Ensemble smoother with multiple data assimilation for reverse flow routing

Valeria Todaro[a,b,*], Marco D'Oria[a], Maria Giovanna Tanda[a], J. Jaime Gómez-Hernández[a,b]

[a] Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy

[b] Institute for Water and Environmental Engineering, Universitat Politècnica de València, Valencia, Spain

* Corresponding author.

E-mail addresses: valeria.todaro@studenti.unipr.it (V. Todaro), marco.doria@unipr.it (M. D'Oria), mariagiovanna.tanda@unipr.it (M.G. Tanda); jaime@dihma.upv.es (J.J. Gómez-Hernández)

Authorship Statement:

Valeria Todaro gave substantial contributions to the development of ideas, numerical experiments, writing the manuscript;

Marco D'Oria gave substantial contributions to the development of ideas, the conception of experiments and the numerical experiments;

Maria Giovanna Tanda and J. Jaime Gómez-Hernández gave substantial contributions to the development of ideas, the conception of experiments and the critical revision of the manuscript.

Highlights:

- Reverse flow routing is solved by means of an ensemble Kalman filter technique
- Inflow hydrographs in ungauged sites can be estimated using downstream information
- Covariance localization and inflation techniques improve performance
- Uncertainty of both parameters and reproduced observations can be quantified
- The proposed technique can be coupled with almost any forward model

Abstract:

The reverse flow routing is an inverse procedure aimed at estimating the inflow to a hydraulic system based on information collected downstream. The hydraulic system can be a river reach or a water reservoir. In this paper, we propose a new approach for the solution of the reverse flow routing problem based on the Ensemble Smoother with Multiple Data Assimilation (ES-MDA). The objective is the estimation of an unknown inflow hydrograph discretized in time by coupling ES-MDA with a given forward routing model that relates inflow hydrograph and downstream observations.

Two realistic synthetic examples are presented to show the capabilities of the methodology. The first case is an application of the reverse flow routing problem to a linear reservoir, where the outflow hydrograph and the reservoir characteristics are known; the second one focuses on the estimation of the inflow hydrograph to an open channel from water level information recorded downstream. We also investigate the performance of the inverse algorithm, by looking at different ensemble sizes, and using covariance localization and inflation techniques.

Our tests show that the proposed approach provides good results, comparable with those of other optimization methods presented in the recent literature. It accurately reproduces the inflow hydrographs, as well as the observations, with narrow confidence intervals. Although ES-MDA yields better results increasing the ensemble size, significant improvements in the solution are obtained for small ensemble sizes when covariance localization and inflation techniques are applied. The proposed approach can compete in accuracy and speed with other approaches, with the advantage that it is conceptually simple and can be used with almost any forward routing code.

Keywords: Ensemble Kalman filter technique; hydrograph estimation**;** ungauged sites**;** temporal localization**;** covariance inflation.

1 INTRODUCTION

The knowledge of discharge hydrographs at specific river sections is essential for flood-risk assessment, planning and management of water resource systems, or optimization of existing hydraulic infrastructures and design of new ones, among others. However, only few river sections are equipped to record data; therefore, an indirect determination of discharge hydrographs is often required. When a flood wave propagates along a river reach or passes through a reservoir, it usually experiences a delay and an attenuation. Although the forward flow routing (estimation of downstream discharge hydrographs based on information available upstream) is common and widely used by practitioners, the estimation of discharge hydrographs at ungauged sections that do not have reliable data upstream is still challenging. Discarding the use of rainfall-runoff models, due to their high uncertainty, a technique that could overcome this problem is the reverse flow routing process that couples the information recorded downstream (discharges or water levels) and the channel or reservoir characteristics to estimate the upstream inflow. The two main approaches to solve this problem in open channels are the application of hydrological routing models (see e.g. Das, 2009; Koussis and Mazi, 2016) in a reverse form, and the backward solution in time of the de Saint Venant equations (see e.g. Eli et al., 1974; Szymkiewicz, 1993; Bruen and Dooge, 2007). A more recent approach makes use of optimization procedures to determine the hydrograph that, once propagated downstream, reproduces the available observations. Saghafian et al. (2015) and Zucco et al. (2015) coupled a Genetic algorithm with a one-dimensional forward hydraulic model and with a simplified routing model, respectively. D'Oria and Tanda (2012), D'Oria et al. (2014) and Ferrari et al. (2018) applied a Bayesian Geostatistical Approach (BGA) to perform the reverse flow routing in combination with hydraulic models that solve the one-dimensional or two-dimensional shallow water equations. Zoppou (1999) and Aldama and Aguilar (2007) faced the problem of reverse routing of flood hydrographs in reservoirs inverting a simple storage equation under a level pool approximation. Spurious oscillations arise in some circumstances; D'Oria et al. (2012) and Leonhardt et al. (2014) solved this problem applying a stochastic approach based on BGA.

We propose a new inverse procedure based on ensemble Kalman filtering (EnKF). Since the introduction of the EnKF by Evensen (1994a, 1994b), the method has been widely applied for data assimilation and the estimations of system states and parameters. EnKF has been applied in many fields, such as oceanography (Bertino et al., 2003; Keppenne and Rienecker, 2003), meteorology (Houtekamer and Zhang, 2016), hydrology (Chen and Zhang, 2006; Li et al., 2012; Moradkhani et al., 2005; Reichle et al., 2002; Xu et al., 2013; Xu and Gómez-Hernández 2016, 2018; Xue and Zhang., 2014) and petroleum engineering (Aanonsen et al., 2009; Gu et al., 2007). The main advantages of the EnKF methods, useful for our purposes, are summarized in three aspects. First, they are more computationally efficient than other Monte Carlo inverse modeling methods due to the way the covariances are computed (Hendricks Franssen and Kinzelbach, 2009) and their implementation is amenable to parallel computing. Second, they can easily be coupled with different forward models for the solution of inverse problems. And third, the ensemble-based methods inherently allow to assess the uncertainty associated with the estimations due to the generation of multiple alternative realizations that can be used for this purpose.

In this paper, we handle a parameter estimation problem, where the parameters are represented by the temporal discretization of an upstream unknown discharge hydrograph and the observations are water levels or a flow hydrograph observed downstream. In addition of the novelty application of the EnKF to the reverse flood routing problem, another innovation of the proposed approach is the identification of parameters that are time dependent, in contrast with previous applications that only estimate time-invariant parameters.

Among the ensemble-based methods (Hamill and Snyder, 2000; Moradkhani et al., 2005; Sakov et al., 2012; Zhou et al., 2011), we consider suitable for the solution of the reverse flow routing problems the Ensemble Smoother with Multiple Data Assimilation (ES-MDA). This method is a valid alternative to the EnKF, for the case in which the time sequence of state observation is all available in full at the time of the analysis; we refer to Li et al. (2018a) for a comparison between

EnKF and ES methods. ES-MDA, introduced by Emerick and Reynolds (2012, 2013), is a variant of the Ensemble Smoother, initially proposed by van Leeuwen and Evensen (1996). ES-MDA iteratively assimilates the same data multiple times in order to improve the results of the ES, which assimilates all data simultaneously in a single update step. The purpose of the multiple assimilation is to avoid the problems detected by Evensen and van Leeuwen (2000) and Crestani et al. (2013) with the ES on its application to highly nonlinear problems with a single global update.

We also test how some modifications on the algorithm may improve the ES-MDA performance and overcome the well-known problem of undersampling in ensemble-based methods. Undersampling occurs when the size of the ensemble is so small that it is not statistically representative of the variability of the unknowns; this leads to two main problems: filter divergence and the appearance of long-range spurious correlations. The filter divergence can be handled by covariance inflation techniques; whereas, covariance localization methods help in removing long-range spurious correlations and, at the same time, in increasing the effective ensemble size expanding the degrees of freedom available to assimilate data (Houtekamer and Mitchell 1998; Hamill et al. 2001). In the literature, localization techniques are applied to cutoff spurious correlations among spatial dependent variables; however, we deal with time series and, therefore, a temporal localization is applied considering time lapses rather than spatial distances.

The paper is organized as follows: in the next section, we describe the ES-MDA implementation and the procedure for temporal localization. Then, two realistic synthetic examples are presented: (1) a reverse flow routing problem within a linear reservoir; (2) an application of the reverse flow routing in an open channel. For the second problem, which is nonlinear, the impact of the ensemble size and the localization technique are also tested; moreover, ES-MDA is coupled with the widely used HEC-RAS river analysis system (Brunner, 2010) that solves the one-dimensional shallow water equations. Finally, a comparison between ES-MDA and BGA is reported. The last section presents the conclusions of the paper.

2 METHODS

The objective of this paper is to show the applicability of ensemble Kalman-based methods, specifically the ES-MDA, for the solution of reverse flow routing problems. The ES-MDA is an iterative data assimilation method that updates the unknown parameters (discretized-in-time upstream hydrograph in our case) maintaining consistency with the observations (downstream hydrograph or water levels observed at specific times). The relationship between parameters and observations must be known and a forward model must be available. We do not focus on the setup of the flow and storage models used for the forward routing, due to the capability of the ES-MDA to be coupled with almost any model.

In the following, we present an overview of the ES-MDA procedure, extensively described by Emerick and Reynolds (2013) and Evensen (2018), and the algorithm adaptation to use a temporal localization and a covariance inflation.

2.1 ES-MDA

Consider the following discrete form of a forward model

$$\mathbf{Y} = \mathrm{g}(\mathbf{X}); \tag{1}$$

given a realization of the model parameters $\mathbf{X} \in \mathcal{R}^{N_P}$, the operator $\mathrm{g}(\mathbf{X})$ predicts the system state at measurement locations, $\mathbf{Y} \in \mathcal{R}^{m}$. Here, N_P is the number of parameters, which, in our case, depends on the duration of the unknown discharge hydrograph and the time step selected for its discretization and m is the number of available observations. The inverse problem aims at finding the parameter vector $\mathbf{X}$ making use of a set of observations $\mathbf{D} \in \mathcal{R}^{m}$ of the system state $\mathbf{Y}$. The solution of this problem by means of ES-MDA requires an initialization phase (step 0) and two main iterative steps: (1) a forecast step in which predictions are made (by means of the forward model) for an ensemble of parameters $\mathbf{X}$; (2) an update step in which parameters are corrected based on the misfit between observations $\mathbf{D}$ and corresponding model predictions $\mathbf{Y}$.

The iterative procedure of ES-MDA avoids the problem of overcorrection that may occur in the single global update of ES, leading to unrealistic estimates of some parameters especially when the misfit between prediction and observations is large. ES-MDA overcomes this problem assimilating the same data multiple times with an inflated measurement error covariance matrix. This is accomplished defining a coefficient α_i that, at each iteration i, applies to the measurement error and its covariance matrix and must satisfy the condition:

$$\sum_{i=1}^{N} \frac{1}{\alpha_i} = 1. \tag{2}$$

where N is the total number of iterations. This condition guarantees an exact equivalence between single and multiple data assimilation methods at least for linear models (Evensen, 2018).

In the following, the ES-MDA scheme is presented.

0. Initialization step

   At the beginning of the procedure, an initial ensemble of parameters must be defined. The ensemble realizations should be generated using all the available information, but often no prior data are available. In our specific case, since a flow hydrograph is a continuous function of time, imposing some degree of continuity in the prior information (initial ensemble) can lead to a smooth solution consistent with the available data.

   The second preliminary step is the choice of the number of iterations N and the coefficients α_i.

   We follow the scheme proposed by Evensen (2018) for the computation of the α_i, which ensures that the constraint of Eq. (2) is satisfied. The procedure starts selecting any nonzero value for α'_1, then the following α'_i are computed as:

$$\alpha'_{i+1} = \alpha'_i / \alpha_{geo}, \tag{3}$$

where the constant α_{geo} controls the extent of the change of α_i from one iteration to the next. At the end, the values from Eq (3) are scaled to obtain the final coefficients:

$$\alpha_i = \alpha_i' \left(\sum_{i=1}^{N} \frac{1}{\alpha_i'} \right). \tag{4}$$

The simplest choice is to consider α_{geo}=1 that leads to a constant $\alpha_i = N$. However, a gradual decrease of α_i, obtained with α_{geo}>1, can improve the performance of the method, since it reduces the magnitude of the initial updates in which the misfit between observations and model predictions is usually larger.

After the initialization step, a loop is started on the number of iterations.

1. Forecast step

The forward model is run on each realization j of the parameter ensemble. For the first iteration, predictions are generated using the initial ensemble of parameters; for the following iterations, predictions are generated using the parameters updated at the end of the previous iteration (i is the iteration index):

$$\mathbf{Y}_{j,i} = \mathbf{g}(\mathbf{X}_{j,i}). \tag{5}$$

2. Update step

The vector of parameters for realization j is updated at each iteration i as follows:

$$\mathbf{X}_{j,i+1} = \mathbf{X}_{j,i} + \frac{\mathbf{C}_{\mathbf{XY}}^{i}}{\mathbf{C}_{\mathbf{YY}}^{i} + \alpha_i \mathbf{R}} \times \left(\mathbf{D} + \sqrt{\alpha_i}\boldsymbol{\varepsilon}_j - \mathbf{Y}_{j,i} \right), \tag{6}$$

where $\boldsymbol{\varepsilon}_j$ is the measurement error, which is assumed to follow a Gaussian distribution of mean zero and covariance matrix $\mathbf{R} \in \mathcal{R}^{m\times m}, \mathcal{N}(0, \mathbf{R})$. $\mathbf{C}_{\mathbf{XY}}^{i} \in \mathcal{R}^{n\times m}$ is the cross covariance matrix between the vectors of parameters and predictions and $\mathbf{C}_{\mathbf{YY}}^{i} \in \mathcal{R}^{m\times m}$ is the auto-covariance matrix of the prediction vector.

The covariance matrices are computed from the ensemble of realizations at each iteration i as:

$$\mathbf{C}_{\mathbf{XY}}^{\mathrm{i}} = \frac{1}{\mathrm{N_e}-1}\sum_{\mathrm{j}=1}^{\mathrm{N_e}}\left(\mathbf{X}_{\mathrm{j,i}} - \overline{\mathbf{X}}_{\mathrm{i}}\right)\left(\mathbf{Y}_{\mathrm{j,i}} - \overline{\mathbf{Y}}_{\mathrm{i}}\right)^{\mathrm{T}}, \tag{7}$$

$$\mathbf{C}_{\mathbf{YY}}^{\mathrm{i}} = \frac{1}{\mathrm{N_e}-1}\sum_{\mathrm{j}=1}^{\mathrm{N_e}}\left(\mathbf{Y}_{\mathrm{j,i}} - \overline{\mathbf{Y}}_{\mathrm{i}}\right)\left(\mathbf{Y}_{\mathrm{j,i}} - \overline{\mathbf{Y}}_{\mathrm{i}}\right)^{\mathrm{T}}, \tag{8}$$

where $\mathrm{N_e}$ is the total number of ensemble realizations and $\overline{\mathbf{X}}_{\mathrm{i}}$ and $\overline{\mathbf{Y}}_{\mathrm{i}}$ are the ensemble means of parameters and predictions, respectively.

The update step can be performed in a transformed space, such as the log-space, in order to prevent the appearance of unphysical negative values. In these cases, the vector of parameters is log transformed before the update step and back transformed into the parameter space after the updating. Covariances and cross-covariances must be computed in the transformed space, too.

Then, return to the forecast step considering $\mathbf{X}_{\mathrm{j,i}} = \mathbf{X}_{\mathrm{j,i-1}}$ and repeat until the last iteration.

It should be noted that the computation of the prediction vectors $\mathbf{Y}_{\mathrm{j,i}}$ requires $\mathrm{N_e}$ simulations at each iteration. Therefore, the number of ensemble realizations should be kept as small as possible to reduce computational time. However, a small ensemble size could create spurious correlation in the covariance matrices $\mathbf{C}_{\mathbf{XY}}^{\mathrm{i}}$ and $\mathbf{C}_{\mathbf{YY}}^{\mathrm{i}}$ and lead to filter divergence. In the following section, we discuss covariance localization and covariance inflation techniques, since they are the standard approaches to overcome this problem.

2.2 Covariance localization and covariance inflation

Covariance localization (CL) is a technique developed to mitigate the problem of long-range spurious correlations that could arise when the covariances are approximated from a small number of ensemble realizations not accurately reflecting the statistics of the underlying population. At the same time, CL expands the degrees of freedom available to assimilate data since it increases the rank of the ensemble derived covariance matrices —usually rank deficient— even more so when the ensemble size is lower than the number of unknown parameters or observations.

CL is done by element-wise multiplication (Schur product or Hadamard product) of the original covariance matrix and a distance-dependent correlation function ρ that smoothly reduces the correlations between points for increasing distances and cuts off long-range correlations above a specific distance. CL modifies the update step of the ES-MDA by modifying the covariances in Eq. (7) and (8) as follows:

$$\tilde{\mathbf{C}}_{\mathbf{XY}}^{\mathrm{i}} = \boldsymbol{\rho}_{\mathbf{XY}} \circ \mathbf{C}_{\mathbf{XY}}^{\mathrm{i}}, \tag{9}$$

$$\tilde{\mathbf{C}}_{\mathbf{YY}}^{\mathrm{i}} = \boldsymbol{\rho}_{\mathbf{YY}} \circ \mathbf{C}_{\mathbf{YY}}^{\mathrm{i}}, \tag{10}$$

where $\circ$ represents the Schur product and the elements of the matrices $\boldsymbol{\rho}_{\mathbf{XY}} \in \mathcal{R}^{n\times m}$ and $\boldsymbol{\rho}_{\mathbf{YY}} \in \mathcal{R}^{m\times m}$ are based on temporal distances between parameters and observations (δ_{XY}) and between observations and observations (δ_{YY}), respectively. In this work, we use the fifth-order distance dependent localization function as defined in Gaspari and Cohn (1999):

$$\rho = \begin{cases} -\frac{1}{4}\left(\frac{\delta}{b}\right)^5 + \frac{1}{2}\left(\frac{\delta}{b}\right)^4 + \frac{5}{8}\left(\frac{\delta}{b}\right)^3 - \frac{5}{3}\left(\frac{\delta}{b}\right)^2 + 1, & 0 \le \delta \le b; \\ \frac{1}{12}\left(\frac{\delta}{b}\right)^5 - \frac{1}{2}\left(\frac{\delta}{b}\right)^4 + \frac{5}{8}\left(\frac{\delta}{b}\right)^3 + \frac{5}{3}\left(\frac{\delta}{b}\right)^2 - 5\left(\frac{\delta}{b}\right) + 4 - \frac{2}{3}\left(\frac{\delta}{b}\right)^{-1}, & a \le \delta \le 2b; \\ 0 & \delta \ge 2b; \end{cases} \tag{11}$$

where the coefficient b characterizes the time distance at which the covariances become zero; we have chosen the same value of b for both the parameter-observation and observation-observation covariance functions.

Covariance inflation is a technique developed to overcome the problem of filter divergence. The filter divergence may occur when the variance is underestimated leading to overconfidence in prior estimates and, as a consequence, the ensemble collapses into a set of too similar realizations, which could be different from the true solution. This reduces the weight given to subsequent updates and can lead to a divergence of the ensemble since the filter is not able to adjust an incorrect estimation.

Covariance inflation can be achieved by different ways (see e.g. Anderson, 2007; Li et al., 2009; Liang et at., 2011; Wang and Bishop, 2003; Zheng, 2009); in this work we follow the scheme introduced by Anderson and Anderson (1999). Each realization of the ensemble (Eq. 4) at the end of each update step, $\mathbf{X}_{\mathrm{j,i}}$, is linearly inflated around its mean, $\overline{\mathbf{X}}_{\mathrm{i}}$, using:

$$\widetilde{\mathbf{X}}_{\mathrm{j,i}} = r\left(\mathbf{X}_{\mathrm{j,i}} - \overline{\mathbf{X}}_{\mathrm{i}}\right) + \overline{\mathbf{X}}_{\mathrm{i}}, \tag{12}$$

where r is an inflation factor slightly larger than 1.

3 APPLICATIONS AND RESULTS

The application of the ES-MDA for solving reverse flow routing problems is demonstrated by means of two realistic synthetic examples. First, in order to show the capabilities of the methodology, we consider a linear inverse problem: the estimation of the inflow hydrograph to a linear reservoir based on the knowledge of the outflow hydrograph and the reservoir characteristics. In the second test we apply the methodology to a nonlinear problem of reverse flow routing in an open channel, where the unknown parameters are the discharge values of the discretized upstream hydrograph, the observations are the water levels recorded in a downstream section and a calibrated numerical model for the computation of the forward routing is available.

For both applications, the inflow hydrograph, I, to be estimated is a multi-peak wave modeled as the summation of M gamma functions, that is:

$$\mathrm{I(t)} = \mathrm{A} + \sum_{\mathrm{r=1}}^{\mathrm{M}} \mathrm{B_r} \cdot \mathrm{f_r(t|n_r, k_r)}, \tag{13}$$

where t is the time, A [L^3T^{-1}] represents the base flow, B [L^3] the flood volume of each gamma wave r, and f [T^{-1}] is a gamma distribution function of coefficients n (shape) and k (scale):

$$f(t|n,k) = \frac{1}{k^n\,\Gamma(n)} t^{n-1} e^{-t/k}, \tag{14}$$

where $\Gamma(n)$ is the gamma function.

The synthetic test cases allow the comparison between the results of the inverse algorithm and the reference solution. In our study, the performance of the methodology is evaluated using three different metrics: the root mean square error (RMSE), the Nash-Sutcliffe efficiency criterion (Nash and Sutcliffe, 1970) (NSE) and the relative error in the peak discharge (E_P). RMSE is computed as:

$$RMSE = \sqrt{\frac{\sum_{d=1}^{N_P}(I_d - \overline{X}_d)^2}{N_P}}, \tag{15}$$

where N_P is the number of parameters, I_d is the d-th true inflow discharge and $\overline{X}_d$ is the ensemble mean of the d-th estimated inflow discharge.

NSE is defined as:

$$NSE = \left(1 - \frac{\sum_{d=1}^{N_P}(I_d - \overline{X}_d)^2}{\sum_{d=1}^{N_P}(I_d - \overline{I}_d)^2}\right) \cdot 100, \tag{16}$$

where $\overline{I}_d$ is the mean of the true inflow hydrograph. NSE=100% indicates a perfect match between estimated and actual discharges.

E_P is evaluated as:

$$E_P = \left(\frac{I_p}{\overline{X}_P} - 1\right) \cdot 100, \tag{17}$$

where I_p and $\overline{X}_P$ represent the true and estimated (ensemble mean) peaks of the inflow hydrographs, respectively.

The results of the second synthetic example are also compared with those obtained applying the Bayesian Geostatistical Approach (BGA) proposed by D'Oria and Tanda (2012). BGA needs multiple iterations to reach an optimal solution due to the nonlinearity of the forward problem and the need to estimate the hyperparameters of the prior covariance model, which control the structure of the unknown hydrograph, in addition to the discharge values (parameters). At each inner linearization iteration, the Jacobian matrix (sensitivity of observations to unknown parameters) must be calculated and it requires, in a finite difference approximation, as many forward model runs as the number of parameters, N_P, plus 1. Therefore, the total number of forward model runs, N_t, required by BGA is:

$$\mathrm{N_t} = (\mathrm{N_P} + 1)\mathrm{N_o}\mathrm{N_i} + 1, \tag{18}$$

where $\mathrm{N_o}$ and $\mathrm{N_i}$ are the numbers of BGA iterations needed for hyperparameters (outer loop) and parameters estimation (inner loop), respectively.

3.1 Case 1: Linear reservoir

In this example, we try to estimate the inflow hydrographs to a reservoir based on the knowledge of the outflow hydrograph and the reservoir characteristics.

Under the level pool routing approximation (reservoir dynamics are negligible and water surface inside the reservoir is horizontal), the inflow $\mathrm{I(t)}$ and the outflow $\mathrm{Q(t)}$ in a reservoir are related by a simple continuity equation:

$$\mathrm{I(t)} - \mathrm{Q(t)} = \frac{\mathrm{dS}}{\mathrm{dt}}, \tag{19}$$

where S is the instantaneous volume stored in the reservoir and t is the time. The outflow discharge is related to the storage; for a linear reservoir it can be expressed as:

$$\mathrm{S(t)} = \mathrm{KQ(t)}, \tag{20}$$

where the constant proportionality factor K [T] is known as the storage coefficient.

The solution of the continuity equation for the linear reservoir (starting from a steady state condition) on a continuous time scale is represented by the following convolution integral (Chow et al., 1988):

$$\mathrm{Q(t)} = \int_0^{\mathrm{t}} \frac{1}{\mathrm{K}} e^{-(\mathrm{t}-\tau)/\mathrm{K}}\, \mathrm{I}(\tau) d\tau. \tag{21}$$

A solution at discrete intervals of time can be obtained by means of a discrete convolution equation.

The synthetic test considers a reservoir with storage coefficient K=3 h and an inflow hydrograph with two peaks as defined by Eq. (13) (M=2) and the coefficients reported in Table 1. The resulting hydrograph has a first peak of about 500 m^3/s at 3.5 h and a second peak with a discharge of about 240 m^3/s at 11.4 h.

The total simulation time is 30h. The inflow hydrograph is discretized in equal interval of 9 min resulting in a number of parameters to be estimated $\mathrm{N_P}$=201.

Table 1 - Case 1: coefficients of the two gamma functions used for the description of the inflow hydrograph.

	A [m^3/s]	B [m^3]	n [-]	k [h]
f_1	50	$5.5 \cdot 10^6$	8	0.5
f_2		$4.5 \cdot 10^6$	20	0.6

Preliminarily, the actual inflow hydrograph is forward routed through Eq. (21) to obtain the true outflow hydrograph; this last one was observed every 6 min for a total of 301 observations (m=301) to be sued in the inverse procedure. In applying the ES-MDA procedure, we considered an observation error ε equal to 5% of the true discharge values.

The initial ensemble (Fig. 1) is composed of 200 realizations of the inflow hydrograph. They are all individual gamma functions generated using Eq. (13) with M=1 and the other coefficients selected randomly over a wide range of values. In particular, the range is [10, 150] m^3/s for A, [$1.5 \cdot 10^5$,

$5.0 \cdot 10^7$] m^3 for B, [3, 10] for n and [0.7, 4.5] h for k; the extremes of the ranges, selected on the basis of expert knowledge, guarantee that all the realizations are consistent with the considered problem.

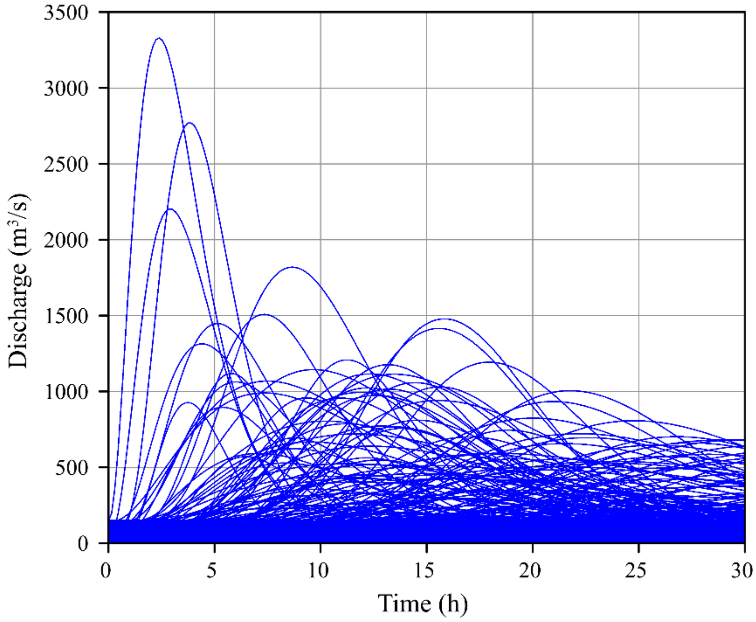

Fig. 1. Case 1: initial ensemble of inflow hydrograph (200 realizations).

For the ES-MDA, we choose to perform 5 iterations with a constant α equal to 5 (α_{geo}=1, Eq. (3)-(4)). In this case, no localization or inflation are applied, and a large ensemble is considered with the aim to show the capability of the method.

Fig. 2 presents the results of the inversion at the end of the iterative process: it shows the ensemble mean of the estimated inflow and outflow hydrographs with their 95% confidence interval computed from the ensemble; the actual inflow and outflow hydrographs are reported for comparison.

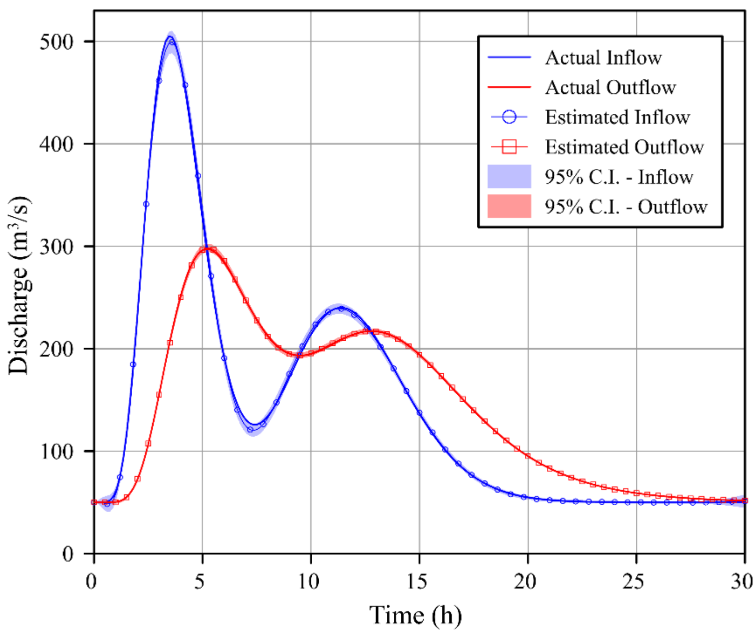

Fig. 2. Case 1: actual and estimated inflow and outflow hydrographs with 95% credibility intervals.

The ES-MDA method accurately reproduces the inflow hydrograph (NSE = 99.94%) with a very narrow confidence interval, as well as the simulated outflow hydrograph. The RMSE at each iteration, shown in Fig. 3, slightly decreases during the procedure reaching the lowest value of 2.9 m^3/s at the end of the simulation. The two inflow peaks and their timing are properly reproduced with a slight underestimation ($E_{P,1}$=-1.1%; $E_{P,2}$=-0.4%, where the subscript 1 stands for the first peak and 2 for the second one).

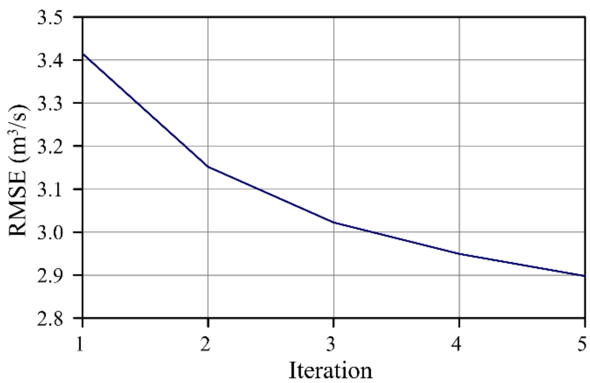

Fig. 3. Case 1: root-mean-square error (RMSE) of the estimated inflow hydrograph at each iteration.

1.1 Case 2: Open channel

The second test focuses on the estimation of the inflow hydrograph to an open channel based on water level information collected in a downstream section using a given numerical model for the forward routing. In this work, we used the Hydrologic Engineering Center's River Analysis System (HEC-RAS), developed by the US Army Corps of Engineers (Brunner, 2010), that simulates one-dimensional unsteady flow by solving the Saint-Venant equations.

We have considered a prismatic channel, 20 km long, with a longitudinal slope of 0.0005 and compound cross sections spaced by 250 m consisting of a trapezoidal main channel and two symmetric floodplains (Fig. 4). The main channel has a bottom width of 50 m, a side slope of 2 and a depth of 6 m; each floodplain has a width of 50 m, horizontal bottom and vertical banks. Manning coefficients of 0.05 $m^{-1/3}/s$ and 0.1 $m^{-1/3}/s$ are adopted for the main channel and the floodplain, respectively.

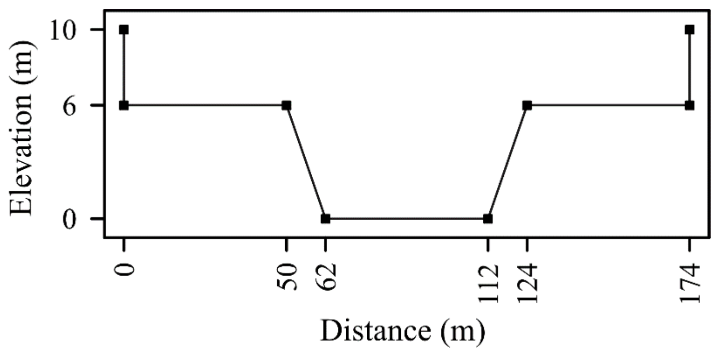

Fig. 4. Case2: compound cross section of the prismatic channel.

The true upstream hydrograph is defined by Eq. (13) (M=2) with the coefficients reported in Table 2. The hydrograph has a first peak of about 1000 m^3/s at 5 h, a second peak with a discharge of about 500 m^3/s at 14 h and a base flow of 50 m^3/s. The total simulation time is 30 h and the upstream hydrograph is discretized in equal intervals of 30 min (N_P=61).

Table 2 - Case 2: coefficients of the two gamma functions used for the description of the inflow hydrograph.

	A $[m^3/s]$	B $[m^3]$	n [-]	k [h]
f_1	50	$1.6 \cdot 10^7$	8	0.7
f_2		$1.4 \cdot 10^7$	18	0.8

The initial condition is obtained from a steady-state simulation according to the first inflow discharge value, assuming a steady-state condition before the flood event. The upstream and downstream boundary conditions are represented by the inflow hydrograph and the normal depth based on the Manning's equation, respectively.

The actual inflow hydrograph has been forward routed by means of HEC-RAS to obtain the water levels used as observations, which are recorded in the section in the middle of the channel, located 10 km downstream from the upstream section, every 30 min (m=61). We consider a random observation error ε with normal distribution, zero mean and variance $2.8 \cdot 10^{-4}$ m^2, that results in the 99.7% of the cases in errors in the range ±0.05m.

In this work, different settings of the inverse algorithm have been tested in the estimation of the upstream hydrograph; we analyzed the impact of the ensemble size, the choice of the coefficient α during the iteration process, the covariance localization and the covariance inflation techniques.

We tested three ensemble sizes equal to: half the number of parameters (N_e=31), the number of parameters (N_e=61) and three times the number of parameters (N_e=183). All the realizations of the initial ensembles are individual gamma functions generated using Eq. (13) with M=1 and coefficients selected randomly over the same wide range of values ([2, 180] m^3/s for A, [$8 \cdot 10^4$, $8 \cdot 10^7$] m^3 for B, [3, 18] for n, [0.6, 4.8] h for k).

For each ensemble size, we carry out four tests: the first test (T1) is performed with a constant coefficient α (Eq. (4)) used for all iterations and without other modifications on the inverse algorithm; the second one (T2) attempts to evaluate the effect of decreasing coefficient α as iterations progress; the third one (T3) studies the effect of covariance localization and covariance inflation keeping α constant; and the last test (T4), combines covariance modification (localization and inflation) with a decreasing α.

We decided to perform 6 iterations for each test, with a constant α equal to 6 (α_{geo}=1, Eq. (3)-(4)), for test T1 and T3 and a decreasing α = [364; 121.33; 40.44; 13.48; 4.49; 1.50], obtained with α_{geo}=3 (Eq. (3)-(4)), for T2 and T4 (recall that the sum of the inverses of α values should add up to 1 (Eq. (2)). Covariance localization and covariance inflation are applied using the coefficient *b* equal to 6 h (Eq. (9)) and the inflation factor equal to 1.01 (Eq. (12)), respectively. In this case, the update step is performed in logarithmic space to avoid the appearance of negative values.

The results of all tests are compared in term of the root mean squared error (RMSE) between the estimated hydrograph and the reference solution (Fig. 4). In all cases, the RMSE significantly decreases at each iteration, reaching low values at the end of the inversion. For the smaller ensemble size (Fig. 5a) the method performs better when a decreasing α (T2) is used and when covariance inflation and localization techniques are used (T3), with the best results obtained when both options are used simultaneously (T4). For the larger ensemble size (Fig. 5b; Fig. 5c), the final RMSE is always smaller than for the smaller ensemble, and in all four experiments the final hydrograph is very close to the real one. Yet, the best performance, at the last iteration, is obtained for the experiment T4. Table 3 reports the RMSEs at the end of each test, together with the Nash-Sutcliffe efficiency criterion and the relative errors in the peak discharge.

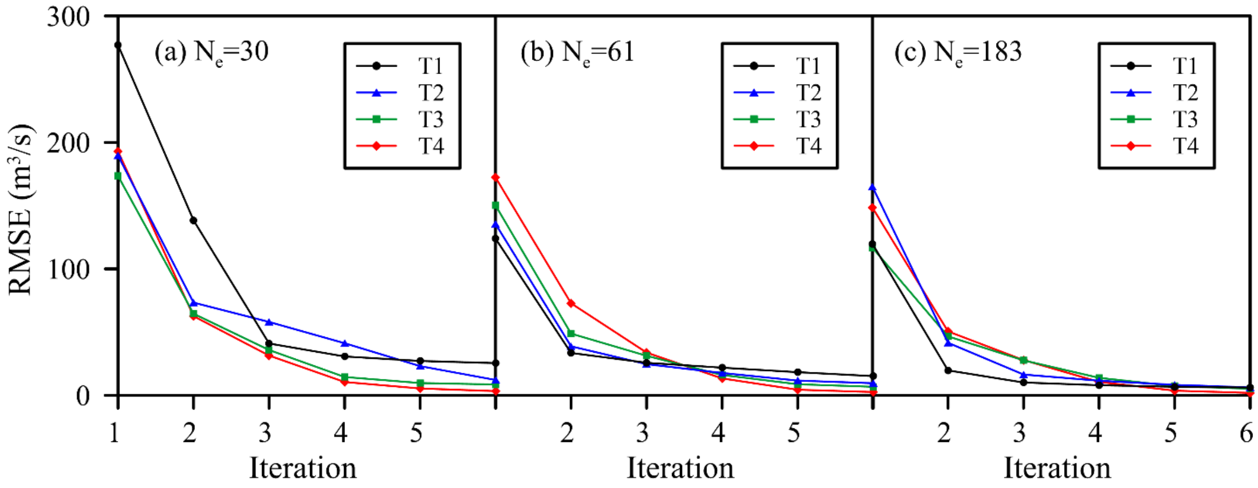

Fig. 5. Case 2: RMSE of the estimated inflow hydrograph for ensemble size N_e=30 (a), N_e=61 (b) and N_e=138 (c).

All the NSE values are above 99% indicating an accurate reproduction of the shape of the upstream hydrograph; the peaks are properly reproduced with E_P values lower than 2.15%, with only an exception (T1, N_e=30). Like RMSE, the metrics NSE and E_P confirm that decreasing α during the iterative process and adopting covariance modification techniques improve the performance of the ES-MDA especially when a small ensemble size is used.

Table 3 - Case 2: root mean square error (RMSE), Nash-Sutcliffe efficiency criterion (NSE) and relative error in the peak discharge (E_P) between estimated and true inflow hydrographs for the four different tests (T1-T4) and for ensemble sizes N_e=30, 61, 183 at the end of the iterative process.

	N_e=30				N_e=61				N_e=183			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
RMSE [m^3/s]	25.47	12.12	8.57	3.32	15.17	9.53	6.71	2.56	6.11	6.17	5.17	1.89
NSE[%]	99.06	99.78	99.89	99.98	99.67	99.87	99.93	99.99	99.95	99.94	99.96	99.99
$\mathbf{E_{P,1}}$[%]	8.15	-0.18	1.79	-0.20	0.73	2.13	0.27	-0.28	1.28	1.53	-1.01	-0.66
$\mathbf{E_{P,2}}$[%]	-0.11	1.61	2.09	-0.24	0.62	0.46	1.47	0.36	0.65	0.73	0.87	-0.10

For the sake of brevity, we show only the hydrographs resulting from the inversion obtained when the ensemble size is small and for two of the experiments, the one with no modifications of the ES-

MDA algorithm (T1) and the one using a decreasing α and covariance localization and inflation techniques (T4). In Fig. 6 the true values and the ensemble means of the estimated inflow hydrographs with their 95% confidence intervals are depicted. Fig. 7 shows the true observations and the ensemble means of the estimated water levels with their 95% confidence interval. In both figures the residuals between actual and estimated values are also shown.

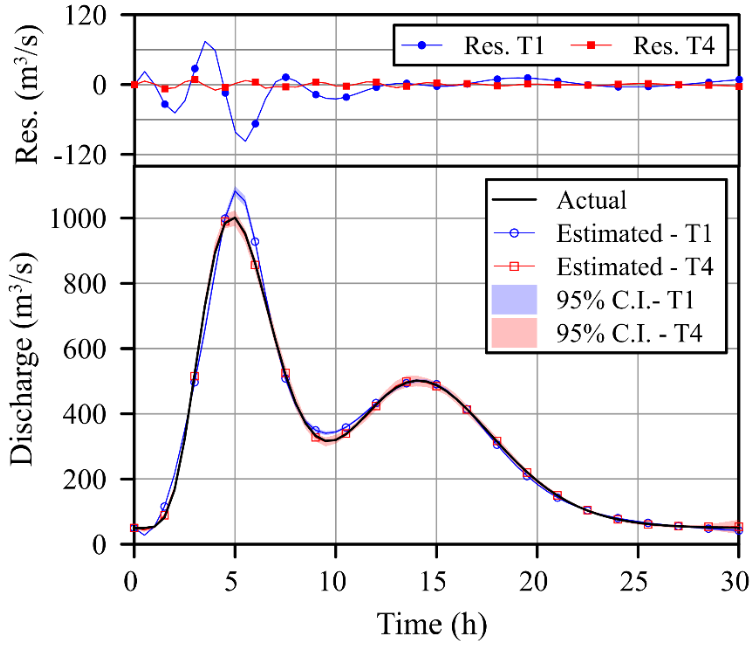

Fig. 6. Case 2: actual and estimated upstream hydrographs with 95% confidence intervals (bottom) and residuals between actual and estimated values (top) resulting from tests T1 and T4 with N_e=30.

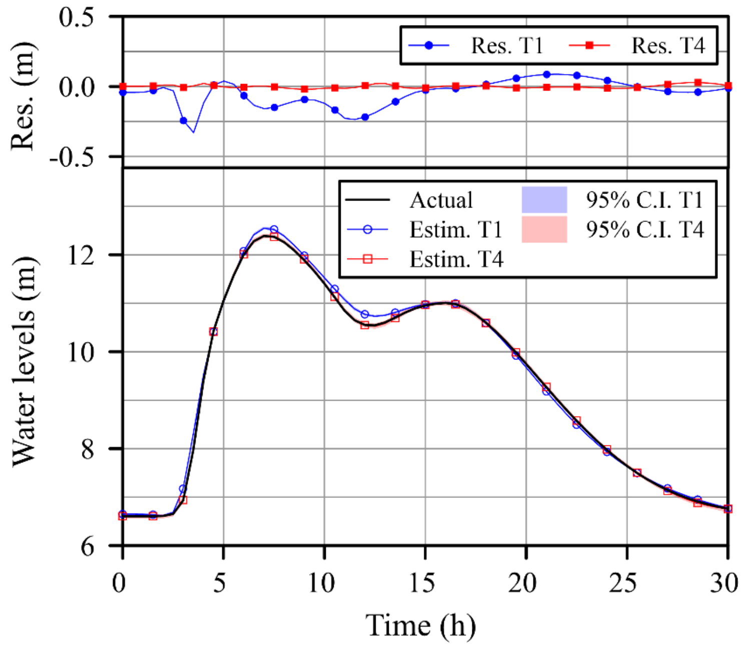

Fig. 7. Case 2: actual and estimated water levels with 95% confidence intervals (bottom) and residuals between actual and estimated values (top) resulting from tests T1 and T4 with N_e=30.

Test T1 reproduces the shape of the inflow hydrograph quite well (NSE = 99.06%), but with a larger error on the first peak ($E_{P,1}$=8.15); the observations are not perfectly reproduced everywhere and the residuals are high in some points. Meanwhile, test T4 leads to a good match between the true and estimated inflow hydrograph (NSE = 99.98%) and the true and estimated water levels with very small residuals. The inflow peaks and their timing are properly reproduced with negligible errors ($E_{P,1}$=-0.2%; $E_{P,2}$=-0.4%).

Finally, we compare the results of test T4, obtained with the smaller ensemble size, with those of the Bayesian Geostatistical Approach. The test is performed coupling BGA with the same forward model used for the solution of Case 2, considering the same simulation time (30 h) and discretization of the unknown hydrograph (N_P=61). The true observations were perturbed with random errors with zero mean and variance $2.8 \cdot 10^{-4}$ m^2. We selected a number of iterations for the linearization process (inner loop) equal to N_i=5 and equal to N_o=4 for the outer loop required to estimate the hyperparameters.

The results of the comparison are reported in Fig 8. The BGA method accurately estimates the inflow hydrograph (RMSE=4.2 m^3/s, NSE=99.97%) with small residuals and small errors in the estimation of the peaks ($E_{P,1}$=-1.0%; $E_{P,2}$=-0.3%). The two approaches show fully comparable results, which are confirmed by a very similar residual range and the almost equal values of the performance metrics. However, ES-MDA outperform BGA in terms of total number of forward model runs required and hence computational time: 1241 (Eq 18) runs for BGA and 182 ($N_e \cdot N_i$) for ES-MDA.

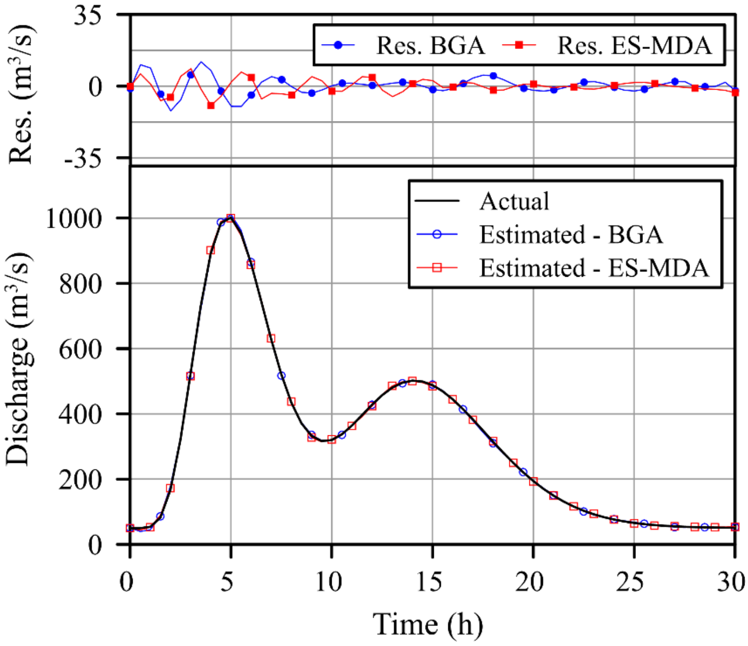

Fig. 8 Case 2: actual and estimated upstream hydrographs (bottom) and residuals between actual and estimated values (top) resulting from BGA and ES-MDA (T4, N_e=30) approaches.

4 SUMMARY AND CONCLUSIONS

In this paper, we propose a new approach for the solution of the reverse flow routing problem using an ensemble Kalman filter technique: the Ensemble Smoother with Multiple Data Assimilation (ES-MDA). The unknown parameters, represented by the temporal discretization of the upstream hydrograph, are estimated based on discharge or water level information recorded downstream. Two synthetic examples are considered to test the methodology. The first case shows the capability of the inverse procedure in estimating the inflow hydrograph to a linear reservoir, where the outflow hydrograph and the reservoir characteristics are known. It is noteworthy that for linear problems the ensemble smoother methods should lead to the exact solution in a single update step, provided that the observations are free of errors and the initial ensemble is statistically representative of the variability of the unknowns. In our case, due to the presence of corrupted observations, the ES-MDA updates the vector of parameters in multiple iterations. At the end of the process, the true inflow hydrograph is accurately reproduced; the Nash-Sutcliffe efficiency criterion (NSE) is

99.94%, the errors in the peak discharges are less than 1.1% and the RMSE reaches the small value of 2.9 m^3/s.

The second case study validates the method for non-linear problems by estimating the inflow hydrograph to an open channel based on water level information collected in a downstream section and for given forward routing model. Four tests were performed to investigate the effect of different settings of the inverse algorithm: the ensemble size, the decreasing α during the iterative process and the temporal localization and inflation of the covariances. In all tests, the NSE exceeds 99% and, as expected, the ES-MDA reaches a better solution increasing the ensemble size. However, as the ensemble becomes larger, the computational time increases, since, at each iteration, the method requires a number of forward runs equal to the number of realizations. The results of our tests show that a significant improvement in the inverse solution is obtained if a decreasing α and the covariance modifications are applied, the ensemble size being equal. This is particularly clear working with small ensemble sizes, since covariance localization and inflation overcome the problem of undersampling that occurs when a low number of realizations is used. The test performed with the smaller ensemble size using a decreasing α and the covariance modifications reproduces very well the inflow hydrograph with negligible errors. The NSE is 99.98 % and the relative error in the peak discharges are less than 0.3%; these values are fully comparable with those obtained with the larger ensemble size. The RMSE is 3.32 m^3/s, which corresponds to a reduction of about 87% compared to test with constant α and the basic algorithm for the same ensemble size.

In summary, the modified ES-MDA method allows to solve the reverse flow routing problems using also small ensemble sizes (with a total number of realizations less than the number of parameters) leading to a significant reduction of the computational burden. The modified algorithm provides results comparable with those of the other optimization methods presented in the recent literature, although ES-MDA achieves the solution with a lower number of forward runs. In addition, the forward runs related to the ensemble realizations can be easily parallelized allowing an additional reduction of the computational time. Moreover, another important advantage of the method is the capability to assess the uncertainty in the estimations from the realizations of the ensemble. It allows to quantify the uncertainty associated with both the unknown parameters and the reproduction of the observations, which is a novelty in the solution of the reverse flow routing problem.

It is noteworthy to point out that one can handle non-Gaussian distributed parameters, and it is well known that the ensemble Kalman filter methods are optimal for multiGaussian distributed variables. Our results, for the analyzed case studies, show that ES-MDA was able to reach a good solution in all cases. However, for those cases in which the method may fail due to the non-Gaussianity of the parameters, different approaches are presented in the literature to overcome the problem; we suggest to couple ES-MDA with the Normal-Score transformation, which it is shown to work properly with ensemble Kalman filter methods (Zhou et al., 2011; Li et al., 2018b).

Finally, another aspect that should be taken into account is the uncertainty in the forward model. Since the inverse methodology requires a numerical model able to accurately describe the forward processes, the errors in the model structure (likely systematic and correlated, see e.g. Gaganis and Smith, 2001) could add to the measurement noise. Therefore, in real applications, a proper and calibrated forward model is crucial to obtain a reliable inverse solution and an examination of the most uncertain model parameters is advisable. Approaches that quantitatively account for structural model uncertainties will be investigated in a future work.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous Reviewers for their valuable and constructive comments. The TeachinParma initiative, co-funded by Fondazione Cariparma and University of Parma (http://www.teachinparma.com/about/), supported Prof. J. Jaime Gómez-Hernández as Visiting Professor at the University of Parma.

REFERENCES

- Aanonsen, S.I., Nævdal, G., Oliver, D. S., Reynolds, A. C., Vallès, B., 2009. The Ensemble Kalman Filter in Reservoir Engineering--a Review. Society of Petroleum Engineers. https://doi.org/10.2118/117274-PA.
- Aldama, A., Aguilar, E., 2007. Flow Rate Estimation via Inverse Flood Routing and Spectral Error Control. American Geophysical Union, Spring Meeting, abstract #H41A-03.
- Anderson J.L., 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. Tellus A: Dynamic Meteorology and Oceanography 59(2), 210-24. https://doi.org/10.1111/j.1600-0870.2006.00216.x.
- Anderson J.L., Anderson S.L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Monthly Weather Review 127, 2741-2758. https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.
- Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential Data Assimilation Techniques in Oceanography. International Statistical Review 71, 223–241. https://doi.org/10.1111/j.1751-5823.2003.tb00194.x.
- Bruen, M., Dooge, J.C.I., 2007. Harmonic analysis of the stability of reverse routing in channels. Hydrology and Earth System Sciences 11, 559–568. https://doi.org/10.5194/hess-11-559-2007.
- Brunner G.W., 2010. HEC-RAS, River analysis system hydraulic reference manual version 4.1. Davis, CA: US Army Corps of Engineers, Institute for Water resource, Hydrologic Engineering Center.
- Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Advances in Water Resources 29(8), 1107-1122. https://doi.org/10.1016/j.advwatres.2005.09.007.
- Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied Hydrology, McGraw Hill, 572 p.
- Crestani, E., Camporese, M., Baú, D., Salandin, P., 2013. Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation, Hydrology and Earth System Sciences 17, 1517-1531. https://doi.org/10.5194/hess-17-1517-2013.
- Das, A., 2009. Reverse stream flow routing by using Muskingum models, Sādhanā 34(3), 483-489. https://doi.org/10.1007/s12046-009-0019-8.
- D'Oria, M., Tanda, M.G., 2012. Reverse flow routing in open channels: A Bayesian Geostatistical Approach. Journal of Hydrology 460, 130-135. https://doi.org/10.1016/j.jhydrol.2012.06.055.
- D’Oria, M., Mignosa, P., Tanda, M.G., 2012. Reverse level pool routing: Comparison between a deterministic and a stochastic approach. Journal of Hydrology 470-471, 28-35. https://doi.org/10.1016/j.jhydrol.2012.07.045.
- D'Oria, M., Mignosa, P., Tanda, M.G., 2014. Bayesian estimation of inflow hydrographs in ungauged sites of multiple reach systems. Advances in Water Resources, 63, 143-151. https://doi.org/10.1016/j.advwatres.2013.11.007.

- Eli R.N., Wiggert J.M., Contractor D.N., 1974. Reverse flow routing by the implicit method. Water Resources Research 10(3), 597-600. https://doi.org/10.1029/WR010i003p00597.
- Emerick, A.A., Reynolds, A.C., 2012. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. Computers & Geosciences 16(3), 639–659. https://doi.org/ 10.1007/s10596-012-9275-5.
- Emerick, A.A., Reynolds, A.C., 2013. Ensemble Smoother with multiple data assimilation. Computers & Geosciences 55, 3-15. http://dx.doi.org/10.1016/j.cageo.2012.03.011.
- Evensen, G., van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics. Monthly Weather Review 128, 1852–1867. https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.
- Evensen, G., 1994a. Inverse methods and data assimilation in nonlinear ocean models. Physica D: Nonlinear Phenomena 77(1), 108–129. https://doi.org/10.1016/0167-2789(94)90130-9.
- Evensen, G., 1994b. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99(C5), 10143-10162. https://doi.org/10.1029/94JC00572.
- Evensen, G., 2018. Analysis of iterative ensemble smoothers for solving inverse problems. Computational Geosciences 22, 885-908. https://doi.org/10.1007/s10596-018-9731-y.
- Ferrari, A., D'Oria, M., Vacondio, R., Dal Palù, A., Mignosa, P., Tanda, M.G., 2018. Discharge hydrograph estimation at upstream-ungauged sections by coupling a Bayesian methodology and a 2-D GPU shallow water model. Hydrology and Earth System Sciences 22, 5299-5316. https://doi.org/10.5194/hess-22-5299-2018.
- Gaganis, P., Smith, L., 2001. A Bayesian Approach to the quantification of the effect of model error on the predictions of groundwater models. Water Resources Research 37(9), 2309–2322. http://dx.doi.org/10.1029/2000WR000001.
- Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Quarterly Journal of the Royal Meteorological Society 125, 723-757. http://dx.doi.org/10.1002/qj.49712555417.
- Gu, Y., Oliver, D.S., 2007. An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation. Society of Petroleum Engineers 12. https://doi.org/10.2118/108438-PA.
- Hamill, T.M., Snyder, C., 2000. A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme. Monthly Weather Review 128(8), 2905–2919. http://dx.doi.org/10.1175/1520-0493(2000)128<2905:ahekfv>2.0.co;2.
- Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review 129, 2776–2790. http://dx.doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

- Hendricks Franssen, H.J., Kinzelbach, W., 2009. Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems, Journal of Hydrology 365(3), 261–74. https://doi.org/10.1016/j.jhydrol.2008.11.033.
- Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filtering technique. Monthly Weather Review 126, 796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Houtekamer, P.L., Mitchell, H.L., 2005. Ensemble Kalman filtering. Quarterly Journal of the Royal Meteorological Society 131, 3269-3289. https://doi.org/10.1256/qj.05.135.
- Houtekamer, P.L., Zhang, F., 2016. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. Monthly Weather Review 144. https://doi.org/10.1175/MWR-D-15-0440.
- Keppenne, C.L., Rienecker, M.M., 2003. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. Journal of Marine Systems 40–41, 363–380. https://doi.org/10.1016/S0924-7963(03)00025-3.
- Koussis, A.D., Mazi, K., 2016. Reverse flood and pollution routing with the lag-and-route model. Hydrological Sciences Journal 61, 1952–1966.
- Leonhardt, G., D'Oria, M., Kleidorfer, M., Rauch, W., 2014. Estimating inflow to a combined sewer overflow structure with storage tank in real time: evaluation of different approaches. Water Science and Technology 70, 1143–1151. https://doi.org/10.2166/wst.2014.331.
- Li H., Kalnay E., Miyoshi T., 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. Quarterly Journal of the Royal Meteorological Society 135(639), 523–33. https://doi.org/10.1002/qj.371.
- Li, L., Zhou, H., Hendricks Franssen, H., Gómez-Hernández, J.J., 2012. Groundwater flow inverse modeling in non-multigaussian media: performance assessment of the normal-score ensemble Kalman filter. Hydrology and Earth System Sciences 16(2), 573. https://doi.org/10.5194/hess-16-573-2012.
- Li, L., Puzel, R, Davis, A., 2018a. Data assimilation in groundwater modelling: ensemble Kalman filter versus ensemble smoothers. Hydrological Processes 32, 2020– 2029. https://doi.org/10.1002/hyp.13127.
- Li, L., Stetler, L., Cao, Z., Davis, A., 2018b. An iterative normal-score ensemble smoother for dealing with non-Gaussianity in data assimilation. Journal of Hydrology 567, 759-766. https://doi.org/10.1016/j.jhydrol.2018.01.038.
- Liang, X., Zheng, X., Zhang, S., Wu, G., Dai, Y., Li, Y., 2011. Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. Quarterly Journal of the Royal Meteorological Society 138(662), 263–73. https://doi.org/10.1002/qj.912.

- Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P. R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources 28(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models Part I - a discussion of principles. Journal of Hydrology 10(3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6.
- Reichle, R.H., McLaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. Monthly Weather Review 130, 103–114. https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.
- Saghafian, B., Jannaty, M., Ezami, N., 2015. Inverse hydrograph routing optimization model based on the kinematic wave approach. Engineering Optimization 47, 1031–1042. https://doi.org/10.1080/0305215X.2014.941289.
- Sakov, P., Oliver, D.S., Bertino, L., 2012. An Iterative EnKF for Strongly Nonlinear Systems. Monthly Weather Review 140 (6), 1988–2004. https://doi.org/10.1175/mwrd-11-00176.1.
- Szymkiewicz, R., 1993. Solution of the inverse problem for the Saint Venant equations. Journal of Hydrology 147, 105–120. https://doi.org/10.1016/0022-1694(93)90077-M.
- van Leeuwen, P.J., Evensen, G., 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. Monthly Weather Review 124, 2898–2913. https://doi.org/10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.
- Xu, T., Gómez-Hernández, J.J., Zhou, H., Li, L., 2013. The power of transient piezometric head data in inverse modeling: an application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field. Advances in Water Resources 54, 100-118. https://doi.org/10.1016/j.advwatres.2013.01.006.
- Xu, T., Gómez-Hernández, J.J., 2016. Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resources Research 52 (8), 6587-6595. http://doi.org/10.1002/2016WR019111.
- Xu, T., Gómez-Hernández, J.J., 2018. Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter. Advances in Water Resources 112, 106-123. https://doi.org/10.1016/j.advwatres.2017.12.011.
- Xue, L., Zhang, D., 2014. A multimodel data assimilation framework via the ensemble Kalman filter. Water Resources Research 50, 4197–4219. https://doi.org/10.1002/2013WR014525.

- Wang, X., Bishop, C., 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. Journal of the Atmospheric Sciences 60(9), 1140–58. https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.
- Zhou, H., Gómez-Hernández, J.J., Hendricks Franssen, H.-J., Li, L., 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Advances in Water Resources 34 (7), 844–864. https://doi.org/10.1016/j.advwatres.2011.04.014.
- Zheng, X., 2009. An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. Advances in Water Resources 26(1), 154–60. https://doi.org/10.1007/s00376-009-0154-5.
- Zoppou, C., 1999. Reverse routing of flood hydrographs using level pool routing. Journal of Hydrologic Engineering 4 (2), 184–188. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(184).
- Zucco, G., Tayfur, G., Moramarco, T., 2015. Reverse flood routing in natural channels using genetic algorithm. Water Resources Management 29, 4241–4267. https://doi.org/10.1007/s11269-015-1058-z.