

## Hierarchical Bayesian model averaging for hydrostratigraphic modeling: Uncertainty segregation and comparative evaluation

Frank T.-C. Tsai<sup>1</sup> and Ahmed S. Elshall<sup>1</sup>

Received 21 December 2012; revised 30 June 2013; accepted 15 July 2013; published 9 September 2013.

[1] Analysts are often faced with competing propositions for each uncertain model component. How can we judge that we select a correct proposition(s) for an uncertain model component out of numerous possible propositions? We introduce the hierarchical Bayesian model averaging (HBMA) method as a multimodel framework for uncertainty analysis. The HBMA allows for segregating, prioritizing, and evaluating different sources of uncertainty and their corresponding competing propositions through a hierarchy of BMA models that forms a BMA tree. We apply the HBMA to conduct uncertainty analysis on the reconstructed hydrostratigraphic architectures of the Baton Rouge aquifer-fault system, Louisiana. Due to uncertainty in model data, structure, and parameters, multiple possible hydrostratigraphic models are produced and calibrated as base models. The study considers four sources of uncertainty. With respect to data uncertainty, the study considers two calibration data sets. With respect to model structure, the study considers three different variogram models, two geological stationarity assumptions and two fault conceptualizations. The base models are produced following a combinatorial design to allow for uncertainty segregation. Thus, these four uncertain model components with their corresponding competing model propositions result in 24 base models. The results show that the systematic dissection of the uncertain model components along with their corresponding competing propositions allows for detecting the robust model propositions and the major sources of uncertainty.

**Citation:** Tsai, F. T.-C., and A. S. Elshall (2013), Hierarchical Bayesian model averaging for hydrostratigraphic modeling: Uncertainty segregation and comparative evaluation, *Water Resour. Res.*, 49, 5520–5536, doi:10.1002/wrcr.20428.

### 1. Introduction

[2] When developing a conceptual model to represent a subsurface formation, uncertainties in model data, structure, and parameters always exist. To accommodate for different sources of uncertainty, strategies as model selection, model elimination, model reduction, model discrimination, and model combination are commonly used to reach a robust model, using single-model approaches [Cardiff and Kitanidis, 2009; Demissie et al., 2009; Engdahl et al., 2010; Feyen and Caers, 2006; Kitanidis, 1986; Gagani and Smith, 2001, 2006, 2008; Irving and Singha, 2010; Nowak et al., 2010; Wingle and Poeter, 1993] or multimodel approaches [Doherty and Christensen, 2011; Li and Tsai, 2009; Morales-Casique et al., 2010; Neuman, 2003; Refsgaard et al., 2006; Rojas et al., 2008, 2009, 2010a, 2010b, 2010c; Singh et al., 2010; Trolldborg et al., 2010; Tsai and Li, 2008a, 2008b; Tsai, 2010; Ye et al., 2004, 2005; Wöhling and Vrugt, 2008].

[3] Although single-model approach is commonly used for model prediction and uncertainty assessment of hydrologic systems, yet it has several flaws. *Beven and Binley* [1992] and *Beven* [1993] bring the concept of equifinality by pointing to model nonuniqueness of catchment models, which is the possibility that the same final solution can be obtained by many potential model propositions. This concept as coined by *von Bertalanffy* [1968] means that unlike a closed system, which final state is unequivocally determined by the initial conditions, the final state of an open system may be reached from different initial conditions and in different ways. The problem of model nonuniqueness is salient to almost any field-scale hydrogeological model due to uncertainty about data, model structure, and model parameters. Thus, a single model may result in failing to accept a true model or failing to reject a false model [Neuman and Wierenga, 2003; Neuman, 2003]. In addition, even if a single model can still explicitly segregate and quantify different sources of uncertainty, *Neuman* [2003] points out to an important observation that adopting one model can lead to statistical bias and underestimation of uncertainty. The hierarchical treatment in this study clearly illustrates this point.

[4] Multimodel approach aims at overcoming the aforementioned shortcomings of the single-model approach by utilizing competing conceptual models that adequately fit the data. Multimodel methods aim at ranking or averaging the considered models through their posterior model

<sup>1</sup>Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, Louisiana, USA.

Corresponding author: F. T.-C. Tsai, Department of Civil and Environmental Engineering, Louisiana State University, 3418G Patrick F. Taylor Hall, Baton Rouge, LA 70803, USA. (ftsai@lsu.edu)

probabilities. The most general multimodel method is the generalized likelihood uncertainty estimation (GLUE) [Beven and Binley, 1992], which is based on the equifinality [Beven, 1993, 2005]. In the first step, different models are generated by Monte Carlo simulation and are behavioral according to a user-defined threshold based on their residual errors. In the second step, the posterior model probability for each of accepted models is calculated based on observation data for a given likelihood function.

[5] Variant GLUE methods can be developed by modifying the first step of model generation and acceptance. For example, to move from equifinality to optimality, Mugunthan and Shoemaker [2006] show that calibration performs better than GLUE both in terms of identifying more behavioral samples for a given threshold and in matching the output. However, this is a debatable point. For example, Rojas *et al.* [2008] remarked that by including a calibration step in multimodel approaches, errors in the conceptual models will be compensated by biased parameter estimates during the calibration and the calibration result will be at the risk of being biased toward unobserved variables in the model [Refsgaard *et al.*, 2006]. This study proposes a hierarchical Bayesian averaging approach to address this concern by explicitly segregating different sources of uncertainty.

[6] Variant GLUE methods can also be developed by modifying the second step by using different likelihood functions for model averaging. Formal GLUE [Beven and Binley, 1992] uses inverse weighted variance likelihood function, but the method is flexible allowing for diverse statistical likelihood functions such as exponential function [Beven, 2000] or even possibilistic functions [Jacquin and Shamseldin, 2007]. Exponential and inverse weighted variance likelihood functions do not account for model complexity and number of data points and may lack statistical bases [Singh *et al.*, 2010]. Rojas *et al.* [2008, 2010a, 2010b, 2010c] introduce Bayesian model averaging (BMA) in combination with GLUE to maintain equifinality. Although using BMA is statistically rigorous, yet a typical problem with BMA is that it tends to favor only few best models [Neuman, 2003; Troldborg *et al.*, 2010]. For example, several studies [Rojas *et al.*, 2010c; Singh *et al.*, 2010; Ye *et al.*, 2010b] show that model averaging under formal BMA criteria (AIC, AICc, BIC, and KIC) tends to eliminate most of the alternative models, which may underestimate prediction uncertainty and bias the predictions, while GLUE probabilities are more evenly distributed across all models resulting superior prediction. To maintain the use of statistically meaningful functions, while avoiding underestimating uncertainty, Tsai and Li [2008a, 2008b] propose a variance window to allow selection of more models, but may simultaneously enlarge the magnitude of uncertainty, while satisfying the constraints imposed by the background knowledge.

[7] All the previously cited studies are collection multimodel methods, in which all models are at one level. Wagnier and Gupta [2005] remark that an uncertainty assessment framework should be able to account for the level of contribution of the different sources of uncertainty to the overall uncertainty. In the groundwater area, to advance beyond collection multimodel methods, Li and Tsai [2009] and Tsai [2010] present a BMA approach that can separate two sources of uncertainty, which arise from different conceptual models and different parameter esti-

mation methods. These were the first two studies to extend the collection BMA formulation of Hoeting *et al.* [1999] to two levels. The current study generalizes the work of Li and Tsai [2009] and Tsai [2010] to a fully hierarchical BMA method. To our knowledge, this is the first work that extends the BMA formulation in Hoeting *et al.* [1999] to any number of levels for analyzing individual contributions of each source of uncertainty with respect to model data, structure, and parameters.

[8] The hierarchical BMA provides more insight than collection BMA on the model selection, model averaging, and uncertainty propagation through a BMA tree. Each level of uncertainty represents an uncertain model component with its different competing discrete model propositions. For example, the variogram model selection can be one source of uncertainty and its competing propositions could be exponential, Gaussian, and pentaspherical variogram models. The proposed HBMA method serves as a framework for evaluating competing propositions of each source of uncertainty, to prioritize different sources of uncertainty and to understand the uncertainty propagation through dissecting uncertain model components.

[9] We test the HBMA method on an indicator hydrostratigraphy model to characterize the Baton Rouge aquifer-fault system in Louisiana. The outline of the study is as follows. Section 2 shows the derivation of HBMA under maximum likelihood estimation. Section 3 describes the indicator hydrostratigraphy model and the segregation of its uncertain model components, which are calibration data, variogram model, geological stationarity assumption, and fault conceptualization. Through the BMA tree of the hydrostratigraphic models, section 4 presents the evaluation of the competing model propositions, the uncertainty propagation, and the prioritization of the uncertain model components. Section 5 draws conclusions about the main features of the HBMA.

## 2. Hierarchical Bayesian Model Averaging

### 2.1. Terminology and Notation

[10] We start by defining some basic terminologies of the BMA tree. Figure 1 shows a BMA tree, which is a hierarchical structure of models at different levels. The growing of the BMA tree reflects the expansion of the number of sources of model uncertainties, which entails the expansion of the number of models. Each source of uncertainty is represented by one level. A *collection* is a set of all models at one level. A *superior* or *subordinate* level is a level that is ranked higher or lower, respectively. The top level of the hierarchy consists of one model, which is termed the *hierarchy* BMA model and its level number is zero. The immediate subordinate level of the top level is level 1 that tackles the first source of uncertainty; the immediate subordinate level of level 1 is level 2 that tackles the second source of uncertainty and so forth. *Ranking* is the arrangement of levels. Ranking of different sources of uncertainty in the BMA tree depends on analysts' preference.

[11] The systematic segregation of different sources of model uncertainties is the central idea of the HBMA method. The base level of the hierarchy is a collection of all candidate models that are resulted from all considered sources of uncertainty. The base models can be viewed as basic

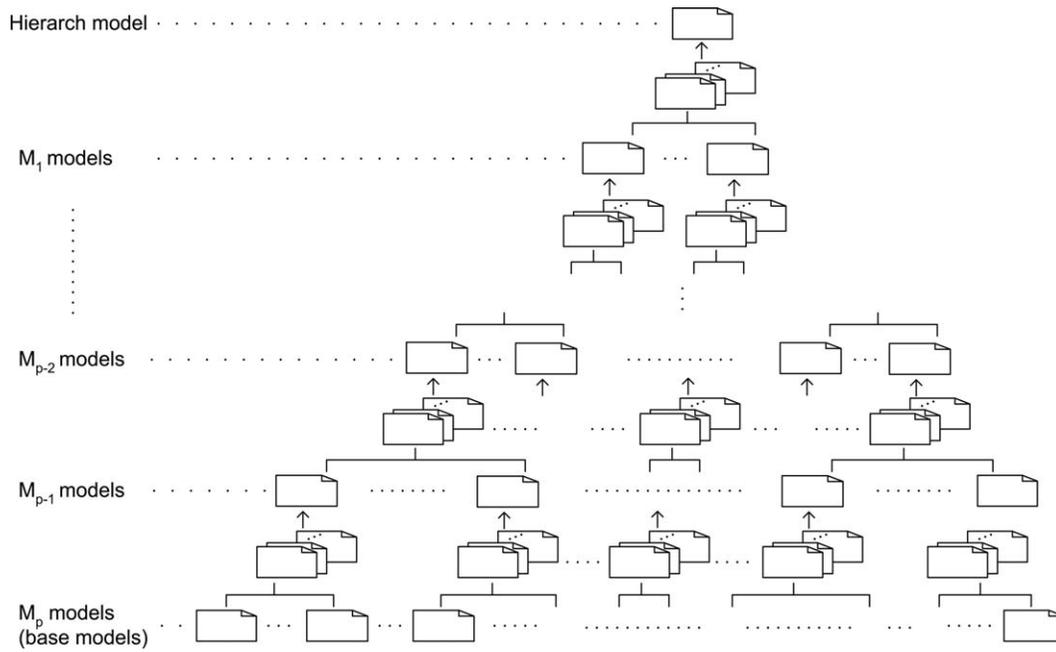


Figure 1. A BMA tree.

elements, and are the same in either the collection BMA method or hierarchical BMA method. The only exception is that the base models of hierarchical BMA are developed following a combinatorial design to achieve a systematic representation of the competing propositions of all sources of uncertainty. The base level tackles the last source of uncertainty. The Bayesian model averaging starts from the base level. All models superior to the base levels are BMA models. A *parent* model is a vertex in the BMA tree, which averages of its *child* models. Each set of child models represent competing propositions for each source of uncertainty.

[12] The subscript is an important index to determine the hierarchy and branch relationship among models. Consider the model  $M_{(ij\dots lm)} \in \mathbf{M}_p$  at level  $p$ . The subscript  $(ij\dots lm)$  <sub>p</sub>

locates the model hierarchically top down from the first level, to the second level, and so forth to reach to level  $p$ . 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 BMA model  $M_0$  is reached. Following these notations, next we formulate the hierarchical BMA posterior model probabilities, hierarchical BMA prediction means, and hierarchical BMA prediction covariances.

**2.2. Posterior Model Probability and Conditional Posterior Model Probability**

[13] Selecting base models for both collection BMA and hierarchical BMA is based on the acknowledged sources of uncertainty. Both collection BMA and hierarchical BMA can

deal with base models which are mutually exclusive and collectively exhaustive if all sources of uncertainty and all propositions become available. However, it is practically impossible to exhaust all uncertain model components and all possible propositions. Accordingly, uncertainty arising from uncertain model components not accounted for cannot be evaluated by either the collection BMA and hierarchical BMA. Therefore, it is understood that the number of considered propositions are not exhaustive. The collection BMA and hierarchical BMA still can deal with nonexhaustive models, but require the base models to be mutually exclusive, which can be achieved in practice by not including nested models.

[14] Consider base models at level  $p$ . According to the law of total probability, the posterior probability for predicted quantity  $\Delta$  given data  $\mathbf{D}$  is

$$E_{\mathbf{M}_p}(\Delta|\mathbf{D}) = E_{\mathbf{M}_1}E_{\mathbf{M}_2} \cdots E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_p)], \tag{1}$$

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  $\Delta$  given data  $\mathbf{D}$  and models  $\mathbf{M}_p$  at level  $p$ . The expectation  $E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_p)]$  is posterior probability averaging at level  $p$ . That is

$$E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_p)] = \sum_m \Pr \left( \Delta|\mathbf{D}, M_{(ij\dots lm)} \right) \Pr \left( M_{(ij\dots lm)}|\mathbf{D}, M_{(ij\dots lm)} \right) \tag{2}$$

where  $\Pr \left( \Delta|\mathbf{D}, M_{(ij\dots lm)} \right) = \Pr(\Delta|\mathbf{D}, \mathbf{M}_p)$  and  $m$  is the number of child models at level  $p$  under the branch of the parent model  $(ij \cdots l)$  at level  $p - 1$ .

$$[15] \Pr(M_{(ij\dots lm)}|\mathbf{D}, M_{(ij\dots lm)}) = \Pr(\mathbf{M}_p|\mathbf{D}, \mathbf{M}_{p-1})$$

is the conditional posterior model probability of model  $M_{(ij\dots lm)}$  at level  $p$  under model  $M_{(ij\dots l)}$  at level  $p-1$ .

$\Pr(\mathbf{M}_p|\mathbf{D}, \mathbf{M}_{p-1})$  also represents the conditional posterior model probabilities and will be used to develop a BMA tree of posterior model probabilities. Note that model  $M_{(ij\dots lm)}$  is a child model under the parent model  $M_{(ij\dots l)}$  because both have the same subscript for the first  $p-1$  levels. Equation (2) is the BMA at level  $p$ , which can be written as

$$\Pr(\Delta|\mathbf{D}, \mathbf{M}_{p-1}) = E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_p)]. \quad (3)$$

[16] According to equation (3), one can derive the posterior probability of prediction using BMA over models at any level, say level  $n$ :

$$\Pr(\Delta|\mathbf{D}, \mathbf{M}_n) = E_{\mathbf{M}_{n+1}}E_{\mathbf{M}_{n+2}}\cdots E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_p)]. \quad (4)$$

[17] For collection BMA, only one level of models is considered. Given the law of total probability, equation (1) becomes [Hoeting et al., 1999]

$$\Pr(\Delta|\mathbf{D}) = E_{\mathbf{M}_1}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_1)] = \sum_i \Pr(\Delta|\mathbf{D}, M_{(i)})\Pr(M_{(i)}|\mathbf{D}). \quad (5)$$

where  $\mathbf{M}_1$  are the base models at level 1. Equation (5) is the model averaging of all base models  $M_{(i)}$ ,  $\{i = 1, 2, \dots\}$ . To develop a multilevel method that separates different sources of uncertainty, we represent BMA in its general hierarchical form. Then equation (1) for hierarchical BMA becomes

$$\begin{aligned} \Pr(\Delta|\mathbf{D}) &= E_{\mathbf{M}_1}E_{\mathbf{M}_2}\cdots E_{\mathbf{M}_p}[\Pr(\Delta|\mathbf{D}, \mathbf{M}_p)] \\ &= \sum_i \sum_j \cdots \sum_l \sum_m \Pr(\Delta|\mathbf{D}, M_{(ij\dots lm)})\Pr(M_{(ij\dots lm)}|\mathbf{D}, M_{(ij\dots l)})\Pr(M_{(ij\dots l)}|\mathbf{D}). \end{aligned} \quad (6)$$

[18] Based on the Bayes rule, the posterior model probability for the base models is

$$\Pr(\mathbf{M}_p|\mathbf{D}) = \frac{\Pr(\mathbf{D}|M_{(ij\dots lm)})\Pr(M_{(ij\dots lm)})}{\sum_i \sum_j \cdots \sum_l \sum_m \Pr(\mathbf{D}|M_{(ij\dots lm)})\Pr(M_{(ij\dots lm)})}, \quad (7)$$

where  $\Pr(\mathbf{D}|M_{(ij\dots lm)})$  is the likelihood of a base model.  $\Pr(\mathbf{M}_p|\mathbf{D})$  often refers to the model weight in BMA.  $\Pr(M_{(ij\dots lm)})$  is the prior model probability of a base model. The conditional posterior model probability of a base model under their parent models is

$$\Pr(\mathbf{M}_p|\mathbf{D}, \mathbf{M}_{p-1}) = \frac{\Pr(\mathbf{D}|M_{(ij\dots lm)})\Pr(M_{(ij\dots lm)}|M_{(ij\dots l)})}{\sum_m \Pr(\mathbf{D}|M_{(ij\dots lm)})\Pr(M_{(ij\dots lm)}|M_{(ij\dots l)})}, \quad (8)$$

where  $\Pr(M_{(ij\dots lm)}|M_{(ij\dots l)})$  is the conditional prior model probability of a base model  $M_{(ij\dots lm)} \in \mathbf{M}_p$  under its parent

model  $M_{(ij\dots l)} \in \mathbf{M}_{p-1}$ . Equation (8) is also referred to as the conditional model weight. The likelihood for parametric base models is

$$\Pr(\mathbf{D}|\mathbf{M}_p) = \int_{\beta_p} \Pr(\mathbf{D}|\mathbf{M}_p, \beta_p)\Pr(\beta_p|\mathbf{M}_p)d\beta_p, \quad (9)$$

where  $\beta_p$  is a vector of model parameters for base models  $\mathbf{M}_p$ . The likelihood,  $\Pr(\mathbf{D}|\mathbf{M}_{p-1})$  is

$$\Pr(\mathbf{D}|M_{(ij\dots l)}) = \sum_m \Pr(\mathbf{D}|M_{(ij\dots lm)})\Pr(M_{(ij\dots lm)}|M_{(ij\dots l)}). \quad (10)$$

[19] By considering equal conditional prior model probabilities for  $\Pr(M_{(ij\dots lm)}|M_{(ij\dots l)})$ , we calculate conditional posterior model probability for the base models under their parent models as follows

$$\Pr(\mathbf{M}_p|\mathbf{D}, \mathbf{M}_{p-1}) = \frac{\Pr(\mathbf{D}|M_{(ij\dots lm)})}{\sum_m \Pr(\mathbf{D}|M_{(ij\dots lm)})}. \quad (11)$$

[20] Using equations (10) and (11) under the consideration of equal conditional prior model probabilities, the conditional posterior model probability for models at level  $n$  under their parent models is

$$\Pr(\mathbf{M}_n|\mathbf{D}, \mathbf{M}_{n-1}) = \frac{\sum_{n+1} \cdots \sum_l \sum_m \Pr(M_{(ij\dots lm)}|\mathbf{D}, M_{(ij\dots l)})}{\sum_n \sum_{n+1} \cdots \sum_l \sum_m \Pr(M_{(ij\dots lm)}|\mathbf{D}, M_{(ij\dots l)})}. \quad (12)$$

[21] And the posterior model probability at level  $n$  is

$$\Pr(\mathbf{M}_n|\mathbf{D}) = \frac{\sum_{n+1} \cdots \sum_l \sum_m \Pr(\mathbf{D}|M_{(ij\dots lm)})}{\sum_i \sum_j \cdots \sum_l \sum_m \Pr(\mathbf{D}|M_{(ij\dots lm)})}. \quad (13)$$

[22] Therefore, each model at any level in Figure 1 has its own posterior model probabilities as in equation (13) and conditional posterior model probabilities as in equation (12). As a result, a BMA tree of posterior model probabilities can be obtained.

### 2.3. Prediction Means and Prediction Covariances

[23] Based on the law of total expectation, the expectation of prediction over  $p$  levels of models for hierarchical BMA is

$$E(\Delta|\mathbf{D}) = E_{\mathbf{M}_1}E_{\mathbf{M}_2}\cdots E_{\mathbf{M}_p}[E(\Delta|\mathbf{D}, \mathbf{M}_p)]. \quad (14)$$

where  $E(\Delta|\mathbf{D}, \mathbf{M}_p)$  is the expectation of prediction for given data  $\mathbf{D}$  and models at level  $p$ . Moreover, the hierarchical BMA not only shows the total expectation of

prediction over all levels of models, but also shows the expectation of prediction at a desired level where models are used. According to equation (4), the expectation of prediction using models at level  $n$  is

$$E(\Delta|\mathbf{D}, \mathbf{M}_n) = E_{\mathbf{M}_{n+1}} E_{\mathbf{M}_{n+2}} \cdots E_{\mathbf{M}_p} [E(\Delta|\mathbf{D}, \mathbf{M}_p)]. \quad (15)$$

where  $n < p$ . Equation (15) provides thorough information for analysts who can have flexibility to see all possible averaged predicted quantities using various BMA models at different levels, while typical (one-level) BMA only provides one overall expectation of all models. Using equation (15) at any level in Figure 1, a BMA tree of prediction means can be obtained.

[24] The law of total covariance for hierarchical BMA is

$$\begin{aligned} \text{Cov}(\Delta|\mathbf{D}) &= E_{\mathbf{M}_1} E_{\mathbf{M}_2} \cdots E_{\mathbf{M}_p} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_p)] \\ &+ \sum_{n=p}^{n=1} E_{\mathbf{M}_1} E_{\mathbf{M}_2} \cdots E_{\mathbf{M}_{n-1}} \text{Cov}_{\mathbf{M}_n} [E(\Delta|\mathbf{D}, \mathbf{M}_n)], \end{aligned} \quad (16)$$

where  $\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_p)$  is the covariance of prediction for given data  $\mathbf{D}$  and base models at level  $p$  and  $\text{Cov}_{\mathbf{M}_n}$  is the covariance operator with respect to model  $\mathbf{M}_n$  at level  $n$ .  $\text{Cov}_{\mathbf{M}_n}$  is the between-model covariance:

$$\begin{aligned} \text{Cov}_{\mathbf{M}_n} [E(\Delta|\mathbf{D}, \mathbf{M}_n)] \\ = E_{\mathbf{M}_n} \left[ [E(\Delta|\mathbf{D}, \mathbf{M}_n) - E(\Delta|\mathbf{D}, \mathbf{M}_{n-1})][E(\Delta|\mathbf{D}, \mathbf{M}_n) - E(\Delta|\mathbf{D}, \mathbf{M}_{n-1})]^T \right], \end{aligned} \quad (17)$$

where  $T$  is the transpose operator.

[25] When considering the collection BMA that has only one level ( $p - 1$ ), the covariance in equation (16) returns to the usual form [Hoeting *et al.*, 1999]

$$\text{Cov}(\Delta|\mathbf{D}) = E_{\mathbf{M}_1} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_1)] + \text{Cov}_{\mathbf{M}_1} [E(\Delta|\mathbf{D}, \mathbf{M}_1)]. \quad (18)$$

where  $\mathbf{M}_1$  are the base models at level 1.

[26] When considering two levels ( $p - 2$ ) as in Li and Tsai [2009] and Tsai [2010], the covariance in equation (16) is

$$\begin{aligned} \text{Cov}(\Delta|\mathbf{D}) &= E_{\mathbf{M}_1} E_{\mathbf{M}_2} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_2)] + E_{\mathbf{M}_1} \text{Cov}_{\mathbf{M}_2} [E(\Delta|\mathbf{D}, \mathbf{M}_2)] \\ &+ \text{Cov}_{\mathbf{M}_1} [E(\Delta|\mathbf{D}, \mathbf{M}_1)]. \end{aligned} \quad (19)$$

[27] For this study, we will consider four sources of uncertainty, i.e., four levels ( $p = 4$ ). The hierarchical BMA formulation following equations (15) and (16) is

$$E(\Delta|\mathbf{D}) = E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} E_{\mathbf{M}_4} [E(\Delta|\mathbf{D}, \mathbf{M}_4)], \quad (20)$$

$$\begin{aligned} \text{Cov}(\Delta|\mathbf{D}) &= E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} E_{\mathbf{M}_4} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_4)] \\ &+ E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} \text{Cov}_{\mathbf{M}_4} [E(\Delta|\mathbf{D}, \mathbf{M}_4)] \\ &+ E_{\mathbf{M}_1} E_{\mathbf{M}_2} \text{Cov}_{\mathbf{M}_3} [E(\Delta|\mathbf{D}, \mathbf{M}_3)] \\ &+ E_{\mathbf{M}_1} \text{Cov}_{\mathbf{M}_2} [E(\Delta|\mathbf{D}, \mathbf{M}_2)] \\ &+ \text{Cov}_{\mathbf{M}_1} [E(\Delta|\mathbf{D}, \mathbf{M}_1)]. \end{aligned} \quad (21)$$

[28] Similarly, the hierarchical BMA permits the evaluation of the prediction covariance when different BMA models at different levels are proposed. The basic informa-

tion from base models is their covariance of prediction  $\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_p)$  at level  $p$  and their mean of prediction  $E(\Delta|\mathbf{D}, \mathbf{M}_p)$ . Then, the covariance of prediction using individual models at level  $n < p$  is

$$\begin{aligned} \text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_n) &= E_{\mathbf{M}_{n+1}} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_{n+1})] \\ &+ \text{Cov}_{\mathbf{M}_{n+1}} [E(\Delta|\mathbf{D}, \mathbf{M}_{n+1})]. \end{aligned} \quad (22)$$

[29] Therefore, the within-model covariance of prediction using models at level  $n < p$  is

$$\begin{aligned} E_{\mathbf{M}_n} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_n)] &= E_{\mathbf{M}_n} E_{\mathbf{M}_{n+1}} [\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_{n+1})] \\ &+ E_{\mathbf{M}_n} \text{Cov}_{\mathbf{M}_{n+1}} [E(\Delta|\mathbf{D}, \mathbf{M}_{n+1})]. \end{aligned} \quad (23)$$

[30] The between-model covariance of prediction using models at level  $n$  is equation (17). The within-model covariance in equation (23) contains the sum of the within-model covariance and the between-model covariance at level  $n + 1$ . If stepping into the within-model covariance  $\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_{n+1})$  in equation (23), one can see that this term is composed of the within-model covariance and the between-model covariance at level  $n + 2$ . In other words, except for the base models, the within-model covariance at level  $n$  is composed of the within-model covariances and the between-model covariances at levels  $n + 1$ ,  $n + 2$ , and so forth, up to  $p$ . Using equation (22) for each model at any level in Figure 1, a BMA tree of prediction covariances can be developed.

[31] From the calculation procedure, one needs to first obtain the expectation and covariance for all base models, i.e.,  $E(\Delta|\mathbf{D}, \mathbf{M}_p)$  and  $\text{Cov}(\Delta|\mathbf{D}, \mathbf{M}_p)$  because the base models are the basic elements to either the collection BMA or the hierarchical BMA. Then, the expectation of prediction using models at level  $n$  in equation (15) needs to be calculated starting from level  $p$ , then to level  $p-1$ , then to level  $p-2$ , and so forth until it reaches level  $n$ . Similarly, the within-model covariance in equation (23) and the between-model covariance in equation (17) at level  $n$  need to start from level  $p$ , then to level  $p-1$ , then to level  $p-2$ , and so forth until it reaches level  $n$ .

[32] These derivations show that the calculation of the posterior model probabilities for the hierarchical BMA is the same as collection BMA since all models above the base level are BMA models. However, the conditional posterior model probability calculation in the hierarchical BMA is different since it only takes place for child models under their parent models, allowing for the segregation of the competing model propositions and the segregation of the uncertain model components. This is different from collection BMA, in which all child models are treated as one set.

#### 2.4. Computation of Posterior Model Probability With Variance Window

[33] Computation of posterior model probability can be done through sampling techniques or information-theoretic criteria. Markov Chain Monte Carlo (MCMC) simulation has been the most common tool to infer posterior distributions [Wöhling and Vrugt, 2008; Rojas *et al.*, 2010b]. Although accurate, the MCMC simulation requires a large ensemble to achieve stable convergence, which will be computationally expensive. Alternatively, information-

theoretic criteria such as Akaike Information Criterion [Poeter and Anderson, 2005; Singh et al., 2010], Bayesian information criterion (BIC), and Kashyap information criterion (KIC) [Neuman, 2003; Ye et al., 2004; Tsai and Li, 2008a; Singh et al., 2010] are inexpensive, fair estimators to evaluate posterior model probability. We note that due to the differences in their basic statistical assumptions, different information-theoretic criteria can often lead to different model ranking and posterior model probabilities [Tsai and Li, 2008a; Rojas et al., 2010c; Singh et al., 2010; Foglia et al., 2013]. A debate on the selection of BIC and KIC under the Bayesian paradigm is given by Ye et al. [2010a] and Tsai and Li [2010]. Assuming a large data size and a Gaussian distribution for prior model parameters [Raftery, 1995], this study adopts the Bayesian information criterion (BIC). Nevertheless, other sampling techniques or information-theoretic criteria can be considered in HBMA. Following skips BIC derivation and readers are referred to Draper [1995] and Raftery [1995].

[34] The likelihood for a base model in equation (9) is

$$\Pr(\mathbf{D}|M_{(ij \dots lm)}) \approx \exp\left[-\frac{1}{2}\text{BIC}_{(ij \dots lm)}\right]. \quad (24)$$

[35] The Bayesian information criterion is

$$\text{BIC}_{(ij \dots lm)} = -2\ln \Pr(\mathbf{D}|M_{(ij \dots lm)}, \hat{\beta}_{(ij \dots lm)}) + m_{(ij \dots lm)} \ln n, \quad (25)$$

where  $\hat{\beta}_{(ij \dots lm)}$  are the maximum-likelihood estimated model parameters in model  $M_{(ij \dots lm)}$ ,  $m_{(ij \dots lm)}$  is the dimension of the model parameters  $\hat{\beta}_{(ij \dots lm)}$ , and  $n$  is the size of data set  $\mathbf{D}$ .  $\Pr(\mathbf{D}|M_{(ij \dots lm)}, \hat{\beta}_{(ij \dots lm)})$  is the maximum likelihood value. By considering equal prior parameter probabilities for  $\hat{\beta}_{(ij \dots lm)}$  and a multi-Gaussian distribution for fitting errors to observation data  $\Delta^{obs}$ , the BIC in equation (25) is simplified to [Tsai and Li, 2008a; Li and Tsai, 2009]

$$\text{BIC}_{(ij \dots lm)} = Q_{(ij \dots lm)} + n \ln 2\pi + m_{(ij \dots lm)} \ln n, \quad (26)$$

where

$$Q_{(ij \dots lm)} = \left(\Delta^{cal}(\hat{\beta}_{(ij \dots lm)}) - \Delta^{obs}\right)^T C_{\Delta}^{-1} \left(\Delta^{cal}(\hat{\beta}_{(ij \dots lm)}) - \Delta^{obs}\right), \quad (27)$$

is the sum of the weighted squared fitting errors between calculated  $\Delta^{cal}$  and observed  $\Delta^{obs}$ .  $C_{\Delta}$  is the error covariance matrix.

[36] Substituting BIC in equation (24) into equation (7) and assuming equal prior model probability for  $\Pr(M_{(ij \dots lm)})$ , the posterior model probability for the base model is

$$\Pr(M_{(ij \dots lm)}|\mathbf{D}) = \frac{\exp\left[-\frac{1}{2}\Delta\text{BIC}_{(ij \dots lm)}\right]}{\sum_i \sum_j \dots \sum_l \sum_m \exp\left[-\frac{1}{2}\Delta\text{BIC}_{(ij \dots lm)}\right]}, \quad (28)$$

where  $\Delta\text{BIC}_{(ij \dots lm)} = \text{BIC}_{(ij \dots lm)} - \text{BIC}_{\min}$ .  $\text{BIC}_{\min}$  is the minimum BIC value among all the base models. Using

$\Delta\text{BIC}$  in equation (28) is a common practice [e.g., Neuman, 2003; Li and Tsai, 2009] to avoid numerical difficulty when  $\text{BIC}_{(ij \dots lm)}$  are large numbers. The conditional posterior model probabilities of the base models under their parent models are

$$\Pr(\mathbf{M}_p|\mathbf{D}, \mathbf{M}_{p-1}) = \frac{\Pr(M_{(ij \dots lm)}|\mathbf{D})}{\sum_m \Pr(M_{(ij \dots lm)}|\mathbf{D})}. \quad (29)$$

[37] Once we obtain the posterior model probabilities and conditional posterior model probabilities for base models, the posterior model probabilities and conditional posterior model probabilities at any level can be obtained via equations (13) and (12), respectively.

[38] This study adopts the variance window [Tsai and Li, 2008a, 2008b; Li and Tsai, 2009] to calculate the posterior model probabilities for base models. The variance window introduces a scaling factor

$$\alpha = \frac{s_1}{s_2 \sigma_D}, \quad (30)$$

into equation (28) as follows

$$\Pr(M_{(ij \dots lm)}|\mathbf{D}) = \frac{\exp\left[-\frac{1}{2}\alpha\Delta\text{BIC}_{(ij \dots lm)}\right]}{\sum_i \sum_j \dots \sum_l \sum_m \exp\left[-\frac{1}{2}\alpha\Delta\text{BIC}_{(ij \dots lm)}\right]}. \quad (31)$$

[39] The parameter  $s_1$  is the  $\Delta\text{BIC}$  value corresponding to the significance level in Occam's window, and  $s_2$  is the width of the variance window in the unit of  $\sigma_D = \sqrt{2n}$ . The selection of significance level for  $s_1$  and the selection of the window size for  $s_2$  are subjected to an analyst's preference. If the scaling factor is zero, then all base models are weighted equally. If the scaling factor is unity, then the base models are weighted according to Occam's window. If the scaling factor is smaller than unity, then we enlarge Occam's window to accept more models. The scaling factor can be seen as analogous to the smoothed information criteria [Hjort and Claeskens, 2003, 2006]. For more details, we refer the readers to groundwater studies that compare the use of the variance window and Occam's window [Tsai and Li, 2008a, 2008b; Li and Tsai, 2009; Singh et al., 2010].

[40] The posterior model probability is an epistemic probability [Ellison, 2004; Williamson, 2005]. 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]. Thus, the validity of posterior model probability is subject to our knowledge, and the estimated posterior model probabilities will be subject to revision should new knowledge become available. The term knowledge here is not merely limited to our knowledge about the different propositions of the model data, structure, parameters or processes, but also extends to the statistical matrices that are used to calculate the posterior model probabilities.

[41] This provides a complete formulation for hierarchical BMA with variance window, from which we can conclude the following basic concepts. Similar to the

collection BMA, the base level of the hierarchical BMA represents the individual models given the full array of different propositions with corresponding posterior model probabilities. All models above the base level are BMA models. For each level, the posterior model probability, conditional posterior model probability, prediction mean and covariances (within-model covariances, between-model covariances, and total model covariances) can be obtained for each BMA model and presented through a BMA tree. From the base level to the level 1, each level distinguishes uncertainty arising from one source of uncertainty. The top level of the hierarchy contains the hierarch BMA model, which contains information from all subordinate models. In other words, the hierarch BMA model is identical to the BMA model of the collection BMA.

### 2.5. Similarities and Differences Between Collection BMA and Hierarchical BMA

[42] The previous analysis shows that the hierarchical BMA provides the general form for the BMA in *Hoeting et al.* [1999]. The result of the hierarch model of the hierarchical BMA is identical to the result of the collection BMA. However, *Gupta et al.* [2012] comment 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.” While the hierarchical BMA can be used for model averaging similar to collection BMA, GLUE, or other averaging methods [e.g., *Seifert et al.*, 2012], yet on the other hand it facilitates a different purpose altogether that is to learn about the individual model components with their competing propositions. This is in line with *Gupta et al.* [2012] conclusion that “a systematic characterization of different aspects of model structural adequacy will help by explicitly recognizing the role of each aspect in shaping the overall adequacy of the model.” In other words, hierarch BMA can serve as “multiple hypothesis methodology” as proposed by *Clark et al.* [2011] in which competing hypotheses can be systematically constructed and evaluated, providing a learning tool and can lead to considerably more scientifically defensible model. To serve this purpose, hierarchical BMA offers four additional features to collection BMA as follows.

[43] First, through model dissection following a combinatorial design, the hierarchical BMA provides a systematic representation of the competing propositions of all sources of uncertainty. This can also be done with collection BMA. Yet since this is not a prerequisite to collection BMA, thus it is not a common practice. Second, model dissection allows the evaluation of competing model propositions of each uncertain model component through using the BMA tree of posterior model probabilities. Although this can be directly inferred from model ranking (e.g., *Foglia et al.* [2013] and this study), yet the BMA tree of posterior model probabilities provides more detailed information. Third, the segregation of the between-model variance for each uncertain model component allows for the prioritization of different sources of uncertainty. Fourth, hierarchical BMA facilitates the illustration of the change of the BMA prediction due to the addition of each source of uncertainty, while collection BMA (one-level) only provides one overall BMA prediction of all models. Similarly, the total

model variance for each uncertain model component depicts the uncertainty propagation resulting from adding up different sources of uncertainty. Thus, the hierarchical BMA allows for uncertainty segregation, for comparative evaluation of competing model propositions, for prioritizing the uncertain model components, and for depicting the prediction and uncertainty propagation. These features, which advance our knowledge about the uncertain model components, are not readily possible to obtain through collection BMA. We illustrate these four features in the following case study.

## 3. Case Study

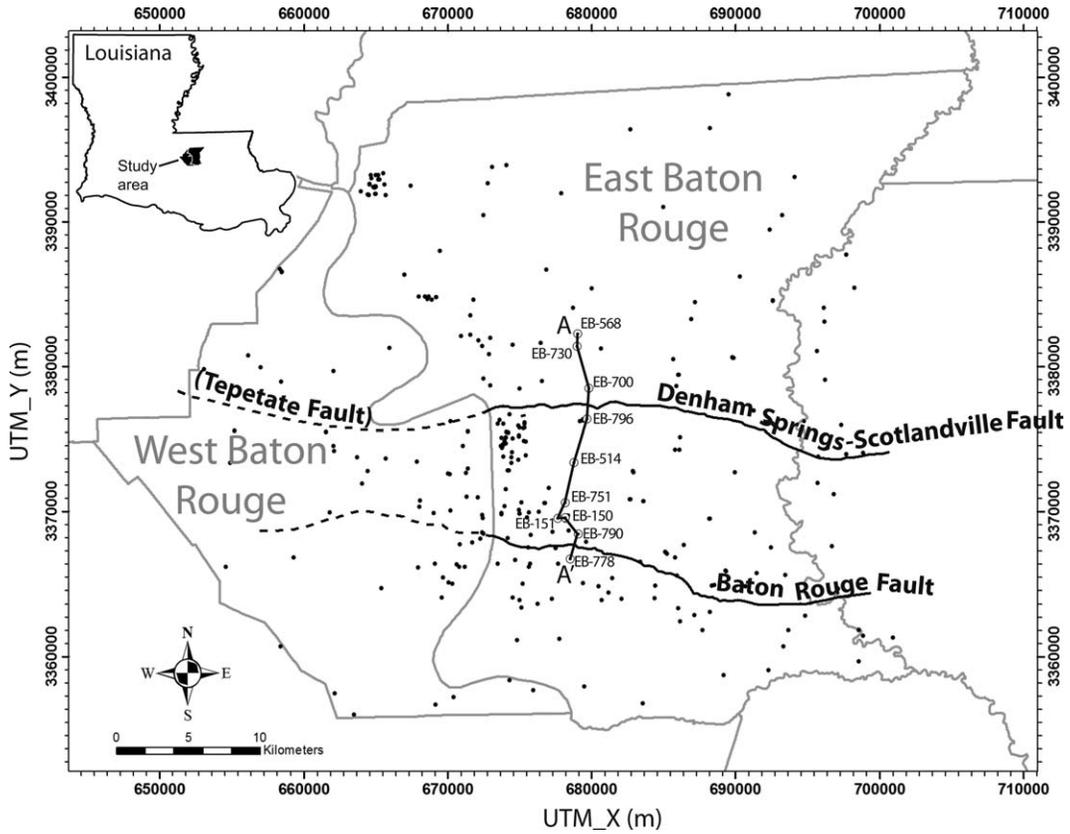
### 3.1. Baton Rouge Aquifer-Fault System, Louisiana

[44] The Baton Rouge aquifer system in Louisiana is located at the south limit of the Southern Hills regional aquifer system [*Buono*, 1983]. The aquifer system consists of complexly interbedded series of sand and clay formations that gently dip south [*Tomaszewski*, 1996]. This sequence of aquifers/aquicludes extends to a depth of 3000 ft (914.4 m) in the Baton Rouge area. The study area shown in Figure 2 focuses on the Miocene-Pliocene deposits [*Griffith*, 2003] of the “1200 foot” sand, the “1500 foot” sand, the “1700 foot” sand, the “2000 foot” sand, and the “2400 foot” sand. These sand units were classified and named by their approximate depth below ground level in Baton Rouge industrial district [*Meyer and Turcan*, 1955].

[45] The Baton Rouge fault system consists of the Baton Rouge fault and the Denham Springs-Scotlandville fault (Tepetate fault) [*McCulloh and Heinrich*, 2012]. This east-west trending fault system crosscuts the aquifer/aquiclude sequences in the study area as shown in Figure 2. The low permeable Baton Rouge fault is important from a resource point of view since it separates the sequence of freshwater and brackish aquifers at the north and south of the fault, respectively. Heavy ground water pumping reversed the flow direction, which was originally southward, resulting in saltwater encroachment from south of the Baton Rouge fault [*Rollo*, 1969; *Tomaszewski*, 1996; *Tsai*, 2010]. No detailed hydrogeological information is available on the Denham Springs-Scotlandville fault in literature. To better understand the hydrogeological settings of the Baton Rouge aquifer-fault system, it is important to study detailed hydrostratigraphic architecture of the system.

### 3.2. Hydrostratigraphic Model and Its Uncertainty

[46] Due to uncertainty of the model data, structure, and parameters, multiple potential hydrostratigraphic models are resulted and calibrated. The central idea of the HBMA method is to segregate different uncertain model components with their corresponding competing model propositions. We illustrate these concepts of uncertain model components and competing model propositions in Figure 3. In this case study as shown in Figure 3, we consider four uncertain model components in the hydrostratigraphic model, which are two calibration data sets, three variogram models, two geological stationarity assumptions, and two conceptualizations with respect to the Denham Springs-Scotlandville fault. Alternatively, Figure 3 shows that the hydrofacies data interpretation and parameter estimation technique are not considered as uncertain model

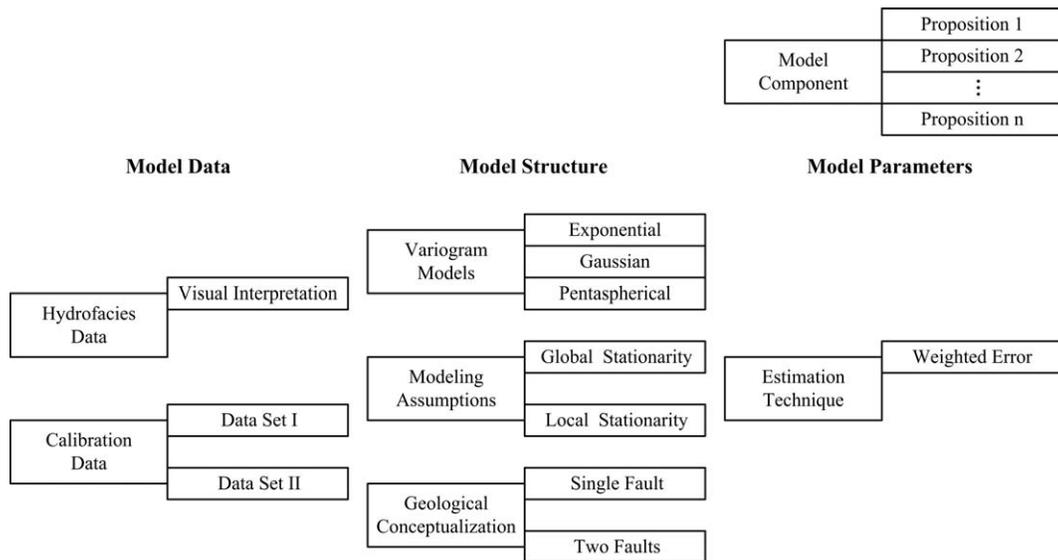


**Figure 2.** The study area. Black dots represent the location of electrical well logs. The dashed lines in West Baton Rouge Parish represent approximated location of the faults. *AA'* is a cross section for Figure 5.

components since only one proposition is considered for each of these model components.

[47] The four uncertain model components with their corresponding competing propositions result in 24 calibrated models. We use these models to perform hierarchi-

cal multimodel characterization of the hydrostratigraphy of the Baton Rouge aquifer-fault system to present the main features of the HBMA method. In this section, we present detailed description of the hydrostratigraphy model with its uncertain model components.



**Figure 3.** Uncertainty segregation through dissection of model components with their competing modeling propositions.

[48] We define our hydrostratigraphy model on a scale that is smaller than the regional scale and larger than the channel scale to describe strongly bimodal spatial distribution of intercalated pervious and impervious units. This scale is the same as the hydrofacies assemblages complex scale [Rubin, 2003] and depositional environment scale [Koltermann and Gorelick, 1996]. This strong bimodal scale of characterization fits our objective of delineating the thickness, lateral extent and depth of sand and clay units underneath Baton Rouge. For reconstructing the hydrostratigraphic architecture, we adopt the indicator hydrostratigraphy method [Johnson and Dreiss, 1989; Johnson, 1995]. The study analyzes 288 electrical well logs and interprets the sand-clay bimodal sequence for each log based on electrical resistivity, spontaneous-potential, and gamma ray.

[49] For model calibration, we use lithologic data from 33 driller’s logs. However, different interpretations of drillers’ logs lead to multiple calibration data sets. Sand and gravel are considered as sand facies with indicator 1. Silt and clay are considered as clay facies with indicator 0. The interpretation uncertainty arises from indistinct lithologic terms such as “sand with shale,” “shaly sand,” “sand with strikes of shale,” and so forth. Two data sets are proposed. Data set I interprets the indistinct lithologic terms clay facies with indicator 0. The data set II interprets the indistinct lithologic terms as sand facies with indicator 1.

[50] With respect to the hydrostratigraphic model structure, the first uncertain model component is the choice of the spatial correlation function of the hydrofacies units. In this study, we use three competing propositions, which are exponential, pentaspherical and Gaussian variogram models. The second source of uncertainty concerning the model structure is the geological stationarity assumption. If geological stationarity is shown to be inappropriate, it is helpful to divide the system into zones that are likely to be stationary [Koltermann and Gorelick, 1996; Rubin, 2003; Deutsch, 2007]. For the uncertainty analysis, we adopt two geological stationarity propositions. Global stationarity proposition assumes geological stationarity over the entire modeling domain resulting in one global variogram model. Local stationarity proposition assumes stationarity for each model zone as separated by the fault system resulting in local variogram model for each model zone. For the global variogram model proposition, the correlation between the data across the faults is still prevented, yet the experimental

variograms from all zones are used to fit one theoretical variogram model. Beside the aforementioned mathematical structure uncertainty, model structure uncertainty also includes geological conceptualization uncertainty. For example, different fault characterizations can lead to different model structures [Chester et al., 1993; Bredehoeft, 1997; Salve and Oldenburg, 2001; Fairley et al., 2003; Nishikawa et al., 2009]. This study investigates the geological effect due to the Denham Springs-Scotlandville fault. While the Baton Rouge fault is significant to fluid flow, the Denham Springs-Scotlandville fault was not considered in many groundwater models [Torak and Whiteman, 1982; Huntzinger et al., 1985; Tsai and Li, 2008a; Li and Tsai, 2009; Tsai, 2010] due to the presence of no significant evidence of hydraulic discontinuity across the fault.

[51] We test two geological conceptualization propositions, which are the two-zone proposition and the three-zone proposition. The two-zone proposition does not consider the Denham Springs-Scotlandville fault similar to Rollo [1969], and thus the model domain is separated into two zones by the Baton Rouge fault. The correlation between the well log data across Denham Springs-Scotlandville fault is allowed. The three-zone proposition explicitly accounts for the Denham Springs-Scotlandville fault, thus the model domain is separated into three zones. The correlation between the well log data across the Denham Springs-Scotlandville fault is prevented.

### 3.3. Model Parameters and Calibration

[52] This section presents the inverse procedure to estimate the unknown model parameters. The first model parameter is the formation dip, which establishes data correlation. Different formation dips have a significant effect on the variogram structure and selection of data points. To obtain prior information to constrain the search space, we calculate the formation dip to be  $\phi = 0.30^\circ \pm 0.06^\circ$  from the U.S. Geological Survey (USGS) cross-sectional map in the area [Griffith, 2003]. We assign a range of  $0.06^\circ \leq \phi \leq 0.57^\circ$  for the formation dip. The second model parameter is the sand-clay cutoff  $\theta$ , which rounds the indicator estimate  $\Delta$  to a binary value. We set the range of the cutoff to  $0.3 \leq \theta \leq 0.7$ .

[53] To estimate the unknown model parameters, we formulate the inverse problem by minimizing the fitting errors between the estimated and observed facies as follows

$$\min_{\phi, \theta} \frac{1}{2} \left\{ \frac{1}{M_{sand}} \sum_{i=1}^{M_{sand}} \frac{[\Delta^{i,est}(\mathbf{x}) - \Delta_{sand}^{i,obs}(\mathbf{x})]^2}{\sigma^2(\mathbf{x})} + \frac{1}{M_{clay}} \sum_{i=1}^{M_{clay}} \frac{[\Delta^{i,est}(\mathbf{x}) - \Delta_{clay}^{i,obs}(\mathbf{x})]^2}{\sigma^2(\mathbf{x})} \right\} \quad (32)$$

where,  $M_{sand}$  and  $M_{clay}$  are the data size of the sand facies and clay facies, respectively,  $\Delta_{sand}^{i,obs}(\mathbf{x})$ ,  $\Delta_{clay}^{i,obs}(\mathbf{x})$ , and  $\Delta^{i,est}(\mathbf{x})$  are the observed sand facies indicator, the observed clay facies indicator and the indicator estimate at a location  $\mathbf{x}$ , respectively. To make the calibration consistent with equation (27), equa-

tion (32) includes the variance term  $\sigma^2(\mathbf{x})$ , which is the sum of the data variance and the kriging variance at location  $\mathbf{x}$ . The data variance for the two calibration data sets is 0.128 as calculated from the differences between electrical and driller’s logs when both are available at the same locations.

**Table 1.** Calibrated Model Parameters, Fitting Errors (Equation (32)), Q, ΔBIC, and Posterior Model Probabilities for Base Models<sup>a</sup>

Base Model	Parameters and Results					Posterior Model Probabilities						
	Dip (deg.)	Cutoff	Fitting Error	Q	ΔBIC	Occam's Window	Variance Window 1%			Variance Window 5%		
							1σ <sub>D</sub>	2σ <sub>D</sub>	3σ <sub>D</sub>	1σ <sub>D</sub>	2σ <sub>D</sub>	3σ <sub>D</sub>
Z2D1GExp	0.23	0.39	2.010	10,010	854	0.00	0.00	0.04	1.42	0.00	0.52	3.54
Z2D1GGus	0.44	0.44	2.181	10,620	1464	0.00	0.00	0.00	0.09	0.00	0.01	0.57
Z2D1GPen	0.30	0.41	2.208	10,927	1771	0.00	0.00	0.00	0.02	0.00	0.00	0.23
Z2D1LExp	0.44	0.42	2.080	10,062	906	0.00	0.01	0.79	6.50	0.19	3.74	9.51
Z2D1LGus	0.32	0.41	2.111	10,243	1087	0.00	0.00	0.00	0.49	0.00	0.13	1.77
Z2D1LPen	0.44	0.42	2.005	9679	523	0.00	0.00	0.02	1.12	0.00	0.38	3.03
Z2D2GExp	0.19	0.40	2.100	10,190	1034	0.00	0.00	0.01	0.62	0.00	0.18	2.07
Z2D2GGus	0.44	0.42	2.160	10,362	1206	0.00	0.00	0.00	0.28	0.00	0.06	1.24
Z2D2GPen	0.18	0.41	2.621	11,243	2087	0.00	0.00	0.00	0.00	0.00	0.00	0.09
Z2D2LExp	0.44	0.42	2.087	10,132	976	0.00	0.00	0.01	0.81	0.00	0.25	2.46
Z2D2LGus	0.28	0.40	2.151	10,349	1193	0.00	0.00	0.00	0.30	0.00	0.07	1.29
Z2D2LPen	0.17	0.42	2.372	11,226	2070	0.00	0.00	0.00	0.01	0.00	0.00	0.09
Z3D1GExp	0.29	0.40	2.087	10,087	931	0.00	0.00	0.02	1.00	0.00	0.33	2.81
Z3D1GGus	0.18	0.41	2.219	10,928	1772	0.00	0.00	0.00	0.02	0.00	0.00	0.23
Z3D1GPen	0.30	0.41	2.000	9544	388	0.00	0.08	2.73	12.09	0.96	8.37	14.23
Z3D1LExp	0.45	0.41	1.934	9156	0	100.00	99.91	96.32	71.80	98.84	84.95	45.34
Z3D1LGus	0.34	0.40	2.105	10,194	1038	0.00	0.00	0.01	0.61	0.00	0.17	2.04
Z3D1LPen	0.28	0.41	2.150	10,253	1097	0.00	0.00	0.00	0.47	0.00	0.12	1.71
Z3D2GExp	0.29	0.40	2.092	10,159	1003	0.00	0.00	0.01	0.72	0.00	0.21	2.27
Z3D2GGus	0.18	0.42	2.188	10,845	1689	0.00	0.00	0.00	0.03	0.00	0.00	0.29
Z3D2GPen	0.30	0.41	2.249	11,118	1962	0.00	0.00	0.00	0.01	0.00	0.00	0.13
Z3D2LExp	0.45	0.42	2.067	10,042	886	0.00	0.00	0.03	1.23	0.00	0.43	3.22
Z3D2LGus	0.30	0.40	2.164	10,394	1238	0.00	0.00	0.00	0.24	0.00	0.05	1.12
Z3D2LPen	0.44	0.42	2.174	10,548	1392	0.00	0.00	0.00	0.12	0.00	0.02	0.71

<sup>a</sup>Z3D1LExp is the best model.

[54] Given two fault conceptualizations, two calibration data sets, two geological stationarity assumptions, and three variogram models combinatorial design results in 24 base models. The unknown model parameters ( $\phi, \theta$ ) are independently estimated for each of the 24 models. The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [Hansen *et al.*, 2003], which is a global-local derivative-free optimization algorithm, is used to solve the inverse problem in equation (32) according to the following procedure. First, the CMA-ES generates candidate solutions ( $\phi, \theta$ ). For each candidate solution, the experimental variograms for each zone are calculated given the formation dip  $\phi$ . Then a theoretical variogram model is automatically fitted to the experimental variograms using the direct search method of Hooke and Jeeves [1961]. Third, indicator kriging is used to estimate facies at the locations of observation data. The indicator kriging estimates are then rounded to indicators by the sand-clay cutoff  $\theta$ . Fourth, the fitting error is calculated by comparing the estimated indicators to the observation data set, which is data set I or data set II, according to equation (32). This procedure is repeated until the fitting error is minimized.

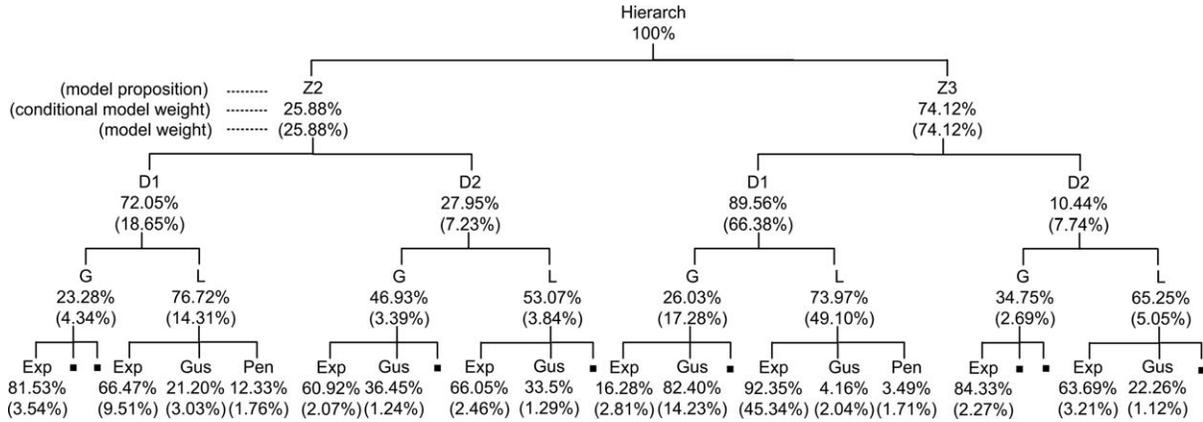
## 4. Results and Discussion

### 4.1. Calibration and BIC

[55] For results and discussion, we use the following short forms. The first level of uncertainty is about the conceptualization of the Denham Springs-Scotlandville fault resulting into two-zone (Z2) and three-zone (Z3) propositions. The second level is for calibration data containing the data set I (D1) and the data set II (D2). The third level

has the global (G) and the local (L) stationarity assumptions. The fourth level of uncertainty has three propositions, which are Exponential (Exp), Gaussian (Gus), and Pentaspherical (Pen) variogram models. The short forms of each proposition form the name of the 24 base models and their corresponding hierarchical BMA models. For example, Z3D1LExp is the name of a base model with three-zone (Z3), using the calibration data set I (D1), local stationarity assumption (L), and Exponential variogram model (Exp). The name Z3D1L represents a BMA model of the Z3D1LExp model, the Z3D1LGus model, and Z3D1LPen model under the propositions Z3, D1, and L. Similarly, the Z3D1 model represents a BMA model of the Z3D1L model and the Z3D1G model under the propositions Z3 and D1. The Z3 model is the BMA of the Z3D1 model and the Z3D2 model under the hierarch model. At the top-most level, the hierarch model is a BMA of the Z2 and Z3 models.

[56] Table 1 shows the calibration results of the 24 models to obtain the formation dip and the sand-clay cutoff. The mean sand-clay cutoff 0.41 is in agreement with the calculated sand proportion 0.40 from the electrical logs, which implies that the sand-clay cutoff can be interpreted as the probability of occurrence [Chilès and Delfiner, 1999]. While previous studies [Johnson and Dreiss, 1989; Falivene *et al.*, 2007] consider a sand-clay cutoff of 0.5 as a reasonable assumption. The calibration results show that a fixed cutoff 0.5 will result in an underestimation of sand proportion in this case. The minimum, mean, and maximum formation dip for the 24 models are 0.17°, 0.32°, and 0.45°, respectively. This agrees with the geological information that the aquifer system gently dips south



**Figure 4.** The BMA tree of the posterior model probabilities (model weights) and the conditional posterior model probabilities (conditional model weights) for the four uncertain model components. Models with posterior model probabilities less than 1% are not shown in the figure.

[Tomaszewski, 1996] and with the estimated dip  $0.30^\circ \pm 0.06^\circ$  from Griffith [2003].

[57] Given two unknown model parameters and the fitting residual  $Q$ , we use equation (26) to calculate  $BIC_{(ij...lm)}$ . To obtain the BMA tree, we calculate the posterior model probabilities using  $\Delta BIC_{(ij...lm)} = BIC_{(ij...lm)} - BIC_{min}$  for both Occam's window and different variance windows.  $BIC_{min}$  is the minimum BIC value among all models, which is  $BIC_{min} = 67070$  for the best base model Z3D1LExp. The number of data points is 31,500. Table 1 shows  $\Delta BIC$  and posterior model probabilities for base models using Occam's window and different variance windows based on the scaling factors of 1% and 5% significance levels and three different standard deviations  $\sigma_D$  of the fitting residual  $Q$  [Tsai and Li, 2008a, 2008b]. Due to the large data size, Occam's window as expected singles out only the best model. Posterior model probabilities of less influential models increase as the significance levels and the standard deviation  $\sigma_D$  increase, which consecutively decrease the posterior model probabilities of the best models. Adjusting the scaling factor of the variance windows is subject to the analyst decision; and posterior model probabilities are changed as shown in Table 1. However, adjusting the scaling factor does not change the model ranking, but just increases the inclusion of base models [Tsai and Li, 2008a, 2008b; Li and Tsai, 2009; Singh et al., 2010]. Nevertheless, propositions of different variance windows are not mutually exclusive. To illustrate the variance propagation from the base models to the hierarch model, we use a large variance window of 5% and  $3\sigma_D$  for the successive analysis.

#### 4.2. Model Propositions Evaluation Using the BMA Tree

[58] Figure 4 shows the BMA tree for the four uncertain model components with their corresponding propositions. The best branch starts from the hierarch model to Z3 model, to Z3D1 model, and to Z3D1L model. The best branch coincides with the branch of the best base model because best base model has dominant posterior model probability. Two outcomes can be drawn from the BMA tree. First, model dissection through the BMA tree allows for spotting the propositions that result in good models. By looking at the propositions of the best model, three-zone

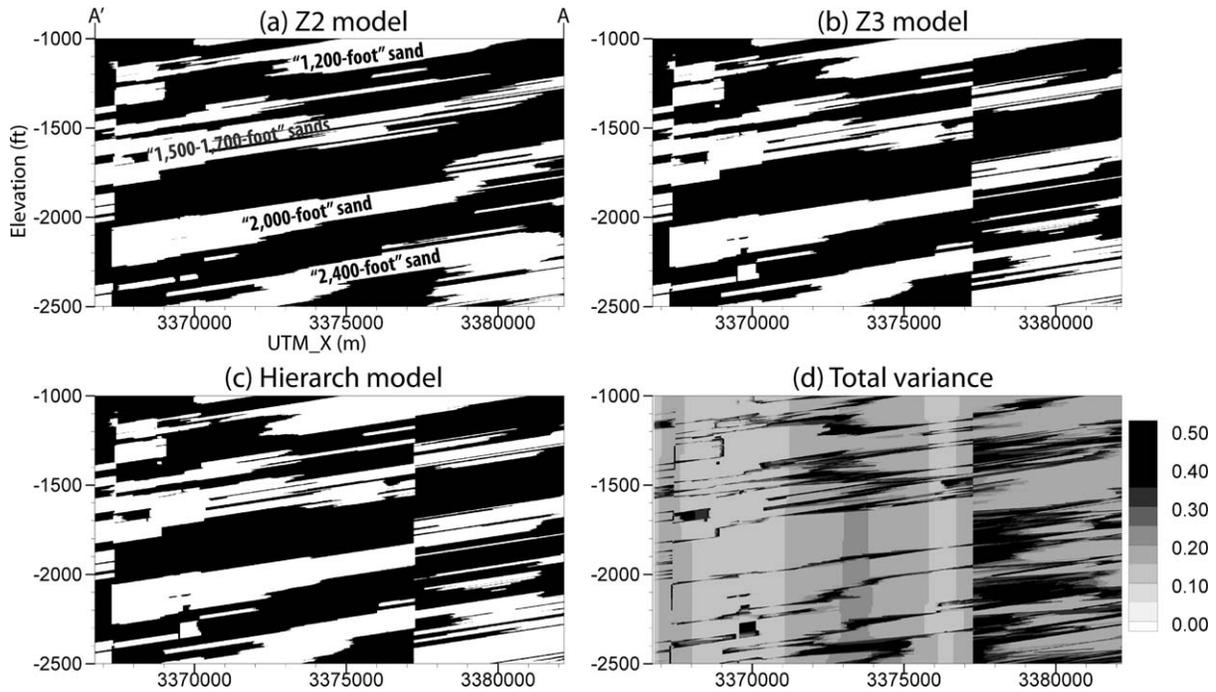
(Z3), data set I (D1), local stationarity (L), and exponential variogram (Exp) generally show higher posterior model probabilities than other competing propositions. As expected, the worst model Z2D2GPen does not share a single proposition with the best model. The second worst model Z2D2LPen shares only local stationarity (L) proposition with the best model.

[59] Second, since the posterior model probabilities in the BMA tree is based on the evidence of data, this may provide an opportunity to recognize the robust propositions. In other words, we examine if the posterior model probabilities can relate to our understanding of the model under study. Starting with the base level of the BMA tree as shown in Figure 4, models with exponential variogram propositions (Exp) have higher posterior model probabilities in most branches, followed by the Gaussian variogram proposition (Gus) and finally the Pentaspheical variogram proposition (Pen). This is not surprising since exponential model is an indicative of a sharp transitions occurring between blocks of different values [Rubin, 2003]. Thus, the exponential function honors this binary conceptualization of sand and clay.

[60] The third level of the BMA tree in Figure 4, which represents the global (G) and local (L) stationarity propositions, shows that the local proposition has consistently higher conditional posterior model probabilities, yet generally the conditional posterior model probabilities of the local proposition and global proposition are not largely different. To pool data for common processing for reasonably defined geological region is not refutable from data a priori, but it can be shown inappropriate a posteriori [Deutsch, 2007]. However, Z2D2G and Z2D2L can be regarded as possible a posteriori since their conditional posterior model probabilities are relatively similar.

[61] The second level of the BMA tree indicates that calibrating the models against the calibration data set I (D1) is more robust than data set II (D2). This is anticipated because D1 is in agreement with the electrical logs interpretation that identifies sand and gravel sequences to belong to sand facies with indicator 1.

[62] The first level of the BMA tree compares the two-zone proposition (Z2) and the three-zone proposition (Z3). The posterior model probability of the Z3 proposition that



**Figure 5.** The BMA model estimates for the cross section  $AA'$  (see Figure 3): (a) Z2 model, (b) Z3 model, and (c) Hierarch model. White areas are sand and black areas are clay. (d) Total model variance for the hierarch model.

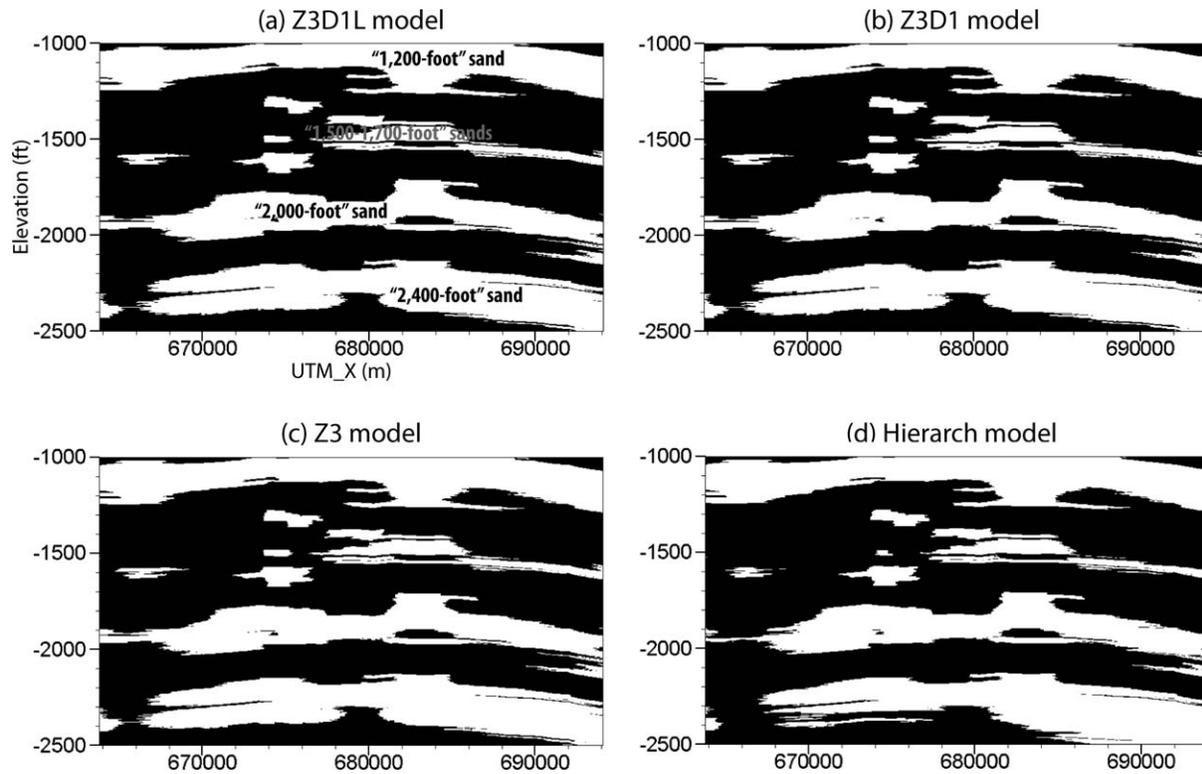
explicitly accounts for the Denham Springs-Scotlandville fault is relatively higher than the Z2 proposition. In Figure 5, we visually evaluate whether the Denham Springs-Scotlandville fault causes sand displacement along the fault plane. The Z2 model in Figure 5a implies a fault throw in the “2000 foot” sand, but shows no obvious displacement in the “1500 foot” sand. The Z3 model in Figure 5b defines the displacement in the “1500 foot” sand and “2000 foot” sand suggesting that the Denham Springs-Scotlandville fault causes sand units displacement along the fault plane. This is in agreement with the higher posterior model probability of the Z3 proposition. It is interesting to see that the hierarch model in Figure 5c is very similar to the Z3 model, yet showing high total model variance around the Denham Springs-Scotlandville fault in Figure 5d due to the Z2 proposition.

### 4.3. Uncertainty Propagation and Prioritization

[63] The total uncertainty as expressed by the total model variance is the summation of the between-model variance and within-model variance. The between-model variance depicts the estimation differences between competing models. By moving to the superior level this total model variance becomes the within-model variance for that level. This section presents the variance propagation of the within-model variance, between-model variance, and total model variance, and aims at prioritizing the uncertain model components based on their corresponding between-model variance. For this purpose, we use the south cross section of the Denham Springs-Scotlandville fault as shown in Figure 6 that follows the fault line shown in Figure 2, but rendered in two dimensions for clarity. The grid spacing is 50 m along the fault line and 1 ft (0.304 m) in the vertical direction.

[64] To trace and understand the patterns of uncertainty propagation, Table 2 shows the mean values of the variances for all BMA models in the BMA tree. Table 2 shows the prediction variances and conditional posterior model probabilities for the BMA models at given levels, which are obtained from child models in the subordinate level. For example, Level 3 shows the results from different variogram propositions; Level 2 shows the results from different stationarity propositions; Level 1 shows the results from different calibration data propositions; and the hierarch level shows the results from different fault propositions. Following the best branch starting from the Z3D1L model to the hierarch model, as expected the total model variance is monotonically increasing because the variances are adding up. This is not necessarily the case for other branches. For example, if the model has high total model variance and lower posterior model probabilities as Z2D2G model, then at the next superior level the between-model variance that averages Z2D2G model and Z2D2L model will be less than the total model variance of Z2D2G model. Similar to the total model variance, the within-model variance depends on its subordinate levels and has a tendency to increase as moving up to superior levels, since it is adding up between-model variances at its subordinate levels. Yet unlike the total model variance, the within-model variance is not necessarily monotonically increasing for the best branch and the within-model variance can decrease depending on the posterior model probability. The best branch in Table 2 illustrates this observation in which the within-model variance of Z3 model is less than the hierarch model.

[65] Unlike the within-model variance and the total model variance, the between-model variance is independent of subordinate levels as indicated by equation (17) and as illustrated in Table 2. This feature allows for prioritizing



**Figure 6.** The BMA model estimates for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model, and (d) Hierarch model. White areas are sand and black areas are clay.

the relative impact of each uncertain model component on the overall model uncertainty. For example, the small between-model variance at level 1 as shown in Table 2 indicates that the calibration data set propositions have insignificant contribution to the overall model uncertainty. The between-model variances of the three other uncertain model components in Table 2 are high indicating that each of them has a large contribution to the overall model uncertainty since the between-model variances are additive as shown in the following discussion. To further understand the uncertainly propagation, we plot the model estimation, within-model variance, between-model variance and total model variance for the best branch.

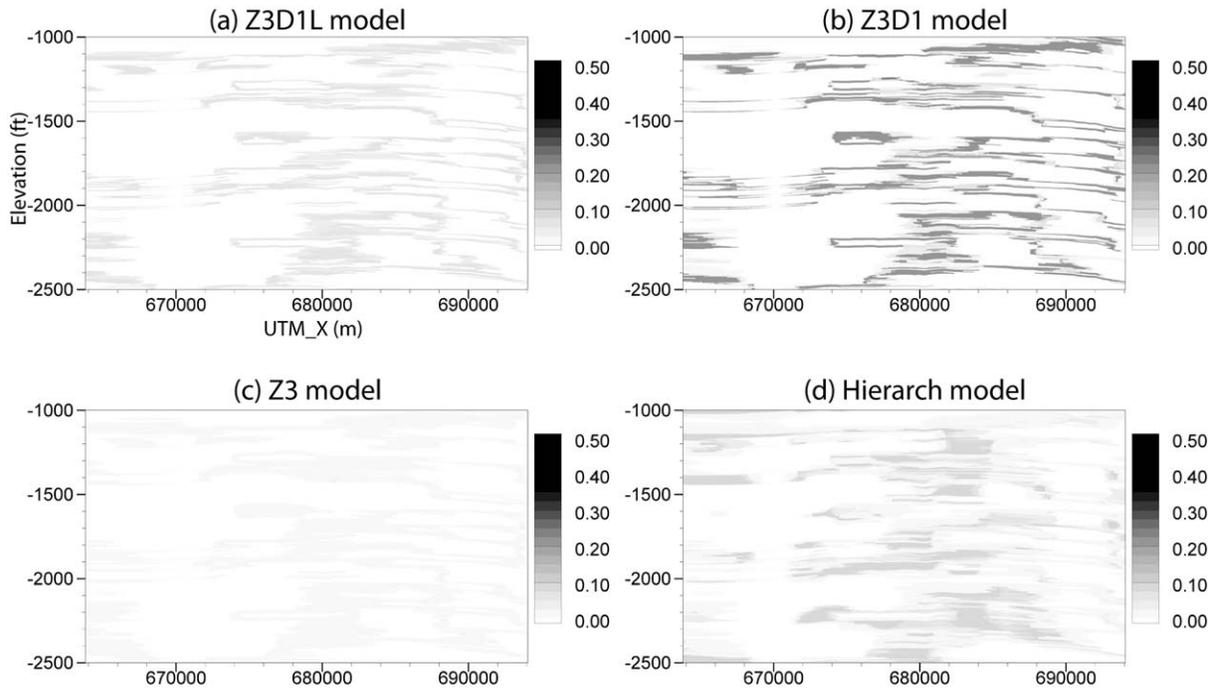
[66] The BMA models in Figure 6 shows the model estimation for sand-clay distribution south of the Denham Springs-Scotlandville. The prediction of Z3D1L model,

Z3D1 model, and Z3 model are almost identical. This indicates that the BMA model prediction with respect to these sources of uncertainty is stable because these models are relatively similar and the best base model dominates the results. Unlike these three BMA models, the hierarch model as shown in Figure 6d is marginally different particularly for the “2400 foot” sand because the Z2 and Z3 propositions produce different estimations as previously illustrated in Figures 5a and 5b.

[67] Figure 7 shows the between-model variance of the four uncertain model components. The Z3D1L model, Z3D1 model, and Z3 model have similar variance patterns, yet with different values. The similar variance patterns indicate again that the best base model dominates the results. High between-model variance indicates that competing propositions are important and the competing

**Table 2.** Mean Values of the Within-Model Variance (WMV), the Between-Model Variance (BMV), and the Total Model Variance (TMV), and the Conditional Posterior Model Probabilities (cPr.) for the Cross Section South of the Denham Springs-Scotlandville Fault

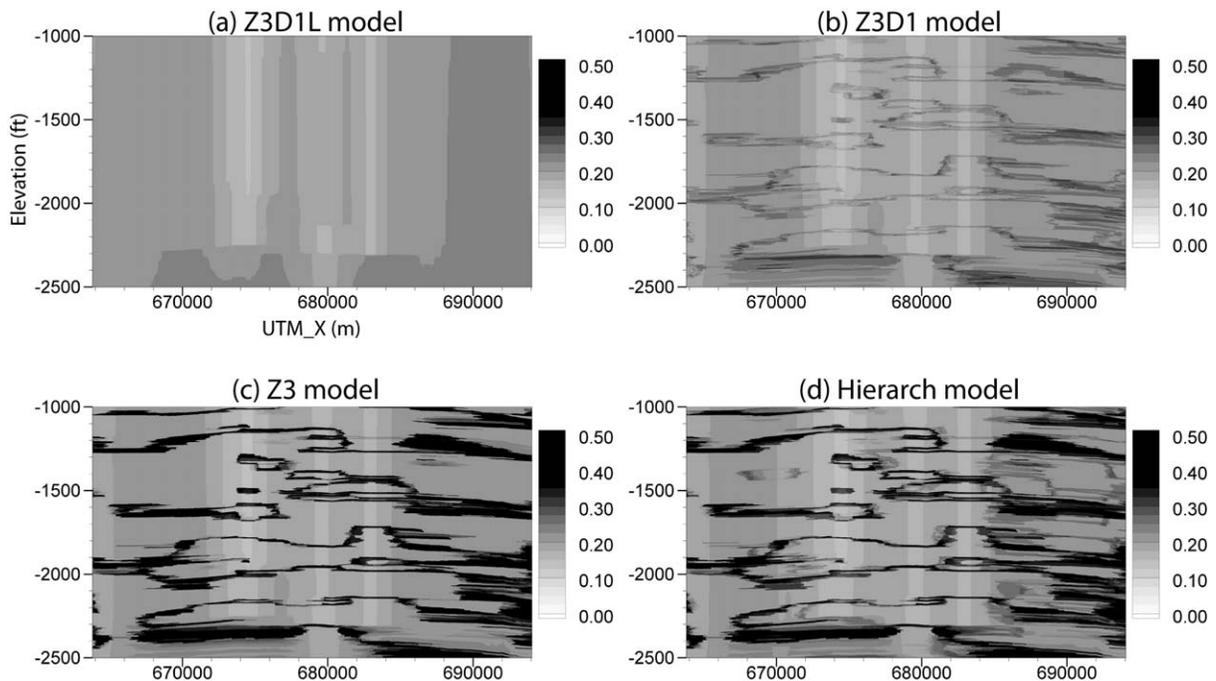
BMA Model	Level 3				Level 2				Level 1				Hierarch Level		
	WMV	BMV	TMV	cPr.	WMV	BMV	TMV	cPr.	WMV	BMV	TMV	cPr.	WMV	BMV	TMV
Z2D1G	0.183	0.020	0.203	0.23											
Z2D1L	0.184	0.020	0.204	0.77	0.204	0.015	0.218	0.72							
Z2D2G	0.187	0.039	0.226	0.47					0.222	0.004	0.226	0.26			
Z2D2L	0.187	0.031	0.218	0.53	0.222	0.010	0.232	0.28							
Z3D1G	0.185	0.017	0.202	0.26									0.244	0.028	0.271
Z3D1L	0.208	0.011	0.220	0.74	0.215	0.033	0.248	0.90							
Z3D2G	0.180	0.013	0.193	0.35					0.248	0.001	0.250	0.74			
Z3D2L	0.209	0.034	0.243	0.65	0.226	0.024	0.250	0.10							



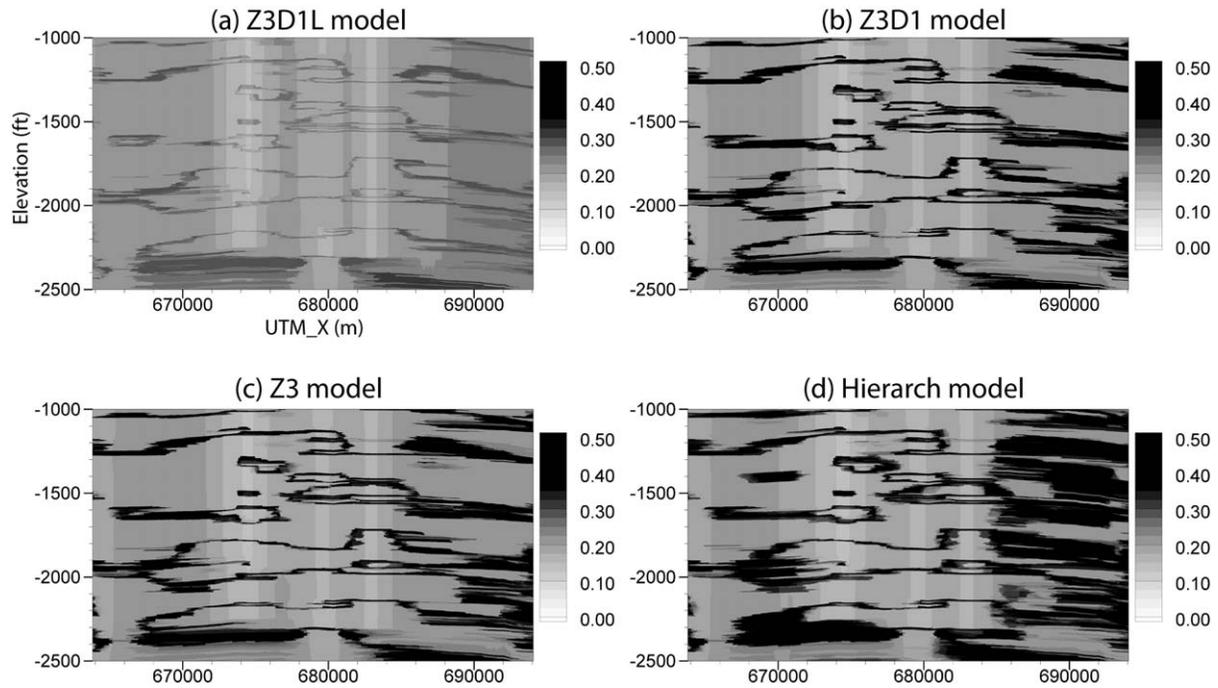
**Figure 7.** The between-model variance for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model, and (d) Hierarch model.

models are considerably different. For example, the local and global stationarity assumptions are both good propositions as indicated by their posterior model probabilities and thus resulted in high between-model variance as shown in Figure 7b. Also, the averaging of the Z2 model and Z3 model as shown in Figure 7d resulted in high between-

model variance, since their estimations are noticeably different. Similarly, small between-model variance is due to the similarity of the competing models or the presence of a dominant competing proposition. For example, using different calibration data sets resulted in similar models with the same dip and cutoff, yet with different fitting errors as



**Figure 8.** The within-model variance for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model, and (d) Hierarch model.



**Figure 9.** The total model variance for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model, and (d) Hierarch model.

shown in Table 1. Also, D1 proposition outperforms D2 proposition as shown by their posterior model probabilities in Figure 4. Thus, the insignificant impact of the calibration data set propositions on overall model uncertainty is due to the combined effect of the two factors.

[68] As shown in Figure 8a, the within-model variance of the Z3D1L model is the average of the variances of the three base models Z3D1LExp, Z3D1LGus, and Z3D1LPen. Regions close to the electrical logs have lower variance. Figure 8b illustrates that the within-model variance at the next level for model Z3D1 is the weighted average of total model variances of its subordinate models Z3D1G and Z3D2L. Similarly, we can obtain the within-model variance for the Z3 model and the hierarch model as shown in Figures 8c and 8d, respectively. Comparing Figure 8c with Figure 8d illustrates the previous remark that the within-model variance does not monotonically increase in value, yet more uncertain regions occur. This provides an interesting remark that the areas of uncertainty will always increase by adding more sources of uncertainty, yet the variance magnitude can decrease.

[69] Figure 9 shows the total model variance that is the sum of the between-model variance in Figure 7 and the within-model variance in Figure 8. Figure 9 shows the monotonic variance increase in value and area for the best branch. Another noticeable remark is that the between-model variance is taking over the within-model variance, which indicates that uncertainty arising from the uncertain model components is higher than the uncertainty arising from the kriging variance.

[70] Model propositions evaluation and prioritization of the uncertain model components as previously discussed are features of the hierarchical BMA, which are not possible to obtain through the collection BMA. However, the

estimation (Figure 6d) and total model variance (Figure 9d) from the hierarch BMA model are identical to those from the collection BMA.

### 5. Conclusions

[71] The hierarchical Bayesian model averaging (HBMA) provides a framework for incorporating competing knowledge about the model data, structure, and parameters to advance our understanding about model prediction and uncertainty. Since uncertainty arises because models are not perfect simulators of reality, it is common to consider multiple models. Similar to collection BMA, the hierarchical BMA utilizes multiple base models for model prediction under Bayesian statistical framework such that the model importance is based on the evidence of data. Unlike collection BMA that results in a single BMA model, the hierarchical BMA develops a hierarchy of BMA models through systematic dissection of uncertain model components. The hierarchy of BMA models allows for uncertainty segregation, comparative evaluation of competing model propositions, and prioritization of uncertain model components.

[72] The HBMA supports the rejection of a single representation of the system in favor of many system representations. The HBMA method illustrates the fact that model uncertainty is likely to be underestimated if only the best model is used. HBMA method explains this observation by distinguishing the within-model uncertainty and the between-model uncertainty for each uncertain model component. Analyzing the uncertainty propagation across different uncertain model components in the BMA tree shows that the within-model uncertainty can increase or decrease depending on the posterior model probabilities. However,

the contribution of the between-model uncertainty is cumulative to the total model variance. Therefore, by adding more sources of uncertainty, which increases the number of uncertain model components and/or the number of corresponding competing propositions, the total uncertainty is increased through the between-model variance. The between-model variance presents an important uncertainty origin that cannot be discarded.

[73] The advantages of using the hierarchical BMA over the collection BMA have been illustrated in the hydrostratigraphic modeling of the complex Baton Rouge aquifer-fault system in Louisiana. Comparative evaluation of the competing propositions for each uncertain model component is completed through the posterior model probabilities and the conditional posterior model probabilities in the BMA tree. The conditional posterior model probabilities of the BMA tree suggest that explicit expression of the Denham Springs-Scotlandville fault is favorable, hydrofacies interpretation for the calibration data set I is considerably better, local geological stationarity due to the presence of the faults is favorable, and the exponential variogram model is preferable. The prioritization of different sources of uncertainty can be carried out through the between-model uncertainty. For this case study, uncertainty arising from different conceptualizations with respect to the Denham Springs-Scotlandville fault is more prominent, which is followed by uncertainty arising from variogram models and stationarity assumptions. Uncertainty arising from different calibration data sets appears insignificant due to small between-model variance. The HBMA model is an epistemic model that heavily relies on data evidence and knowledge advancement. Thus, the understanding of the current Baton Rouge aquifer-fault system is subject to revision shall new data, expert knowledge, model propositions, sources of uncertainty, calibration parameters or statistical inference methods become available.

[74] A key feature of the hierarchical BMA is knowledge update. One mean of knowledge update is to drop 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 propositions at the same level. A second mean of knowledge update is to conduct data collection for further model evaluation until inappropriate propositions show insignificant posterior model probabilities. These are potential applications of the hierarchical BMA.

[75] **Acknowledgments.** This study was supported in part by the National Science Foundation under grant 1045064 and grant/cooperative agreement G10AP00136 from the United States Geological Survey. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS. The authors acknowledge two anonymous reviewers for providing constructive comments which improved the manuscript.

## References

- Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrogeological modeling, *Adv. Water Resour.*, 16(1), 41–51.
- Beven, K. (2000), On model uncertainty, risk and decision making, *Hydrol. Processes*, 14(14), 2605–2606.
- Beven, K. (2005), On the concept of model structural error, *Water Sci. Technol.*, 52(6), 167–175.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279–298.
- Bredehoeft, J. D. (1997), Fault permeability near Yucca Mountain, *Water Resour. Res.*, 33(11), 2459–2463.
- Buono, A. (1983), The Southern Hills regional aquifer system of southeastern Louisiana and southwestern Mississippi, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 83-4189, 43.
- Cardiff, M., and P. K. Kitaniadis (2009), Bayesian inversion for facies detection: An extensible level set framework, *Water Resour. Res.*, 45, W10416, doi:10.1029/2008WR007675.
- Chester, F. M., J. P. Evans, and R. L. Biegel (1993), Internal structure and weakening mechanisms of the San Andreas fault, *J. Geophys. Res.*, 98(B1), 771–786.
- Chilès, J. P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, John Wiley, New York.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010WR009827.
- Demissie, Y. K., A. J. Valocchi, B. S. Minsker, and B. A. Bailey (2009), Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions, *J. Hydrol.*, 364(3-4), 257–271.
- Deutsch, C. V. (2007), A Review of Geostatistical Approaches to Data Fusion, in *Hydrology: Data Integration for Properties and Processes*, edited by D. W. Hyndman, F. D. Day-Lewis, and K. Singha, pp. 7–18, AGU, Washington, D. C.
- Doherty, J., and S. Christensen (2011), Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, *Water Resour. Res.*, 47, W12534, doi:10.1029/2011WR010763.
- Draper, D. (1995), Assessment and propagation of model uncertainty, *J. R. Stat. Soc., Ser. B*, 57(1), 45–97.
- Ellison, A. M. (2004), Bayesian inference in ecology, *Ecol. Lett.*, 7(6), 509–520.
- Engdahl, N. B., G. S. Weissmann, and N. D. Bonal (2010), An integrated approach to shallow aquifer characterization: Combining geophysics and geostatistics, *Comput. Geosci.*, 14(2), 217–229.
- Fairley, J., J. Heffner, and J. Hinds (2003), Geostatistical evaluation of permeability in an active fault zone, *Geophys. Res. Lett.*, 30(18), 1962, doi:10.1029/2003GL018064.
- Falivene, O., L. Cabrera, and A. Saez (2007), Large to intermediate-scale aquifer heterogeneity in fine-grain dominated alluvial fans (Cenozoic As Pontes Basin, northwestern Spain): Insight based on three-dimensional geostatistical reconstruction, *Hydrogeol. J.*, 15(5), 861–876.
- Feyen, L., and J. Caers (2006), Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations, *Adv. Water Resour.*, 29(6), 912–929.
- Foglia, L., S. W. Mehl, M. C. Hill, and P. Burlando (2013), Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, southern Switzerland, *Water Resour. Res.*, 49, 260–282, doi:10.1029/2011WR011779.
- Gaganis, P., and L. Smith (2001), A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resour. Res.*, 37(9), 2309–2322.
- Gaganis, P., and L. Smith (2006), Evaluation of the uncertainty of groundwater model predictions associated with conceptual errors: A per-datum approach to model calibration, *Adv. Water Resour.*, 29(4), 503–514.
- Gaganis, P., and L. Smith (2008), Accounting for model error in risk assessments: Alternatives to adopting a bias towards conservative risk estimates in decision models, *Adv. Water Resour.*, 31(8), 1074–1086.
- Griffith, J. M. (2003), Hydrogeologic framework of southeastern Louisiana, *Water Resour. Tech. Rep. 72*, La. Dep. of Transp. and Dev., Baton Rouge, La.
- Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, 48, W08301, doi:10.1029/2011WR011044.
- Hansen, N., S. D. Muller, and P. Koumoutsakos (2003), Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evol. Comput.*, 11(1), 1–18.
- Hjort, N. L., and G. Claeskens (2003), Frequentist model average estimators, *J. Am. Stat. Assoc.*, 98(464), 879–899.
- Hjort, N. L., and G. Claeskens (2006), Focused information criteria and model averaging for the Cox hazard regression model, *J. Am. Stat. Assoc.*, 101(476), 1449–1464.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Stat. Sci.*, 14(4), 382–401.
- Hooke, R., and T. A. Jeeves (1961), Direct search solution of numerical and statistical problems, *J. Assoc. Comput. Mach.*, 8(2), 212–229.

- Huntzinger, T. L., C. D. Whiteman Jr., and D. D. Knochenmus (1985), Simulation of ground-water movement in the "1,500-and 1,700 foot" aquifer of the Baton Rouge area, Louisiana, *Water Resour. Tech. Rep.* 34, 52 pp., La. Dep. of Transp. and Dev., Off. of Public Works.
- Irving, J., and K. Singha (2010), Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *Water Resour. Res.*, 46, W11514, doi:10.1029/2009WR008340.
- Jacquín, A. P., and A. Y. Shamseldin (2007), Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, *Water Resour. Res.*, 43, W04425, doi:10.1029/2006WR005072.
- Johnson, N. M. (1995), Characterization of alluvial hydrostratigraphy with indicator semivariograms, *Water Resour. Res.*, 31(12), 3217–3227.
- Johnson, N. M., and S. J. Dreiss (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., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resour. Res.*, 32(9), 2617–2658.
- Li, X., and F. T.-C. Tsai (2009), Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multithreshold, *Water Resour. Res.*, 45, W09403, doi:10.1029/2008WR007488.
- McCulloh, R. P., and P. V. Heinrich (2012), Surface faults of the south Louisiana growth-fault province, the *Geological Society of America Special Papers*, 493, 37–49.
- Meyer, R. R., and A. N. Turcan Jr. (1955), Geology and ground-water resources of the Baton Rouge area, Louisiana, U.S. Geol. Surv. Water Supply Pap., 1296, 138.
- Morales-Casique, E., S. P. Neuman, and V. V. Vesselinov (2010), Maximum likelihood Bayesian averaging of airflow models in unsaturated fractured tuff using Occam and variance windows, *Stochastic Environ. Res. Risk Assess.*, 24(6), 863–880.
- Mugunthan, P., and C. A. Shoemaker (2006), Assessing the impacts of parameter uncertainty for computationally expensive groundwater models, *Water Resour. Res.*, 42, W10428, doi:10.1029/2005WR004640.
- Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, 17(5), 291–305.
- Neuman, S. P., and P. J. Weirenga (2003), A comprehensive strategy of hydrogeologic modelling and uncertainty analysis for nuclear facilities and sites, NUREG/CR-6805, U.S. Nucl. Regul. Comm., Off. of Nucl. Regul. Res., Washington, D. C.
- Nishikawa, T., A. J. Siade, E. G. Reichard, D. J. Ponti, A. G. Canales, and T. A. Johnson (2009), Stratigraphic controls on seawater intrusion and implications for groundwater management, Dominguez Gap area of Los Angeles, California, USA, *Hydrogeol. J.*, 17(7), 1699–1725.
- Nowak, W., F. P. J. de Barros, and Y. Rubin (2010), Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain, *Water Resour. Res.*, 46, W03535, doi:10.1029/2009WR008312.
- Poeter, E., and D. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Ground Water*, 43(4), 597–605.
- Raftery, A. E. (1995), Bayesian model selection in social research, *Sociol. Methodol.*, 25, 111–163.
- Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, 29(11), 1586–1597.
- Rojas, R., L. Feyen, and A. Dassargues (2008), Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging, *Water Resour. Res.*, 44, W12418, doi:10.1029/2008WR006908.
- Rojas, R., L. Feyen, and A. Dassargues (2009), Sensitivity analysis of prior model probabilities and the value of prior knowledge in the assessment of conceptual model uncertainty in groundwater modelling, *Hydrol. Processes*, 23(8), 1131–1146.
- Rojas, R., O. Batelaan, L. Feyen, and A. Dassargues (2010a), Assessment of conceptual model uncertainty for the regional aquifer Pampa del Tamaraugal—North Chile, *Hydrol. Earth Syst. Sci.*, 14(2), 171–192.
- Rojas, R., L. Feyen, O. Batelaan, and A. Dassargues (2010b), On the value of conditioning data to reduce conceptual model uncertainty in groundwater modeling, *Water Resour. Res.*, 46, W08520, doi:10.1029/2009WR008822.
- Rojas, R., S. Kahunde, L. Peeters, O. Batelaan, L. Feyen, and A. Dassargues (2010c), Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling, *J. Hydrol.*, 394(3-4), 416–435.
- Rollo, J. R. (1969), Salt-water encroachment in aquifers of the Baton Rouge area, Louisiana, *Water Resour. Bull.* 13, 45 pp., La. Geol. Surv. Baton Rouge, La.
- Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, New York.
- Salve, R., and C. M. Oldenburg (2001), Water flow within a fault in altered nonwelded tuff, *Water Resour. Res.*, 37(12), 3043–3056.
- Seifert, D., T. O. Sonnenborg, J. C. Refsgaard, A. L. Hojberg, and L. Trolborg (2012), Assessment of hydrological model predictive ability given multiple conceptual geological models, *Water Resour. Res.*, 48, W06503, doi:10.1029/2011WR011149.
- Singh, A., S. Mishra, and G. Ruskauff (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, *Water Resour. Tech. Rep.* 59, 44 pp., La. Dep. of Transp. and Dev. Baton Rouge, La.
- Torak, L. J., and C. D. Whiteman (1982), Applications of digital modeling for evaluating the ground-water resources of the "2,000-foot" sand of the Baton Rouge area, Louisiana, *Water Resour. Tech. Rep.* 27, 87 pp., La. Dep. of Transp. and Dev., Off. of Public Works.
- Trolborg, M., W. Nowak, N. Tuxen, P. L. Bjerg, R. Helmgig, and P. J. Binning (2010), Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework, *Water Resour. Res.*, 46, W12552, doi:10.1029/2010WR009227.
- Tsai, F. T.-C. (2010), Bayesian model averaging assessment on groundwater management under model structure uncertainty, *Stochastic Environ. Res. Risk Assess.*, 24(6), 845–861.
- Tsai, F. T.-C., and X. Li (2008a), Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window, *Water Resour. Res.*, 44, W09434, doi:10.1029/2007WR006576.
- Tsai, F. T.-C., and X. Li (2008b), Multiple parameterization for hydraulic conductivity identification, *Ground Water*, 46(6), 851–864.
- Tsai, F. T.-C., and X. Li. (2010), Reply to comment by Ming Ye et al. on "Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window," *Water Resour. Res.*, 46, W02802, doi:10.1029/2009WR008591.
- von Bertalanffy, L. (1968), *General System Theory: Essays on its Foundation and Development*, revised ed., George Braziller, New York.
- Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochastic Environ. Res. Risk Assess.*, 19(6), 378–387.
- Williamson, J. (2005), *Bayesian Nets and Causality: Philosophical and Computational Foundations*, Oxford Univ. Press, New York.
- Wingle, W. L., and E. P. Poeter (1993), Uncertainty associated with semivariograms used for site simulation, *Ground Water*, 31(5), 725–734.
- Wöhling, T., and J. A. Vrugt (2008), Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models, *Water Resour. Res.*, 44, W12432, doi:10.1029/2008WR007154.
- Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, 40, W05113, doi:10.1029/2003WR002557.
- Ye, M., S. P. Neuman, P. D. Meyer, and K. Pohlmann (2005), Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff, *Water Resour. Res.*, 41, W12429, doi:10.1029/2005WR004260.
- Ye, M., D. Lu, S. P. Neuman, and P. D. Meyer (2010a), Comment on "Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window" by F. T.-C. Tsai and X. Li, *Water Resour. Res.*, 46, W02801, doi:10.1029/2009WR008501.
- Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohl, and D. M. Reeves (2010b), A model-averaging method for assessing groundwater conceptual model uncertainty, *Ground Water*, 48(5), 716–728.