Advances in Water Resources 34 (2011) 844–864

Contents lists available at ScienceDirect

Advances in Water Resources

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

An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering

Haiyan Zhou [a,b,*], J. Jaime Gómez-Hernández [b], Harrie-Jan Hendricks Franssen [c], Liangping Li [b]

[a] School of Water Resources and Environment, China University of Geosciences, 29 Xueyuan Lu, 100083 Beijing, China
[b] Group of Hydrogeology, Department of Hydraulics and Environmental Engineering, Universitat Politècnica de València, 46022 Valencia, Spain
[c] Agrosphere, IBG-3, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

ARTICLE INFO

Article history:
Received 29 March 2011
Accepted 1 April 2011
Available online 23 April 2011

Keywords:
Non-Gaussian
Ensemble Kalman filter
Parameter identification
Data assimilation
Uncertainty
Groundwater modeling

ABSTRACT

The ensemble Kalman filter (EnKF) is a commonly used real-time data assimilation algorithm in various disciplines. Here, the EnKF is applied, in a hydrogeological context, to condition log-conductivity realizations on log-conductivity and transient piezometric head data. In this case, the state vector is made up of log-conductivities and piezometric heads over a discretized aquifer domain, the forecast model is a groundwater flow numerical model, and the transient piezometric head data are sequentially assimilated to update the state vector. It is well known that all Kalman filters perform optimally for linear forecast models and a multiGaussian-distributed state vector. Of the different Kalman filters, the EnKF provides a robust solution to address non–linearities; however, it does not handle well non-Gaussian state-vector distributions. In the standard EnKF, as time passes and more state observations are assimilated, the distributions become closer to Gaussian, even if the initial ones are clearly non-Gaussian. A new method is proposed that transforms the original state vector into a new vector that is univariate Gaussian at all times. Back transforming the vector after the filtering ensures that the initial non-Gaussian univariate distributions of the state-vector components are preserved throughout. The proposed method is based in normal-score transforming each variable for all locations and all time steps. This new method, termed the normal-score ensemble Kalman filter (NS-EnKF), is demonstrated in a synthetic bimodal aquifer resembling a fluvial deposit, and it is compared to the standard EnKF. The proposed method performs better than the standard EnKF in all aspects analyzed (log-conductivity characterization and flow and transport predictions).

1. Introduction

Groundwater modeling and prediction plays a critical role in decision making for groundwater management and environmental protection. In order to make reliable groundwater flow model predictions, it is necessary to account for all measured data. Although important information can be obtained from field work, it is in practice impossible to characterize an aquifer exhaustively. To best account for the state information—such as flows, hydraulic heads or concentrations—in the characterization of aquifer parameters, numerous methods of parameter identification have been proposed (e.g., for an overview see [1–6]).

In the early days of parameter identification, the aim was to obtain a single "best" estimate of the aquifer parameters. However, it has been proven that such a "best" estimate is always much smoother than the real aquifer and that flow and transport predictions performed in such estimate are very poor [7]. The alternative is to resort to Monte Carlo analysis in which multiple realizations of the aquifer parameters are built accounting for all the measured data. Each realization represents a possible case of the unknown reality. Flow and transport are modeled in each realization and all model predictions are collected to characterize uncertainty and to define an optimal prediction (e.g., pilot point method [8,9], self-calibration [10] or gradual deformation [11]).

One such Monte Carlo-based method is the ensemble Kalman filter (EnKF) proposed by Evensen [12] and subsequently clarified by Burgers et al. [13]. The EnKF is an extension of the Kalman filter to deal with nonlinear state equations. The Kalman filter [14] proved to be a very powerful data assimilation algorithm for systems in which the relation between parameters and state is linear. This linearity allowed an exact propagation of the state covariance in time. The first attempt to deal with nonlinear transfer functions, such is the case in groundwater modeling, was the extended Kalman filter, e.g., [15,16]. In the extended Kalman filter, the

* Corresponding author. Address: Group of Hydrogeology, Department of Hydraulics and Environmental Engineering, Universitat Politècnica de València, 46022 Valencia, Spain.

E-mail addresses: haizh@upvnet.upv.es (H. Zhou), jaime@dihma.upv.es (J.J. Gómez-Hernández), h.hendricks-franssen@fz-juelich.de (H.-J. Hendricks Franssen), liali@upvnet.upv.es (L. Li).

0309-1708/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2011.04.014

nonlinear transfer function is linearized after a Taylor expansion and this linearization is used for the covariance propagation. For highly nonlinear transfer functions, extended Kalman filter tends to deteriorate as time progresses, since the errors in the covariance propagation accumulate; at the same time, extended Kalman filter is time consuming when the aquifer is finely discretized [17].

The EnKF circumvents the problem of covariance propagation in time by working with an ensemble of realizations. In each realization the state equation is solved, and the ensemble of states is used to compute the covariance explicitly and efficiently. The EnKF has gained popularity in diverse disciplines such as oceanography, meteorology and hydrology, e.g., [18–24]. The advantages of the EnKF can be summarized as follows: first, CPU consumption is reduced mostly because of the way the state covariance is computed (Hendricks Franssen and Kinzelbach [25] documented a reduction of the needed CPU time by a factor of 80 as compared with other Monte-Carlo type inverse modeling); second, the EnKF can easily be combined with different forecast models; third, the EnKF is capable of incorporating the observations sequentially in time without the need to store all previous states. In addition, although Evensen [12] developed the EnKF to obtain a single optimal estimate of the system state, the EnKF provides, as a by-product, the entire ensemble of states, which can be used to assess uncertainty.

Our objective with this paper is to use the EnKF for the generation of an ensemble of hydraulic conductivity realizations which are conditional to hydraulic conductivity measurements and to piezometric head data. This approach has already been described in the literature, for instance by Hendricks Franssen and Kinzelbach [26], who generated realizations of transmissivity and recharge, and by Liu et al. [27], who focused on hydraulic conductivity, dispersivity, mass transfer rate and mobile porosity ratio for a dual-domain mass transfer model at the MADE site. Example applications in petroleum engineering can be found in the works by Wen and Chen [28] and Gu and Oliver [29]. Our contribution is to develop a new approach applicable to non-Gaussian distributions of hydraulic conductivities, with an application to a bimodal distribution typical, for instance, of fluvial deposits.

It can be shown that the EnKF provides an optimal solution when the parameter vector follows a multiGaussian distribution and the state transfer function is linear [30]. In most practical applications in groundwater and petroleum engineering neither the parameters can be modeled as multiGaussian nor the transfer function is linear. The importance of accounting for the nonGaussianity of hydraulic conductivity has already been demonstrated in the literature [31–33]. To circumvent the problems of the EnKF, some authors have concentrated in the problem of non-Gaussianity in the parameters and others have focused on reparameterizing the state equation so that the relationship between model parameters and state variables is closer to linear. Sun et al. [34] worked on the non-Gaussianity aspect and took advantage of localization techniques with a Gaussian mixture model to update the parameters of a multimodal distribution. Chen et al. [35] addressed the nonlinearity problem by reparameterization.

In this work we take the route of transforming parameters and state variables so that they both follow marginal Gaussian distributions. These transformations, which are themselves highly nonlinear, will make the state transfer function even more nonlinear than for the untransformed variables, but, in return, the Kalman filtering equations will be applied on Gaussian variates. We demonstrate this approach on a synthetic aquifer resembling a highly channelized fluvial deposit. It will be shown how this approach improves the results obtained using a standard implementation of the EnKF even though the univariate transformation applied ensures marginal Gaussian distributions but does not ensure multiGaussianity of the joint state vector.

We are aware that in another study [36] similar transformation techniques as proposed in this paper were applied but only to the state variables, not to the parameters, in the context of hydraulic tomography. The authors focus on the improvements that can be achieved with different transformation techniques, analyze under which conditions improvements can be obtained, and demonstrate a pseudo-linearizing effect of their transformations. For reference, they compared their improved method to an exhaustive particle filter. Similar applications can be found in other disciplines; for instance, in reservoir modeling, transformation from non-Gaussian distribution to Gaussian distribution is applied to the state variables, such as saturation, by Gu and Oliver [37] and, in ocean ecosystem modeling, a similar transformation is applied to chlorophyll-*a* concentration by Simon and Bertino [38]. Other transformation algorithms can be found in the literature such as in Béal et al. [39] and Bocquet et al. [40]. In contrast with all of these applications, the method proposed in this paper focuses on transforming not only the non-Gaussian distributed state variables but, most importantly, the non-Gaussian distributed model parameters, i.e., the hydraulic conductivities, which are commonly assumed to follow a log-normal distribution. These transformation of parameters and state variables will make the relationship between the transformed variables even more nonlinear (instead of more linear as pursued in the previously cited works), but the Kalman filtering equations will be applied to Gaussian variables. To our knowledge, no such transformation algorithm has been applied to handle the non-Gaussian distribution of hydraulic conductivities in the scope of the EnKF in hydrogeology.

Throughout the paper we use interchangeably the terminology from geostatistics, hydrogeology and Kalman filtering, most noticeably, (a) piezometric head data assimilation is equivalent to (inverse) data conditioning, in the sense that the solution of the flow equation in each of the final realizations of hydraulic conductivity will match the measured piezometric heads, (b) by forecast model, transfer function or state transition model we refer to the transient groundwater flow equation and its corresponding numerical model, and (c) when referring to the flow equation we distinguish between parameters (i.e., hydraulic conductivities) and state variables (i.e., piezometric heads), whereas in the EnKF we will talk about a (joint) state vector that includes realizations of both parameters and state variables at time t. Also, since our goal is the characterization of the hydraulic conductivity spatial variability, we will be using the term hard data to refer to local measurements of hydraulic conductivity, as opposed to measurements of piezometric heads which are soft data since they do not measure directly hydraulic conductivity but serve to characterize its spatial variability.

The rest of the paper is structured as follows. After the standard EnKF is introduced, the new algorithm is explained in detail. Then a synthetic bimodal aquifer is used to evaluate the performance of the proposed method. The paper ends with a discussion and some conclusions.

2. Methodology

We first present the groundwater flow and transport equations that will be used in the synthetic example, then we follow with the description of the standard EnKF and propose the new algorithm with the transformed variables, which will be referred to as the normal-score EnKF (NS-EnKF). The flow model will be used in conjunction with ensemble Kalman filtering to obtain realizations of hydraulic conductivity conditioned to both conductivity and piezometric head data. The transport model will be used only for verification purposes to evaluate how well transport is predicted in the final conductivity fields.