Journal of Hydrology 517 (2014) 105–119

Contents lists available at ScienceDirect

Journal of Hydrology

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

Constructive epistemic modeling of groundwater flow with geological structure and boundary condition uncertainty under the Bayesian paradigm

Ahmed S. Elshall, Frank T.-C. Tsai *

Department of Civil and Environmental Engineering, Louisiana State University, 3418G Patrick F. Taylor Hall, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 4 December 2013
Received in revised form 31 March 2014
Accepted 12 May 2014
Available online 17 May 2014
This manuscript was handled by Peter K. Kitanidis, Editor-in-Chief, with the assistance of Roseanna M. Neupauer, Associate Editor

Keywords:
Subsurface hydrology
Groundwater
Geological uncertainty
Uncertainty analysis
Bayesian analysis
Multimodel analysis

SUMMARY

Constructive epistemic modeling is the idea that our understanding of a natural system through a scientific model is a mental construct that continually develops through learning about and from the model. Using hierarchical Bayesian model averaging (BMA), this study shows that segregating different uncertain model components through a BMA tree of posterior model probability, model prediction, within-model variance, between-model variance and total model variance serves as a learning tool. First, the BMA tree of posterior model probabilities permits the comparative evaluation of the candidate propositions of each uncertain model component. Second, systemic model dissection is imperative for understanding the individual contribution of each uncertain model component to the model prediction and variance. Third, the hierarchical representation of the between-model variance facilitates the prioritization of the contribution of each uncertain model component to the overall model uncertainty. We illustrate these concepts using the groundwater flow model of a siliciclastic aquifer-fault system. We consider four uncertain model components. With respect to geological structure uncertainty, we consider three methods for reconstructing the hydrofacies architecture of the aquifer-fault system, and two formation dips. We consider two uncertain boundary conditions, each having two candidate propositions. Through combinatorial design, these four uncertain model components with their candidate propositions result in 24 base models. The study shows that hierarchical BMA analysis helps in advancing knowledge about the model rather than forcing the model to fit a particularly understanding or merely averaging several candidate models.

1. Introduction

Our belief about a system is the core element of Bayesian modeling. From this perspective, a groundwater flow model can be viewed as a mental construct that aims at simulating our empirical, theoretical and abstract understanding of the flow field in the natural aquifer. In other words, we do not simulate the natural flow field, but rather we are simulating our current degree of knowledge about the flow field of the natural system. Accordingly, the treatment of uncertainty is essential since several candidate knowledge propositions exist about the model data, structure, parameters and processes.

Data uncertainty arises from different measurement techniques, measurement errors and mathematical expressions for data interpretation (Singha et al., 2007). Model structural uncertainty arises because the model approximate representation of the complex environment is not unique, which is due to several reasons. First, the characteristics of the spatial variability remain "imperfectly known" (Cardiff and Kitanidis, 2009). Second, different heterogeneity conceptualizations lead to diverse mathematical expressions for quantitative spatial relationships (Kitanidis, 1986; Koltermann and Gorelick, 1996; Refsgaard et al., 2012). Third, due to the scarcity of subsurface data, quantitative methods cannot generally afford a precise description of the complex spatial subsurface geological variations (e.g., Sakaki et al., 2009; Li et al., 2012). Parameter uncertainty arises from the precision of the estimated model parameters. This precision is a factor of maximum likelihood estimation in a rugged, nonseparable and noisy search landscape. A second inherent challenge of parameter estimation is ill-posedness that arises mainly from nonuniqueness and insensitivity (Yeh, 1986; Carrera and Neuman, 1986). The situation is even more intricate since model structure inadequacy can be compensated by biased parameter estimation, and the model solution can be biased toward unobserved variables in the model (Refsgaard et al., 2006). For a current discussion on the uncertainty of groundwater model simulation

* Corresponding author. Tel.: +1 225 578 4246.
E-mail addresses: aelsha1@lsu.edu (A.S. Elshall), ftsai@lsu.edu (F.T.-C. Tsai).

http://dx.doi.org/10.1016/j.jhydrol.2014.05.027

and prediction, the reader is referred to Gupta et al. (2012). Yet based on this brief account, we bring a fundamental question of how to bridge the gap between synthetic mental principles such as mathematical expressions, and empirical observations such as site observation data, when uncertainty exists on both sides.

Using multiple models to account for uncertainty resulting from model data, structure, parameters and processes, strategies as model selection (Poeter and Anderson, 2005), model elimination (Refsgaard et al., 2006), model reduction (Doherty and Christensen, 2011), model combination (Neuman, 2003; Neuman and Wierenga, 2003; Ye et al., 2004; Tsai and Li, 2008a,2008b, Rojas et al., 2008, 2010; Wöhling and Vrugt, 2008; Li and Tsai, 2009; Singh et al., 2010; Troldborg et al., 2010; Seifert et al., 2012) and model discrimination (Usunoff et al., 1992; Tsai et al., 2003; Ye et al., 2010; Foglia et al., 2013; Tsai and Elshall, 2013) are commonly used. A main concern among these different strategies is the incorporation of different candidate knowledge propositions and the uncertainty quantification. A more critical and less acknowledged concern is epistemic uncertainty (Refsgaard et al., 2006, 2007; Clark et al., 2011; Gupta et al., 2012; Beven and Young, 2013), which refers to "the uncertainty due to imperfect knowledge" (Refsgaard et al., 2007). To account for our ignorance, which is the lack of knowledge and the incorrect understanding, epistemic uncertainty is commonly addressed through possibility theory, imprecise probability or pedigree analysis (Agarwal et al., 2004; Refsgaard et al., 2006; Baudrit et al., 2007; He et al., 2008).

In this study we present a complementing prospective on epistemic uncertainty through hierarchical Bayesian model averaging (BMA) analysis (Tsai and Elshall, 2013). The basic element of the hierarchical BMA analysis is the base models, which are all of the considered models. The base models are developed following a combinatorial design to represent the candidate propositions of all sources of uncertainty. Selecting the base models in hierarchical BMA is flexible since new propositions for an uncertain model component can be readily incorporated. However, if we are interested in obtaining a BMA solution based on all the base models, this brings the question of how to select the base models such that to have a collectively exhaustive set of models. Fundamentally, the hierarchical BMA does not overcome this problem since in principal it presents the general form of the collection BMA in Hoeting et al. (1999). However, unlike the collection BMA in which our modeling approach is oriented toward obtaining a BMA solution (i.e., BMA prediction and variance), the hierarchical BMA aims at shifting to a constructive epistemic modeling approach, in which candidate model propositions are tested to learn about individual model components and potentially model adequacy.

The notion "constructive" is basically that "to know the truth means essentially to construct such a truth" (Primiero, 2008). Constructive epistemology is a "meta science" way of thinking that assumes that the mental world is actively constructed, in which there is a developmental path from some initial state, rather than a teleological progress towards some final state (Riegler, 2012). From this prospective, the hierarchical BMA treatment acknowledges epistemic uncertainty, which is mainly that the base models are incomplete since they do not collectively exhaust the space of possible models. The hierarchical BMA treatment acknowledges as well that it could be the case that some model propositions can be incorrectly included in the model (Gupta et al., 2012). Accordingly, constructive epistemic modeling is in agreement with what Christakos (2004) proposes that regarding the model solution as epistemic, in which the model describes incomplete knowledge about nature and focuses on knowledge synthesis, can lead to more realistic results than the (conventional) ontological solution that assumes that the model describes nature per se and focuses on form manipulations.

However, acknowledging the use of an incomplete set of base models brings the question of the statistical meaning of the posterior model probabilities. As presented by Renard et al. (2010), since a BMA key assumption is that the supplied set of models is complete, which is difficult to achieve in practice, then "it is unclear what the posterior predictive uncertainty actually represents when this assumption is not met". Following Williamson (2005), one can make the argument that an objective probabilistic decision for a specific model, which has no obvious collective (von Mises, 1964), repeatable experiment (Popper, 1959) or chance fixer (Popper, 1990) concerning its physical probability, one needs to ascribe an "epistemic probability" (Williamson, 2005) to this model as a function of our factual knowledge. Under the epistemic probability stance, probability is viewed as being neither physical mind-independent features of the world nor arbitrary and subjective entities, but rather an objective degree of belief (Williamson, 2005) since it does not vary from one agent to another because it is coherent and honors data. Ellison (2004) states that "posterior probability distributions are an epistemological alternative to *P*-values, and provide a direct measure of the degree of belief that can be placed on models, hypotheses, or parameter estimates". Accordingly, the posterior predictive variance, which is a function of posterior model probabilities, presents under BMA neither the true variance nor a representation of any frequency. It simply represents the uncertainty of our current state of knowledge as this study shows. It is noted that a *P*-value is the significance probability for testing null-hypothesis (Schervish, 1996).

Essentially, true variance can only be known if we know the deviation from the true model, which is almost not possible (Rubin, 2003). Even if the "true model" is known, the question still whether synthetic mental principles – such as mathematical expressions and conceptualization of spatial variability – are statements of what exist externally in nature, or they are mental statements based on relative empirical observation and their inherent shortcomings as pointed out by Jaynes (1990, 2003). Following a similar line of thought, Gupta et al. (2012) propose revising the commonly used term "model structure error" with "model structure adequacy", since the former term "implies the existence of some 'true' value from which the difference can (in principle) be measured". This last point suggests the plausibility of "epistemic probability" (Williamson, 2005), and the plausibility of accommodating different candidate model propositions in a constructive epistemic framework that is guided by scientific reasoning.

This research develops groundwater models of a siliciclastic aquifer-fault system to illustrate the use of hierarchical BMA as a constructive epistemic framework, which advances knowledge about the model rather than forcing the model to fit a particular understanding or merely averaging several candidate models as some final teleological state. In other words, the modeling objective is to use the BMA trees of posterior model probability, prediction and variance to increase learning. The groundwater model construction involves four uncertain model components, which are the hydrofacies architecture reconstruction method, the geological formation dip and two uncertain boundary conditions. Through dissecting the uncertain model components, the hierarchical BMA allows for comparative evaluation of candidate model propositions, for prioritizing the uncertain model components, for depicting the prediction and uncertainty propagation, and finally for updating our knowledge about the model.

2. Methodology

2.1. Hierarchal Bayesian model averaging

In this study we extend the hierarchal Bayesian model averaging (BMA) methodology in Tsai and Elshall (2013) to account for prior model probability. This shall allow the use of geological models as prior information for groundwater flow models to link geological