Advances in Water Resources 86 (2015) 297–310

Contents lists available at ScienceDirect

Advances in Water Resources

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

Two-point or multiple-point statistics? A comparison between the ensemble Kalman filtering and the ensemble pattern matching inverse methods

Liangping Li[a,b,*], Sanjay Srinivasan[a], Haiyan Zhou[a], J. Jaime Gomez-Hernandez[c]

[a] Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin, Austin, TX 78712, USA
[b] Department of Geology and Geological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA
[c] Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022 Valencia, Spain

ARTICLE INFO

Article history:
Available online 21 May 2015

Keywords:
Multiple-point statistics
Rejection sampling
Conditional simulation
Ensemble Kalman filter
Inverse method

ABSTRACT

The Ensemble Kalman Filter (EnKF) has been commonly used to assimilate real time dynamic data into geologic models over the past decade. Despite its various advantages such as computational efficiency and its capability to handle multiple sources of uncertainty, the EnKF may not be used to reliably update models that are characterized by curvilinear geometries such as fluvial deposits where the permeable channels play a crucial role in the prediction of solute transport. It is well-known that the EnKF performs optimally for updating multi-Gaussian distributed fields, basically because it uses two-point statistics (i.e., covariances) to represent the relationship between the model parameters and between the model parameters and the observed response, and this is the only statistic necessary to fully characterize a multiGaussian distribution. The Ensemble PATtern matching (EnPAT) is an alternative ensemble based method that shows significant potential to condition complex geology such as channelized aquifers to dynamic data. The EnPAT is an evolution of the EnKF, replacing, in the analysis step, two-point statistics with multiple-point statistics. The advantages of EnPAT reside in its capability to honor the complex spatial connectivity of geologic structures as well as the measured static and dynamic data. In this work, the performance of the classical EnKF and the EnPAT are compared for modeling a synthetic channelized aquifer. The results reveal that the EnPAT yields a better prediction of transport characteristics than the EnKF because it characterizes the conductivity heterogeneity better. Issues such as uncertainty of multiple variables and the effect of measurement errors on EnPAT results will be discussed.

1. Introduction

Inverse methods have been used extensively in hydrology and petroleum engineering to identify spatial variations of geological parameters conditioned to observed dynamic data such as piezometric head and concentration, in order to improve flow and transport predictions. Inverse methods have evolved from manual trial-and-error approaches to real-time automatic data assimilation approaches; from deterministic estimation to stochastic simulation; from gradient-based minimization approaches to sampling-based approaches; and from multiGaussian-based methods to those without restrictive multiGaussian assumptions. An extensive description of the evolution of inverse methods and recent trends can be found in the work by Zhou et al. [53].

The widely used ensemble Kalman filter (EnKF), an ensemble-based real-time data assimilation inverse method, was first proposed by Evensen [8] as an extension of the extended Kalman filter. In the EnKF, the cross-correlations of the parameters and the state variables are explicitly calculated through an ensemble of realizations rather than approximated through a Taylor series expansion of the transfer function [e.g., 24]. The ensemble Kalman filter has increasingly been used in multiple disciplines such as petroleum engineering and hydrogeology because of its computational efficiency and its real-time data assimilation capability [e.g., 2,4,6,10,17,20,28,34,35–37,47]. For instance, Chen and Zhang [6] applied standard EnKF to a groundwater system in order to evaluate the sensitivity of inverted/updated parameters to factors such as ensemble size and frequency of conditioning data. Hendricks Franssen and Kinzelbach [20] applied EnKF to a field case study and discussed the filter inbreeding issue in detail. Panzeri et al. [35] coupled EnKF with moment equations to circumvent the computational cost needed in the Monte Carlo simulations, and applied this novel approach in a real case study [37]. Gharamti

* Corresponding author. Tel.: +1 512 8158787.
E-mail address: liangpingli@utexas.edu (L. Li).

http://dx.doi.org/10.1016/j.advwatres.2015.05.014

et al. [12] proposed a hybrid formulation of the EnKF and optimal interpolation that integrates both the ensemble sample covariance and a static background covariance in order to reduce the size of the ensemble and to avoid filter divergence. Panzeri et al. [34] developed a two-step updating scheme to integrate dynamic data into reservoir models exhibiting complex geology; specifically, the geometry of facies is first handled using a Markov mesh model, and then the EnKF is applied to calibrate the conductivities within each facies.

In fluvial depositions and fractured systems, hydraulic conductivity is commonly assumed to follow a multi-modal probability distribution. In other words, spatial variation in conductivity can be thought of as an outcome from several random processes that characterize the geo-material (i.e., facies). Winter et al. [50] and Winter et al. [49] discussed this type of “composite medium” in detail, and, for example, Guadagnini et al. [18] and Riva et al. [39] have applied this concept for aquifer modeling. Multiple-point statistics (MPS) methods are becoming popular for characterizing fluvial depositions and fractured aquifers. MPS uses (only) the observed static data (i.e., measured conductivity data) for conditioning geologic models. Compared to traditional two-point covariance-based geostatistical methods, MPS has the capability to effectively reproduce complex structures observed in a conceptual model (i.e., training image). Several MPS algorithms have been described in the literature since SNESIM, the first MPS code, was developed by Strebelle [44]. Hu and Chugunova [21] presented a comprehensive review of MPS methods.

The challenge is to integrate dynamic data into the MPS-based geological modeling procedures. More specifically, the key question is how to condition non-multiGaussian fields to dynamic data that are related to the model parameters through highly non-linear relationships. Recently, a number of authors have tried to apply the EnKF to an ensemble of MPS-based non-multi-Gaussian conductivity fields where the uncertainty is mainly due to the spatial distribution of geologic facies. However, the fact that the analysis equations in EnKF are equivalent to the normal equations (or cokriging equations) implies that the EnKF is optimal for multi-Gaussian fields and linear state equations [1]. In other words, using only two-point covariances and cross-covariances between parameters and state variables in the analysis step of the EnKF, the heterogeneity features that are controlled by higher-order statistics may not be preserved during the updating process. For this reason, a number of variations to the EnKF-based methods have been proposed to ensure that the connectivity prescribed by MPS simulations is preserved. For instance, Jafarpour and Khodabakhshi [23] introduced a probability conditioning method, in which a probability field (i.e., the ensemble mean of the indicator values of conductivity) is first derived by assimilating the dynamic data, and then the MPS conductivity realizations are regenerated using the calculated probability field as soft data. Sun et al. [45] and Dovera and Della Rossa [7] proposed to couple mixture Gaussian models and the EnKF to preserve the spatial structure of MPS conductivity simulations. Sarma and Chen [40] introduced a kernel EnKF approach applied to MPS conductivity simulations. Zhou et al. [51] and Li et al. [29] developed a normal score EnKF (NS-EnKF) approach in which a normal-score transformation is applied to both the non-Gaussian parameters and state variables prior to the analysis step. Hu et al. [22] proposed to update the uniform random numbers that are used to draw the conductivity values from local conditional probability in the context of a sequential MPS simulation (as implemented in SNESIM, for example). Ping and Zhang [38] presented a vector-based level-set parameterization approach for channelized aquifers, and then combined it with the EnKF to match the observed dynamic data. All of the above mentioned EnKF-based methods accomplish the goal of reproducing non-Gaussian reservoir models to varying degrees of success, but they may still result in suboptimal solutions because the analysis step is still based on two-point covariances and cross-covariances.

Unlike the previous variants of the EnKF, Zhou et al. [52] proposed a fully non-Gaussian stochastic inverse method, termed the Ensemble PATtern matching method (EnPAT), which is an evolution of the EnKF to deal with the issue of reproduction of spatial patterns prescribed by MPS simulations. In EnPAT, the correlation between model parameter and state variables is delineated by MPS (i.e., pattern) rather than by traditional two-point covariances, and thus curvilinear heterogeneities can be preserved while the dynamic data are integrated. Li et al. [26] further extended this method to simultaneously estimate parameter and state variables so that a better characterization at multiple scales is achieved. To improve the computational efficiency, Li et al. [25] coupled the EnPAT algorithm with a pilot-point scheme such as in the implementation of the self-calibration inverse method [15,46].

In this work, we highlight the capabilities of the EnPAT method to assimilate dynamic data by comparing its performance to the standard EnKF. First, the EnPAT is extended to handle continuous conductivity fields. Then, the performance of EnKF and EnPAT is compared on a synthetic aquifer example that is characterized by curvilinear channels with high permeability. Also, in order to explore the space of posterior uncertainty, Bayes’ rejection sampling method is applied in a benchmark case. The performance of EnPAT is evaluated in terms of aquifer characterization, and flow and transport predictions. Finally, we discuss the advantages and drawbacks of the EnPAT method.

The paper will continue as follows: the EnKF and EnPAT algorithms are described in Section 2; in Section 3, a synthetic example is analyzed using both the EnKF and the EnPAT methods. There is a discussion of the main results in Section 4; the paper ends with a summary.

2. The EnKF and EnPAT algorithms

2.1. General framework

The main procedure of both algorithms includes two steps: forecast and analysis. The difference between the EnKF and the EnPAT resides in the analysis step, the EnKF is based on two-point covariances and the EnPAT on multiple-point statistics. The specifics are as follows:

1. Initialization step
   Generate a set of initial models conditioned to the measured static data. In complex geological formations such as fluvial deposits, an MPS simulation method is commonly employed, which uses a conceptual model represented by a training image. Examples of MPS algorithms are the single normal equation simulation (SNESIM) [44] and the direct sampling method (DS) [31]. Note that the initial conductivity realizations are the same for both the comparison of the EnKF and EnPAT methods in this paper.
2. Forecast step
   For each conductivity realization $\mathbf{X}$, the groundwater flow equation is solved from time $t = 0$ to $t = k$, i.e.,
   $$\mathbf{Y}_k = f(\mathbf{X}_{k-1}) \tag{1}$$
   where f represents the groundwater flow model, boundary conditions as well as sources and sinks. $\mathbf{Y}_k$ denotes the simulated piezometric head at time $t = k$. The conductivity $\mathbf{X}_{k-1}$ and corresponding head $\mathbf{Y}_k$ will be used in the analysis step to derive an updated conductivity $\mathbf{X}_k$.
3. Analysis step
   Given the mismatch between the observed state $\mathbf{Y}_k^{obs}$ and the forecasted state values, the ensemble of conductivity $\mathbf{X}$ is updated from time $t = k - 1$ to time $t = k$. Specific analysis schemes for each method will be discussed in subsequent subsections for the EnKF and EnPAT.
4. Loop back to step 2 for the next time step. The forecast and analysis loop starts again with the updated conductivity $\mathbf{X}_k$ as the new