distributions of log-conductivities of the reference field, set of prior realizations and sets of posterior realizations at $t=60$ days.

this study in terms of the performance of tEnKF under this constraint since the degree of improvement of the initial approximation may further depend on the configuration of the array of instruments, the times at which observations are available and the length of time series of hydraulic heads.

Further research is needed to fully assess the performance of the EnKF with transformed data. For example, its effect on the reproduction of the multivariate spatial dependence of a non multi-Gaussian reference field using a multi-Gaussian random field model of conductivities in the prior approximation should be explored

Acknowledgments

The authors sincerely thank the four anonymous reviewers for improving the clarity of the original manuscript with their comments. We would also like to thank the people at the Academic Writing Office at the Coordinación de Estudios de Posgrado, UNAM, for their help revising our manuscript.

References

- Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas, *Water Resources Research*, 44, W07412 (1 of 15).
- Bárdossy, A., and Hörning, S. (2016), Gaussian and non-Gaussian inverse modeling of groundwater flow using copulas and random mixing, *Water Resources Research*, 52, pp. 4504–4526.
- Béal, D., Brasseur, P., Brankart, J., Ourmieres, Y. and Verron, J. (2010). Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis, *Ocean Sciences*, 6, pp. 247–62.
- Bertino, L., Evensen, G. and Wackernagel, H. (2003). Sequential Data Assimilation Techniques in Oceanography, *International Statistical Review*, 71(2), pp. 223–241.
- Billingsley, P. (1995). Probability and measure, 3rd ed. Wiley, New York.
- Bodas-Salcedo, A., López-Baeza, E., Martinez, F., Mateu, J. and Montes, F. (2003). Spatiotemporal modeling and prediction of solar radiation, *Journal of Geophysical Research*, 108 (D24), 8777.
- Burgers, G., van Leeuwen, P. J. and Evensen, G. (1998). Analysis scheme in the Ensemble Kalman Filter, *Monthly Weather Review*, 126(6), pp. 1719–1724.
- Capilla, J. and Llopis-Albert, C. (2009). Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data: 1, Theory, *Journal of Hydrology*, 371, pp. 66–74.
- Chilès, C. and Delfiner, P. (1999). Geostatistics: modeling spatial uncertainty, Wiley, New York.
- Dagan, G. (1989). Flow and transport in porous formations, Springer-Verlag, Berlin, Germany.
- Deutsch, C. V. and Journel A. G. (1992). GSLIB, Geostatistical Library and User's Guide, Oxford University Press, Oxford, N. Y., USA.
- Elsheikh, A. H., Wheeler, M. F. and Hoteit, I. (2013). An iterative stochastic ensemble method for parameter estimation of subsurface flow models, *Journal of Computational Physics*, 242, pp. 696–714.
- Emerick, A. A. (2017). Investigation on Principal Component Analysis Parameterizations for History Matching Channelized Facies Models with Ensemble-Based Data Assimilation, *Mathematical Geosciences*, 49, pp. 85–120.
- Erdal, D., Rahman, M. A. and Neuweiler, I. (2015). The importance of state transformations when using the ensemble Kalman filter for unsaturated flow modeling: Dealing with strong nonlinearities, *Advances in Water Resources*, 86, pp. 354–365.
- Evensen, G., (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model, *Journal of Geophysical Research*, 97 (11), pp. 17905–17924.
- 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, pp. 10143–10162.
- Evensen, G. (2003). Data assimilation: the ensemble Kalman Filter, Springer, New York.
- Gelhar, L. W. (1993). Stochastic subsurface hydrology, Pentrice Hall, New Jersey.

- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*, Oxford University Press, New York.
- Gómez-Hernández, J. J. and Wen, X. (1994). Probabilistic assessment of travel times in groundwater modeling, *Stochastic Hydrology and Hydraulics*, 8(1), pp. 19–55.
- Gómez-Hernández, J. and Wen, X. (1998). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology, *Advances in Water Resources*, 21(1), pp. 47–61.
- Gómez-Hernández, J. J., Sahuquillo, A., and Capilla, J. E. (1997). Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data, 1, Theory, *Journal of Hydrology*, 203, pp. 162–174.
- Haslauer, C., Guthke, P., Bárdossy, A. and Sudicky, E. (2012). Effects of non-Gaussian copula-based hydraulic conductivity fields on macrodispersion, *Water Resources Research*, 48, W07507.
- Hendricks F., H. J., and Kinzelbach, W. (2008), Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resources Research*, 44, W09408
- Hu, L. Y. (2000). Gradual deformation and iterative calibration of Gaussian-related stochastic models, *Mathematical Geology*, 32(1), pp. 87–108.
- Hu, L. Y., Zhao, Y., Liu, Y., Scheepens, C., Bouchard, A. (2013). Updating multipoint simulations using the ensemble Kalman filter, *Computers and Geosciences*, 51, pp. 7–15.
- Johns, C. J. and Mandel, J. (2008). A two-stage ensemble Kalman filter for smooth data assimilation, *Environmental Ecological Statistics*, 15, pp. 101–110.
- Journal, A. and Alabert, F. (1989). Non-Gaussian data expansion in the earth sciences, *Terra Nova*, 1, pp. 123–134.
- Journal, A. G. and Deutsch, C. V. (1993). Entropy and spatial disorder, *Mathematical Geology*, 25(3), pp. 329–55.
- Journal, A. G. and Huijbregts, C. (1978). "Mining geostatistics", Academic Press, San Diego, Calif.
- Journal, A. and Zhang, T. (2006). The necessity of a multiple point statistical model, *Mathematical Geology*, 38(5), pp. 591–610.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, 82, pp. 35–45.
- Katzfuss, M., Stroud, J. R. and Wikle, C. K. (2016). Understanding the Ensemble Kalman Filter, *The American Statistician*, 70(4), pp. 350–357.
- Kerrou, J., Renard, P., Hendricks Franssen, H. J. and Lunati, I. (2008). Issues in characterizing heterogeneity and connectivity in non-multiGaussian media, *Advances in Water Resources*, 31, pp. 147–159.
- Knudby, C. and Carrera, J. (2005). On the relationship between indicators of geostatistical flow and transport connectivity, *Advances in Water Resources*, 28, pp. 405–421.
- Leng, C. H., and Yeh, H. D. (2003). Aquifer parameter identification using the extended Kalman filter, *Water Resources Research*, 39(3), 1062.
- Li, L., Zhou, H., Hendricks F., H. J. and 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, pp. 573–590.
- Li, L., Srinivasan, S., Zhou, H., Gómez-Hernández, J. J. (2013). A pilot point guided pattern matching approach to integrate dynamic data into geological modeling, *Advances in Water Resources*, 62, pp. 125–38.
- Li, L., Srinivasan, S., Zhou, H., Gómez-Hernández, J. J. (2014). Simultaneous estimation of geologic and reservoir state variables within an ensemble-based multiple-point statistic framework, *Mathematical Geosciences*, 46(5), pp. 597–623.
- Li, L., Srinivasan, S., Zhou, H. and Gómez-Hernández, J. J. (2015). Two-point or multiple point statistics? A comparison between the ensemble Kalman filtering and the ensemble pattern matching inverse methods, *Advances in Water Resources*, 86, pp. 297–310.

- Liang, X., Zhang, Y. K. and Schilling, K. (2016). Effect of heterogeneity on spatiotemporal variations of groundwater level in a bounded unconfined aquifer, *Stochastic Environmental Research and Risk Assessment*, 30, pp. 1–8.
- Miller, R. N., Carter, E. F. and Blue, S. T. (1999). Data assimilation into nonlinear stochastic models, *Tellus A*, 51, pp. 167–194.
- Moslehi, M. and de Barros, F. (2017). Uncertainty quantification of environmental performance metrics in heterogeneous aquifers with long-range correlations, *Journal of Contaminant Hydrology*, 196, pp. 21–29.
- Nelsen, R. B. (1997). An introduction to copulas, Springer, Berlin.
- Oliver, D., Cunha, L., and Reynolds, A. (1997). Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, *Mathematical Geology*, 29, 61–91.
- RamaRao, B., LaVenue, A., De Marsily, G. and Marietta, M. (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1, theory and computational experiments, *Water Resources Research*, 31(3), pp. 475–93.
- Reichle, R. H., Walker, J. P., Koster, R. D. and Houser, P. R. (2002). Extended versus ensemble Kalman filtering for land data assimilation, *Journal of Hydrometeorology*, 3, pp. 728–740.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation, *The Annals of Mathematical Statistics*, 23(3), pp. 470–472.
- Sahuquillo, A., Capilla, J. E., Gómez-Hernández, J. J., and Andreu, J. (1992). Conditional simulation of transmissivity fields honoring piezometric head data, In: *Hydraulic Engineering Software IV, Fluid Flow Modeling*, edited by: Blair, W. R. and Cabrera, E., Vol. II, Elsevier Applied Science, London, UK, pp. 201–214.
- Sanchez, L., Infante, S., Griffina, V. and Reya, D. (2016). Spatio-temporal dynamic model and parallelized ensemble Kalman filter for precipitation data, *Brazilian Journal of Probability and Statistics*, 30 (4), pp. 653–675.
- Sarma, P. and Chen, W. H. (2009). Generalization of the Ensemble Kalman Filter using kernels for non-Gaussian random fields, *SPE* 119177, pp. 1–20.
- Schöniger, A., Nowak, W. and Hendricks F., H. (2012). Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography, *Water Resources Research*, 48(W04502), pp. 1–18.
- Shinozuka, M. and Zhang, R. (1996). Equivalence between Kriging and CPDF Methods for Conditional Simulation, *Journal of Engineering Mechanics*, 122(6), pp. 530–538.
- Simon, E. and Bertino, L. (2009). Application of the Gaussian anamorphosis to assimilation in a 3D coupled physical ecosystem model of the North Atlantic with the EnKF: a twin experiment, *Ocean Sciences*, 5, pp. 495–510.
- Simon, E. and Bertino, L. (2012). Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: application to a 1D ocean ecosystem model, *Journal of Marine Systems*, 89(1), pp.1–18.
- Smith, I. and Griffiths, D. (2004). Programming the finite element method, John Wiley and Sons Ltd., England.
- Suciu, N. (2014). Diffusion in random velocity fields with applications to contaminant transport in groundwater, *Advances in Water Resources*, 69, pp. 114–133.
- Suciu, N., Schüller, L., Attinger, S. and Knabner, P. (2016). Towards a filtered density function approach for reactive transport in groundwater, *Advances in Water Resources*, 90, pp. 83–98.
- Sun, A. Y., Morris, A. P., and Mohanty, S. (2009). Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques, *Water Resources Research*, 45, W07424.
- Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. SIAM, USA.
- Xu, T., Gómez-Hernández, J., Zhou, H. and 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 heterogeneous bimodal hydraulic conductivity field, *Advances in Water Resources*, 54, pp. 100–118.

- Zhang, D. (2002). Stochastic methods for flow in porous media, Academic Press, USA.
- Zhou, H., Gómez-Hernández, J., Hendricks Franssen, H. and Li, L. (2011). An approach to handling nongaussianity of parameters and state variables in ensemble Kalman filtering, *Advances in Water Resources*, 34(7), pp. 844–864.
- Zhou, H., Gómez-Hernández, J. J. and Li, L. (2012). A pattern-search-based inverse method, *Water Resources Research*, 48, W03505.
- Zhou, H., Gómez-Hernández, J. J. and Li, L. (2014). Inverse methods in hydrogeology: Evolution and recent trends, *Advances in Water Resources*, 63, pp. 22–37.
- Zovi, F., Camporese, M., Hendricks Franssen, H. J., Huisman, J. A. and Salandin, P. (2017). Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter, *Journal of Hydrology*, 548, pp. 208–224.