Journal of Hydrology 538 (2016) 689–704

Contents lists available at ScienceDirect

Journal of Hydrology

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

Assessing Bayesian model averaging uncertainty of groundwater modeling based on information entropy method

Xiankui Zeng [a], Jichun Wu [a,*], Dong Wang [a], Xiaobin Zhu [a], Yuqiao Long [b]

[a] Key Laboratory of Surficial Geochemistry, Ministry of Education, School of Earth Sciences and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing, China
[b] Nanjing Hydraulic Research Institute, Nanjing, China

ARTICLE INFO

Article history:
Received 23 September 2015
Received in revised form 30 January 2016
Accepted 18 April 2016
Available online 26 April 2016
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Okke Batelaan, Associate Editor

Keywords:
Groundwater modeling
Uncertainty assessment
Bayesian model averaging
Variance
Information entropy

SUMMARY

Because of groundwater conceptualization uncertainty, multi-model methods are usually used and the corresponding uncertainties are estimated by integrating Markov Chain Monte Carlo (MCMC) and Bayesian model averaging (BMA) methods. Generally, the variance method is used to measure the uncertainties of BMA prediction. The total variance of ensemble prediction is decomposed into within-model and between-model variances, which represent the uncertainties derived from parameter and conceptual model, respectively. However, the uncertainty of a probability distribution couldn't be comprehensively quantified by variance solely. A new measuring method based on information entropy theory is proposed in this study. Due to actual BMA process hard to meet the ideal mutually exclusive collectively exhaustive condition, BMA predictive uncertainty could be decomposed into parameter, conceptual model, and overlapped uncertainties, respectively. Overlapped uncertainty is induced by the combination of predictions from correlated model structures. In this paper, five simple analytical functions are firstly used to illustrate the feasibility of the variance and information entropy methods. A discrete distribution example shows that information entropy could be more appropriate to describe between-model uncertainty than variance. Two continuous distribution examples show that the two methods are consistent in measuring normal distribution, and information entropy is more appropriate to describe bimodal distribution than variance. The two examples of BMA uncertainty decomposition demonstrate that the two methods are relatively consistent in assessing the uncertainty of unimodal BMA prediction. Information entropy is more informative in describing the uncertainty decomposition of bimodal BMA prediction. Then, based on a synthetical groundwater model, the variance and information entropy methods are used to assess the BMA uncertainty of groundwater modeling. The uncertainty assessments of groundwater BMA prediction are consistent with that of analytical function examples.

1. Introduction

In recent years, a number of multi-model methods have been proposed to account for uncertainties arising from input parameter and the definition of model structure (Ajami et al., 2007; Neuman, 2003; Neuman et al., 2012; Poeter and Anderson, 2005; Rojas et al., 2008; Ye et al., 2010). These methods believe that it is more appropriate to consider multi-model predictive uncertainty than relying on a single conceptual model. In addition, the Bayesian model averaging (BMA) (Draper, 1995; Hoeting et al., 1999) provides an effective framework for integrating the results of proposed conceptual models. BMA is superior to other methods in incorporating previous information, and quantifying the uncertainties of model parameter and structure independently (Diks and Vrugt, 2010; Rojas et al., 2008; Singh et al., 2010).

Generally, the concept of uncertainty means the lack of certainty. Due to limited knowledge to specify a state, it is impossible to exactly describe the state or future outcome, and there is more than one possibility (Retzer et al., 2009). As for the theme of uncertainty analysis, the most fundamental questions should be the definition of uncertainty, as well as the measurement of the composition, propagation, and interaction of uncertainties. However, currently, there is not a universal framework for uncertainty measurement and assessment.

* Corresponding author at: Key Laboratory of Surficial Geochemistry, Ministry of Education, School of Earth Sciences and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing 210023, China. Tel.: +86 25 83595591.

E-mail address: jcwu@nju.edu.cn (J. Wu).

http://dx.doi.org/10.1016/j.jhydrol.2016.04.038
0022-1694/

In present researches on the assessment of groundwater modeling uncertainty, in general, variance is defined as the measurement of uncertainty (Neuman and Wierenga, 2003; Refsgaard et al., 2006; Rojas et al., 2008; Troldborg et al., 2010; Ye et al., 2010). Variance measures how far a set of numbers is spread out. A small variance indicates that the data points tend to be very close to the mean (expected value) and hence to each other. For the variance measuring method, the total variance of BMA predictive distribution is decomposed into two parts which include within-model and between-model variances (Draper, 1995; Hoeting et al., 1999). Correspondingly, the two terms represent the uncertainties derived from model parameter and conceptual model, respectively. However, the uncertainty of a probability distribution could be more meaningful than variance measurement solely.

As mentioned by Ebrahimi et al. (2010), Khinchin (1957), Rényi (1961), Shannon (1948), that a qualified uncertainty function $U(f)$ (where f denotes a variable's probability distribution, $f=f(f_1, f_2,\ldots, f_n)$, n is the number of variable's possibility) should meet several basic properties. These properties include: (1) continuity, $U(f)$ is continuous to f. (2) symmetry, $U(f)$ is invariant with the order of variable's states. (3) monotonicity, $U(f)$ is a monotonically increasing with n when f_i equals to $1/n$. (4) partition invariance, $U(f)$ is invariant when the variable's space is partitioned into subsets. In addition, the above properties are referred as Shannon's axioms (Shannon, 1948). Two important inferences are derived to ensure that $U(f)$ satisfies with Shannon's axioms and obtains unique solution. Ebrahimi et al. (2010) summarized them as (1) $U(f)$ is a concave function of f. (2) the maximum uncertainty is attained when f is a uniform distribution.

Variance is a concave function. However, the second inference of Shannon's axioms is not satisfied by variance function. Variance is conditionally able to measure the characteristics of a probability distribution, e.g., normal distribution. Ebrahimi et al. (2010) provided some examples for comparing the variance and uncertainty degree of probability distributions. For example, the exponential and gamma distributions ($E(1.0)$ and $G(1.0, 1.41)$) which have the same variance but exhibit different levels of concentration degrees. The beta distribution $B(0.5, 0.5)$ displays a larger concentration degree than the uniform distribution $U(0.0, 1.0)$, although the former has a larger variance. Moreover, the variance of a multivariate probability distribution is extended to a covariance matrix which cannot be represented uniquely as a statistic of f (Ebrahimi et al., 2010). Therefore, variance maybe not a perfect uncertainty functions for measuring a probability distribution, e.g. groundwater predictions.

In information theory, the information entropy is defined as the measurement of degree of uncertainty. The uncertainty of a random variable is defined as the amount of information used to describe this variable. The more information required to describe this variable, the more uncertain this variable is (Ebrahimi et al., 2010). In addition, information entropy theory satisfies all the derived properties of $U(f)$ and Shannon's axioms. This method is applicable to any type of probability distribution, e.g., non-normal distribution. In addition, information entropy is feasible for multivariate situation, and it is invariant under the one-to-one transformations of variables (Ebrahimi et al., 2010; Retzer et al., 2009), such as unit transformation. Through two ecological models describing the uncertainty of resource availability for an organism, Smaldino (2013) demonstrated that variance is good at measuring the uncertainty of a small number of discrete samples, but information entropy is a better measure for multimodal or discontinuous distribution. For complicated groundwater system, the research variables' probability distributions are not necessary to be normal distributions, especially for BMA prediction (Raftery et al., 2005; Rojas et al., 2010). Thus, information entropy method has potential excellent properties for describing BMA uncertainties.

Ideally, the foundational BMA working principle is building mutually exclusive collectively exhaustive (MECE) conceptual models (Refsgaard et al., 2012). In fact, MECE is a rather intractable problem for groundwater modeling. For example, under complicated geological conditions, the uncertainty components (used to build alternative conceptual models) could be not fully independent. For saving computing resource, hydrogeologists always use a few of plausible conceptual models to represent unknown groundwater filed in BMA (Poeter and Anderson, 2005; Rojas et al., 2010). Thus, according to the logical process of BMA, the predictive uncertainty could be induced from three steps that include (1) the selection of model structure, (2) the setting of model parameters and boundary conditions for each conceptual model, and (3) the operation of multi-model averaging. Correspondingly, three uncertainties are produced at these three steps, and they are model selection uncertainty, the parameter uncertainty of each conceptual model, and the overlapped uncertainty caused by the combination of correlated predictions. For a specified model, the predictive uncertainty is only induced from model parameters. For building alternative conceptual models, the uncertainty components of groundwater model could be correlated in practice, e.g. the recharge scenarios and the description of hydraulic conductivity field. The proposed conceptual models may be constructed by partially similar model structures, and then produce correlated predictions. Thus, BMA within-model uncertainty should be the weighted sum of parameter uncertainties over all conceptual models minus the overlapped uncertainty among correlated predictions. Overlapped uncertainty will be zero when alternative model structures are fully independent.

For variance method, the overlapped uncertainty among conceptual models' predictions cannot be appropriately represented by its decomposition formula. By contrast, information entropy method can obtain more information on the composition of BMA uncertainty, as shown below. Furthermore, information entropy theory have been widely used for the related researches of groundwater modeling uncertainties, e.g., the sensitivity analysis by mutual entropy (Mishra et al., 2009), the identification of groundwater contaminant release by relative entropy (Woodbury and Ulrych, 1996), the prediction of a spatial random field by principle of maximum entropy (Orton and Lark, 2009), the description of epistemic and aleatory uncertainties of hydrologic modeling (Gong et al., 2013), and describing the evolution of groundwater flow system by field entropy (Xu and Du, 2014).

Variance and information entropy methods represent different research ideas for uncertainty analysis. Variance concerns how far a set of samples are spread out with their mean value. Information entropy concerns the probabilities of samples. "Uncertainty" is still an open concept, and it has different definitions and interpretations in different fields. The result of information entropy method can be regarded as a supplement to the variance method. The purpose of this paper is not to justify which method is more valid for measuring the uncertainty of a random variable, but to explore the feasibility of two uncertainty measurement methods for various probability distributions, such as the predictions from single conceptual model and BMA. It could provide some insights into the assessment of conceptual model uncertainty. However, to the best of our knowledge, information entropy is rarely applied to the uncertainty assessment of BMA of groundwater modeling. In this paper, information entropy and variance methods will be used to assess the uncertainties of analytical functions and groundwater modeling, and the results of these two methods are compared and summarized.

The remainder of this paper is organized as follows. In Section 2, we provide a condensed description of the methods for integrated uncertainty assessment. Section 3 illustrates the application of variance and information entropy methods through 5 analytical