Advances in Water Resources 108 (2017) 69–82

Contents lists available at ScienceDirect

Advances in Water Resources

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

Identification of transmissivity fields using a Bayesian strategy and perturbative approach

Andrea Zanini[a,*], Maria Giovanna Tanda[a], Allan D. Woodbury[b,1]

[a] Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124, Parma, Italy
[b] Department of Civil Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

ARTICLE INFO

Article history:
Received 10 April 2017
Revised 17 July 2017
Accepted 28 July 2017
Available online 29 July 2017

Keywords:
ABIC
Bayes
Groundwater
Transmissivity estimation
Inverse modeling

ABSTRACT

The paper deals with the crucial problem of the groundwater parameter estimation that is the basis for efficient modeling and reclamation activities. A hierarchical Bayesian approach is developed: it uses the Akaike's Bayesian Information Criteria in order to estimate the hyperparameters (related to the covariance model chosen) and to quantify the unknown noise variance. The transmissivity identification proceeds in two steps: the first, called empirical Bayesian interpolation, uses Y^* ($Y = \ln T$) observations to interpolate Y values on a specified grid; the second, called empirical Bayesian update, improve the previous Y estimate through the addition of hydraulic head observations. The relationship between the head and the $\ln T$ has been linearized through a perturbative solution of the flow equation. In order to test the proposed approach, synthetic aquifers from literature have been considered. The aquifers in question contain a variety of boundary conditions (both Dirichelet and Neuman type) and scales of heterogeneities ($\sigma_Y^2 = 1.0$ and $\sigma_Y^2 = 5.3$). The estimated transmissivity fields were compared to the true one. The joint use of Y^* and head measurements improves the estimation of Y considering both degrees of heterogeneity. Even if the variance of the strong transmissivity field can be considered high for the application of the perturbative approach, the results show the same order of approximation of the non-linear methods proposed in literature. The procedure allows to compute the posterior probability distribution of the target quantities and to quantify the uncertainty in the model prediction. Bayesian updating has advantages related both to the Monte-Carlo (MC) and non-MC approaches. In fact, as the MC methods, Bayesian updating allows computing the direct posterior probability distribution of the target quantities and as non-MC methods it has computational times in the order of seconds.

1. Introduction

One of the most interesting problems in groundwater hydrology is the identification of contaminant sources and their release histories. This knowledge is of particular importance to assess blame, forecast the fate of solutes in aquifers and to design efficient remediation actions. During the last two decades a substantial number of approaches aimed at identifying the contaminant source position and/or its release in time were developed (see for example the works of Woodbury and Ulrych, 1996; Butera et al., 2013; Cupola et al., 2015; Koch and Nowak, 2016; Yeh et al., 2016). All the proposed approaches are promising and for the most part efficient, but do require full knowledge of the hydraulic properties of the investigated aquifer, specifically, hydraulic conductivity, hydraulic head, porosity and the like. Unfortunately, the complete spatial distribution of these hydraulic properties is never fully known, never uniform, and is dependent on the knowledge base of an observer. The reader will likely be aware that a large number of methodologies are also aimed at strictly estimating hydraulic parameters with different degrees of reliability and applicability; see the reviews by Kitanidis (1996), McLaughlin and Townley (1996), Zimmerman et al. (1998), Hendricks Franssen et al. (2009) and Zhou et al. (2014). The challenge of estimating hydraulic parameters is still motivating the development of new approaches; see for instance the works of Zanini and Kitanidis (2009), Liu et al. (2010), Majdalani and Akerer (2011), Zhang (2014), Berg and Illman (2015), Xu and Gómez-Hernández (2015), Marinoni et al. (2016) and Riva et al. (2017). These references are by no means exhaustive and serve only to highlight the importance of the overall problem of parameter estimation.

* Corresponding author.
E-mail addresses: andrea.zanini@unipr.it (A. Zanini), mariagiovanna.tanda@unipr.it (M.G. Tanda), allan.woodbury@unimanitoba.ca (A.D. Woodbury).

[1] Visiting Professor, Department of Earth and Environmental Sciences, University of British Columbia Okanagan, Kelowna, BC, Canada.

http://dx.doi.org/10.1016/j.advwatres.2017.07.022
0309-1708/

One class of methods revolves around information measures. Woodbury and Ulrych (2000) proposed a very efficient hierarchical (full) Bayesian approach to the groundwater inverse problem for steady-state groundwater flow. These authors adopt a stochastic conceptual framework for the porous media under consideration and they transformed the non-linear inverse problem into a linear one. The approach consists in two steps: the first one (typical linear interpolation) is by Bayesian updating (Woodbury, 1989) that was used to condition prior estimates of an ln(T) field with ln(T) measurements; then in the second one hydraulic head measurements were incorporated into the updating procedure by adopting a linearized aquifer equation. This total approach requires “soft” geological information in the form of an expected value of a priory probability distribution function and on a correlation function. This approach allows to consider data from different sources (for instance transmissivity values and head levels) and to constrain the inverse problem with knowledge on the system. Kennedy and Woodbury (2002) applied Woodbury and Ulrych’s (2000) approach to estimate the transmissivity of a carbonate and a sandstone aquifer with reasonable results. Jiang et al. (2004) generalized Woodbury and Ulrych’s method implementing different boundary conditions and estimating the transmissivity field of a well-known literature case (Edwards’s aquifer). Jiang and Woodbury (2006) extended the approach updating the transmissivity field incorporating thermal (heat) observations. For the Edwards Aquifer, Painter et al. (2007) augmented the approach with first upscaling local T measurements to block scales and compared the results obtained through the Bayesian approach to standard interpolation methods.

As Riva et al. (2011) highlighted, the estimation of hyperparameters is a crucial issue. In the above-mentioned works, the unknown hyperparameters that characterize the mean, the observation error and the covariance function can be assumed known or inferred from data, (Woodbury and Ulrych, 2000; Jiang et al., 2004; Painter et al., 2007) or estimated (Woodbury and Ulrych, 2000; Kennedy and Woodbury, 2002; Jiang and Woodbury, 2006). The estimation of the hyperparameters has been carried out through a minimum relative entropy approach (Woodbury and Ulrych, 1993, 1996). The method basically assigns a prior without adding any unknown information and is, therefore, considered maximally uncommitted with respect to unknown information, see for details Woodbury and Ulrych (1993) and Woodbury and Ulrych (2000).

Our current work revisits the Bayesian approach proposed by Woodbury and Ulrych (2000) and recasts it in an empirical framework. The idea behind the empirical Bayes approach is that the prior is based on information contained in the input data (Robbins, 1964). In this framework, it is considered that the prior probabilities are not fixed as in classical Bayesian modeling. Here, information on the prior is considered available in the actual data that is measured. The transmissivity field is estimated on the basis of transmissivity and hydraulic head observations using a two phase procedure: (1) estimation using only transmissivity observations; (2) update of the estimate through adding hydraulic head observations. In both phases the hyper-parameters of the prior pdfs can be determined by a special procedure as noted below.

The objectives of this paper are:

1. to amend the Woodbury and Ulrych (2000) procedure with Akaike’s Bayesian Information Criteria (ABIC) for estimating the hyperparameters. This approach treats the overall approximation capability for the unknowns of the entire models (Ulrych et al., 2001; Shibata, 2002; Woodbury and Ferguson, 2006; Zanini and Woodbury, 2016);
2. to extend the Woodbury and Ulrych (2000) procedure to strongly heterogeneous cases;
3. to provide an assessment of the reliability of the results;
4. to test the influence on the performance of the method of different boundary conditions.
5. to evaluate the computational robustness of the proposed approach considering different degrees of heterogeneity;
6. to evaluate the improving in the estimation process as the number of observations increases.

2. Research method

2.1. Empirical Bayesian solution to linear interpolation

In this work we follow an empirical approach to obtain the solution of an inverse problem consisting in the identification of a groundwater hydraulic parameter field with available observations of transmissivity and hydraulic head. Recalling briefly the Bayesian approach for inverse problems in groundwater (more details are available in Zanini and Woodbury (2016), amongst other sources), given the vectors of observed data $\mathbf{d^*}$ and the unknown parameters $\mathbf{m}$, we consider the fundamental equation (Bayes theorem):

$$p(\mathbf{m}|\mathbf{d}^*, I) = \frac{p(\mathbf{d}^*|\mathbf{m}, I)p(\mathbf{m}|I)}{\int p(\mathbf{d}^*|\mathbf{m}, I)p(\mathbf{m}|I)d\mathbf{m}} \quad (1)$$

where $p(\mathbf{m}|I)$ is the prior probability density (pdf) of the model parameters, given the prior information $I;p(\mathbf{d}^*|\mathbf{m}, I)$ is the likelihood of observing $\mathbf{d}^*$ given the model parameters $\mathbf{m}$ and the prior information I, $p(\mathbf{m}|\mathbf{d}^*, I)$ is the posterior probability density of the vector $\mathbf{m}$ after the occurrence of $\mathbf{d^*}$ and in validity of the prior information I. The integral in the denominator of (1) is a normalizing constant, called “predictive distribution” and written also as $p(\mathbf{d^*}|I)$, that represents the pdf of observing the data $\mathbf{d^*}$ with the uncertainty in the model parameters marginalized out of consideration. In the empirical Bayes approach the denominator depends on any hyperparameters in the prior.

Consider now the linear inverse problem:

$$\mathbf{d}^* = \mathbf{Gm} + \nu \quad (2)$$

where $\mathbf{d^*}$ is an observed data vector ($N \times 1$), $\mathbf{m}$ is model vector ($M \times 1$) which contains the unknowns, $\mathbf{G}$ is the kernel matrix ($N \times M$) which transforms model space into data space and $\boldsymbol{\nu}$ is the noise vector ($M \times 1$). Here, the “true” data $\mathbf{d}$ are unknown to us, because they are corrupted by noise with the statistical parameters given below. Given $\mathbf{d^*}$, the object of an inversion is to extract information about the model, $\mathbf{m}$. If one assumes that the errors in the observations and the prior information on the model parameters are adequately described by the Gaussian hypothesis, then the posterior probability in the model space is also Gaussian (Tarantola, 1987).

The assumption of Gaussian hypothesis follows many previous approaches carried out by, for example, Kitanidis and Vomvoris (1983), Carrera and Neuman (1986), Kitanidis (1995), Snodgrass and Kitanidis (1997), Woodbury and Ulrych (2000), and Ulrych et al. (2001).

The errors are considered independent with a correlation matrix $\mathbf{C}_d = \sigma_d^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and σ_d (standard deviations of the error) is a measure of discrepancy between observations and model predictions due to several sources of errors such as: measurement errors, roundoff errors and conceptual model misfit. The component due to the measurement error can be reasonable assumed uncorrelated; but the component due to modeling and conceptual uncertainty is likely systematic and correlated (Gaganis and Smith, 2001). Unfortunately the information to characterize the structure of $\mathbf{C}_\mathrm{d}$ is rarely available; for this reason we assumed the errors uncorrelated in agreement with literature works (See for instance Hoeksema and Kitanidis, 1984, Woodbury et Ulrych, 2000 and Fienen et al., 2009) and in fact is a choice that can be rationalized by Maximum Entropy considerations.