Journal of Hydrology 536 (2016) 255–272

Contents lists available at ScienceDirect

Journal of Hydrology

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

Efficient fuzzy Bayesian inference algorithms for incorporating expert knowledge in parameter estimation

Mohammad Mahdi Rajabi [a,*], Behzad Ataie-Ashtiani [a,b]

[a] Department of Civil Engineering, Sharif University of Technology, PO Box 11155-9313, Tehran, Iran
[b] National Centre for Groundwater Research & Training and School of the Environment, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form 5 February 2016
Accepted 15 February 2016
Available online 2 March 2016
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Wolfgang Nowak, Associate Editor

Keywords:
Groundwater modeling
Expert knowledge
Polymorphic uncertainty
Fuzzy-Bayes
Markov chain Monte Carlo
Polynomial chaos expansions

SUMMARY

Bayesian inference has traditionally been conceived as the proper framework for the formal incorporation of expert knowledge in parameter estimation of groundwater models. However, conventional Bayesian inference is incapable of taking into account the imprecision essentially embedded in expert provided information. In order to solve this problem, a number of extensions to conventional Bayesian inference have been introduced in recent years. One of these extensions is 'fuzzy Bayesian inference' which is the result of integrating fuzzy techniques into Bayesian statistics. Fuzzy Bayesian inference has a number of desirable features which makes it an attractive approach for incorporating expert knowledge in the parameter estimation process of groundwater models: (1) it is well adapted to the nature of expert provided information, (2) it allows to distinguishably model both uncertainty and imprecision, and (3) it presents a framework for fusing expert provided information regarding the various inputs of the Bayesian inference algorithm. However an important obstacle in employing fuzzy Bayesian inference in groundwater numerical modeling applications is the computational burden, as the required number of numerical model simulations often becomes extremely exhaustive and often computationally infeasible. In this paper, a novel approach of accelerating the fuzzy Bayesian inference algorithm is proposed which is based on using approximate posterior distributions derived from surrogate modeling, as a screening tool in the computations. The proposed approach is first applied to a synthetic test case of seawater intrusion (SWI) in a coastal aquifer. It is shown that for this synthetic test case, the proposed approach decreases the number of required numerical simulations by an order of magnitude. Then the proposed approach is applied to a real-world test case involving three-dimensional numerical modeling of SWI in Kish Island, located in the Persian Gulf. An expert elicitation methodology is developed and applied to the real-world test case in order to provide a road map for the use of fuzzy Bayesian inference in groundwater modeling applications.

1. Introduction

One of the key challenges routinely encountered in the estimation of groundwater model input parameters is the lack of field measurements. In real world studies, direct measurements of input parameters such as hydraulic conductivity and transmissivity are often inadequate to fully characterize the subsurface variability (Carrera et al., 2005; Hassan et al., 2008), because the tests required to determine these parameters are generally both time consuming and expensive (Ross et al., 2007). Field data on state variables such as groundwater heads and concentrations can be employed to estimate model input parameters through inverse modeling or model calibration, but field data on these state variables is often also insufficient due to the scarcity of boreholes. One way out of this problem is to incorporate other available sources of information which usually exist in the form of soft data such as expert knowledge (Ross et al., 2009; Krueger et al., 2012). The ability of experts to interpret complex and ambiguous evidence in view of the broader experiences make their knowledge an important and yet often untapped source of information (O'Hagan, 2012).

A formal mechanism for the incorporation of expert knowledge in parameter estimation is provided through Bayesian inference (Coolen and Newby, 1994; Lele and Das, 2000). In this context, Bayesian inference enables the fusion of hard data resulting from field measurement with soft data acquired through expert knowledge.

* Corresponding author.
E-mail addresses: mmrajabi@alum.sharif.edu (M.M. Rajabi), ataie@sharif.edu (B. Ataie-Ashtiani).

http://dx.doi.org/10.1016/j.jhydrol.2016.02.029
0022-1694/

Nomenclature

θ	uncertain input parameter(s)
x, x_i	data point/variable
$\gamma(x)$	fuzzy membership function of x
$P(\theta)$	prior distribution
$P(x\|\theta)$	likelihood of the observations x given a parameter set θ
$P(x)$	proportionality constant
$P(\theta\|x)$	posterior distribution
$\boldsymbol{\beta}, \boldsymbol{\beta}_0, \boldsymbol{\delta}, \boldsymbol{\delta}_0$	fuzzy vectors
$y(x)$	random variable representing the output quantity of interest
ξ	random variable
ω_i	mode strength
ψ_i	mode function
q	number of regression points for construction of polynomial chaos expansions
σ_e^2	error variance
N_x	number of data points
$M_i(\theta)$	model output for the location of x_i
θ^*	proposal density in the Markov chain Monte Carlo algorithm
θ^{**}	second stage proposal density in the Markov chain Monte Carlo delayed rejection algorithm
Cov	covariance
S_d	scaling parameter
N_{MC}	maximum allowable length of the Markov Chain
$\alpha_{\text{acceptance}}$	acceptance probability in the Markov chain Monte Carlo algorithm
μ	mean
σ	standard deviation
κ	excess kurtosis
$\epsilon()$	normalized deviations from the respective reference solutions
d	degree of polynomial chaos expansions
$\widetilde{P}(\theta\|x)$	approximate posterior distributions
μ_{post}	mean of the posterior distributions
RMSE_{ave}	average root mean square error
C_I	salinity concentrations in the monitoring points/wells
$\varepsilon(C_I)$	Gaussian random noise
k_H	horizontal permeability of the aquifer in the circular island test case
k_V	vertical permeability of the aquifer in the circular island test case
k_L	permeability of the lower geological layer in the Kish Island test case
k_U	permeability of the upper geological layer in the Kish Island test case
α_L	longitudinal dispersivity
α_T	transverse dispersivity

This procedure is also called 'Bayesian fusion' (Khaleghi et al., 2013). In Bayesian inference, the subjective belief of an individual expert or the inter-subjective belief of several experts about the value of the parameters is represented by the prior probability distributions using expert elicitation methods (Beer et al., 2013; Rinderknecht et al., 2014). These priors then enter a learning process in which they are updated based on hard data, to obtain the posterior distributions (Choy et al., 2009). The influence of expert knowledge on parameter estimations decays with growing amount of hard data (Beer et al., 2013). The Bayesian inference framework for the incorporation of expert knowledge in parameter estimation has been previously applied to modeling studies in hydrology, hydrogeology and water resources management (e.g. Scholten et al., 2013, see Krueger et al., 2012 for a review).

1.1. Shortcomings of the common approach

In practice, the conventional Bayesian inference has a number of important shortcomings in the assimilation of expert knowledge, which we review in the following. First, it has been argued by many researchers (e.g. Coolen and Newby, 1994; Lele and Das, 2000; O'Hagan and Oakley, 2004; Lele and Allen, 2006; Rinderknecht et al., 2012, 2014) that it is very difficult, and sometimes impossible, for experts to express their knowledge as probability distributions in a precise, clear and consistent way. The reason is that expert knowledge often has the form of imprecisely-defined and ambiguous terms and statements rather than exact probability distributions (Li et al., 2013). So it would be more acceptable to describe expert knowledge as intervals, bounds or sets of probability distributions (Rinderknecht et al., 2012). Moreover, using single probability distributions to describe the intrinsically imprecise expert knowledge can bring new, faulty and unwarranted assumptions to the parameter estimation process (Lele and Allen, 2006; Stein et al., 2013). For example, assume that the expert provides an interval $[a, b]$ in which he 'thinks' that the actual value of the parameter x may occur, but has no reason to believe that some of the values in the interval are more probable than others. In this situation, it is standard practice to represent the prior as a uniform probability distribution. However this practice has been shown by a number of studies (e.g. Journel, 1986; Royall, 1997; Stein et al., 2013) to be a misrepresentation of elicitation results, because the bounds of the interval, a and b, are assumed to be precise, neglecting the uncertainty of the expert about the exact value of these bounds. Hence, assuming a uniform prior implicitly ascribes more information than is actually given by the expert provided information.

When expert knowledge is used in conjunction with actual observations to build the prior, both uncertainty and imprecision occur simultaneously. To highlight the difference between uncertainty and imprecision, we briefly review the definition of the two terms. Data is uncertain when the confidence degree of what is stated by the data is less than one. In contrast, data is imprecise if the implied attribute is not singular, but a well-defined or ill-defined set or interval (Khaleghi et al., 2013). Hence data can be uncertain yet precise, and vice versa. These two types are not distinguishable when both are represented by a single probability distribution and the contribution of each to the outcome of the Bayesian inference procedure becomes unknown (Ross et al., 2009). Distinction between uncertainty and imprecision is important as it allows for proper guidance of the data collection procedure.

In conventional Bayesian inference, expert knowledge can only affect the parameter estimation process through the definition of the prior distribution for the uncertain parameters. However in many instances in groundwater modeling, the expert also has other forms of knowledge apart from the possible values of the uncertain parameters. This knowledge can be used to supplement the available hard data in order to improve the precision of the inference results. These other forms of knowledge may include the following: (1) Hard data on state variables such as groundwater heads and concentrations may not be available in certain part of the modeling domain, but the expert may have some knowledge of