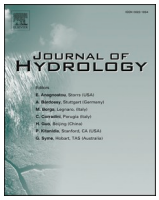
Journal of Hydrology 634 (2024) 131054

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Characterization of conductivity fields through iterative ensemble smoother and improved correlation-based adaptive localization

Chuan-An Xia [a,b,c], Jiayun Li [b], Monica Riva [c,d,*], Xiaodong Luo [e,*], Alberto Guadagnini [c,d]

[a] Zijin School of Geology and Mining, Fuzhou University, Fuzhou, China
[b] Institude of Groundwater and Earth Sciences, Jinan University, Guangzhou, China
[c] Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Milan, Italy
[d] Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, USA
[e] Norwegian Research Centre (NORCE), Bergen, Norway

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Claus Philip Haslauer, Associate Editor

Keywords:
Ensemble-based data assimilation
Iterative ensemble smoother
Correlation-based localization
Adaptive threshold
Generalized sub-Gaussian model
Contamination

ABSTRACT

Localization is critical to the effective use of an (iterative) ensemble Kalman filter or ensemble smoother to estimate uncertain quantities of interest. Here, we propose a novel, fully adaptive, correlation-based localization method (termed FBadap). We embed our FBadap approach within an iterative ensemble smoother to estimate three-dimensional spatially heterogeneous log-conductivity (Y) fields. The latter are characterized through a Generalized sub-Gaussian model, which includes the Gaussian distribution as a particular case. They constitute random fields within which head and concentration observations are collected at monitoring wells screened at multiple depths. To ensure transparent comparisons, we study and analyze the performance of our approach through a wide range of synthetic test cases. These comprise diverse configurations, including (*a*) various ensemble sizes, (*b*) various degrees of departure of the description of the spatial heterogeneity from a Gaussian model, as well as (*c*) different values of the mean and variance of the initial ensemble of Y. Our results show that (*i*) FBadap is robust adaptive approach enabling one to tackle a variety of settings; (*ii*) FBadap exhibits stronger adaptivity to cope with diverse ensemble sizes than FBconst, and can provide improved accuracy of conductivity estimates in comparison with traditional methods; and (*iii*) the quality of conductivity estimates is jointly impacted by the degree of departures of the reference Y field and of the initial ensemble of Y from a description based on a Gaussian model.

1. Introduction

Ensemble-based data assimilation (or history matching) techniques (e.g., Evensen, 2009; Emerick and Reynolds, 2013; Chen and Oliver, 2013; Luo et al., 2015) are widely used in Earth system sciences. Their practical implementation typically involves relying on a limited-size collection of realizations of the random field (e.g., spatially heterogeneous hydraulic conductivities) to be estimated, mainly due to constraints associated with available computational resources. This, in turn, leads to issues associated with rank-deficiency of the covariance matrix of system parameters/states of interest as well as spurious correlations amongst parameters/states/observations. The latter can hamper the variability across the realizations of the collection. It can thus severely affect the overall quality of data assimilation and our ability to characterize the system within which the dynamics of hydrological processes of interest are developing.

Localization has been proposed to mitigate negative impacts of small ensemble size. A variety of studies document development and/or application of localization techniques in diverse fields. These include, e.g., petroleum engineering (Chen and Oliver, 2010; Emerick and Reynolds, 2011; Luo et al., 2018; Soares et al., 2019; Luo and Bhakta, 2020), meteorology (Houtekamer and Mitchell, 1998; Bocquet and Sakov, 2017), and hydrology (Nan and Wu, 2011; Zovi et al., 2017; Rasmussen et al., 2015; Li et al., 2018). Examples of applications in the context of the latter include groundwater (e.g., Zovi et al., 2017; Li et al., 2018 and references therein), land surface (e.g., Vrugt et al., 2005), and coupled subsurface-surface water systems (e.g., Rasmussen et al., 2015) scenarios.

Distance- (Houtekamer and Mitchell, 1998; Sun et al., 2009; Chen and Oliver, 2010; Emerick and Reynolds, 2011; Xia et al., 2018; Li et al.,

* Corresponding authors at: Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Milan, Italy (M. Riva).
E-mail addresses: monica.riva@polimi.it (M. Riva), xluo@norceresearch.no (X. Luo).

https://doi.org/10.1016/j.jhydrol.2024.131054
Received 9 June 2023; Received in revised form 14 January 2024; Accepted 23 February 2024
Available online 14 March 2024

2018) and correlation- (Furrer and Bengtsson, 2007; Luo et al., 2018; Luo and Bhakta, 2020) based localization approaches are characterized by distinct key features. Distance-based localization approaches limit the spatial extent of the influence of available observations on local model parameter values through a specified tapering function. The latter is based on the relative distance between locations associated with an observation and a parameter to be estimated, respectively. Tapering functions that are typically employed (exponential (e.g., Tong et al., 2012; Xia et al., 2018), Gaussian (e.g., Nan and Wu, 2011), or Gaspari-Cohn (e.g., Hamill et al., 2001; Wang et al., 2018) functions) are unimodal and specified/characterized through some characteristic length scales. Applications in the context of hydrological sciences include, e.g., the studies of Sun et al. (2009), Xia et al. (2018), Li et al. (2018) and references therein. Drawbacks associated with distance-based localization methods comprise: (*i*) the need for a high computational cost when estimating a large number of parameters (Wang et al., 2018); (*ii*) the inherent difficulty to adaptively cope with space–time dynamics of the strength of the correlation between observations (or between observations and parameters to be estimated); (*iii*) the documented dependence of the optimal choice of tapering function on the forward model (Arroyo-Negrete et al., 2008; Emerick and Reynolds, 2011; Soares et al., 2019) and/or data type (e.g., Xia et al., 2018); and (*iv*) the inability to cope with parameters and/or observations which are not strictly characterized by a given physical location (e.g., Furrer and Bengtsson, 2007; Luo et al., 2018). With reference to the latter, we recall typical hydrological scenarios associated with (*a*) large scale parameter information such as, e.g., watershed scale soil depth, soil parameters related to a typical scale of a soil profile, or regional scale (effective/equivalent) conductivities; or (*b*) nonlocal (or depth-averaged) information such as, e.g., head and concentration data associated with monitoring wells screened along multiple depth intervals (Konikow et al., 2009; Xia et al., 2021).

In contrast, a correlation-based localization limits the strength of the influence of observations on parameter estimates via a specified tapering function which is based on sample (cross-) correlations between parameters and simulated observations. The value of a so-called tapering coefficient, r, between a simulated observation and a parameter drives the weight that the observation exerts on the parameter to be estimated. Furrer and Bengtsson (2007) propose to evaluate the tapering coefficient through a piece-wise tapering function with a constant threshold that needs to be tuned (see additional details in Section 2.3). Soares et al. (2019) point out that the selection of the (constant) threshold value significantly impacts the variability of the model parameters across the collection of realizations. A proper value of such a threshold may be case-dependent and can range (in terms of order of magnitude) from 10^{-3} to 10^{-1}, even higher values being documented in some cases (e.g., Furrer and Bengtsson, 2007; Soares et al., 2019; Lacerda et al., 2021; and references therein).

Miyoshi (2010) suggests to evaluate r as the product of two adaptive components. One of these components stems from cross-validation relying on two sets of sample correlations from two hierarchical collections (or sub-ensembles) of parameters with an identical number of members (Anderson, 2007). It is expressed as $(1-|c_1 - c_2|/2)^{\beta}$; here, c_1 and c_2 correspond to the degree of correlation between parameters and simulated observations related to the two sub-ensembles of parameters, respectively; and β is a constant that is typically tuned by trial-and-error, thus being case-dependent. The other component considers high levels of the correlation between parameters and simulated observations (Bishop and Hodyss, 2009). It is typically evaluated as $|c|^{\gamma}$, where c and γ are the strength of the correlation assessed across the whole ensemble (which includes the two aforementioned sub-ensembles) and a constant to be appropriately tuned (hence being case-dependent), respectively.

The localization approaches listed above are semi-adaptive, consistent with the observation that the associated hyperparameters (e.g., β or γ) are required to be tuned. Luo and Bhakta (2020) recently propose a fully adaptive approach relying on the use of the Gaspari-Cohn formulation as a tapering function to evaluate the tapering coefficient under some assumptions.

Correlation-based localizations are considered only in a limited number of studies in the context of hydrological sciences. Rasmussen et al. (2015) rely on the approach proposed by Miyoshi (2010) and apply the correlation-based localization within an ensemble Kalman filter approach to estimate parameters of a coupled subsurface-surface hydrological model. Available observations include heads across the groundwater system and stream discharge. These authors compare a correlation- and a distance-based localization approach. They note that the former outperforms the latter because of its enhanced ability to quantify and constrain the influence of observations to each parameter.

Our study aims at developing an original fully adaptive localization method, hereafter termed FBadap. The latter is grounded on the correlation-based approach originally proposed by Furrer and Bengtsson (2007) and adaptively evaluates the threshold value. We embed our scheme in an iterative ensemble smoother and test its potential against (*a*) its original counterpart (hereafter denoted as FBconst; Furrer and Bengtsson, 2007) which relies on a pre-set constant threshold and (*b*) the localization method proposed by Luo and Bhakta (2020; hereafter denoted as LB2020). To provide transparent assessment and comparison of the approaches analyzed, we rest on a suite of computational settings that we take as reference. In this framework, we consider scenarios associated with Gaussian and non-Gaussian heterogeneous log-conductivity fields, $Y(\boldsymbol{x})$ ($\mathbf{x}$ denoting a vector of spatial coordinates). We then explore the impact of the degree of departure from a description of spatial heterogeneity based on a Gaussian model within a three-dimensional subsurface system in the presence of head and concentration observations. Here, we consider the presence of long-screened monitoring wells at which monitored data are representative of (depth-)average system states and can be regarded as information of an integral nature. We refer to these types of information as nonlocal, for simplicity (see also Section 2).

The non-Gaussian random fields we analyze are characterized through a Generalized sub-Gaussian (GSG) model (Riva et al., 2015a). The latter constitutes a modeling approach which enables one to characterize jointly the statistical behavior of a given (spatially heterogeneous) quantity, $Y(\boldsymbol{x})$, and its spatial increments, $\Delta Y(\boldsymbol{s}) = Y(\boldsymbol{x} + \boldsymbol{s}) - Y(\boldsymbol{x})$, evaluated at various separation distances (or lags), $\mathbf{s}$. This approach is capable of interpreting salient statistical features that have been documented through the analysis of a wide range of hydrogeological variables (e.g., Guadagnini et al., 2018; and references therein) and result in a scale-dependent behavior of sample distributions (and associated statistical moments) of spatial increments. Documented manifestations of such a behavior comprise the observation that the distribution of $\Delta Y(\boldsymbol{s})$ tends to be symmetric and to develop heavier tails and sharper peaks as lag decreases. Such a behavior has been displayed (among others) by log-conductivity and permeability (Painter, 1996, 2001; Liu and Molz, 1997; Meerschaert, 2004; Siena et al., 2012, 2017; Riva et al., 2013a, 2013b; Guadagnini et al., 2018), electrical resistivity (Painter, 2001), vadose zone hydraulic properties (Guadagnini et al., 2012, 2013, 2014, 2015), neutron porosity (Riva et al., 2015a), sediment transport quantities (e.g., Ganti et al., 2009), fully developed turbulence (Boffetta et al., 2008), and micro-scale geochemical data (Siena et al., 2020, 2021). The impact of these aspects on flow and transport in porous media whose log-conductivity is characterized through a GSG model has been explored in preliminary analytical and numerical studies (Riva et al., 2017; Libera et al., 2017; Sole-Mari et al., 2021; Ceresa et al., 2022).

The remainder of the work is structured as follows. Section 2 provides details on the methodology, including some background on head and concentration observations related to monitoring wells which are screened along multiple depths (Section 2.1), the main formulations associated with the iterative ensemble smoother (Section 2.2), the three tested localization schemes (Section 2.3), and the theoretical elements underpinning the GSG formulation adopted to describe the three-

dimensional spatial distribution of log-conductivity (Section 2.4). Section 3 illustrates the design of the computational analyses together with three criteria considered to assess the performance of the localization approaches. Results and related analyses are included in Section 4. Main conclusions of the work are provided in Section 5.

2. Methodology

2.1. Groundwater flow - transport and borehole data

We consider a three-dimensional confined aquifer associated with transient flow, as governed by

$$S_s\frac{\partial h(x,t)}{\partial t} = -\nabla\cdot q(x,t) + f(x,t),\ with\ q(x,t) = -K(x)\nabla h(x,t) \tag{1}$$

where S_s [L^{-1}] is specific storage; h [L] is hydraulic head; $\boldsymbol{x} = (x_1, x_2, x_3)$ is a spatial coordinator vector; t [T] is time; K [LT^{-1}] is hydraulic conductivity; $\mathbf{q}$ [LT^{-1}] is Darcy velocity vector; and f [T^{-1}] is a source term.

We take (non-reactive) solute transport to be driven by the classical advection dispersion equation

$$\frac{\partial\theta C(\boldsymbol{x},t)}{\partial t} = \nabla\cdot\left[\theta\underline{\underline{\mathbf{D}}}(\boldsymbol{x},t)\nabla C(\boldsymbol{x},t)\right] - \nabla\cdot[\boldsymbol{q}(\boldsymbol{x},t)C(\boldsymbol{x},t)] + q_s(\boldsymbol{x},t)C_s(\boldsymbol{x},t) \tag{2}$$

Here, θ [–] is effective porosity; C [ML^{-3}] is solute concentration; q_s [T^{-1}] is a source/sink term associated with solute concentration C_s [ML^{-3}], and $\underline{\underline{\mathbf{D}}}$ [L^2T^{-1}] is the dispersion tensor

$$\underline{\underline{D}} = \begin{bmatrix} D_l & 0 & 0 \\ 0 & D_h & 0 \\ 0 & 0 & D_v \end{bmatrix},\ with\ D_\zeta = \beta_\zeta\|V\| + D_0;$$
$$\|V\| = \|q\| \Big/ \theta; \tag{3}$$

Here, D_0 [L^2T^{-1}] is the effective molecular diffusion; $\|\mathbf{q}\|$ is the magnitude of $\mathbf{q}$; and β_ζ [L] is dispersivity along longitudinal ($\zeta = l$) and transverse ($\zeta = h, v$) directions. Eqs. (1)–(3) are then complemented with a set of initial and boundary conditions, as detailed in Section 3.

Borehole information, including head (i.e., water level or pressure), solute concentration, and in some cases fluxes, are typically employed for the purpose of system characterization in terms of, e.g., model parameter estimation. Here, we consider head and concentration to be available from multi-screen monitoring wells, a condition that is very common in practical applications (see, e.g., Post et al., 2007; Konikow et al., 2009; Zhang et al., 2019). In our computational analyses, we represent these boreholes as multi-node monitoring wells. Following Konikow et al. (2009), Zheng (2010), and Xia et al. (2021), and neglecting linear (due to skin effects) and nonlinear (due to turbulent flow) head losses at the well, water level, h_I^w [L], and solute concentration, C_I^w [ML^{-3}], monitored at a multi-node well I are evaluated as

$$h_I^w = \frac{\sum_{i=1}^n b_i K_i h_i}{\sum_{i=1}^n b_i K_i};\quad C_I^w = \frac{\sum_{i=1}^n Q_i C_i}{\sum_{i=1}^n Q_i} \tag{4}$$

where n [–] is the number of nodes in the multi-node well I (i.e., the number of cells according to which the screened interval of the well is discretized); and b_i [L], K_i [LT^{-1}], h_i [L], and C_i [ML^{-3}] are thickness, conductivity, hydraulic head, and concentration associated with the cell of the numerical grid whose centroid corresponds to the i^{th} node in the multi-node well, respectively. Evaluation of Eq. (4) requires computing the flux exchange Q_i [L^2/T] between the i^{th} node of the multi-node well and the aquifer system. Following Konikow et al. (2009), Q_i is assessed by

$$Q_i = a\, b_i K_i\left(h_I^w - h_i\right),\ \ with\ \ a = \frac{2\pi}{\ln(r_0/r_w)}, \tag{5}$$

where $r_0 = 0.14\sqrt{\Delta x_1^2 + \Delta x_2^2}$ [L] and r_w [L] are the effective and actual radius of monitoring well I, respectively, Δx_1 and Δx_2 corresponding to grid spacing along x_1 and x_2 directions. We further note that neglecting linear (due to skin effects) and nonlinear (due to turbulent flow) head losses at the well can lead to computational savings while otherwise introducing a simplified description of the actual flow field in the proximity of the well.

As stated in Section 1, hydraulic heads and/or solute concentrations observed at a long-screened monitoring well can be regarded as information of an integral nature. These are representative of (depth-)averaged system states (e.g., Elci et al., 2001, 2003; Konikow et al., 2009; Zheng, 2010; Zhang et al., 2019), and are here considered as nonlocal information. Elci et al. (2001, 2003) note that data collected at long-screened wells should be used with caution, since spatial distributions of hydraulic heads and/or solute concentrations (as well as estimates of source contaminant locations and plume geometry informed by such (integral/nonlocal) data) can be associated with a considerable degree of ambiguity. We also recall that actual (physical) distances between these types of observations and locations at which model parameters (e.g., conductivity values) need to be assessed cannot be clearly defined. This constitutes an additional layer of difficulty that hampers the applicability of distance-based localization approaches as opposed to their correlation-based counterparts.

2.2. Iterative ensemble smoother

We denote by $\boldsymbol{m} = [m_1, m_2, \ldots, m_P]^T$ the vector (of size P) of log-conductivity, Y, values and by $\boldsymbol{d} = [d_1, d_2, \ldots, d_O]^T$ the vector (of size O) collecting all observations available from multi-node monitoring wells (e.g., heads and concentrations), superscript T denoting transpose.

Considering a collection of realizations (i.e., an ensemble) of size N and assuming that the prior probability density function (pdf) of Y is Gaussian, the objective function to be minimized is expressed as (Luo et al., 2015; Chen and Oliver, 2013):

$$f_{obj}(\boldsymbol{m}) = \frac{1}{2}(g(\boldsymbol{m}) - \boldsymbol{d})^T\underline{\underline{\boldsymbol{C}}}_d^{-1}(g(\boldsymbol{m}) - \boldsymbol{d}) + \frac{1}{2}\gamma(\boldsymbol{m} - \boldsymbol{m}^{pri})^T\underline{\underline{\boldsymbol{C}}}_m^{-1}(\boldsymbol{m} - \boldsymbol{m}^{pri}) \tag{6}$$

Here, $g(\boldsymbol{m}^k)$ is the operator associated with the groundwater flow and transport model; $\boldsymbol{m}^{pri}$ is the prior counterpart of $\mathbf{m}$; γ is a positive scalar which corresponds to the relative weight of the two terms included in Eq. (6); $\underline{\underline{\boldsymbol{C}}}_m$(of size $P \times P$) is the covariance matrix associated with $\boldsymbol{m}^{pri}$; and $\underline{\underline{\boldsymbol{C}}}_d$ (of size $O \times O$) is the covariance matrix of observation errors. For the purpose of our study, observation errors are assumed to be uncorrelated. Thus, $\underline{\underline{\boldsymbol{C}}}_d$ is diagonal, its entries being equal to the variance of observation errors.

The spatial distribution of Y can be estimated upon relying on an iterative ensemble smoother (iES; e.g., Luo et al., 2015; Chen and Oliver, 2013), i.e.,

$$\begin{cases} m^{k+1} = m^k + \underline{\underline{K}}_{Gain}^k\Delta d^k \\ \underline{\underline{K}}_{Gain}^k = \underline{\underline{S}}_m^k\left(\underline{\underline{S}}_d^k\right)^T\left(\underline{\underline{S}}_d^k\left(\underline{\underline{S}}_d^k\right)^T + \gamma^k\underline{\underline{I}}\right)^{-1}, \\ \Delta d^k = g\left(m^k\right) - d \end{cases}$$
$$with\ \gamma^i = \xi^i trace\left(\underline{\underline{S}}_d^i\left(\underline{\underline{S}}_d^i\right)^T\right)\Big/ O \tag{7}$$

Here, superscript k represents the index of the iteration step; $\mathbf{I}$ is the identity matrix (of size $O \times O$); and ξ^k is an adaptive coefficient (Luo

et al., 2015). The latter is employed in the context of the adaptive step size associated with each iteration of the Levenberg-Marquardt (LM) algorithm we consider to assess Eq. (7). This algorithm is first introduced by Levenberg (1944). It is commonly employed to solve non-linear least squares problems in several contexts, including history matching (Li and Reynolds, 2009; Chen and Oliver, 2013; Luo et al., 2015). We set $\xi^0 = 10$ in our showcase application examples (see Section 3) and follow Luo and Bhakta (2020) to update its value for the remaining iteration steps. Matrices $\underline{\underline{\boldsymbol{S}}}_m^k$ (of size $P \times N$) and $\underline{\underline{\boldsymbol{S}}}_d^k$ (of size $O \times N$) are evaluated upon making use of the ensemble of Y fields associated with the k^{th} iteration step and of the corresponding ensemble of model simulation results as

$$\underline{\underline{\boldsymbol{S}}}_m^k = \frac{1}{\sqrt{N-1}}\left[\boldsymbol{m}_1^k - \overline{\boldsymbol{m}}^k, \cdots, \boldsymbol{m}_N^k - \overline{\boldsymbol{m}}^k\right]$$
$$\underline{\underline{\boldsymbol{S}}}_d^k = \frac{1}{\sqrt{N-1}}\left[g\left(\boldsymbol{m}_1^k\right) - g\left(\overline{\boldsymbol{m}}^k\right), \cdots, g\left(\boldsymbol{m}_N^k\right) - g\left(\overline{\boldsymbol{m}}^k\right)\right] \quad , \text{ with } \overline{\boldsymbol{m}}^k = \sum_{j=1}^{N} \boldsymbol{m}_j^k \Big/ N \tag{8}$$

Note that both inner and outer iterations are required in the context of the LM algorithm during optimization. In the numerical analysis illustrated in Sections 3 and 4 we (*i*) set the inner iteration number equal to 5 (see also Luo and Bhakta, 2020) and (*ii*) analyze the convergence of our results upon varying the outer iteration number. We set $(\delta_{k-1} - \delta_k)/\delta_{k-1} \times 100\% \leqslant 10^{-6}$ as a stopping criterion, where

$$\delta_k = \frac{1}{N}\sum_{j=1}^{N}\left\{\left(\boldsymbol{d}_j^k - g\left(\boldsymbol{m}_j^k\right)\right)^T \underline{\underline{\boldsymbol{C}}}_d^{-1}\left(\boldsymbol{d}_j^k - g\left(\boldsymbol{m}_j^k\right)\right)\right\}. \tag{9}$$

2.3. Correlation-based adaptive localization

When considering localization, the first line of Eq. (7) is replaced by

$$\boldsymbol{m}^{k+1} = \boldsymbol{m}^k + \left(\underline{\underline{\boldsymbol{R}}}^k \otimes \underline{\underline{\boldsymbol{K}}}_{Gain}^k\right)\Delta\boldsymbol{d}^k, \tag{10}$$

where "$\otimes$" represents the Hadamard product and $\mathbf{R}^k$ (of size $P \times O$) is the tapering matrix where the entry r_{po}^k is evaluated as (Furrer and Bengtsson, 2007)

$$r_{po}^k = \begin{cases} \dfrac{N}{N+1+\left(\widehat{\rho}_{po}^k\right)^{-2}} & \text{if } \left|\rho_{po}^k\right| \geqslant \omega \\ 0 & \text{if } \left|\rho_{po}^k\right| < \omega \end{cases}, \text{ with } \begin{cases} \widehat{\rho}_{po}^k = \left|\rho_{po}^k\right| - 2\left(1 - \left|\rho_{po}^k\right|^2\right)\Big/\sqrt{N} \\ \omega = 2\Big/\sqrt{N} \end{cases} \tag{14}$$

$$r_{po}^k = \frac{N}{N+1+\left(\widehat{\rho}_{po}^k\right)^{-2}}, \quad \text{with} \quad \widehat{\rho}_{po}^k = \frac{\widehat{c}_{po}^k}{\sqrt{\widehat{c}_{pp}^k \widehat{c}_{oo}^k}} \tag{11}$$

Here, $\widehat{c}_{po}^k$ is the covariance between the p^{th} parameter and the o^{th} observation; $\widehat{c}_{pp}^k$ and $\widehat{c}_{oo}^k$ denote the variances of the p^{th} parameter and the o^{th} observation, respectively; and $\widehat{\rho}_{po}^k$ is the corresponding correlation coefficient. As quantities embedded in Eq. (11) are generally unknown, Furrer and Bengtson (2007) suggest to employ a variant of Eq. (11), hereafter denoted as FBconst, to evaluate r_{po}^k as

$$r_{po}^k = \begin{cases} \dfrac{N}{N+1+\left(\rho_{po}^k\right)^{-2}} & \text{if } \left|\rho_{po}^k\right| \geqslant \omega \\ 0 & \text{if } \left|\rho_{po}^k\right| < \omega \end{cases}, \text{ with } \rho_{po}^k = \frac{c_{po}^k}{\sqrt{c_{pp}^k c_{oo}^k}} \tag{12}$$

Here, c_{po}^k is the sample (evaluated across the ensemble of size N) cross-covariance between m_p^k and the o^{th} outcome associated with $g(\boldsymbol{m}^k)$, i.e., $g_o(\boldsymbol{m}^k)$; c_{pp}^k is the sample variance of m_p^k; c_{oo}^k is the sample variance of $g_o(\boldsymbol{m}^k)$; and $0 \leqslant \omega \leqslant 1$ is a constant threshold for the sample cross-correlation coefficient ρ_{po}^k between m_p^k and $g_o(\boldsymbol{m}^k)$. The main concept underlying Eq. (12) is that the o^{th} observation should not be used to update the p^{th} entry of Y when their (cross-) correlation is negligible. Equation (12) is widely employed in many studies (e.g., Maschio and Schiozer, 2016; Soares et al., 2019; Lacerda et al., 2021 and references therein).

According to Anderson (2003), ρ_{po}^k is asymptotically (i.e. for increasing values of N) described by a Gaussian distribution, i.e., $\mathcal{N}\left(\widehat{\rho}_{po}^k, \left(1 - \left(\widehat{\rho}_{po}^k\right)^2\right)^2/N\right)$. Therefore, the link between $\widehat{\rho}_{po}^k$ and ρ_{po}^k can be (asymptotically) described as

$$\rho_{po}^k = \widehat{\rho}_{po}^k + \varepsilon, \text{ with } \varepsilon \sim \mathcal{N}\left(0, \left(1 - \left(\widehat{\rho}_{po}^k\right)^2\right)^2 \Big/ N\right). \tag{13}$$

Eq. (13) suggests that the variance of the noise ε associated with the (sample) cross-correlation coefficient tends to increase as $\widehat{\rho}_{po}^k$ decreases. We then note that ρ_{po}^k does not vanish when $\widehat{\rho}_{po}^k = 0$ (i.e., m_p^k and $g_o(\boldsymbol{m}^k)$ are uncorrelated) and is equal to ε ($\sim \mathcal{N}(0, 1/N)$). It can then be expected that relying on ρ_{po}^k in Eq. (12) instead of $\widehat{\rho}_{po}^k$ can contribute to deteriorate the effect of localizing $\underline{\underline{\boldsymbol{K}}}_{Gain}^k$. This is especially seen when the strength of ρ_{po}^k (i.e., $\left|\rho_{po}^k\right|$) is much higher than that of $\widehat{\rho}_{po}^k$ (i.e., $\left|\widehat{\rho}_{po}^k\right|$), which leads to overestimating r_{po}^k.

Based on these observations, we propose here a new tapering function, hereafter denoted as FBadap. The latter descends from Eq. (12). It is obtained by considering an adaptive selection of ω and rewriting the tapering function as

Here, the term $2\left(1 - \left|\rho_{po}^k\right|^2\right)/\sqrt{N}$ corresponds to an approximation of twice the standard deviation of ε, i.e., $2\left(1 - \left|\widehat{\rho}_{po}^k\right|^2\right)/\sqrt{N}$. Since $\left|\rho_{po}^k\right| > \left|\widehat{\rho}_{po}^k\right|$ when $\varepsilon > 0$, employing the first row of Eq. (12) might yield excessively large values for the tapering coefficients, thus deteriorating the performance of localization approach. Hendricks Franssen and Kinzelbach (2008) show that enforcing a relatively small damping factor (which is a constant tapering coefficient across diverse parameters and/or observations) on the Kalman gain matrix can help to improve the effectiveness of localization, as embedded in an ensemble Kalman filter. Consistent with these findings, here we employ

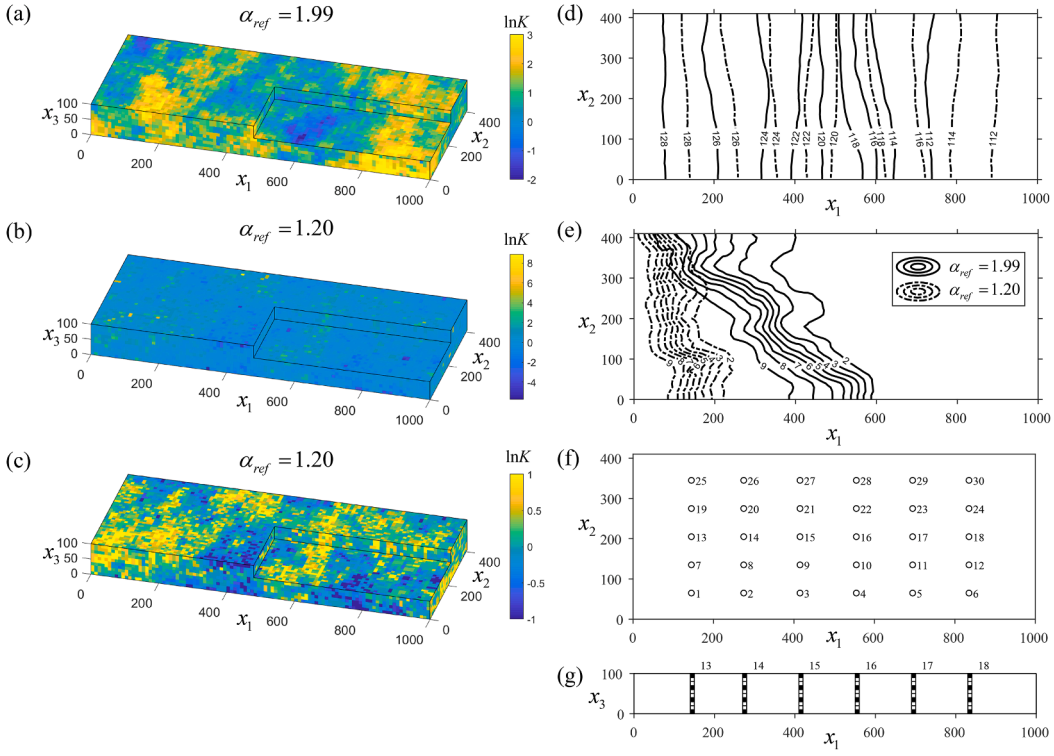


Fig. 1. Reference Y fields with α_{ref} = (a) 1.99 and (b, c corresponding to a color gradation associated with the total and a reduced range of variability, respectively) 1.20; contours of (d) reference heads and (e) reference concentrations evaluated after the 20th time step by considering reference Y fields with α_{ref} = 1.99 (solid black) and 1.20 (dashed black); (f) spatial distribution of 30 monitoring wells (circles); (g) vertical distribution of screens (filled square black) associated with monitoring wells across the transect at x_2 = 205.

$\widehat{\rho}^k_{po} = \left|\rho^k_{po}\right| - 2\left(1 - \left|\rho^k_{po}\right|^2\right)/\sqrt{N}$ in Eq. (14) to obtain smaller values of tapering coefficients than those stemming from Eq. (12). Note that we consider $\omega = 2/\sqrt{N}$ (i.e., twice the standard deviation of ε) when $\widehat{\rho}^k_{po} = 0$. This is related to the observation that one cannot reject the hypothesis that $\rho^k_{po} = 0$ when $\rho^k_{po} < w$. In this case r^k_{po} is set to zero. Otherwise, r^k_{po} is evaluated through the first row of Eq. (14) when the above mentioned hypothesis is rejected with a confidence level of 95 % (i.e., there is high confidence that $\rho^k_{po} \neq 0$).

For the purpose of comparison, we also test the performance of the adaptive localization approach proposed by Luo and Bhakta (2020). The latter is hereafter denoted as LB2020 and is characterized by a tapering function evaluated as

$$r^k_{po} = f_{tap}\left(\frac{1 - \left|\rho^1_{po}\right|}{1 - \sqrt{2(\ln P)/N}}\right) \tag{15}$$

Here, f_{tap} is the commonly used Gaspari-Cohn function with argument $(\cdot)$. It is worth noting that the sample correlation ρ^1_{po} appearing in Eq. (15) is evaluated at the first iteration step, i.e., the tapering function evaluated at the first iteration step remains constant during the remaining steps.

2.4. Generalized sub-Gaussian log-conductivity fields

Applications of data assimilation or history matching approaches to subsurface flow and/or transport scenarios typically rest on the assumption that system attributes (e.g., log-conductivities) be characterized by a (multivariate) Gaussian probability density function (pdf). There is ample evidence that probability distributions and associated statistical moments characterizing a variety of hydrogeological and soil science variables and their spatial increments display distinctive features that are not captured by a typical Gaussian model (see Section 1). The core concept underpinning the Generalized sub-Gaussian (GSG) model we consider in our study (see also Riva et al., 2015a, b) is that the departure of the distribution of a variable and its two-point increments from the Gaussian one is driven by the action of a (spatially uncorrelated) subordinator on an otherwise spatially correlated Gaussian random field. Following Riva et al., (2015a), we model the log-conductivity field, $Y(\mathbf{x}) = \ln K(\mathbf{x})$, as a GSG random field

$$Y'(x) = U(x)G(x) \;\; = y(x) \;\; (y) \tag{16}$$

Here, angular brackets denote expectation; and $G(\boldsymbol{x})$ and $U(\boldsymbol{x})$ correspond to a zero-mean correlated Gaussian random field and a (non-negative) uncorrelated subordinator independent of G, respectively. While a variety of formulations can be employed to characterize the subordinator (Siena et al., 2020), we consider (see Section 3) $U(\boldsymbol{x})$ to be characterized by a lognormal pdf, i.e., $U(\boldsymbol{x}) \sim \ln\mathcal{N}\left[0, (2-\alpha)^2\right]$, where $\alpha < 2$ is the parameter controlling the strength of the departure of $Y'(\boldsymbol{x})$ from the Gaussian distribution. Such a modeling choice has also been adopted in previous analytical and numerical studies (Riva et al., 2017; Libera et al., 2017; Sole-Mari et al., 2021). Note that $Y'(\boldsymbol{x})$ tends to become Gaussian in the limit for $\alpha \to 2$ (see also Riva et al., (2015a) for additional details).

We then consider $G(\boldsymbol{x})$ to be a stationary Gaussian random field constituting a truncated fractional Brownian motion, tfBm. Relying on this formulation is fully consistent with the theoretical geostatistical framework introduced by Di Federico and Neuman (1997) and Neuman and Di Federico (2003) and further extended by Neuman et al. (2008a, b). According to the latter, a Truncated Power Variogram (TPV) model enables one to view a given system attribute as a continuous, overlapping, multiscale hierarchy of stationary random fields (modes), each

Table 1
Input quantities for the generation of the initial collection/ensemble of Y fields.

Symbol	Unit	Shape parameter			
		$\alpha_0 = 1.20$	$\alpha_0 = 1.50$	$\alpha_0 = 1.80$	$\alpha_0 = 1.99$
λ_l	[L]	10			
H	[–]	0.35			
A	[L^{-2H}]	1.77×10^{-3}	5.19×10^{-3}	9.31×10^{-3}	1.04×10^{-2}
λ_u	[L]	881.526	588.250	472.179	452.781

having an exponential (or Gaussian) variogram characterized by a single integral scale. We can then express the variogram of $G(\boldsymbol{x})$, γ_G, as

$$\gamma_G(s;\lambda_u,\lambda_l) = \gamma(s;\lambda_u) - \gamma(s;\lambda_l), \tag{17}$$

where one can write (for exponential modes)

$$\frac{\gamma(s;\lambda_m)}{\sigma^2(\lambda_m)} = 1 - \exp\left(-\frac{s}{\lambda_m}\right) + \left(\frac{s}{\lambda_m}\right)^{2H}\Gamma\left(1-2H;\frac{s}{\lambda_m}\right); \qquad \sigma^2(\lambda_m) = A\frac{\lambda_m^{2H}}{2H}, \quad m = l,u; \tag{18}$$

Here, $s = \|\mathbf{s}\| = \sqrt{s_{x_1}^2 + s_{x_2}^2/a_{x_2}^2 + s_{x_3}^2/a_{x_3}^2}$ [L] is separation distance (lag) in an equivalent isotropic domain, $s_{x_i}^2$ (with $i = 1, 2, 3$) being the separation distance along x_i-axis in the original (anisotropic) domain; A [L^{-2H}] is a coefficient; H [–] is a Hurst scaling coefficient, with $0 < H < 0.5$; λ_l and λ_u [L] are lower and upper cutoff scales, proportional to the measurement scale associated with available data and to the scale of the sampling domain, respectively; and $\Gamma(\cdot;\cdot)$ is the incomplete gamma function. Based on Eqs. (16) and (17), the variance (σ_Y^2), variogram (γ_Y), and integral scale ($I_{Y_{xi}}$; $i = 1, 2, 3$) of Y for a lognormal subordinator are (Riva et al., 2015b)

$$\sigma_Y^2(\lambda_u,\lambda_l,\alpha) = e^{2(2-\alpha)^2}\sigma_G^2, \quad with \quad \sigma_G^2(\lambda_u,\lambda_l) = \frac{A}{2H}\left(\lambda_u^{2H} - \lambda_l^{2H}\right) \tag{19}$$

$$\gamma_Y(s,\lambda_u,\lambda_l,\alpha) = e^{(2-\alpha)^2}\left[\sigma_G^2\left(e^{(2-\alpha)^2} - 1\right) + \gamma_G\right] \tag{20}$$

$$I_{Y_{xi}}(\lambda_u,\lambda_l,\alpha) = e^{-(2-\alpha)^2}I_{G_{xi}}, \; with \; I_{G_{x1}} = \frac{2H}{1+2H}\frac{\lambda_u^{1+2H} - \lambda_l^{1+2H}}{\lambda_u^{2H} - \lambda_l^{2H}}; \qquad I_{G_{x2}} = I_{G_{x1}}a_{x_2}; \quad I_{G_{x3}} = I_{G_{x1}}a_{x_3}; \tag{21}$$

Here, a_{x_2} and a_{x_3} are anisotropy ratios (i.e., ratios between integral scales associated with the x_2 and x_3 directions, respectively, and their counterpart along direction x_1).

3. Computational analyses

To ensure transparent assessment and comparison of the approaches analyzed, we rest on a suite of computational settings that we take as reference. We consider heterogeneous log-conductivity fields across a three-dimensional domain of size 1000 × 410 × 100 (see Fig. 1a). All quantities are hereafter given in consistent (length/mass/time) units. The domain is discretized through a regular mesh formed by blocks of uniform size equal to 10 × 10 × 10 (thus yielding $P = 41{,}000$). Each monitoring well is equipped with four screens (see Fig. 1g). This leads to $b_i = 10$ (with $i = 1, 2, 3, 4$). We consider the reference Y fields and the initial collection of realizations of Y to be characterized by (*a*) $\alpha = \alpha_{ref} =$ 1.99 or 1.20, to encompass two markedly different scenarios, respectively representing settings associated with a near-Gaussian distribution and a strong departure from it; (*b*) $I_{Y_{x1}} = 100$, $I_{Y_{x2}} = 200$ (i.e., $a_{x_2} = 2$), $I_{Y_{x3}} = 50$ (i.e., $a_{x_3} = 0.5$); (*c*) $\langle Y \rangle = 0.5$; (d) $\sigma_Y^2 = 1.0$; and (*e*) $H = 0.35$. We recall that H reflects the degree of persistence (that increases with H) of the field, the intermediate value selected being consistent with estimates obtained by analyzing log-conductivity data at field scales (Neuman, 1990, 1995).

We analyze the impact on the results rendered by considering the initial collection of random fields of Y to be generated according to various degrees of departure from Gaussian. We do so by generating these according to four values of α, i.e., $\alpha = \alpha_0 = 1.20$, 1.50, 1.80, and 1.99, corresponding to strongly, moderately, or slightly non-Gaussian and near-Gaussian Y fields, respectively. The complete list of parameters employed to generate the initial Y fields (selected to comply with the constrains (*b*)–(*e*) illustrated above and evaluated on the basis of Eqs. (19)–(21)) is included in Table 1. Note that we set $\lambda_l = 10$ in all test cases, a value that is consistent with the grid size and with the physical meaning of λ_l. Details about the random field generation scheme are illustrated in Riva et al., (2015b).

Fig. 1a and b depict the two above mentioned $Y(\mathbf{x})$ random reference fields characterized by $\alpha_{ref} = 1.99$ and 1.20. Fig. 1c complements the description by providing a depiction of Fig. 1b focusing on a reduced range of variability of Y values (generated for $\alpha_{ref} = 1.20$; see the corresponding color gradation).

Groundwater flow is driven by a mean uniform flow (from left, $x_1 =$ 0, to right, $x_1 = 1000$, where fixed heads are set to 130 and 110, respectively, yielding a mean uniform gradient equal to 2 %, the remaining domain sides being impervious). Non-reactive transport is solved by setting a constant concentration $C_B = 10$ at $x_1 = 0$ and layers 5 and 6. The remaining sides of the domain are set as no-dispersive-mass-flux boundaries. Initial head across the system is designed as a linear function of x_2 with a solution constrained by fixed heads. A spatially uniform initial solute concentration of $C_I = 1.0$ is considered. The remaining parameters, which are considered to be deterministic and

Table 2
Overview of the test cases (TCs) analyzed. All TCs are characterized by mean and variance of the Y reference field equal to 0.5 and 1.0, respectively; $\langle Y \rangle$ and σ_Y^2 are denoted as the mean and variance of initial ensemble of Y field respectively.

Group												
1	Test Case	TC1, TC1#	TC2, TC2#	TC3, TC3#	*Group* 2	Test Case	TC1*1	TC1*2	TC1*3	TC1#1	TC1#2	TC1#3
	Size of the ensemble N	100	50	500		α_0	1.20	1.50	1.80	1.50	1.80	1.99
	Localization approaches	FBconst, FBadap, LB2020	FBconst, FBadap	FBconst, FBadap		Additional illustrations	$\alpha_{ref} = \alpha_0 = 1.99$ for TCs 1*1–1*3; $\alpha_{ref} = \alpha_0 = 1.20$ for TCs 1#1–1#3; $\langle Y \rangle = 0.5$, $\sigma_Y^2 = 1.0$, and $N = 100$ for *group* 2; both FBadap and LB2020 implemented for every TC in *group* 2					
	Additional illustrations	$\alpha_{ref} = \alpha_0 = 1.99$ for TCs 1–3; $\alpha_{ref} = \alpha_0 = 1.20$ for TCs 1#-3#; $\langle Y \rangle = 0.5$ and $\sigma_Y^2 = 1.0$ for *group* 1										

Group 3	Test Case	TC1@1	TC1@2	TC1@3	TC1@4
	$\langle Y \rangle$	2.5	1.5	−0.5	−1.5
	Additional illustrations	$\alpha_{ref} = \alpha_0 = 1.99$ for TCs 1@1–1@4; $\sigma_Y^2 = 1.0$ and $N = 100$ for *group* 3; both FBadap and LB2020 implemented for every TC in *group* 3			

Group 4	Test Case	TC1$1	TC1$2	TC1$3	TC1$4
	σ_Y^2	2.0	1.5	0.8	0.5
	Additional illustrations	$\alpha_{ref} = \alpha_0 = 1.99$ for TCs 1@1–1@4; $\langle Y \rangle = 0.5$ and $N = 100$ for *group* 4; both FBadap and LB2020 implemented for every TC in *group* 4			

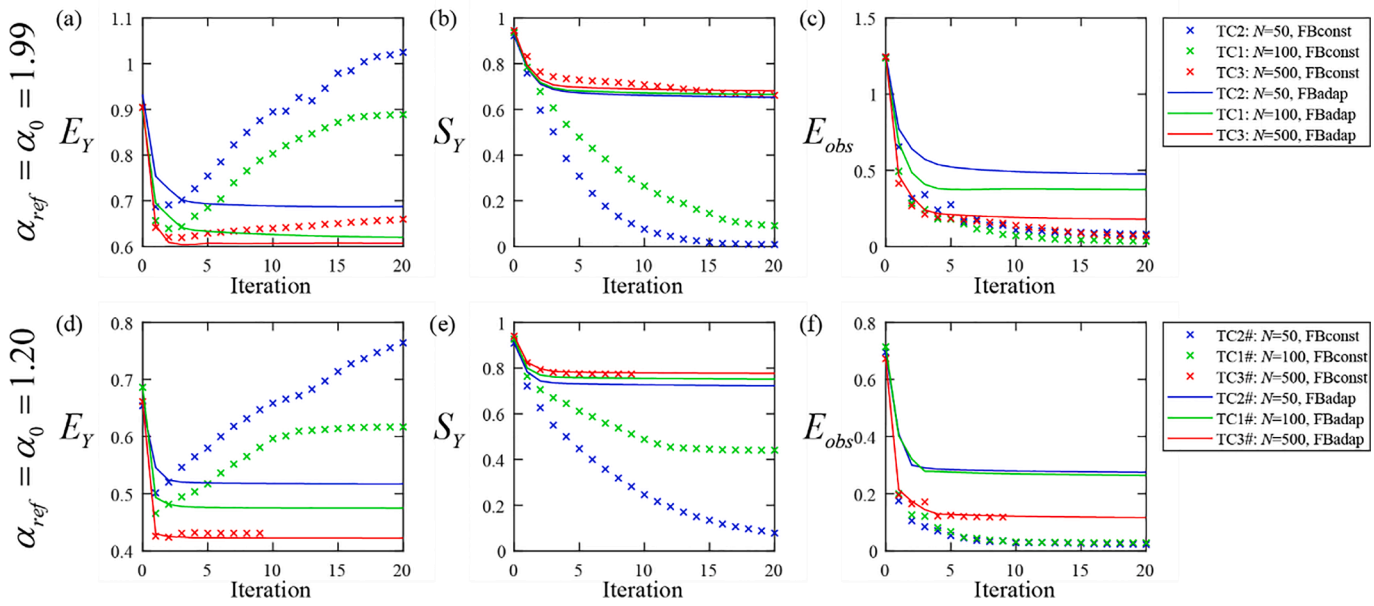


Fig. 2. Values of (a, d) E_Y, (b, e) S_Y, and (c, f) E_{obs} versus the number of outer iterations considering various ensemble sizes (with $N = 50$, 100, and 500) and $\alpha_{ref} = \alpha_0 = 1.99; 1.20$. Results are related to FBconst (crosses) and FBadap (solid curves) for (a, b, c) TC2 (blue), TC1 (green), and TC3 (red) and (d, e, f) TC2# (blue), TC1# (green), and TC3# (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

uniformly distributed in space across all simulations, are set as $S_s = 10^{-3}$, $\theta = 0.3$, $\beta_l = 10$; $\beta_h = \beta_v = 1$, $D_0 = 0$, $r_0 = 1.97$, and $r_w = 0.1$.

The numerical simulation period extends across a temporal window discretized onto 30 intervals, each of uniform size equal to 100. Numerical solutions of head and concentration fields are obtained through the well-known and broadly tested computational codes MODFLOW (Harbaugh, 2005) and MT3DMS (Zheng, 2009, 2010), respectively. We make use of the Multi-Node Well package MNW2 to simulate wells (Konikow et al., 2009). As an example, Fig. 1d and e respectively depict head and concentration isolines evaluated across layer 1 at the 20th time interval for $\alpha_{ref} = 1.99$ (solid curves) and 1.20 (dashed curves).

Reference observations/data used to estimate Y fields are hydraulic head and concentration values evaluated at 30 multi-node monitoring wells (see circles in Fig. 1f), leading to $O = 30$ (number of monitoring wells) $\times$ 30 (number of monitored times) $\times$ 2 (number of monitored quantities, i.e., head and concentration) = 1800. Four screens (i.e., $n = 4$ in Eq. (4)) are installed in each monitoring well at locations corresponding to the first, fourth, seventh, and tenth layer of the numerical model (see black filled squares in Fig. 1g).

To mimic the impact of measurement errors, head and concentration data are corrupted by adding a Gaussian white noise with standard deviation equal to 0.01 (covering a broad range of values of coefficient of variations, i.e., about 0.1 %–1.0 %, depending on the location within the system).

We organize our exemplary settings according to the following four groups, for a total of 20 test cases (TCs), listed in Table 2.

Group 1. It includes six TCs, allowing us to compare the performance of FBconst (Eq. (12), with $w = 0.1$), FBadap (Eq. (14)) and LB2020 (Eq. (15)) upon varying the ensemble size N and the value of α_{ref}, when the latter is known (i.e., $\alpha_0 = \alpha_{ref}$).

Group 2. It includes six TCs, enabling us to explore the impact of the strength of the departure of the (otherwise unknown) pdf of Y from a Gaussian model (i.e., $\alpha_0 \neq \alpha_{ref}$) on Y estimates stemming from FBadap and LB2020.

Each of the groups denoted as **Group 3** and **Group 4** include four TCs, allowing for the assessment of the effect of the mean and variance, respectively, of the initial ensemble of Y on conductivity estimates obtained upon relying on FBadap or LB2020.

We recall that randomly heterogenous conductivity fields of aquifer systems can be represented through a unique distribution that can be sometimes characterized by a large variance (see, e.g., Winter and Tartakovsky, 2003). This corresponds to a conceptual framework according to which conductivity values, which might be associated with diverse geomaterials, are homogenized within a unique system. Otherwise, one can view the domain as structured according to various regions, each associated with a given geomaterial and related (spatially heterogeneous) conductivity (see e.g., Winter et al., 2002; Winter and Tartakovsky, 2000, 2002, 2003; Bianchi Janetti et al., 2019 and references therein). In this sense, the distribution associated with conductivity within each of these regions is typically characterized by low to mild variance (Winter et al., 2002; Winter and Tartakovsky, 2002, 2003 and references therein). The range of values we consider here for the variance of log-conductivity can then be viewed as corresponding to the type of internal variability related to a given geologic unit.

As a metric upon which we appraise the accuracy of conductivity estimates, we consider the average (a) absolute error, E_Y, and (b) estimate of the standard deviation, S_Y, of Y. These are defined as

$$E_Y = \frac{1}{P}\sum_{i=1}^{P} \left| \langle Y_i \rangle^{est} - Y_i^{ref} \right| S_Y = \sqrt{\frac{1}{P}\sum_{i=1}^{P} \left(\sigma^2_{Y,i}\right)^{est}}, \tag{22}$$

where $\langle Y_i \rangle^{est}$, $\left(\sigma^2_{Y,i}\right)^{est}$ and Y_i^{ref} indicate estimated (ensemble) mean and variance as well as reference value of Y at the i^{th} cell of the numerical mesh, respectively. We then consider the average absolute difference

$$E_{obs} = \frac{1}{O}\sum_{i=1}^{O} \left| \langle d_i \rangle^{up} - d_i^{ref} \right| \tag{23}$$

to quantify data mismatch, where $\langle d_i \rangle^{up}$ and d_i^{ref} correspond to the (updated) result of the simulation process and its reference observed counterpart at the i^{th} sampled location, respectively.

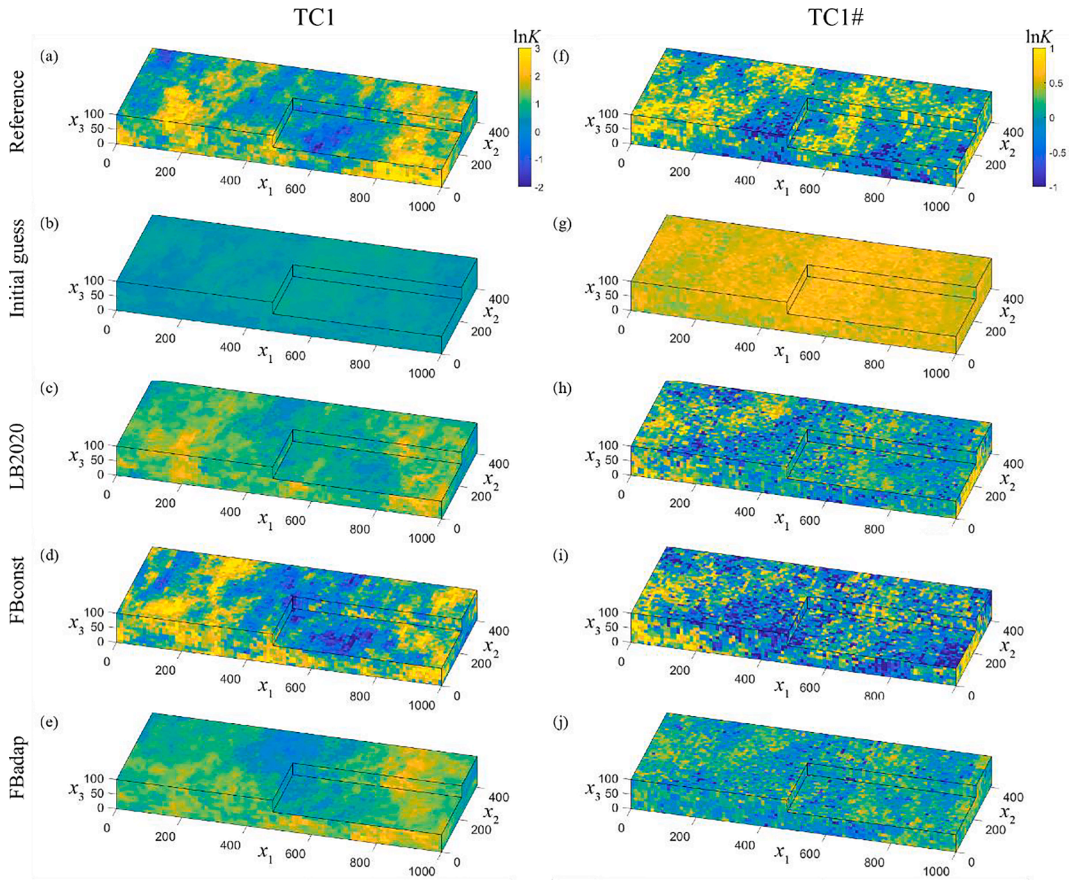


Fig. 3. Reference (top row; a, f) and initial guess (second row; b, g) of the $Y = \ln K$ field together with final estimated (ensemble) mean of Y fields obtained through (c, h) LB2020, (d, i) FBconst, and (e, j) FBadap for TC1 (left column) and TC1# (right).

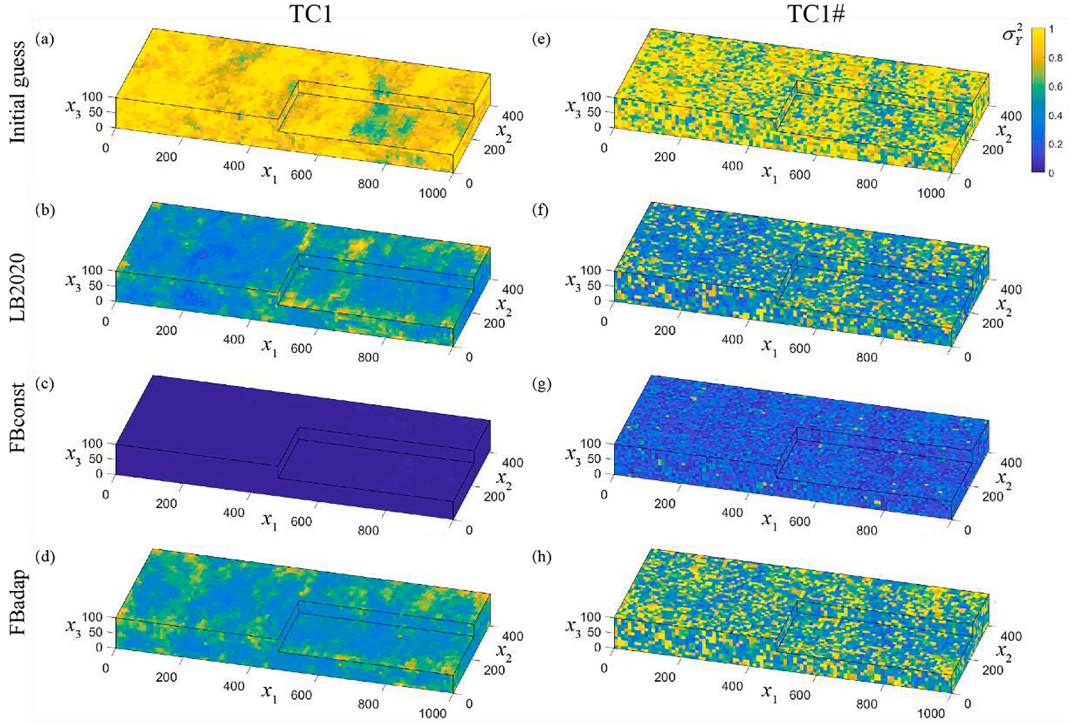


Fig. 4. Initial guesses (top row) and final estimates of the spatial distribution of Y variance obtained through LB2020 (second row), FBconst (third row), and FBadap (fourth row) for TC1 (left column) and TC1# (right).

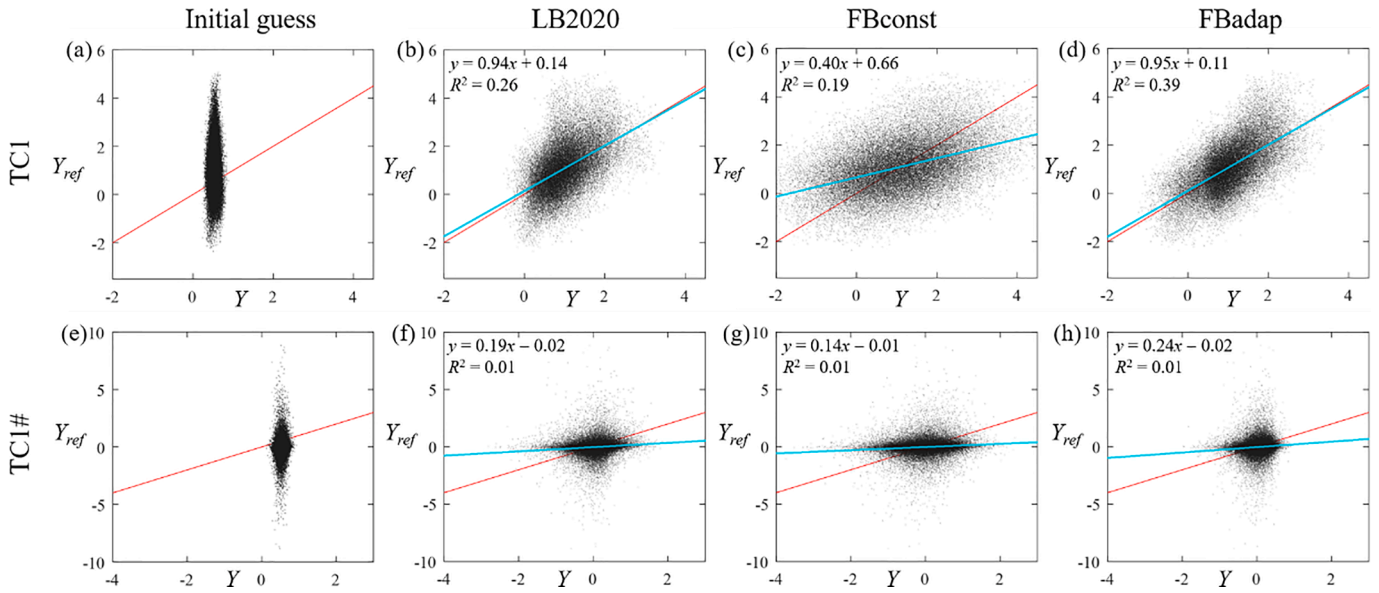


Fig. 5. Scatter plots of reference Y values against Y values of the initial collection/ensemble of realizations (first column) and final estimated values obtained through LB2020 (second column), FBconst (third column), and FBadap (fourth column) for TC1 (top row) and TC1# (bottom row). The 45° line is depicted in red, as a reference; light blue lines are based on linear regression (corresponding values of R^2 are also included). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Results and discussion

4.1. Effect of ensemble size (Group 1)

Fig. 2 depicts E_Y (Fig. 2a, d), S_Y (Fig. 2b, e), and E_{obs} (Fig. 2c, f) versus the number of outer iterations considering various ensemble sizes (N) and $\alpha_{ref} = \alpha_0 = 1.99; 1.20$.

The value of E_Y at the end of the iteration procedure based on FBconst decreases as N increases. It is also worth noting that values of E_Y for the small ensemble size (i.e., $N = 50$; TC2 and TC2#) decrease rapidly across the first two outer iterations and then gradually increase across the remaining ones. The strength of this phenomenon is reduced as N increases. Inspection on Fig. 2b and e reveals that S_Y increases with N. This behavior is related to the use of a constant threshold (w) and the approximation of the correlation coefficient through its sample counterpart. These two elements jointly contribute to a failure in adequately filtering out spurious correlations. The latter (a) can dampen the variability across the ensemble of Y realizations and (b) tend to decrease as N increases. Note that values of E_Y, S_Y, and E_{obs} for TC3 are not depicted after the 9th iteration step, because the stopping criterion has been met.

Considering FBadap (see solid curves in Fig. 2), one can note that E_Y and S_Y decrease with the iteration step and (in general) with increasing N. Moreover, values of S_Y display only minute (almost negligible) variations with N. These results support the ability of FBadap to cope with diverse ensemble sizes in the context of the estimation workflow.

A joint analysis of E_Y, S_Y, and E_{obs} (Fig. 2c and f) reveals that FBconst leads to an overall lower estimation accuracy with respect to conductivity than FBadap, while otherwise yielding higher estimation accuracy with reference to hydraulic heads. This phenomenon is linked to the tendency of FBconst to give rise to overfitting due to the presence of high spurious correlations.

Fig. 3 provides visual depictions of the reference (first row) and initial mean (second row) Y fields, together with the estimated (ensemble) mean of Y fields obtained through LB2020 (third row), FBconst (fourth row), and FBadap (fifth row). Results related to TC1 (left column, $\alpha_{ref} = 1.99$) and TC1# (right column, $\alpha_{ref} = 1.20$) are depicted. We recall that the corresponding E_Y values are depicted in Fig. 2(a, d) and refer to Fig. S1 (see Supplementary Material) where absolute differences between the estimated mean and reference Y field are depicted.

Table 3
Values of E_Y, S_Y, and E_{obs} at the end of the assimilation procedure obtained through FBadap with diverse values of ω ($\omega = 1/\sqrt{N}$, $2/\sqrt{N}$, $3/\sqrt{N}$, and $4/\sqrt{N}$) for TC1 and TC1#.

Test Case	Metrics	ω			
		$1/\sqrt{N}$	$2/\sqrt{N}$	$3/\sqrt{N}$	$4/\sqrt{N}$
TC1	E_Y	0.70	0.62	0.65	0.72
	S_Y	0.35	0.67	0.72	0.93
	E_{obs}	0.10	0.37	0.45	0.73
TC1#	E_Y	0.58	0.48	0.49	0.57
	S_Y	0.59	0.75	0.78	0.83
	E_{obs}	0.06	0.26	0.39	0.51

These results are also complemented by Figs. S2 and S3 (see Supplementary Material), depicting the mean absolute difference between each Y fields of the ensemble and their reference counterpart and the corresponding standard deviation, respectively. Fig. 4 depicts the initial (first row) and final estimated spatial distributions of the variance of Y obtained via LB2020 (second row), FBconst (third row), and FBadap (fourth row) for TC1 (left column) and TC1# (right column). Values of Y variances estimated through FBconst (Fig. 4c, g) are clearly lower than those stemming from FBadap (Fig. 4d, h) and LB2020 (Fig. 4b, f), and are associated with values close to zero. Spatial distributions of estimated Y variance obtained by LB2020 and FBadap are visually similar, overall slightly higher values being associated with FBadap.

To assist quantitative appraisal of the results illustrated above, Fig. 5 depicts scatter plots of reference Y values against Y values of the initial collection of realizations (first column) and final estimated values obtained through LB2020 (second column), FBconst (third column), and FBadap (fourth column) for TC1 (top row) and TC1# (bottom row). The 45° line is depicted in red, as a reference, light blue lines being based on linear regression (corresponding values of the coefficient of determination, R^2, are also included). When considering TC1, note that the regression line obtained on the basis of FBadap results is associated with the highest R^2 value, thus suggesting that FBadap provides the overall highest accuracy with reference to Y estimates when a near-Gaussian Y

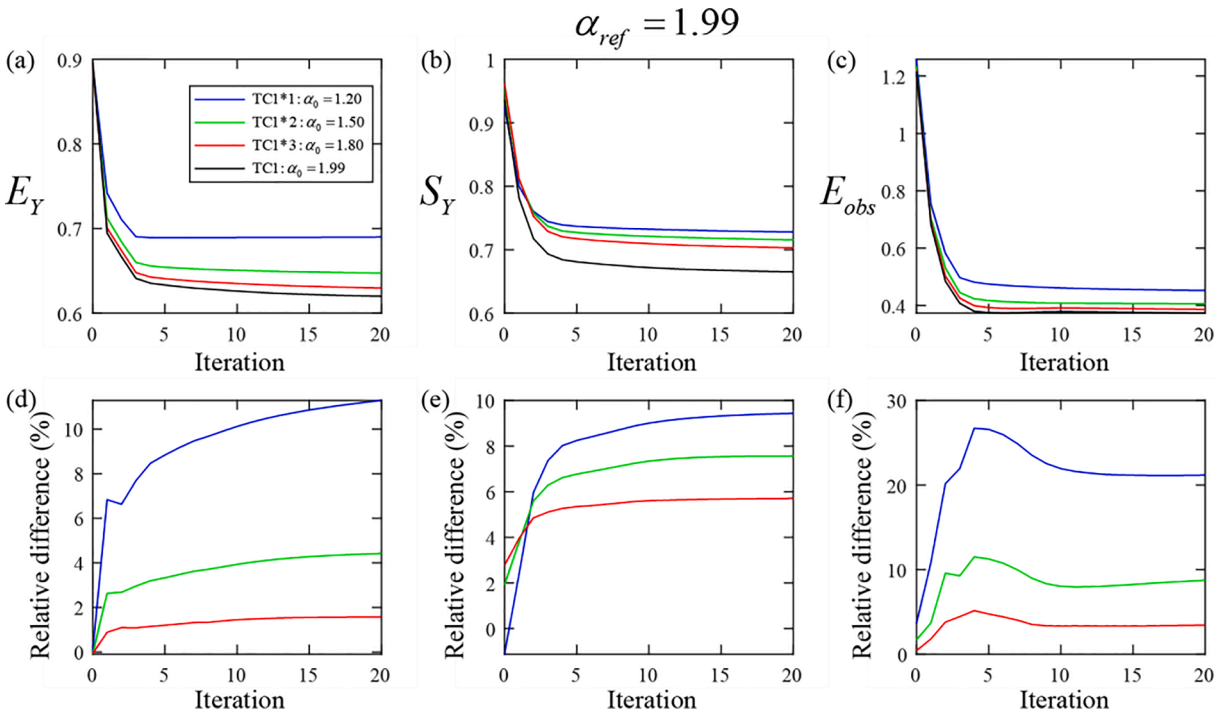


Fig. 6. Variation of (a) E_Y, (b) S_Y, and (c) E_{obs} obtained by FBadap across the outer iterations for TC1*1 (blue), TC1*2 (green), TC1*3 (red), and TC1 (black). Relative (percent) difference between (d) E_Y, (e) S_Y, and (f) E_{obs} obtained for TC1*1 (blue curves), TC1*2 (green), and TC1*3 (red) with respect to their counterparts associated with TC1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

field is considered. Results associated with the strongly non-Gaussian system (TC1#) suggest that additional research efforts are required to obtain reliable estimates of GSG random Y fields on the basis of information of the kind we consider.

The results illustrated above are complemented by Fig. S4 (see Supplementary Material). The latter depicts values of E_Y (Fig. S4a, d), S_Y (Fig. S4b, e), and E_{obs} (Fig. S4c, f) obtained through FBadap upon varying the number of outer iterations and diverse values of ω (i.e., $\omega = 1/\sqrt{N}$, $2/\sqrt{N}$, $3/\sqrt{N}$, and $4/\sqrt{N}$) for TC1 and TC1#. Table 3 lists corresponding values of E_Y, S_Y, and E_{obs} at the end of the assimilation

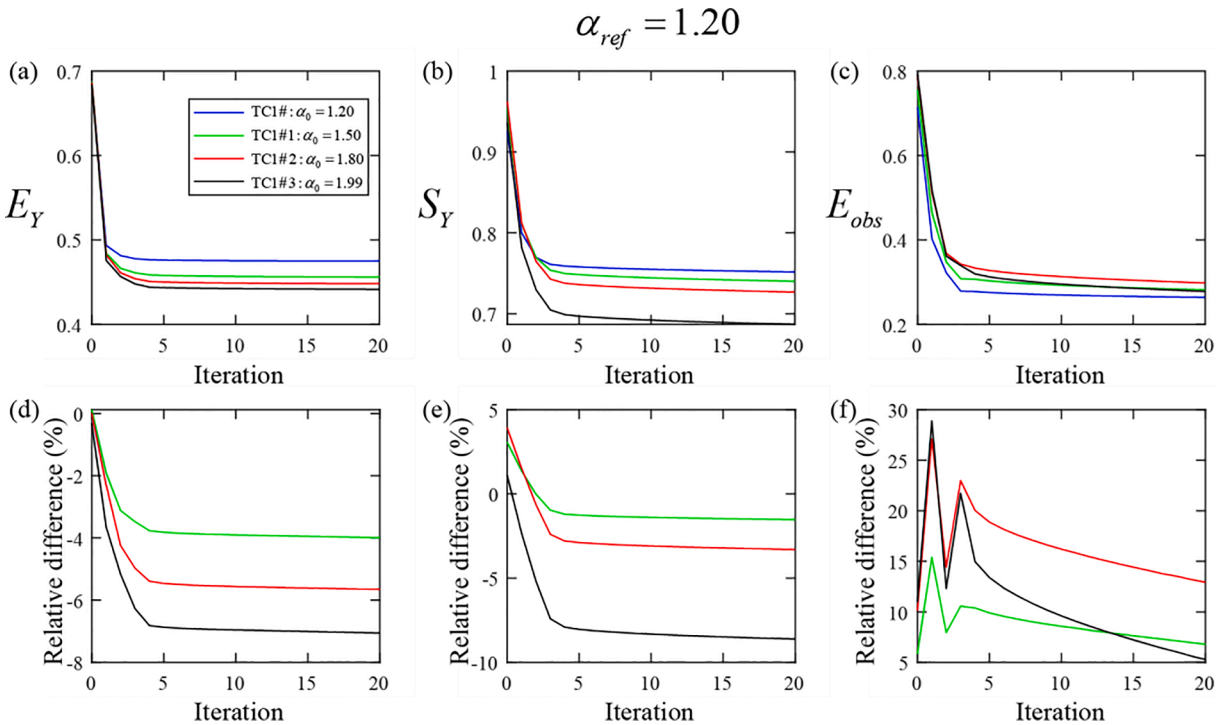


Fig. 7. Variation of (a) E_Y, (b) S_Y, and (c) E_{obs} obtained by FBadap across the outer iterations for TC1# (blue), TC1#1 (green), TC1#2 (red), and TC1#3 (black). Relative (percent) difference between (d) E_Y, (e) S_Y, and (f) E_{obs} obtained for TC1#1 (green curves), TC1#2 (red), and TC1#3 (black) with respect to their counterparts associated with TC1#. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

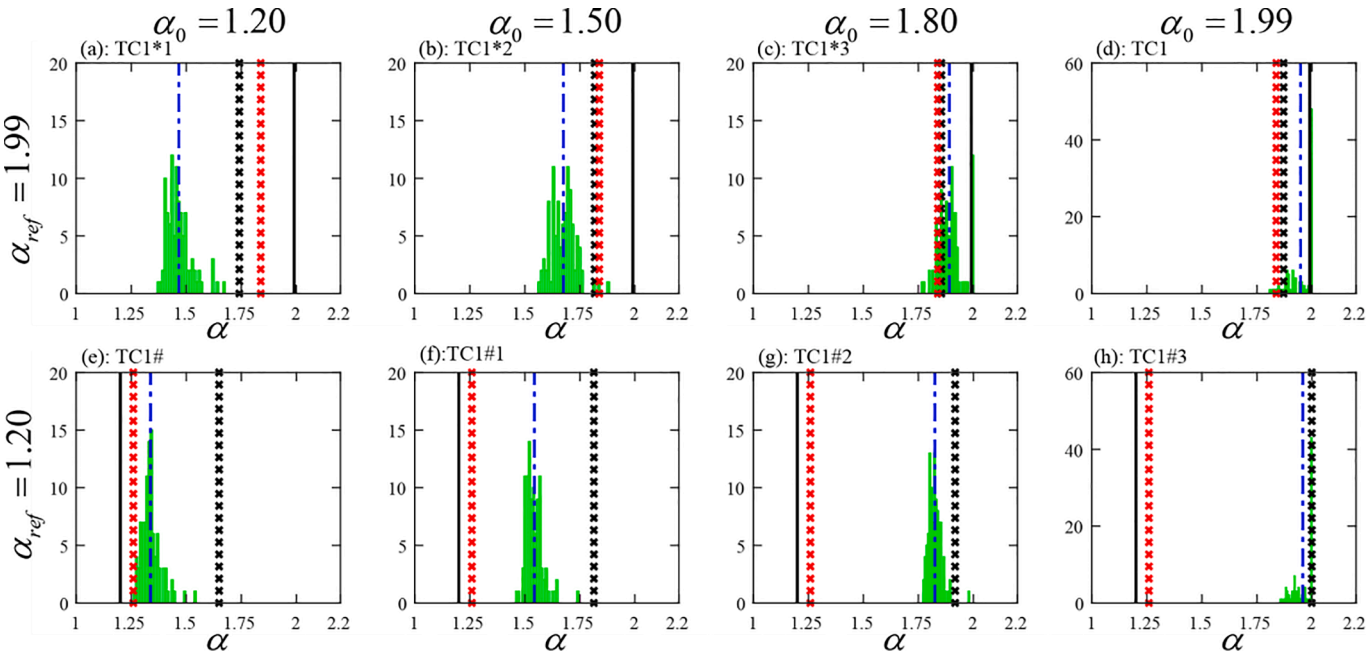


Fig. 8. Histogram (green bars) and mean values (dash-dot blue lines) of the values of α estimated across the 100 realizations of the final ensemble for all TCs of Group 2 and TC1 (Group 1); (a) TC1*1, (b) TC1*2, (c) TC1*3, (d) TC1, (e) TC1#, (f) TC1#1, (g) TC1#2, and (h) TC1#3. Estimated values for α associated with the mean Y fields (black crosses) of the final ensembles, Y reference fields (i.e., α_{ref}, red crosses) and their corresponding theoretical values (black solid) are also depicted. Estimates are obtained through the application of Method A of Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

procedure. When considering $\omega = 1/\sqrt{N}$, E_Y decreases at first and then increases. Otherwise, S_Y and E_{obs} consistently decrease while increasing the number of outer iterations. This phenomenon does not appear for the other values of ω considered. Values of E_Y, S_Y, and E_{obs} at the end of the assimilation procedure consistently increase when ω increases from $2/\sqrt{N}$ to $4/\sqrt{N}$. This is related to the observation that larger values of ω can filter out more spurious correlations while to some extent dampening the effect of some otherwise informative data of hydraulic head and solute concentration.

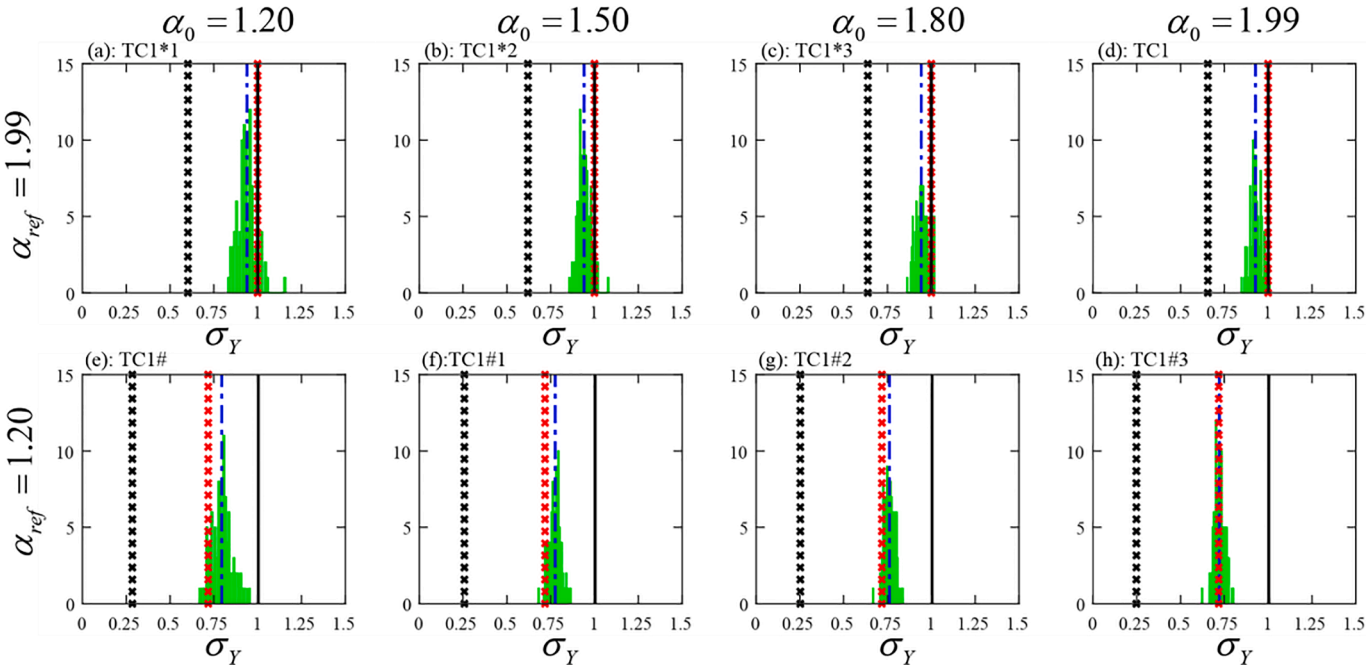


Fig. 9. Histogram (green bars) and mean values (dash-dot blue lines) of the values of σ_Y estimated across the 100 realizations of the final ensemble for all TCs of Group 2 and TC1 (Group 1); (a) TC1*1, (b) TC1*2, (c) TC1*3, (d) TC1, (e) TC1#, (f) TC1#1, (g) TC1#2, and (h) TC1#3. Estimated values for α associated with the mean Y fields (black crosses) of the final ensembles, Y reference fields (red crosses) and their corresponding theoretical values (black solid) are also depicted. Estimates are obtained through the application of Method A of Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Values of E_Y, S_Y, and E_{obs} obtained at the end of the assimilation procedure through FBadap and LB2020 for all TCs of Group 2 and TC1 (from Group 1). Lowest values are indicated in bold.

Shape parameter		FBadap			LB2020		
$\alpha_{ref} = 1.99$		E_Y	S_Y	E_{obs}	E_Y	S_Y	E_{obs}
$\alpha_0 =$	1.20 (TC1*1)	0.69	0.73	0.45	0.79	0.62	0.49
	1.50 (TC1*2)	0.65	0.72	0.41	0.70	0.66	0.49
	1.80 (TC1*3)	0.63	0.70	0.39	0.68	0.69	0.49
	1.99 (TC1)	**0.62**	**0.67**	**0.37**	**0.67**	**0.66**	**0.47**
$\alpha_{ref} = 1.20$		E_Y	S_Y	E_{obs}	E_Y	S_Y	E_{obs}
$\alpha_0 =$	1.20 (TC1#)	0.47	0.75	**0.26**	0.53	0.69	**0.13**
	1.50 (TC1#1)	0.46	0.74	0.28	0.49	0.70	0.16
	1.80 (TC1#2)	0.45	0.73	0.29	0.48	0.71	0.18
	1.99 (TC1#3)	**0.44**	**0.69**	0.28	**0.47**	**0.68**	0.18

4.2. *Effect of lack of knowledge of the shape of the probability density function of the initial ensemble (Group 2)*

Fig. 6 depicts the way E_Y (Fig. 6a), S_Y (Fig. 6b), and E_{obs} (Fig. 6c) obtained by FBadap vary across the outer iterations for TC1*1 (blue), TC1*2 (green), TC1*3 (red), and TC1 (black). Relative differences between the values of E_Y (Fig. 6d), S_Y (Fig. 6e), and E_{obs} (Fig. 6f) for TC1*1 (blue), TC1*2 (green), and TC1*3 (red) and their counterparts related to TC1 are also depicted. Values of E_Y and S_Y increase with the difference between α_{ref} and α_0. The highest relative differences of E_Y, S_Y, and E_{obs} values (as compared against TC1) correspond to TC1*1. At the end of the optimization procedure, these are larger than 10 %, 8 %, and 20 %, respectively. Note that for TC1*3 (i.e., α= 1.80) all metrics analyzed are close to their counterparts evaluated for TC1. In this case, relative differences of E_Y, S_Y and E_{obs} values at the end of the optimization procedure being smaller than 2 %, 6 %, and 5 %, respectively. This is partially consistent with previous observations (Riva et al., 2017), documenting that differences between moments related to key flow and transport quantities associated with GSG and Gaussian log-conductivity fields become virtually unnoticeable for$\alpha \geq 1.80$.

As summarized above, accuracies of Y and state variable estimates improve as the value of α_0 tends to $\alpha_{ref} = 1.99$. This is caused by the joint effect of two elements, corresponding to (*a*) the difference between Y values in the reference field and their mean counterparts evaluated across the (finite) collection of initial guesses (which persists even as the reference value of α is perfectly known) and (*b*) the departure of the initial ensemble of Y from a Gaussian behavior. The latter effect is consistent with the observation that a markedly non-Gaussian behavior of the initial ensemble of Y is associated with a violation of the Gaussianity assumption underlying the formulation of an iES (see Chen and Oliver, 2013). We further comment that the use of an iterative normal-score ensemble smoother (see Li et al., 2018) hardly improves the accuracy of conductivity estimates when considering a small ensemble of, e.g., 100 realizations. This is possibly related to nonnegligible

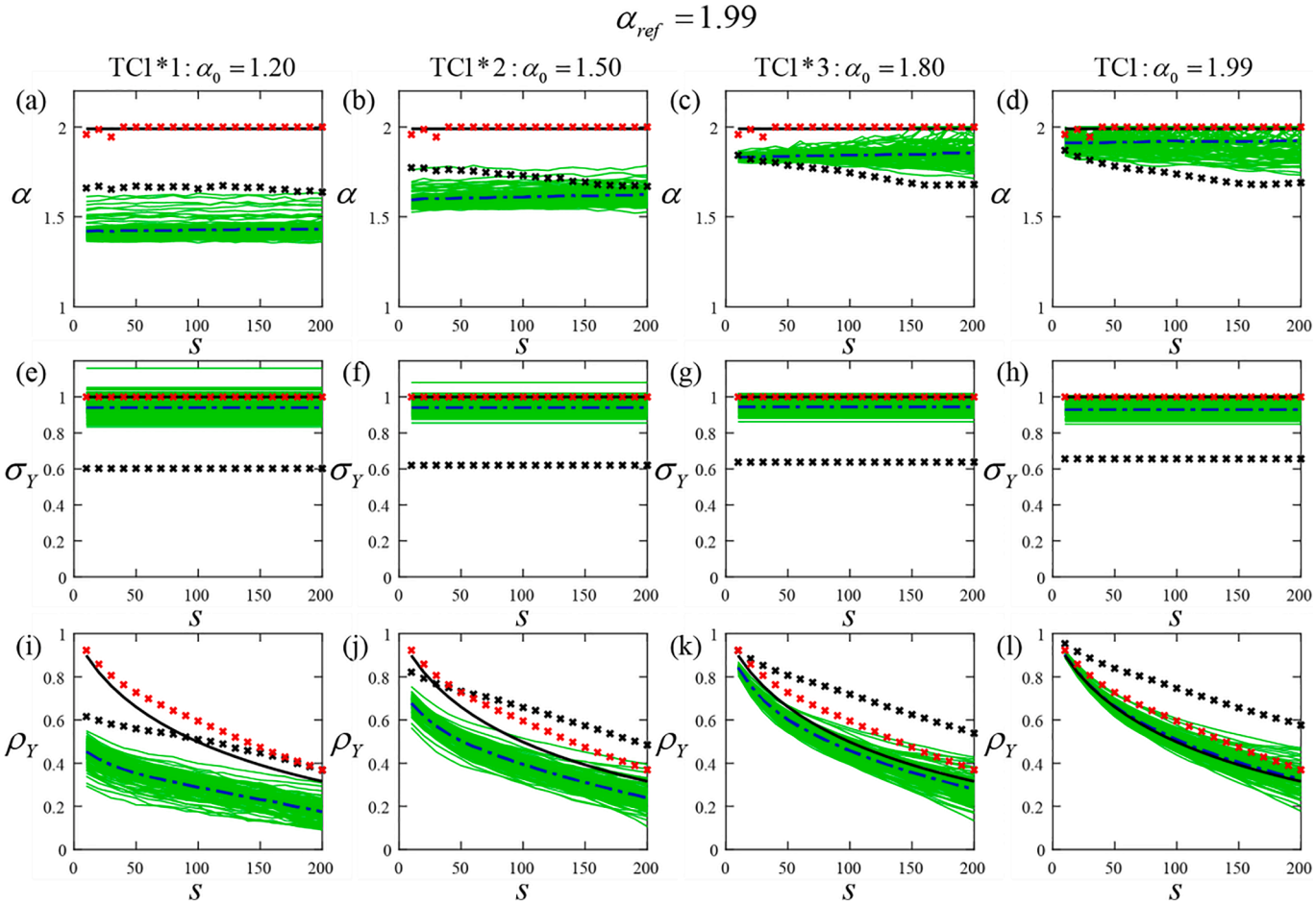


Fig. 10. Estimates of α (top row), σ_Y (middle), and correlation ρ_Y (bottom) versus lag (denoted by s) for 100 realizations (green curves) of the ensemble of Y fields obtained at the end of the assimilation process and characterized by $\alpha_{ref} = 1.99$. Dash-dotted blue curves correspond to the mean of these collection of curves; red and black cross symbols correspond to the results associated with the reference Y field and to the mean Y field of the final ensemble, respectively; and black solid lines correspond to the true (input) values associated with the reference Y field. All estimates are obtained upon relying on estimation Method B suggested by Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

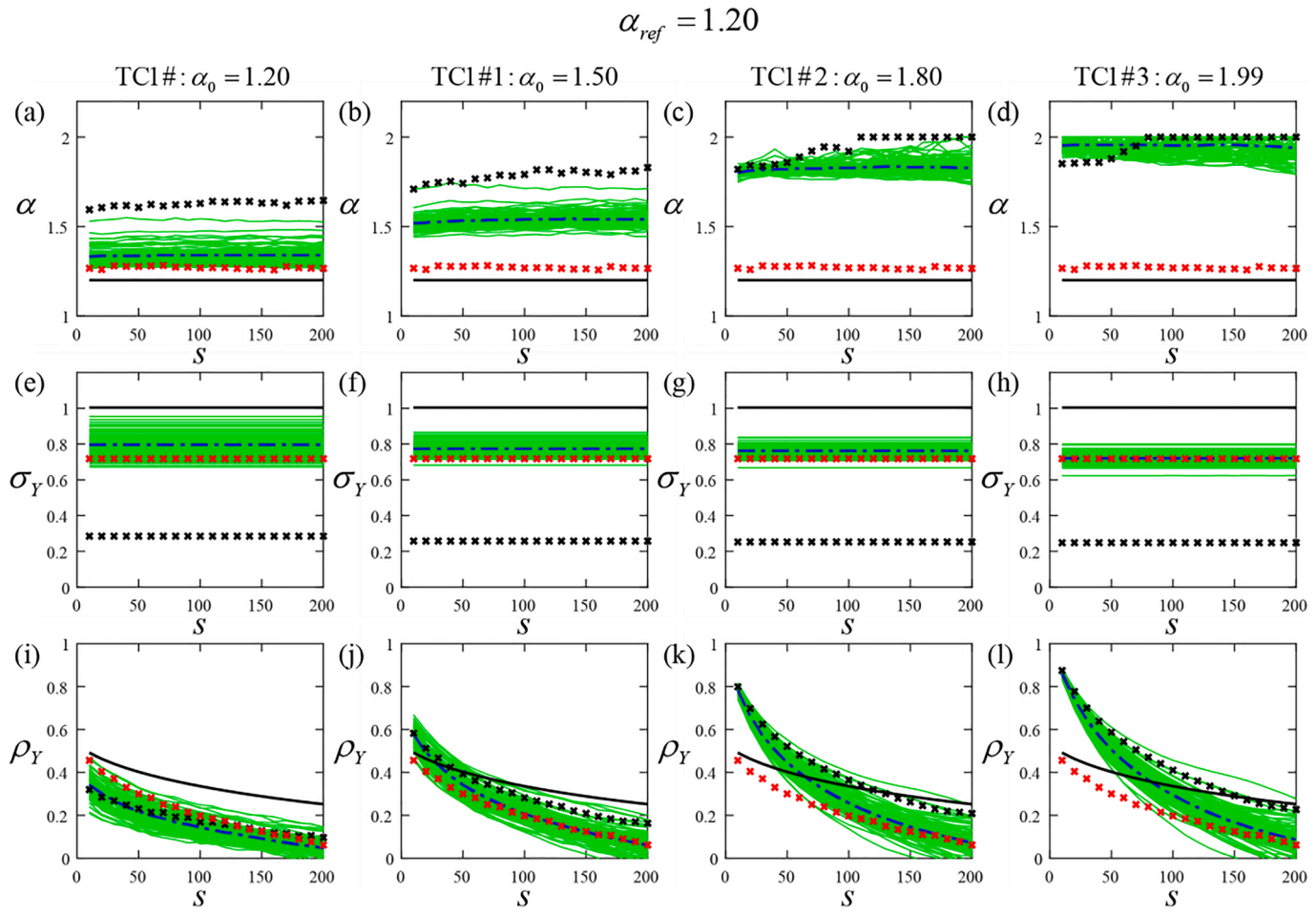


Fig. 11. Estimates of α (top row), σ_Y (middle), and correlation ρ_Y (bottom) versus lag (denoted by s) for 100 realizations (green curves) of the ensemble of Y fields obtained at the end of the assimilation process and characterized by $\alpha_{ref} = 1.20$. Dash-dotted blue curves correspond to the mean of these collection of curves; red and black cross symbols correspond to the results associated with the reference Y field and to the mean Y field of the final ensemble, respectively; and black solid lines correspond to the true (input) values associated with the reference Y field. All estimates are obtained upon relying on estimation Method B suggested by Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

inaccuracies associated with the characterization of the tails of the pdf of *Y*, consistent with the observation that accurate evaluation of sharp peaks and heavy tails of the pdf of a GSG field requires a large number of Monte Carlo samples (see Riva et al., 2015a). These aspects are also exemplified in the Supplementary Material, where it is shown that a large collection of realizations (i.e., of the order of 10^4–10^5) would be required to significantly improve the characterization of the tails of the pdf of *Y* in the presence of large departures from a Gaussian behavior (e.g., when $\alpha = 1.2$). Additional research is otherwise needed in future studies to systematically address the benefits of relying on an iterative normal-score ensemble smoother to assist the analysis.

Fig. 7 depicts results associated with $\alpha_{ref} = 1.20$ and corresponding to those embedded in Fig. 6. The lowest values of E_Y and S_Y correspond to TC1#3, followed by TC1#2, TC1#1, and TC1# (see Fig. 9a). Absolute relative (percent) differences between values of E_Y and S_Y obtained for TC1#1 with respect to their counterparts related to TC1# are the lowest ones (see green curves in Fig. 7d, e). As discussed above, this behavior may be due to the use of the iES formulation in the context of a strongly non-Gaussian field. Otherwise, note that the lowest values of E_{obs} are observed for TC1#, E_{obs} generally increasing as α_0 deviates from α_{ref}.

Table 4 lists values of E_Y, S_Y, and E_{obs} obtained through FBadap and LB2020 at the end of the assimilation procedure for TCs 1*1–1*3 and TC1 (associated with $\alpha_{ref} = 1.99$ and $\alpha_0 = 1.20$, 1.50, 1.80, and 1.99, respectively) and for TCs 1# and 1#1–1#3 (associated with $\alpha_{ref} = 1.20$ and $\alpha_0 = 1.20$, 1.50, 1.80, and 1.99, respectively). Similar to what has been observed with reference to Figs. 6 and 7, one can note that: (*i*) values of E_Y and S_Y display an overall decreasing trend as α_0 increases (i.e., as the initial *Y* field tends to become Gaussian); while (*ii*) E_{obs} values decrease as α_0 tends to α_{ref}. One can additionally note that E_Y values rendered by FBadap are consistently lower than those stemming from LB2020. This result suggests that the former provides higher accuracy of conductivity estimates than the latter. Such a behavior is possibly linked to the distinct types of tapering functions used in these two approaches.

The results embedded in Figs. 6 and 7 and Table 4 suggest that the use of a Gaussian ensemble as an initial guess is a robust choice. It enables one to obtain accurate conductivity estimates (whose quality is similar to the results obtained upon relying on the knowledge of the reference value of α) also in the case where the reference *Y* field is associated with a strongly non-Gaussian behavior.

As an additional investigation, we analyze the statistics of the final estimated *Y* fields corresponding to the TCs included in Figs. 6 and 7. In this context, Figs. 8 and 9 respectively depict frequency distributions of the values of α and σ_Y estimated for each of the 100 realizations of the final ensemble for all TCs of Group 2 and TC1 (from Group 1). These estimated values have been obtained upon relying on the methodology proposed by Riva et al., (2015a; their Method A, see Supplementary Material for additional details). Green bars in Figs. 8 and 9 correspond to histograms of the available 100 estimates of α and σ_Y, respectively. Vertical dashed blue lines denote the corresponding mean values. Red cross symbols correspond to values of α and σ_Y evaluated from the

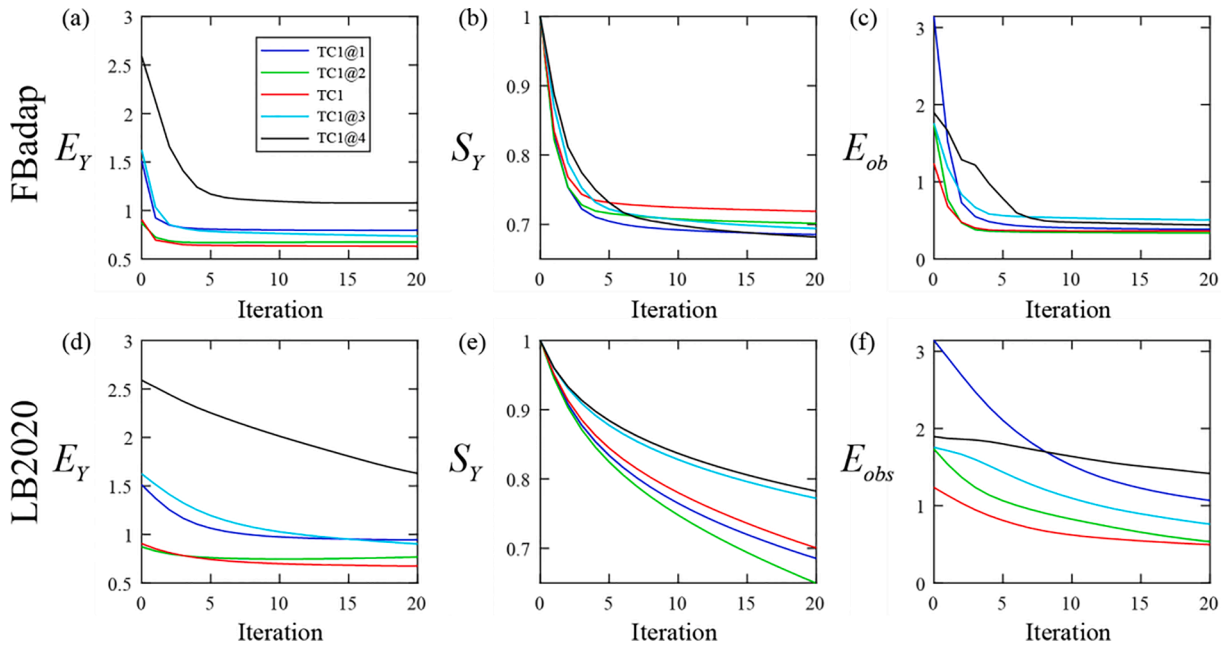


Fig. 12. Values of E_Y (left column), S_Y (middle), and E_{obs} (right) versus the number of outer iterations associated with Fbadap (top row) and LB2020 (bottom) for all TCs of Group 3 and TC1 (Group 1): TC1@1 (blue curves), TC1@2 (green), TC1 (red), TC1@3 (cyan), and TC1@4 (black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reference Y fields. Black cross symbols denote values of α and σ_Y evaluated from the mean Y fields of the final ensembles, black solid lines representing the theoretical (input) values of α (i.e., α_{ref}) and σ_Y associated with the reference Y fields. One can see from Figs. 8 and 9 that α and σ_Y values evaluated from the generated reference Y fields differ from their theoretical counterparts (see Table 1). Such a discrepancy becomes more pronounced for $\alpha_{ref} = 1.20$. This is consistent with the observation that it is difficult to accurately estimate the shape and scale parameter of a GSG distribution from a single realization comprising a limited number of values of Y (see also Riva et al., 2015a). Mean values of α and σ_Y evaluated through the final collection of 100 realizations are generally closer to their true counterparts as α_0 approaches α_{ref}. This finding suggests that the combination of a limited number of realizations and the initial guess for α can have a marked impact on the accuracy of the estimation of these parameters. Additionally, the histogram of the available 100 estimates of α displays a sharp peak located around a value

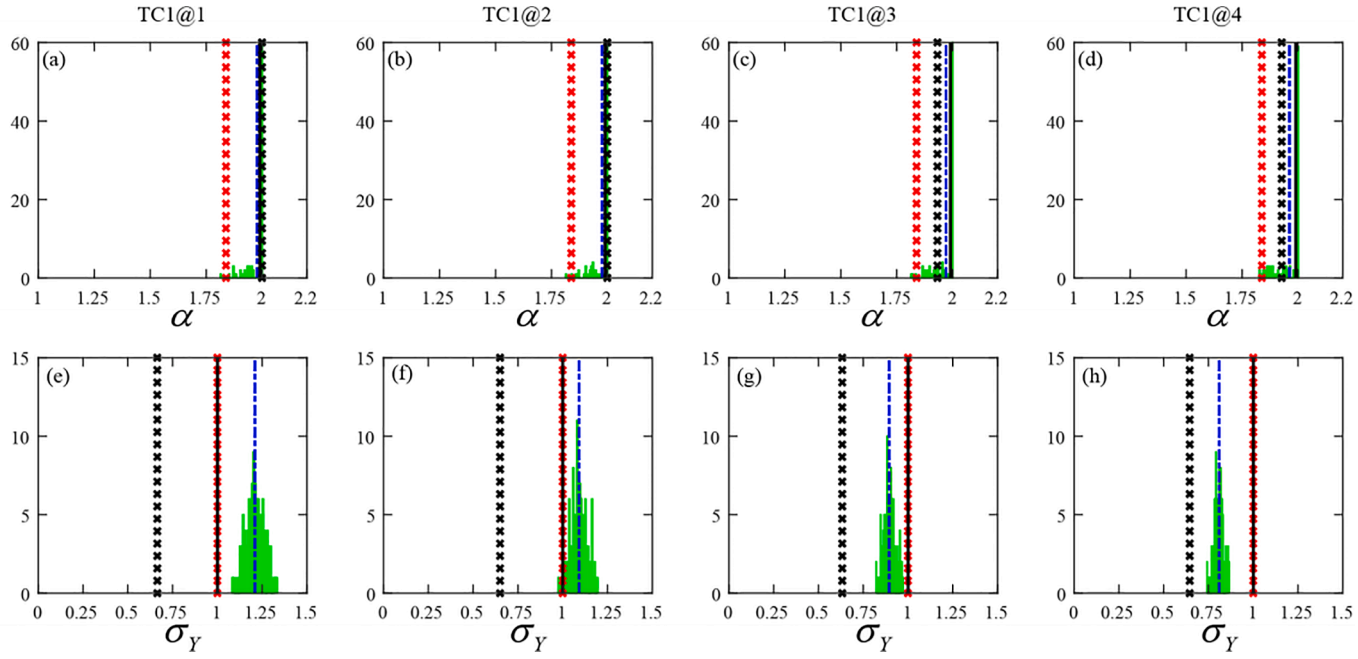


Fig. 13. Histogram (green bars) and mean values (dash-dot blue lines) of the values of α and σ_Y estimated across the 100 realizations of the final ensemble of Y for TC1@1 (first column), TC1@2 (second), TC1@3 (third), and TC@4 (fourth). Estimated values for α and σ_Y associated with the mean Y fields (black crosses) of the final ensembles, Y reference fields (red crosses) and their corresponding theoretical values (black solid) are also depicted. Estimates are obtained through the application of Method A of Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

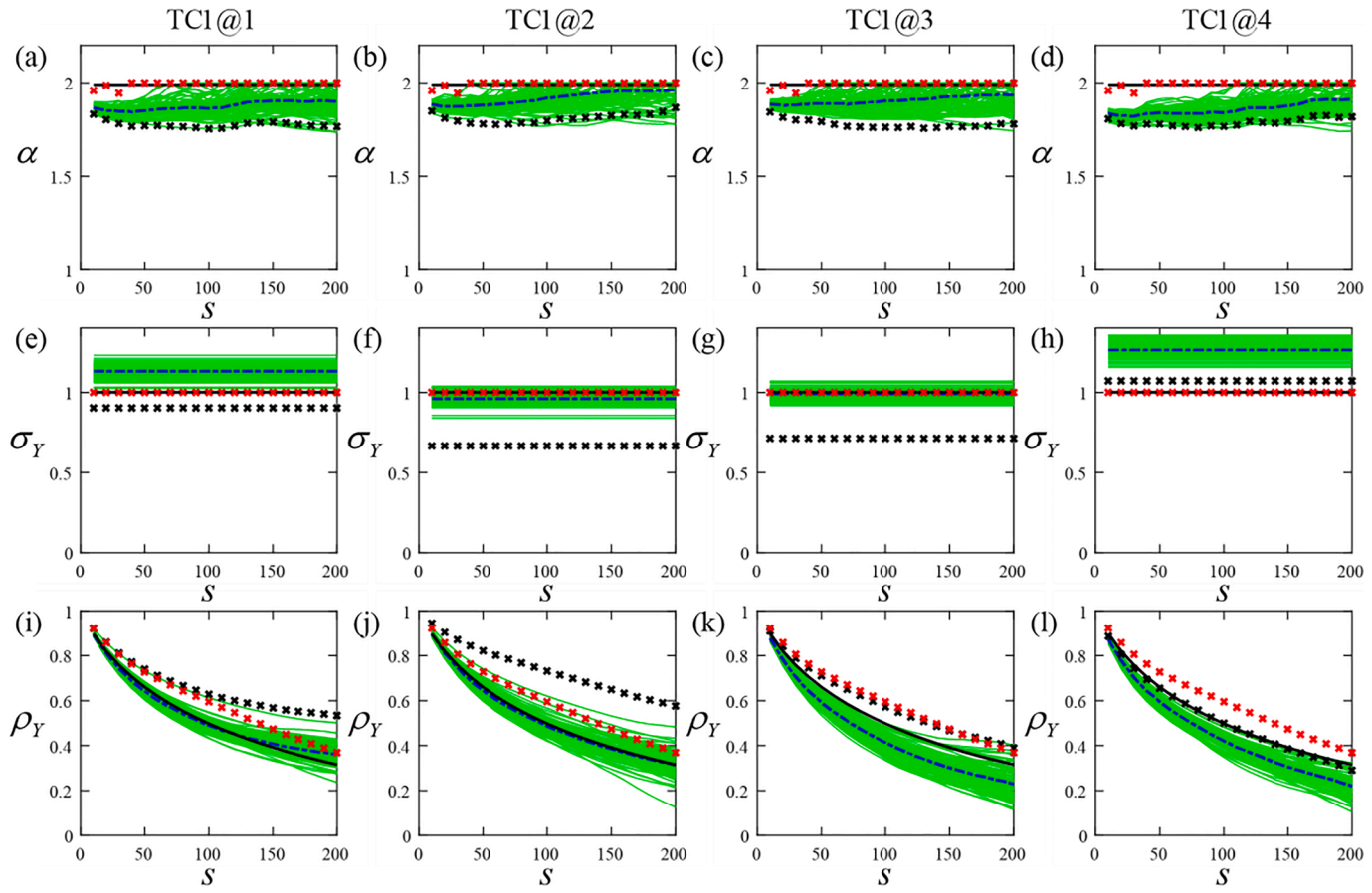


Fig. 14. Estimates of α (top row), σ_Y (middle), and correlation ρ_Y versus lag (denoted by s) for 100 realizations (green curves) of the ensemble of Y fields obtained at the end of the assimilation process for TC1@1 (first column), TC1@2 (second), TC1@3 (third), and TC1@4 (fourth). Dash-dotted blue curves correspond to the mean of these collection of curves; red and black cross symbols correspond to the results associated with the reference Y field and to the mean Y field of the final ensemble, respectively; and black solid lines correspond to the true (input) values associated with the reference Y field. All estimates are obtained upon relying on estimation Method B suggested by Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of α that is close to 2.0. This finding is consistent with the observation that $\alpha_{ref} = \alpha_0 = 1.99$ in TC1 (see also in Table 2).

To complement the analysis, we also estimate α, σ_Y, and the correlation coefficient of Y (i.e., ρ_Y) at 20 lags ($s = 10, 20, \ldots, 200$) for each realization through Method B of Riva et al., (2015a; see Supplementary Material for additional details) for all TCs of Group 2 and TC1 (Group 1). Figs. 10 and 11 depict the corresponding results for $\alpha_{ref} = 1.99$ and 1.20, respectively. Green curves in each subplot correspond to the values of either α, σ_Y, or ρ_Y versus lag (s) obtained for the collection of 100 realizations of Y at the end of the assimilation process; dash-dotted blue curves correspond to the means of these collection of curves; red and black cross symbols correspond to the results associated with the reference Y fields and with the mean Y fields of the final ensembles, respectively; and black solid lines correspond to the true (input) values linked to the reference Y field. The patterns related to the results included in the first two rows of Figs. 10 and 11 are similar to what can be observed in Figs. 8 and 9, i.e., the values obtained for α and σ_Y on the basis of the mean Y field of the final ensemble (*i*) are generally closer to the corresponding values linked to the reference fields as α_0 approaches α_{ref} and (*ii*) do not change with lag, consistent with the underlying theoretical framework. When $\alpha_{ref} = 1.99$, all tested values of α_0 would yield similar accuracy of ρ_Y estimates. Otherwise, when $\alpha_{ref} = 1.20$ the accuracy of ρ_Y estimate improves as α_0 approaches α_{ref}.

Based on the above results, we conclude that, regardless of the strength of the departure from Gaussian of the reference Y field, an initial Gaussian ensemble of Y can yield a level of accuracy of conductivity estimates (in terms of E_Y values) that is of similar quality as the one that can be obtained when the true value of α (i.e., α_{ref}) is known. Otherwise, in the strongly non-Gaussian settings we consider, estimation of system states (i.e., heads and concentrations) as well as shape (i.e., α) and scale (i.e., σ_Y) parameters of the realizations of the final ensemble of Y can be impacted by the selected α_0.

4.3. *Effect of incomplete knowledge of mean and variance of the initial ensemble (Groups 3 and 4)*

Fig. 12 depicts E_Y (left column), S_Y (middle), and E_{obs} (right) versus the number of outer iterations associated with Fbadap (top row) and LB2020 (bottom) for all TCs of Group 3 and TC1 (from Group 1). The highest values of E_Y and E_{obs} correspond to TC1@4 (where the percentage difference between the mean of the initial ensemble of Y and the true one is equal to 400 %), the lowest value being associated with TC1 (where the mean of the initial ensemble coincides with the true value). Notably, values of E_Y and E_{obs} evaluated via LB2020 are larger and are characterized by a much higher variability among diverse TCs than their counterparts obtained via Fbadap. Note that Fbadap and LB2020 yield different rankings for the results of S_Y associated with the various TCs considered. The results summarized above indicate that the mean of the initial ensemble of Y impacts conductivity estimation. They also suggest that Fbadap is generally associated with a lower negative impact than LB2020 on the quality of the final estimates.

Fig. 13 depicts the frequency distribution of the values of α (top row)

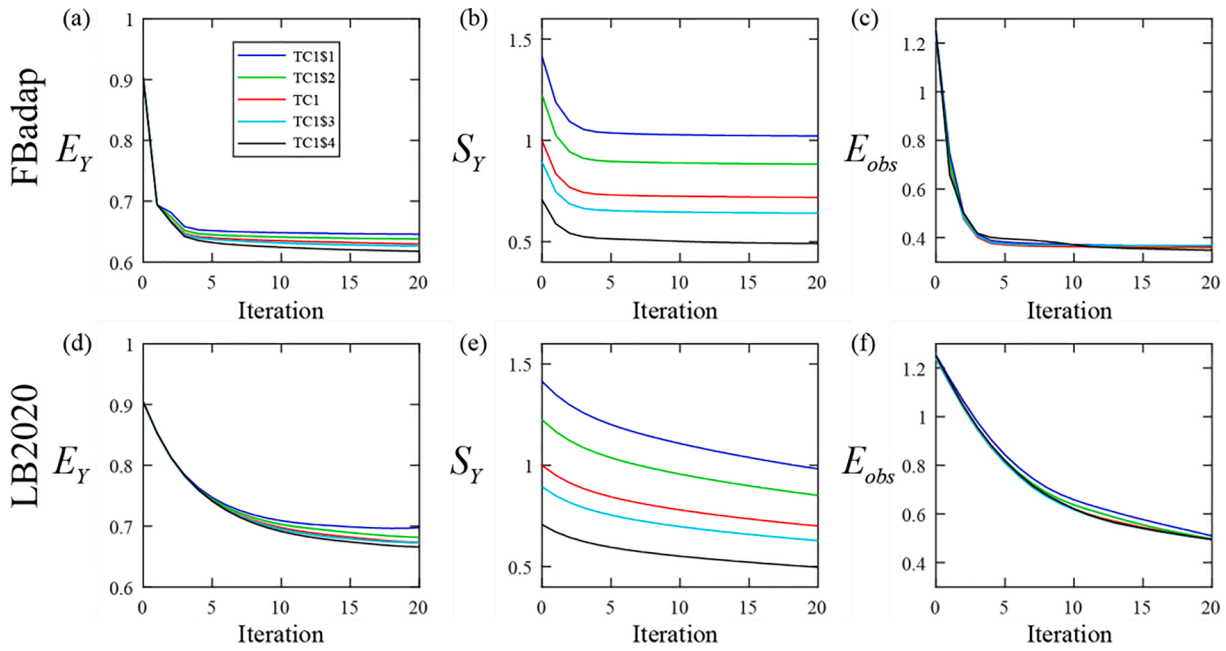


Fig. 15. Values of E_Y (left column), S_Y (middle), and E_{obs} (right) versus the number of outer iterations associated with Fbadap (top row) and LB2020 (bottom) for TC1$1 (blue curves), TC1$2 (green), TC1 (red), TC1$3 (cyan), and TC1$4 (black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

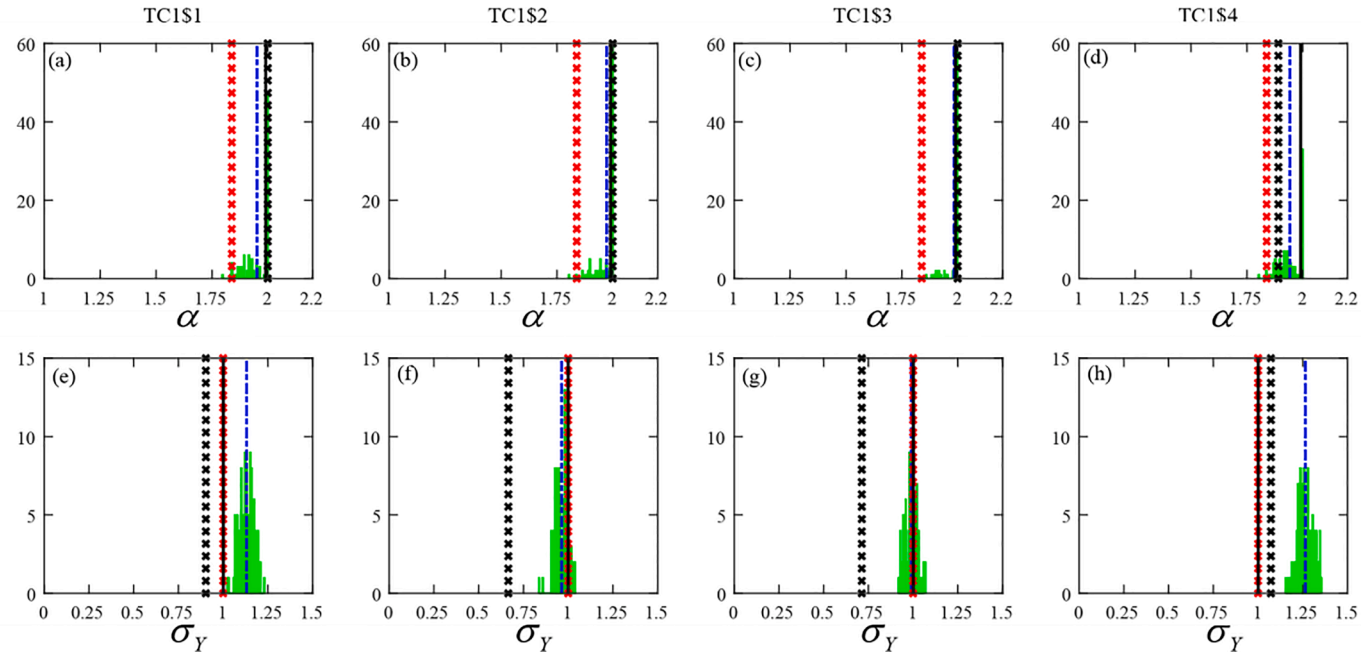


Fig. 16. Histogram (green bars) and mean values (dash-dot blue lines) of the values of α and σ_Y estimated across the 100 realizations of the final ensemble of Y for (a, e) TC1$1, (b, f) TC1$2, (c, g) TC1$3, and (d, h) TC1$4. Estimated values for α and σ_Y associated with the mean Y fields (black crosses) of the final ensembles, Y reference fields (red crosses) and their corresponding theoretical values (black solid) are also depicted. Estimates are obtained through the application of Method A of Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and σ_Y (bottom row) estimated for each of the 100 realizations of the final ensemble. It also shows the corresponding values estimated upon relying on (*i*) the mean *Y* field resulting from the final ensemble or (*ii*) the reference *Y* field through Method A of Riva et al., (2015a) for all TCs of Group 3, together with their theoretical counterparts. A joint analysis on Fig. 8d, 9d, and 13 suggests that the mean of the initial ensemble of *Y* has only a modest impact on the estimation of α and σ_Y.

Fig. 14 depicts the frequency distribution of the values of α (top row), σ_Y (middle row), and ρ_Y (bottom row) at 20 lags (s = 10, 20, …, 200) estimated for each of the 100 realizations of the final ensemble. It also includes the corresponding values estimated upon relying on (*i*) the mean *Y* field associated with the final ensemble and (*ii*) the reference *Y* field for all TCs of Group 3, together with their theoretical counterparts. Estimates are based on Method B of Riva et al., (2015a). From visual inspection it is difficult to assess which TC is associated with the (overall) best estimates of α, σ_Y, or ρ_Y, no systematic variation of these

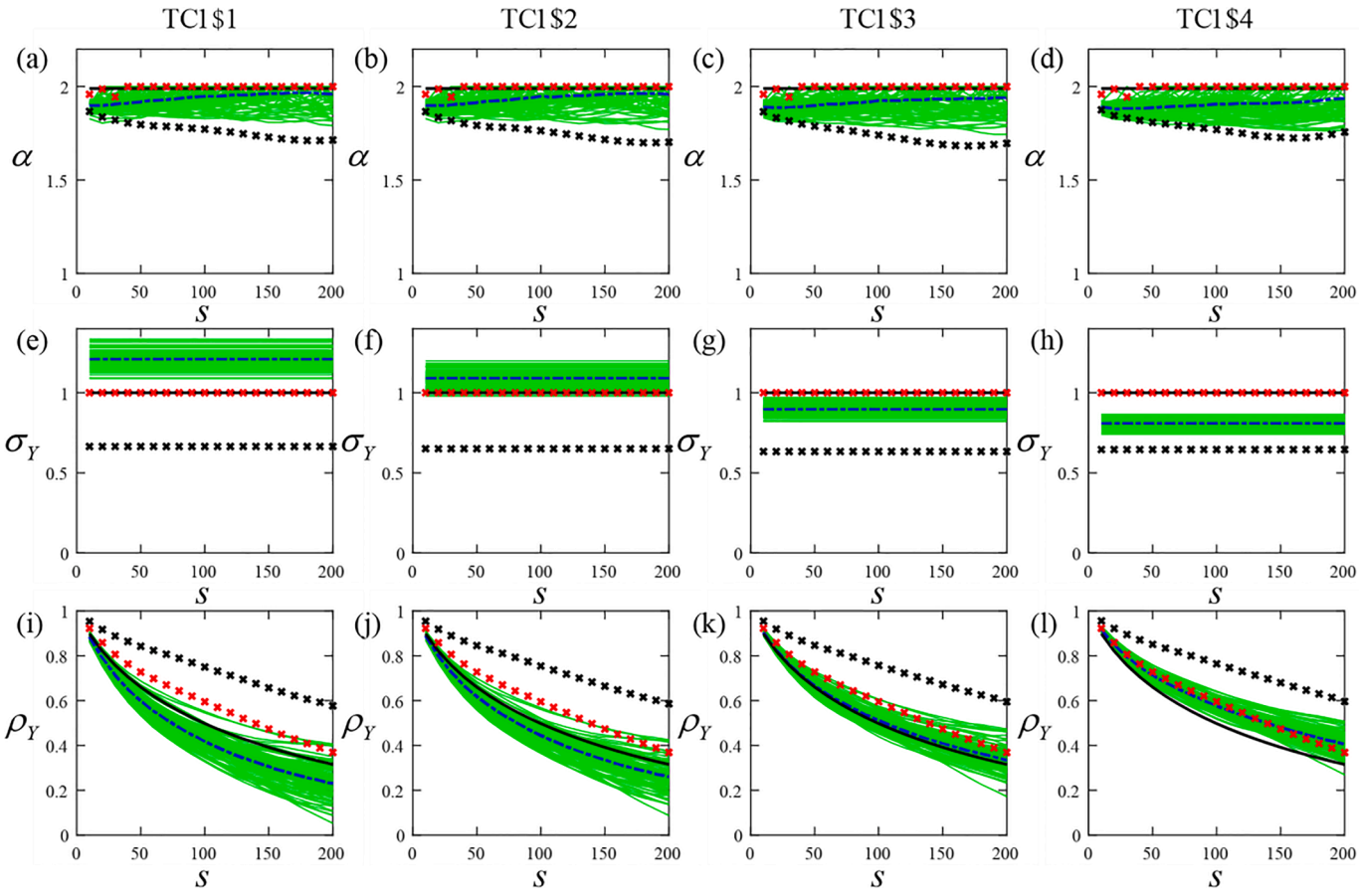


Fig. 17. Estimates of α (top row), σ_Y (middle), and correlation ρ_Y versus lag (denoted by s) for 100 realizations (green curves) of the ensemble of Y fields obtained at the end of the assimilation process for TC1$1 (first column), TC1$2 (second), TC1$3 (third), and TC1$4 (fourth). Dash-dotted blue curves correspond to the mean of these collection of curves; red and black cross symbols correspond to the results associated with the reference Y field and to the mean Y field of the final ensemble, respectively; and black solid lines correspond to the true (input) values associated with the reference Y field. All estimates are obtained upon relying on estimation Method B suggested by Riva et al., (2015a; see also Supplementary Material). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

values against the mean of the initial ensemble of Y being evidenced.

Fig. 15 depicts E_Y (left column), S_Y (middle), and E_{obs} (right) versus the number of outer iterations associated with FBadap (top row) and LB2020 (bottom) for all test cases in Group 4 and TC1. All TCs yield similar values of E_Y upon relying on FBadap (Fig. 15a), similar to what could be noted for E_{obs} in Fig. 15c. Values of E_Y (Fig. 15d) and E_{obs} (Fig. 15f) obtained through LB2020 are also characterized by only mild variations across TCs. These are otherwise generally higher than their counterparts obtained through FBadap. These results indicate that the imperfect knowledge of the variance of the initial ensemble of Y does not significantly impact conductivity and state variable estimation, regardless of the methodology applied. The highest values of S_Y obtained through FBadap and LB2020 correspond to TC1$1 (characterized by a variance of the initial Y ensemble that is twice the true one), values of S_Y generally decreasing with the variance of the initial ensemble.

Similar to Fig. 13, Fig. 16 depicts the frequency distribution of the values of α (top row) and σ_Y (bottom row) estimated for each of the 100 realizations of the final ensemble. It also depicts the corresponding values estimated from the mean Y field of the final ensemble and from the reference Y field through Method A of Riva et al., (2015a) for all TCs of Group 4, together with their theoretical counterparts. Joint analysis of Fig. 8d, 9d, and 16 reveals that there is no systematic variation of α or σ_Y with the values considered for the variance of the initial ensemble of Y. Similar to Fig. 14, Fig. 17 depicts estimates of α, σ_Y, or ρ_Y obtained through Method B of Riva et al., (2015a) for all TCs of Group 4, together with their theoretical counterparts. It is visually hard to assess which TC yields the (overall) highest accuracy of α, σ_Y, or ρ_Y estimates. No systematic variation of α, σ_Y, or ρ_Y estimates can be observed for the different values of variance considered for the initial ensemble of Y.

Elements of interest for additional analyses can include the study of the effects of (*a*) frequency of data collection, (*b*) number of monitoring wells, and (*c*) impact of the depth at which data are collected on estimation accuracy of conductivity. In this context, we recall that Xia et al. (2018) indicate that relying on a high frequency of data collection and a large number of monitoring wells can further enhance negative effect of spurious correlation on conductivity estimates through distance-based localized iterative ensemble Kalman filter.

5. Conclusions

A novel correlation-based localization approach (here denoted as FBadap) is developed to estimate spatially heterogeneous log-conductivity ($Y = \ln K$) fields. The latter form three-dimensional generalized sub-Gaussian random fields within which head and concentration observations are collected from monitoring wells screened at multiple depths. We investigate the reliability of FBadap to estimate Y in comparison to (*i*) its traditional counterpart (denoted by FBconst) associated with a constant threshold and (*ii*) a recently proposed adaptive approach (denoted by LB2020; Luo and Bhakta, 2020). Based on our results and comprehensive analyses, our work yields the following major conclusions.

1. We find that FBadap yields high quality results (in terms of average absolute difference, E_Y, between estimated Y values and their reference counterparts; square root of the average estimation variance, S_Y; and average absolute difference, E_{obs}, between the updated simulated observations and their reference values) in the context of a variety of settings. These include various ensemble sizes, strength of the departure of the description of the spatial heterogeneity from a Gaussian model, as well as values of the mean and variance of the initial collection/ensemble of Y.
2. FBadap exhibits adaptivity to cope with diverse ensemble sizes when estimating nearly Gaussian as well as strongly non-Gaussian random conductivity fields (in terms of the employed evaluation metrics, i.e., E_Y, S_Y, and E_{obs}) and yields (overall) lower values of E_Y and E_{obs}, as compared against LB2020.
3. An initial virtually Gaussian ensemble of Y (i.e., associated with $\alpha_0 = 1.99$) can always (even when the reference field is non-Gaussian) yield the highest accuracy of conductivity estimates in terms of E_Y.
4. Knowledge of the shape parameter (i.e., of α) associated with the distribution of Y yields higher accuracy in the estimation of the system states (in terms of heads and solute concentrations) as well as of the scale parameter (i.e., σ_Y) of the final collection of Y fields.
5. The mean and variance of the initial ensembles of Y exert some impacts on the accuracy of the estimates of Y and of the system state (i.e., hydraulic heads and concentrations). Varying the mean of the initial ensembles of Y yields stronger impacts on both E_Y and E_{obs} than varying their variance. Otherwise, no clear systematic influence of the mean or variance of the initial ensemble of Y can be documented on the estimates of α, σ_Y, and ρ_Y.

CRediT authorship contribution statement

Chuan-An Xia: Conceptualization, Funding acquisition, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Jiayun Li:** Software, Visualization, Writing – original draft. **Monica Riva:** Methodology, Supervision, Writing – original draft, Writing – review & editing. **Xiaodong Luo:** Methodology, Supervision, Writing – review & editing. **Alberto Guadagnini:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the National Nature Science Foundation of China (Grant No. 42002247) and by the Nature Science Foundation of Guangdong Province, China (Grant No. 2020A1515111054). Xiaodong Luo acknowledges financial support from the Research Council of Norway through the Petromaks-2 project DIGIRES (RCN no. 280473) and the industrial partners AkerBP, Wintershall DEA, Vår Energi, Petrobras, Equinor, Lundin, and Neptune Energy, and the National Centre for Sustainable Subsurface Utilization of the Norwegian Continental Shelf (NCS2030), which is funded by industry partners and the Research Council of Norway (project number: 331644). Chuan-An Xia would like to acknowledge Dr. Zhi-yuan Zhang for his help on building the very early groundwater model. A. Guadagnini acknowledges funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005) in the context of the RETURN Extended Partnership.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2024.131054.

References

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons 752.
Anderson, J.L., 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D 230(1), 99–111.
Arroyo-Negrete, E., Devegowda, D., Datta-Gupta, A., 2008. Streamline-assited Ensemble Kalman Filter for rapid and continuous reservoir model updating. SPE Reserv. Eval. Eng. 11 (6), 1046–1060.
Bianchi Janetti, E., Guadagnini, L., Riva, M., Guadagnini, A., 2019. Global sensitivity analyses of multiple conceptual models with uncertain parameters driving groundwater flow in a regional-scale sedimentary aquifer. J. Hydrol. 574, 544–556.
Bishop, C.H., Hodyss, D., 2009. Ensemble covariances adaptively localized with ECO-RAP, Part 1: Tests on simple error models. 61: 84-96.
Bocquet, M., Sakov, P., 2014. An iterative ensemble Kalman smoother. Quart. J. Roy. Meteor. Soc. 140(682), 1521–1535. DOI:10.1002/qj.2236.
Boffetta, G., Mazzino, A., Vulpiani, A., 2008. Twenty-five years of multifractals in fully developed turbulence: a tribute to Giovanni Paladin. J. Phys. A Math. Theor. 41 (36), 363001 https://doi.org/10.1088/1751-8113/41/36/363001.
Ceresa, L., Guadagnini, A., Riva, M., Porta, G.M., 2022. Macro dispersion in generalized sub-Gaussian heterogeneous porous media. Int. J. Heat Mass Transf. 195 (123117), 1–11. https://doi.org/10.1016/j.ijheatmasstransfer.2022.123117.
Chen, Y., Oliver, D.S., 2010. Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput. Geosci. 14, 579–601. https://doi.org/10.1007/s10596-009-9174-6.
Chen, Y., Oliver, D.S., 2013. Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput. Geosci. 17, 689–703.
Di Federico, V., Neuman, S.P., 1997. Scaling of random fields by means of truncated power variograms and associated spectra. Water Resour. Res. 33, 1075–1085.
Elçi, A., Flach, G.P., Molz, F.J., 2003. Detrimental effects of natural vertical head gradients on chemical and water level measurements in observation wells: identification and control, J. Hydrol. 281, 70–81.
Elci, A., Molz III, F.J., Waldrop, W.R., 2001. Implications of Observed and Simulated Ambient Flow in Monitoring Wells. Ground Water 39, 853–862.
Emerick, A., Reynolds, A., 2011. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. Comput. Geosci. 15, 251–269. https://doi.org/10.1007/s10596-010-9198-y.
Emerick, A.A., Reynolds, A.C., 2013. Ensemble smoother with multiple data assimilation. Comput. Geosci. 55, 3–15. https://doi.org/10.1016/j.cageo.2012.03.011.
Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter, Data Assimilation: The Ensemble Kalman Filter.
Furrer, R., Bengtsson, T., 2007. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivar. Anal. 98 (2), 227–255. https://doi.org/10.1016/j.jmva.2006.08.003.
Ganti, V., Singh, A., Passalacqua, P., Foufoula-Georgiou, E., 2009. Subordinated Brownian motion model for sediment transport. Phys. Rev. E 80 (1), 011111. https://doi.org/10.1103/PhysRevE.80.011111.
Guadagnini, A., Riva, M., Neuman, S.P., 2012. Extended power-law scaling of heavy tailed random air-permeability fields in fractured and sedimentary rocks. Hydrol. Earth Syst. Sci. 16, 3249–3260. https://doi.org/10.5194/hess-16-3249-2012.
Guadagnini, A., Neuman, S.P., Schaap, M.G., Riva, M., 2013. Anisotropic statistical scaling of vadose zone hydraulic property estimates near Maricopa, Arizona. Water Resour. Res. 49, 8463–8479. https://doi.org/10.1002/2013WR014286.
Guadagnini, A., Neuman, S.P., Schaap, M.G., Riva, M., 2014. Anisotropic statistical scaling of soil and sediment texture in a stratified deep vadose zone near Maricopa, Arizona. Geoderma 217–227. https://doi.org/10.1016/j.geoderma.2013. 09.008.
Guadagnini, A., Neuman, S.P., Nan, T., Riva, M., Winter, C.L., 2015. Scalable statistics of correlated random variables and extremes applied to deep borehole porosities. Hydrol. Earth Syst. Sci. 19, 729–745. https://doi.org/10.5194/hess-19-729-2015.
Guadagnini, A., Riva, M., Neuman, S.P., 2018. Recent advances in scalable non-Gaussian geostatistics: the generalized sub-Gaussian model. J. Hydrol. 562, 685–691. https://doi.org/10.1016/j.jhydrol.2018.05.001.
Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Wea. Rev. 129 (11), 2776–2790.
Harbaugh, A.W., 2005. MODFLOW-2005, the US Geological Survey Modular Groundwater Model-the Groundwater Flow Process. U.S. Geological Survey, Reston, Virginia.
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 Resour. Res. 44, W09408. https://doi.org/10.1029/2007WR006505.
Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique. Mon. Weather Rev. 126 (3), 796–811.

Konikow, L.F., Hornberger, G.Z., Halford, K.J., Hanson, R.T., Harbaugh, A., 2009. W.: Revised multi-node well (MNW2) package for MODFLOW ground-water flow model, Report 6-A30.
Lacerda, J.M., Emerick, A.A., Pires, A.P., 2021. Using a machine learning proxy for localization in ensemble data assimilation. Comput. Geosci. 25, 931–944.
Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 2 (2), 164–168. https://doi.org/10.1090/qam/10666.
Li, G., Reynolds, A.C., 2009. Iterative ensemble Kalman filters for data assimilation. SPE J. 14 (3), 496–505.
Li, L., Stetler, L., Cao, Z., Davis, A., 2018. An iterative normal-score ensemble smoother for dealing with non-Gaussianity in data assimilation. J. Hydrol. 567, 759–766.
Libera, A., de Barros, F.P.J., Riva, M., Guadagnini, A., 2017. Solute concentration at a well in non-Gaussian aquifers under constant and time-varying pumping schedule. J. Contam. Hydrol. 205, 37–46. https://doi.org/10.1016/j.jconhyd.2017.08.006.
Liu, H.H., Molz, F.J., 1997. Comment on "Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations" by Scott Painter. Water Resour. Res. 33 (4), 907–908. https://doi.org/10.1029/96WR03788.
Luo, X., Bhakta, T., 2020. Automatic and adaptive localization for ensemble-based history matching. J. Pet. Sci. Eng. 184 https://doi.org/10.1016/j.petrol.2019.106559.
Luo, X., Lorentzen, R., Valestrand, R., Evensen, G., 2018. Correlation-Based Adaptive Localization for Ensemble-Based History Matching: Applied to the Norne Field Case Study.
Luo, X., Stordal, A.S., Lorentzen, R.J., Naevdal, G., 2015. Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: theory and applications. SPE J. 20, 962–982.
Maschio, C., Schiozer, D.J., 2016. Probabilistic history matching using discrete Latin Hypercube sampling and nonparametric density estimation. J. Pet. Sci. Eng. 147, 98–115.
Meerschaert, M.M., 2004. Fractional Laplace model for hydraulic conductivity. Geophys. Res. Lett. 31 (8), L08501. https://doi.org/10.1029/2003GL019320.
Miyoshi, T., 2010. An adaptive covariance localization method with the LETKF. In: 14th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS).
Nan, T., Wu, J., 2011. Groundwater parameter estimation using the ensemble Kalman filter with localization. Hydrgeol. J. 19, 547–561.
Neuman, S.P., Di Federico, V., 2003. Multifaceted nature of hydrogeologic scaling and its interpretation. Rev. Geophys. 41 (3) https://doi.org/10.1029/2003RG000130.
Neuman, S.P., Riva, M., Guadagnini, A., 2008a. On the geostatistical characterization of hierarchical media. Water Resour. Res. 44, W02403. https://doi.org/10.1029/2007WR006228.
Neuman, S.P., Riva, M., Guadagnini, A., 2008b. On the geostatistical characterization of hierarchical media. Water Resour. Res. 44 (2) https://doi.org/10.1029/2007WR006228.
Neuman, S.P., 1990. Universal scaling of hydraulic conductivities and dispersivities in geologic media. Water Resour. Res. 26(8), 1749–1758. DOI:10.1029/WR026i008p01749.
Neuman, S.P., 1995. On advective transport in fractal permeability and velocity-fields. Water Resour. Res. 31(6), 1455–1460. DOI:10.1029/95wr00426.
Painter, S., 1996. Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations. Water Resour. Res. https://doi.org/10.1029/96WR00286.
Painter, S., 2001. Flexible scaling model for use in random field simulation of hydraulic conductivity. Water Resour. Res. 37 (5), 1155–1163. https://doi.org/10.1029/2000WR900394.
Post, V., Kooi, H., Simmons, C., 2007. Using hydraulic head measurements in variable-density ground water flow analyses. Ground Water 45 (6), 664–671. https://doi.org/10.1111/j.1745-6584.2007.00339.x.
Rasmussen, J., Madsen, H., Jensen, K.H., Refsgaard, J.C., 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydrol. Earth Syst. Sci. 19, 2999–3013. https://doi.org/10.5194/hess-19-2999-2015.
Riva, M., Guadagnini, A., Neuman, S.P., 2017. Theoretical analysis of non-Gaussian heterogeneity effects on subsurface flow and transport 53, 2998–3012.
Riva, M., Neuman, S.P., Guadagnini, A., 2013a. Sub-Gaussian model of processes with heavy tailed distributions applied to permeabilities of fractured tuff. Stoch. Env. Res. Risk A. 27, 195–207. https://doi.org/10.1007/s00477-012-0576-y.
Riva, M., Neuman, S.P., Guadagnini, A., Siena, M., 2013b. Anisotropic scaling of Berea sandstone log air permeability statistics. Vadose Zone J. 12 (3), 1–15. https://doi.org/10.2136/vzj2012.0153.
Riva, M., Neuman, S.P., Guadagnini, A., 2015a. New scaling model for variables and increments with heavy-tailed distributions. Water Resour. Res. 51, 4623–4634.
Riva, M., Panzeri, M., Guadagnini, A., Neuman, S.P., 2015b. Simulation and analysis of scalable non-Gaussian statistically anisotropic random functions. J. Hydrol. 531, 88–95.
Siena, M., et al., 2021. Statistical characterization of heterogeneous dissolution rates of calcite from insitu and real-time AFM imaging. Transport Porous Media 140 (1), 291–312. https://doi.org/10.1007/s11242-021-01624-z.
Siena, M., Guadagnini, A., Bouissonnié, A., Ackerer, P., Daval, D., and Riva, M., 2020. Generalized Sub-Gaussian Processes: Theory and Application to Hydrogeological and Geochemical Data, 56, e2020WR027436.
Siena, M., Guadagnini, A., Riva, M., Neuman, S.P., 2012. Extended power-law scaling of air permeabilities measured on a block of tuff. Hydrol. Earth Syst. Sci. 16, 29–42. https://doi.org/10.5194/hess-16-29-2012.
Siena, M., Riva, M., Giamberini, M., Gouze, P., Guadagnini, A., 2017. Statistical modeling of gas-permeability spatial variability along a limestone core. Spatial Statistics. https://doi.org/10.1016/j.spasta.2017.07.007.
Soares, R.V., Maschio, C., Schiozer, D.J., 2019. A novel localization scheme for scalar uncertainties in ensemble-based data assimilation methods. J. Petrol. Explor. Prod. Technol. https://doi.org/10.1007/s13202-019-0727-5.
Sole-Mari, G., Riva, M., Fernàndez-Garcia, D., Sanchez-Vila, X., Guadagnini, A., 2021. Solute transport in bounded porous media characterized by generalized sub-Gaussian log-conductivity distributions. Adv. Water Resour. 147, 103812.
Sun, A.Y., Morris, A.P., Mohanty, S., 2009. Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. Water Resour. Res. 45, W07424. https://doi.org/10.1029/2008WR007443.
Tong, J., Hu, B.X., Yang, J., 2012. Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field. Stoch. Environ. Res. Risk A 26, 467–478.
Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resour. Res. 41 (W01017), 2005. https://doi.org/10.1029/2004WR003059.
Wang, B., Liu, J.J., Liu, L., Xu, S.M., Huang, W.Y., 2018. An approach to localization for ensemble-based data assimilation. PLoS One 13.
Winter, C.L., Tartakovsky, D.M., 2000. Mean flow in composite porous media. Geophys. Res. Lett. 27 (12), 1759–1762. https://doi.org/10.1029/1999gl011030.
Winter, C.L., Tartakovsky, D.M., 2003. Moment Differential Equations for Flow in Highly Heterogeneous Porous Media. Surv. Geophys. 24 (1), 81–106. https://doi.org/10.1023/A:1022277418570.
Winter, C.L., Tartakovsky, D.M., Guadagnini, A., 2002. Numerical solutions of moment equations for flow in heterogeneous composite aquifers. Water Resour. Res. 38 (5), 1–8. https://doi.org/10.1029/2001WR000222.
Winter, C.L., Tartakovsky, D.M., 2002. Groundwater flow in heterogeneous composite aquifers. Water Resour. Res. 38 (8) https://doi.org/10.1029/2001wr000450.
Xia, C.-A., Hu, B.X., Tong, J., Guadagnini, A., 2018. Data Assimilation in Density-Dependent Subsurface Flows via Localized Iterative Ensemble Kalman Filter. Water Resour. Res. 54, 6259–6281.
Xia, C.-A., Luo, X., Hu, B.X., Riva, M., Guadagnini, A., 2021. Data assimilation with multiple types of observation boreholes via the ensemble Kalman filter embedded within stochastic moment equations. Hydrol. Earth Syst. Sci. 25, 1689–1709. https://doi.org/10.5194/hess-25-1689-2021.
Zhang, Z.Y., Jiang, X.W., Wang, X.S., Wan, L., Wang, J.Z., 2019. Why mixed groundwater at the outlet of open flowing wells in unconfined-aquifer basins can represent deep groundwater: implications for sampling in long-screen wells. Hydrgeol. J. 27 (1), 409–421. https://doi.org/10.1007/s10040-018-1842-y.
Zheng, C., 2009. MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems: Supplemental User's Guide. University of Alabama, Birmingham, AL.
Zheng, C., 2010. a modular three-dimensional multispecies transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems, Supplemental User's Guide, 2010.
Zovi, F., Camporese, M., Franssen, H.-J.-H., Huisman, J.A., Salandin, P., 2017. Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter. J. Hydrol. 548, 208–224. https://doi.org/10.1016/j.jhydrol.2017.02.056.