Advances in Water Resources 86 (2015) 311–326

Contents lists available at ScienceDirect

Advances in Water Resources

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

Inverse sequential simulation: Performance and implementation details

Teng Xu*, J. Jaime Gómez-Hernández

Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022 Valencia, Spain

ARTICLE INFO

Article history:
Available online 18 June 2015

Keywords:
Inverse modeling
Normal-score transform
Non-Gaussianity
Simple co-kriging
Data assimilation
Non-stationary covariance

ABSTRACT

For good groundwater flow and solute transport numerical modeling, it is important to characterize the formation properties. In this paper, we analyze the performance and important implementation details of a new approach for stochastic inverse modeling called inverse sequential simulation (iSS). This approach is capable of characterizing conductivity fields with heterogeneity patterns difficult to capture by standard multiGaussian-based inverse approaches. The method is based on the multivariate sequential simulation principle, but the covariances and cross-covariances used to compute the local conditional probability distributions are computed by simple co-kriging which are derived from an ensemble of conductivity and piezometric head fields, in a similar manner as the experimental covariances are computed in an ensemble Kalman filtering. A sensitivity analysis is performed on a synthetic aquifer regarding the number of members of the ensemble of realizations, the number of conditioning data, the number of piezometers at which piezometric heads are observed, and the number of nodes retained within the search neighborhood at the moment of computing the local conditional probabilities. The results show the importance of having a sufficiently large number of all of the mentioned parameters for the algorithm to characterize properly hydraulic conductivity fields with clear non-multiGaussian features.

1. Introduction

In groundwater flow and mass transport the characterization of the formation properties is important if we wish to predict the state of the aquifer, i.e., the spatiotemporal distributions of piezometric heads and solute concentrations. This characterization is generally made on the basis of a few direct (hard) measurements of the parameters that control the aquifer state, such as hydraulic conductivities and porosities, some indirect (soft) measurements, such as those derived from geophysical surveys, and a few observation of the state of the aquifer, such as piezometric heads.

Incorporating the state observations to characterize the parameters of the system is the object of inverse modeling, and it is a powerful but difficult task that has been the subject of extensive research during the last decades. See the work by Zhou et al. [1] for a recent review on inverse methods in hydrogeology.

Many inverse modeling approaches have been developed. Just to name a few, there are the gradual deformation method, the sequential self-calibration, variants of the Markov chain Monte Carlo method, the representer method, the pilot points method, etc. [2–13].

Besides the above mentioned methods, the ensemble Kalman filter (EnKF) [14] is the method that has attracted most attention recently. Although not an inverse method by conception, the inclusion of the parameters governing the state equation as a part of an extended state vector has turned the EnKF in the most favored inverse method nowadays [15,16]. The EnKF is well known for its flexibility to be applied to virtually any inverse problem, for its simple implementation and usage, and for its efficiency in producing realizations of the geological parameters that are consistent with the observed state variable data.

None of the above mentioned methods has been able to address the problem of characterizing conductivity fields with clear non-multiGaussian features, including the EnKF [17,18]. The main reason why the EnKF fails for non-multiGaussian fields is that it is optimal only for multiGaussian variates and linear state-transfer functions [19]. For this reason, nowadays, the main focus of inverse modeling, at least in the fields of hydrogeology and petroleum engineering, is on how to reproduce non-multiGaussian patterns.

Outside of the inverse modeling realm, the development of the single normal-equation simulation [20,21] has solved the problem of incorporating hard and soft data for the characterization of spatial patterns using statistics higher than order two, and thus, the inclusion of the spatial features that cannot be characterized simply by a covariance function. The algorithms that are capable to account for statistics higher-than-order-two are broadly termed as multiple-point statistics (MPS). They rely on the existence of a training image

* Corresponding author. Tel.: +34 963879615; fax: +34 963879492.
E-mail addresses: tenxu@posgrado.upv.es (T. Xu), jaime@dihma.upv.es (J.J. Gómez-Hernández).

http://dx.doi.org/10.1016/j.advwatres.2015.04.015
0309-1708/

exhibiting the types of patterns to be present in the final models, from which to infer the higher-order statistics. Some available algorithms and codes are SNESIM [22], FILTERSIM [23], SIMPAT [24], DS [25], EnPAT [26], and others [27–31]. But these algorithms were not devised for inverse modeling; therefore, they cannot incorporate state variable information.

There have been some attempts to combine MPS and inverse modeling: Hu et al. [32] used, with moderate success, the realizations of uncorrelated random numbers needed for the drawing of the conductivity value from the local distribution function on the sequential simulation implementation of the single normal equations as the state variables to be updated during the analysis st ep of the EnKF; Zhou et al. [27], Li et al. [26] developed a new MPS algorithm (termed EnPAT) which blends direct simulation [25] and the EnKF to generate inverse conditional realizations of conductivity in channelized bimodal aquifers; EnPAT works well, but it is still very CPU-time consuming.

Some authors are against the use of MPS arguing that MPS is too dependent on the choice of the parameters controlling the algorithm [33]. Other approaches to address the issue of non-multiGaussianity in inverse modeling include the works by Sun et al. [17], who combined the EnKF with a Gaussian mixture model, or by Liu and Oliver [34], Gu and Oliver[35,36] who used an iterative EnKF; plus a set of works who combine the normal-score transform (sometimes referred as anamorphosis) and the EnKF [18,37–44]. None of these methods can be considered as the definite solution of inverse modeling for conductivity fields that display non-multiGaussian features.

In this paper we describe a new method for inverse stochastic modeling applicable for non-multiGaussian fields. We have called this method inverse sequential simulation, and it is inspired on the standard multivariate sequential simulation algorithm [45,46] with normal-score transforms [47] and the Monte Carlo concept of the EnKF. The paper describes the algorithm and its implementation, and then performs a sensitivity analysis of the key parameters controlling the algorithm; the paper ends with a post-audit of the generated ensemble of realizations to check how they would perform in a solute transport prediction exercise. The algorithm has been benchmarked against the normal-score ensemble Kalman filter, with excellent results, in the paper by Xu and Gómez-Hernández [48]; therefore, this paper will not focus on a comparison with other methods, but on the implementation and performance of the algorithm.

2. Methodology

The key idea of inverse sequential simulation (iSS) is to use multivariate multi-Gaussian sequential simulation [45] to generate realizations of normal scores of conductivity, conditioned on conductivity and observed head data. The main difference with standard sequential simulation is that the method does not use an analytical, stationary model for the auto- and cross-covariances, but rather, as in the EnKF, non-stationary auto- and cross-covariances are derived from an ensemble of conductivity realizations and their associated piezometric heads (obtained by solving a groundwater flow model).

Before describing the whole algorithm, recall the main steps in any sequential simulation algorithm:

1. Define a random path to visit all nodes of the grid on which the realization will be generated.
2. Visit the random path sequentially.
   (a) At each node, collect the conditioning data for all variables (in our case, we will have two variables: conductivity and piezometric heads) within a user-defined search neighborhood centered at the point to simulate (the size and orientation of the search neighborhood, and the number of data of each variable to keep within it are parameters that must be specified by the user).
   (b) Compute the local conditional distribution function. If we adopt a multivariate multiGaussian random function model, the local conditional distribution is a Gaussian distribution with mean and variance given by the simple co-kriging estimate and the simple co-kriging variance.
   (c) Draw, randomly, a value from the local conditional distribution function.
   (d) Include the simulated value in the set of conditioning data for the simulation of the next nodes and move to the next node.

Our proposal is to use this algorithm to generate conductivity fields conditioned to piezometric heads. For this purpose we need the auto-covariances of both conductivity and head, and their cross-covariance. These covariances, particularly the ones involving the piezometric heads, but also the conductivity auto-covariance when there are conditioning conductivity data, are clearly non-stationary. Some authors have developed analytical expressions relating these covariances by approximating the solution of the groundwater flow equation [49]. We propose to use experimentally-derived covariances obtained from an ensemble of realizations, much like it is done in the EnKF.

At any time t, we could derive all necessary covariances experimentally as follows:

1. Generate an ensemble of N_e realizations of conductivity. Each realization contains N nodes. $K_i(j)$ refers to the conductivity value at realization i and node j.
2. Given initial and boundary conditions, sources and sinks, solve the groundwater flow equation Eq. (1) [50] for each realization until time t and obtain an ensemble of piezometric heads.

$$S_s \frac{\partial H}{\partial t} - \nabla \cdot (K \nabla H) = W, \tag{1}$$

   where S_s is specific storage coefficient $[L^{-1}]$, H is the hydraulic head $[L]$, K is the hydraulic conductivity $[LT^{-1}]$, W denotes sources and sinks per unit volume $[T^{-1}]$; t is the time $[T]$, $\nabla \cdot$ is the divergence operator, and ∇ is the gradient operator.
3. The cross-covariance between conductivity K at location j and piezometric head H at location l is given by

$$C_{K,H}(j,l) = \frac{1}{N_e} \sum_{i=1}^{N_e} (K_i(j) - \overline{K}(j))(H_i(l) - \overline{H}(l)), \tag{2}$$

   where the overbar indicates ensemble average, i.e.,

$$\overline{K}(j)) = \frac{1}{N_e} \sum_{i=1}^{N_e} K_i(j). \tag{3}$$

The auto-covariances for heads and conductivities are computed similarly.

In addition, since we are planning to work with multiGaussian sequential simulation, it is more convenient to work with the normal-score transform [47] of the variable of interest, in our case conductivity. Therefore, the sequential simulation algorithm is performed on a new variable $\widetilde{K}$ which is obtained by the normal-score transform of K according to the following expression:

$$\widetilde{K}_i(j) = G^{-1}(F_j(K_i(j))) \tag{4}$$

where $F_j(K_i(j))$ is the local cumulative distribution at node j computed (numerically) from the N_e conductivity values of all realizations at node j, and $G(\cdot)$ is the standard Gaussian cumulative distribution function. Auto-covariances and cross-covariances are computed, as described above, for the normal-score transformed conductivities, not for the untransformed ones as in the description. These covariances will be different from the ones corresponding to the untransformed conductivity.