Environmental Modelling & Software 70 (2015) 55–64

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A local–global pattern matching method for subsurface stochastic inverse modeling

Liangping Li [a, *], Sanjay Srinivasan [a], Haiyan Zhou [a], J. Jaime Gómez-Hernández [b]

[a] Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin, 78712, Austin, USA
[b] Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022, Valencia, Spain

ARTICLE INFO

Article history:
Received 21 January 2015
Received in revised form
8 April 2015
Accepted 10 April 2015
Available online 15 May 2015

Keywords:
Multiple-point geostatistics
Conditional simulation
Inverse modeling
Global matching
Uncertainty assessment

ABSTRACT

Inverse modeling is an essential step for reliable modeling of subsurface flow and transport, which is important for groundwater resource management and aquifer remediation. Multiple-point statistics (MPS) based reservoir modeling algorithms, beyond traditional two-point statistics-based methods, offer an alternative to simulate complex geological features and patterns, conditioning to observed conductivity data. Parameter estimation, within the framework of MPS, for the characterization of conductivity fields using measured dynamic data such as piezometric head data, remains one of the most challenging tasks in geologic modeling. We propose a new local–global pattern matching method to integrate dynamic data into geological models. The local pattern is composed of conductivity and head values that are sampled from joint training images comprising of geological models and the corresponding simulated piezometric heads. Subsequently, a global constraint is enforced on the simulated geologic models in order to match the measured head data. The method is sequential in time, and as new piezometric head become available, the training images are updated for the purpose of reducing the computational cost of pattern matching. As a result, the final suite of models preserve the geologic features as well as match the dynamic data. This local–global pattern matching method is demonstrated for simulating a two-dimensional, bimodally-distributed heterogeneous conductivity field. The results indicate that the characterization of conductivity as well as flow and transport predictions are improved when the piezometric head data are integrated into the geological modeling.

1. Introduction

Inverse modeling is a mathematical approach to identify parameters such as permeability or hydraulic conductivity at unsampled locations such that flow and transport modeling using the estimated parameters match observed state variables such as piezometric head or concentration data. Predictions for groundwater flow and solute transport made using the estimated parameters would then be more accurate. The fact that the number of observed state variables is much smaller than the number of unknown parameters implies that the solution of inverse problem will be non-unique (Carrera and Neuman, 1986) especially when heterogeneous subsurface systems are considered. In order to represent this non-uniqueness, stochastic inverse modeling seeks to generate multiple likely representations of parameter fields that are all conditioned to both direct measurements of the parameters at specific locations and dynamic data (Gómez-Hernández et al., 1997). The multiple calibrated models obtained by applying stochastic inversion methods could be used to assess the uncertainty in predictions based on the available data. Reliable models for uncertainty are required by decision-makers. For a review of the evolution and recent trends of inverse methods in hydrogeology, the reader is referred to Zhou et al. (2014).

In cross-bedded aquifers or fluvial geologies, aquifer properties such as hydraulic conductivity exhibit connectivity along curvilinear paths. This complex connectivity significantly affects the flow and transport of fluids and chemical species (Gómez-Hernández and Wen, 1998; Renard and Allard, 2011). Reproduction of the curvilinear geometry can be achieved using Multiple-Point Statistics (MPS) based stochastic simulation methods (Strebelle, 2002). MPS simulation was developed to overcome the limitation of traditional two-point variogram-based methods, which cannot capture strong

* Corresponding author.
E-mail addresses: liangpingli@utexas.edu (L. Li), sanjay.srinivasan@engr.utexas.edu (S. Srinivasan), haiyanzhou@utexas.edu (H. Zhou), jaime@dihma.upv.es (J.J. Gómez-Hernández).

http://dx.doi.org/10.1016/j.envsoft.2015.04.008
1364-8152/

connectivities in the subsurface aquifer. The higher moments (i.e., multiple-point statistics) are introduced into the simulation by borrowing patterns from a training image (Guardiano and Srivastava, 1993). Although MPS provides an avenue to simulate complex formations, stochastic inverse modeling within the framework of MPS simulations is extremely challenging because of the difficulty in maintaining the complex curvilinear connectivity geological structures while simultaneously honoring dynamic data that are related to conductivity through a strongly non-linear transfer function.

In the literature, stochastic inverse methods can be classified into two groups. In the first group, an objective function is first constructed based on the discrepancy between observed data and simulated values. This objective function is subsequently minimized by iteratively perturbing the parameter values until a sufficiently close match is attained. Preservation of the prior geological structures is not explicitly considered during this process of optimization. Examples of this data-driven stochastic inverse method are sequential self-calibration (Gómez-Hernández et al., 1997; Hendricks Franssen et al., 2003), the pilot-point method (de Marsily, 1978) and the ensemble Kalman filter (EnKF) (Evensen, 2003). It has been proven that these methods yield optimal estimates for multiGaussian conductivity fields. Some variants were proposed to handle non-multiGaussian conductivity fields. For example, Capilla et al. (1999) proposed the application of self-calibration method to local conditional probabilities defining the uncertainty in conductivity, instead of calibrating the conductivities directly. Later, Capilla and Llopis-Albert (2009) coupled the gradual deformation method and the optimization of the probability fields in order to improve the efficiency of the previous proposal. In a similar way, Hu et al. (2013) proposed to consider the uniform random number used to draw the MPS realizations as part of the state variable set in EnKF. Sun et al. (2009) coupled Gaussian mixture models and EnKF to handle non-Gaussian conductivity fields. Jafarpour and Khodabakhshi (2011) proposed to first update the ensemble of MPS-generated conductivities to derive local probabilities, and then, to re-simulate the conductivities using the probability maps as soft data. Zhou et al. (2011) developed a normal-score EnKF to handle non-Gaussianity within the ensemble Kalman filtering framework.

In the second group of inverse modeling approaches, data integration is achieved using Bayes' theorem. The posterior models are sampled from the prior models by assessing first a likelihood function. A typical example of this model-driven stochastic inverse method is rejection sampling (Tarantola, 2005). The likelihood of a model sampled from a prior set is assessed, and the model is rejected depending on a likelihood threshold. The prior geological structures will be preserved in this process, because the posterior set of models is simply a subset of the prior set. However, like the particle filtering approach, this method is computationally expensive and is inapplicable in most practical cases because tens of thousands of models need to be evaluated. To improve the computational efficiency, Mariethoz et al. (2010a) proposed an iterative spatial resampling method in which the candidate models are generated by conditioning to data sampled from previous accepted models, thus resulting in less computational cost because of faster convergence to a posterior set that exhibits the desired dynamic characteristics. Another popular Bayesian approach to inverse modeling is the Markov chain Monte Carlo method (McMC) (Metropolis et al., 1953; Oliver et al., 1997) in which the parameter model is first locally perturbed for a gridblock or for a set of gridblocks (i.e., the transition kernel) and then the forecast model is run to judge whether the new candidate model will be accepted (e.g., the Metropolis–Hastings rule). The problems with these McMC methods are: (1) the acceptance rate of new models is dependent on the transition kernel used; (2) a long chain is usually required before the posterior distribution can be correctly sampled, and (3) a large number of perturbed models have to be generated and evaluated. An extensive description of the mathematical framework for the McMC method and recent advances can be found in the review paper by Liu et al. (2010).

The Ensemble PATtern matching (EnPAT) stochastic inverse method was first proposed by Zhou et al. (2012) with the aim to create multiple conductivity fields honoring both measured conductivity and piezometric head data as well as the prior geological structures. The EnPAT is inspired by the Direct Sampling (DS) MPS method developed by Mariethoz et al. (2010b). In DS, the conductivity patterns are directly sampled from a training image without storing the entire pattern database in memory. This results in fast simulation and the possibility to simulate continuous variables such as hydraulic conductivity. Zhou et al. (2012) borrows the concept of DS and expands the conductivity pattern to include the pattern of piezometric heads for the purpose of inverse modeling. Correspondingly, multiple MPS-simulated conductivity models and the corresponding head models obtained by running the forward simulator are jointly used as the training images for learning during the simulation. Conductivities are simulated by matching joint patterns from the training image sets. As a result, the simulated conductivity models are not only conditioned to the measured conductivity and piezometric data, but also preserve the prior geological structures. Li et al. (2013a) developed a hybrid of the EnPAT and the pilot point/self-calibration method (Gómez-Hernández et al., 1997) to reduce the computational cost and to improve the characterization of conductivity connectivity during the dynamic data assimilation process.

In this paper, we propose a local–global pattern matching method to integrate dynamic data into geologic models. In the previous implementation of the EnPAT, a local pattern is considered for ensemble matching, but that does not guarantee that the updated model matches the observed global dynamic data because of the non-linearity of the forecast function as well as the existence of complex boundary conditions. To address this issue, we implement an additional step in which we simulate the global response of the updated models and select those that best fit the observed data after the process of local pattern matching. As a consequence, updated models will preserve the geological structures and the dynamic data, although at a computational cost because of the additional forward simulations in the rejected models. In order to mitigate the computational demand and to accelerate the learning process, the training image sets are refined by progressively replacing the worst models in the prior training set with the newly accepted models. The method therefore borrows the concept of iterative resampling proposed by Mariethoz et al. (2010a). A ranking scheme is implemented to identify the poor initial models. The proposed methodology is demonstrated on a synthetic example for which predictions of flow and transport are considered.

The remainder of the paper will be organized as follows. In Section 2, the implementation of the ensemble pattern matching method is described, with emphasis on the significance of global constraints on the predictions of flow and transport. In Section 3, a synthetic example is used to demonstrate the effectiveness of the proposed method. Then, in Section 4, we discussed the computational efficiency of the EnPAT by continuously refining the training images. In Section 5, there is a general discussion. The paper ends with a summary and conclusions.

2. Methodology

In the EnPAT method two steps are performed at each time step: the forecast step (i.e., solving the flow equation based on the current hydraulic conductivities to derive the piezometric head) and updating step (i.e., updating both conductivity and head through a pattern matching approach).