Journal of Hydrology 564 (2018) 1074–1084

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter

Zi Chen[a], J. Jaime Gómez-Hernández[a], Teng Xu[a,*], Andrea Zanini[b]

[a] Institute of Water and Environmental Engineering, Universitat Politècnica de València, Valencia, Spain
[b] Dipartimento di Ingegneria e Architettura, Università degli Studi di Parma, Parma, Italy

ARTICLE INFO

Keywords:
Inverse modeling
Observation error
Groundwater laboratory experiment
Stochastic hydrogeology

ABSTRACT

Contaminant source identification is a key problem in handling groundwater pollution events. The ensemble Kalman filter (EnKF) is used for the spatiotemporal identification of a point contaminant source in a sandbox experiment, together with the identification of the position and length of a vertical plate inserted in the sandbox that modifies the geometry of the system. For the identification of the different parameters, observations in time of solute concentration are used, but not of piezometric head data since they were not available. A restart version of the EnKF is utilized because it is necessary to restart the forecast from time zero after each parameter update. The results show that the restart EnKF is capable of identifying both contaminant source information and aquifer-geometry-related parameters together with an uncertainty estimate of such identification.

1. Introduction

The problem of identifying a contaminant source in an aquifer using solute concentration data has been the subject of attention for many years (e.g., Atmadja and Bagtzoglou, 2001; Michalak and Kitanidis, 2004; Bagtzoglou and Atmadja, 2005; Sun et al., 2006 and references therein). Briefly, the proposed methods could be grouped into two categories: optimization approaches and probabilistic approaches. The main difference between the two approaches is that the optimization approaches cast the problem as a deterministic one in which parameters are found that minimize a given objective function, whereas the probabilistic approaches cast the problem in a stochastic framework and the parameters to estimate become random variables. In the first category, Gorelick et al. (1983) identified the groundwater pollution source information through an optimization model using linear programming and multiple regression; Wagner (1992) employed a non-liner maximum likelihood method to estimate source location and flux; Mahar and Datta (2000) used a nonlinear optimization model for estimating the magnitude, location and duration of groundwater pollution sources with binding equality constraints; Yeh et al. (2007) developed a hybrid approach, which combines simulated annealing, tabu search and a three-dimensional groundwater flow and solute transport model to solve the source identification problem; and Ayvaz (2010) utilized a harmony search-based simulation-optimization model to determine the source location and release histories by using an implicit solution procedure. In the second category, Bagtzoglou et al. (1992) applied a particle method to estimate, probabilistically, source location and spill-time history; Woodbury and Ulrych (1996) used a minimum relative entropy approach to recover the release and evolution histories of a groundwater contaminant plume in a one-dimensional system; Neupauer and Wilson (1999) employed a backward location model based on adjoint state method (BPM-ASM) to identify a contaminant source; Butera et al. (2013) utilized a simultaneous release function and source location identification (SRSI) method to identify the release history and source location of an injection in a groundwater aquifer; and Koch and Nowak (2016) derived and applied a Bayesian reverse-inverse methodology to infer source zone architectures and aquifer parameters.

The ensemble Kalman filter (EnKF), which could be included in the group of probabilistic approaches mentioned above, has recently addressed the problem of contaminant source identification. The EnKF introduced by Evensen (2003) has gained much popularity in recent years for its efficiency in solving inverse problems in different fields such as oceanography, meteorology and hydrology (Houtekamer and Mitchell, 2001; Li et al., 2012a; Xu et al., 2013b). The advantages of the EnKF can be summarized as follows (Chen and Zhang, 2006; Zhou et al., 2011): computational efficiency when compared with other inverse approaches, easy integration with different forecast models, ability to account for model and observation errors, and easy uncertainty characterization since the final outcome is always an ensemble of

* Corresponding author.
E-mail address: tenxu@posgrado.upv.es (T. Xu).

https://doi.org/10.1016/j.jhydrol.2018.07.073

Available online 01 August 2018
0022-1694/ © 2018 Elsevier B.V. All rights reserved.

realizations. In hydrogeology, the EnKF has been mainly applied for the identification of aquifer parameters such as hydraulic conductivity or porosity (Li et al., 2012b; Xu et al., 2013a; Zhou et al., 2014; Xu and Gómez-Hernández, 2015; Xu and Gómez-Hernández, 2016a). Recently, Xu and Gómez-Hernández (2016b) demonstrated the possibility to apply the EnKF for the identification of a contaminant source in a deterministic synthetic aquifer, and later Xu and Gómez-Hernández (2018) showed that the method can be also applied for the simultaneous identification of hydraulic conductivities and the parameters defining a contaminant source also in a synthetic aquifer. All the works mentioned above were tested in synthetic cases. Only a few works can be found in the literature for laboratory or field cases. Woodbury et al. (1998) extended the minimum relative entropy (MRE) method to recover the release history of a contaminant and applied it to reconstruct the release history of a 1,4-dioxane plume observed at the Gloucester Landfill in Ontario, Canada. Michalak (2003, 2004) employed a Bayesian inverse formulation to estimate the contaminant history of trichloroethylene (TCE) and perchloroethylene (PCE) in an aquifer at the Dover Air Force Base, Delaware, a site that had already been analyzed by Liu and Ball, 1999 in the same context of source identification. Cupola et al. (2015,) compared the source location identification (SRSI) method to the backward probability model based on the adjoint state method (BPM-ASM) with data taken from a sandbox experiment. Zanini and Woodbury (2016) also used data from a sandbox experiment to apply an empirical Bayesian method combined with Akaike's Bayesian Information Criterion (ABIC) to deduce the release history of a groundwater contaminant.

The main objective of this paper is to assess the performance of the restart EnKF (r-EnKF) for the identification of contaminant source parameters and aquifer geometry with data from a sandbox experiment. The source parameters of interest are the release location, release starting and ending times, and contaminant load, and regarding the geometry the method should try to retrieve the position and length of a plate that is inserted about the center of the sandbox and induces a deflection of the flowlines towards the bottom of the sandbox. The state information assimilated by the r-EnKF is limited to concentration data at a few observation points, since no piezometric head data were available.

The paper is organized as follows, first, the state equations and the fundamentals of the r-EnKF will be recalled, second, the sandbox characteristics are described together with the numerical model used to reproduce its behavior, third, the r-EnKF is tested with data from a synthetic experiment that mimics the sandbox experiment with the aim to verify if the r-EnKF is capable of identifying the kind of parameters sought, and four, the r-EnKF is applied with observation values taken from the sandbox experiment, the problems encountered are analyzed, alternative approaches are discussed and the final results presented. The paper ends with a summary and conclusions on the main findings.

2. Methodology

2.1. Groundwater flow and solute transport equation

The sandbox will be modeled as a two-dimensional system in the XZ plane, where an inert contaminant spreads due to advection and dispersion under a steady-state flow. The dimension of the sandbox in the y direction is small enough to assume that the state variables are constant along any line for any given (x, z) value. The governing equations are:

$$\frac{\partial}{\partial x}\left(K_x\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial z}\left(K_z\frac{\partial h}{\partial z}\right)+w=0 \tag{1}$$

$$\frac{\partial(\theta C)}{\partial t}=\nabla\cdot(\theta D\cdot\nabla C)-\nabla\cdot(\theta vC)-q_sC_s \tag{2}$$

where K_x and K_z are the principal components of the hydraulic conductivity tensor in the x and z spatial coordinates respectively $[LT^{-1}]$ which are assumed aligned with the coordinate system of reference in the entire domain; h is the hydraulic head $[L]$;w represents distributed sources or sinks $[T^{-1}]$;t is time $[T]$; θ represents the porosity of the medium; C is dissolved concentration $[ML^{-3}]$;$\nabla\cdot$ is the divergence operator; ∇ is the gradient operator; D represents the hydrodynamic dispersion coefficient tensor $[L^2T^{-1}]$;v is the flow velocity vector $[LT^{-1}]$ derived from the solution of the flow model; q_s represents volumetric flow rate per unit volume of aquifer associated with a fluid source or sink $[T^{-1}]$ and C_s is the concentration of the source or sink $[ML^{-3}]$.

The flow equation is solved using MODFLOW (McDonald and Harbaugh, 1988), and the transport equation is solved using MT3DS (Zheng and Wang, 1999).

2.2. The ensemble Kalman filter

The ensemble Kalman filter was first introduced by Evensen (2003) to circumvent the difficulty of propagating covariances in time in the original and extended Kalman filter formulations. The restart EnKF (r-EnKF) has proven its capacity for contaminant source identification in synthetic cases (Xu and Gómez-Hernández, 2016b; Xu and Gómez-Hernández, 2018); now, we propose to test the r-EnKF in a sandbox experiment. For this specific case, there will be eight parameters to identify, six related to the contaminant source, and two related to aquifer geometry. In the first group, they are the contaminant source location (Xs, Zs), the injection concentration Ic, the injection rate Ir, plus the starting Ts and ending Te release times. In the second group, the algorithm will try to identify the position along the x direction Xb and the total depth Zb of a vertical plate inserted about the center of the sandbox to deflect the flowlines. The rest of the parameters defining the flow and transport conditions in the sandbox are not subject to identification and are equal to their observed values as explained in the description of the experiment in the next section. The r-EnKF is shortly described next.

In the ensemble Kalman filter with extended state vector, we deal with two types of variables, the system parameters subject of identification, of which there could be observations or not, and the state of the system, of which there will be observations. The state is forecasted in time solving the corresponding state equations, with the latest parameter update, up to the specific time steps when observations are collected; these observations are assimilated by the filter and serve to update the parameters and the state of the system. In the restart filter, state variables are not updated, only system parameters are, because the system state forecast for the next observation time is restarted from time zero to make sure that the forecasted system state is fully coherent with the state equations, and, in our case, with the updated contaminant source. (In the original implementation of the filter, both state and parameters are updated, and the state system is forecasted from the last updated state values using the last updated parameters.) The r-EnKF is an iterative algorithm that cycles forecast and data assimilation (with the corresponding parameter update) until all observations have been accounted for. The implementation of the r-EnKF for the identification of the eight parameters described above can be summarized as follows (Evensen, 2003; Xu and Gómez-Hernández, 2016b):

1. Generate an initial ensemble of parameter values. An ensemble of N_e realizations of eight-tuples of the parameters to be identified is generated. Parameter values are drawn, independently, from uniform distributions defined between first-guess minimum and maximum values—there are no restrictions on these uniform distributions, their range can be wider or narrower than the one used in this paper, and they do not have to necessarily contain the "real" value, they are simply used to initialize the algorithm. We build N_e vectors S_i with the eight parameters for each realization:

$$S_i=[Xs_i,\ Zs_i,\ Xb_i,\ Zb_i,\ Ic_i,\ Ir_i,\ Ts_i,\ Te_i]^T \tag{3}$$

Download English Version:

https://daneshyari.com/en/article/8894533

Download Persian Version:

https://daneshyari.com/article/8894533

Daneshyari.com