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
0022-1694/

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

related uncertainty with groundwater flow related uncertainty. The geological models are hydrofacies architecture models with different conceptualizations as described in Sections 3.2 and 3.3.

To permit systematic segregation of the uncertain model components with their candidate propositions, the hierarchical BMA (Tsai and Elshall, 2013) provides the general form of the collection BMA in Hoeting et al. (1999) that treats all models at one level. The base level of the hierarchy is a collection of all candidate models that are produced following a combinatorial design. For example, if the first uncertain model component is two candidate boundary condition propositions, then two models with the two different boundary condition propositions are produced and calibrated. If a second uncertain model component is three candidate hydrofacies architecture reconstruction propositions, then the combinatorial design results in six base models. The hierarchical BMA analysis can include any number of sources of uncertainty. This allows the construction of a BMA tree of posterior model probabilities, prediction means, prediction within-model variance and prediction between-model variance. Thus, each level in the hierarchy distinguishes uncertainty arising from an uncertain model component, allowing for comparative evaluation of different propositions and uncertainty prioritization.

Each level in the hierarchy represents an uncertain model component. Considering p sources of uncertainty, a base model is denoted as $M_{\underbrace{(ij\ldots lm)}_{p}} \in \mathbf{M}_p$ at level p. The subscript $\underbrace{(ij\ldots lm)}_{p}$ locates the model hierarchically top down from the first level, to the second level and so forth to reach to level p. Each index represents one source of uncertainty. For example, $M_{(i)} \in \mathbf{M}_1$ is model i at level 1, $M_{(ij)} \in \mathbf{M}_2$ is model j at level 2, which is a child model to parent model i at level 1. $M_{(ijk)} \in \mathbf{M}_3$ is model k at level 3, which is a child model to the parent model j at level 2 and the grandparent model of model i at level 1. From bottom up, parent models $\mathbf{M}_{p-1}$ at level $p-1$ is composed of the child models $\mathbf{M}_p$ at level p. Models $\mathbf{M}_{p-2}$ at level $p-2$ are composed of models $\mathbf{M}_{p-1}$ at level $p-1$ and so forth until the hierarch level is reached. The top level (hierarch level) of the hierarchy consists of one model, which is termed the hierarch BMA model. The hierarch level is level zero. Eventually, the hierarch BMA model averages information of all base models and thus it is the same at the BMA model of the collection BMA.

Consider models at any level, say n, according to the law of total probability, the posterior probability for predicted quantity Δ given data $\mathbf{D}$ is

$$\Pr(\Delta|\mathbf{D},\ \mathbf{M}_n) = E_{\mathbf{M}_{n+1}} E_{\mathbf{M}_{n+2}} \cdots E_{\mathbf{M}_p}\ [\Pr(\Delta|\mathbf{D},\ \mathbf{M}_p)] \tag{1}$$

where $E_{\mathbf{M}_p}$ is the expectation operator with respect to models $\mathbf{M}_p$ at level p; $\Pr(\Delta|\mathbf{D},\ \mathbf{M}_p)$ is the posterior probability of predicted quantity Δ given data $\mathbf{D}$ and models $\mathbf{M}_p$ at level p; and the expectation $E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D},\ \mathbf{M}_p)]$ is the posterior probability averaging at level p.

The total expectation and the total variance are two important statistical moments for Δ in BMA. The law of total expectation that derives the mean Δ at level n given data $\mathbf{D}$ is

$$E(\Delta|\mathbf{D},\ \mathbf{M}_n) = E_{\mathbf{M}_{n+1}} E_{\mathbf{M}_{n+2}} \cdots E_{\mathbf{M}_p}\ [E(\Delta|\mathbf{D},\ \mathbf{M}_p)]. \tag{2}$$

where $n < p$. Eq. (2) provides thorough information for analysts, who can see all possible averaged predicted quantities using various BMA models at different levels, while the collection BMA (one-level) only provides one overall expectation of all models. Using Eq. (2) at any level, a BMA tree of mean predictions can be obtained.

Using the law of total variance, the variance of prediction Δ at level $n < p$ is

$$\begin{aligned}\mathrm{Var}(\Delta|\mathbf{D},\ \mathbf{M}_n) = {} & E_{\mathbf{M}_{n+1}}[\mathrm{Var}(\Delta|\mathbf{D},\ \mathbf{M}_{n+1})] \\ & + \mathrm{Var}_{\mathbf{M}_{n+1}}[E(\Delta|\mathbf{D},\ \mathbf{M}_{n+1})]\end{aligned} \tag{3}$$

The first and second terms at the right side of Eq. (3) are the within-model variance and the between-model variance of prediction, respectively, using models at level $n+1$. Since Eq. (3) is a recursive equation, the within-model variance is obtained by

$$\begin{aligned}E_{\mathbf{M}_{n+1}}[\mathrm{Var}(\Delta|\mathbf{D},\ \mathbf{M}_{n+1})] = {} & E_{\mathbf{M}_{n+1}}[E_{\mathbf{M}_{n+2}}[\mathrm{Var}(\Delta|\mathbf{D},\ \mathbf{M}_{n+2})] \\ & + \mathrm{Var}_{\mathbf{M}_{n+2}}[E(\Delta|\mathbf{D},\ \mathbf{M}_{n+2})]]\end{aligned} \tag{4}$$

The between-model variance $\mathrm{Var}_{\mathbf{M}_{n+1}}[E(\Delta|\mathbf{D},\ \mathbf{M}_{n+1})]$ is obtained as follows:

$$\mathrm{Var}_{\mathbf{M}_{n+1}}[E(\Delta|\mathbf{D},\ \mathbf{M}_{n+1})] = E_{\mathbf{M}_{n+1}}[[E(\Delta|\mathbf{D},\ \mathbf{M}_{n+1}) - E(\Delta|\mathbf{D},\ \mathbf{M}_n)]^2] \tag{5}$$

Eq. (5) distinguishes the between-model variance for each uncertain model component addressing the remark of Gupta et al. (2012) that "while model averaging provides a framework for explicitly considering (conceptual) model uncertainty, it currently lumps all errors into a single misfit term and does not provide insights into model structural adequacy". For this study, we will consider four sources of uncertainty, i.e., $p = 4$. The expectation of prediction at the hierarch level is

$$E[\Delta|\mathbf{D}] = E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} E_{\mathbf{M}_4}[E(\Delta|\mathbf{D},\ \mathbf{M}_4)] \tag{6}$$

The total variance of prediction at the hierarch level is

$$\begin{aligned}\mathrm{Var}[\Delta|\mathbf{D}] = {} & E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} E_{\mathbf{M}_4}[\mathrm{Var}(\Delta|\mathbf{D},\ \mathbf{M}_4)] \\ & + E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} \mathrm{Var}_{\mathbf{M}_4}[E(\Delta|\mathbf{D},\ \mathbf{M}_4)] \\ & + E_{\mathbf{M}_1} E_{\mathbf{M}_2} \mathrm{Var}_{\mathbf{M}_3}[E(\Delta|\mathbf{D},\ \mathbf{M}_3)] \\ & + E_{\mathbf{M}_1} \mathrm{Var}_{\mathbf{M}_2}[E(\Delta|\mathbf{D},\mathbf{M}_2)] + \mathrm{Var}_{\mathbf{M}_1}[E(\Delta|\mathbf{D},\mathbf{M}_1)]\end{aligned} \tag{7}$$

The first term on the right hand side in Eq. (7) is the prediction variance due to uncertain model parameters. The second term is the between-model variance for level 3. The third term is the between-model variance for level 2. The fourth term is the between-model variance for level 1. The last term is the between-model variance for the hierarch level.

2.2. Posterior model probabilities for hierarchical BMA

To calculate $E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D},\ \mathbf{M}_p)]$ in Eq. (1), the posterior model probabilities $\Pr(\mathbf{M}_p|\mathbf{D})$ for the base models are needed. Based on the Bayes rule, we have

$$\Pr(\mathbf{M}_p|\mathbf{D}) = \frac{\Pr(\mathbf{D}|M_{(ij\cdots lm)})\Pr(M_{(ij\cdots lm)})}{\sum_i \sum_j \cdots \sum_l \sum_m \Pr(\mathbf{D}|M_{(ij\cdots lm)})\Pr(M_{(ij\cdots lm)})} \tag{8}$$

where $\Pr(\mathbf{D}|M_{(ij\cdots lm)})$ is the likelihood of a base model and $\Pr(M_{(ij\cdots lm)})$ is the prior model probability of a base model. The posterior model probabilities $\Pr(\mathbf{M}_p|\mathbf{D})$ are also referred to as model weights. The posterior model probabilities at level n are

$$\Pr(\mathbf{M}_n|\mathbf{D}) = \frac{\sum_{n+1} \cdots \sum_l \sum_m \Pr(\mathbf{D}|M_{(ij\cdots lm)})\Pr(M_{(ij\cdots lm)})}{\sum_i \sum_j \cdots \sum_l \sum_m \Pr(\mathbf{D}|M_{(ij\cdots lm)})\Pr(M_{(ij\cdots lm)})} \tag{9}$$

The conditional posterior model probabilities of base models under their parent models are

$$\Pr(\mathbf{M}_p|\mathbf{D},\ \mathbf{M}_{p-1}) = \frac{\Pr(\mathbf{D}|M_{(ij\cdots lm)})\Pr(M_{(ij\cdots lm)}|M_{(ij\cdots l)})}{\sum_m \Pr(\mathbf{D}|M_{(ij\cdots lm)})\Pr(M_{(ij\cdots lm)}|M_{(ij\cdots l)})} \tag{10}$$

where $\Pr(M_{(ij\cdots lm)}|M_{(ij\cdots l)})$ is the conditional prior model probability of a base model $M_{\underbrace{(ij\ldots lm)}_{p}} \in \mathbf{M}_p$ under its parent model $M_{\underbrace{(ij\ldots lm)}_{p-1}} \in \mathbf{M}_{p-1}$. Eq. (10) is also referred to as the conditional model weights. The conditional posterior model probabilities for models at level n under their parent models are

$$\Pr(\mathbf{M}_n|\mathbf{D},\mathbf{M}_{n-1})=\frac{\sum_{n+1}\cdots\sum_l\sum_m\Pr(\mathbf{M}_{(ij\cdots lm)}|\mathbf{D},\mathbf{M}_{(ij\cdots l)})\Pr(M_{(ij\cdots lm)}|M_{(ij\cdots l)})}{\sum_n\sum_{n+1}\cdots\sum_l\sum_m\Pr(\mathbf{M}_{(ij\cdots lm)}|\mathbf{D},\mathbf{M}_{(ij\cdots l)})\Pr(M_{(ij\cdots lm)}|M_{(ij\cdots l)})} \tag{11}$$

Therefore, each model at any level has its own posterior model probability as in Eq. (9) and conditional posterior model probability as in Eq. (11). As a result, a BMA tree of posterior model probabilities can be obtained.

2.3. Approximation of posterior model probabilities

The posterior model probabilities for the base models can be computed through sampling techniques (Wöhling and Vrugt, 2008; Rojas et al., 2010) or can be approximated using information-theoretic criteria (Neuman, 2003; Ye et al., 2004; Poeter and Anderson, 2005; Singh et al., 2010; Ye et al., 2010; Foglia et al., 2013). In this study, we estimate the posterior model probabilities by the Bayesian information criterion (BIC) (Schwarz, 1978; Draper, 1995; Raftery, 1995):

$$\Pr(\mathbf{D}|M_{(ij\ldots lm)})\approx\exp\left[-\frac{1}{2}\mathrm{BIC}_{(ij\ldots lm)}\right] \tag{12}$$

where

$$\mathrm{BIC}_{(ij\ldots lm)}=Q_{(ij\ldots lm)}+L\ln 2\pi+m_{(ij\ldots lm)}\ln L \tag{13}$$

and

$$Q_{(ij\ldots lm)}=(\Delta^{cal}(\hat{\boldsymbol{\beta}}_{(ij\ldots lm)})-\Delta^{obs})^T\ C_\Delta^{-1}(\Delta^{cal}(\hat{\boldsymbol{\beta}}_{(ij\ldots lm)})-\Delta^{obs}) \tag{14}$$

where $Q_{(ij\ldots lm)}$ is the sum of the weighted squared fitting errors between calculated Δ^{cal} from model $m_{(ij\ldots lm)}$ and observed Δ^{obs}; $m_{(ij\ldots lm)}$ is the number of the model parameters $\hat{\boldsymbol{\beta}}_{(ij\ldots lm)}$; L is the number of observation data $\mathbf{D}$; $\hat{\boldsymbol{\beta}}_{(ij\ldots lm)}$ are the maximum-likelihood estimated model parameters in model $M_{(ij\ldots lm)}$; and C_Δ is a diagonal matrix of variances of Δ errors, which is calculated as follows from a sufficient number of realizations:

$$\sigma_k^2=\mathrm{var}\left[\Delta_{r,q,k}^{cal}(\hat{\boldsymbol{\beta}}_{(ij\ldots lm)})-\Delta_k^{obs}\right] \tag{15}$$

where k is a data point; r is the number of realizations; and q is the number of base models.

To avoid underestimating posterior probabilities of good models, we use the variance window (Tsai and Li, 2008a; Singh et al., 2010). We substitute Eq. (12) into Eq. (8) and divided by $\exp\left[-\frac{1}{2}\mathrm{BIC}_{\min}\right]$ to calculate the posterior model probability for a base model:

$$\Pr(M_{(ij\ldots lm)}|\mathbf{D})=\frac{\exp\left[-\frac{1}{2}\alpha\Delta\mathrm{BIC}_{(ij\ldots lm)}\right]\Pr(M_{(ij\cdots lm)})}{\sum_i\sum_j\cdots\sum_l\sum_m\exp\left[-\frac{1}{2}\alpha\Delta\mathrm{BIC}_{(ij\ldots lm)}\right]\Pr(M_{(ij\cdots lm)})} \tag{16}$$

where $\Delta\mathrm{BIC}_{(ij\ldots lm)}=\mathrm{BIC}_{(ij\ldots lm)}-\mathrm{BIC}_{\min}$; $\mathrm{BIC}_{\min}$ is the minimum BIC value among all the base models; and α is the scaling factor for the variance window. If the scaling factor α is zero, then all models are weighted equally; if it is unity, then the models are weighted according Occam's window; if it is smaller than unity, then we enlarge Occam's window to accept more models. We discuss the calculation of the prior model probability $\Pr(M_{(ij\cdots l\ m)})$ based on the geological models in the next section.

Similarly, substituting Eq. (12) into Eq. (10) we obtain the conditional posterior model probabilities of the base models under their parent models.

3. Numerical study

3.1. Baton Rouge aquifer-fault system, southeastern Louisiana

The aquifer system in Baton Rouge consists of a sequence of aquifers and aquicludes extending to a depth of 900 m (Tomaszewski, 1996). The aquifer units are amalgamated fluvial sand bodies (Chamberlain, 2012). Meyer and Turcan (1955) classified and named these sand units by their approximate depth below ground level in the Baton Rouge industrial district. The study area includes the east–west trending Baton Rouge fault and the Denham Springs-Scotlandville fault (Tepetate fault) shown in Fig. 1, which crosscuts the aquifer/aquiclude sequence (McCulloh and Heinrich, 2012). The Baton Rouge fault historically separates a sequence of fresh and brackish aquifers at the north and south of the fault, respectively. Prior to heavy pumping, groundwater flow in the vicinity of the Baton Rouge fault was southward along the topographic gradient. However, extensive anthropogenic pumping starting 1940s reversed the flow direction and resulted in salt water intrusion from south of the Baton Rouge fault (Tomaszewski, 1996; Tsai, 2010). This study focuses on the "2000-foot" sand between the two faults. The model area shown in Fig. 1 includes the "2000-foot" sand and the "2400-foot" sand in the north of the Denham Springs-Scotlandville fault, and the "2000-foot" and the "1700-foot" sand in the south of the Baton Rouge fault. The connections of the "2000-foot" sand between the two faults to other sands are due to sand unit displacement at the faults.

3.2. Geological structure uncertainty

To depict the complexity of the aquifer-fault system in the Baton Rouge area, Elshall et al. (2013) used indicator geostatistics at the dispositional environmental scale (Koltermann and Gorelick, 1996; Rubin, 2003) to reconstruct the hydrofacies architecture. This scale of characterization depicts the strong bimodal heterogeneity of the pervious sand units and impervious clay units. Accordingly, the binary hydrofacies architecture delineates the spatial extent of the sand units, the interconnections between the sand units, and the flow pathways across the faults. However, since the characteristics of spatial variability remain elusive, the best quantitative method to reconstruct the subsurface hydrofacies architecture is prior unknown. Moreover, given multiple hydrofacies models, it is unclear whether the best hydrofacies model in terms of model probability would generally result in the best flow model. In this study we test three indicator geostatistics methods for reconstructing the hydrofacies architecture, which are indictor zonation (IZ) (Tsai, 2009), generalized parameterization (GP) (Tsai and Yeh, 2004; Tsai, 2006) and indicator kriging (IK) (Johnson and Dreiss, 1989; Johnson, 1995; Trevisani and Fabbri, 2010). For using these three hydrofacies architecture reconstruction propositions in this study, the reader is referred to Elshall et al. (2013).

The second uncertain model component with respect to geological structure is the formation dip. This study adopts two formation dip propositions of 0.29° and 0.35° from Elshall et al. (2013), which were estimated using an inverse modeling approach and a clustering approach, respectively. Using 491 electric well logs for hydrofacies architecture reconstruction, Fig. 2 shows the flow pathways to the "2000-foot" sand in the middle domain through of the Baton Rouge fault and the Denham-Springs Scotlandville fault, given the propositions IZ, GP and IK and the formation dip proposition D1 = 0.29°. The propositions IZ, GP and IK provide sharp edged conditional estimates, non-smooth conditional estimates and smooth conditional estimates, respectively. Given the second formation dip proposition D2 = 0.35° and the propositions IZ, GP and IK, Fig. 3 shows the architectures of the Baton Rouge fault and the Denham Springs-Scotlandville fault. Comparing Denham

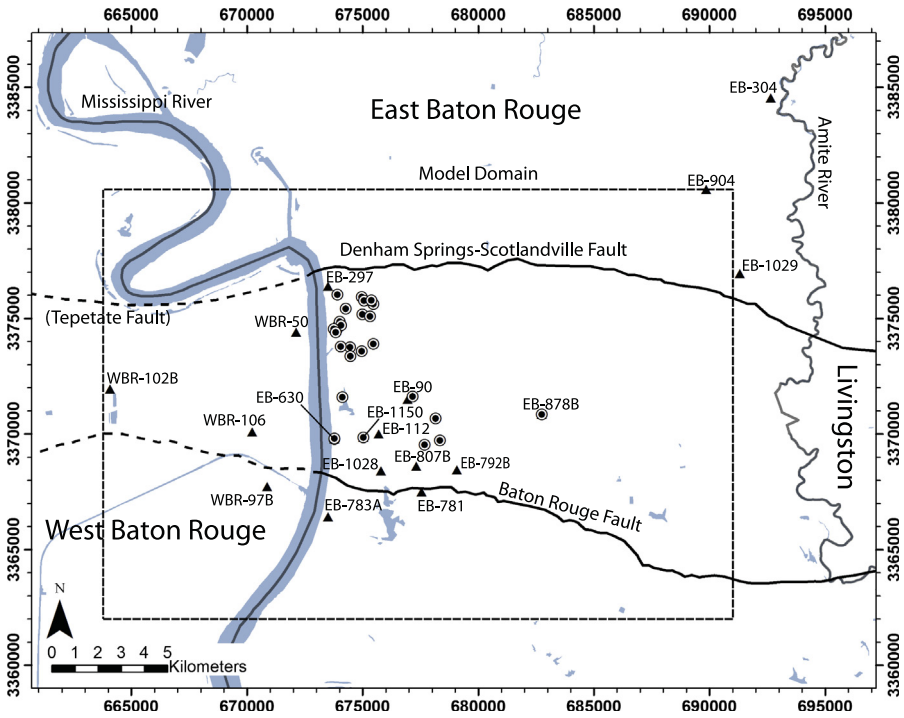

Fig. 1. The map of study area showing the locations of the USGS observation wells (triangles) and pumping wells (circles) for the "2000-foot" sand. The bold solid lines are fault lines identified by the surface expression and the bold dashed lines are the approximate surface locations of the faults (McCulloh and Heinrich, 2012). The gray lines and areas are water bodies.

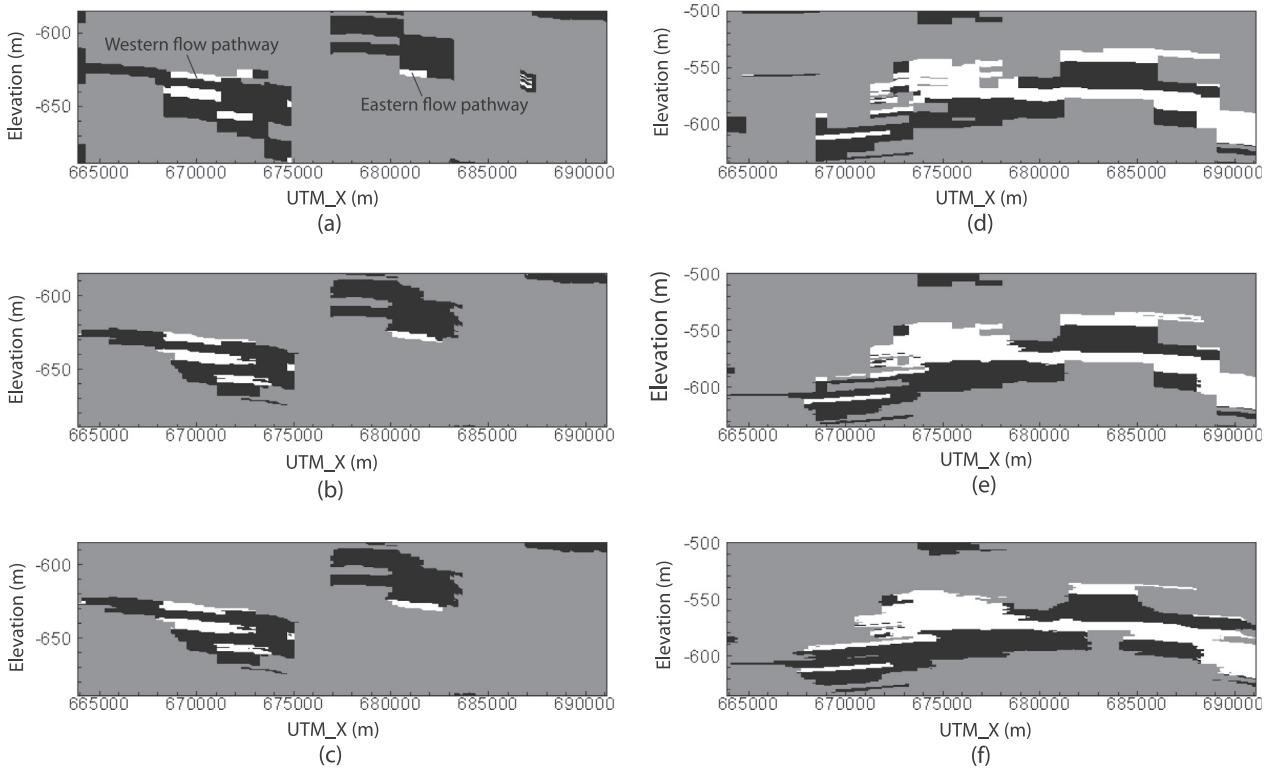

Fig. 2. For formation dip proposition D1 = 0.29°: the architecture of the Baton Rouge fault using (a) indicator zonation, (b) generalized parameterization and (c) indicator kriging, and the architecture of the Denham Springs-Scotlandville fault using (d) indicator zonation, (e) generalized parameterization and (f) indicator kriging. Black areas are clay units north of the fault. Gray areas are clay units south of the fault. White areas are potential flow pathways to the "2000-foot" sand through the faults. The fault cross sections follow the fault line in Fig. 1, but are rendered in 2D for clarity.

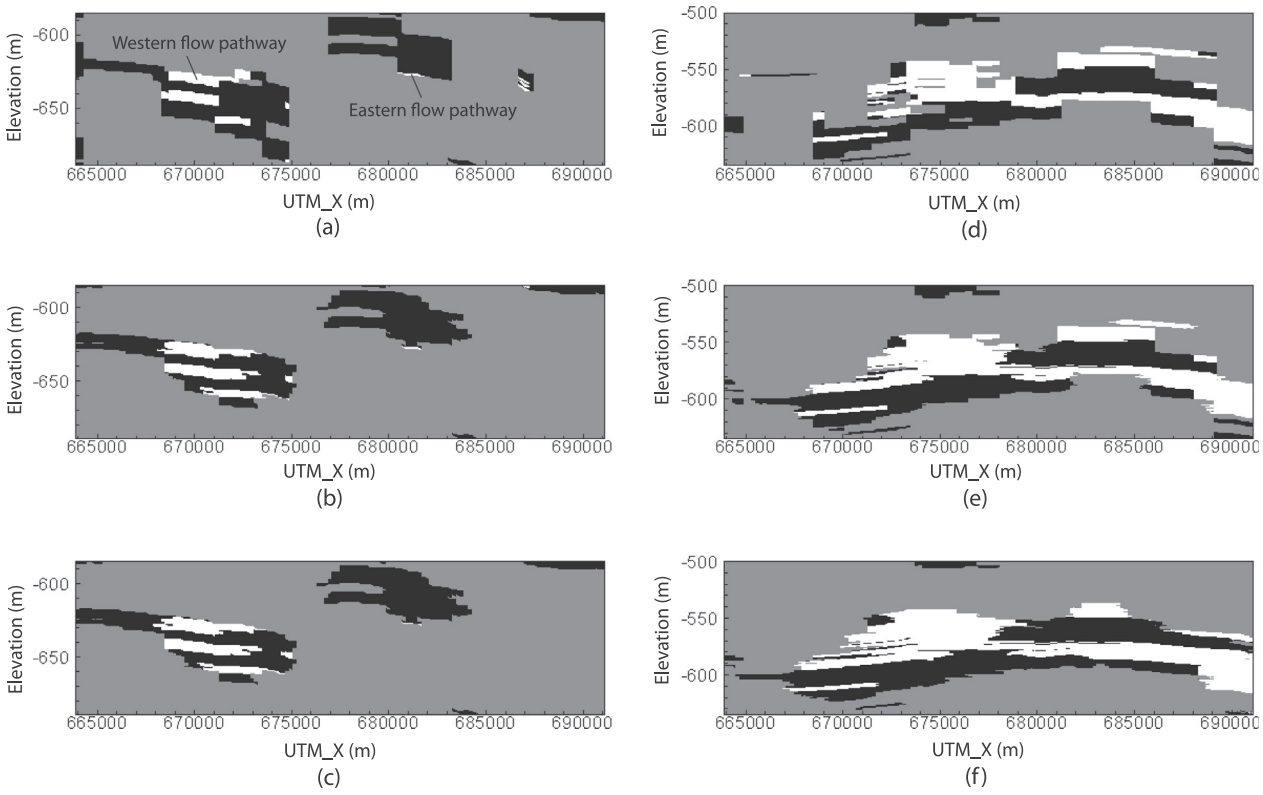

Fig. 3. For formation dip proposition D2 = 0.35°: the architecture of the Baton Rouge fault using (a) indicator zonation, (b) generalized parameterization and (c) indicator kriging, and the architecture of the Denham Springs-Scotlandville fault using (d) indicator zonation, (e) generalized parameterization and (f) indicator kriging. Black areas are clay units north of the fault. Gray areas are clay units south of the fault. White areas are potential flow pathways to the "2000-foot" sand through the faults. The fault cross sections follow the fault line in Fig. 1, but are rendered in 2D for clarity.

Springs-Scotlandville fault architectures in Figs. 2 and 3 show that using different formation dips produces relatively different hydrofacies architectures. The Baton Rouge fault architectures in Figs. 2 and 3 show an important remark that the flow pathway at the east becomes thinner by increasing the formation dip.

3.3. Prior model probabilities from geological models

Given the three methods for hydrofacies architecture reconstruction and the two formation dips, combinatorial design results in six hydrofacies architecture models. Following the same procedure as Tsai and Elshall (2013), we calibrated these six hydrofacies architecture models using lithological data and calculated the model probabilities as shown in Table 1. The results show that the hydrofacies models with proposition D2 have higher model probabilities than those with proposition D1. In addition, models with proposition GP have higher model probabilities. The best hydrofacies model is GPD2. The calculated hydrofacies model probabilities are used as prior model probabilities for groundwater flow models.

Table 1
Model probabilities of the six hydrofacies architecture models: Q is the sum of weighted errors between the estimated and observed hydrofacies (Tsai and Elshall, 2013), ΔBIC is the BIC value of each hydrofacies architecture model minus the minimum BIC value among all the hydrofacies architecture models, and the model probabilities of the six hydrofacies architecture models represent the prior probabilities of the aquifer structures of the groundwater models.

Hydrofacies model	Q	ΔBIC	Model probability
IZD1	6092	733	0.007
IZD2	5619	259	0.100
GPD1	5845	497	0.029
GPD2	5360	0	0.424
IKD1	5839	485	0.028
IKD2	5365	5	0.412

Note that the vertical discretization for the hydrofacies models is at one-foot (0.304 m) intervals. For developing groundwater flow models, we vertically aggregate the detailed vertical discretization of the hydrofacies architecture. In order to maintain an error of less than 0.1% between the original and the aggregated hydrofacies architectures, the aggregated hydrofacies architectures resulted into 29 layers with variable thickness from 1 to 6 m.

Fig. 4 shows the six hydrofacies architectures and their averaged architectures, using simple model averaging and Bayesian model averaging, for a selected layer that has a top elevation of −556 m NGVD29 at northeast corner and top elevation of −667 m NGVD29 at the southwest corner. The two methods IZ and GP produce slightly different architectures as shown in Fig. 4(a)–(d). Yet GP and IK methods produce relatively similar architectures as shown in Fig. 4(c)–(f), which is mainly because of the large electric well log data set. Fig. 4(a)–(f) shows that propositions D1 and D2 produce relatively different architectures, particularly in the north domain. For model averaging as shown in Fig. 4(g)–(h), the gray areas with indicator values between 0 and 1 represent uncertain regions for clay hydrofacies and sand hydrofacies. The result of simple model averaging in Fig. 4(g) shows large uncertainty about the clay and sand hydrofacies distribution. However, the Bayesian model averaging results in less uncertainty because the propositions IK and GP are similar and have much higher hydrofacies model probabilities than the proposition IZ, and the proposition D2 has relatively higher hydrofacies model probabilities than proposition D1.

3.4. Boundary condition uncertainty

Given multiple geological structure propositions, the study assigns no-flow boundary condition to clay hydrofacies and a time-varied constant head boundary condition to sand hydrofacies.

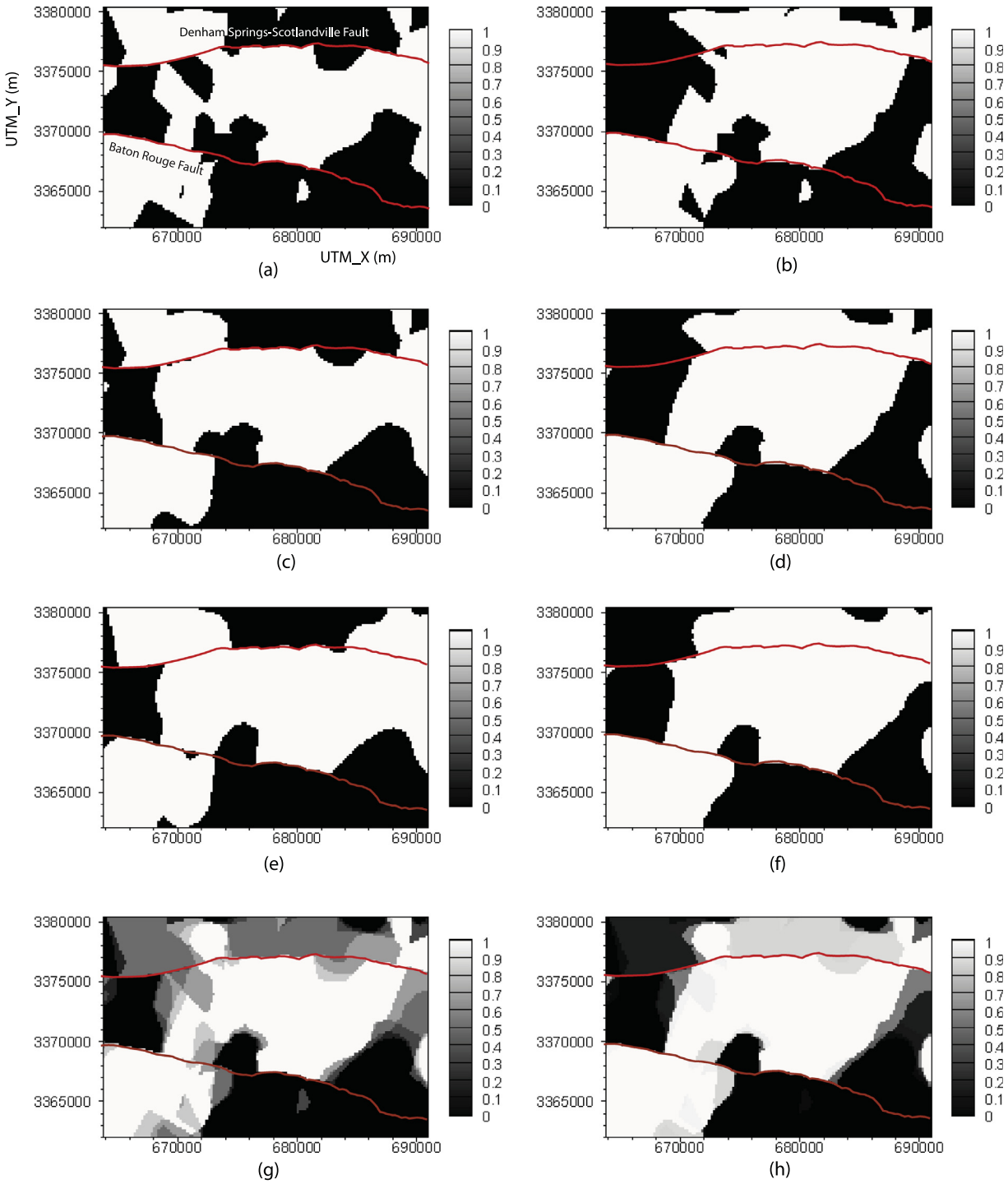

Fig. 4. Hydrofacies architectures for a selected layer for (a) IZD1 model, (b) IZD2 model, (c) GPD1 model, (d) GPD2 model, (e) IKD1 model, (f) IKD2 model, (g) simple model average of the six hydrofacies architectures and (h) Bayesian model average of the six hydrofacies architectures. White areas are sand unit and black areas are clay unit.

Yet different definitions of the boundary conditions can result in different groundwater flow models (Rojas et al., 2008, 2010).

The study aims at simulating the groundwater heads from January 1975 to December 2010 with monthly discretization resulting in 432 stress periods. Accordingly, 432 time-varied constant-head values need to be defined for each sand boundary cell. Assigning time-varied constant-head boundary values is uncertain when very limited head observation data is available near the boundaries. This is the case at the boundaries in the north domain in which only four head observations are available from the USGS observation wells EB-904 and EB-1029 (see Fig. 1 for location). We consider two candidate propositions to determine boundary values for the north domain boundaries. The first proposition (N1) uses linear interpolation of the four available data points as shown in Fig. 5. The second proposition (N2) adjusts the head variation trend of EB-304 (see Fig. 1 for location) to the head elevations of the four data points as shown in Fig. 5.

Assigning time-varied constant head boundary values could also be uncertain when clusters of observation wells are available and do not show the same head behaviors. Then it is unclear which

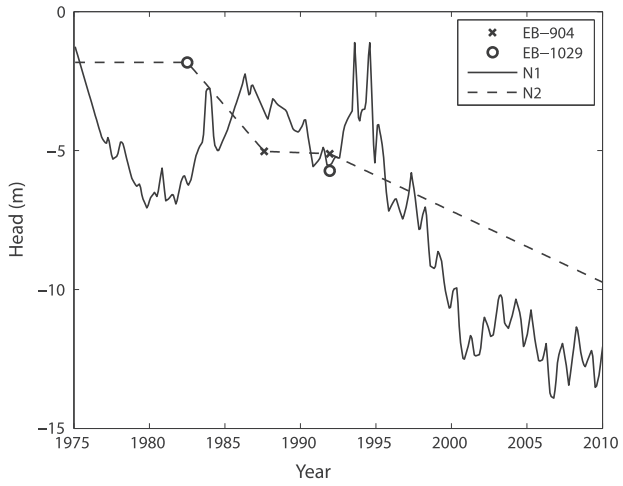

Fig. 5. Two boundary head propositions N1 and N2 for the northern boundaries.

cluster to select to extrapolate to the boundary. This is the case with the eastern boundary condition in the middle domain, in which we categorized two clusters of observation wells to determine the eastern boundary head values. The first proposition (E1) uses the USGS observation wells EB-781, EB-792B, EB-807B and EB-1028. The second proposition (E2) uses the USGS observation wells EB-297 and WBR-106. Note that while the head boundary elevations of the propositions N1 and N2 are fixed, the propositions E1 and E2 have an elevation adjustment factor to be determined by inverse modeling.

The western boundary condition of the middle domain is for an isolated sand unit and is determined by the observation well WBR-102B. The time-varied boundary head values at the south domain are determined using WBR-97B and EB-783A, which have sufficient observation points.

3.5. Groundwater model development and calibration

Given the six aquifer-fault hydrofacies architectures, the two northern boundary condition propositions N1 and N2, and the two eastern boundary condition propositions E1 and E2, through combinatorial design we obtain 24 base models for the hierarchical BMA analysis. We used MODFLOW-2005 (Harbaugh, 2005) to simulate groundwater heads. Each model has 29 layers. Each layer has 12,741 cells and the cell size is 200 m $\times$ 200 m. Detailed pumpage data is available from Louisiana Capital Area Ground Water Conservation Commission.

Each groundwater model is calibrated for six unknown model parameters. The sand hydrofacies has three unknown parameters, the hydraulic conductivity (m/d), specific storage (1/m) and vertical anisotropy ratio. The other three unknown model parameters are hydraulic characteristics (1/d) of the Baton Rouge fault and the Denham Springs-Scotlandville fault, and the elevation adjustment factor (m) for the eastern boundary condition. The hydraulic characteristic of the fault is defined as the hydraulic conductivity of the fault divided by the width of the fault. Flow model calibration is based on 1285 head data between 1975 and 2010 from 17 USGS observation wells (see Fig. 1 for locations). The inverse problem is to minimize the root mean squared error (RMSE) between the simulated and observed heads.

For solving the inverse problem, we used the covariance matrix adaptation evolution strategy (CMA-ES) (Hansen and Ostermeier, 2001; Hansen et al., 2003). CMA-ES is a leading stochastic and derivative-free algorithm for solving continuous optimization problems, and is considered to be based on a statistical principle (Akimoto et al., 2012). The CMA-ES adapts a covariance matrix representing the pair-wise dependency between unknown model parameters, which approximates the inverse of the Hessian matrix up to a certain factor. The candidate solutions, which are vectors of unknown model parameters generated from a multivariate normal distribution, are updated using the covariance matrix and an adaptable step size, which are adapted through two conjugates that implement heuristic control terms. The covariance matrix is empirically calculated in which the covariance matrix adaptation uses information from the current iteration and from the previous search path. Reviewing the CMA-ES algorithm is beyond the scope of this work and reader is referred to Hansen (2006) and Akimoto et al. (2012).

3.6. Within-model variance calculation

In order to calculate the variance term in Eq. (15), the head prediction variance for each model needs to be calculated. For each model we use the maximum likelihood estimates and their covariance matrix to generate 512 samples. A sample is a random vector of the six unknown model parameters chosen from the multivariate normal distribution using the full covariance matrix, and is used to generate one realization of the head prediction. It is noted that CMA-ES is our implementation choice for model calibration and within-model uncertainty quantification, and can be replaced by other methods.

3.7. High performance computing

The model calibration and the Monte Carlo realizations of the 24 models were carried out using SuperMike-II at Louisiana State University, which is a cluster of 146 TFlops peak performance 440 nodes with each node containing two 8-core processors operating at a core frequency of 2.6 GHz. For each of the 24 models, the calibration algorithm requires about 59 ± 16 iterations to reach the stopping criterion. An iteration contains 32 candidate solutions (i.e., groundwater flow model simulations). Thus, using an embarrassingly parallel master–slave technique, each iteration requires two nodes (32 processors) on the SuperMike-II. The mean iteration running time is 1.18 ± 0.28 h. The iteration running time is the maximum of the running times of the candidate solutions in an iteration. Since the candidate solutions do not communicate and accordingly the parallelization overhead is negligible, the model calibration time is the sum of all the iterations run times. The calibration of the 24 groundwater flow models can be done simultaneously and takes around 72 h. Generating the Monte Carlo realizations is more flexible since all the realizations for all the models are independent. Thus, both the calibration and Monte Carlo realizations can be finished for all the models in one week.

4. Results and discussion

4.1. Model calibration and Monte Carlo realizations

Table 2 shows the calibration results for the 24 models. The base models are named according to the hierarchical order of propositions. For example, the base model IZD1N1E1 contains the indicator zonation proposition IZ, the formation dip proposition D1, the northern boundary condition proposition N1 and the eastern boundary condition proposition E1. The best model IKD2N2E1 and the worst model IZD2N1E2 have RMSE of 2.95 m and 4.06 m, respectively. The boundary condition adjustment factor for the eastern boundary for the 24 models have a narrow range of −2.61 m to 2.76 m, where the calibration bound is ±10 m. This

Table 2
Calibration results: boundary condition adjustment factor (BC (m)), hydraulic conductivity (K (m/d)), anisotropic ratio (Kh/kV), specific storage (Ss (1/m)), hydraulic characteristics of the Baton Rouge fault (BR (1/d)), hydraulic characteristics of the Denham Springs-Scotlandville fault (DSS (1/d)), root mean square error (RMSE (m)) of the base models. BMA results: Q, ΔBIC, prior model probability (priorPr) and posterior model probability (postPr) for the base models.

Base model	BC (m)	K (m/d)	Kh/kV	Ss (1/m)	DSS (1/d)	BR (1/d)	RMSE (m)	Q	ΔBIC	PriorPr	PostPr
IZD1N1E1	2.07	164	1.31	2.00E−05	8.41E−06	9.35E−03	3.24	1808	265	1.75E−03	9.82E−05
IZD1N1E2	0.38	170	1.55	2.84E−05	9.77E−05	1.02E−02	4.01	2526	983	1.75E−03	2.41E−09
IZD1N2E1	2.06	165	1.02	2.04E−05	5.64E−06	9.09E−03	3.24	1805	262	1.75E−03	1.02E−04
IZD1N2E2	0.40	170	3.82	2.59E−05	1.07E−04	1.04E−02	4.01	2527	984	1.75E−03	2.39E−09
IZD2N1E1	2.76	161	1.02	1.98E−05	2.32E−06	7.60E−03	3.25	1820	277	2.50E−02	3.32E−04
IZD2N1E2	0.10	170	3.17	2.49E−05	9.99E−05	9.98E−03	4.06	2585	1042	2.50E−02	4.04E−09
IZD2N2E1	2.73	161	1.00	2.04E−05	1.45E−06	7.71E−03	3.25	1821	278	2.50E−02	3.27E−04
IZD2N2E2	0.11	170	1.01	2.56E−05	9.97E−05	9.98E−03	4.05	2579	1036	2.50E−02	4.44E−09
GPD1N1E1	0.67	154	3.37	1.83E−05	5.78E−06	7.41E−03	3.17	1757	214	7.25E−03	2.97E−03
GPD1N1E2	−1.89	161	3.06	2.29E−05	9.61E−05	1.03E−02	3.83	2324	781	7.25E−03	6.83E−07
GPD1N2E1	0.70	155	2.75	1.91E−05	2.20E−06	7.26E−03	3.17	1752	209	7.25E−03	3.20E−03
GPD1N2E2	−2.05	162	1.47	1.94E−05	8.25E−05	9.91E−03	3.83	2333	790	7.25E−03	5.92E−07
GPD2N1E1	2.32	145	1.01	1.93E−05	4.50E−05	4.58E−03	2.98	1571	28	1.06E−01	1.99E−01
GPD2N1E2	−0.48	163	1.07	2.35E−05	9.76E−05	5.52E−03	3.79	2268	726	1.06E−01	6.58E−06
GPD2N2E1	2.48	147	1.23	1.82E−05	2.21E−06	4.46E−03	2.98	1566	23	1.06E−01	2.12E−01
GPD2N2E2	−0.52	163	1.80	2.65E−05	9.73E−05	5.41E−03	3.79	2260	717	1.06E−01	7.44E−06
IKD1N1E1	0.11	149	1.29	1.83E−05	7.23E−06	6.81E−03	3.13	1713	170	7.00E−03	1.67E−03
IKD1N1E2	−2.59	156	1.14	2.21E−05	2.11E−05	9.98E−03	3.72	2208	665	7.00E−03	1.10E−06
IKD1N2E1	0.05	150	1.08	1.94E−05	1.02E−06	6.86E−03	3.12	1712	169	7.00E−03	1.69E−03
IKD1N2E2	−2.61	157	1.05	2.06E−05	1.43E−05	9.99E−03	3.72	2207	664	7.00E−03	1.11E−06
IKD2N1E1	1.29	145	1.01	1.82E−05	1.30E−06	4.16E−03	2.95	1544	1	1.03E−01	2.87E−01
IKD2N1E2	−1.04	157	1.00	2.31E−05	9.92E−05	4.98E−03	3.71	2196	653	1.03E−01	1.87E−05
IKD2N2E1	1.32	145	1.07	1.90E−05	1.11E−06	4.17E−03	2.95	1543	0	1.03E−01	2.91E−01
IKD2N2E2	−1.04	157	1.02	2.20E−05	9.95E−05	4.98E−03	3.71	2191	648	1.03E−01	2.00E−05

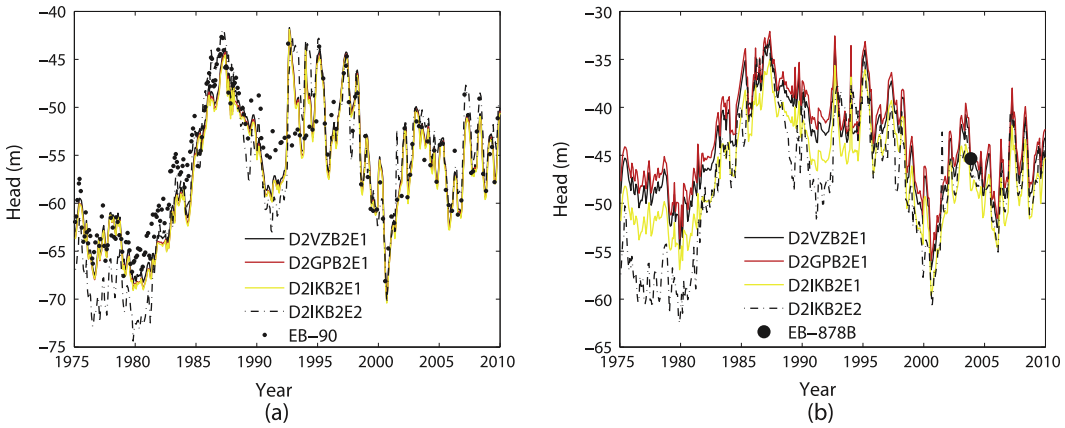

Fig. 6. Head predictions and head observations of four selected base models at observation wells (a) EB-90 and (b) EB-878B.

narrow range for all the models indicates that the prior boundary head elevation of the E1 and E2 was well estimated.

The range of the estimated hydraulic conductivity 145–170 m/d, specific storage 1.82×10^{-5}–2.84×10^{-5} 1/m, and vertical anisotropy 1.00–3.82 is narrow. However, the range of the estimated hydraulic characteristics of the Denham Springs-Scotlandville fault 1.04×10^{-6}–1.07×10^{-4} 1/d and the Baton Rouge fault 4.16×10^{-3}–1.04×10^{-2} 1/d is relatively wide, particularly the hydraulic characteristic of the Denham Springs-Scotlandville fault. The wide range of the hydraulic characteristics of the Denham Springs-Scotlandville fault might be due to the lack of data for the northern boundary condition.

Fig. 6 looks at the calibration results in more details by showing the maximum likelihood head prediction and the prediction variance for four selected models at two observation wells. The four models were selected to show the difference between propositions IZ, GP and IK and the difference between propositions E1 and E2. First, we selected observation well EB-90 because it is a long-term USGS observation well. The RMSE of EB-90 for the 24 models is (2.58 ± 0.52) m with the minimum RMSE = 1.82 m for the base model D1VZB2E1 and maximum RMSE = 3.20 m for the base model D2IKB1E2. From Fig. 6(a), it is clear that the propositions E1 and E2 produce relatively different head predictions. Second, we selected the observation well EB-878B since it has only one head observation and thus the model prediction at this location is not conditioned on the head observations. As shown in Fig. 6(b), the IK and GP produce similar predictions, which are different from the IZ prediction. This is not surprising because the hydrofacies models D2GP and D2IK have similar posterior model probabilities (see Table 1) and similar hydrofacies architectures (see Figs. 3–5). Comparing Fig. 6(a) and (b) shows that the absence of conditioning on head observations results in higher between-model prediction variance for observation well EB-878B.

4.2. BIC calculation

We use the calibration and the Monte Carlo results to calculate ΔBIC. The larger the variance windows, the more models are selected. To allow for more model selection, we use $s_1 = 5\%$ significance level and variance width $s_2 = 4\sigma_D$, where $\sigma_D = \sqrt{2n}$. Thus, the scaling factor in Eq. (16) is $\alpha = 1.06/\sqrt{n}$ (Tsai and Li, 2008a).

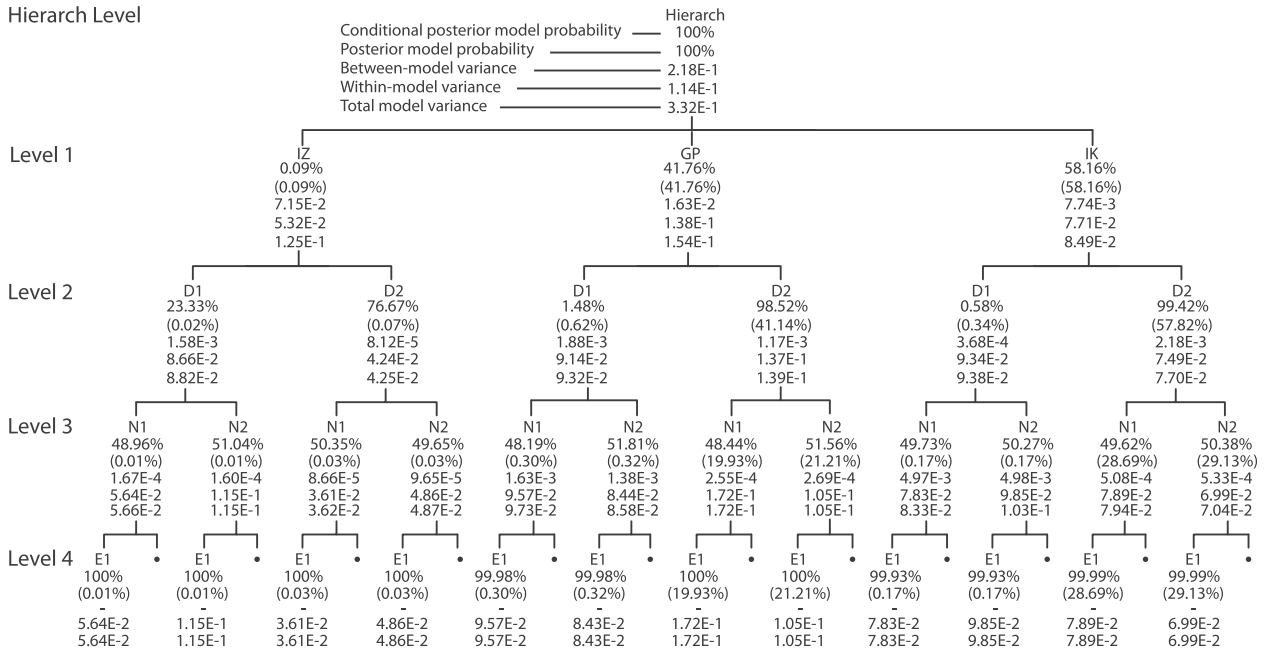

Fig. 7. BMA tree of the posterior model probabilities and prediction variances of the four uncertain model components: hydrofacies architecture reconstruction method (IZ, GP and IK), formation dip (D1 and D2), northern boundary conditions (N1 and N2) and eastern boundary condition in the middle domain (E1 and E2). Models that have posterior model probabilities less than 0.01% are not shown.

Table 3
Between-model variance (BMV), within-model variance (WMV) and total model variance for each source of uncertainty.

Level	Sources of model uncertainty	BMV	WMV	TMV
4 (base level)	Model parameters	–	1.00E−01	1.00E−01
3	Eastern boundary condition	4.35E−04	1.00E−01	1.01E−01
2	Northern boundary condition	1.75E−03	1.01E−01	1.03E−01
1	Formation dip	1.14E−02	1.03E−01	1.14E−01
Hierarch	Hydrofacies reconstruction method	2.18E−01	1.14E−01	3.32E−01

The BIC results are given in Table 2. Substituting Q as calculated from Eq. (14) into Eq. (13) we obtain the BIC for all base models. The best base model IKD2N2E1 has $BIC_{min} = 3948$ from which we obtain the ΔBIC for all base models. Finally, we calculate the posterior model probabilities for all base models as shown in Table 2 by substituting ΔBIC and the prior model probabilities into Eq. (16). The prior model probabilities are obtained from the normalized model probabilities of the hydrofacies architecture models (see Table 1).

4.3. Model propositions evaluation

The first feature of hierarchical BMA analysis is that model dissection allows the evaluation of candidate model propositions of each uncertain model component. Although this can be directly inferred from posterior model probabilities of the base models (e.g., Foglia et al., 2013), yet the BMA tree of posterior model probability and conditional posterior model probability, as shown in Fig. 7, provides more detailed information.

Starting from the conditional posterior model probabilities at the base level of the BMA tree, it is clear that the proposition E1 is consistently supported better by the data than the proposition E2. Moving to level 3, the similar conditional posterior model probabilities of the propositions N1 and N2 indicate that they are both equally supported by the data. At level 2, the geological formation dip propositions D1 and D2 have different performance

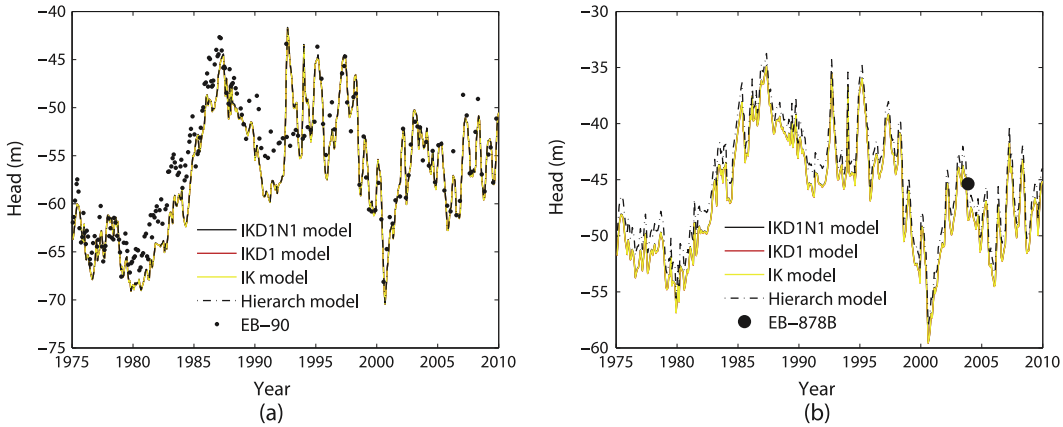

Fig. 8. BMA head predictions for the best branch of the BMA tree for observation wells (a) EB-90 and (b) EB-878B.

under different hydrofacies architecture reconstruction propositions. Under the propositions IK and GP, proposition D2 is considerably supported better by the data than propositions D1, yet under the proposition IZ the proposition D1 has some considerable weight. This indicates that both D1 and D2 are relevant propositions. However, proposition IZ is out. Finally at level 1, the proposition IK is generally supported better by the data than the proposition GP. Note that since boundaries between sand and clay units are neither smooth nor blocky as a result of fluvial depositional processes, thus the GP is a better geological conceptual model since the GP estimation is neither as sharp-edged as the IZ estimation nor as smooth as the IK estimation. In addition, the GP models have slightly higher prior probabilities (see Table 1). However, the results show that the best hydrofacies architecture model (i.e., GPD2) does not necessarily lead to the best groundwater flow model (i.e., IKD2N2E1).

4.4. Uncertainty propagation and prioritization

The second feature is that the hierarchical BMA facilitates the depiction of the uncertainty propagation. For the illustration purpose, we calculate averaged prediction variances at all the head observation locations for all the time steps as shown in Fig. 7. Hierarchical BMA separates the uncertainty contribution of each source of uncertainty through providing the between-model variance for each source of uncertainty. As shown by Eq. (7), the between-model variance at any given level of the BMA tree is independent of between-model variances at other levels, which is illustrated in Fig. 7. Alternatively, the within-model variance is dependent on the total model variances at its subordinate level. The total model variance is the summation of the between-model variance and the within-model variance. The total model variance at the hierarch level is the overall model variance.

Level 4 of the BMA tree in Fig. 7, shows the within-model variance of the base models as calculated from the Monte Carlo realizations. There is no between-model variance for the base models since the base level consists of the simulation models, not the BMA models. Thus, the within-model variance is the total model variance at the base level.

Tracing the uncertainty propagation starts with the BMA models at level 3 of the BMA tree in Fig. 7. Although propositions E1 and E2 produce very different estimation as previously shown in Fig. 6, yet since posterior model probability of the proposition E1 is substantially higher then proposition E2, the between-model variance is negligible. At level 2, even though the propositions N1 and N2 have similar conditional posterior model probabilities, still the between-model variance is negligible due to the similar prediction of propositions N1 and N2. Level 1 shows that the propositions D1 and D2 introduce small within-model variance for the IZ and GP branches. The hierarch level shows that the hydrofacies architecture reconstruction methods introduce considerable within-model variance.

The total model variance for each uncertain model component depicts the uncertainty propagation resulting from adding up different sources of uncertainty. The BMA tree in Fig. 7 shows that generally the total variance increases by adding more uncertain components, yet this is not necessarily the case. For example, the IZD1N2 model has higher posterior model probability and higher within-model variance than IZD1N2 model, yet its superior IZD1 model has lower total model variance because the between-model variance of IZD1N2 model is lower.

The third feature of the hierarchical BMA analysis is that, while the collection BMA only provides one overall between-model variance of all base models, the segregation of the between-model variance at different levels permits the prioritization of the relative impact of each uncertain model component on the overall model uncertainty. Given the between-model variances in the BMA tree in Fig. 7, the hydrofacies architectures reconstruction method has the most contribution to the overall model uncertainty. The formation dip, northern boundary condition and eastern boundary condition have minor contributions. This can also be seen from Table 3 that lists the variance contributions of individual sources of uncertainty to the total variance using Eq. (7). The hydrofacies reconstruction method contributes the most variance, followed by model parameters.

4.5. Temporal and spatial distribution of head prediction and variance

We further illustrate these three features by looking at the temporal and spatial distribution of the groundwater head prediction and variance of the BMA models of the best branch, which are

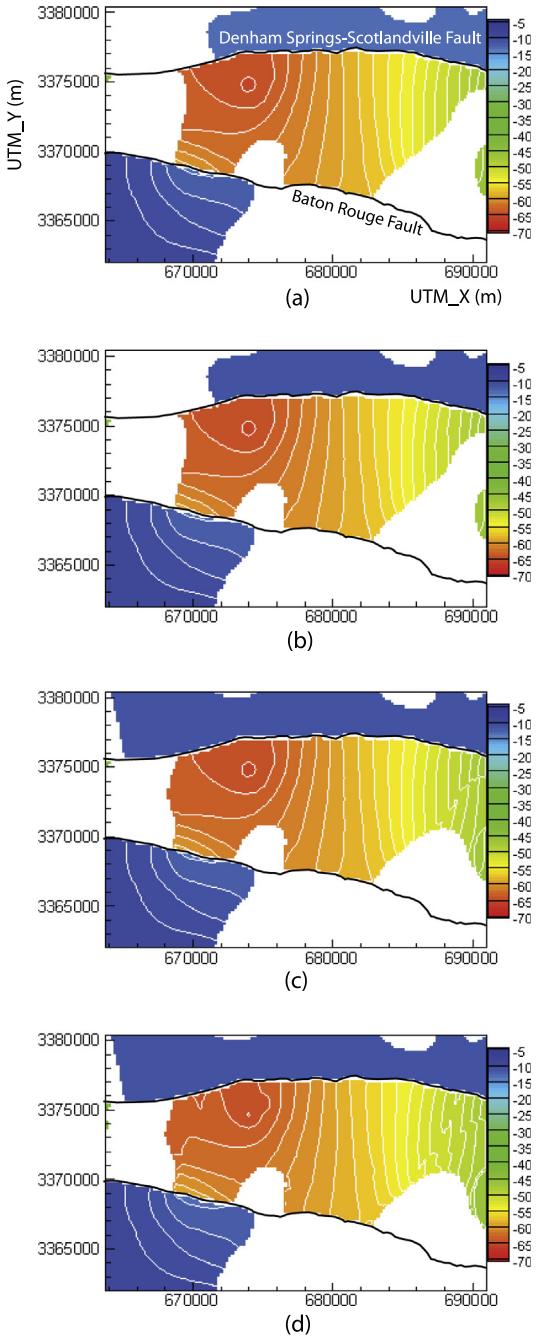

Fig. 9. BMA head predictions (meters) for the selected layer in Fig. 4 for the best branch: (a) IKD1N1 model, (b) IKD1 model, (c) IK model and (d) hierarch model.

the hierarch, IK, IKD2 and IKD2N1 models. The BMA predictions over the simulation period from 1975 to 2010 for observation wells EB-90 and EB-878B are shown in Fig. 8. Fig. 8(a) shows very similar head predictions at different levels for EB-90 that has many observations. However, due to only one observation datum at EB-878B, the head prediction at this well changes at the hierarch level as shown in Fig. 8(b). Since conditioning on head observations reduces the between-model variance, for EB-90 the within-model variance is similar to the total model variance. This is not the case for EB-878B due to the lack of data.

We show the spatial distribution of head prediction and variance for the selected layer that is shown in Fig. 4 for the last simulation period December 2010. Given the six hydrofacies models, a cell at a given location could be a sand cell for all the models or for some models. In this study, only the head predictions at the sand cells are averaged by BMA. The probability of whether a cell is sand or clay is given in Fig. 4(h). The head predictions in the middle domain in Fig. 9(a)–(c) are similar due to small between-model prediction variance. This occurs by one of the following two reasons. (1) A single model with a dominant posterior model probability results in small between-model prediction variance even though all models have very different head predictions. (2) Similar posterior model probabilities with similar head predictions also result in small between-model prediction variance. The head predictions in the middle domain in Fig. 9(a)–(c) are different from Fig. 9(d). This confirms the previous remark that different hydrofacies reconstruction methods result in significantly different head predictions. This occurs because similar posterior model probabilities with different head predictions result in large between-model prediction variance.

The between-model prediction variance as depicted in Fig. 10(a)–(d) illustrates the contribution of each uncertain model component to the overall model variance. The variance contributions from the eastern boundary condition and formation dip as shown in Fig. 10(a) and (c), respectively, are negligible. The variance contribution from the northern boundary condition is large in the north domain and negligible in the middle and south

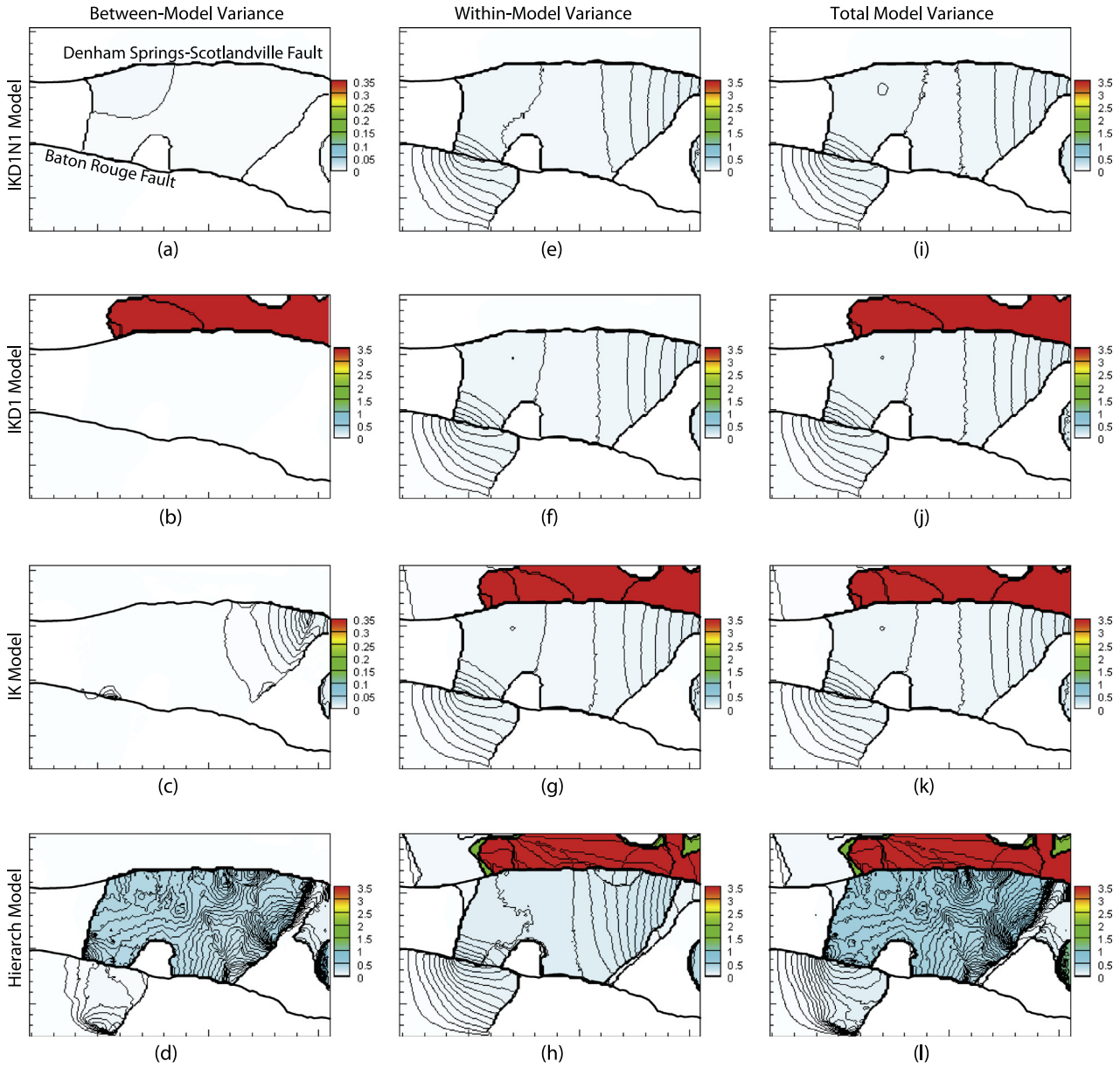

Fig. 10. Between-model variance (a)–(d), within-model variance (e)–(h) and total model variance (i)–(l) for the best branch that contains IKD1N1 model, IKD1 model, IK model and the hierarch model for the selected layer in Fig. 4 and the last time step.

domains, which is due to the very low permeability of the Denham Springs-Scotlandville fault. The hydrofacies reconstruction method has the most variance contribution in middle and south domains as shown in Fig. 10(d).

The within-model variance and total model variance as shown in Fig. 10(e)–(h) and (i)–(l), respectively, show the construction of uncertainty. Fig. 10(e) and (i) at level 3 are similar, and Fig. 10(g) and (k) at level 1 are similar because the eastern boundary condition and formation dip result in the small between-model variances, respectively. Alternatively, the high total variance in Fig. 10(j) in the north domain is due to high between-model variance from the northern boundary conditions at level 2. By adding more uncertain model components, Fig. 10(i)–(l) introduce more uncertain regions. However, the variance magnitude can decrease as shown in the north domain in Fig. 10(k) and (l).

4.6. Knowledge update

Based on what was learned from the previous analysis, a key component of constructive epistemic modeling is knowledge update. The hierarchical BMA allows for the segregation of uncertain model components and thus provides a framework that facilities knowledge update. One mean of knowledge update is to oust a level of uncertainty after having sufficient evidences from the posterior model probabilities, model solution and expert knowledge that one model proposition is more robust than other model propositions.

The study shows that the proposition E1 consistently has substantially higher posterior model probabilities than E2 under all superior propositions. In addition, looking more closely at the model geological structure shows that the observation wells that were used to develop the proposition E1 are directly connected to the eastern boundary condition. Thus, this level of uncertainty can be dropped.

4.7. Challenges for Bayesian multimodel analysis

Several theoretical and practical challenges pertain to Bayesian multimodel analysis. Quantifying the posterior model probabilities still requires extensive treatment. First, a major practical and theoretical concern is the ability to infer the quantities of interest from the available data in order to correctly discriminate between candidate propositions (Beven, 2006; Renard et al., 2010; Clark et al., 2011; Gupta et al., 2012). This is mainly because it involves the inherent challenges of non-identifiability or ill-posed inference, which is the inability to infer some or all quantities of interest from the available data (Renard et al., 2010). Although in theory the collection BMA and the hierarchical BMA have no problem with ill-posed inference, yet in practice ill-posed inference has an impact on the posterior model probabilities. Second, adding new model parameters or state variables in the calibration process would result in new posterior model probabilities. The variability of the posterior model probabilities when other state variables are predicted or new model parameters are added is a practical concern in Bayesian multimodel analysis that requires further development. Third, more critical issue is obviously the selection of statistical functions (e.g., Gaussian or non-Gaussian) and statistical inference methods (e.g., Markov Chain Monte Carlo methods, information-theoretic criteria, etc.) to estimate the posterior model probabilities.

The evaluation of the adequacy of the candidate model propositions with respect to model simulation and prediction is not just limited to the posterior model probabilities. More generally "a general hierarchical system of metrics that covers the dimensions of space, time, state/process, and application" as suggested by Gupta et al. (2012) is needed. Actually, as Clark et al. (2011) note that "model comparison studies are still a long way from reliably elucidating the appropriateness of different model representations".

The aforementioned concerns are mainly epistemic and result in epistemic uncertainty. As suggested by Beven and Young (2013), epistemic uncertainty could be reduced in principle by having more or better measurements or by new knowledge (science). It is important to stress that the term knowledge does not only refers to our knowledge about the different propositions of the model data, structure, parameters and processes, but also refers to the statistical methods that facilitate the evaluation of these different propositions.

5. Conclusions

Using the hierarchical Bayesian model averaging, we have studied the geological and flow uncertainties of the "2000-foot" of the Baton Rouge aquifer fault system. We consider the geological uncertainty arising from different conceptualizations of hydrofacies architectures and flow uncertainty arising from different conceptualizations of the boundary conditions. For geological structure uncertainty, we consider three indicator methods and two formation dips to reconstruct six hydrofacies architecture models of the aquifer-fault system. The probabilities of the six hydrofacies models represent the prior model probabilities of the six aquifer structures of the groundwater models. For flow uncertainty, we consider two uncertain boundary conditions, each having two candidate propositions. This results in 24 groundwater models that were calibrated using head data. The posterior model probabilities for these models are estimated using Bayesian information criterion.

The results show that the hierarchical Bayesian model averaging is a learning tool about model construction and model uncertainty. First, through uncertainty segregation, the hierarchical BMA facilitates prioritizing the uncertain model components. In the numerical study, the analysis shows that uncertainty arising from boundary conditions is minor in comparison to geological structure uncertainty. Second, the hierarchical BMA permits comparative evaluation of candidate model propositions. With respect to hydrofacies architecture reconstruction method, the indicator kriging proposition is supported better by the data than generalized parameterization proposition, indicating that robust hydrofacies architecture does not necessarily lead to the best groundwater flow model. Third, hierarchical BMA depicts the change of the BMA prediction and variance due to the addition of each source of uncertainty. Results shows that head predictions at observation wells are very similar when long-term head observation data are available from the wells. On the other hand, head predictions at observation wells that have limited observation data are considerably different at different levels in the BMA tree. The variance propagation along a branch of the BMA tree depicts model structure uncertainty increases in both the magnitude and regions of uncertainty. Finally, as a constructive epistemic framework, our current understanding about the "2000-foot" sand flow model is subject to revision shall new knowledge become available.

We discussed the term constructive epistemic modeling. Constructive means that our perception of reality is being constructed through a development path. Although this development path under hierarchical BMA can be computational expensive since combinatorial design results in factorial increase in the number of base models, yet such computational issues can be resolved with high performance computing as this study shows that the model calibration and the Monte Carlo realizations run time of the 24 models is about a week. Moreover, based on the analyst's knowledge, not all branches in the BMA tree need to be considered. In addition, this development path does not only aim at just accumulating new pieces of information, but also aims at ousting unsound propositions. For example, this numerical study shows that one

specific proposition about the eastern boundary condition is substantially supported better by the data, thus this level of uncertainty can be dropped.

From a constructive epistemic modeling prospective, uncertainty would mean the uncertainty of our current state of knowledge. The explicit differentiation between within-model variance and between-model variance through the hierarchical BMA has an important implication. Given data and a model structure, the within-model variance is mainly a measure of data uncertainty such as kriging uncertainty and the capability of the calibration algorithm to reach a precise solution in a rugged and noisy search landscape as affected by data error. Yet more importantly is the between-model variance, which is a measure of the uncertainty resulting from candidate knowledge propositions about the natural system. Although the between-model variance contribution to the overall uncertainty is additive at each level, the between-model variance for a given uncertain model component can increase or decrease by testing more propositions or by adding new uncertain model components. Thus, through the learning path, our uncertainty about the model solution can increase or decrease.

Acknowledgements

This study was supported in part by the National Science Foundation under Grant No. 1045064 and Grant/Cooperative Agreement Number G10AP00136 from the United States Geological Survey. The authors gratefully acknowledge two reviewers, Ming Ye and Wolfgang Nowak, for their constructive comments.

References

Agarwal, H., Renaud, J.E., Preston, E.L., Padmanabhan, D., 2004. Uncertainty quantification using evidence theory in multidisciplinary design optimization. Reliab. Eng. Syst. Saf. 85 (1–3), 281–294.
Akimoto, Y., Nagata, Y., Ono, I., Kobayashi, S., 2012. Theoretical foundation for CMA-ES from information geometry perspective. Algorithmica 64 (4), 698–716.
Baudrit, C., Guyonnet, D., Dubois, D., 2007. Joint propagation of variability and imprecision in assessing the risk of groundwater contamination. J. Contam. Hydrol. 93 (1–4), 72–84.
Beven, K., 2006. On undermining the science? Hydrol. Process. 20 (14), 3141–3146.
Beven, K., Young, P., 2013. A guide to good practice in modeling semantics for authors and referees. Water Resour. Res. 49 (8), 5092–5098.
Cardiff, M., Kitanidis, P.K., 2009. Bayesian inversion for facies detection: an extensible level set framework. Water Resour. Res. 45 (10), W10416.
Carrera, J., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms. Water Resour. Res. 22 (2), 211–227.
Chamberlain, E.L., 2012. Depositional environments of upper miocene through pleistocene siliciclastic sediments, baton rouge aquifer system, Southeastern Louisiana, Louisiana State University, Department of Geology and Geophysics, M.Sc. Thesis, p. 66.
Christakos, G., 2004. The cognitive basis of physical modelling. In: Miller, T.M., Farthing, M.W., Gray, W.G., Pinder, G.F. (Eds.), Computational Methods in Water Resources, Developments in Water Science 55, vol. 1. Elsevier, Amsterdam, The Netherlands, pp. 661–669.
Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour. Res. 47 (9), W09301.
Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. Water Resour. Res. 47 (12), W12534.
Draper, D., 1995. Assessment and propagation of model uncertainty. J. R. Stat. Soc. Ser. B (Methodol.) 57 (1), 45–97.
Ellison, A.M., 2004. Bayesian inference in ecology. Ecol. Lett. 7 (6), 509–520.
Elshall, A.S., Tsai, F.T.-C., Hanor, J.S., 2013. Indicator geostatistics for reconstructing Baton Rouge aquifer-fault hydrostratigraphy, Louisiana, USA. Hydrogeol. J. 21 (8), 1731–1747.
Foglia, L., Mehl, S.W., Hill, M.C., Burlando, P., 2013. Evaluating model structure adequacy: the case of the Maggia Valley groundwater system, southern Switzerland. Water Resour. Res. 49 (1), 260–282.
Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G., Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. Water Resour. Res. 48 (8), W08301.
Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. Evol. Comput. 9 (2), 159–195.
Hansen, N., Müller, S.D., Koumoutsakos, P., 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evol. Comput. 11 (1), 1–18.
Hansen, N., 2006. The CMA evolution strategy: a comparing review. In: Lozano, J.A., Larrañga, P., Inza, I., Bengoetxea, E. (Eds.), Towards A New Evolutionary Computation. Springer, Berlin, Germany, pp. 75–102.
Harbaugh, A.W. 2005. MODFLOW-2005, The US geological survey modular groundwater model—the ground-water flow process: US geological survey techniques and methods 6–A16, US geological survey, Reston, Virginia.
He, L., Huang, G.H., Lu, H.W., 2008. A simulation-based fuzzy chance-constrained programming model for optimal groundwater remediation under uncertainty. Adv. Water Resour. 31 (12), 1622–1635.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14 (4), 382–401.
Jaynes, E.T., 1990. Probability theory as logic. In: Fougère, P.F. (Ed.), Maximum Entropy and Bayesian Methods. Kluwer Academic Publishers, Dartmouth.
Jaynes, E.T., 2003. Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK.
Johnson, N.M., 1995. Characterization of alluvial hydrostratigraphy with indicator semivariograms. Water Resour. Res. 31 (12), 3217–3227.
Johnson, N.M., Dreiss, S.J., 1989. Hydrostratigraphic interpretation using indicator geostatistics. Water Resour. Res. 25 (12), 2501–2510.
Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour. Res. 22 (4), 499–507.
Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. Water Resour. Res. 32 (9), 2617–2658.
Li, L., Zhou, H., Franssen, H.J.H., Gomez-Hernandez, J.J., 2012. Groundwater flow inverse modeling in non-MultiGaussian media: performance assessment of the normal-score ensemble kalman filter. Hydrol. Earth Syst. Sci. 16 (2), 573–590.
Li, X., Tsai, F.T.-C., 2009. Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod. Water Resour. Res. 45 (9), W09403.
McCulloh, R.P., Heinrich, P.V., 2012. Surface faults of the South Louisiana Growth-Fault Province. In: Cox, R.T., Tuttle, M., Boyd, O., Locat, J. (Eds.), Recent Advances in North American Paleoseismology and Neotectonics East of The Rockies and Use of The Data in Risk Assessment and Policy. Geological Society of America, Boulder, Colorado.
Meyer, R.R., Turcan Jr., A.N., 1955. Geology and ground-water resources of the Baton Rouge area, Louisiana, US Geological Survey, Water Supply Paper 1296, p. 137.
Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stoch. Environ. Res. Risk Assess. 17 (5), 291–305.
Neuman, S.P., Wierenga, P.J., 2003. A comprehensive strategy of hydrogeologic modelling and uncertainty analysis for nuclear facilities and sites, NUREG/CR-6805. US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Washington, DC.
Poeter, E., Anderson, D., 2005. Multimodel ranking and inference in ground water modeling. Ground Water 43 (4), 597–605.
Popper, K., 1959. The propensity interpretation of probability. Br. J. Philos. Sci. 10 (37), 25–42.
Popper, K.R., 1990. A World of Propensities. Thoemmes, Bristol, UK.
Primiero, G., 2008. Information and knowledge: a constructive type-theoretical approach. In: Rahman, S., Symons, J. (Eds.), Logic, Epistemology, and the Unity of Science, vol. 10. Springer, Dordrecht, The Netherlands.
Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 25, 111–163.
Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process: a framework and guidance. Environ. Model. Softw. 22 (11), 1543–1556.
Refsgaard, J.C., van der Sluijs, J.P., Brown, J., van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. Adv. Water Resour. 29 (11), 1586–1597.
Refsgaard, J.C., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A.L., Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Adv. Water Resour. 36 (2), 36–50.
Riegler, A., 2012. Constructivism. In: L'Abate, L. (Ed.), Paradigms, in Theory Construction. Springer, New York, pp. 235–256.
Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. 46 (5), W05521.
Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. Water Resour. Res. 44 (12), W12418.
Rojas, R., Feyen, L., Batelaan, O., Dassargues, A., 2010. On the value of conditioning data to reduce conceptual model uncertainty in groundwater modeling. Water Resour. Res. 46 (8), W08520.
Rubin, Y., 2003. Applied Stochastic Hydrogeology. Oxford University Press, New York.
Sakaki, T., Frippiat, C.C., Komatsu, M., Illangasekare, T.H., 2009. On the value of lithofacies data for improving groundwater flow model accuracy in a three-dimensional laboratory-scale synthetic aquifer. Water Resour. Res. 45 (11), W11404.
Schervish, M.J., 1996. P-values: What they are and what they are not. Am. Stat. 50 (3), 203–206.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.

Seifert, D., Sonnenborg, T.O., Refsgaard, J.C., Højberg, A.L., Troldborg, L., 2012. Assessment of hydrological model predictive ability given multiple conceptual geological models. Water Resour. Res. 48 (6), W06503.

Singha, K., Hyndman, D.W., Day-Lewis, F.D., 2007. Introduction. In: Hyndman, D.W., Day-Lewis, F.D., Singha, K. (Eds.), Subsurface Hydrology: Data Integration for Properties and Processes, Geophysical Monograph Series, vol. 171. American Geophysical Union, Washington, DC, pp. 1–5.

Singh, A., Mishra, S., Ruskauff, G., 2010. Model averaging techniques for quantifying conceptual model uncertainty. Ground Water 48 (5), 701–715.

Tomaszewski, D.J. 1996. Distribution and movement of saltwater in aquifers in the Baton Rouge area, Louisiana, 1990–92, Louisiana Department of Transportation and Development, Water Resources Technical Report No. 59, Baton Rouge, Louisiana.

Trevisani, S., Fabbri, P., 2010. Geostatistical modeling of a heterogeneous site bordering the Venice Lagoon, Italy. Ground Water 48 (4), 614–623.

Troldborg, M., Nowak, W., Tuxen, N., Bjerg, P.L., Helmig, R., Binning, P.J., 2010. Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework. Water Resour. Res. 46 (12), W12552.

Tsai, F.T.-C., 2006. Enhancing random heterogeneity representation by mixing the kriging method with the zonation structure. Water Resour. Res. 42 (8), W08428.

Tsai, F.T.-C., 2009. Indicator generalized parameterization for interpolation point selection in groundwater inverse modeling. J. Hydrol. Eng. 14 (3), 233–242.

Tsai, F.T.C., 2010. Bayesian model averaging assessment on groundwater management under model structure uncertainty. Stoch. Environ. Res. Risk Assess. 24 (6), 845–861.

Tsai, F.T.-C., Elshall, A.S., 2013. Hierarchical Bayesian model averaging for hydrostratigraphic modeling: uncertainty segregation and comparative evaluation. Water Resour. Res. 49 (9), 5520–5536.

Tsai, F.T.-C., Sun, N.-Z., Yeh, W.W.-G., 2003. A combinatorial optimization scheme for parameter structure identification in ground water modeling. Ground Water 41 (2), 156–169.

Tsai, F.T.-C., Yeh, W.W.-G., 2004. Characterization and identification of aquifer heterogeneity with generalized parameterization and Bayesian estimation. Water Resour. Res. 40 (10), W10102.

Tsai, F.T.-C., Li, X., 2008a. Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window. Water Resour. Res. 44 (9), W09434.

Tsai, F.T.-C., Li, X., 2008b. Multiple Parameterization for hydraulic conductivity identification. Ground Water 46 (6), 851–864.

Usunoff, E., Carrera, J., Mousavi, S.F., 1992. An approach to the design of experiments for discriminating among alternative conceptual models. Adv. Water Resour. 15 (3), 199–214.

von Mises, R., 1964. Mathematical Theory of Probability and Statistics. Academic Press, New York.

Williamson, J., 2005. Bayesian Nets and Causality: Philosophical and Computational Foundations. Oxford University Press, Oxford, UK.

Wöhling, T., Vrugt, J.A., 2008. Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models. Water Resour. Res. 44 (12), W12432.

Ye, M., Neuman, S.P., Meyer, P.D., 2004. Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. Water Resour. Res. 40 (5), W05113.

Ye, M., Pohlmann, K.F., Chapman, J.B., Pohll, G.M., Reeves, D.M., 2010. A model-averaging method for assessing groundwater conceptual model uncertainty. Ground Water 48 (5), 716–728.

Yeh, W.W.-G., 1986. Review of parameter identification procedures in groundwater hydrology: the inverse problem. Water Resour. Res. 22 (2), 95–108.