

CHARACTERIZATION AND UNCERTAINTY ANALYSIS
OF SILICICLASTIC AQUIFER-FAULT SYSTEM

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Civil and Environmental Engineering

by

Ahmed Saad Elshall

B.S. The American University in Cairo, 2003

M.S. University of Tübingen, 2009

December 2013

Acknowledgements

Dr. Frank Tsai provided the generalized parameterization method (see Section 3.1) and the hierarchical BMA Bayesian model averaging method (see Section 3.3). In addition, Dr. Frank Tsai carried out all the geophysical data interpretation (see Section 4.1.1) and prepared the GIS maps (see Figure 1, Figure 3 and Figure 33). The PhD students at the Civil and Environment Engineering Department Mr. Nima Chitsazan, Mr. Ihsan Beigi, Mr. Hai Pham and Mr. Amir Mani assisted with the geophysical data processing. The PhD student Mr. Hai Pham developed the grid of the groundwater flow models (see Section 6.2.1, Figure 32). Dr. Le Yan from the High Performance Computing parallelized the model calibration algorithm (see Section 6.2.2). I acknowledge Dr. Frank Tsai, Dr. Jeffrey Hanor, Dr. Le Yan and Mr. Hai Pham for their contribution to and for being coauthors on the four peer-reviewed manuscripts and the 13 conference proceedings, which were carried out during the study period. Acknowledgment is due to the PhD committee members Dr. John Pardue, Dr. Jeffrey Hanor, Dr. Carol Wicks and Dr. Robert Reigh for their support. I am thankful to the staff members Mrs. Janet Labatut and Mrs. Julie Mueller for their administrative support and to Mr. Dave Robertson and Mr. Al Pawlowski for their technical support. I acknowledge LSU HPC and LONI for permitting the use of their HPC clusters. This work is funded by the National Science Foundation–Hydrologic Sciences, the U.S. Geological Survey–National Institute for Water Resources, the Louisiana Water Resources Research Institute and Louisiana State University–Graduate School Supplementary Award. I am very thankful to Dr. Frank Tsai for his meticulous supervision and generous assistance without which this work would not have been possible. Finally, I am indebted to my parents and family for their priceless guidance. Thank you very much.

Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iii
Abstract.....	vi
1 Introduction.....	1
2 Literature review.....	8
2.1 Baton Rouge aquifer-fault system.....	8
2.2 Hydrofacies architecture modeling using indicator geostatistics.....	9
2.3 Model calibration and uncertainty quantification using CMA-ES.....	12
2.4 Constructive epistemic modeling using hierarchical Bayesian model averaging.....	15
2.4.1 Hierarchical Bayesian model averaging.....	15
2.4.2 Constructive epistemic modeling under Bayesian paradigm.....	19
3 Methods.....	24
3.1 Indicator geostatistics.....	24
3.2 CMA-ES.....	25
3.2.1 CMA-ES algorithm.....	25
3.2.2 Review of CMA-ES with respect to comparison algorithms.....	30
3.3 Hierarchical Bayesian model averaging.....	34
3.3.1 Terminology and notation.....	34
3.3.2 Posterior model probability and conditional posterior model probability.....	36
3.3.3 Prediction means and prediction covariances.....	40
3.3.4 Computation of posterior model probability with variance window.....	42
3.3.5 Similarities and differences between collection BMA and hierarchical BMA.....	47
4 Constructive epistemic modeling of hydrofacies architecture under Bayesian paradigm.....	49
4.1 Case Study: Hydrofacies architecture model of the Baton Rouge aquifer-fault system.....	49
4.1.1 Model data.....	49
4.1.2 Model data and model structure uncertainty.....	52
4.1.3 Model parameters and calibration.....	55
4.2 Results and Discussion.....	57
4.2.1 Calibration and BIC.....	57
4.2.2 Model propositions evaluation using the BMA tree.....	60
4.2.3 Uncertainty propagation and prioritization.....	63
4.3 Conclusions.....	69
5 Hydrogeological characterization of the Baton Rouge aquifer-fault system.....	72
5.1 Case Study: Hydrofacies architecture model of the Baton Rouge aquifer-fault system.....	72
5.1.1 Hydrofacies architecture model.....	72
5.1.2 Model parameters and calibration.....	73
5.2 Results and discussion.....	76
5.2.1 Calibration results.....	76

5.2.2 Leaky faults	80
5.2.3 Quantification of structural geology parameters	82
5.2.4 Interconnections between aquifer units	89
5.2.5 Baton Rouge aquifer-fault connections for saltwater intrusion	91
5.3 Conclusions	94
6 Groundwater flow model calibration and uncertainty quantification using CMA-ES	97
6.1 Synthetic groundwater flow problem	97
6.1.1 Design of the synthetic problem	97
6.1.2 Ill-posedness and search difficulties	98
6.1.3 Model parameters and calibration	100
6.1.4 Algorithms tuning	101
6.1.5 Performance comparison	104
6.1.6 Parallel versus sequential implementation	106
6.1.7 Covariance matrix for Monte Carlo sampling	108
6.2 “2,000-foot” sand groundwater flow problem	112
6.2.1 Model parameters and calibration	112
6.2.2 Parallel calibration using high performance computing	115
6.2.3 Speedup of parallel runs	116
6.2.4 Covariance matrix for Monte Carlo sampling	118
6.3 Conclusions	119
7 Constructive epistemic modeling of groundwater flow under Bayesian paradigm	122
7.1 Case Study: Groundwater flow model of the “2,000-foot” sand	122
7.1.1 Geological structure uncertainty	122
7.1.2 Prior model probabilities from geological models	124
7.1.3 Boundary condition uncertainty	127
7.1.4 Model parameters and calibration	128
7.1.5 Quantification of within-model variance	129
7.1.6 High performance computing for model calibration and variance quantification ...	129
7.2 Results and discussion	130
7.2.1 Model calibration and within-model variance quantification	130
7.2.2 BIC calculation	133
7.2.3 Model propositions evaluation	133
7.2.4 Uncertainty propagation and prioritization	135
7.2.5 Temporal and spatial distribution of head prediction and variance	136
7.2.6 Knowledge update	140
7.2.7 Critical issues in implementing hierarchical BMA	141
7.3 Conclusions	142
8 What do we mean by groundwater model uncertainty?	145
9 Conclusions	147
10 References	152

Appendix: Open access permissions.....	168
Vita.....	179

Abstract

The complex siliciclastic aquifer system underneath the Baton Rouge area, Louisiana, USA, is fluvial in origin. The east-west trending Baton Rouge fault and Denham Springs-Scotlandville fault cut across East Baton Rouge Parish and play an important role in groundwater flow and aquifer salinization. To better understand the salinization underneath Baton Rouge, it is imperative to study the hydrofacies architecture and the groundwater flow field of the Baton Rouge aquifer-fault system. This is done through developing multiple detailed hydrofacies architecture models and multiple groundwater flow models of the aquifer-fault system, representing various uncertain model propositions. The hydrofacies architecture models focus on the Miocene-Pliocene depth interval that consists of the “1,200-foot” sand, “1,500-foot” sand, “1,700-foot” sand and the “2,000-foot” sand, as these aquifer units are classified and named by their approximate depth below ground level. The groundwater flow models focus only on the “2,000-foot” sand. The study reveals the complexity of the Baton Rouge aquifer-fault system where the sand deposition is non-uniform, different sand units are interconnected, the sand unit displacement on the faults is significant, and the spatial distribution of flow pathways through the faults is sporadic. The identified locations of flow pathways through the Baton Rouge fault provide useful information on possible windows for saltwater intrusion from the south. From the results we learn that the “1,200-foot” sand, “1,500-foot” sand and the “1,700-foot” sand should not be modeled separately since they are very well connected near the Baton Rouge fault, while the “2,000-foot” sand between the two faults is a separate unit. Results suggest that at the “2,000-foot” sand the Denham Springs-Scotlandville fault has much lower permeability in comparison to the Baton Rouge fault, and that the Baton Rouge fault plays an important role in the aquifer salinization.

1 Introduction

The water withdrawal in Baton Rouge, Louisiana in 2010 was approximately 629,000 m³/day such that approximately 88% is groundwater and the rest is surface water [Sargent, 2012]. Baton Rouge relies on high-quality and low-cost groundwater for both municipal and industrial use. Municipal water supply in Baton Rouge is 100% dependent on groundwater, and approximately 78% of the industrial water use in Baton Rouge is groundwater [Sargent, 2012]. The Southern Hills regional aquifer system covers Baton Rouge and the surrounding parishes. The aquifer system consists of sequence of aquifers and aquicludes extending to a depth of 3000 ft. (900 m) [Tomaszewski , 1996]. These aquifer units are amalgamated fluvial sand bodies [Chamberlain, 2012]. Meyer and Turcan [1955] classified and named these aquifer units by their approximate depth below ground level in Baton Rouge industrial district. The Baton Rouge fault system, which consists of the Baton Rouge fault and the Denham Springs-Scotlandville fault (Tepetate fault) as shown in Figure 1, is an east-west trending fault system that crosscuts the aquifer and aquiclude sequence [McCulloh and Heinrich, 2012]. The Baton Rouge fault crosses the aquifer system separating a sequence of fresh and brackish aquifers at the north and south of the fault, respectively. Prior to heavy pumping water flow in the aquifer system was from north to south following the natural gradient [Elshall et al., 2013]. However, heavy ground water pumping reversed the flow direction resulting in salt water intrusion from south of Baton Rouge fault [Morgan and Winner, 1964; Anderson, 2012], suggesting that Baton Rouge fault is currently acting as a conduit and barrier fault [Bense and Person, 2006; Hanor et al., 2011]. To better understand the salinization underneath Baton Rouge city, it is imperative to study the hydrofacies architecture and the groundwater flow field of the Baton Rouge aquifer-fault system. The ultimate goal of this study is to develop a scientific sound groundwater model for the further salt water intrusion study. This is done in this study through developing multiple detailed

hydrofacies architecture models and multiple groundwater flow models of the aquifer-fault system representing several uncertain model propositions, according to the following four research steps.

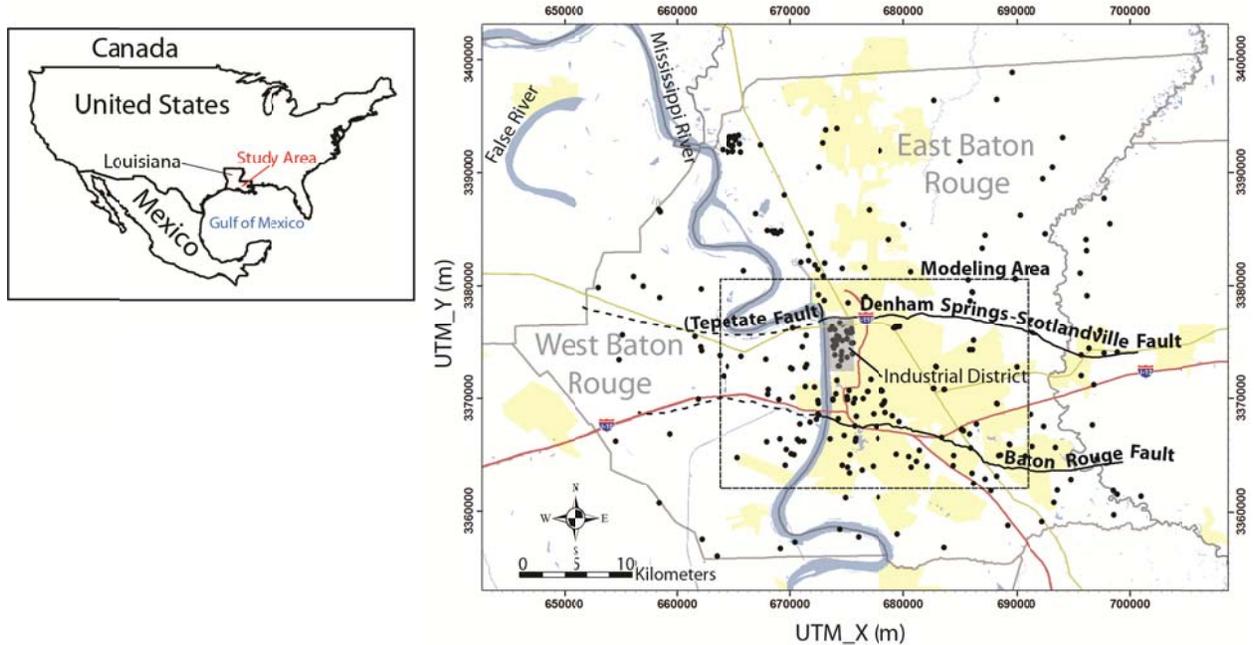


Figure 1 Map of the study area in the Universal Transverse Mercator (UTM) coordinate system. Black dots represent the location of electrical well logs, which are used for the hydrofacies architecture reconstruction. The bold solid lines are fault lines identified by the surface expression [McCulloh and Heinrich, 2012]. The bold dashed lines are the approximate surface locations of the faults [Griffith, 2003]. The yellow areas are urban areas, the grey lines are parish borders, the red lines are interstate freeways, the green lines are US highways, and the blue areas and lines are water bodies [Elshall, et al., 2013].

The first research step is the constructive epistemic modeling of the hydrofacies architecture of the Baton Rouge aquifer-fault system [Tsai and Elshall, 2013]. Analysts are often faced with various candidate 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? 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 BMA Bayesian model averaging (BMA), the study shows that segregating different uncertain model components through a BMA trees 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 BMA representation of the between-model variance facilitates the prioritization of the contribution of each uncertain model component to the overall model uncertainty.

The study illustrates these concepts using the hydrofacies architecture model of the of the Baton Rouge aquifer-fault system, which is based on indicator geostatistics. Due to uncertainty in model data, structure and parameters, multiple possible hydrofacies architecture 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 uncertainty, 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 candidate model propositions result in 24 base models. The study shows that the systematic dissection of the uncertain model components along with their corresponding candidate propositions allows for detecting the robust model propositions and the major sources of uncertainty.

The second research step is the hydrogeological characterization of the Baton Rouge aquifer-fault system [Elshall et al., 2013]. The complex siliciclastic aquifer system underneath the Baton Rouge area is fluvial in origin [Chamberlain, 2012] and is characterized by strongly

binary heterogeneity of sand units and mudstones as pervious and impervious hydrofacies. Using the robust model propositions as identified from the first research step, the study reconstructs the Baton Rouge aquifer-fault system architecture for the Miocene-Pliocene depth interval that consists of the “1,200-foot” sand to the “2,000-foot” sand. The study will provide essential information on the Baton Rouge aquifer-fault system, which has never been studied in such detail in the past. First, the resulting hydrofacies architecture will provide a detailed distribution of the thickness, lateral extent and depth of different sand units. The formation dip, sand offset on the faults, and volumetric sand proportion can be quantified. The hydrofacies architecture will also improve the understanding of potential interconnections among different sand units resulting from the complexity of fluvial deposition. Second, the study will provide essential information on the flow pathways across the Baton Rouge fault and the Denham Springs-Scotlandville fault. Mapping the architecture of the two faults has never been done before. In addition, the result will provide essential information on identifying potential flow pathways through the Baton Rouge fault with regard to saltwater encroachment. Third, the reconstructed hydrofacies architecture is used as the geological structure of the groundwater flow model, which is the subject of the next two research steps.

The third research step is the calibration and uncertainty quantification of the groundwater flow model [Elshall et al., submitted] using the Covariance Matrix Adaptation - Evolution Strategy (CMA-ES) [Hansen and Ostermeier, 2001; Hansen et al., 2003]. The inverse groundwater problem is a rugged, nonseparable and noisy function since it involves solving second order nonlinear partial differential equations with sources and sinks. Derivative calibration algorithms may fail to reach a near global solution due to stagnation at a local solution. This study presents the Covariance Matrix Adaptation-Evolution Strategy (CMA-ES)

as a global-local calibration algorithm that avoids entrapment at a local solution, and enhances the search properties. Evaluation of CMA-ES with five commonly used calibration algorithms on a synthetic groundwater calibration problem shows that CMA-ES improves the solution precision. Second, the study shows that the empirically estimated covariance matrix is precise and can be used for Monte Carlo sampling to quantify the parameters related uncertainty. Third, the CMA-ES is readily amendable to embarrassingly parallel master-slave computation. The parallel CMA-ES, which substantially reduced the calibration, permitted the use of a realistic groundwater model that is based on the actual geology. Note that while the hydrofacies architecture models covers the “1,200-foot” sand to the “2,000-foot” sand, the groundwater model considers only the “2,000-foot” sand.

The fourth research step is the constructive epistemic modeling of the groundwater flow in the “2,000-foot” sand [Elshall and Tsai, submitted]. The hierarchical BMA allows for segregating, prioritizing, and evaluating different sources of uncertainty and their corresponding candidate propositions through a hierarchy of BMA models. The study considers four uncertain model components. With respect to geological structure uncertainty, the study considers three candidate methods for reconstructing the hydrofacies architecture of the aquifer-fault system, and two different formation dips. The study considers 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 particular understanding, as BMA trees of model weights, prediction and variance serves as a learning tool. For example, the study shows that the geological related uncertainty is larger than boundary condition uncertainty; the model structure uncertainty is

larger than parameter uncertainty; and the best hydrofacies architecture model does not necessarily yield the best groundwater flow model.

The aforesaid brief discussion of the four research steps shows that the study has three main methods. The first method is the indicator geostatistics for hydrofacies architecture reconstruction [Elshall et al., 2013]. Indicator geostatistics is used in the four research steps. Indicator geostatistics is used in first and second research steps to reconstruct the hydrofacies architectures for the constructive epistemic modeling of the hydrofacies architectures and for the hydrogeological characterization of the Baton Rouge aquifer-fault system, respectively. These hydrofacies architectures are then used in the third and fourth research steps as the geological structure of the “2,000-foot” sand groundwater flow model. The second method is the CMA-ES algorithm for model calibration and uncertainty quantification [Elshall et al., submitted]. CMA-ES algorithm is used in the four research steps for model calibration, and is used in the last three research steps for uncertainty quantification. The third method is the hierarchical BMA for constructive epistemic modeling [Tsai and Elshall, 2013; Elshall and Tsai, submitted]. The hierarchical BMA is used in the first and fourth research steps for the constructive epistemic modeling of the hydrofacies architectures of the Baton Rouge aquifer-fault system and the groundwater flow in the “2,000-foot” sand, respectively.

The dissertation is organized as follows. Section 2 presents a literature review about the Baton Rouge aquifer-fault system and the three methods used. Section 3 presents the mathematical formulations and a critical evaluation for each method. Section 4 presents the first research step that is the constructive epistemic modeling of hydrofacies architecture under Bayesian paradigm. Section 5 presents the second research step that is the hydrogeological characterization of the Baton Rouge aquifer-fault system. . Section 6 presents the third research

step that is the groundwater flow model calibration and uncertainty quantification using CMA-ES. Section 7 presents the fourth research step that is the constructive epistemic modeling of groundwater flow under Bayesian paradigm . Section 8 and Section 9 provide general discussion and conclusions about the study as a whole.

Most of this work is published or submitted for publication in Tsai and Elshall [2013], Elshall et al. [2013], Elshall et al. [submitted] and Elshall and Tsai [submitted]. Section 2.4.1, Section 3.3 and Section 4 are published with some modifications in Tsai and Elshall [2013] (see Appendix for permission). Section 2.1, Section 2.2, Section 3.1 and Section 5 are published with some modifications in Elshall et al. [2013] (see Appendix for permission). Section 2.3, Section 3.2 and Section 6 are submitted for publication [Elshall et al., submitted]. Section 2.4.2 and Section 7 are submitted for publication [Elshall and Tsai, submitted].

2 Literature review

2.1 Baton Rouge aquifer-fault system

This section is reproduced with modifications from Elshall et al. [2013].

The Baton Rouge aquifer system in southeastern Louisiana, USA, is part of the Southern Hills regional aquifer system [Buono, 1983] and is a siliciclastic aquifer system consisting of a complexly interbedded series of fluvial sand and clay units [Chamberlain, 2012] that thicken and dip southward [Tomaszewski, 1996]. This sequence of aquifers and aquitards extends to a depth of 3,000 feet (914.4 m) in the Baton Rouge area. According to Chamberlain [2012] the vertical alternation of sand-dominated units and clay-dominated units reflects cyclic variations in sea-level, with amalgamated fluvial sand bodies having been generally deposited during sea-level lowstands and mudstones during transgressive highstands. The sand units have variable thicknesses ranging from 20-300 feet (6.10-91.44 m) [Griffith, 2003]. The study area shown in Figure 1 focuses on late Miocene-Pliocene deposits of the “1,200-foot” sand, the “1,500-foot” sand, the “1,700-foot” sand and the “2,000-foot” sand. The Baton Rouge fault system, which consists of the Baton Rouge fault and the Denham Springs-Scotlandville fault (Tepetate fault), is an east-west trending listric fault system that crosscuts this aquifer and aquitard sequence [McCulloh and Heinrich, 2012]. The low permeability of the Baton Rouge fault historically separates the sequence of freshwater and brackish aquifers immediately north and south of the fault, respectively. The natural direction of water flow in the aquifer system is southward. However, heavy public supply and industrial groundwater pumping reversed the flow direction near the Baton Rouge fault and has resulted in saltwater encroachment across the fault [Morgan and Winner, 1964; Meyer and Rollo, 1965; Rollo, 1969; Whiteman, 1979; Tomaszewski and Anderson, 1995; Tomaszewski, 1996; Griffith and Lovelace, 2003; Prakken, 2004; Tsai and Li,

2008a, 2008b, Li and Tsai 2009; Tsai, 2010], suggesting that Baton Rouge fault is currently acting as a conduit-barrier fault [Bense and Person, 2006; Hanor et al., 2011].

The Baton Rouge fault system is composed of the Baton Rouge fault and the Denham Springs-Scotlandville fault. The Baton Rouge fault is listric growth fault [McCulloh and Heinrich, 2012] that crosscuts the aquifer units causing the aquifers to be offset up to 344 ft. (105 m) at the top of the “2,000-foot” sand [Durham and Peeples, 1956]. The Baton Rouge fault was originally active from Late Eocene-Early Oligocene until the Late Oligocene [Murray, 1961; McCulloh and Heinrich, 2012]. The fault was reactivated in the Plio–Pleistocene [Durham and Peeples, 1956; Murray, 1961; McCulloh and Heinrich, 2012]. Little is known about the Denham Springs-Scotlandville fault, and the displacement of the aquifer units on this fault is not well characterized. Rollo’s [1969] hydrofacies mapping of the Baton Rouge aquifer system did not recognize the presence of the Denham Springs-Scotlandville fault, and thus the aquifer units north of the Baton Rouge fault appear continuous on his cross sections.

2.2 Hydrofacies architecture modeling using indicator geostatistics

This section is reproduced with modifications from Elshall et al., [2013].

Constructing hydrofacies architecture depends on the type and density of hydrofacies data and the scale of heterogeneity characterization. Different scales include the sequence hydrostratigraphic scale [Miller et al., 2000; Scharling et al., 2009; Faunt et al., 2010], the hydrofacies assemblage scale [Weissmann et al., 1999; Trevisani and Fabbri, 2010], the hydrofacies unit scale [Zappa et al., 2006; Engdahl et al., 2010] and combinations of different heterogeneity scales [Weissmann and Fogg, 1999; Proce et al., 2004; Comunian et al., 2011]. This study focuses on the sequence hydrostratigraphic scale to obtain a detailed distribution of the thickness, lateral extent and depth of sand units underneath Baton Rouge. Following the

classification of scales [Koltermann and Gorelick, 1996], this scale is the same as the depositional environment scale, which is larger than the channel scale but smaller than the basin scale. This is also the same as the hydrofacies assemblages complex of Rubin [2003], which exhibits strong bimodal heterogeneity. A bimodal heterogeneity of pervious and impervious formations is conceptualized for the Baton Rouge aquifer system, in which sand assemblages complex and clay assemblages complex exhibit strong bimodal heterogeneity. For detailed descriptions of the depositional environmental scale of characterization and the concept of strong bimodal heterogeneity, the reader is Rubin [2003, Figure 2.9].

The indicator geostatistics are particularly helpful in the Baton Rouge aquifer setting, since they are able to handle strongly bimodal heterogeneity. For the depositional environment scale of characterization, variogram-based geostatistics can still be a choice over the multiple-point training images geostatistics [Caers, 2001; Strebelle, 2002] when there are no predefined patterns of the shapes of the aquifer units in practice [Li et al., 2012a], as it is the case in this study area. Chamberlain [2012] interpreted these aquifer units as zones of amalgamated sand bodies that were created by fluvial aggradation following changes in sea levels and thus they are morphologically complex sand units with highly variable erosional unconformities. Since these sand units have irregular depositional and erosional patterns, indicator variogram-based geostatistics [Johnson and Dreiss, 1989; Desbarats and Bachu, 1994; Johnson, 1995; Trevisani and Fabbri, 2010] is used for indicator hydrofacies architecture modeling in this study. The indicator variograms as described by Journel [1983] are structurally informative [Johnson and Dreiss, 1989]. By empirically acknowledging the random and structured qualities of geological geometry, indicator variograms can depict sharp transitions in the spatial field [Johnson, 1995].

This study employs the generalized parameterization method [Tsai and Yeh, 2004; Tsai, 2006] through an inversion scheme to obtain the hydrofacies architecture. The generalized parameterization (GP) is a combination of indicator kriging (IK) and indicator zonation (IZ) for providing flexible nonsmooth conditional estimates. Indicator zonation divides the space into a number of non-overlapping zones based on an indicator function and provides sharp edged estimations [e.g. Tsai, 2009]. On the other hand, indicator kriging provides smooth estimations. Since boundaries between sand and clay units are neither smooth, nor blocky as a result of fluvial depositional processes, the GP is able to estimate the nonsmooth distribution of sand and clay units by combining both features of indicator kriging and indicator zonation through weighting coefficients. A second problem, which is peculiar to indicator geostatistics methods, is that the facies cutoff that rounds the model estimates into binary values to produce the indicators is unknown. To simplify this problem previous studies [Johnson and Dreiss, 1989; Falivene et al., 2007] have considered a cutoff value of 0.5 as a reasonable assumption. Yet fixed cutoff value 0.5 results in an underestimation of the facies that exists in less proportion. Thus, this unknown model parameter needs to be calibrated. Thirdly, to calculate the structure of the experimental variogram, it is important to establish correct correlations among well logs to account for the spatial continuity of the deposits. Different formation dips have a significant effect on the selection of data points and the variogram structure, and thus the formation dip is considered as an unknown model parameter. Estimating the weighting coefficients of the GP method along with two other unknown model parameters, which are the cutoff and the formation dip, through an inversion scheme, addresses these three aforesaid issues of the variogram-based geostatistics.

Several studies have utilized abundant hydrofacies data to reconstruct sedimentary architecture from geophysical logs and lithologic logs. This includes the use of electrical resistivity data [Schulmeister et al., 2003; Tartakovsky et al., 2008], multiple geophysical data types [Linde et al., 2006; Wiederhold et al., 2008], and combined geophysical data and lithologic data [Ezzedine et al., 1999; Chen and Rubin, 2003; Bersezio et al., 2007]. This study uses binary sand and clay hydrofacies data from electric well logs for reconstructing images of the subsurface and lithologic data from drillers' logs as the calibration data.

2.3 Model calibration and uncertainty quantification using CMA-ES

This section is reproduced with modifications from Elshall et al. [submitted].

The use of optimization algorithms for solving the inverse groundwater problem is a common practice. The classes of optimization algorithms include local derivative algorithms, global heuristic algorithms, hybrid global-heuristic local-derivative algorithms, and global-local heuristic algorithms. While the local derivative algorithms are of computational efficiency and have ability to handle larger number of unknown model parameters, yet this can be at the cost of finding local solutions instead of a near global solution. The second class of algorithm is global heuristic algorithms, which are generally implemented when gradient search is not successful. Heuristic algorithms are experience-based techniques that utilize a simple to complex forms of learning to escape local optima and improve the solutions. Few studies use global heuristic algorithms such as genetic algorithm [ElHarrouni, 1996; Karpouzou et al., 2001; Solomatine et al., 1999; Bastani et al., 2010] or particle swarm optimization [Scheerlinck et al., 2009; Jiang et al., 2010] to avoid entrapment at local minima. The third class of algorithms for solving the inverse problem in subsurface modeling is to use a hybrid global-heuristic local-derivative algorithm [Tsai et al., 2003a,b; Blasone et al., 2007; Matott and Rabideau, 2008a,b; Zhang et al.,

2009], which runs a global heuristic algorithm for exploring the search landscape followed by a local derivative algorithm for exploiting favorable search regions. The fourth class of algorithms is the global-local heuristic algorithm, which can perform both global search and local convergence without the need of combining two different algorithms. For solving the inverse groundwater problem and quantifying model parameter uncertainty, this study uses the covariance matrix adaptation evolution strategy (CMA-ES) [Hansen and Ostermeier, 2001; Hansen et al., 2003] as a global-local stochastic derivative free algorithm, which is readily amendable for embarrassingly parallel computation.

The enhanced search properties of CMA-ES stems from its complex learning techniques with high level of abstract description. The CMA-ES adapts a covariance matrix representing the pair-wise dependency between decision variables, which approximates the inverse of the Hessian matrix up to a certain factor. The solution is updated with the covariance matrix and an adaptable step size, which are adapted by two conjugates that implement heuristic control terms. The covariance matrix adaptation uses information from the current population and from the previous search path. Since such an elaborate search mechanism is not common in other heuristic algorithms, the first objective of the study is to evaluate the CMA-ES with respect to other commonly used global heuristic and local derivative algorithms. For the evaluation purpose, four global population-based algorithms are considered, which are ant colony optimization for real domain [Socha and Dorigo, 2008], particle swarm optimization [Iwasaki et al., 2006], modified differential evolution [Babu and Angira, 2006] and genetic algorithm [Haupt and Haupt, 2004]. The ant colony optimization for real domain (ACOR) is selected since it shares the feature of probability distribution estimation with CMA-ES. The particle swarm optimization (PSO) is selected since it is famous for its computational efficiency and it is the second most published

heuristic algorithm after the genetic algorithm (GA). The modified differential evolution (mDE) is selected since it belongs to the same class of evolutionary computation of CMA-ES.

Heuristic algorithms are more commonly used than local derivative algorithms in the subsurface design optimization problem since they generally outperform local derivative algorithms [Aly and Peralta, 1999; Yoon and Shoemaker, 1999; Matott and Rabideau, 2008a] although at a higher computational cost [Yoon and Shoemaker, 1999; Matott and Rabideau, 2008a]. Yet heuristic algorithms are seldom used for solving the inverse groundwater problem of the higher computational cost and the curse of dimensionality. However, algorithms that utilizes multiple solutions in iteration that do not exchange information allows for embarrassingly parallel computation [Vrugt et al., 2006; Tang et al., 2007; Vrugt et al., 2008; Tang et al., 2010]. This is the most efficient parallel technique since the solutions in iteration do not communicate. The second objective of this study is to show that parallel CMA-ES superiorly improves the calibration speed over the sequential CMA-ES. In addition, the speedup of parallel runs scales variably with increasing the number of processors up to a certain limit.

In addition to the global-local search capabilities and parallelization, the third favorable feature of CMA-ES is to quantify model parameter uncertainty due to estimation error. The solution of the CMA-ES, which consist of a maximum likelihood estimate and a full covariance matrix, can be used for Monte Carlo sampling. Several algorithms have utilized the covariance matrix for Monte Carlo sampling [Haario et al., 1999, 2001; Qi and Minka, 2002; Kavetski et al., 2006a, b; Smith and Marshall, 2008; Bardenet and Kégl, 2009; Cui et al., 2011; Zhang and Sutton, 2011]. As pointed out by Müller and Sbalzarini [2010] and Müller [2010], the CMA-ES shares many common concepts and features with the derivative free Markov chain Monte Carlo sampling algorithms [Haario et al., 1999, 2001; Andrieu and Thoms, 2008; Haario et al., 2006;

Müller and Sbalzarini, 2010]. This study shows that the adapted covariance matrix of the maximum likelihood estimation is precise and can be used for Monte Carlo sampling. To the best of my knowledge this is the first study that examines the use of CMA-ES to quantify model parameter uncertainty.

2.4 Constructive epistemic modeling using hierarchical Bayesian model averaging

2.4.1 Hierarchical Bayesian model averaging

This section is reproduced with modifications from Tsai and Elshall [2013]

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; Gaganis 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,b,c; Singh et al., 2010; Trolldborg et al., 2010; Tsai and Li, 2008a,b; Tsai, 2010; Ye et al. 2004, 2005; Wöhling and Vrugt, 2008].

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 non-uniqueness 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 non-uniqueness 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 BMA treatment in this study clearly illustrates this point.

Multimodel approach aims at overcoming the aforementioned shortcomings of the single-model approach by utilizing candidate conceptual models that adequately fit the data. Multimodel methods aim at averaging the considered models through their posterior model probabilities. The most general model averaging 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.

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 BMA Bayesian averaging approach to address this concern by explicitly segregating different sources of uncertainty.

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,b,c] 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,b] 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.

All the previously cited studies are collection multimodel methods, in which all models are at one level. Wagener 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 estimation methods. These were the first two studies to extend the collection BMA formulation of Hoeting et al. [1999] to two levels. Tsai and Elshall [2013] study generalizes the work of Li and Tsai [2009] and Tsai [2010] to a fully hierarchical BMA method. Tsai and Elshall [2013] 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 in relation to model calibration, selection or prediction.

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 candidate discrete model propositions. For example, the variogram model selection can be one source of uncertainty and its candidate propositions could be exponential, Gaussian and pentaspherical variogram models. The proposed hierarchical BMA method serves as a framework for evaluating candidate propositions of each source of uncertainty, to prioritize different sources of uncertainty and to understand the uncertainty propagation through dissecting uncertain model components.

The study uses the hierarchical BMA method for constructive epistemic modeling of the hydrofacies architecture and groundwater flow of the Baton Rouge aquifer-fault system. The concept of constructive epistemic modeling is the subject of the following section.

2.4.2 Constructive epistemic modeling under Bayesian paradigm

This section is reproduced with modifications from Elshall and Tsai [submitted].

A groundwater flow model, for example, could 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 [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. The second inherent challenge of parameter estimation is ill-posedness, which 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 detailed discussion on the uncertainty of groundwater model prediction, the reader is referred to Gupta et al. [2012]. Yet based on this brief account, one can bring the 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,b; Rojas et al., 2008, 2009, 2010a,b,c; Wöhling and Vrugt, 2008; Singh et al., 2010; Troldborg et al., 2010; Seifert et al. 2012] and model discrimination [Usunoff et al., 1992; Li and Tsai, 2009; Tsai, 2010; 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 secondary concern that only few studies acknowledge is epistemic uncertainty [Refsgaard et al., 2006, 2007; Beven, 2006; Clark et al., 2011; Gupta et al. 2012], which is a term that refers to the uncertainty due lack of knowledge. To account for our ignorance, epistemic uncertainty is commonly addressed through possibility theory, imprecise probability and pedigree analysis [Agarwal et al., 2004; Baudrit et al., 2007; He et al., 2008; Refsgaard et al., 2006].

This study presents a complementing prospective on epistemic uncertainty through hierarchical BMA analysis. The basic element of the hierarchical BMA analysis is the base models. 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 is merely the general form of collection BMA in Hoeting [1999]. However, the main aim of the hierarchical BMA is that unlike the collection BMA in which our modeling approach is oriented toward obtaining a BMA solution (i.e. BMA prediction and BMA prediction 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 – or the experienced reality – 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 BMA key assumption is that the supplied set of model 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 that 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.

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.

3 Methods

3.1 Indicator geostatistics

This section is reproduced with modifications from Elshall et al. [2013].

Note that for the following discussion parameterization is conducted in the two-dimensional planar direction along the geological formation dip for every one-foot vertical interval. Three-dimensional aquifer-fault architecture is reconstructed by assembling all two-dimensional slices.

This study utilizes a generalized parameterization (GP) method [Tsai and Yeh, 2004; Tsai, 2006], which combines the indicator kriging (IK) and indicator zonation (IZ) through a set of data weighting coefficients to obtain nonsmooth conditional estimates. The indicator function $\{I(\mathbf{x}, \Delta) : \mathbf{x} \in \text{study area}\}$ is a random function with the indicator random variable Δ describing the spatial extent of sand or clay facies. For a given sand-clay cutoff α , the random function of the indicator random variable Δ for sand facies is defined as

$$I(\mathbf{x}, \Delta) = \begin{cases} 1 & \Delta \in \text{Sand}, \Delta(\mathbf{x}) \geq \alpha \\ 0 & \Delta \notin \text{Sand}, \Delta(\mathbf{x}) < \alpha \end{cases} \quad (1)$$

From equation (1) the indicator outcome (one or zero) indicates the presence of sand facies or clay facies, respectively. The indicator variogram has the same definition as the normal variogram except that the real random function is replaced by the indicator random function $I(\mathbf{x}, \Delta)$. To calculate the expected value $\Delta^*(\mathbf{x}_0)$ at location \mathbf{x}_0 , the GP is

$$\Delta^*(\mathbf{x}_0) = I(\mathbf{x}_k) + \sum_{i=1}^N \lambda_i [I(\mathbf{x}_i) - I(\mathbf{x}_k)] \beta_i \quad (2)$$

where N is the number of electric well logs, $I(\mathbf{x}_i)$ is the indicator data, λ_i is the indicator kriging weight, and β_i is the data weighting coefficient for a data point of a well log at location

\mathbf{x}_i . $I(\mathbf{x}_k)$ is indicator data for a zone defined by well log k . Equation (2) shows that GP estimate at unknown location is similar to IK estimate $\Delta^*(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i I(\mathbf{x}_i)$ or the IZ estimate $\Delta^*(\mathbf{x}_0) = I(\mathbf{x}_k)$ except for the introduction of β_i such that $\forall \beta_i = 1$ gives the IK estimate, $\forall \beta_i = 0$ gives the IZ estimate and $0 < \beta_i < 1$ gives the in-between GP estimate.

The indicator variance using the GP is

$$\sigma_{\Delta}^2(\mathbf{x}_0) = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j R(\mathbf{x}_i, \mathbf{x}_j) \beta_i \beta_j - 2 \sum_{i=1}^N R(\mathbf{x}_i, \mathbf{x}_0) \beta_i + 2\gamma(\mathbf{x}_0, \mathbf{x}_k) \quad (3)$$

where $R(\mathbf{x}_a, \mathbf{x}_b) = \gamma(\mathbf{x}_a, \mathbf{x}_k) + \gamma(\mathbf{x}_b, \mathbf{x}_k) - \gamma(\mathbf{x}_a, \mathbf{x}_b)$ and γ is the variogram.

For zonal delineation, this study uses two-dimensional Voronoi tessellation [Sibson, 1980]. This is a simple mathematical technique for dividing a space into a number of Voronoi zones, given a set of coplanar points, which are electric well logs data. A Voronoi zone, which is drawn based on bi-sectors for each data point, is a boundary enclosing all the intermediate space lying nearest to that data point than to other data points in the plane. The Voronoi tessellation is considered a neutral and unbiased approach to define the neighborhood of a data point [Tsai and Yeh, 2004; Tsai, 2006].

3.2 CMA-ES

3.2.1 CMA-ES algorithm

This section is reproduced with modifications from Elshall et al. [submitted].

The difference between the heuristic and derivative algorithms is that the former algorithms update the search distribution through utilizing stochastic or probabilistic components, while the later are hill descent algorithms that utilizes the gradient or the Hessian matrix through calculating derivatives. Other components of the heuristic algorithm (e.g.

step size, conjugate steps and covariance matrix adaptation) can be similar to derivative algorithms. This is particularly true with the CMA-ES, which empirically calculates the covariance matrix as presented in this section. To avoid terminology confusion, the study uses general terms. For example, generation size in mDE, ant colony size in ACOR and particles swarm size in PSO are referred to as population size. Algorithm specific jargons are presented in parentheses. In addition, the terms solution, search point, unknown model parameters, search space dimensions, and target vector are synonyms.

The study implements the (μ_w, λ) -CMA-ES [Hansen and Ostermeier, 2001; Hansen et al., 2003] in which the weighted recommendations of the best solutions μ_w (parents population size) out of all the solutions λ (offspring population size) in each iteration (generation) are used to update the search distribution parameters. For a problem with search space dimension n , each iteration g with size λ consists of sampling new solutions $\mathbf{v}_i \in \mathbb{R}^n$ for $i = 1 \dots \lambda$, sorting them in an ascending order according to their fitting error and updating the optimization parameters accordingly. The five self-adaptive optimization parameters are the distribution-mean $\mathbf{m}^{(g)} \in \mathbb{R}^n$, the step size $\sigma^{(g)} > 0$, the symmetric and positive definite $n \times n$ covariance matrix \mathbf{C}^g with $\mathbf{C}_0 = \mathbf{I}$ and the two search paths (evolution paths) $\mathbf{p}_\sigma \in \mathbb{R}^n$ and $\mathbf{p}_c \in \mathbb{R}^n$ that are initialized as zero vectors.

The new solutions of iteration $g+1$ are generated by the perturbation (mutation) of the current favorite solution, which is the distribution mean vector $\mathbf{m}^{(g)}$

$$\mathbf{v}_i^{(g+1)} = \mathbf{m}^{(g)} + \sigma^{(g)} \mathbf{y}_i^{(g+1)} \quad i = 1 \dots \lambda \quad (4)$$

in which $\mathbf{y}_i^{(g+1)} \sim \mathbb{N}(\mathbf{0}, \mathbf{C}^{(g)})$ donates a realization of a normally distributed random vector with zero mean and covariance \mathbf{C}^g . The three major components of the CMA-ES are the mean vector

$\mathbf{m}^{(g)}$ representing the favorite solution, the step-size $\sigma^{(g)}$ that controls the step length and the covariance matrix $\mathbf{C}^{(g)}$ that determines the shape of the search distribution ellipsoid.

To update these three optimization state variables, the first step is to evaluate the solutions $\mathbf{x}_i^{(g+1)}$ on the objective function to be minimized, and then rank them in an ascending order by the fitting error such that $\mathbf{v}_{i:\lambda}^{(g+1)}$ becomes the i th solution vector with the corresponding random vector realization $\mathbf{y}_i^{(g+1)}$. The mean of the new distribution (selection and recombination) is carried out by taking a weighted mean of the best solutions of the current iteration (parent population)

$$\mathbf{m}^{(g+1)} = \sum_{i=1}^{\mu} w_i \mathbf{v}_{i:\lambda}^{(g+1)} \quad (5)$$

where the positive (recombination) weight coefficients $w_1 \geq \dots \geq w_{\mu} > 0$ sum to one. The measure termed the variance effective selection mass μ_{eff} is used.

$$\mu_{eff} = \frac{1}{\sum_{i=1}^{\mu} w_i^2} \geq 1 \quad (6)$$

From the definition of w_i the variance effective selection mass is $1 \leq \mu_{eff} \leq \mu$. In case of equal recombination weights, μ_{eff} is equal to μ . Usually $\mu_{eff} \approx \lambda / 4$ indicates a reasonable setting of w_i .

After updating $\mathbf{m}^{(g)}$, the next step is to update the step-size $\sigma^{(g)}$ of the covariance matrix $\mathbf{C}^{(g)}$. To achieve competitive change rates for the overall step size, the search path (evolution path) $\mathbf{p}_{\sigma}^{(g)} \sim \mathbb{N}(\mathbf{0}, \mathbf{I})$ acts as a conjugate to the step size $\sigma^{(g)}$. The length of the search path, which is the cumulation of a fading iteration (population) mean in the iteration sequence, is used to control to control the step size through comparing the search path with its expected length under random selection. If the search path $\mathbf{p}_{\sigma}^{(g)}$ is short indicating that single steps are canceling one

another, then the step size $\sigma^{(g)}$ decreases. On the other hand, if the single steps are pointing to the similar direction and thus the search path $\mathbf{p}_\sigma^{(g)}$ is long, then the step size $\sigma^{(g)}$ increases. Initialized with $\mathbf{p}_\sigma^{(0)} = \mathbf{0}$, the new conjugate search path $\mathbf{p}_\sigma^{(g+1)} \in \mathbb{R}^n$ is

$$\mathbf{p}_\sigma^{(g+1)} = (1 - c_\sigma) \mathbf{p}_\sigma^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)} \mu_{\text{eff}} \frac{1}{\sqrt{\mathbf{C}^{(g)}}} \langle \mathbf{y} \rangle \quad (7)$$

$$\sigma^{(g+1)} = \sigma^{(g)} \cdot \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma^{(g+1)}\|}{E\|\mathbb{N}(\mathbf{0}, \mathbf{I})\|} - 1\right)\right) \quad (8)$$

where $1/c_\sigma \geq 1$ is the backward time horizon of the search path, the term $\sqrt{c_\sigma(2 - c_\sigma)} \mu_{\text{eff}}$ is a normalized constant and $d_\sigma \approx 1$ controls the change magnitude of the step-size.

$E\|\mathbb{N}(\mathbf{0}, \mathbf{I})\| \approx \sqrt{2} \left(1 - \frac{1}{4n} + \frac{1}{21n^2}\right)$ is the expected length of a random variable distribution

according to $\mathbb{N}(\mathbf{0}, \mathbf{I})$. The weighted mean of the best μ ranked vectors \mathbf{y}_i is denoted by

$$\langle \mathbf{y} \rangle \sim \mathbb{N}(\mathbf{0}, \mathbf{C}^{(g)}) = \mathbf{m}^{(g+1)} - \mathbf{m}^{(g)} / \sigma^{(g)} = \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}^{(g+1)} \quad (9)$$

The term $\mathbf{C}^{(g)} = \mathbf{B}^{(g)} (\mathbf{D}^{(g)})^2 \mathbf{B}^{(g)T}$ is an eigendecomposition of $\mathbf{C}^{(g)}$ in which $\mathbf{B}^{(g)}$ is an orthonormal basis of eigenvectors that defines the coordinate system, and the diagonal element of the diagonal matrix $\mathbf{D}^{(g)}$ are the square roots of the corresponding positive eigenvalues that defines the scaling. The transformation $1/\sqrt{\mathbf{C}^{(g)}} = \mathbf{B}^{(g)} (\mathbf{D}^{(g)})^{-1} \mathbf{B}^{(g)T}$ rescales the step $\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}$ to the isotropic coordinate system given by $\mathbf{B}^{(g)}$. In other words, $\mathbf{B}^{(g)T}$ rotates the principal axes of the distribution $\mathbb{N}(\mathbf{0}, \mathbf{C}^{(g)})$ into the coordinate axes; $(\mathbf{D}^{(g)})^{-1}$ apply rescaling such that all axes becomes equally sized; and $\mathbf{B}^{(g)}$ rotates the results back into the original coordinate system to ensure that the principal axes of the distribution are not rotated by the transformation to allow the

directions of the consecutive steps to be comparable. Thus, even if $\mathbf{y}_{i,\lambda}^{(g+1)}$ is distributed according to $\mathbb{N}(\mathbf{0}, \mathbf{C}^{(g)})$, yet the transformation $1/\sqrt{\mathbf{C}^{(g)}}$ makes the conjugate search path flow normal distribution all the time $\mathbf{p}_\sigma^{(g)} \sim \mathbb{N}(\mathbf{0}, \mathbf{I})$, and thus remains independent of its orientation to allow the comparison of $\mathbf{p}_\sigma^{(g+1)}$ with $E\|\mathbb{N}(\mathbf{0}, \mathbf{I})\|$.

The last step is to update the covariance matrix \mathbf{C}^g , which learns all pairwise dependencies between the decision variables. The updating of the \mathbf{C}^g increases the likelihood of the successful steps $\langle \mathbf{y} \rangle$ to reappear. The adaptation follows a natural gradient approximation of the expected fitness since there is a close relation between the covariance matrix and the Hessian matrix. Just as the search path $\mathbf{p}_\sigma^{(g)}$ is calculated to update the step-size $\sigma^{(g)}$, similarly the other search path $\mathbf{p}_c^{(g)}$, which captures the relation between consecutive steps, is calculated to update the covariance matrix \mathbf{C}^g .

$$\mathbf{p}_c^{(g+1)} = (1 - c_c) \mathbf{p}_c^{(g)} + h_\sigma \sqrt{c_\sigma (2 - c_\sigma) \mu_{\text{eff}}} \mathbf{y}_{i,\lambda}^{(g+1)} \quad (10)$$

$$\mathbf{C}^{(g+1)} = (1 - c_1 - c_\mu) \mathbf{C}^{(g)} + c_1 \underbrace{\mathbf{p}_c^{(g+1)} \mathbf{p}_c^{(g+1)T}}_{\text{rank-one update}} + c_\mu \underbrace{\sum_{i=1}^{\mu} w_i \mathbf{y}_{i,\lambda}^{(g+1)} (\mathbf{y}_{i,\lambda}^{(g+1)})^T}_{\text{rank-}\mu \text{ update}} \quad (11)$$

where $c_c \leq 1$, $c_1 \leq 1 - c_\mu$ and $c_\sigma < 1$ are learning rates. The Heaviside h_σ function

$$h_\sigma = \begin{cases} 1 & \text{if } \frac{\|\mathbf{p}_\sigma\|}{\sqrt{1 - (1 - c_\sigma)^{2(g+1)}}} < (1.4 + \frac{2}{n+1}) E\|\mathbb{N}(\mathbf{0}, \mathbf{I})\| \\ 0 & \text{Otherwise} \end{cases} \quad (12)$$

pauses the update of $\mathbf{p}_\sigma^{(g)}$ if $\|\mathbf{p}_\sigma\|$ is large to prevent the fast increase of the axes of \mathbf{C} when the step size is very small. Using the search path $\mathbf{p}_\sigma^{(g)}$ for the rank-one update of the covariance matrix reduces the number of function evaluations to adapt to a straight ridge from $O(n^2)$ to

$O(n)$ [Hansen et al., 2003]. Thus, important parts of the model can be learned in time order of n . The rank- μ update extends the update rule for large population sizes by using μ vectors to update C at each iteration step. The rank- μ update increases the learning rate for large population sizes and thus reduces the number of necessary iterations, which has an important implication for parallel CMA-ES as explained in the results. Rank-one and rank- μ update are used in combination.

3.2.2 Review of CMA-ES with respect to comparison algorithms

This section is produced with modifications from Elshall et al. [submitted].

This section presents a review of five algorithms with respect to CMA-ES. First, the ant colony optimization for real domain (ACOR) [Socha and Dorigo, 2008] is inspired by a foraging ant colony that indirectly communicates through the pheromone trail to identify and minimize the route to the food source. The term ant-colony-size is the number of solutions λ in an iteration. The information of the pheromone trail is stored in archive with size k , which is used by each ant to construct the next solution. The archive size k is greater than the problem dimensions and should be greater than the number of solutions λ (ants). After each function evaluation the archive is updated.

One can notice many similarities between the CMA-ES and ACOR. Both algorithms utilize a Gaussian probability distribution for recombination of new solutions. ACOR uses a Gaussian kernel to depict more than one peck while CMA-ES uses Gaussian probability distribution with maximum entropy. Secondly, the concept of the step size control appears in both algorithms. However, the implementation of step size appears rudimentary in ACOR in comparison to CMA-ES since the ACOR implements only one parameter ξ on the control of step size. The parameter ξ is left for manual tuning along with the parameter q that is related to

the standard deviation of the Gaussian function. Large ξ leads to slower convergence rate. If the parameter ξ is not carefully tuned, then this can lead to premature convergence particularly for large q .

The third similar point is that both algorithms sample a multi-variate normal distribution with a second order approach that changes the neighborhood metric to allow linear transformation of the problem. This is particularly important for tackling ill-conditioned problems (i.e. high conditioning number), in which first order gradient information is insufficient and second order Hessian information is needed. This is clearly implemented in the CMA-ES through adapting the covariance matrix with the eigendecomposition $\mathbf{C}^{(g)} = \mathbf{B}^{(g)}(\mathbf{D}^{(g)})^2\mathbf{B}^{(g)T}$ for the forward and backward transformation of the search space. This is implemented in ACOR through the use of the Gaussian functions associated with a single chosen solution. This property is important for the non-separable problems, which have dependency between the decision variables. It is clear that the covariance matrix $\mathbf{C}^{(g)}$ can further explore the problem structure particularly through its non-diagonal elements.

The second algorithm is particle swarm optimization (PSO), which was first proposed by Kennedy and Elberhart [1995]. This algorithm utilizes a group of interacting particles. Each particle has a position and velocity. In this study a PSO algorithm [Iwasaki et al., 2006] that adapts the particle average velocity through the inertia w , which allows the step size (average velocity) to change to a convergent or a divergent value through comparing it with the ideal step size:

$$v_{avg} = \frac{1}{n \cdot \lambda} \sum_{i=1}^{\lambda} \sum_{j=1}^n |v_{ij}| \quad (13)$$

in which v_{ij} is the step size (velocity) of solution (particle) $i = 1 \dots \lambda$ along dimension $j = 1 \dots n$.

The ideal step size appears as a primitive conjugate in comparison to the step size control $\mathbf{p}_\sigma^{(g+1)}$ of CMA-ES. The neighborhood solution is generated through a stochastic weighted recombination of the particle best solution, the best current solution of all particles and the global best solution. Such weighted recombination updates the current solution by biasing toward the solution and global best solutions, while CMA-ES invokes the pair-wise dependency between decision variables to create a new solution.

While CMA-ES replaces the current solution by a new one, differential evolution (DE) [Storn and Price, 1997] is a constructive algorithm that keeps updating the current solution. Similar to PSO, DE updates the current solution by mixing it with some randomly chosen solutions from the same population. Unlike ACOR that retains the good solutions in the archive, PSO and DE retain the good solution by replacing the old solution with the new one if the new solution is better. The constructive algorithm can hinder parallelism because after each function evaluation the archive in ACOR or the populations in PSO and DE are updated based on the fitting errors. To allow for parallelism, this update after each function evaluation has to be postponed until all the function evaluations of the population are completed. Constructive algorithms need amendment techniques for parallelization. Among these techniques is the utilization of two populations of solutions. This degrades the performance of the optimization since an instant update of solution archive after each function evaluation is more robust than a delayed update after the evaluation of the whole iteration (population). Thus, Babu and Angira [2006] modified the original DE [Storn and Price, 1997] by utilizing only one population allowing for instant update. The modified version reduces the computation effort of the original version, yet scarifies the parallelization capability. Unlike constructive algorithms, CMA-ES

generates new solutions for new iteration based on the learning history. Thus, it can be readily used in parallel without any modification to the algorithm or any change in performance. The mDE is used in this study, since all the algorithm comparisons are done for sequential runs only.

The fourth algorithm is genetic algorithm (GA) [Haupt and Haupt, 2004]. Evolution strategies computation was developed in parallel with the genetic algorithm with the major difference that while genetic algorithm was initially developed with binary encoding, the evolution algorithms were originally developed with real-coding to achieve fast computations. A comparison between CMA-ES and GA on a subsurface design optimization problem [Bayer and Finkel, 2004] shows that CMA-ES outperformed GA, and is a more robust algorithm due to its self-adaptive parameters and since solutions of repeated runs have minor variability. For critical evaluation of GA algorithm with respect to CMA-ES algorithm the reader is referred to Bayer and Finkel [2004].

Finally, the Levenberg–Marquardt algorithm [Marquardt, 1963] is compared with CMA-ES. The Levenberg–Marquardt algorithm is commonly used for solving the inverse groundwater problem [Sun and Yeh, 1990; Weiss and Smith, 1998; Inoue et al., 2000; Li and Tsai, 2009; Zidane et al., 2012]. The Levenberg–Marquardt algorithm stabilizes the Gauss-Newton algorithm by adding a constant to proceed steadily from a poor initial solution, and switches to Gauss-Newton algorithm near the optimum solution as this constant becomes zero for rapid convergence. Nützmann et al. [1998] reported for an inverse problem of a transient flow model that the Levenberg–Marquardt algorithm did not converge in several cases and that divergence occurs with poor initial guesses. To avoid divergence, Nowak and Cirpka [2004] proposed a modified Levenberg–Marquardt algorithm that adapts the step size to reduce the risk of oscillation or overshooting of the solution.

However, the major problem of Levenberg–Marquardt algorithm is the local minimum entrapment. Bledsoe et al. [2011] compared CMA-ES to Levenberg–Marquardt algorithm for solving several inverse transport problems in cylindrical radioactive source/shield systems. They found that for problems with more than three unknown model parameters the CMA-ES performance is superior due to the entrapment of Levenberg–Marquardt algorithm at a local minimum. For Levenberg–Marquardt algorithm evaluation, an initial solution is selected in the vicinity of the global optimum to mitigate the local entrapment problem. In addition, multiple restarts are used.

3.3 Hierarchical Bayesian model averaging

3.3.1 Terminology and notation

This section is reproduced with modifications from Tsai and Elshall [2013].

This section starts by defining some basic terminologies of the BMA tree. Figure 2 shows a BMA tree, which is a hierarchical BMA 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 *hierarch* BMA model with level number 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.

The systematic segregation of different sources of model uncertainties is the central idea of the hierarchical BMA 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 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 candidate 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 candidate propositions for each source of uncertainty.

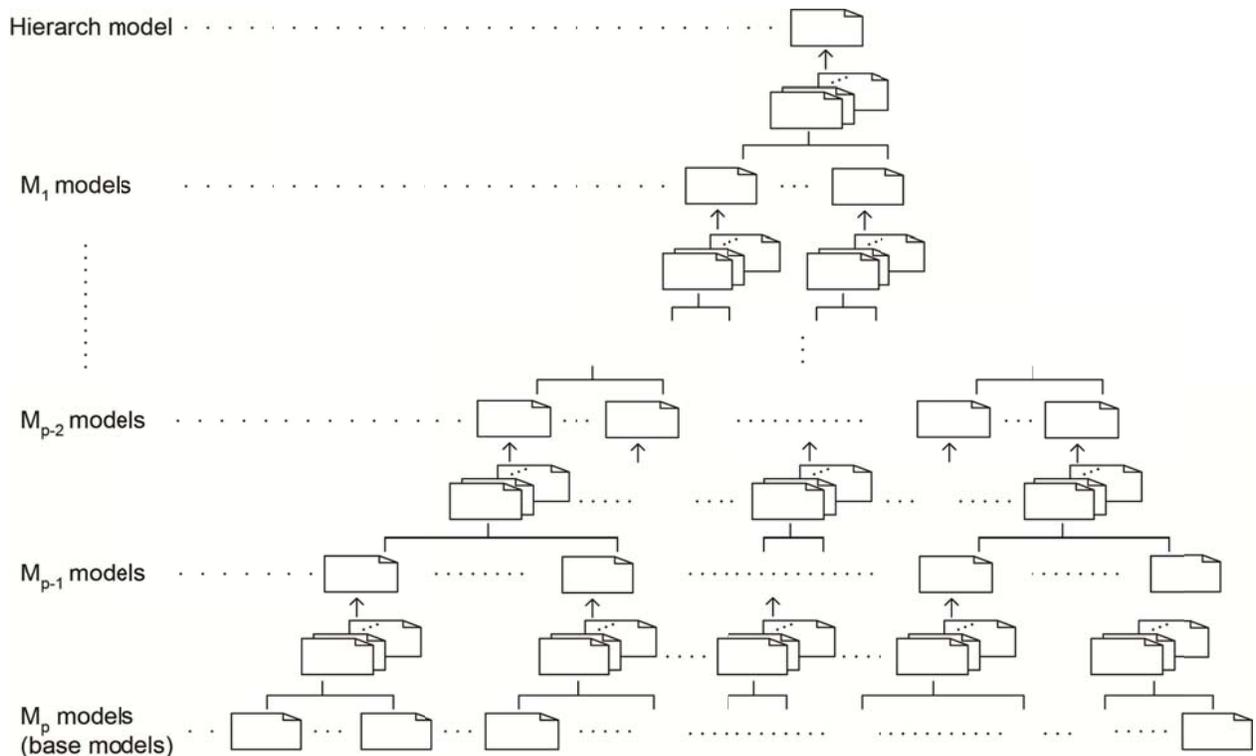


Figure 2 A BMA Tree. This figure is reproduced from Tsai and Elshall [2013].

The subscript is an important index to determine the hierarchy and branch relationship among models. Consider the model $M_{\underbrace{(ij\dots lm)}_p} \in \mathbf{M}_p$ at level p . The subscript $\underbrace{(ij\dots lm)}_p$ 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 the hierarchical BMA posterior model probabilities, hierarchical BMA prediction means, and hierarchical BMA prediction covariances are formulated.

3.3.2 Posterior model probability and conditional posterior model probability

This section is reproduced with modifications from Tsai and Elshall [2013].

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 (MECE) 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 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

non-exhaustive models, but require the base models to be mutually exclusive, which can be achieved in practice by not including nested models.

Consider base models at level p . According to the law of total probability, the posterior probability for predicted quantity Δ (e.g. given data \mathbf{D} is

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

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 (e.g. facies indicator, groundwater head, solute concentration, etc.) given data \mathbf{D} and models \mathbf{M}_p at level p . The expectation

$E_{\mathbf{M}_p} \left[\Pr(\Delta | \mathbf{D}, \mathbf{M}_p) \right]$ is posterior probability averaging at level p . That is

$$E_{\mathbf{M}_p} \left[\Pr(\Delta | \mathbf{D}, \mathbf{M}_p) \right] = \sum_m \Pr \left(\Delta | \mathbf{D}, \underbrace{M_{(ij \cdots lm)}_p}_p \right) \Pr \left(\underbrace{M_{(ij \cdots lm)}_p}_p | \mathbf{D}, \underbrace{M_{(ij \cdots l)}_{p-1}}_{p-1} \right) \quad (15)$$

where $\Pr \left(\Delta | \mathbf{D}, \underbrace{M_{(ij \cdots lm)}_p}_p \right) = \Pr(\Delta | \mathbf{D}, \mathbf{M}_p)$ and m is the number of child models at level p under

the branch of the parent model $(ij \cdots l)$ at level $p-1$.

$\Pr \left(\underbrace{M_{(ij \cdots lm)}_p}_p | \mathbf{D}, \underbrace{M_{(ij \cdots l)}_{p-1}}_{p-1} \right) = \Pr(\mathbf{M}_p | \mathbf{D}, \mathbf{M}_{p-1})$ is the conditional posterior model probability of

model $M_{(ij \cdots lm)}_p$ at level p under model $M_{(ij \cdots l)}_{p-1}$ 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 $M_{(ij \cdots lm)}$ is a child model under the parent model $M_{(ij \cdots l)}$

because both have the same subscript for the first $p-1$ levels. Equation(15) is the Bayesian

model averaging (BMA) at level p , which can be written as

$$\Pr(\Delta | \mathbf{D}, \mathbf{M}_{p-1}) = E_{\mathbf{M}_p} \left[\Pr(\Delta | \mathbf{D}, \mathbf{M}_p) \right] \quad (16)$$

According to equation(16), one can derive the posterior probability of prediction using

BMA over models at any level, say level n :

$$\Pr(\Delta | \mathbf{D}, \mathbf{M}_n) = E_{\mathbf{M}_{n+1}} E_{\mathbf{M}_{n+2}} \cdots E_{\mathbf{M}_p} \left[\Pr(\Delta | \mathbf{D}, \mathbf{M}_p) \right] \quad (17)$$

For collection BMA, only one level of models is considered. Given the law of total probability, equation (14) becomes [Hoeting et al., 1999]

$$\Pr(\Delta | \mathbf{D}) = E_{\mathbf{M}_1} \left[\Pr(\Delta | \mathbf{D}, \mathbf{M}_1) \right] = \sum_i \Pr(\Delta | \mathbf{D}, \mathbf{M}_{(i)}) \Pr(\mathbf{M}_{(i)} | \mathbf{D}) \quad (18)$$

where \mathbf{M}_1 are the base models at level 1 in the second term in equation(18). The third term in equation(18) is the model averaging of all base models $\mathbf{M}_{(i)}$, $\{i = 1, 2, \dots\}$. To develop a multi-

level method that separates different sources of uncertainty, BMA in its general hierarchical BMA form is presented. Then equation(14) for hierarchical BMA becomes

$$\begin{aligned} \Pr(\Delta | \mathbf{D}) &= E_{\mathbf{M}_1} E_{\mathbf{M}_2} \cdots E_{\mathbf{M}_p} \left[\Pr(\Delta | \mathbf{D}, \mathbf{M}_p) \right] \\ &= \sum_i \sum_j \cdots \sum_l \sum_m \Pr(\Delta | \mathbf{D}, \mathbf{M}_{(ij \cdots lm)}) \Pr(\mathbf{M}_{(ij \cdots lm)} | \mathbf{D}, \mathbf{M}_{(ij \cdots l)}) \Pr(\mathbf{M}_{(ij \cdots l)} | \mathbf{D}) \end{aligned} \quad (19)$$

Based on the Bayes rule, the posterior model probability for the base model is

$$\Pr(\mathbf{M}_p | \mathbf{D}) = \frac{\Pr(\mathbf{D} | \mathbf{M}_{(ij \cdots lm)}) \Pr(\mathbf{M}_{(ij \cdots lm)})}{\sum_i \sum_j \cdots \sum_l \sum_m \Pr(\mathbf{D} | \mathbf{M}_{(ij \cdots lm)}) \Pr(\mathbf{M}_{(ij \cdots lm)})} \quad (20)$$

where $\Pr(\mathbf{D} | \mathbf{M}_{(ij \cdots lm)})$ is the likelihood of a base model. $\Pr(\mathbf{M}_p | \mathbf{D})$ often refers to the model

weight in BMA. $\Pr(\mathbf{M}_{(ij \cdots lm)})$ is the prior model probability of the base models. The conditional

posterior model probability of a base model under their parent models is

$$\Pr(\mathbf{M}_p | \mathbf{D}, \mathbf{M}_{p-1}) = \frac{\Pr(\mathbf{D} | \mathbf{M}_{(ij \cdots lm)}) \Pr(\mathbf{M}_{(ij \cdots lm)} | \mathbf{M}_{(ij \cdots l)})}{\sum_m \Pr(\mathbf{D} | \mathbf{M}_{(ij \cdots lm)}) \Pr(\mathbf{M}_{(ij \cdots lm)} | \mathbf{M}_{(ij \cdots l)})} \quad (21)$$

where $\Pr(M_{(ij\cdots lm)} | M_{(ij\cdots l)})$ is the conditional prior model probability of a base model

$M_{\underbrace{(ij\cdots lm)}_p} \in \mathbf{M}_p$ under its parent model $M_{\underbrace{(ij\cdots lm)}_{p-1}} \in \mathbf{M}_{p-1}$. Equation(21) is also referred to as the

conditional model weight. The likelihood for parametric base models is

$$\Pr(\mathbf{D} | \mathbf{M}_p) = \int_{\beta_p} \Pr(\mathbf{D} | \mathbf{M}_p, \beta_p) \Pr(\beta_p | \mathbf{M}_p) d\beta_p \quad (22)$$

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

$$\Pr(\mathbf{D} | M_{(ij\cdots l)}) = \sum_m \Pr(\mathbf{D} | M_{(ij\cdots lm)}) \Pr(M_{(ij\cdots lm)} | M_{(ij\cdots l)}) \quad (23)$$

By considering equal conditional prior model probabilities for $\Pr(M_{(ij\cdots lm)} | M_{(ij\cdots l)})$, the conditional posterior model probabilities are calculated for the base models under their parent models as follows

$$\Pr(\mathbf{M}_p | \mathbf{D}, \mathbf{M}_{p-1}) = \frac{\Pr(\mathbf{D} | M_{(ij\cdots lm)}) \Pr(M_{(ij\cdots lm)} | M_{(ij\cdots l)})}{\sum_m \Pr(\mathbf{D} | M_{(ij\cdots lm)}) \Pr(M_{(ij\cdots lm)} | M_{(ij\cdots l)})} \quad (24)$$

Using equations(23) and (24) the conditional posterior model probabilities for models at level n under their parent models are

$$\Pr(\mathbf{M}_n | \mathbf{D}, \mathbf{M}_{n-1}) = \frac{\sum_{n+1} \cdots \sum_l \sum_m \Pr(\mathbf{M}_{(ij\cdots lm)} | \mathbf{D}, \mathbf{M}_{(ij\cdots l)}) \Pr(\mathbf{M}_{(ij\cdots l)})}{\sum_n \sum_{n+1} \cdots \sum_l \sum_m \Pr(\mathbf{M}_{(ij\cdots lm)} | \mathbf{D}, \mathbf{M}_{(ij\cdots l)}) \Pr(\mathbf{M}_{(ij\cdots l)})} \quad (25)$$

And the posterior model probabilities at level n is

$$\Pr(\mathbf{M}_n | \mathbf{D}) = \frac{\sum_{n+1} \cdots \sum_l \sum_m \Pr(\mathbf{D} | M_{(ij\cdots lm)}) \Pr(M_{(ij\cdots l)})}{\sum_i \sum_j \cdots \sum_l \sum_m \Pr(\mathbf{D} | M_{(ij\cdots lm)}) \Pr(M_{(ij\cdots l)})} \quad (26)$$

Therefore, each model at any level in Figure 2 has its own posterior model probabilities as in equation(26) and conditional posterior model probabilities as in equation(25). As a result, a BMA tree of posterior model probabilities can be obtained.

3.3.3 Prediction means and prediction covariances

This section is reproduced with modifications from Tsai and Elshall [2013].

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

$$E(\Delta | \mathbf{D}) = E_{M_1} E_{M_2} \cdots E_{M_p} [E(\Delta | \mathbf{D}, \mathbf{M}_p)] \quad (27)$$

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(17), the expectation of prediction using models at level n is

$$E(\Delta | \mathbf{D}, \mathbf{M}_n) = E_{M_{n+1}} E_{M_{n+2}} \cdots E_{M_p} [E(\Delta | \mathbf{D}, \mathbf{M}_p)] \quad (28)$$

where $n < p$. Equation(28) provides thorough information for decision makers 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(28) at any level in Figure 2, a BMA tree of prediction means can be obtained.

The law of total covariance for hierarchical BMA is

$$E_{M_1} E_{M_2} \cdots E_{M_p} [Cov(\Delta | \mathbf{D}, \mathbf{M}_p)] + \sum_{n=p}^{n=1} E_{M_1} E_{M_2} \cdots E_{M_{n-1}} Cov_{M_n} [E(\Delta | \mathbf{D}, \mathbf{M}_n)] \quad (29)$$

where $Cov(\Delta | \mathbf{D}, \mathbf{M}_p)$ is the covariance of prediction for given data \mathbf{D} and base models at level

p and Cov_{M_n} is the covariance operator with respect to model \mathbf{M}_n at level n . Cov_{M_n} is the

between-model covariance:

$$\begin{aligned} & Cov_{M_n} [E(\Delta | \mathbf{D}, \mathbf{M}_n)] \\ &= E_{M_n} \left[[E(\Delta | \mathbf{D}, \mathbf{M}_n) - E(\Delta | \mathbf{D}, \mathbf{M}_{n-1})][E(\Delta | \mathbf{D}, \mathbf{M}_n) - E(\Delta | \mathbf{D}, \mathbf{M}_{n-1})]^T \right] \end{aligned} \quad (30)$$

where T is the transpose operator.

When considering the collection BMA that has only one level ($p-1$), the covariance in equation(30) returns to the usual form [Hoeting et al., 1999]

$$\text{Cov}(\Delta | \mathbf{D}) = E_{\mathbf{M}_1} [\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_1)] + \text{Cov}_{\mathbf{M}_1} [E(\Delta | \mathbf{D}, \mathbf{M}_1)] \quad (31)$$

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

When considering two levels ($p-2$) as in Li and Tsai [2009] and Tsai [2010], the covariance in equation(29) is

$$\begin{aligned} \text{Cov}(\Delta | \mathbf{D}) &= E_{\mathbf{M}_1} E_{\mathbf{M}_2} [\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_2)] + E_{\mathbf{M}_1} \text{Cov}_{\mathbf{M}_2} [E(\Delta | \mathbf{D}, \mathbf{M}_2)] \\ &+ \text{Cov}_{\mathbf{M}_1} [E(\Delta | \mathbf{D}, \mathbf{M}_1)] \end{aligned} \quad (32)$$

This study considers four sources of uncertainty, i.e., four levels ($p = 4$). The hierarchical BMA formulation following equation(28) and (29) is

$$E(\Delta | \mathbf{D}) = E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} E_{\mathbf{M}_4} [E(\Delta | \mathbf{D}, \mathbf{M}_4)] \quad (33)$$

$$\begin{aligned} \text{Cov}(\Delta | \mathbf{D}) &= E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} E_{\mathbf{M}_4} [\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_4)] + E_{\mathbf{M}_1} E_{\mathbf{M}_2} E_{\mathbf{M}_3} \text{Cov}_{\mathbf{M}_4} [E(\Delta | \mathbf{D}, \mathbf{M}_4)] \\ &+ E_{\mathbf{M}_1} E_{\mathbf{M}_2} \text{Cov}_{\mathbf{M}_3} [E(\Delta | \mathbf{D}, \mathbf{M}_3)] + E_{\mathbf{M}_1} \text{Cov}_{\mathbf{M}_2} [E(\Delta | \mathbf{D}, \mathbf{M}_2)] \\ &+ \text{Cov}_{\mathbf{M}_1} [E(\Delta | \mathbf{D}, \mathbf{M}_1)] \end{aligned} \quad (34)$$

Similarly, the hierarchical BMA permits the evaluation of the prediction covariance when different BMA models at different levels are proposed. The basic information 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

$$\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_n) = E_{\mathbf{M}_{n+1}} [\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_{n+1})] + \text{Cov}_{\mathbf{M}_{n+1}} [E(\Delta | \mathbf{D}, \mathbf{M}_{n+1})] \quad (35)$$

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

$$E_{\mathbf{M}_n} [\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_n)] = E_{\mathbf{M}_n} E_{\mathbf{M}_{n+1}} [\text{Cov}(\Delta | \mathbf{D}, \mathbf{M}_{n+1})] + E_{\mathbf{M}_n} \text{Cov}_{\mathbf{M}_{n+1}} [E(\Delta | \mathbf{D}, \mathbf{M}_{n+1})] \quad (36)$$

The between-model covariance of prediction using models at level n is equation(30). The within-model covariance in equation(36) 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(36), 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(35) for each model at any level in Figure 2, a BMA tree of prediction covariances can be developed.

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 usual BMA or the hierarchical BMA. Then, the expectation of prediction using models at level n in equation(28) 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(36) and the between-model covariance in equation(30) 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 .

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 candidate 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.

3.3.4 Computation of posterior model probability with variance window

This section is reproduced with modifications from Tsai and Elshall [2013].

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. 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; 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 hierarchical BMA. Following skips BIC derivation and readers are referred to Draper [1995] and Raftery [1995].

The likelihood for a base model $\Pr(\mathbf{D}|\mathbf{M}_p)$ in equation(22) is

$$\Pr(\mathbf{D} | M_{(ij..lm)}) \approx \exp\left[-\frac{1}{2}\text{BIC}_{(ij..lm)}\right] \quad (37)$$

The Bayesian information criterion (BIC) is

$$\text{BIC}_{(ij..lm)} = -2\ln \Pr(\mathbf{D} | M_{(ij..lm)}, \hat{\boldsymbol{\beta}}_{(ij..lm)}) + m_{(ij..lm)} \ln n \quad (38)$$

where $\hat{\boldsymbol{\beta}}_{(ij\dots lm)}$ are the maximum-likelihood estimated model parameters in model $M_{(ij\dots lm)}$, $m_{(ij\dots lm)}$ is the dimension of the model parameters $\hat{\boldsymbol{\beta}}_{(ij\dots lm)}$, and n is the size of data set \mathbf{D} . $\Pr(\mathbf{D} | M_{(ij\dots lm)}, \hat{\boldsymbol{\beta}}_{(ij\dots lm)})$ is the maximum likelihood value. By considering equal prior parameter probabilities for $\hat{\boldsymbol{\beta}}_{(ij\dots lm)}$ and a multi-Gaussian distribution for fitting errors to observation data

Δ^{obs} , the BIC in equation(38) is simplified to [Tsai and Li, 2008a; Li and Tsai, 2009]

$$\text{BIC}_{(ij\dots lm)} = Q_{(ij\dots lm)} + n \ln 2\pi + m_{(ij\dots lm)} \ln n \quad (39)$$

where

$$Q_{(ij\dots lm)} = \left(\Delta^{\text{cal}}(\hat{\boldsymbol{\beta}}_{(ij\dots lm)}) - \Delta^{\text{obs}} \right)^T C_{\Delta}^{-1} \left(\Delta^{\text{cal}}(\hat{\boldsymbol{\beta}}_{(ij\dots lm)}) - \Delta^{\text{obs}} \right) \quad (40)$$

is the sum of the weighted squared fitting errors between calculated Δ^{cal} and observed Δ^{obs} . C_{Δ} is the error covariance matrix. The variance of the errors can be estimated by running sufficient number of realizations

$$\sigma_k^2 = \text{var} \left[\Delta_{r,q,k}^{\text{cal}}(\hat{\boldsymbol{\beta}}_{(ij\dots lm)}) - \Delta_k^{\text{obs}} \right] \quad (41)$$

where $k = 1, \dots, n$ is a data point; r is the number of realizations that could vary for each model; and q is the number of models.

Substituting BIC in equation(37) into equation(20) the posterior model probability for the base model is

$$\Pr(M_{(ij\dots lm)} | \mathbf{D}) = \frac{\exp \left[-\frac{1}{2} \Delta \text{BIC}_{(ij\dots lm)} \right] \Pr(M_{(ij\dots lm)})}{\sum_i \sum_j \dots \sum_l \sum_m \exp \left[-\frac{1}{2} \Delta \text{BIC}_{(ij\dots lm)} \right] \Pr(M_{(ij\dots lm)})} \quad (42)$$

where $\Delta \text{BIC}_{(ij\dots lm)} = \text{BIC}_{(ij\dots lm)} - \text{BIC}_{\min}$. BIC_{\min} is the minimum BIC value among all the base models. Using ΔBIC in equation(42) is a common practice [Neuman, 2003; Li and Tsai, 2009]

to avoid numerical difficulty when $BIC_{(ij\dots lm)}$ are large numbers. BIC The conditional posterior model probabilities of the base models under their parent models are

$$\Pr(\mathbf{M}_p | \mathbf{D}, \mathbf{M}_{p-1}) = \frac{\Pr(\mathbf{D} | M_{(ij\dots lm)}) \Pr(M_{(ij\dots lm)} | M_{(ij\dots l)})}{\sum_m \Pr(\mathbf{D} | M_{(ij\dots lm)}) \Pr(M_{(ij\dots lm)} | M_{(ij\dots l)})} \quad (43)$$

Once the posterior model probabilities and conditional posterior model probabilities for base models are obtained, the posterior model probabilities and conditional posterior model probabilities at any level can be obtained via equations(26) and (25) respectively.

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

$$\alpha = \frac{s_1}{s_2 \sigma_D} \quad (44)$$

into equation(42) as follows

$$\Pr(M_{(ij\dots lm)} | \mathbf{D}) = \frac{\exp\left[-\frac{1}{2} \alpha \Delta BIC_{(ij\dots lm)}\right] \Pr(M_{(ij\dots lm)})}{\sum_i \sum_j \dots \sum_l \sum_m \exp\left[-\frac{1}{2} \alpha \Delta BIC_{(ij\dots lm)}\right] \Pr(M_{(ij\dots lm)})} \quad (45)$$

The parameter s_1 is the ΔBIC 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 Occam's window is enlarged 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 the readers is referred to groundwater studies that

compare the use of the variance window and Occam's window [Tsai and Li, 2008a, b; Li and Tsai, 2009; Singh et al., 2010].

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 are subject to revision shall 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.

This provides a complete formulation for hierarchical BMA with variance window, from which the following basic concepts can be concluded. 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 that is obtained through the hierarchical BMA analysis is identical to the BMA model that is obtained through the collection BMA analysis.

3.3.5 Similarities and differences between collection BMA and hierarchical BMA

This section is reproduced with modifications from Tsai and Elshall [2013].

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 candidate 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 candidate 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.

First, through model dissection following a combinatorial design, the hierarchical BMA provides a systematic representation of the candidate 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 candidate 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; Tsai and Elshall, 2013], 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 candidate 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. The study illustrates these four features in the following case study. Constructive epistemic modeling of hydrofacies architecture under Bayesian paradigm.

4 Constructive epistemic modeling of hydrofacies architecture under Bayesian paradigm

4.1 Case Study: Hydrofacies architecture model of the Baton Rouge aquifer-fault system

4.1.1 Model data

This section is reproduced with modifications from Elshall et al. [2013] and Tsai and Elshall [2013].

Electric and drillers' well log data that are used in this case study cover elevations between -1000 ft (-304.8 m) to -3000 ft (-914.4m) NGVD29. The vertical discretization is at one-foot (0.304 m) intervals. For geostatistical reconstruction of subsurface formation, the study uses 288 electric well logs as shown in Figure 3 with 237 water well logs and 51 oil/gas well logs. The case study uses indicator kriging as explained in Section 3.1.

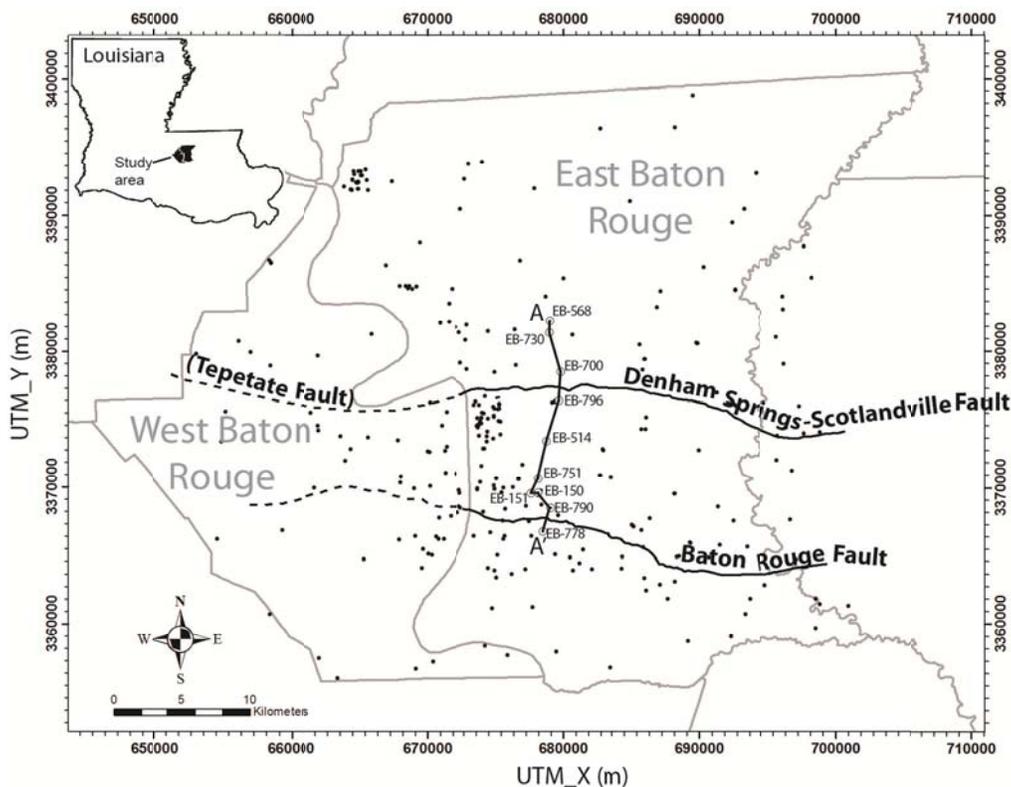


Figure 3 The study area. Black dots represent the location of electrical well logs. The dashed lines in West Baton Rouge Parish represent approximated location of the faults. AA' is a cross section for Figure 7 [Tsai and Elshall, 2013].

The sand-clay sequence interpretation for each log is based on electrical resistivity, spontaneous potential, and gamma. In general, shallow electrical resistivity (e.g., short-normal resistivity, medium induction resistivity, etc.) of 20 ohm-m is a good threshold for water-well logs to identify sand units for the freshwater formations in southeastern Louisiana. When salty water is present instead, the spontaneous-potential response helps to identify sand units. When available, the gamma ray response is used as a guide along with resistivity and spontaneous potential to identify sand units. For example, in Figure 4 the saline sands are identified in well log EB-783 located at the south of the Baton Rouge fault using SP and resistivity. For a saline sand, the SP response is pronounced and the long normal resistivity is less than the short normal resistivity. Also, the presence of salt water can be seen at the bottom of the sand in the depth of 2200 feet. Freshwater sands are identified in well log EB-1317 (south of the Baton Rouge fault) and WBR-128 (south of the Baton Rouge fault) based on resistivity. SP response is not pronounced in these two well logs. Low gamma ray in EB-1317 correlates sand units.

While inverse modeling is generally utilized for hydrofacies reconstruction by conditioning on piezometric head or on concentration data [e.g. Zhou et al., 2010, Li et al., 2012b], this study uses lithologic data from drillers' logs as the calibration data. For model calibration lithologic data from 93 drillers' logs are used. For each well log, lithologic data represents the actual lithology with depth. Interpreting drillers' logs can be subjective and thus assigning binary indicator values to the drillers' logs is uncertain. To achieve a consistent interpretation of drillers' logs, lithologic descriptions are categorized into three categories: sand facies, clay facies and undetermined facies as listed in Table 1. The assignment of the drillers' logs facies depends on the used data sets as explained in the next section.

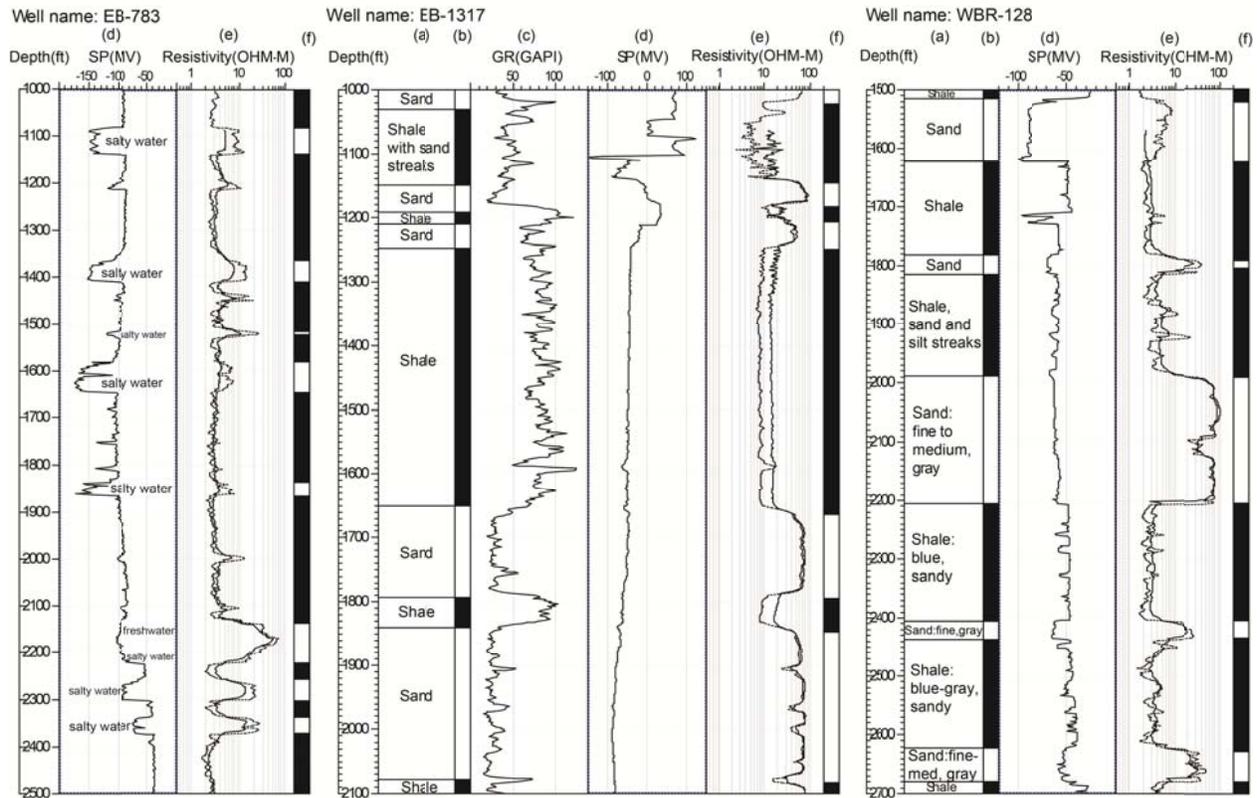


Figure 4 Wells logs for water well EB-1317, north of the Baton Rouge fault, and water wells EB-783 and WRB-128, south of the Baton Rouge fault. Column index is as follows: (a) drillers' log, (b) binary interpretation of drillers' log (white for sand and black for clay), (c) gamma ray (GR), (d) spontaneous potential (SP), (e) short normal resistivity (dotted line) and long normal resistivity (solid line), and (f) binary interpretation of electrical log (white for sand and black for clay) [Elshall et al., 2013].

Table 1 Interpretation of drillers' logs into three lithological facies and indicator assignment [Elshall et al., 2013].

Facies	Clay	Undetermined	Sand
Lithologic description	Clay: blue, hard, soft, gray green, brown, dark brown highly organic, tan, red-brown, green, with sand strings Shale: heavy, sandy, hard, red, brown, sticky, yellow, with mixed gravel, with streaks of sand, with some sand breaks	Clay and sand, shale and sand, streaks of sand and shale, shaly sand, poor sand and streaks of shale, sand and hard sandy shale	Sand: fine, packed, very fine, good, medium, coarse, loose, yellow, hard packed, packed, pay, sandstone, gray, lightly gray, tight, with shell fragments, with wood, gray-white, blue-gray, with gravel
Indicator	0	0/1	1

To achieve consistency with the electric well log interpretation, sand and gravel are considered to belong to the sand facies indicator 1 and other materials belong to the clay facies indicator 0. This point is illustrated in Figure 4, which shows lithology columns where both the drillers' logs and electric logs are available. For observation well WBR-128, drillers' terms such as "sand", "sand: fine, medium, gray" and "sand: fine, gray" are easily interpreted as sand facies indicator 1. Similarly, terms such as "shale", "shale: blue, gray, sandy" are easily interpreted as clay facies indicator 0. Indistinct terms such as "shale, sand, and silt streaks" are interpreted as clay facies indicator 0. Similarly, for observation well EB-1317 the indistinct term "shale with sand streaks" is interpreted as clay facies indicator 0. This is to maintain consistency with the electric logs interpretation in which distinct sand only is assigned sand facies indicator 1. For the well logs EB-1317 and WBR-128 in Figure 4, the interpretation of the drillers' log shows very good match with the interpretation of the electric logs. The mismatch of the interpreted indicators from the drillers' log and the electric logs is 3.0 % for WBR-128 and 4.6 % for EB-1317. The average mismatch for the 19 well logs in the used data set where both drillers' logs and electric logs are available is $7.12 \pm 2.44\%$. This indicates that the selected 93 drillers logs tend to have adequate quality and that the interpretation and the indicator assignment for the drillers' logs and electric logs are consistent. However, this is the first calibration data set. The second calibration data set is explained in the following section.

4.1.2 Model data and model structure uncertainty

This section is reproduced with modifications from Tsai and Elshall [2013] and Elshall et al. [2013].

Due to uncertainty of the model data, structure and parameters, multiple potential hydrofacies models are resulted and calibrated. The central idea of the hierarchical BMA

method is to segregate different uncertain model components with their corresponding candidate model propositions. These concepts of uncertain model components and candidate model propositions are illustrated in Figure 5. This case study as shown in Figure 5 considers four uncertain model components in the hydrofacies 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 5 shows that the hydrofacies data interpretation and parameter estimation technique are not considered as uncertain model components since only one proposition is considered for each of these model components.

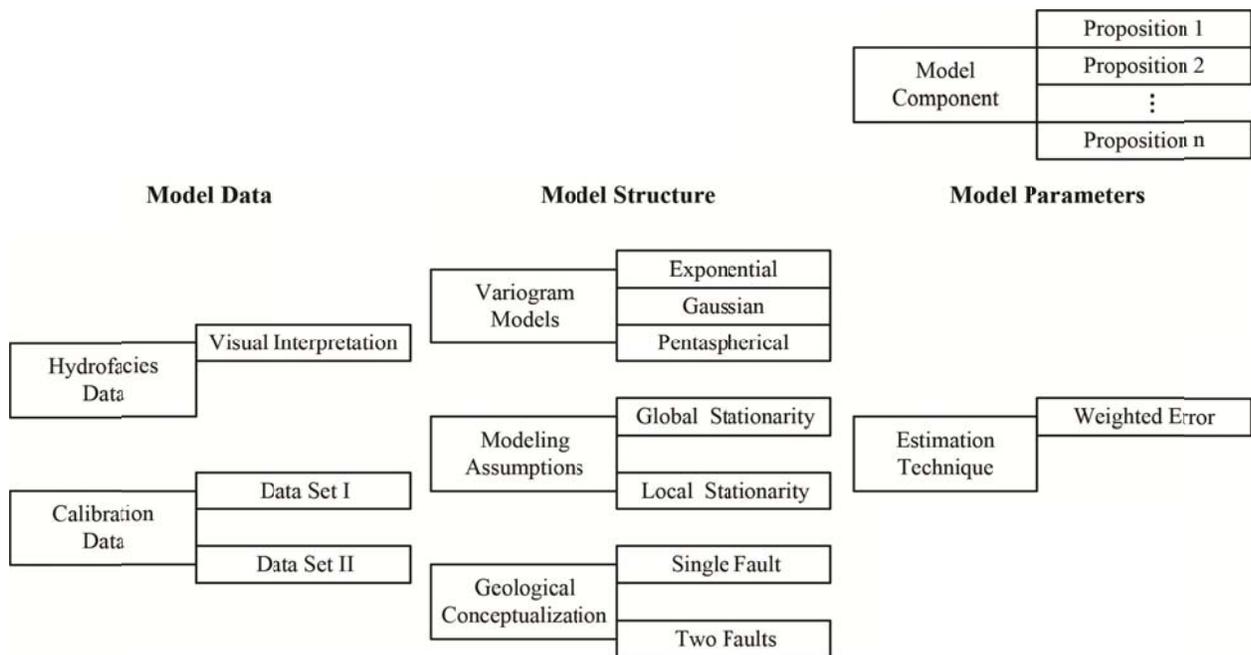


Figure 5 Uncertainty segregation through dissection of model components with their candidate modeling proposition [Tsai and Elshall, 2013].

The four uncertain model components with their corresponding candidate propositions result in 24 calibrated models. These models are used to perform hierarchical BMA multimodel characterization of the hydrofacies architecture of the Baton Rouge aquifer-fault system to

present the main features of the hierarchical BMA method. This section presents detailed description of the hydrofacies architecture model with its uncertain model components.

For model calibration lithologic data from 33 driller's logs are used. However, different interpretations of drillers' logs lead to multiple calibration data sets (see Table 1). 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.

With respect to the hydrofacies model structure, the first uncertain model component is the choice of the spatial correlation function of the hydrofacies units. This study uses three candidate 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, two geological stationarity propositions are adopted. 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 domain as separated by the fault system resulting in local variogram model for each model domain. 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.

Two geological conceptualization propositions, which are the two-domain proposition and the three-domain proposition, are tested. Similar to Rollo [1969] the two-domain proposition does not consider the Denham Springs-Scotlandville fault, 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-domain 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.

4.1.3 Model parameters and calibration

This section is reproduced with modifications from Tsai and Elshall [2013] and Elshall et al., 2013].

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, the formation dip $\phi = 0.30^\circ \pm 0.06^\circ$ is calculated from the USGS cross-sectional map in the area [Griffith, 2003]. A range of

$0.06^\circ \leq \phi \leq 0.57^\circ$ is assigned for the formation dip. The second model parameter is the sand-clay cutoff θ , which rounds the indicator estimate Δ to a binary value. The range of the cutoff is set to $0.3 \leq \theta \leq 0.7$.

To estimate the unknown model parameters, the inverse problem is formulated by minimizing the fitting errors between the estimated and observed facies as follows

$$\min_{\phi, \theta} \frac{1}{2} \left\{ \frac{1}{M_{sand}} \frac{\sum_{i=1}^{M_{sand}} [\Delta^{i,est}(\mathbf{x}) - \Delta_{sand}^{i,obs}(\mathbf{x})]^2}{\sigma^2(\mathbf{x})} + \frac{1}{M_{clay}} \frac{\sum_{i=1}^{M_{clay}} [\Delta^{i,est}(\mathbf{x}) - \Delta_{clay}^{i,obs}(\mathbf{x})]^2}{\sigma^2(\mathbf{x})} \right\} \quad (46)$$

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(40), equation(46) 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.

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 (ϕ, θ) are independently estimated for each of the 24 models. The CMA-ES [Hansen et al., 2003] is used to solve the inverse problem in equation(46) according to the following procedure. First, the CMA-ES generates candidate solutions (ϕ, θ) . For each candidate solution, the experimental variograms for each domain are calculated given the formation dip ϕ . 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 θ . Forth, 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.

4.2 Results and Discussion

4.2.1 Calibration and BIC

This section is reproduced with modifications from Tsai and Elshall [2013] and Elshall et al. [2013].

For results and discussion, the following short forms are used. The first level of uncertainty is about the conceptualization of the Denham Springs-Scotlandville fault resulting into two-domain (Z2) and three-domain (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-domain (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.

Table 2 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, 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 [Thomaszewski, 1996] and with the estimated dip $0.30^\circ \pm 0.06^\circ$ from Griffith [2003].

Given two unknown model parameters and the fitting residual Q , I use equation(39) to calculate $BIC_{(ij..lm)}$. To obtain the BMA tree, the posterior model probabilities are calculated 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 31500. Table 2 shows ΔBIC 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 σ_D of the fitting residual Q [Tsai and Li, 2008a,b].

Table 2 Calibrated model parameters, fitting errors (equation(46)), Q, Δ BIC and posterior model probabilities for base models. Z3D1LExp is the best model [Tsai and Elshall, 2013].

Base Model	Parameters			Q	Δ BIC	Occam's Window	Posterior Model Probabilities					
	Dip [deg.]	Cutoff	Fitting Error				Variance Window 1%			Variance Window 5%		
							$1\sigma_D$	$2\sigma_D$	$3\sigma_D$	$1\sigma_D$	$2\sigma_D$	$3\sigma_D$
Z2D1GExp	0.23	0.39	2.010	10010	854	0.00	0.00	0.04	1.42	0.00	0.52	3.54
Z2D1GGus	0.44	0.44	2.181	10620	1464	0.00	0.00	0.00	0.09	0.00	0.01	0.57
Z2D1GPen	0.30	0.41	2.208	10927	1771	0.00	0.00	0.00	0.02	0.00	0.00	0.23
Z2D1LExp	0.44	0.42	2.080	10062	906	0.00	0.01	0.79	6.50	0.19	3.74	9.51
Z2D1LGus	0.32	0.41	2.111	10243	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	10190	1034	0.00	0.00	0.01	0.62	0.00	0.18	2.07
Z2D2GGus	0.44	0.42	2.160	10362	1206	0.00	0.00	0.00	0.28	0.00	0.06	1.24
Z2D2GPen	0.18	0.41	2.621	11243	2087	0.00	0.00	0.00	0.00	0.00	0.00	0.09
Z2D2LExp	0.44	0.42	2.087	10132	976	0.00	0.00	0.01	0.81	0.00	0.25	2.46
Z2D2LGus	0.28	0.40	2.151	10349	1193	0.00	0.00	0.00	0.30	0.00	0.07	1.29
Z2D2LPen	0.17	0.42	2.372	11226	2070	0.00	0.00	0.00	0.01	0.00	0.00	0.09
Z3D1GExp	0.29	0.40	2.087	10087	931	0.00	0.00	0.02	1.00	0.00	0.33	2.81
Z3D1GGus	0.18	0.41	2.219	10928	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	10194	1038	0.00	0.00	0.01	0.61	0.00	0.17	2.04
Z3D1LPen	0.28	0.41	2.150	10253	1097	0.00	0.00	0.00	0.47	0.00	0.12	1.71
Z3D2GExp	0.29	0.40	2.092	10159	1003	0.00	0.00	0.01	0.72	0.00	0.21	2.27
Z3D2GGus	0.18	0.42	2.188	10845	1689	0.00	0.00	0.00	0.03	0.00	0.00	0.29
Z3D2GPen	0.30	0.41	2.249	11118	1962	0.00	0.00	0.00	0.01	0.00	0.00	0.13
Z3D2LExp	0.45	0.42	2.067	10042	886	0.00	0.00	0.03	1.23	0.00	0.43	3.22
Z3D2LGus	0.30	0.40	2.164	10394	1238	0.00	0.00	0.00	0.24	0.00	0.05	1.12
Z3D2LPen	0.44	0.42	2.174	10548	1392	0.00	0.00	0.00	0.12	0.00	0.02	0.71

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 σ_D increase, which consecutively decrease the weights of the best models. Adjusting the scaling factor of the variance windows is subject to the analyst decision; and model weights are changed as shown in Table 2. However, adjusting the scaling factor does not change the model ranking, but just increases the inclusion of base models [Tsai and Li, 2008a, b; 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, a large variance window of 5% and $3\sigma_D$ is used for the successive analysis

4.2.2 Model propositions evaluation using the BMA tree

This section is reproduced with modifications from Tsai and Elshall [2013].

Figure 6 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-domain (Z3), data set I (D1), local stationarity (L) and exponential variogram (Exp) generally show higher weights than other candidate 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.

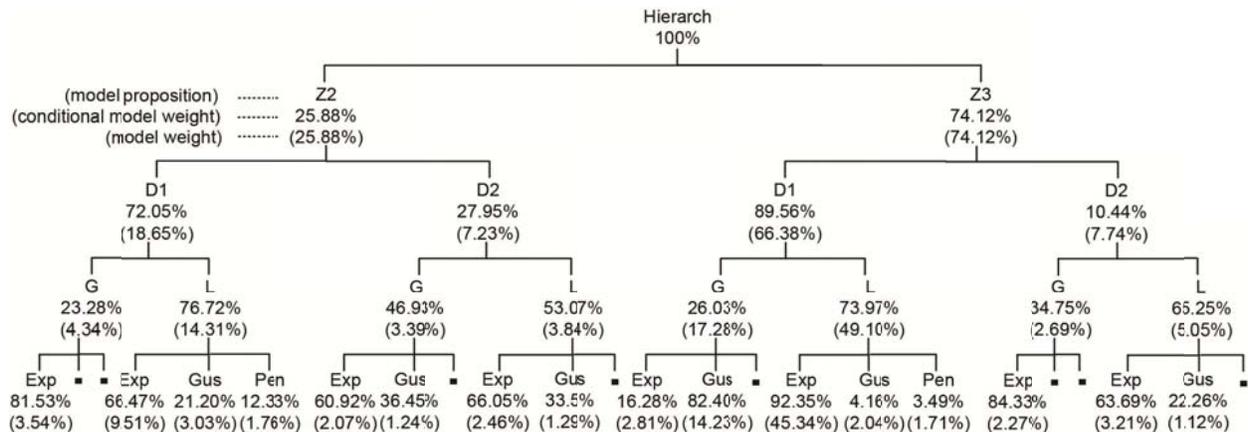


Figure 6 The BMA tree of the posterior model probabilities (model weights) and the conditional posterior model probabilities (conditional model weights) for the four uncertain model components. Models with posterior model probabilities less than 1% are not shown in the figure [Tsai and Elshall, 2013].

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, the study examines if the models weights can relate to our understanding of the model under study. Starting with the base level of the BMA tree as shown in Figure 6, models with exponential variogram propositions (Exp) have higher weights in most branches, followed by the Gaussian variogram proposition (Gus) and finally the Pentaspherical 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.

The third level of the BMA tree in Figure 6 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.

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.

The first level of the BMA tree compares the two-domain proposition (Z2) and the three-domain proposition (Z3). The posterior model probability of the Z3 proposition that explicitly

accounts for the Denham Springs-Scotlandville fault is relatively higher than the Z2 proposition. Figure 7 permits the visually evaluation of whether the Denham Springs-Scotlandville fault causes sand units displacement along the fault plane. The Z2 model in Figure 7(a) implies a fault throw in the “2,000-foot” sand, but shows no obvious displacement in the “1,500-foot” sand. The Z3 model in Figure 7(b) defines the displacement in the “1,500-foot” sand and “2,000-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 7(c) is very similar to the Z3 model, yet showing high total model variance around the Denham Springs-Scotlandville fault in Figure 7(d) due to the Z2 proposition.

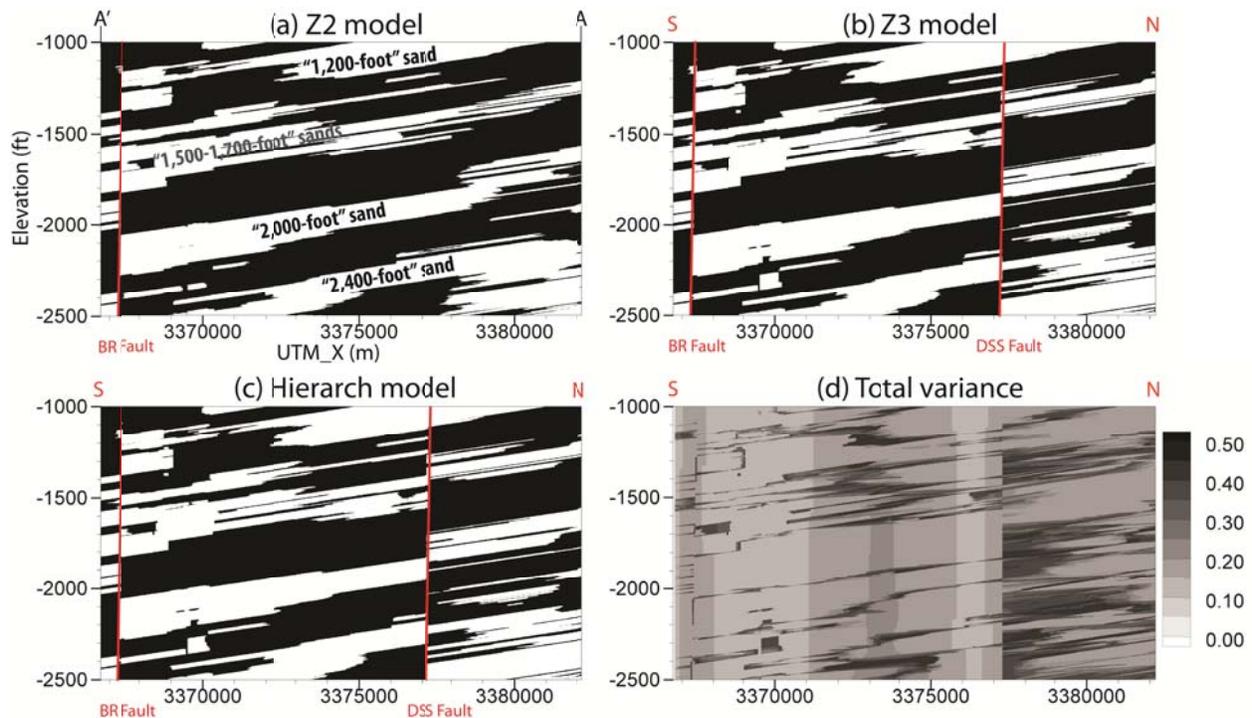


Figure 7 The BMA model estimates for the cross section $\overline{AA'}$ (see Figure 3): (a) Z2 model, (b) Z3 model and (c) Hierarch model. White areas are sand and black areas are clay. (d) is the total model variance for the hierarch model. The locations of the Baton Rouge (BR) fault and the Denham Springs-Scotlandville (DSS) fault are marked [modified from Tsai and Elshall, 2013].

4.2.3 Uncertainty propagation and prioritization

This section is reproduced with permission from Tsai and Elshall [2013].

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 candidate 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, the study uses the south cross section of the Denham Springs-Scotlandville fault as shown in Figure 8 that follows the fault line shown in Figure 3 but rendered in two dimensions for clarity. The grid spacing is 50 m along the fault line and 1 foot (0.304 m) in the vertical direction.

Before discussing Figure 8, tracing and understanding the patterns of uncertainty propagation is first discussed. Table 3 shows the mean values of the variances for all BMA models in the BMA tree. Table 3 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 weight 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 3 illustrates this observation in which the within-model variance of Z3 model is less than the hierarch model.

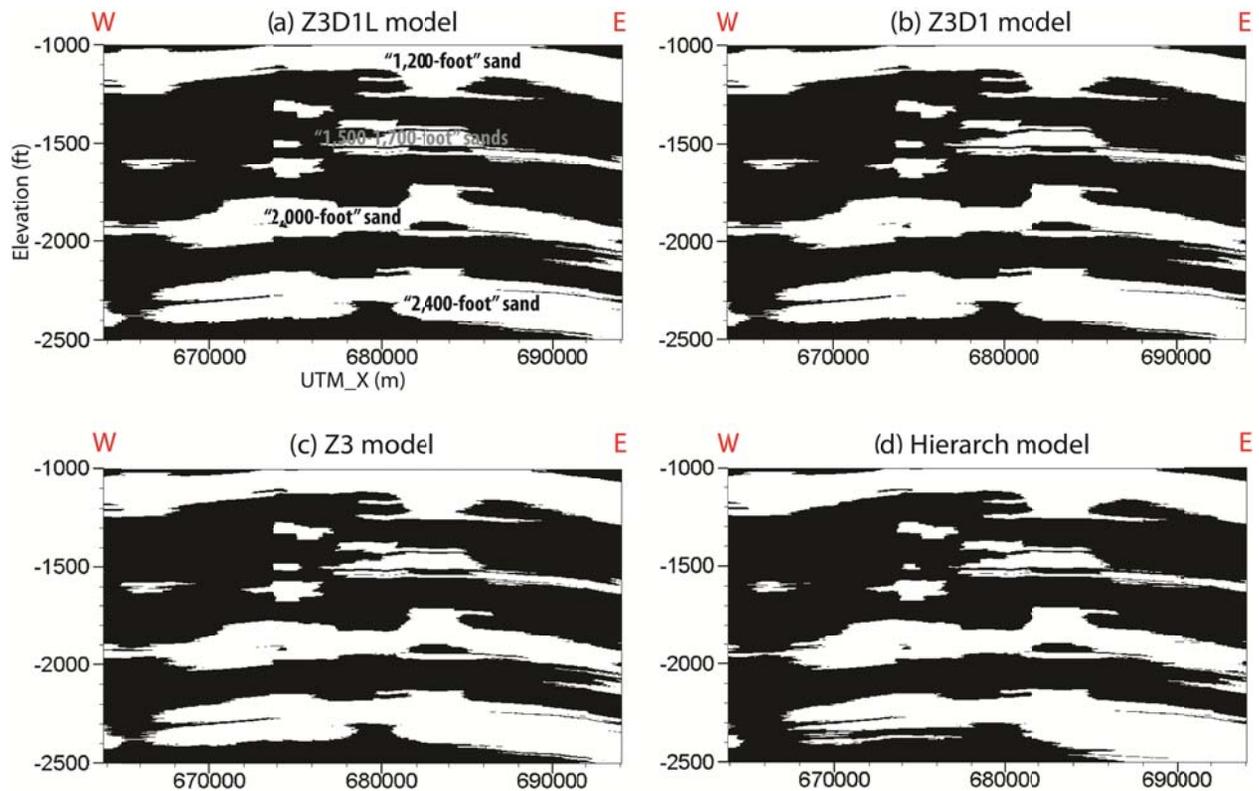


Figure 8 The BMA model estimates for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model and (d) Hierarch model. White areas are sand and black areas are clay [modified from Tsai and Elshall, 2013].

Table 3 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 [Tsai and Elshall, 2013].

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	0.204	0.015	0.218	0.72	0.222	0.004	0.226	0.26	0.244	0.028	0.271
Z2D1L	0.184	0.020	0.204	0.77											
Z2D2G	0.187	0.039	0.226	0.47	0.222	0.010	0.232	0.28	0.248	0.001	0.250	0.74			
Z2D2L	0.187	0.031	0.218	0.53											
Z3D1G	0.185	0.017	0.202	0.26	0.215	0.033	0.248	0.90	0.248	0.001	0.250	0.74			
Z3D1L	0.208	0.011	0.220	0.74											
Z3D2G	0.180	0.013	0.193	0.35	0.226	0.024	0.250	0.10							
Z3D2L	0.209	0.034	0.243	0.65											

Unlike the within-model variance and the total model variance, the between-model variance is independent of subordinate levels as indicated by equation (30) and as illustrated in Table 3. This feature allows for prioritizing the relative impact of each uncertain model component on the overall model uncertainty. For example, the small between-model variance at level one as shown in Table 3 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 3 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 uncertainty propagation, the model estimation, within-model variance, between-model variance and total model variance for the best branch are plotted.

The BMA models in Figure 8 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(d) is marginally different particularly for the “2,400-foot” sand because the Z2 and Z3 propositions produce different estimations as previously illustrated in Figure 7(a) and Figure 7(b).

Figure 9 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 candidate propositions are important and the candidate 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 Figure 9(b). Also, the averaging of the Z2 model and Z3 model as shown in Figure 9(d) resulted in high between-model variance, since their estimations are noticeably different. Similarly, small between-model variance is due to the similarity of the candidate models or the presence of a dominant candidate proposition. For example, using different calibration data sets resulted in similar models with the same dip and cutoff, yet with different (see Table 2). Also, D1 proposition outperforms D2 proposition as shown by their posterior model probabilities in Figure 7. Thus, the insignificant impact of the calibration data set propositions on overall model uncertainty is due to the combined effect of the two factors.

As shown in Figure10(a), 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. Figure10 (b) 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, the within-model variance for the Z3 model

and the hierarch model can be obtained as shown in Figure10 (c) and Figure10 (d), respectively. Comparing Figure10 (c) with Figure10 (d) 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.

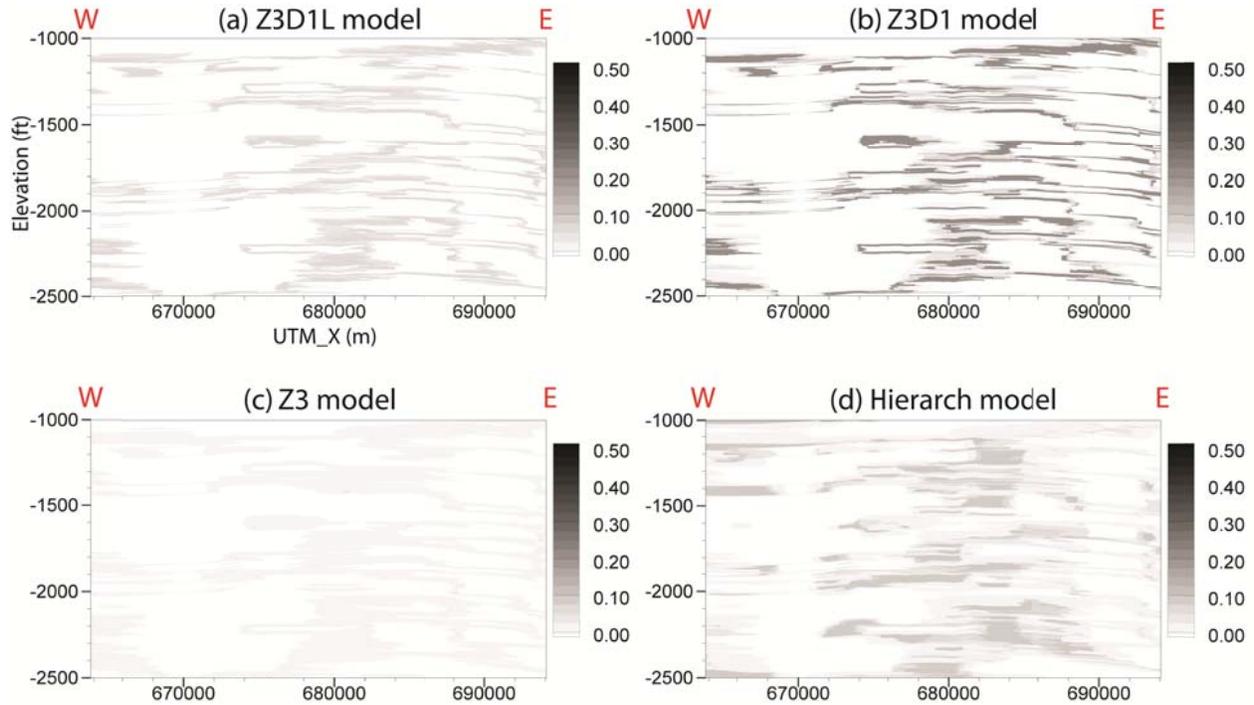


Figure 9 The between-model variance for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model and (d) Hierarch model [modified from Tsai and Elshall, 2013].

Figure 11 shows the total model variance that is the sum of the between-model variance in Figure 9 and the within-model variance in Figure 10. Figure 11 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.

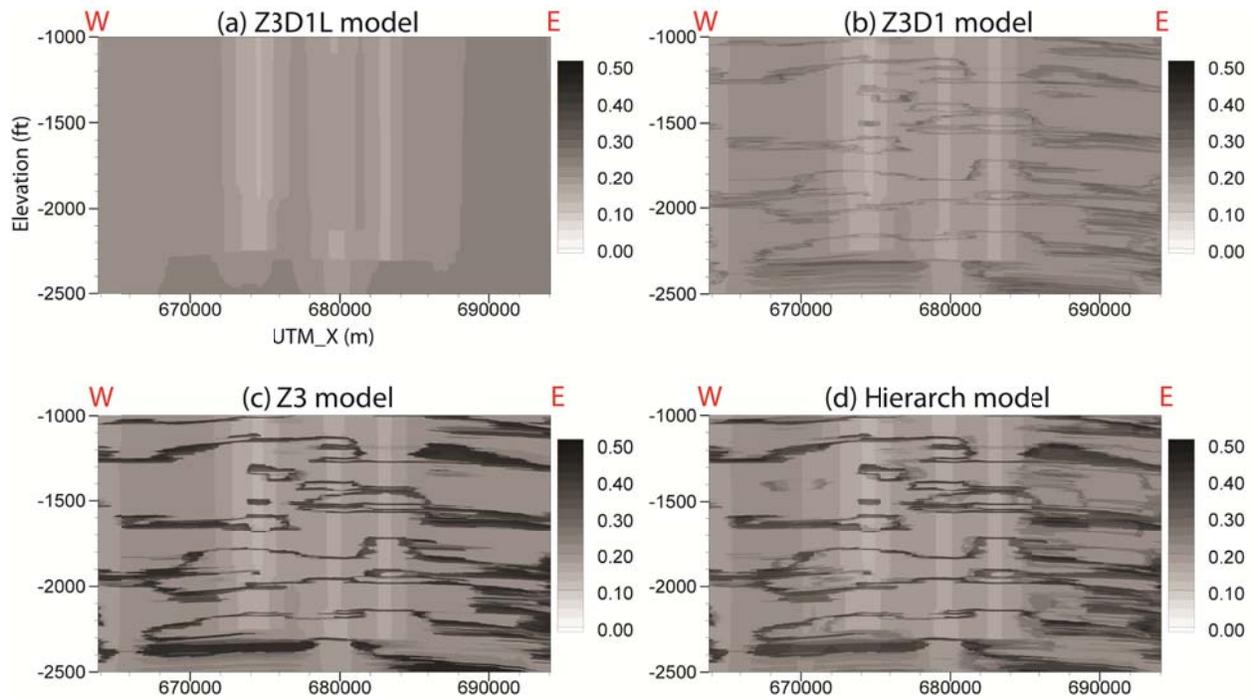


Figure 10 The within-model variance for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model and (d) Hierarch model [modified from Tsai and Elshall, 2013].

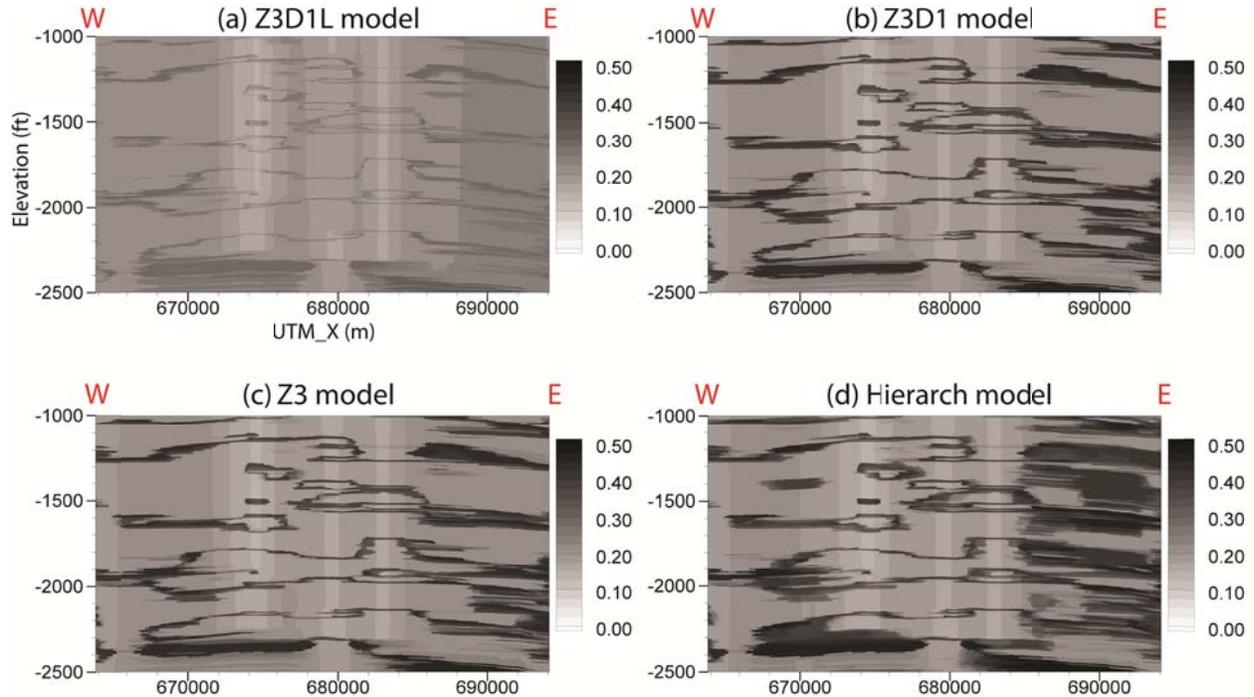


Figure 11 The total model variance for the cross section south of the Denham Springs-Scotlandville fault for the best branch: (a) Z3D1L model, (b) Z3D1 model, (c) Z3 model and (d) Hierarch model [modified from Tsai and Elshall, 2013].

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 8d) and total model variance (Figure 11d) from the hierarch BMA model are identical to those from the collection BMA.

4.3 Conclusions

The hierarchical BMA Bayesian model averaging (hierarchical BMA) provides a framework for incorporating candidate 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 candidate model propositions, and prioritization of uncertain model components.

The hierarchical BMA supports the rejection of a single representation of the system in favor of many system representations. The hierarchical BMA method illustrates the fact that model uncertainty is likely to be underestimated if only the best model is used. hierarchical BMA 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 candidate propositions, the total uncertainty is increased through the between-model variance. The between-model variance presents an important uncertainty origin that cannot be discarded.

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 candidate 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 for the indicators interpreted from the electrical logs. 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 as it has small between-model variance. The hierarchical BMA 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.

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.

5 Hydrogeological characterization of the Baton Rouge aquifer-fault system

5.1 Case Study: Hydrofacies architecture model of the Baton Rouge aquifer-fault system

5.1.1 Hydrofacies architecture model

This entire chapter is reproduced with modifications from Elshall et al. [2013].

To better understand salinization underneath Baton Rouge and the role of the geological faults as conduits and barriers [Bense and Person, 2006; Hanor et al., 2011] to fluid flow, it is imperative to study the detailed hydrofacies architecture of the Baton Rouge aquifer-fault system. The proposed hydrofacies architecture model is similar to the hydrofacies architecture model in Section 4.1. This includes the model data, the indicator geostatistics method and the inversion scheme to estimate the unknown model parameters. One difference is that this model uses 89 drillers' logs for model calibration. A second difference is that in the previous case study the weighting coefficients β were set to one (i.e. indicator kriging) to reduce the calibration run time. In this case study fixed weighting coefficients β of zero (i.e. indicator zonation) and one (i.e. indicator kriging) are used. In addition, variable weighting coefficients β (i.e. generalized parameterization) are used.

This case study uses the robust model propositions as identified by the first reach step in Section 4. With respect to the fault conceptualization proposition, by recognizing the presence of the Denham Springs-Scotlandville fault as a robust proposition, the modeling area in Figure 1 results in three geographic modeling domains: a domain south of the Baton Rouge fault, a middle domain between the Baton Rouge fault and the Denham Springs-Scotlandville fault, and a domain north of the Denham Springs-Scotlandville fault. Aquifer architectures are independently reconstructed for the three individual domains. The fault architecture is produced by juxtaposing the reconstructed aquifer architectures immediately north and south of the fault plane. For both

the Baton Rouge fault and Denham Springs-Scotlandville fault, it is assumed that the fault planes are vertical, although deeper in the sections the faults are known to be dipping to the south [McCulloh and Heinrich, 2012]. With respect to the other three positions, this case study considers the calibration data set I proposition, the exponential variogram proposition and the global stationarity proposition. The details of these propositions are provided in Section 4.1.2.

5.1.2 Model parameters and calibration

The first unknown model parameter is the formation dip, which establishes data correlation. The Baton Rouge aquifer system gently dips south. Prior geological studies did not quantify the formation dip. The dip was calculated to be $0.30^\circ \pm 0.06^\circ$ from the USGS cross-sectional maps in the area [Griffith, 2003] as prior information. To constrain the search space, the dip is set within the range $0.06^\circ \leq \phi \leq 0.60^\circ$. The vertical tolerance of the dip is not reported in any study, and no vertical tolerance is considered.

The second unknown model parameter is the sand-clay cutoff θ . The estimated Δ values, which cutoff value rounds to produce an indicator, could be viewed as the conditional probability with respect to the binary variables [Chilès and Delfiner, 1999]. The limits of the cutoff θ are constrained to a realistic but flexible range of $0.3 \leq \theta \leq 0.7$.

Other unknown model parameters are the data weighting coefficients β of the well logs. The model used 288 geophysical well logs in which the south, middle and north domains have 61, 129 and 98 well logs, respectively, to reconstruct the hydrofacies architecture. When the spacing of well logs is dense, the estimates by the GP method in these areas become insensitive to the data weighting coefficients β , since the indicator kriging estimates and indicator zonation estimates are similar. To reduce the computational cost of the inverse problem, insensitive values of β are identified through sensitivity analysis. Performing the sensitivity analysis starts by

calibrating the hydrofacies architecture model only with respect to the dip and cutoff for given $\forall \beta = 1$ to all well logs, and then using the calibrated model as a reference for the fitting error. Then taking one well log at a time, its data weighting coefficient is evaluated from 0 to 1 incrementally by 0.1 to calculate new fitting errors. Any well log that results in an error difference less than $\pm 0.05\%$ from the calibrated model is considered to have a fixed data weighting coefficient $\beta = 1$. The sensitivity analysis shows that 48 well logs have sensitive β coefficients with their number in the south, middle and north domains being 6, 34 and 8 well logs, respectively.

The data weighting coefficients β along with the dip ϕ and sand-clay cutoff θ are the unknown model parameters to be estimated using an inversion scheme. The inversion scheme for the IZ, IK and GP is the same except for the size of the unknown parameters. The IZ inversion has only one unknown parameter that is the dip. The unknown parameters of the IK inversion are the dip and the cutoff. The unknown parameters of the GP inversion are the dip, the cutoff and the data weighting coefficients. The inverse problem is formulated by minimizing the mean squared error between the estimated and observed facies as follows:

$$\frac{1}{2} \left\{ \frac{1}{M_{sand}} \sum_{i=1}^{M_{sand}} [\Delta^{i,est}(\mathbf{x}) - \Delta_{sand}^{i,obs}(\mathbf{x})]^2 + \frac{1}{M_{clay}} \sum_{i=1}^{M_{clay}} [\Delta^{i,est}(\mathbf{x}) - \Delta_{clay}^{i,obs}(\mathbf{x})]^2 \right\} \quad (47)$$

where M_{sand} and M_{clay} are the number of data points of the sand facies and clay facies, respectively. The $\Delta^{i,est}(\mathbf{x})$, $\Delta_{sand}^{i,obs}(\mathbf{x})$ and $\Delta_{clay}^{i,obs}(\mathbf{x})$ are the indicator estimate, the observed sand facies indicator and the observed clay facies indicator at a location \mathbf{x} , respectively. The mean squared error is separated into two error terms with one for each facies to avoid calibration bias toward favoring the fitting of clay over sand since the well logs indicate a clay proportion of

about two-third by volume within the study area. The proportion of sand calculated from the electric logs is 0.338 and is 0.339 from the drillers' logs. This separation underlines that reducing the sand error is equally important as reducing the clay error.

CMA-ES [Hansen et al., 2003] is used to solve the inverse problem. The inversion scheme steps are as follows. First, the CMA-ES generates candidate solutions, which are sets of unknown model parameters. Second, for each proposed solution the experimental variograms and a theoretical variogram are calculated based on the proposed dip. With respect to experimental variograms it is important to clarify one precaution with respect to location dependence of data correlation is accounted for. The correlation between the data across the faults is prevented, but all the experimental variograms of each domain are grouped together to calculate one theoretical variogram. The theoretical variograms is fitted to the experimental variograms automatically through using the pattern search method of Hooke and Jeeves [1961]. It performs direct directional search for the correlation parameters, which are the nugget, sill and effective range, to minimize the squared root error between the experimental and the theoretical variograms. Third, interpolation function in equation(2) is used to estimate facies distribution at unknown locations. For the inversion purpose the unknown locations are the drillers' logs locations. For the IZ inversion all the β values are set to zero, and thus the cutoff is not needed. Contrariwise, for the IK inversion the β values are set to one, thus the estimated facies is rounded to the indicator value by the cutoff. For the GP inversion β values are used by the interpolation function in equation(2) to estimate facies distribution at the unknown locations and the cutoff is used to round the indicator. Fourth, the estimated facies are compared to the observed facies to calculate the mean squared error for individual solutions. Then, step 1 is repeated until the mean squared error is minimized.

The outcome of the inversion is the best unknown model parameters set that fits the observed facies. This parameters set (dip, cutoff and β values) can be used to plot any 2-dimensional or 3-dimensional diagrams according to the desired grid size. For example, in this study all the cross sections of the faults have a grid size of 50 m along the fault lines. The 3-dimensional diagrams of the aquifer system have a grid size of 200 m in the X and Y directions. The discretization in Z direction is 1-foot (0.34-m) interval.

5.2 Results and discussion

5.2.1 Calibration results

This inversion scheme is used to obtain the optimal data weighting coefficients, dip, and cutoff for the hydrofacies architecture model. The calibration results are shown in Table 4. The variogram structure and cutoff are similar for the indicator zonation (IK), generalized parameterization (GP), and indicator kriging (IZ) methods. The three methods also show the same dip around 0.29° and the same sand proportion around 0.35. The GP shows less fitting error than the IK and IZ methods due to the flexibility of the method.

Table 4 Estimated Variogram Structural Parameters and Model Parameters for the Three Methods [Elshall et al., 2013]

Method	Nugget	Sill	Range [m]	Dip [Deg.]	Cutoff	Sand Proportion	Sand Error[%]	Clay Error[%]	Total Error[%]
IZ	0.062	0.16	8400	0.276	-	0.34	13.02	12.79	12.91
GP	0.083	0.14	8400	0.289	0.404	0.347	11.96	12.9	12.43
IK	0.084	0.14	8600	0.286	0.404	0.347	12.04	12.96	12.5

To show the differences between the three methods, the architecture of the Denham Springs-Scotlandville fault and the Baton Rouge fault are used as examples. Figure 12 and Figure 13 show the juxtaposition at the fault cross sections. Black areas are clay units north of the fault. Gray areas are clay units south of the fault. White areas show potential hydraulic connections formed by sand units from both sides of the fault. It is noted that the faults are three-

dimensional zones of deformation, not two-dimensional planes. Determination of permeability of the fault zone is suggested by Bense and Person [2006] and Hanor et al. [2011]. Nevertheless, the detailed architecture of the fault zone is not the scope of this study. The three methods succeeded in capturing the main flow pathways through the faults. The facies shape using the GP estimate is neither as sharp-edged as the IZ estimate nor as smooth as the IK estimate, which is the main advantage of the GP method. The following results and discussion are based on the GP estimate.

Besides using three methods to understand the uncertainty of the estimates, the study further assesses the uncertainty of the unknown model parameters, which are the dip, cutoff and β values. The CMA-ES provides the maximum likelihood estimate with a full covariance for the unknown model parameters. Hundred samples are obtained, which are random vectors each with the size of the unknown model parameters set chosen from the multivariate normal distribution with mean, and covariance. Figure 14 shows the results of the uncertainty analysis. Figure 14 (a) and Figure 14 (b) show the realizations that have the minimum generated dip 0.23° and maximum generated dip 0.41° in all the realizations. It is important to note that for the steeper dip the flow pathway at the east in the “2,000-foot” sand disappears. Figure 14 (c) and Figure 14 (d) show the realizations that have the minimum generated cutoff 0.39 and maximum generated cutoff 0.42 in all the realizations. These two realizations have dips of 0.29° and 0.30° , respectively. Since these two figures are relatively similar particularly with respect to main flow pathways across the fault, it can be concluded that the dip is the most sensitive parameter with respect to the impact of the results. Figure 14 (e)- Figure 14 (h) show a randomly selected realization, ensemble average of the realizations, ensemble variance of the realizations at south of the fault and ensemble variance of the realizations at north of the fault. Again the emendable

average Figure 14 (f) is comparable with the maximum likelihood estimate Figure 13(b) with respect to the main flow pathways across the fault.

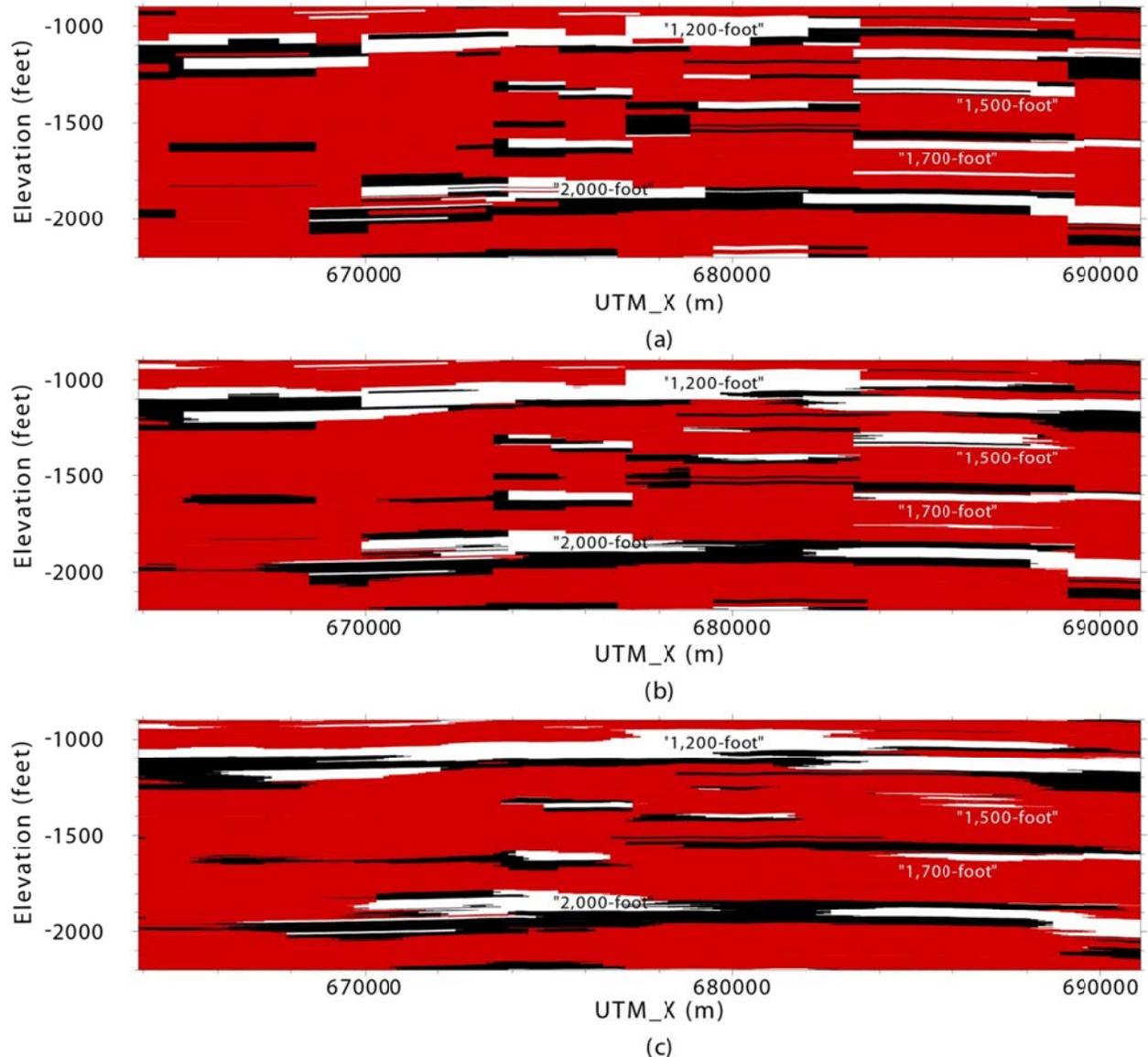


Figure 12 The architecture of the Denham Springs-Scotlandville fault in the modeling area using (a) indicator zonation, (b) generalized parameterization, and (c) indicator kriging. Black areas are clay units north of the fault. Gray areas are clay units south of the fault. White areas show potential hydraulic connections formed by sand units from both sides of the fault. The fault cross sections are based on 3D estimates that follow the UTM_X and UTM_Y coordinates of the fault line in Figure 3. Elevation is feet above NGVD29 [modified from El shall et al., 2013]

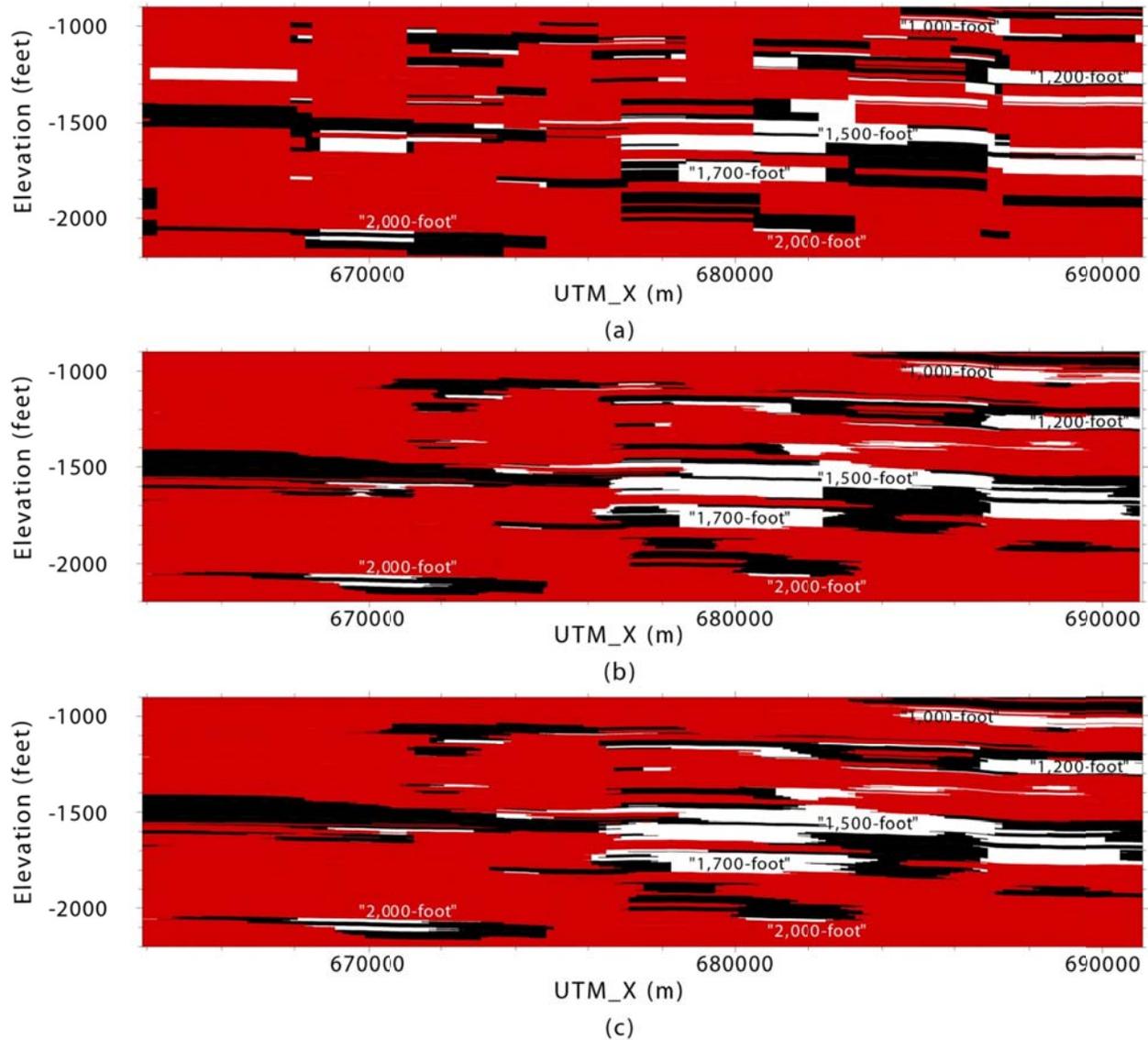


Figure 13 The architecture of the Baton Rouge fault in the modeling area using (a) indicator zonation, (b) generalized parameterization, and (c) indicator kriging. Black areas are clay units north of the fault. Gray areas are clay units south of the fault. White areas show potential hydraulic connections formed by sand units from both sides of the fault. The fault cross sections are based on 3D estimates that follow the UTM_X and UTM_Y coordinates of the fault line in Figure 3 [modified from Elshall et al., 2013]

The ratio of the sensitive β coefficients with respect to the number geophysical well logs for south of the fault is $61/6 \approx 0.10$ and for north of the fault is $34/129 \approx 0.26$. Thus, it is expected that the variance south of the fault as shown in Figure 14(g) is lower than the variance north of the fault as shown in Figure 14 (h).

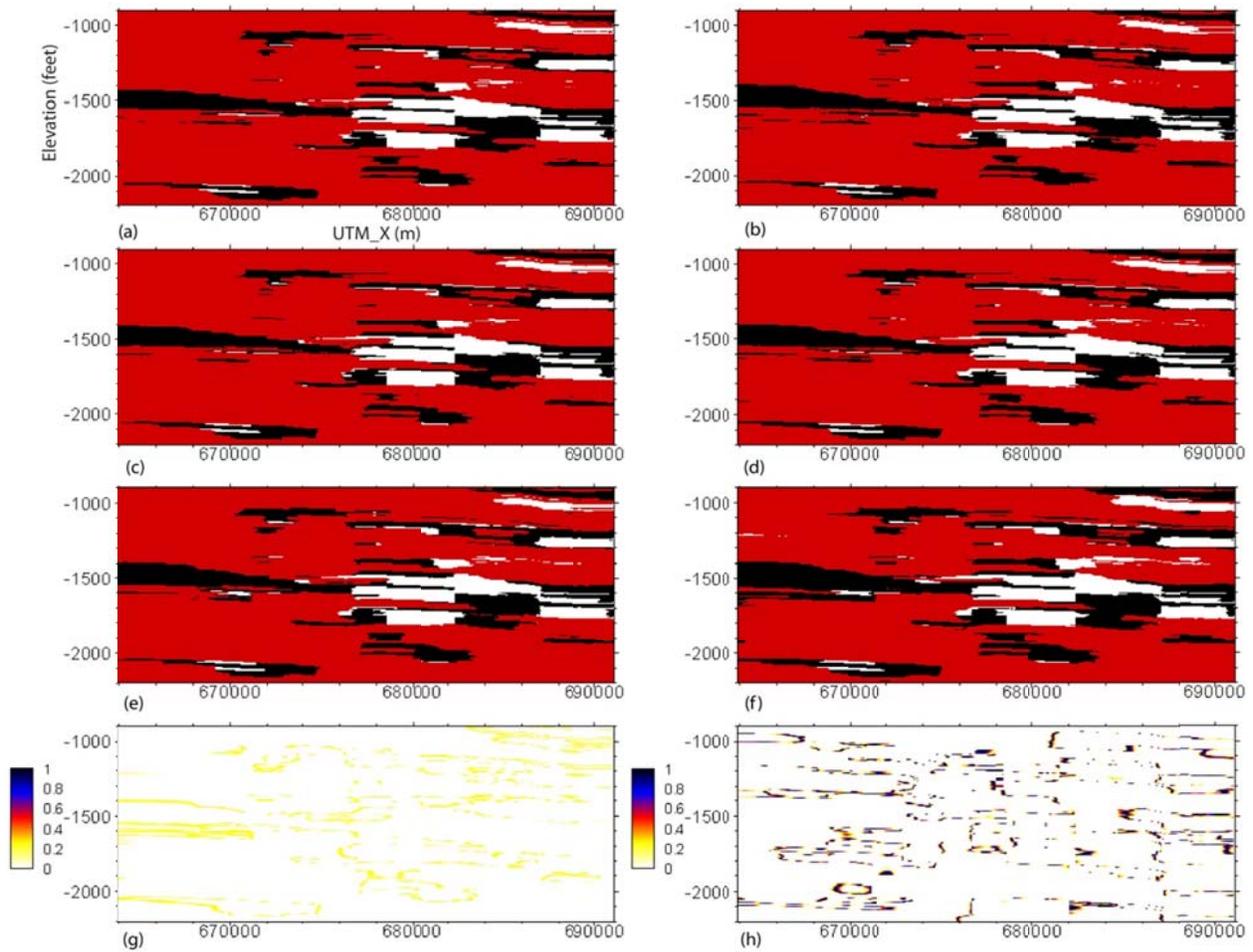


Figure 14 The architecture of the Baton Rouge fault in the modeling area (see Figure 3): (a)-(e) different realizations, (f) ensemble average of 100 realizations, (g) ensemble variance of 100 realizations for south of the fault, (h) ensemble variance of 100 realizations for north of the fault. Black areas in (a)-(f) are clay units north of the fault. Gray areas are clay units south of the fault. White areas show potential hydraulic connections formed by sand units from both sides of the fault. The fault cross sections are based on 3D estimates that follow the UTM_X and UTM_Y coordinates of the fault line in Figure 3 [modified from Elshall et al., 2013]

5.2.2 Leaky faults

The architecture of the Denham Springs-Scotlandville fault in Figure 12 shows the role of the fault in controlling lateral hydraulic continuity. The conjoined areas of sand and clay units at the fault naturally create horizontal flow barriers. Potential horizontal flow pathways may be identified at the places where sands are conjoined at the fault. The results in Figure 12 show a wide extent of horizontal flow pathways through the Denham Springs-Scotlandville fault to the

“1,200-foot” sand and the “2,000-foot” sand in the middle domain. On the other hand, the “1,500-foot” and “1,700-foot” sands in the middle domain show fewer flow pathways for hydraulic connection through the Denham Springs-Scotlandville fault. Figure 13 and Figure 14 show that there are a moderate number of horizontal flow pathways to the “1,200-foot” sand and very limited horizontal flow pathways to the “2,000-foot” sand in the middle domain through the Baton Rouge fault. However, the horizontal flow pathways through the Baton Rouge fault to the “1,500-foot” sand and the “1,700-foot” in the middle domain sand are extensive.

The actual fault permeability for the flow pathways depends on the clay content in the fault zone [Bense and Person, 2006], which directly impacts on groundwater head across the fault. The USGS groundwater data suggested that the Baton Rouge fault is laterally a low-permeability fault zone as reflected by significant head differences across the fault [Meyer and Rollo, 1965; Rollo, 1969]. The groundwater modeling results confirmed a low fault permeability for the Baton Rouge fault at the “2,000-foot” sand [Torak and Whiteman, 1982] and at the “1,500-foot” sand [Tsai and Li, 2008; Tsai 2010]. The literature shows that the Denham Springs-Scotlandville fault has not been well studied. The current groundwater levels (EB-168 and EB-652) from the USGS database do not suggest significant hydraulic discontinuities across the Denham Springs-Scotlandville fault for the “1,500-foot” sand.

Both faults are leaky faults. The flow pathways of the Denham Springs-Scotlandville fault allow groundwater to flow into the middle domain, which is heavily pumped. The flow pathways of the Baton Rouge fault allow saltwater intrusion into several freshwater sands [Tomaszewski, 1996; Lovelace, 2007, 2009; Anderson, 2012].

5.2.3 Quantification of structural geology parameters

Figure 15 shows the simulated aquifer architecture from the “1,200-foot” sand to the “2,000-foot” sand in the middle domain based on the GP method. The “1,200-foot” sand connects vertically to the “1,500-foot” sand. The “2,000-foot” sand is clearly separated from the “1,700-foot” sand by a confining layer. There are four sand units between the “1,200-foot” sand and the “2,000-foot” sand, which are generally classified as the “1,500-foot” sand and the “1,700-foot” sand [Griffith, 2003]. Unlike the distinguishable “1,200-foot” sand and the “2,000-foot” sand, the separation of the “1,500-foot” sand from the “1,700-foot” sand is not well-defined in the published cross sections [Rollo, 1969; Griffith, 2003]. The findings of this study also show that they are not clearly separable. Therefore, in this study the “1,500-foot” sand and the “1,700-foot” sand are together treated as a single unit.

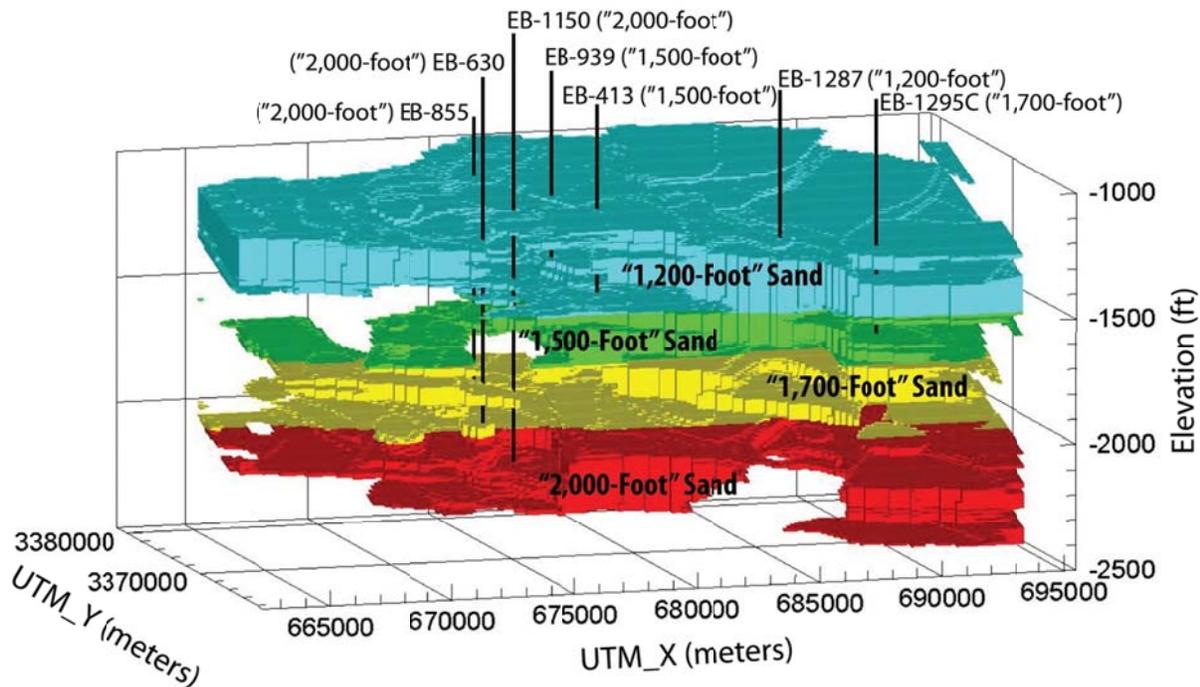


Figure 1 Hydrofacies architecture of the “1,200-foot” sand to the “2,000-foot” sand in the middle domain. The wells in the figure are public supply and industrial wells [modified from Elshall et al., 2013]

This section presents a regression technique and a clustering technique to quantify sand proportion, formation dip and sand unit offset on the faults.

The main purpose of the regression technique is to determine the dip of a sand unit. Each sand unit shown in Figure 16 is first visually distinguished. Given an east-west cross section, the top, middle and bottom elevations of sand units for the cross section are averaged. Repeating it for different east-west cross sections along the dip direction, the elevation points for the sands shown in Figure 17 are obtained. Then, linear regression method weighted by the sample size is used to fit straight lines to the top, middle and bottom elevation points; and therefore, the dip is determined by the slope. Based on Figure 17, the displacements of a sand unit on the faults can be calculated by the line dislocation. Moreover, based on Figure 16 the proportion of sand in each domain can be calculated.

The clustering method is the second technique that is used to quantify the same geological structure parameters. Chamberlain [2012] determines the proportions of sand in clay with depth in a series of strike sections perpendicular to dip to identify sand-dominated zones in a sequence with a high degree of lateral stratigraphic variation. A similar approach is used to calculate the sand proportion with depth in east-west using 200-foot (61 m) wide stripes immediately north and south of the faults. Four strips are used such that two strips are along the Baton Rouge fault from south and north and two strips are along the Denham Springs-Scotlandville fault from south and north. Then the variation in the proportion of the sand with depth north and south adjacent to the Baton Rouge fault and the Denham Springs-Scotlandville fault is determined as shown in Figure 18.

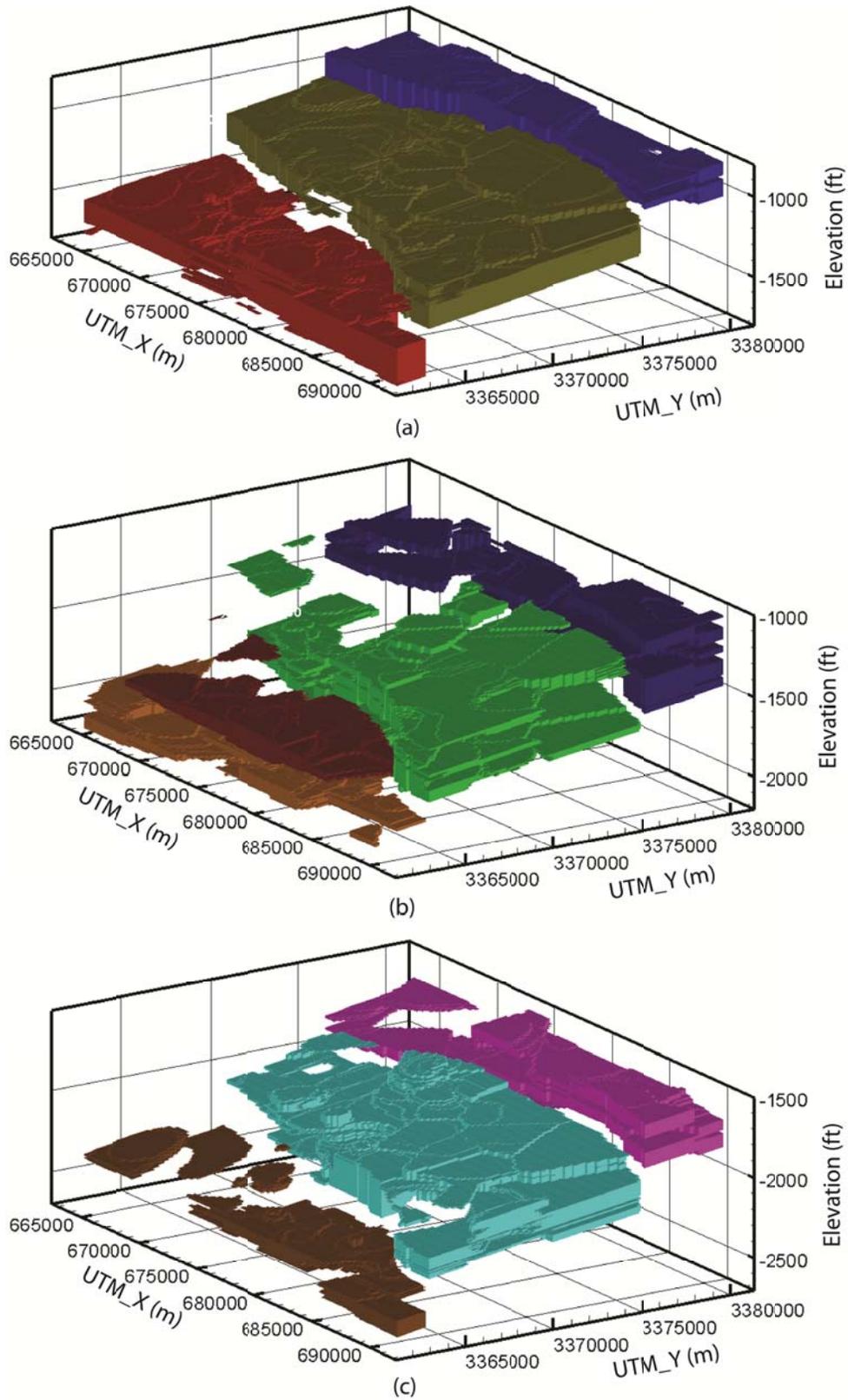


Figure 16 Sand units and displacements on the faults for (a) the “1,200-foot” sand, (b) the “1,500-1,700-foot” sands, and (c) the “2,000-foot” sand [modified from Elshall et al., 2013]

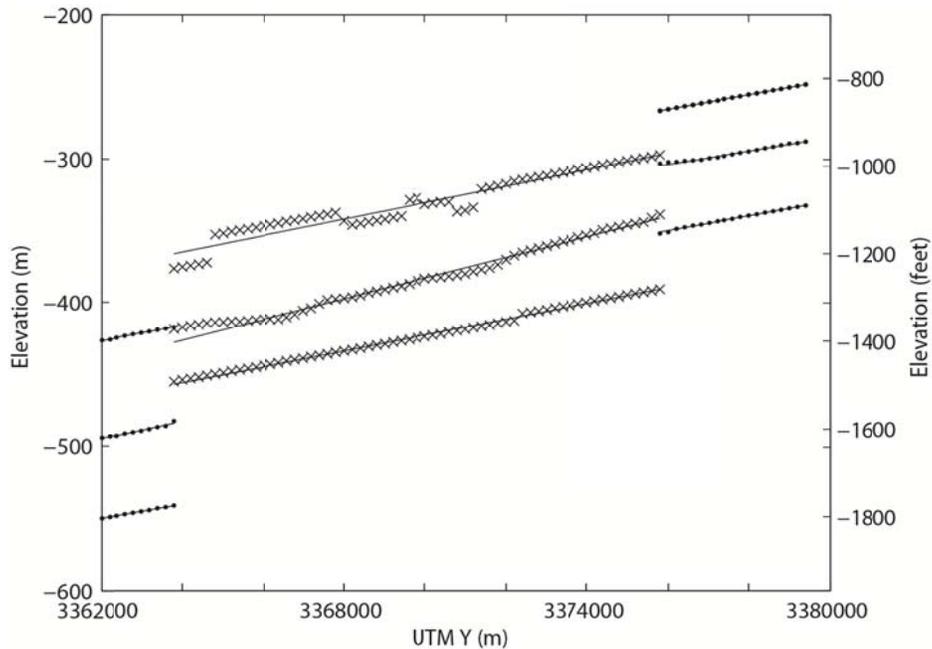


Figure 17 Top, middle and bottom elevations of the “1,200-foot” sand using the regression method [Elshall et al., 2013]

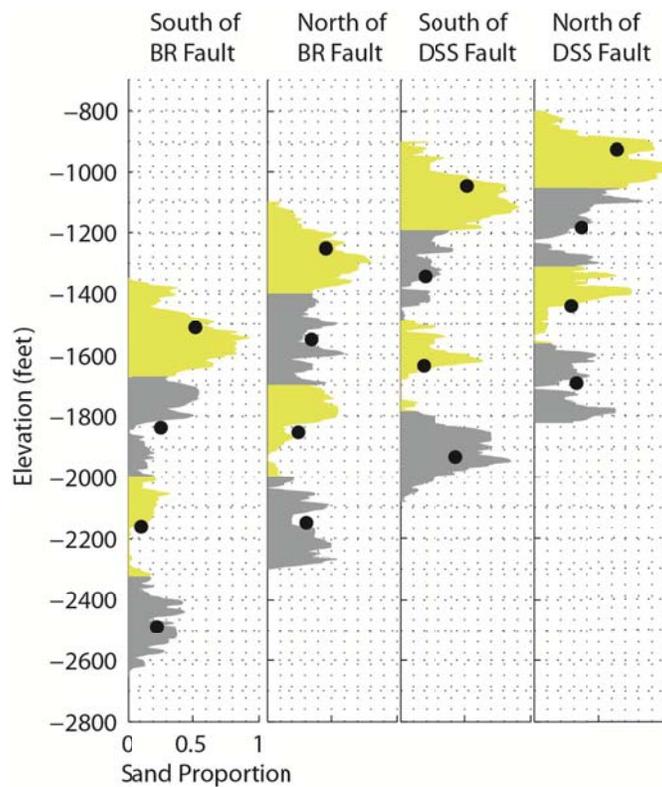


Figure 18 Variation in the proportion of sand with depth immediately south and north of the Baton Rouge and Denham Springs-Scotlandville faults using the clustering method. Black dots are centroids of the clusters [modified from Elshall et al., 2013].

The k-means clustering [Lloyd, 1982] is used to analyze Figure 18 and determine the cluster centers and boundaries. The decision variable is the cluster elevation. The k-means automatically designate different clusters representing different aquifer units. The k-means provides the first moment of each cluster, which represents the mean elevation of a slice at a distance of 200 feet (60.96 m) along the fault and the mean sand ratio of this cluster. Since the number of clusters needs to be predefined for k-means method, several numbers of clusters were tested. It was found that using four clusters give excellent solution in terms of being able to accurately delineate the bottom elevation of the four aquifer units for the considered slice. This can be verified for example by comparing the cluster elevations in Figure 18 (Middle-South) to the aquifer units elevations in Figure 15. The four clusters in the depth sequence represent the “1,200-foot” sand, the “1,500-foot” sand, the “1,700-foot” sand, and the “2,000-foot” sand.

According to Figure 18, the cluster centers and boundaries can be used to determine sand displacement on the faults and formation dip for the sands in the middle domain. Although the clustering method was able to identify the “1,500-foot” sand and “1,700-foot” sand as two separate sand units, they are represented as one unit with a mean value in order to compare the results with the regression method.

The calculated sand proportions of individual sands given the vertical intervals are shown in Table 5. Both methods show a high sand proportion for the “1,200-foot” sand, which has more than 13% of sand for the south and middle domain and has more than 16% of sand in the north domain. The high sand proportion reflects the massive nature of the “1,200-foot” sand, as shown in Figure 16(a). The “1,500-1,700-foot” sands have more than 12% of sand estimated by the regression method. However, the clustering method estimates a low sand proportion of 9.4% for the south domain and high sand proportion 16.1% for the north domain. Although most of the

“1,500-foot” sand in the industrial district and in West Baton Rouge Parish is missing as shown in Figure 16(b), which is potentially due to erosion as suggested by Chamberlain [2012], the sand proportion is similar to that of the “1,200-foot” sand in the middle domain. This is because of a significant proportion of the “1,500-1,700-foot” sands extends over an extensive depth interval in East Baton Rouge Parish. The “2,000-foot” sand has a low sand proportion of around 10% for the middle domain and less than 8% in the north and south domains due to a significant amount of sand missing in West Baton Rouge Parish, as shown in Figure 16(c). In summary, the total sand proportion in the south domain is less than 30% and is more than 34% in the middle domain and is around 40% in the north domain. The calculated total sand proportion for the entire modeling area is 34%, which is consistent with the total proportion of the sand indicators of the electric logs data 33.8% and drillers’ logs data 33.9% for the considered range.

Table 5 Estimated Sand Proportions [Elshall et al., 2013]

Sand	Regression Method			Clustering Method		
	South	Middle	North	South	Middle	North
“1,200-foot” sand	0.139	0.131	0.162	0.130	0.123	0.161
“1,500-1,700-foot” sands	0.112	0.122	0.128	0.094	0.122	0.161
“2,000-foot” sand	0.045	0.120	0.104	0.059	0.092	0.079
All sands	0.296	0.373	0.394	0.282	0.337	0.402

The calculated dips are shown in Table 6. The mean dip for the “1,200-foot” sand and the “1,500-foot” sand is 0.30° , and for the “2,000-foot” sand is 0.38° using the regression method. The mean dips for the middle domain using the cluster method are similar for all sands, which vary from 0.33° to 0.35° . Moreover, the dip increases with depth [Griffith, 2003]. The average dip for all sand units is 0.33° . This is comparable to the dip 0.30° estimated from the cross sections in Griffith [2003] and 0.29° from the inverse solutions in Table 4.

Table 6 Estimated Formation Dip (degrees) for Sand Units [Elshall et al., 2013].

Sand	Regression Method			Clustering Method
	South	Middle	North	Middle
“1,200-foot” sand	0.28±0.03	0.32±0.01	0.29±0.03	0.33±0.05
“1,500-1,700-foot” sands	0.28±0.02	0.32±0.03	0.31±0.03	0.34±0.05
“2,000-foot” sand	0.41±0.22	0.38±0.06	0.34±0.14	0.35±0.06

The sand displacement on the faults is shown in Table 7. The clustering method estimates sand displacements on the Baton Rouge fault which increase from 262 ft (79.2 m) to 337 ft (102.7 m) for the “1,200-foot” sand to the “2,000-foot” sand, and are 20 to 30 ft (6.1 to 9.1 m) more than displacements calculated by the regression method. Durham and Peeples [1956] estimated a 344-ft (104.9 m) displacement on the Baton Rouge fault for the “2,000-foot” sand, which is close to the result of the clustering method. Both methods have similar estimated sand displacements on the Denham Springs-Scotlandville fault for the “1,200-foot” and the “1,500-1,700-foot” sands, which are 120 ft. (36.6 m) and 179 ft. (54.6 m), respectively. The sand displacement on the “2,000-foot” sand is estimated to be 239 ft. (72.8 m) using the clustering method, which is 50 ft. (15.2 m) more than the regression method. In summary, the sand displacement on the Baton Rouge fault is 100 ft. (30.5 m) to 140 ft. (42.7 m) more than that on the Denham Springs-Scotlandville fault. Also, the fault throw appears to increase with depth.

Table 7 Estimated Sand Unit Displacements in Feet (Meters) on the Baton Rouge (BR) fault and the Denham Springs-Scotlandville (DSS) fault [Elshall et al., 2013]

Sand	Regression Method		Clustering Method	
	BR Fault	DSS Fault	BR Fault	DSS Fault
“1,200-foot” sand	241±62 (73.4±18.9)	114±54 (34.7±16.5)	262±12 (79.9±3.7)	120±20 (36.6±6.1)
“1,500-1,700-foot” sands	290±59 (88.4±18.0)	173±50 (52.7±15.2)	298±17 (90.8±5.2)	180±28 (54.9±8.5)
“2,000-foot” sand	307±38 (93.6±11.6)	187±57 (57.0±17.4)	337±14 (102.7±4.3)	239±20 (72.8±6.1)

5.2.4 Interconnections between aquifer units

Since most of the industrial and public supply wells in Baton Rouge are screened in sand units in the middle domain, it is important to understand the interconnections between sand units in this domain. As shown in Figure 19(a), the “1,200-foot” sand in the middle domain receives groundwater from the “1,200-foot” sand and the “1,500-1,700-foot” sands at the north due to the throw on the Denham Springs-Scotlandville fault. The flow pathways through the Denham Springs-Scotlandville fault are extensive according to Figure 12. The “1,200-foot” sand connects to the lower portion of the “1,000-foot” sand and upper portion of the “1,200-foot” sand south of the Baton Rouge fault, where the extent of flow pathways are moderate, as shown in Figure 12. It is interesting to see the connection of the “1,200-foot” sand to the “1,500-foot” sand in the southeastern area of the middle domain, which indicates partial recharge to the “1,500-foot” sand.

The “1,500-1,700-foot” sands in the middle domain shown in Figure 19(b) connect to the same sand unit north of the Denham Springs-Scotlandville fault. The extent of lateral flow pathways through the Denham Springs-Scotlandville fault are not significant as shown Figure 12, which indicates the importance of the “1,200-foot” sand at the top to supply groundwater to these sands. The “1,500-1,700-foot” sands extensively connect to the “1,200-foot” sand and the “1,500-foot” sand in the south as shown Figure 13 due to significant fault throw on the Baton Rouge fault.

The “2,000-foot” sand in the middle domain shown in Figure 19(c) connects to the same sand and upper portion of the “2,400-foot” sand north of the Denham Springs-Scotlandville fault. The connections are significant as shown in Figure 12 due to significant fault throw. The “2,000-foot” sand has a very limited connection to the lower portion of the “1,700-foot” sand south of

the Baton Rouge fault. As shown in the following discussion, the limited pathways still create enough avenues for saltwater into the “2,000-foot” sand.

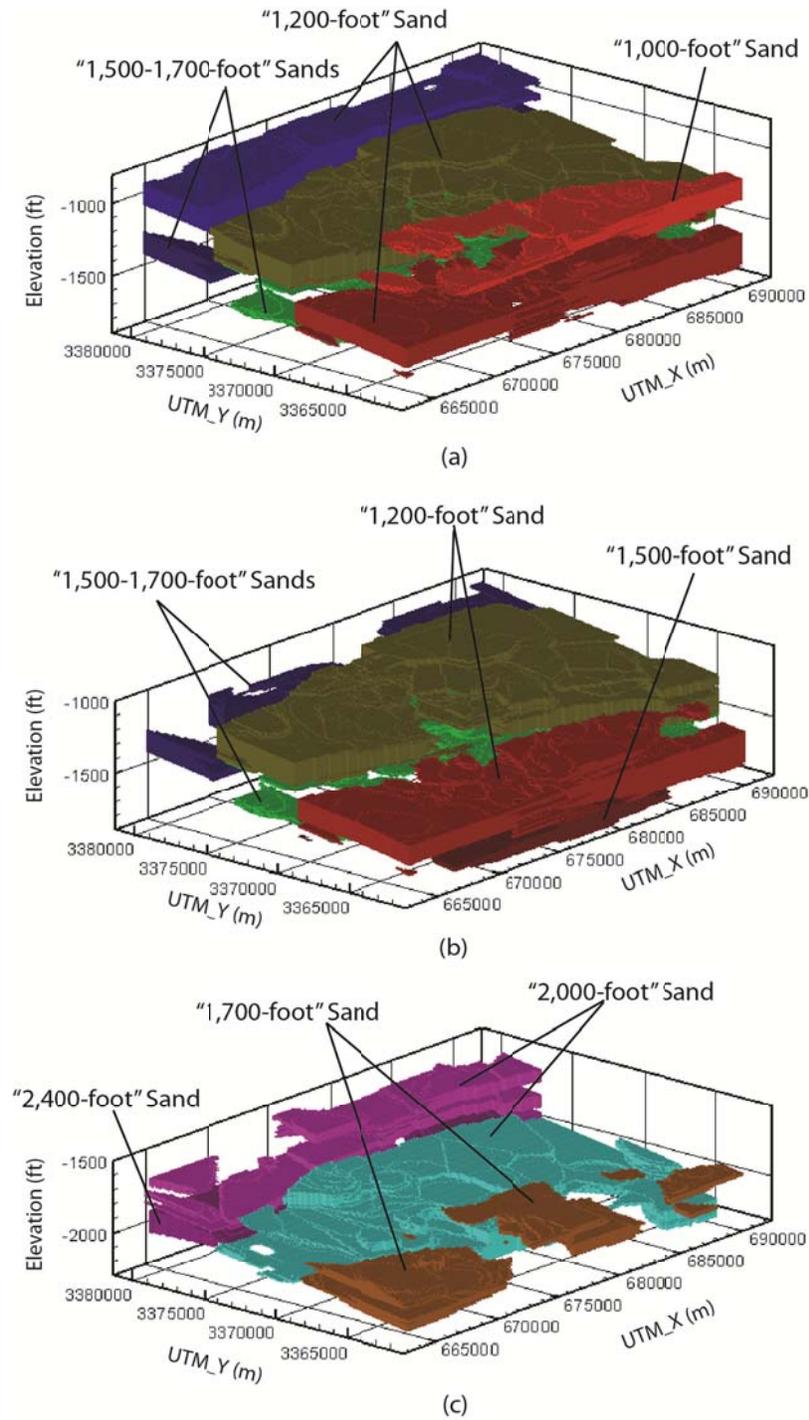


Figure 19 Interconnections of sand units to the sand units in middle domain for (a) the “1,200-foot” sand, (b) the “1,500-1,700-foot” sands, and (c) the “2,000-foot” sand [modified from Elshall et al., 2013].

5.2.5 Baton Rouge aquifer-fault connections for saltwater intrusion

The vulnerability of the aquifer system to saltwater intrusion is assessed by mapping the potential flow pathways across the Baton Rouge fault with respect to the locations of municipal and industrial wells in the “1,200-foot” sand, the “1,500-foot” sand and the “2,000-foot” sand, which are currently under the threat of saltwater encroachment [Lovelace, 2007]. Using two-dimensional cross sections that are based on the three-dimensional hydrofacies model is adequate for showing that the pumping wells are connected to the source of the saline water south of the Baton Rouge fault. Figure 20 shows the merger of the “1,200-foot” sand and the “1,500-foot” sand, which Figure 15 depicts in three dimensions. The two sand units connect to the “1,200-foot” sand south of the Baton Rouge fault. Three public supply wells are active along this cross section. There is no report of saltwater encroachment within the area of this cross section.

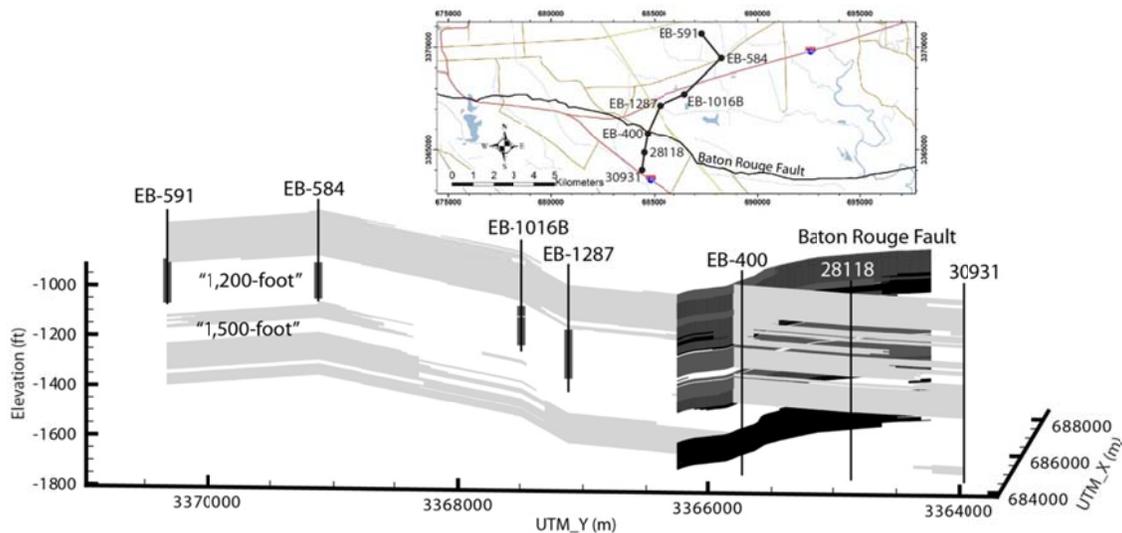


Figure 20 A cross section shows the merger of the “1,200-foot” sand and the “1,500-foot” sand and their connection to the “1,200-foot” sand south of the Baton Rouge fault. Sand units are transparent. EB-1287, EB-1016B, and EB-584 are public supply wells. The color lines and areas in the inset map are defined in Figure 1 [Elshall et al., 2013].

Figure 21 shows a sand connection from the “1,200-foot” sand south of the fault to municipal wells EB-413 and EB-939 screened in the “1,500-foot” sand. High chloride concentrations have been observed at observation well EB-917. The flow pathway is consistent

with the results of saltwater intrusion modeling for the “1,500-foot” sand [Tsai, 2010]. The identified leaky area also explains the salinity distribution in the depth around 1,500 feet below land surface documented by Anderson [2012], where relatively low chloride concentrations are observed in the south of the leaky area. Prior to development, the leaky area used to act as a natural outlet to discharge fresh groundwater to the south of the Baton Rouge fault. The groundwater level data in the 1930s from the online USGS National Water Information System shows southward flow direction. Well EB-326 had a water level of 64 ft (19.51 m) above NGVD29 in October 1936 in the “1,200-foot” sand south of the fault. The head data at EB-84, EB-89, EB-311, and EB-312 indicates a water level 75 ft (22.86 m) above NGVD29 in October 1936 in the “1,500-foot” sand north of the fault. This difference in water levels confirms that during pre-development pumping groundwater level in the “1,500-foot” sand north of the fault was higher than that in the “1,200-foot” sand south of the fault. However, heavy pumping in the “1,500-foot” sand at Lula station and Government Street station reversed the flow gradient causing brackish water to flow northward into the “1,500-foot” sand [Morgan and Winner, 1964; Meyer and Rollo, 1965; Rollo, 1969; Tomaszewski, 1996].

Two leaky areas connected to the “2,000-foot” sand through the Baton Rouge fault are identified in Figure 13. Figure 22(a) shows a saltwater intrusion path starting in East Baton Rouge Parish to production well EB-1150 [Lovelace, 2009]. Figure 22(b) shows the detailed cross section that illustrates a potential saltwater intrusion path in West Baton Rouge Parish to production wells EB-630 and EB-1263. Again, these two pathways explain the spatial variations in salinity at a depth around 2,000 feet below land surface documented by Anderson [2012], where low groundwater salinities are found in the south of the leaky areas. For details on saltwater concentrations, the interested reader can compare the main flow pathways in the

“1,500-foot” sand in the middle, and the east and west flow pathways in the “2,000-foot” sand with the saltwater concentration maps of Anderson [2012, Figure 4.9(c-d)]. Since these two studies were conducted independently, it is important to note that the potential pathways that are identified from the results of this study coincide spatially with leaky areas of high salinities along the fault. This also suggests that fresh groundwater flowed southward from the “2,000-foot” sand north of the Baton Rouge fault into the “1,700-foot” sand south of the Baton Rouge fault prior to heavy pumping. Current high groundwater withdrawals from the “2,000-foot” sand by the water company and the industries have reversed the flow direction and have caused saltwater intrusion into the “2,000-foot” sand [Lovelace, 2007].

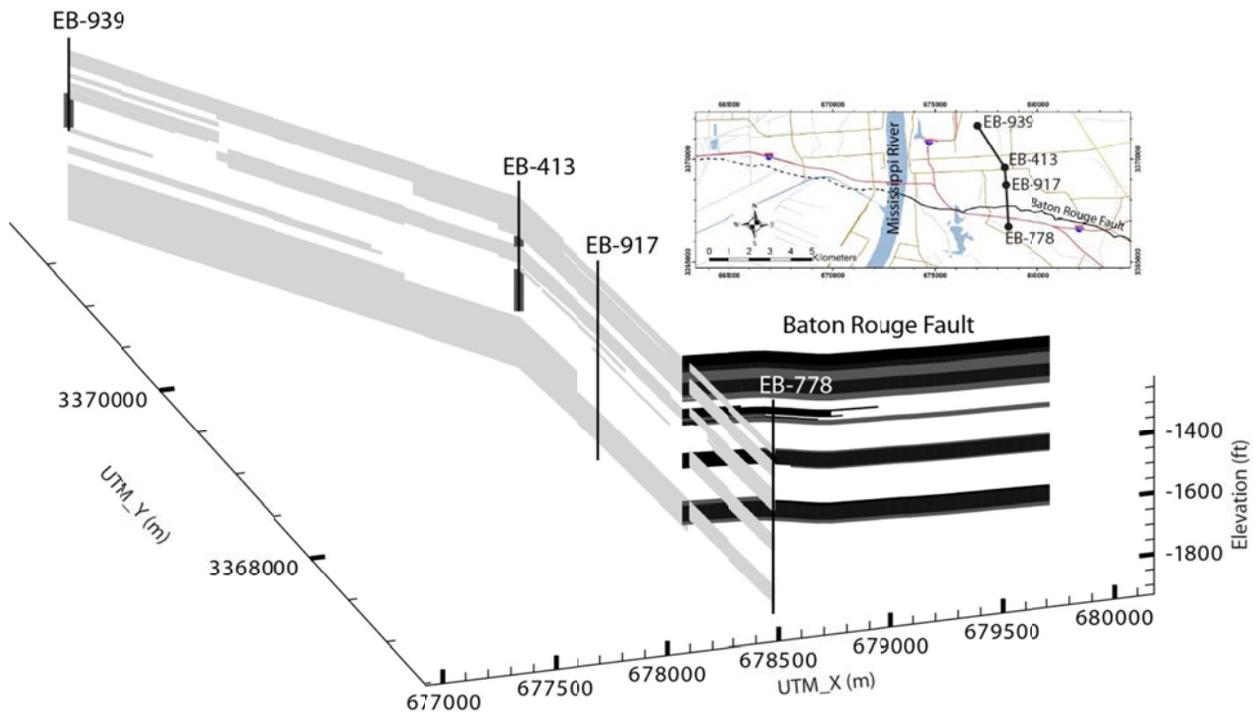


Figure 21 A cross section shows the connection of the “1,500-1,700-foot” sands to the “1,200-foot” sand south of the Baton Rouge fault. Sand units are transparent. EB-413 in the Government Street station and EB-939 in the Lula station are public supply wells. The color lines and areas in the inset map are defined in Figure 1 [Elshall et al., 2013]

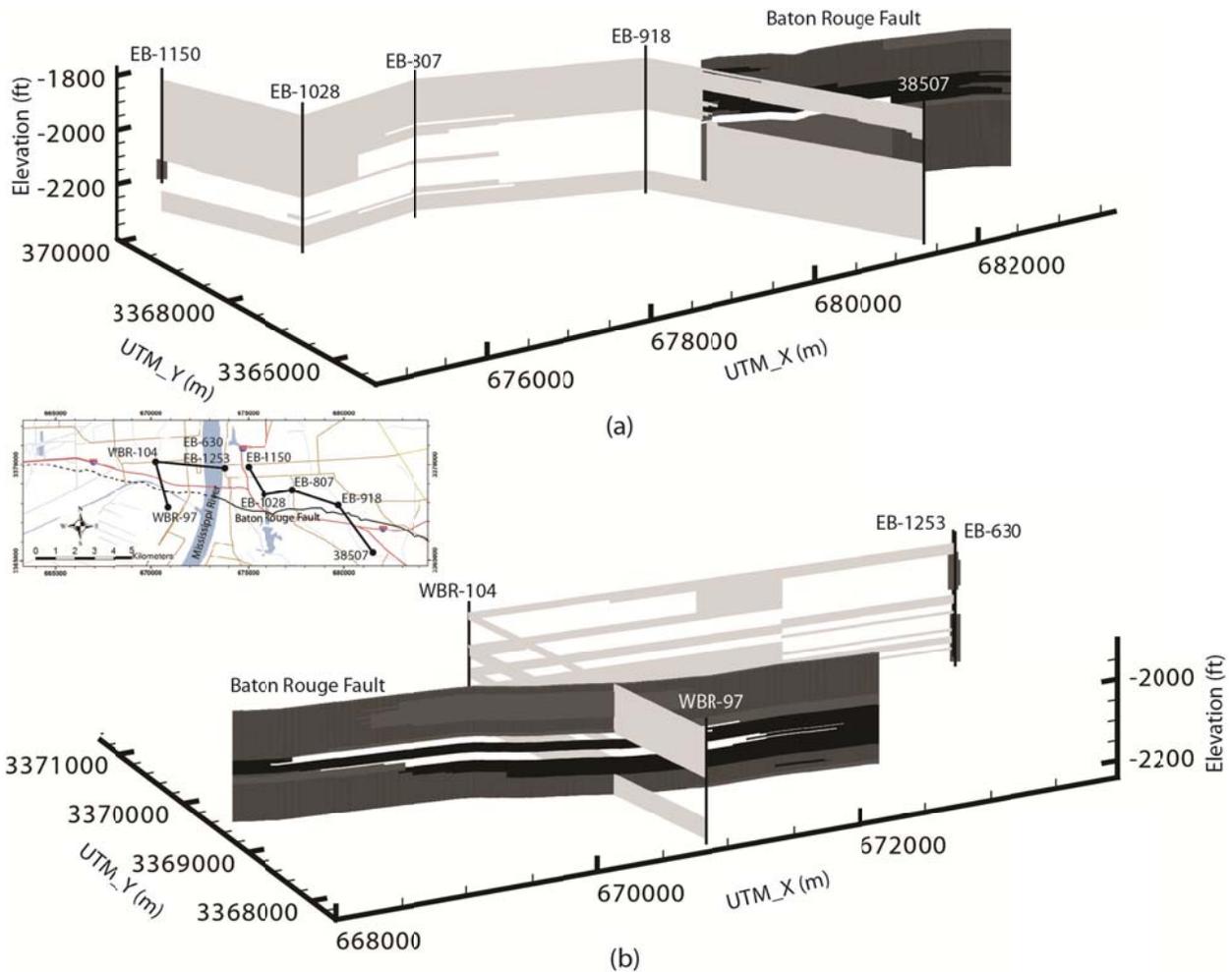


Figure 22 Two cross sections show the connection of the “2,000-foot” sand to the “1,700 foot” sand south of the Baton Rouge fault. Sand units are transparent. EB-1150, EB-1253, and EB-630 are public supply wells. The color lines and areas in the inset map are defined in Figure 1. [Elshall et al., 2013].

5.3 Conclusions

The generalized parameterization (GP) method is shown to be an effective indicator geostatistical method for reconstructing hydrofacies architecture of a complex fluvial binary siliciclastic aquifer system. By depicting the spatial extent of sand units, the derived geological architecture shows interconnections among different sand units, flow pathways across faults, and connections of production wells to the potential leaky areas of the Baton Rouge fault. The regression method and the clustering method are effective methods for post-analyzing important

geological parameters such as formation dip, sand proportion, and sand unit displacement on the fault.

The study finds strong hydraulic connection between the “1,200-foot” sand and the “1,500-foot” sand. Merger of the sand units indicates groundwater recharge from the “1,200-foot” sand to the “1,500-foot” sand. However, there is a distinct clay confining layer to separate the “2,000-foot” sand from the “1,700-foot” sand. The hydrofacies architecture also reveals four sand deposits that compose the “1,500-foot” sand and the “1,700-foot” sand. In general, sand deposition is not uniform, due to spatial and temporal variations in fluvial processes [Chamberlain, 2012]. The study shows that there is large amount of missing sand in “1,500-foot” sand in the industrial district and in West Baton Rouge Parish, which is possibly due to the presence of an erosional unconformity [Chamberlain, 2012].

The sand unit displacement on the Baton Rouge fault and the Denham Springs-Scotlandville fault is significant. The Baton Rouge fault has higher sand displacement than the Denham Springs-Scotlandville fault. Displacement increases over depth. Due to non-uniform fault throw and sand deposition, the study reveals non-uniform flow pathways that connect different sand units at the fault planes. In particular, the identified flow pathways through the Baton Rouge fault provide important information for understanding patterns of salinization of freshwater aquifers in the East Baton Rouge Parish.

Establishing the detailed 3-dimensional fault-aquifer sedimentary architecture of the Baton Rouge aquifer system is a prerequisite to future work on saltwater intrusion in the study area. The detailed fault-aquifer architecture provides information about connections between the aquifer units, which have significant implications on the salt-water intrusion problem. For example, the simulation of the salt-water intrusion in “1,200-foot” sand and “1,500-1,700-foot”

sand should not be done separately, since they are very well connected in the middle domain. On the other hand, the industrial aquifer unit “2,000-foot” sand is not connected to any of the units above. More importantly, the identified flow pathways through the Baton Rouge fault are prerequisites for modeling salt-water intrusion from the south to the north of the Baton Rouge fault. For example, without the fine discretization that this sedimentary architecture model provides especially in the vertical direction, the narrow connection in the “2,000-foot” sand at the east that allows major leakage from the south would have been missed. Finally, by accounting for the geometry and locations flow pathways across the faults and the interconnections of different aquifer units, the sedimentary architecture makes the model structure of the salt-water intrusion model consistent with the real geologic structure of the aquifer which shall improve the salt-water intrusion model accuracy.

6 Groundwater flow model calibration and uncertainty quantification using CMA-ES

6.1 Synthetic groundwater flow problem

6.1.1 Design of the synthetic problem

This study uses CMA-ES algorithm to solve the inverse groundwater problem and to quantify the parameter related uncertainty. A synthetic steady-state groundwater flow problem is designed to compare CMA-ES with the other five algorithms to evaluate the robustness of CMA-ES in handling the search difficulties. The numerical model consists of an unconfined aquifer with a thickness of 400 m, a confined aquifer with a thickness of 100 m and an aquitard in between with a thickness of 100 m. The model top elevation is 200 m. The horizontal domain is 4500 m by 4500 m and is discretized into 9 by 9 cells as shown in Figure 23(a). The unconfined aquifer has a fixed head 1 m at the western boundary and is impervious for other three boundaries. The aquitard and confined aquifer have impervious boundaries. Hydraulic conductivity [m/s] for the unconfined aquifer is of two zones in Figure 23(b):

$$\mathbf{K}(x, y) = \begin{cases} 1 \times 10^{-2} & \text{for } x \leq 2500 \text{ m} \\ 7 \times 10^{-2} & \text{for } x > 2500 \text{ m} \end{cases} \quad (48)$$

The confined aquifer has a heterogeneous transmissivity field [m²/s]

$$\mathbf{T}(x, y) = -20\pi \cos(\pi x) \sin(\pi y) - 20\pi \sin(\pi x) \cos(\pi y) + 40\pi^2 (1 + x + y) \cos(\pi x) \cos(\pi y) \quad (49)$$

The vertical hydraulic conductance of the aquitard is 5×10^{-8} m²/s. Three wells are located in the unconfined aquifer as shown in Figure 23(a). Two injection wells are located in the low conductivity zone with injection rate of 10 m³/d for each well. One pumping well is located in the high conductivity zone with pumping rate of 20 m³/d. The model has a uniform surficial recharge of 5×10^{-5} m/s to the unconfined aquifer. The hydraulic gradient is 1.51% and 1.37%

for the unconfined and confined aquifers, respectively. MODFLOW-2005 [Harbaugh, 2005] is used to solve the steady-state flow problem.

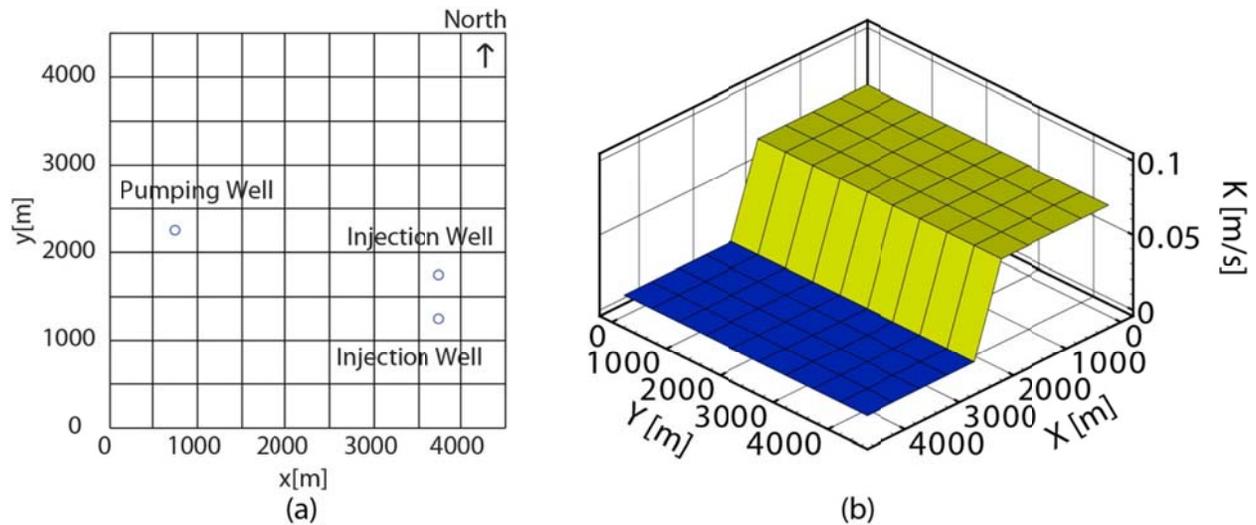


Figure 23 Synthetic problem: (a) Boundary conditions, pumping well and injection wells, (b) true hydraulic conductivity field K [m/s] for the unconfined aquifer.

6.1.2 Ill-posedness and search difficulties

A complex intersection of the unknown model parameters and the state variables, which is due to ill-posedness and search difficulties, underlies the objective function. Ill-posedness is due to the solution nonexistence, instability, insensitivity and nonuniqueness. The existence and instability of the solution are not a problem for an optimization algorithm. Note the difference between an adequate simulation model and a precise optimization solution. A simulation model is said to be more adequate if it corresponds better to the natural system. An optimization solution is said to be precise if it is near the global solution. Thus, given data and defined model structure, a global solution for estimated model parameters always exists regardless of the adequacy of the simulation model.

However, a more critical issue is the low sensitivity of the state variables to unknown model parameters. For example, the head data may contain little information about hydraulic conductivity particularly if the head difference is small. A second issue is nonuniqueness, which

arises in the absence of sufficient data to limit the problem to the true parameter set. This is particularly apparent in steady-state models. These two issues can be addressed by increasing the number of data points or by introducing new data types. To minimize ill-posedness arising from insensitivity and non-uniqueness, dense observation data set is used.

Since the objective is to evaluate the capabilities of different algorithms in obtaining a precise solution in a difficult search landscape, the synthetic example is designed to minimize the ill-posedness problem while increasing the search difficulties. Ruggedness, ill-conditioning, inseparability, noise and high dimensionality are the main search difficulties. A rugged function is a highly nonlinear, multimodal, nonsmooth or discontinuous function. Ill-conditioning occurs when the conditioning number, which is the ratio of the largest to smallest eigenvalues of the covariance matrix, is large such that the surfaces of the objective function have high curvature. First-order information such as the gradient direction is sufficient when conditioning number is small; otherwise second-order information such as covariance matrix is necessarily [Auger and Hansen, 2012]. Inseparability refers to the dependency between the model parameters such that the objective function cannot be minimized as a sequence of one-dimensional minimization problems of the unknown model parameters. In this synthetic problem, the identification of a simple two-zone hydraulic conductivity structure is actually challenging for many search algorithms due to strong correlation among gridded K values in the same zone and low correlation in the different zone. A large number of forward model runs to reach good solution precision is anticipated for more random and less correlated search algorithms.

Another critical challenge for the optimization algorithm performance in the inverse problem is the issue of incorporating ineffectual data, which may lead to imprecise inverse solutions. The ineffectual data is seen as unimportant signals or noises in the objective function

and can conceal the useful signals needed for the optimization process when the useful signals and unimportant signals overlap. Thus, an algorithm that can avoid the fitting of noises is more.

Finally, the curse of dimensionality, which is nonlinear increase of forward model evaluations with the increase of the number of unknowns, is a major search challenge for heuristic algorithms. This is mainly due to the power increase in search space. To amplify this challenge the synthetic inverse problem has 81 dimensions. Thus, a search strategy that is successful in small dimensions might fail in a problem with large dimensions. Another issue, which is indirectly related to the precision of the solution, is the high computational cost associated with the power increase in the search space.

6.1.3 Model parameters and calibration

The inverse problem is solved to estimate 81 unknown hydraulic conductivity values for each computational cell of the unconfined aquifer by minimizing the square root of sum of squared errors:

$$\min_{\mathbf{K} \in \mathbb{R}^n} f = \|\Delta^{obs} - \Delta\|_2 = \sqrt{\sum_{j=1}^L (\Delta_j^{2,obs} - \Delta_j^2)^2} \quad (49)$$

where Δ^{obs} is the vector of observed groundwater heads; Δ is a vector of simulated groundwater heads; $n = 81$ is the number of unknown model parameters; and $L=162$ is the number of head data consisting of 81 head data from the unconfined aquifer and the 81 head data from the confined aquifer. A complete error-free head data set is used to minimize the ill-posedness in order to compare the algorithm performance in terms of reaching a precise solution. The search range is from $K= 0.001$ to 0.1 m/s.

Algorithm performance comparison is carried out by the number of function evaluations and the number of iterations to reach a designated fitting error. A fitting error $f = 1 \times 10^{-3}$ is set

as the stopping criterion. If the algorithm cannot reach this value within 5×10^5 function evaluations, then the optimization terminates.

6.1.4 Algorithms tuning

To allow fair algorithm comparison, parameter tuning is needed to use each algorithm with its optimal parameter to achieve its most effective and efficient performance for the given problem. The assessment of effectiveness is defined as the ability of algorithm to reach a certain function value and efficiency is defined as number of function evaluations to reach to this value for sequential run and the number of iterations for parallel run. This section presents the initialization and parameter tuning results for CMA-ES, ACOR, , mDE, GA, PSO and L-M, respectively.

For all calibration runs, the initial values of the CMA-ES parameters are $\mathbf{p}_\sigma^{(0)} = \mathbf{p}_c^{(0)} = 0$, $\mathbf{p}_\sigma^{(0)} = \mathbf{p}_c^{(0)} = 0$, $\mathbf{C}^{(0)} = \mathbf{I}$, $\mathbf{v} = rand(n)$ and $\sigma^{(0)} = 0.5$ with the default strategy parameters [Hansen et al., 2003]. CMA-ES is quasi-parameter free with the population size $\lambda = 4 + 3\ln(n)$ being the only parameter to be tuned by the user. CMA-ES is a local search, which can become more global by increasing the population size λ . Thus, the tuning of CMA-ES is unproblematic. For a sequential run, it is recommended to start with the default population size and increase it in case that the desired fitting error is not reached. The default population size $\lambda = 17$ converged at fitting error 8.8×10^{-2} and did not reach the desired fitting error 1×10^{-3} . Increasing the population size to $\lambda = 50$ improves the fitting error to 1.4×10^{-2} . The third trial with $\lambda = 100$ the desired fitting error is reached. The optimal λ tuning for both sequential and parallel runs is described in details in a later section.

Tuning ACOR, mDE and GA are relatively easy since it has only two tuning parameters. For ACOR [Socha and Dorigo, 2008] the step size control parameter ζ and the ranking parameter q have clear roles. The parameter ζ is the most critical in terms of its impact on the algorithm performance as previously discussed and compared to step size $\sigma^{(g+1)}$ with path length control $p_{\sigma}^{(g+1)}$ of the CMA-ES. For tuning ACOR, $\zeta = 0.5, 0.6, 0.7, \text{ and } 0.9$ are first tested. Upon finding the optimal $\zeta = 0.6$, three q values of 0.6, 0.7 and 0.8 were tested for the optimal ζ . These tests showed that ACOR performed optimally for this problem at $\zeta = 0.6$ and $q = 0.7$ reaching a fitting error 22.14 at the stopping criterion. The ranking parameter q , which controls the diversification or intensification of the solution as discussed in Socha and Dorigo [2008], appears to have a small impact on the quality of the solution. In addition, different population size $\lambda = 100, 200, \text{ and } 300$ were tested for the optimal parameter set and found that λ has minimal effect. The mDE [Babu and Angira, 2006] has two parameters to tune, which are the crossover constant $CR \in [0,1]$ and weighting coefficient $F \in [0,2]$. Storn and Price [1997] recommended crossover constant $CR = 0.1$ for best result, and $CR = 1$ for fast convergence. Given $CR = 0.1$, five weighting coefficients $F = 0.1, 0.5, 1.0, 1.5, \text{ and } 2.0$ are tested. The optimum parameters for mDE are $CR = 0.1$ and $F = 0.5$, yielding a fitting error 35.80. Similarly, the tuning of GA [Haupt and Haupt, 2004] is relatively easy. First, mutation rates $\mu \in [0.05, 0.9]$ with an increments of 0.05 were tested. A mutation is an operator that randomly alters the different dimensions of the current solution to produce a new solution. Having determined the optimum mutation rate, then 0.4, 0.5 and 0.6 selection fractions N_{good} of the solutions in an iteration to be kept for generating new solutions are tested. GA performed optimally at $\mu = 0.6$ and $N_{good} = 0.5$, yielding a fitting error 24.80.

Unlike the straightforward tuning of CMA-ES or the relatively easy tuning of ACOR, mDE and GA, the tuning of PSO [Socha and Dorigo, 2008] is complex since it has at least four parameters to tune, which are the population size λ , the number of function evaluations, the weight w and the initial partial velocity vi . Moreover, the large amount of randomness in the search strategies adds to the complexity of the tuning task since for repeated runs with the same parameter set, the optimal solution can be different by several orders of magnitude. The result of the solutions of 5 repeated runs of CMA-ES with the parameter set $\lambda = 300$ tends to be comparable while the solutions from 5 repeated runs of PSO are largely different given the same parameter set $(\lambda, w, vi) = (300, 0.7, 2.0)$. However, to simplify the tuning process the number of function evaluations is kept fixed to the second stopping criterion. In addition, increasing the population size would generally improve the solution, thus the population size λ is fixed at 300. Therefore, the only two parameters left for tuning are the weight w and the initial velocity vi . Since the weight is more critical than the initial velocity. The tuning strategy is to find the optimal weight and then to find the optimal velocity for this weight. After testing different parameter sets, the optimal parameters were found $w = 0.5$ and $vi = 0.1$ with fitting error 3.65 at the stopping criterion. Similarly, the tuning of L-M algorithm is problematic since several initial solutions need to be evaluated. Although both the step size and the Marquardt constant are adapted, the initial solution and the constant $dk \in [0.01, 1]$ for incrementing the Jacobian matrix need to be tuned. The L-M performs optimally at $dk = 0.1$, yielding a fitting error 19.48 in only 10 iterations that is 900 function evaluations. Note that unlike the other algorithms which have $\lambda = 300$, L-M has a smaller number of function evaluations $\lambda = 90$ per iteration. These are $n + 1$ solutions to calculate the Jacobian matrix and the other 8 solutions to adapt the step size and the Marquardt constant.

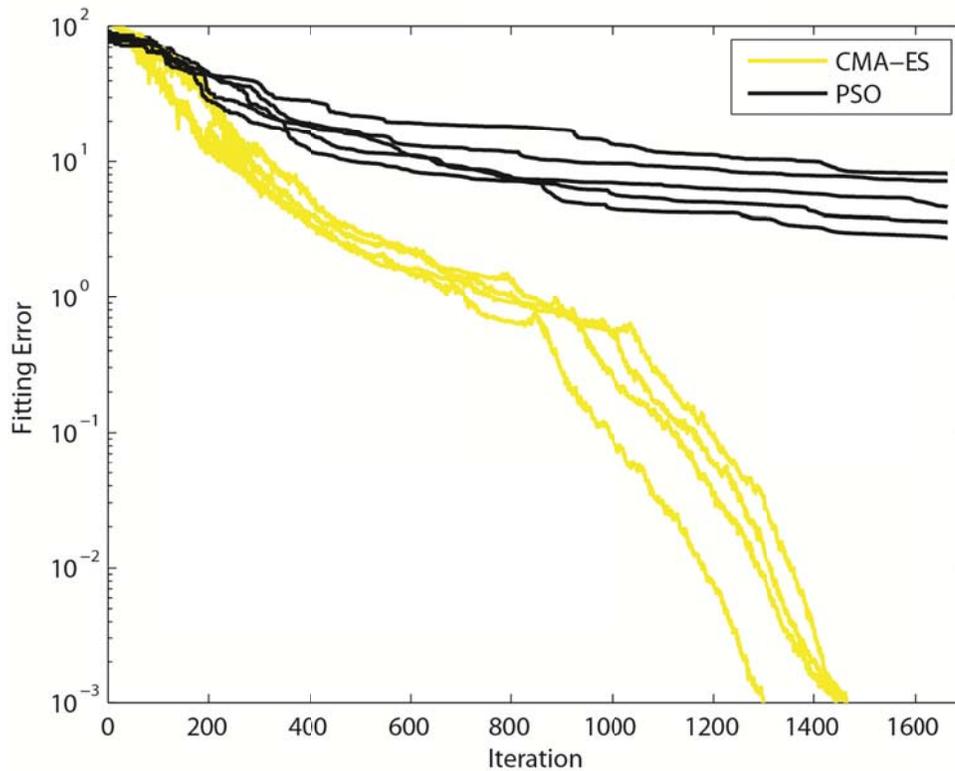


Figure 24 Convergence profiles of five runs to show the performance consistency of (a) CMA-ES and (b) Particle Swarm Optimization. Note that the CMA-ES shows only 4 profiles because two solutions are identical.

6.1.5 Performance comparison

For a population size $\lambda = 300$ except for the L-M that has $\lambda = 90$ function evaluations, the convergence profiles of the six algorithms are shown in Figure 24. Only CMA-ES succeeded in reaching the desired fitting error. PSO was able to reach a fitting error 3.65. The poor performance of ACOR is unexpected since theoretically the ACOR can handle non-separable functions by invoking correlation between decision parameters and can adapt to a rotating search space. However, the result shows a poor performance, which can be attributed to the fixed step size of ACOR as previously discussed. CMA-ES can perform global search similar to other heuristic algorithms, yet unlike other heuristic algorithms the CMA-ES is capable of systematic local convergence as shown in Figure 24.

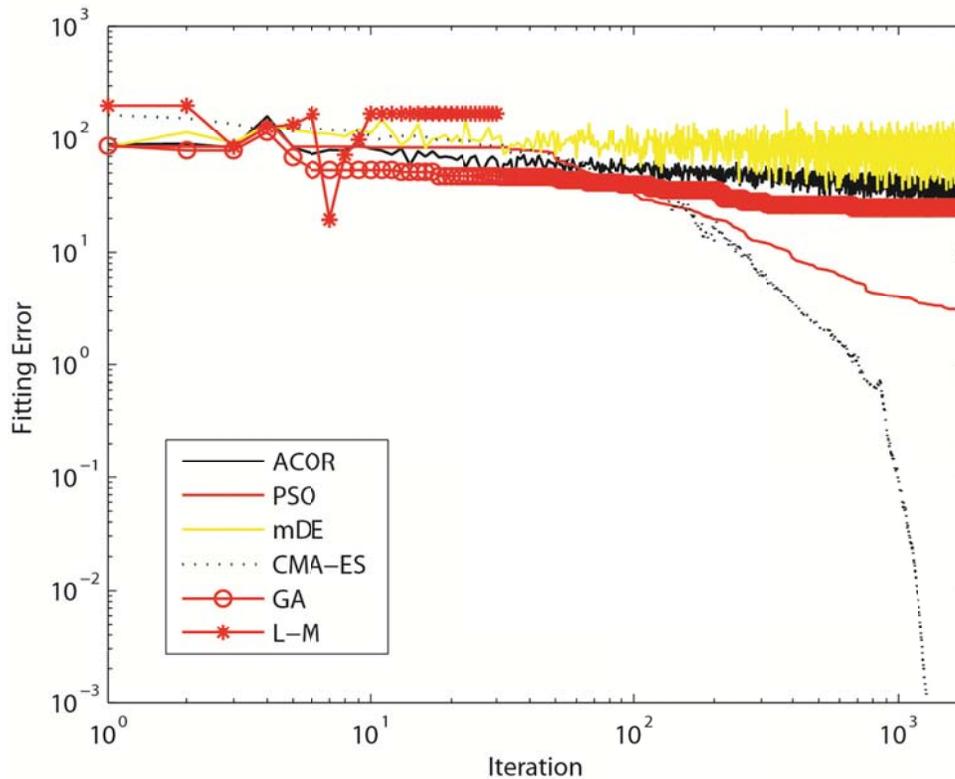


Figure 25 Convergence profiles of Ant Colony Optimization for Real Domain (ACOR), Particle Swarm Optimization (PSO), modified Differential Evolution (mDE), CMA-ES, Genetic Algorithm (GA) and Levenberg-Marquardt (L-M).

Figure 26 shows the best solutions of the six algorithms in a descending order according to their fitting errors. PSO and CMA-ES succeeded in recognizing the two hydraulic conductivity zones. However, the PSO did not succeed in overcoming the noise, which is created by the injection and pumping wells, resulting in imprecise hydraulic conductivity estimation. Second, since PSO cannot effectively exploit correlation between hydraulic conductivity values, the high-conductivity zone did not smooth out. The CMA-ES overcame these two pitfalls. That is mainly due to the utilization of second-order learning through the adaptation of the covariance matrix along with the careful adaptation of the step size to allow for systematic convergence.

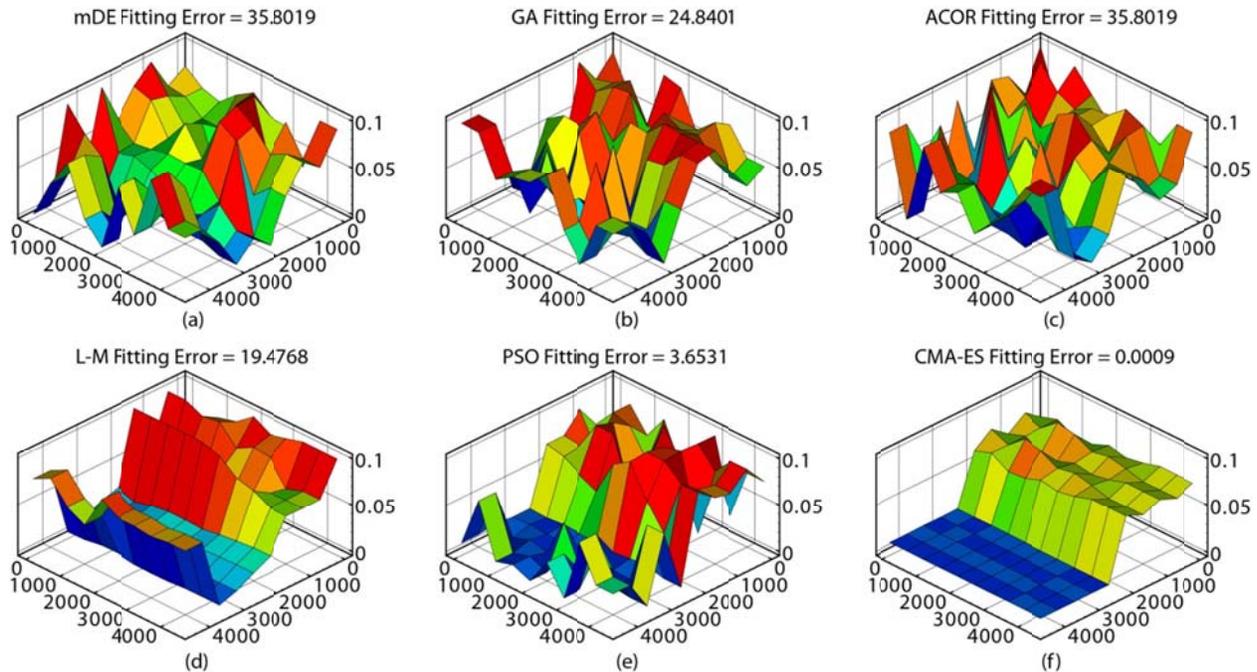


Figure 26 Hydraulic conductivity solutions for the unconfined aquifer: (a) modified Differential Evolution (mDE), (b) Genetic Algorithm (GA), (c) Ant Colony Optimization for Real Domain (ACOR), (d) Levenberg-Marquardt (L-M), (e) Particle Swarm Optimization (PSO) and (f) CMA-ES.

6.1.6 Parallel versus sequential implementation

To analyze the parallel and sequential performance, the second stopping criterion, maximum 5×10^5 function evaluations, is dropped, and the optimization terminates only after reaching a fitting error 1×10^{-3} or upon reaching stagnation. The CMA-ES is a local search process, yet it can detect the global topology by increasing the population size [Hansen and Kern, 2004]. Figure 27(a) shows that a sequential run with population size $\lambda \leq 50$ could not converge to the stipulated fitting error. Another observation is that the number of function evaluations to reach the stipulated fitting error monotonically increases with increasing the population size from $\lambda = 100$ requiring 3.02×10^5 function evaluations to $\lambda = 700$ requiring 6.88×10^5 function evaluations. Thus, $\lambda = 100$ provides the optimum computational cost for sequential implementation. For the parallel run, increasing the population size is advantageous.

Since rank- μ -update can effectively exploit the information contained in large population sizes, then this can significantly reduce the number of iterations to reach a certain fitting error. For example, Figure 5(b) shows the default population size $\lambda = 17$ can reach a fitting error 1×10^{-2} in 18,200 iterations, yet for a population size $\lambda = 600$ the number of iterations is reduced by two orders of magnitude to 869.

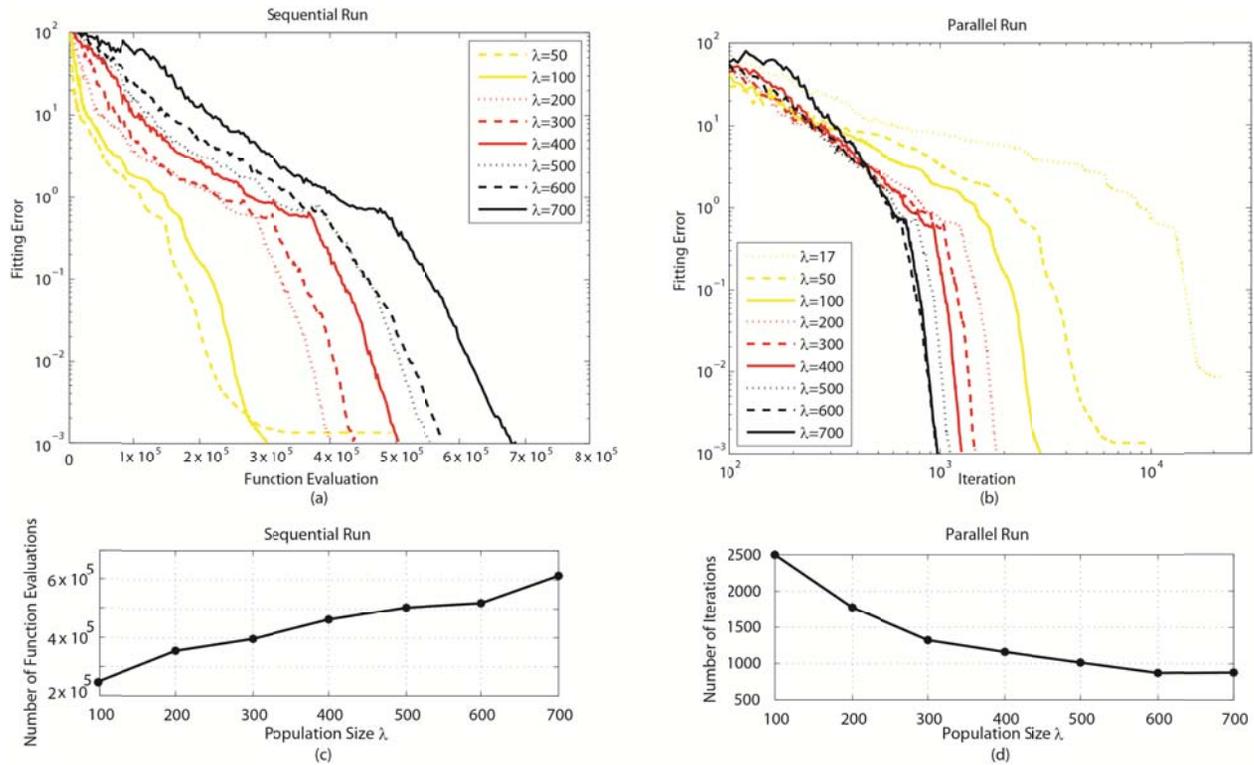


Figure 27 Convergence profiles of different population sizes λ for (a) sequential run and (b) parallel run. (c) The number of function evaluations for different population sizes λ to reach the stopping criterion for sequential run. (d) The number of iterations for different population sizes λ to reach the stopping criterion for parallel run.

The reduction of the number of iterations by increasing the population size has an important implication on the parallel implementation. For sequential runs, Figure 27(c) illustrates the previous remarks that the number of function evaluations monotonically increases with increasing the population size. In contrast, if all the solutions of size λ per iteration are distributed over a number of processors λ , Figure 27(d) shows that the parallel CMA-ES scales

favorably with increasing the number of processors. For example, Figure 27(d) shows that given $\lambda = 100$ processors, the number of iterations required to reach a fitting error 1×10^{-3} is 2499; yet given $\lambda = 600$ processors, the number of iterations is reduced to 964. However, the favorable scaling with increasing the number of processors is up to a certain limit. For example, to reach the stipulated fitting error, $\lambda = 600$ requires 964 iterations, while $\lambda = 700$ requires 984 iterations as shown in Figure 27(d). This result is consistent with Hansen and Kern [2004] results on eight test functions, which show that the scaling could have a convex shape.

The aforementioned analysis shows that optimal population size for sequential runs is different from parallel runs. In this case, $\lambda = 100$ resulting in 3.02×10^5 function evaluations is the optimal population size in the sequential run while $\lambda = 600$ resulting in 964 iterations is the optimal population size for the parallel run. The optimal parallel run is more than 300 times faster than the optimal sequential run. In general, the tuning of the population size λ for CMA-ES is unproblematic for both sequential and parallel runs since it follows a general pattern. For a sequential run it is recommended to start with the default population size, and increase it in case that the desired fitting error is not reached. For a parallel run, it is recommended to start with a relatively large population size and then tune it up or down as needed. The result shows that the optimum population size for the parallel run is about $7.4n$ for the synthetic problem.

6.1.7 Covariance matrix for Monte Carlo sampling

This section shows that the adaptation of the variance, covariance and step size as the solution progresses. This is needed to interpret the meaning of the quantified uncertainty through sampling with the full covariance matrix as empirically estimated by the CMA-ES. The algorithm is allowed to progress to 5000 iterations. Note that the estimation, variance and covariance results are presented according to the unscaled CMA-ES matrix. The solution

progress in Figure 28 shows that the CMA-ES first detects the hydraulic conductivity structure and then overcomes the noise through careful adaptation of the step size. At iteration 5000 the root square error is minimal, yet the estimated hydraulic conductivity field relatively different from the true field (see Figure 23(b)), which is mainly due to ill-posedness that is non-uniqueness in this case.

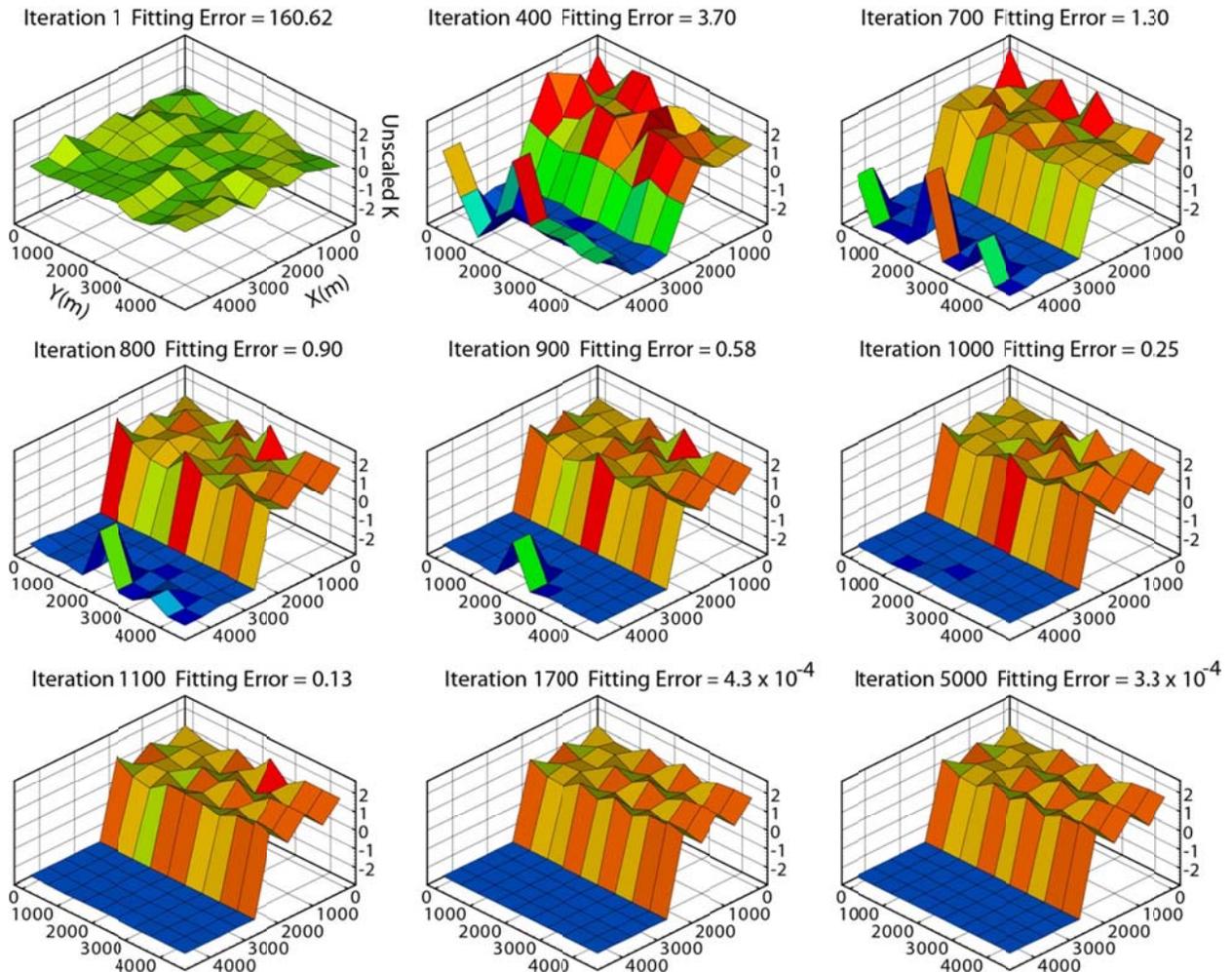


Figure 28 Estimation progress of the unscaled hydraulic conductivity K and fitting error at successive iterations.

Figure 29 shows the progress of the estimated hydraulic conductivity variance. As expected the variance decreases as the solution improves. If the global solution is reached, the diagonal elements of the covariance matrix will be zeros. Iteration 5000 has minimal root square

error, yet the variances are not zero. This shows that the estimated variances reproduce the parameter estimation error. Figure 30 shows the progress of the covariances of hydraulic conductivities with respect to the hydraulic conductivity at the top right corner, which is blanked. The careful adaption of the step size has a clear role during both the global search process and local convergence process. The progress shows that while the hydraulic conductivity estimates from iteration 1100 to iteration 5000 are almost the same (see Figure 28), yet the estimation of covariances is improving due to the adaption of the step size. The covariance values at iteration 5000 are well estimated.

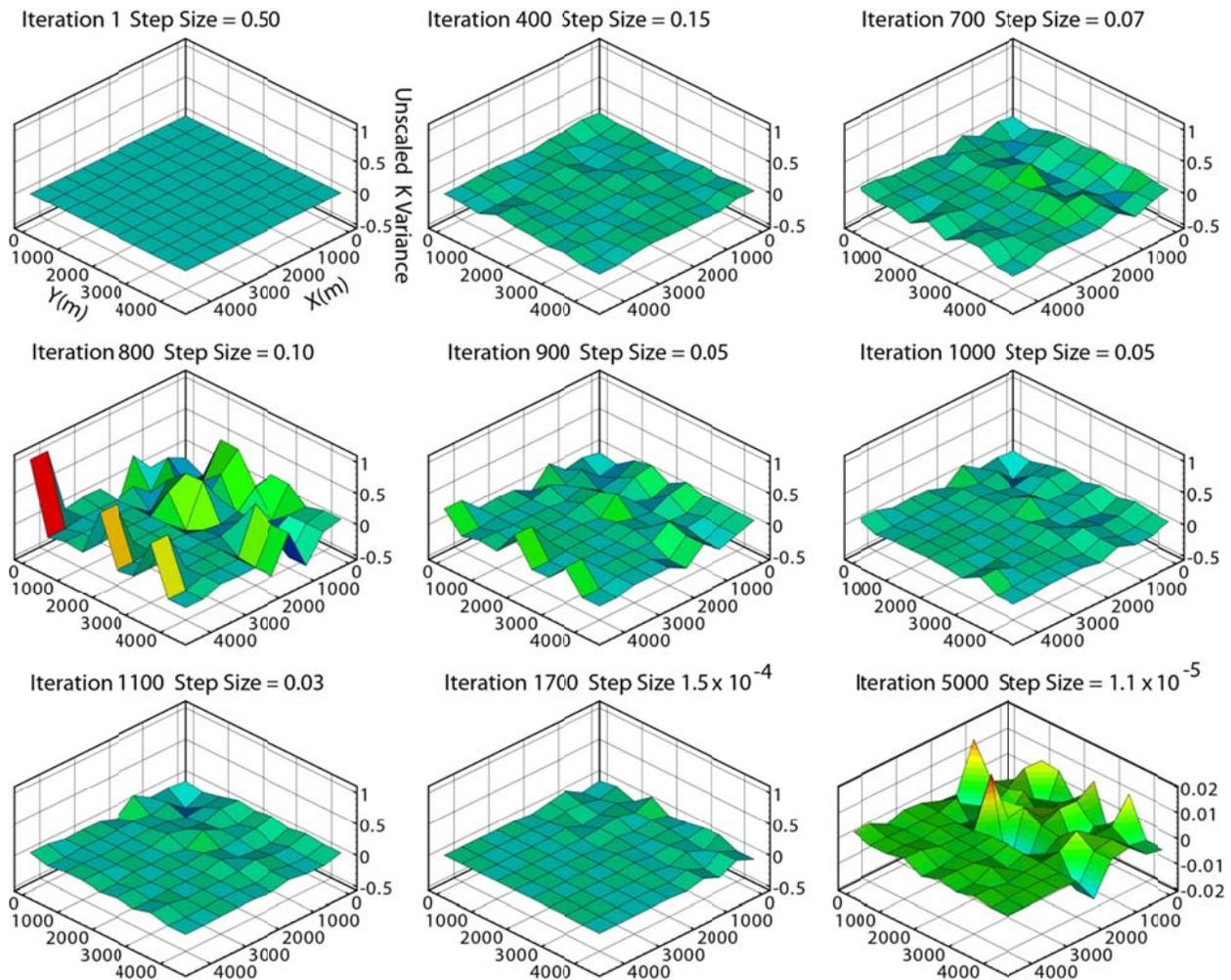


Figure 29 Estimation progress of the variance of the unscaled hydraulic conductivity K and step size at successive iteration.

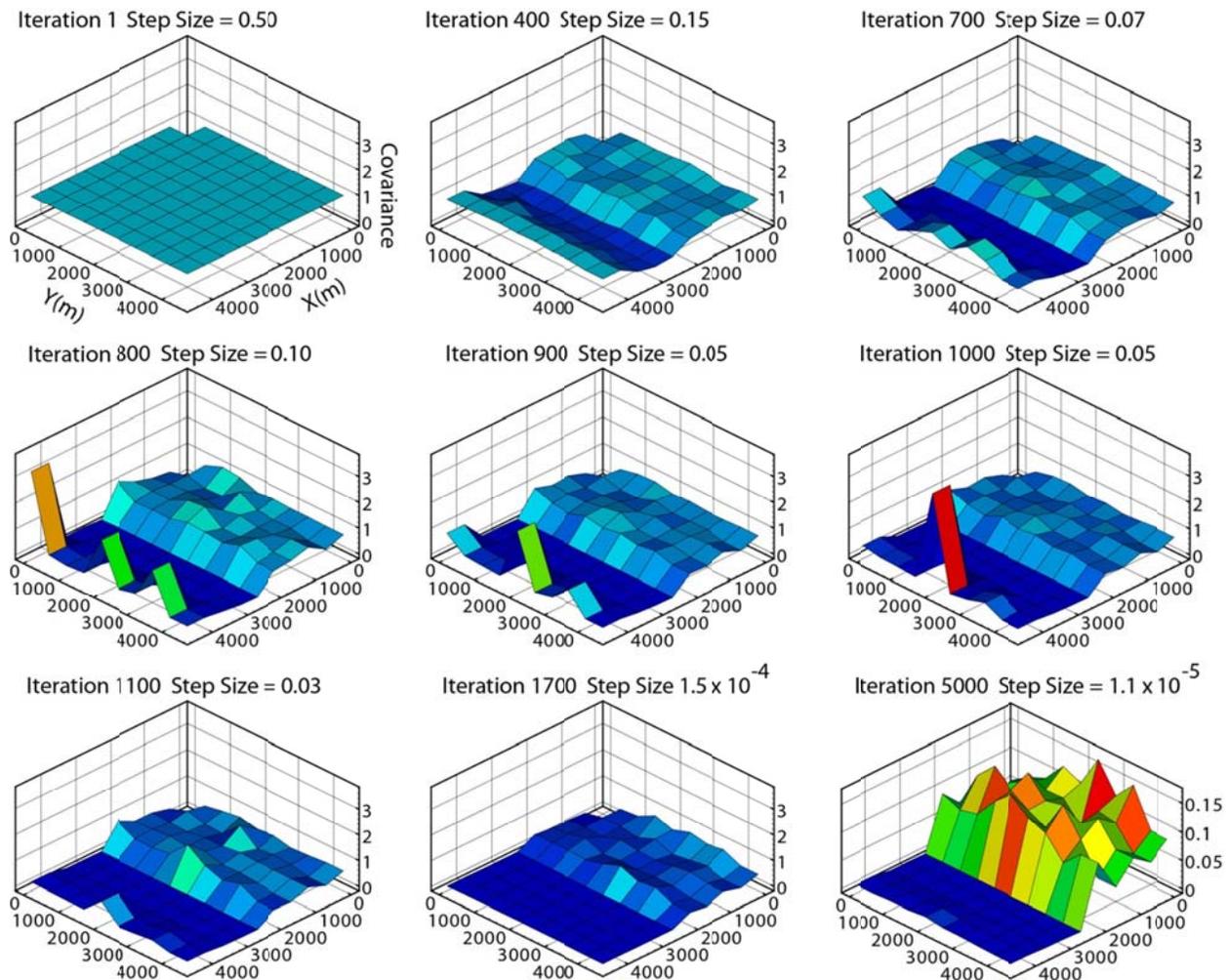


Figure 30 Estimation progress of the covariance of the unscaled hydraulic conductivity K values with respect to the K value at the top-right corner ($X=4250$ m, $Y=4250$ m) at successive iteration.

Given the data and model structure, the empirically calculated covariance matrix can be used for sampling to calculate the head variance due to parameter estimation errors. CMA-ES adapts the covariance matrix not to maximize the entropy of the search distribution, but rather to increase the likelihood of generating successful search directions [Müller, 2010]. This results in precise covariance values that accurately establish the correlation between the unknown model parameters and variance values that quantify the estimation error. Figure 31 addresses the head variance in relation to the covariance matrix of estimated hydraulic conductivity. Figure 31(a) shows the mean head variance [m^2] of the 81 cells of the unconfined aquifer for 10,000

realizations at 100-iteration interval. The mean head variance decreases at the solution progresses until achieving the target distribution after 3000 iterations. For the early iterations the mean head variance is not necessarily monotonically decreasing, which is mainly because different local minima are sampled along the iterations. Figure 31(b) shows the mean head variance with respect to the sample size for different iteration intervals. For the early iterations, convergence is not reached with size as large as 10000 samples. However, after reaching the target distribution as shown by iteration 3000 and iteration 5000 for example in Figure 31(b), about 100 samples are required for variance convergence. This is mainly because the estimated parameters and covariance matrix are precise. Thus, the small head variances, which are due to small parameter estimation error, can be sampled by few realizations.

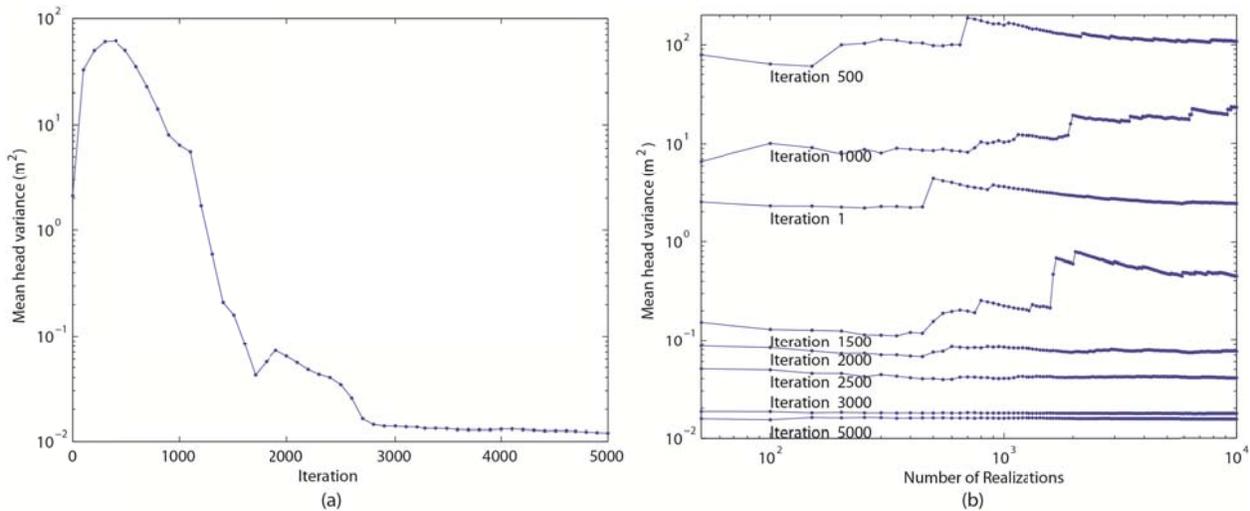


Figure 31 (a) Mean head variance for the unconfined aquifer based on 10,000 realizations at different sampling intervals. (b) Convergence profiles of mean head variance for the unconfined aquifer for different sampling intervals.

6.2 “2,000-foot” sand groundwater flow problem

6.2.1 Model parameters and calibration

The previous hydrogeological characterization in Section 5.2.4 shows that the “1,200-foot”, “1,500-foot”, and “1,700-foot” sands between the two faults are interconnected and should modeled together, while the “2,000-foot” sand between the two faults is a separate aquifer. This

case study focuses on the “2,000-foot” sand. This case study focuses of the “2,000-foot” sand, which compromise of the “2,000-2,400-foot” sand in the north domain north of the Denham Springs-Scotlandville fault, “2,000-foot” sand in the middle domain between the two faults, and the “1,700-foot” sand in the south domain south of the Baton Rouge fault. The complex aquifer-fault architecture for the considered sand in Figure 32 is obtained following the same procedure in Section 5.1 with the only exception that 491 electric well logs are used. The grid generation method in Pham and Tsai [2013] is used to reduce the number of vertical layers. The model discretization consists of 93 rows and 137 columns with a cell size 200 m by 200 m. In the vertical direction, the model grids for “2,000-foot” sand consist of 29 layers with layer thickness ranging from 1 m to 6 m.

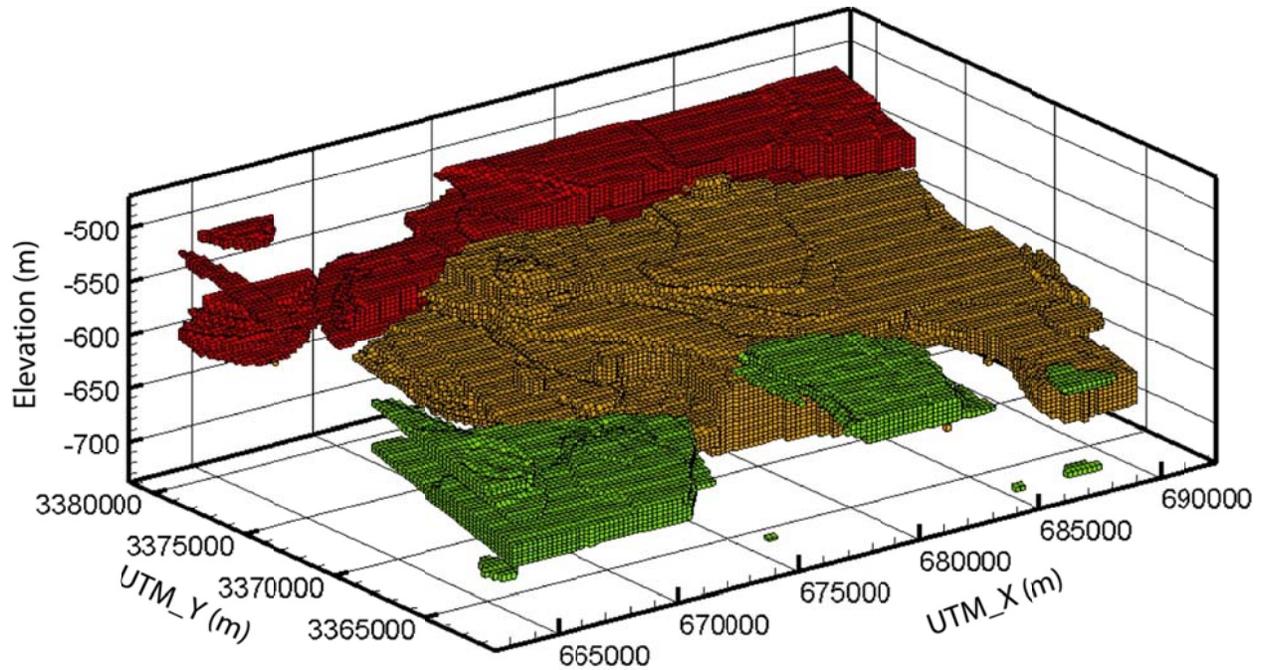


Figure 32 Model grid of the “2,000-foot” sand model.

A time-varied constant head boundary condition is assigned for all active cells through extrapolation of the nearby head observation data. No-flow boundary condition is assigned to clay unit at model boundaries. Detailed pumpage data is available from the Louisiana Capital

Area Ground Water Conservation Commission. The “2,000-foot” sand model has 29 pumping wells extracting about 78,457 m³/day in December 2010. The “2,000-foot” sand model uses 1285 head observations from 17 USGS observation wells for the same period. The locations of the pumping and observation wells are shown in Figure 33.

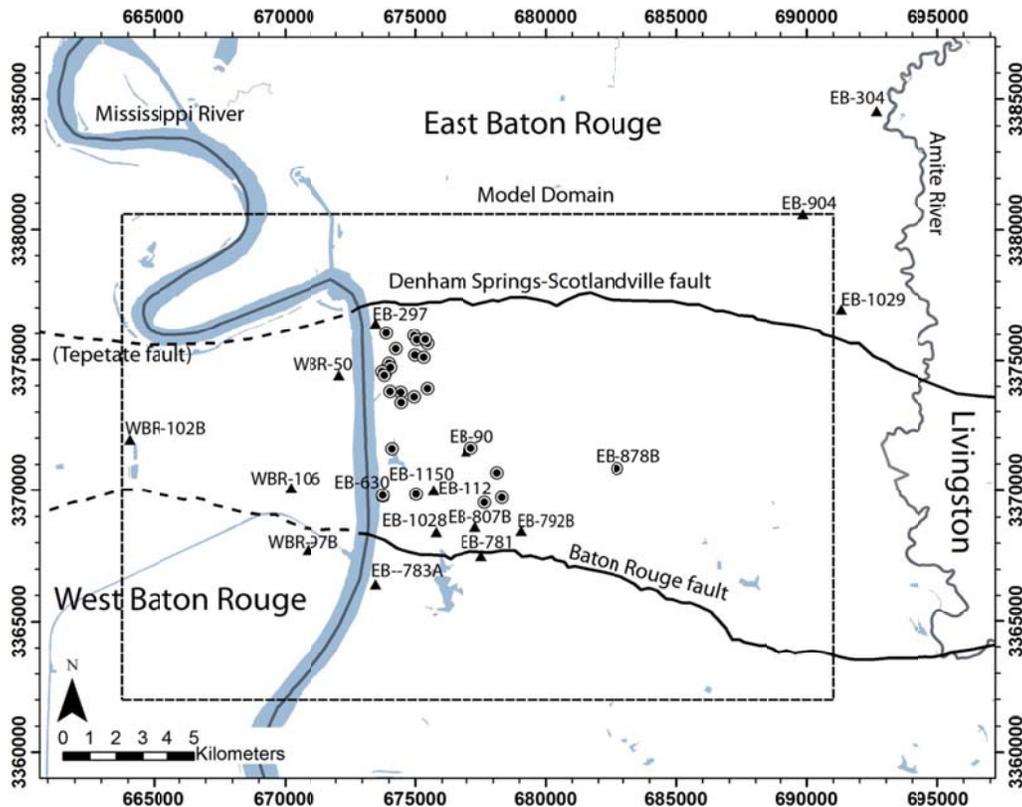


Figure 33 The map of study area showing the locations of the USGS observation wells (triangles) and pumping wells (circles) for the “2,000-foot” sand. The bold solid lines are fault lines identified by the surface expression and the bold dashed lines are the approximate surface locations of the faults [McCulloh and Heinrich, 2012]. The blue lines and areas are surface water bodies.

The inverse problem is to minimize the root mean squared error between the simulated and observed heads. The “2,000-foot” sand model has 6 unknown parameters that are hydraulic conductivity [m/d], vertical anisotropy ratio, specific storage, two hydraulic characteristics for the two faults, and a boundary head adjustment factor for the eastern boundary between the two faults. The calibration parameters ranges are shown in Table 8.

Table 8 The ranges and estimated values of the unknown model parameters for the "2,000-foot" sand model.

Parameter	Parameter Range		CMA-ES
	Minimum	Maximum	Estimated
Hydraulic conductivity [m/d]	70.00	170.00	144.86
Specific storage [1/m]	1.00×10^{-5}	3.00×10^{-5}	1.86×10^{-5}
Vertical anisotropy ratio [-]	1.00	5.00	1.00
Hydraulic characteristic for BR ^a fault [1/d]	1.00×10^{-4}	1.00×10^{-2}	4.20×10^{-3}
Hydraulic characteristic for DSS ^b fault [1/d]	1.00×10^{-6}	1.00×10^{-3}	1.34×10^{-6}
Boundary condition adjustment factor [m]	-5.00	5.00	1.36

^aBaton Rouge ^bDenham Springs-Scotlandville

6.2.2 Parallel calibration using high performance computing

For complex groundwater model that generally takes hours to run, using sequential CMA-ES for solving the groundwater inverse problem is impractical due to the prohibitive computational cost. This study resolves this computational issue by implementing the CMA-ES in a high performance computing (HPC) cluster using an embarrassingly parallel master/slave technique.

Embarrassingly parallel master/slave technique treats the individual solutions as explicit tasks that do not communicate with each other, and assigns each task to a processor. The implantation of the master/slave technique using CMA-ES is straightforward. First, the CMA-ES generates solutions at the master node. These solutions are distributed to the slave nodes to run the simulation models and calculate the fitting errors accordingly. Then the computing nodes pass the fitting errors to the CMA-ES at the master node to generate new solutions for the next iteration, until reaching the stopping criterion. The scheme can be implemented with a simple Bash script without the need of any shared or distributed memory programming languages such as OpenMP or MPI. Thus, embarrassingly parallel problems are the easiest to parallelize, and has the minimal parallelization overhead since the individual tasks do not communicate.

The parallel computation of CMA-ES was carried using SuperMike-II, a HPC cluster at Louisiana State University with 440 compute nodes and a peak performance of 146 TFlops (trillion floating point operations per second). Each compute node is equipped with two 8-Core processors operating at a core frequency of 2.6 GHz.

6.2.3 Speedup of parallel runs

The performance of a parallel algorithm can be evaluated based on parallelization speedup that is the ratio of sequential execution time over parallel execution time, and the scaling that is the speedup due to increasing the number of processors. For this case study, the execution time for a single model simulation is around 1.28 ± 0.1 hours for “2,000-foot” sand model. Given minimal parallelization overhead, the speedup is roughly equal to population size. For example, given $\lambda = 64$ and a stopping criterion $2.95m$ fitting error for the “2,000-foot” sand model, the parallel CMA-ES reached the stopping criterion in 16 iterations in about 21 hours. The sequential CMA-ES execution time would have been roughly $64 \text{ solutions} \times 21 \text{ hours} = 1344 \text{ hours}$. Note that calculated sequential execution time is slightly overestimated since the iteration time for parallel CMA-ES is the maximum of the running times of all the solutions in the iteration. No sequential runs are tested for these case study and the following results and discussion are for parallel runs.

The study demonstrates the speedup of the parallel CMA-ES with increasing the population size λ , which is equal to the number of processors. The optimal population size is determined for the models by performing calibration runs with different population sizes $\lambda = 16, 32, 48$ and 64 . For the best performance, Hansen and Ostermeier [2001] and Hansen et al. [2003] recommended $4 + \lfloor 3 \ln(n) \rfloor \leq \lambda \leq 10n$. Thus, $\lambda = 64$ is selected as the maximum population size for the “2,000-foot” sand model given $n = 6$. Note that the population size $\lambda = 64$ for the “2,000-foot” sand is slight over the recommended range of $10n$. This is because each node in

SuperMike-II has 16 processors. To maximize the efficiency of the parallel implementation, the number of processors is equal to the population size λ .

Figure 34 shows the number of iterations required to meet different stopping criteria of RMSE versus the number of processors. More iterations are needed for smaller RMSE criteria. Smaller RMSE criteria may not be met using smaller population sizes as the search becomes less global. For example, as shown in Figure 34 the population size $\lambda = 16$ did not reach the RMSE 2.948 m criterion. Increasing the population size will always reduce the number of iterations. Thus, the optimal population size for the “2,000-foot” sand model is $\lambda = 64 \approx 10.67n$. The optimal population size is at the upper limit $\lambda = 10n$ given by Hansen et al. [2003]. As shown in Figure 34 given RMSE 2.95m for the “2,000-foot” sand model, speedup 2.29 is achieved by the optimal population size $\lambda = 64$ with respect to the default population size $\lambda = 16$. The optimal population size and the speedup for the “1,200-1,500-1,700-foot” sands model and “2,000-foot” sand model, which is upper limit $\lambda = 10n$, show the invariance properties of the CMA-ES [Elshall et al., submitted].

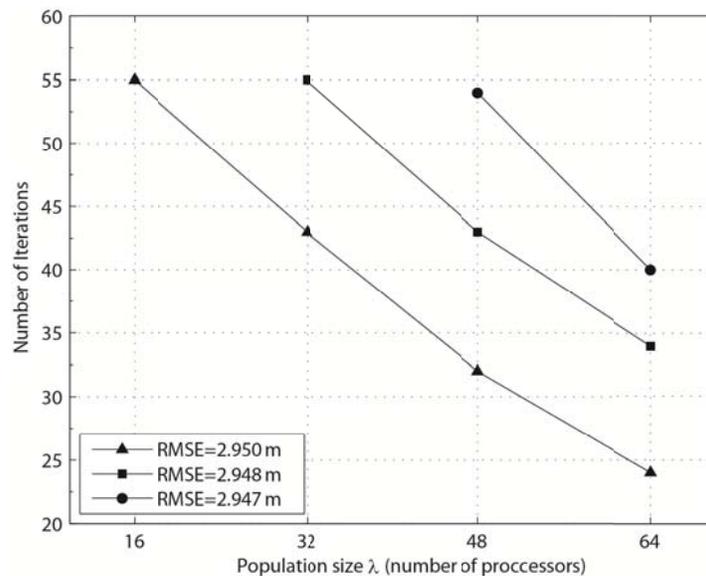


Figure 34 The number of iterations for different population sizes λ required to reach several target fitting errors for the “2,000-foot” sand model.

6.2.4 Covariance matrix for Monte Carlo sampling

Using the optimum population size, estimated model parameter values are converged within 100 iterations for the “2,000-foot” sand model. At the last iteration, covariance matrices of estimated parameters of the two models are obtained. The covariance matrix is used for Monte Carlo simulation to quantify head uncertainty due to parameter estimation error. For the illustration purpose, an observation point at WBR-106 in September 1976 is selected for the “2,000-foot” sand model, which has the highest head standard deviation among all observation data. The head standard deviation σ [m] is calculated based on 200 realizations at a sequence of sampling intervals for the selected observation point. As shown in Figure 35(a) although the improvement of the RMSE is minimal after about 30 iterations for the “2,000-foot” sand model, yet CMA-ES continues to ensure the convergence of the covariance matrix estimation. Results confirm the finding from the synthetic case study that the covariance matrix convergence requires more iterations than the maximum likelihood estimation convergence. Figure 35(b) shows the head prediction and the head prediction standard deviation for the selected observation points at different iteration intervals. For each iteration interval, 200 realizations are generated. The results show that the close we get to the target distribution the less number of realizations are required for the head and standard deviation to converge. Figure 35 (c) shows that the magnitude of the standard deviation generally decreases.

The Monte Carlo simulation results of the groundwater model illustrate the conceptual difference between an adequate model and a precise inverse solution. The results show that the retrieved head variance for the groundwater model, which is quantified by a precise covariance matrix, is very small in comparison to the RMSE. This is because the quantified variance is only due to parameter estimation error, which is a measure of the precision of the inverse solution

regardless of the adequacy of the model. Model structure related variance is generally higher than model parameter related variance. Other sources of variance could also be due to model data and processes. The overall model variance is the summation of these different variance terms.

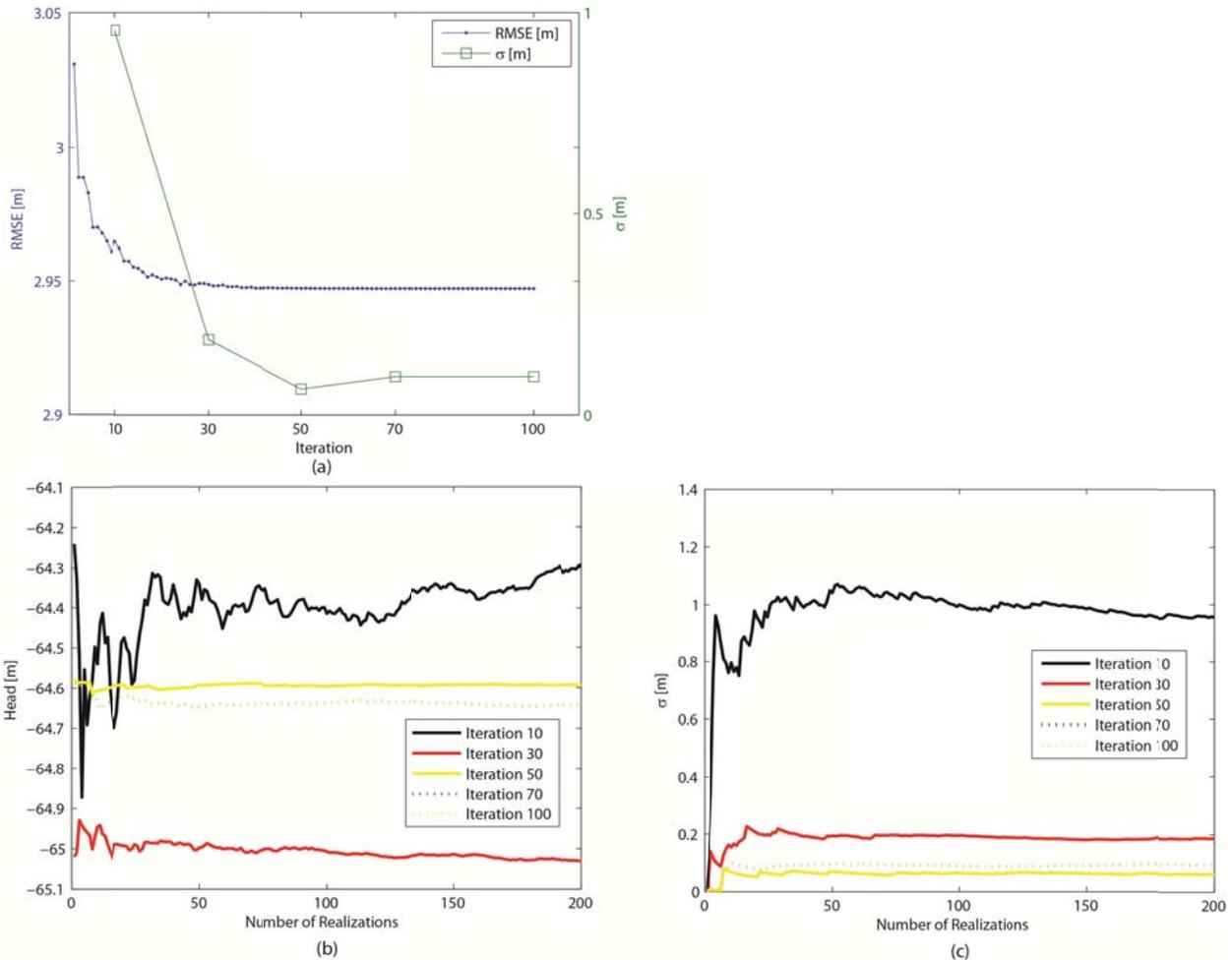


Figure 35 For the “2,000-foot” sand (a) the convergence profiles of the RMSE [m] and head standard deviation σ [m] for a selected observation point based on 200 realizations at successive sampling intervals, (b) Head [m] convergence profiles at successive sampling intervals and (c) head standard deviation [m] convergence profiles at successive sampling Note that for subplots (b) and (c) the convergence profiles for iteration 70 and 100 are similar.

6.3 Conclusions

This study evaluates the robustness of the CMA-ES to solve the groundwater inverse problem. Since such an elaborate search mechanism is not common in the other heuristic algorithms, CMA-ES is shown to be more robust than 4 heuristic algorithms and Levenberg–

Marquardt algorithm in terms of reaching a near global solution for a rugged, nonseparable and noisy function in the synthetic inverse problem.

Using CMA-ES algorithm is convenient in comparison with Levenberg–Marquardt algorithm and particle swarm optimization. The Levenberg–Marquardt algorithm requires multi-start with several initial solutions, and its parameters generally need manual tuning. The PSO, which has the second best performance after CMA-ES, requires significant manual tuning as well. The use of CMA-ES is convenient since CMA-ES adopts a systematic learning process with less stochastic terms, and thus solutions of repeated runs are relatively similar. In addition, since the parameters of the CMA-ES are adaptable, the only parameter that needs to be tuned is the population size. Tuning of CMA-ES is simple since search becomes more global by increasing the population size as shown by the synthetic problem and the two Baton Rouge groundwater models.

The high computational cost is a common drawback for heuristic algorithms. Except for CMA-ES, the performance of the tested heuristic algorithms is similar to the performance of Levenberg–Marquardt algorithm. Thus, the additional immense computational cost of heuristic algorithms in comparison to Levenberg–Marquardt algorithm seems unworthy. However, CMA-ES substantially outperforms the Levenberg–Marquardt algorithm, yet at considerably higher computational cost. Nevertheless, the computational efficiency of Levenberg–Marquardt comes at the cost of obtaining a local solution.

The study shows that the high computational cost of CMA-ES algorithm can be alleviated by parallelization. CMA-ES is readily amenable for embarrassingly parallel computation without any modification that would scarify its performance. In addition, the study shows that increasing the population size reduce the number of iterations that is required to meet

the stopping criteria. The parallel CMA-ES has been successfully implemented to a high performance computing cluster. Results show that the speedup can be more than doubled at the optimum population size in comparison to the default population size.

The study tested the use of the empirically calculated covariance matrix for Monte Carlo simulation to quantify the parameter related uncertainty. Results show that after reaching the target distribution, only a small number of realizations are required for the convergence of the head prediction and head prediction variance. That is mainly due to the use of a full covariance matrix that has both the variances and covariances of the estimated model parameters.

7 Constructive epistemic modeling of groundwater flow under Bayesian paradigm

7.1 Case Study: Groundwater flow model of the “2,000-foot” sand

7.1.1 Geological structure uncertainty

This case study uses the same groundwater flow model that is previously discussed in Section 6.2. However, in this case study multiple groundwater flow models are used in this case study to represent numerous candidate model propositions that are proposed. These candidate propositions are the sources of uncertainty.

The first uncertain model component is the hydrofacies architecture reconstruction method. 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 would generally result in the best flow model. In this case study, the three indicator geostatistics methods for reconstructing the hydrofacies architecture, which are indicator 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], are considered. For using these three hydrofacies architecture reconstruction propositions, the reader is referred to Section 5.1.

The second uncertain model component with respect to geological structure is the formation dip. This case study adopts two formation dip propositions of 0.29° and 0.35° , which were estimated using the inverse modeling approach and a clustering approach, respectively, as shown in Section 5.2.3. Using 491 electric well logs for hydrofacies architecture reconstruction, Figure 36 shows the flow pathways through the “2,000-foot” sand in the middle domain through of the Baton Rouge fault and the Denham-Springs Scotlandville fault, given the IZ, GP, and IK

propositions and the formation dip proposition $D1=0.29^0$. The IZ, GP and IK propositions provide sharp edged conditional estimates, non-smooth conditional estimates and smooth conditional estimates, respectively.

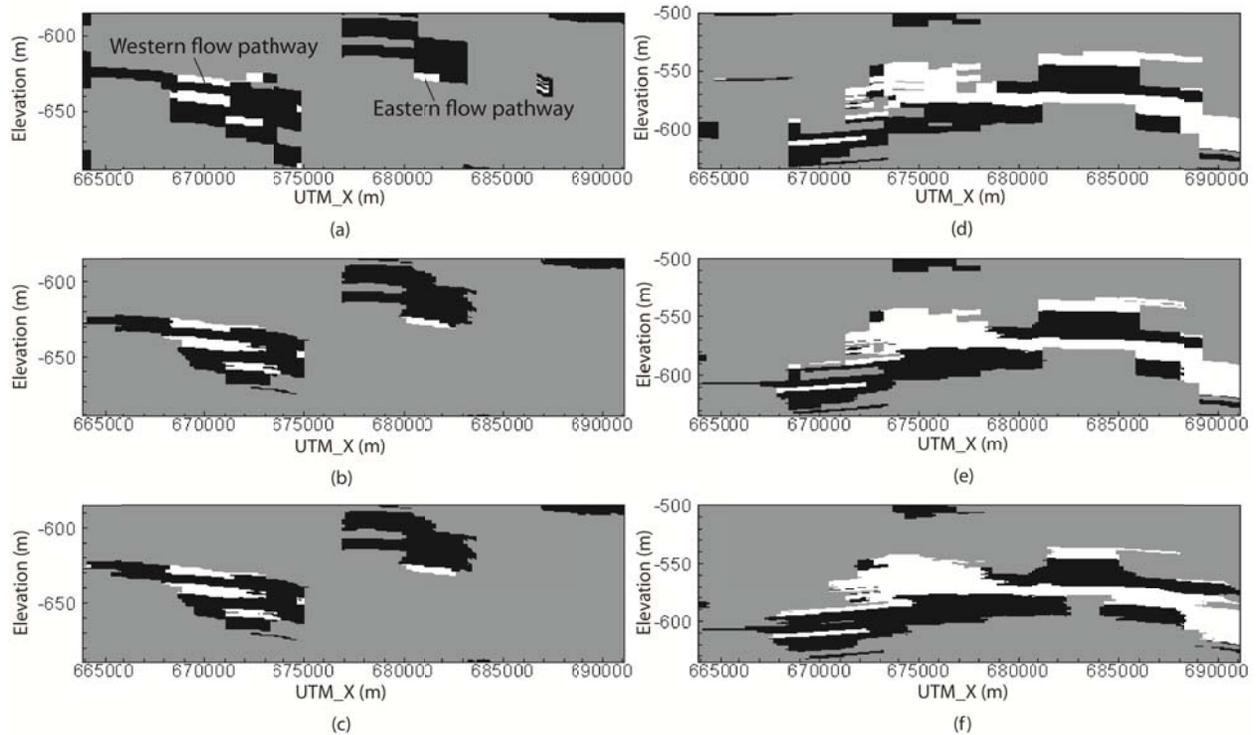


Figure 36 For formation dip proposition $D1=0.29^0$: the architecture of the Baton Rouge fault for the “2,000-foot” sand using (a) indicator zonation, (b) generalized parameterization and (c) indicator kriging, and the architecture of the Denham Springs-Scotlandville fault using (d) indicator zonation, (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 flow pathways to the “2,000-foot” sand through the faults.

Given the second formation dip proposition $D2=0.35^0$ and the IZ, GP and IK propositions, Figure 37 shows the architectures of the Baton Rouge fault and the Denham Springs-Scotlandville fault. Comparing Denham Springs-Scotlandville fault architectures in Figure 36 and Figure 37 shows that using different formation dips produces relatively different hydrofacies architectures. The Baton Rouge fault architectures in Figure 36 and Figure 37 show an important observation that the flow pathway in the east becomes narrower as a result of increasing the formation dip.

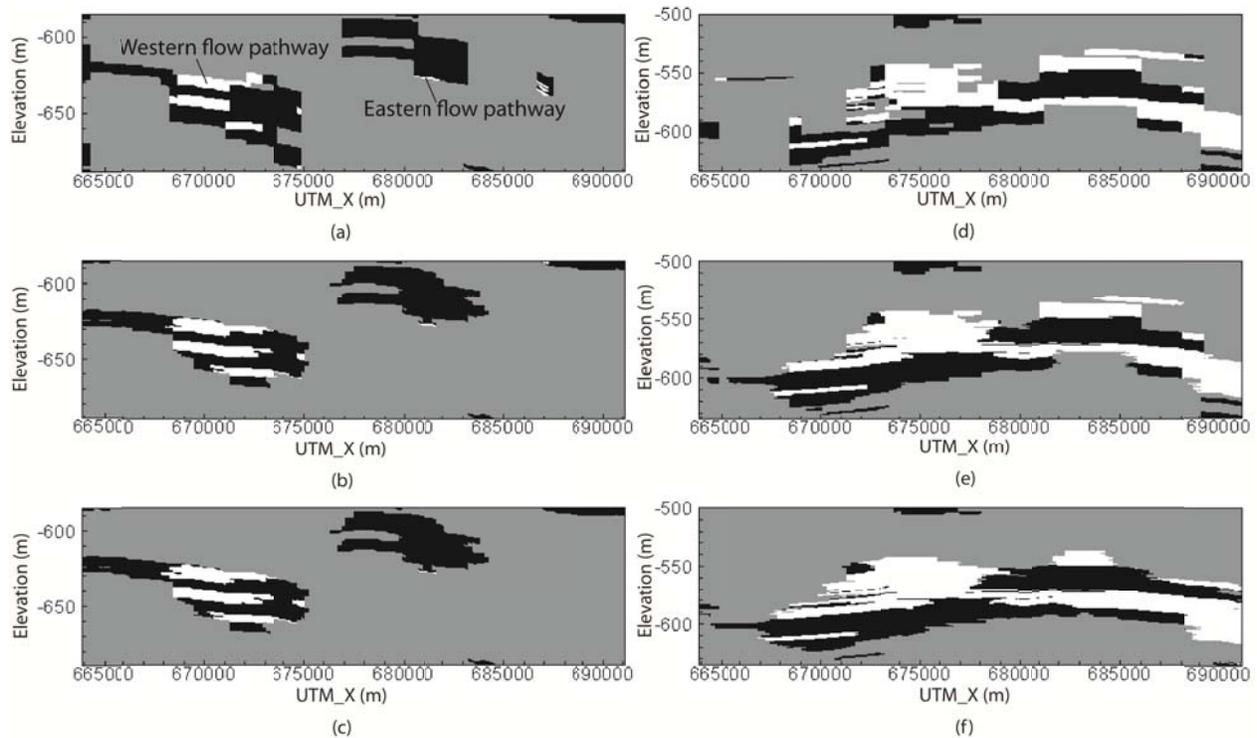


Figure 37 For formation dip proposition $D2=0.35^0$: the architecture of the Baton Rouge fault for the “2,000-foot” sand using (a) indicator zonation, (b) generalized parameterization and (c) indicator kriging, and the architecture of the Denham Springs Scotlandville fault using (d) indicator zonation, (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 flow pathways to the “2,000-foot” sand through the faults.

7.1.2 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 in Section 4.1, the model probabilities are calculated for these six hydrofacies architecture models as shown in Table 9. The results show that the hydrofacies models with D2 proposition have higher model probabilities than those with D1 proposition. In addition, models with GP proposition 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 9 Model probabilities of the six hydrofacies architecture models.

Hydrofacies Model	Q	Δ BIC	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, the detailed vertical discretization of the hydrofacies architecture are vertically aggregated into 29 layers with variable thickness from 1~6 m using the method developed by Pham and Tsai [2013].

Figure 38 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 Figure 38(a)-4(d). Yet GP and IK methods produce relatively similar architectures as shown in Figure 38(c)-4(f), which is mainly because of the large electric well log data set. Figure 38(a)-4(f) show that D1 and D2 propositions produce relatively different architectures, particularly in the north domain. For model averaging as shown in Figure 38(g)-4(h), the grey areas with indicator values between 0 and 1 represent uncertain regions for clay hydrofacies and sand hydrofacies. The result of simple model averaging in Figure 38(g) shows large uncertainty about the clay and sand hydrofacies distribution. However, the Bayesian model averaging results in less uncertainty because the IK and GP propositions are similar and have much higher hydrofacies model probabilities than the IZ proposition, and the D2 proposition has relatively higher hydrofacies model probabilities than D1 proposition.

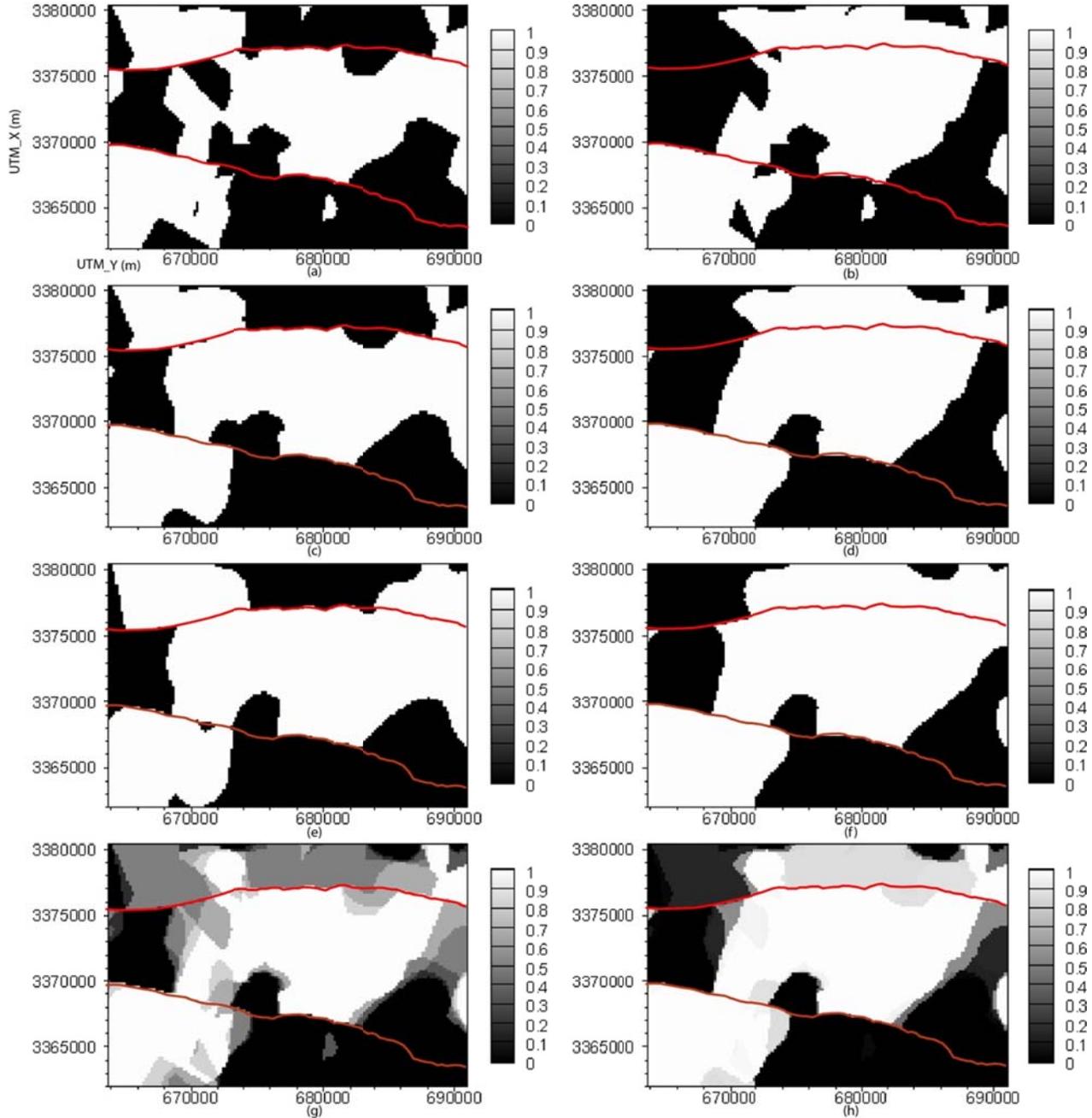


Figure 38 Hydrofacies architectures for a selected horizontal layer at the “2,000-foot” sand that has a top elevation of -556 m NGVD29 at northeast corner and top elevation of -667 m NGVD29 at the southwest corner for (a) IZD1 model, (b) IZD2 model, (c) GPD1 model, (d) GPD2 model, (e) IKD1 model, (f) 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.

7.1.3 Boundary condition uncertainty

Given multiple geological structure propositions, the study assigns no-flow boundary conditions to the clay hydrofacies and a time-varied constant head boundary condition to the sand hydrofacies. Yet different definitions of the boundary conditions can result in different groundwater flow models [Rojas et al., 2008b, 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 Figure 33 for location). Two candidate propositions to determine boundary values for the north domain boundaries are considered. The first proposition (N1) uses linear interpolation of the four available data points as shown in Figure 39. The second proposition (N2) adjusts the head variation trend of EB-304 (see Figure 33 for location) to the head elevations of the four data points as shown in Figure 39.

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 cluster to select to extrapolate to the boundary. This is the case with the eastern boundary condition in the middle domain, in which two clusters of observation wells are categorized 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 N1 and N2 propositions are fixed, the E1 and E2 propositions have an elevation adjustment factor to be determined by inverse modeling.

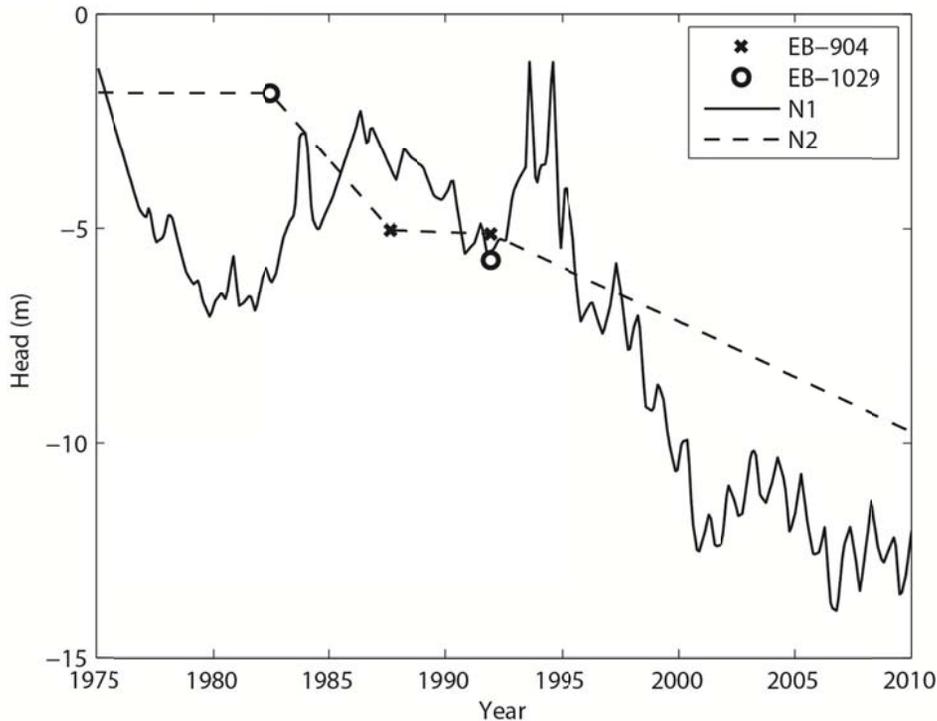


Figure 39 Two boundary head propositions N1 and N2 for the northern boundaries

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.

7.1.4 Model parameters 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 24 base models are obtained for the hierarchical BMA analysis. MODFLOW-2005 [Harbaugh , 2005] is used to simulate groundwater heads. Each model has 29 layers. Each layer has 12,741 cells and the cell size is 200 m×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. Flow model calibration is based on 1285 head data between 1975 and 2010 from 17 USGS observation wells (see Figure 33 for locations). The inverse problem is to minimize the root mean squared error (RMSE) between the simulated and observed heads. CMA-ES [Hansen et al., 2003] algorithm is used for solving the inverse problem.

7.1.5 Quantification of within-model variance

In order to calculate the variance term in equation(41), the head prediction variance for each model needs to be calculated. For each model the maximum likelihood estimates and their covariance matrix are used to generate 512 samples. A sample is 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.

7.1.6 High performance computing for model calibration and variance quantification

The model calibration and the Monte Carlo realizations of the 24 models were carried out using SuperMike-II at Louisiana State University. 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 hours. 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 minimal, thus 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 hours. 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.

7.2 Results and discussion

7.2.1 Model calibration and within-model variance quantification

Table 10 shows the calibration results for the 24 models. The base models are named according 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, indicating that the prior boundary head elevation of the E1 and E2 is well estimated. The ranges 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 are 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.

Table 10 Calibration results: boundary condition adjustment factor (BC [m]), hydraulic conductivity (K [m/d]), anisotropic ratio (K_h/K_v [-]), 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, Δ BIC, prior model probability (priorPr) and posterior model probability (postPr) for the base models.

Base Model	BC [m]	K [m/d]	K_h/K_v [-]	Ss [1/m]	DSS [1/d]	BR [1/d]	RMSE [m]	Q	Δ BIC	priorPr	postPr
IZD1N1E1	2.07	164	1.31	2.00×10^{-5}	8.41×10^{-6}	9.35×10^{-3}	3.24	1808	265	0.007	9.82×10^{-5}
IZD1N1E2	0.38	170	1.55	2.84×10^{-5}	9.77×10^{-5}	1.02×10^{-2}	4.01	2526	983	0.007	2.41×10^{-9}
IZD1N2E1	2.06	165	1.02	2.04×10^{-5}	5.64×10^{-6}	9.09×10^{-3}	3.24	1805	262	0.007	1.02×10^{-4}
IZD1N2E2	0.40	170	3.82	2.59×10^{-5}	1.07×10^{-4}	1.04×10^{-2}	4.01	2527	984	0.007	2.39×10^{-9}
IZD2N1E1	2.76	161	1.02	1.98×10^{-5}	2.32×10^{-6}	7.60×10^{-3}	3.25	1820	277	0.100	3.32×10^{-4}
IZD2N1E2	0.10	170	3.17	2.49×10^{-5}	9.99×10^{-5}	9.98×10^{-3}	4.06	2585	1042	0.100	4.04×10^{-9}
IZD2N2E1	2.73	161	1.00	2.04×10^{-5}	1.45×10^{-6}	7.71×10^{-3}	3.25	1821	278	0.100	3.27×10^{-4}
IZD2N2E2	0.11	170	1.01	2.56×10^{-5}	9.97×10^{-5}	9.98×10^{-3}	4.05	2579	1036	0.100	4.44×10^{-9}
GPD1N1E1	0.67	154	3.37	1.83×10^{-5}	5.78×10^{-6}	7.41×10^{-3}	3.17	1757	214	0.029	2.97×10^{-3}
GPD1N1E2	-1.89	161	3.06	2.29×10^{-5}	9.61×10^{-5}	1.03×10^{-2}	3.83	2324	781	0.029	6.83×10^{-7}
GPD1N2E1	0.70	155	2.75	1.91×10^{-5}	2.20×10^{-6}	7.26×10^{-3}	3.17	1752	209	0.029	3.20×10^{-3}
GPD1N2E2	-2.05	162	1.47	1.94×10^{-5}	8.25×10^{-5}	9.91×10^{-3}	3.83	2333	790	0.029	5.92×10^{-7}
GPD2N1E1	2.32	145	1.01	1.93×10^{-5}	4.50×10^{-5}	4.58×10^{-3}	2.98	1571	28	0.424	1.99×10^{-1}
GPD2N1E2	-0.48	163	1.07	2.35×10^{-5}	9.76×10^{-5}	5.52×10^{-3}	3.79	2268	726	0.424	6.58×10^{-6}
GPD2N2E1	2.48	147	1.23	1.82×10^{-5}	2.21×10^{-6}	4.46×10^{-3}	2.98	1566	23	0.424	2.12×10^{-1}
GPD2N2E2	-0.52	163	1.80	2.65×10^{-5}	9.73×10^{-5}	5.41×10^{-3}	3.79	2260	717	0.424	7.44×10^{-6}
IKD1N1E1	0.11	149	1.29	1.83×10^{-5}	7.23×10^{-6}	6.81×10^{-3}	3.13	1713	170	0.028	1.67×10^{-3}
IKD1N1E2	-2.59	156	1.14	2.21×10^{-5}	2.11×10^{-5}	9.98×10^{-3}	3.72	2208	665	0.028	1.10×10^{-6}
IKD1N2E1	0.05	150	1.08	1.94×10^{-5}	1.02×10^{-6}	6.86×10^{-3}	3.12	1712	169	0.028	1.69×10^{-3}
IKD1N2E2	-2.61	157	1.05	2.06×10^{-5}	1.43×10^{-5}	9.99×10^{-3}	3.72	2207	664	0.028	1.11×10^{-6}
IKD2N1E1	1.29	145	1.01	1.82×10^{-5}	1.30×10^{-6}	4.16×10^{-3}	2.95	1544	1	0.412	2.87×10^{-1}
IKD2N1E2	-1.04	157	1.00	2.31×10^{-5}	9.92×10^{-5}	4.98×10^{-3}	3.71	2196	653	0.412	1.87×10^{-5}
IKD2N2E1	1.32	145	1.07	1.90×10^{-5}	1.11×10^{-6}	4.17×10^{-3}	2.95	1543	0	0.412	2.91×10^{-1}
IKD2N2E2	-1.04	157	1.02	2.20×10^{-5}	9.95×10^{-5}	4.98×10^{-3}	3.71	2191	648	0.412	2.00×10^{-5}

Although the “2,000-2,400-foot” sand at the north domain and the “2,000-foot” sand in middle domain have wide areas of sand-sand contact as previously shown in Figure 36 and Figure 37, it has very low permeability. This suggests that the Denham Springs-Scotlandville fault at the “2,000-foot” sand is a three-dimensional zone of deformation [Bense and Person, 2006; Hanor et al., 2011] and not a two-dimensional plane.

Figure 40 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 IZ, GP and IK propositions and the difference between E1 and E2 propositions. First, observation well EB-90 is selected 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=3.20m for the base model D2IKB1E2. From Figure 40(a), it is clear that the E1 and E2 propositions produce relatively different head predictions. Second, the observation well EB-878B is selected 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 Figure 40(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 9) and similar hydrofacies architectures (see Figure 36 to Figure 38). Comparing Figure 40(a) and Figure 40(b) shows that the absence of conditioning on head observations results in higher between-model prediction variance for observation well EB-878B.

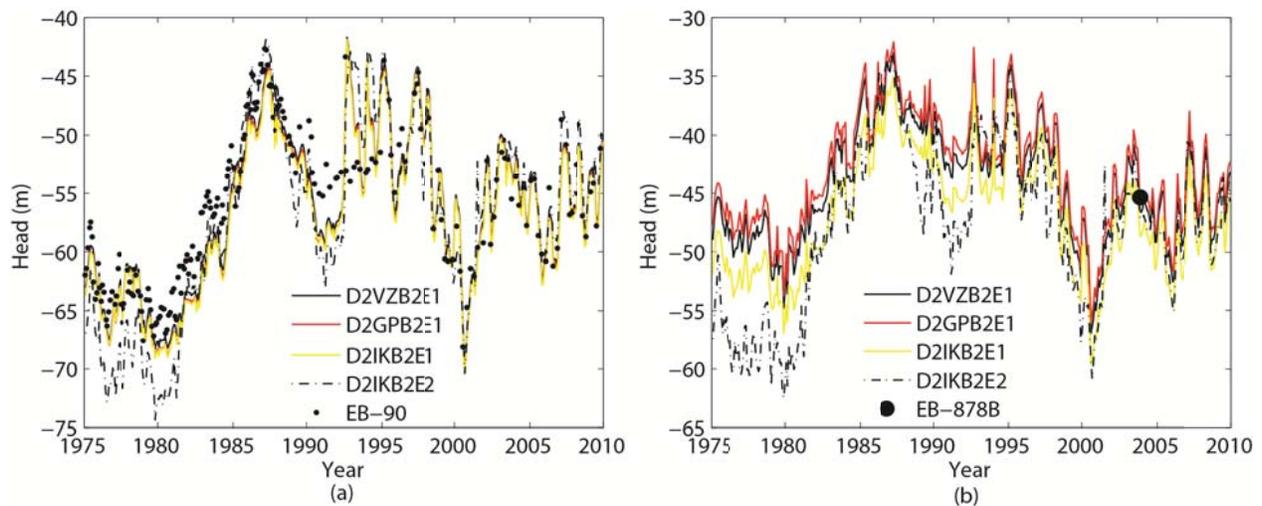


Figure 40 Head predictions and head observations of four selected base models at observation wells (a) EB-90 and (b) EB-878B.

7.2.2 BIC calculation

The calibration and the Monte Carlo results are used to calculate ΔBIC . The larger the variance windows, the more models are selected. To allow for more model selection, $s_1 = 5\%$ significance level and variance width $s_2 = 4\sigma_D$ are used, and thus the scaling factor in equation(44) is $\alpha = 1.06 / \sqrt{n}$ [Tsai and Li, 2008a].

The BIC results are given in Table 10. Substituting Q as calculated from equation (40) into equation(39) the BIC for all base models are obtained. The best base model IKD2N2E1 has $BIC_{\min} = 3948$ from which the ΔBIC for all base models are obtained. Finally, the posterior model probabilities for base models are calculated as shown in Table 10 by substituting ΔBIC and the prior model probabilities into equation(45). The prior model probabilities are obtained from the hydrofacies architecture models (see Table 9).

7.2.3 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 [Foglia et al., 2013], yet the BMA tree of posterior model probability and conditional posterior model probability, as shown in Figure 41 provides more detailed information.

Starting from the conditional posterior model probabilities at the base level of the BMA tree, it is clear that the E1 proposition is consistently robust than the E2 proposition. Moving to level 3, the similar conditional posterior model probabilities of the N1 and N2 propositions indicate that they are both robust. At level 2, the geological formation dip propositions D1 and D2 have different performance under different hydrofacies architecture reconstruction propositions. Under the IK and GP propositions, D2 proposition is considerably more robust than

D1 propositions, yet under the IZ proposition the D1 proposition has some considerable weight. This indicates that both D1 and D2 are relevant propositions. Finally at level 1, the IK proposition is generally more robust than the GP proposition.

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 9). However, the results show that the best hydrofacies architecture model does not necessarily lead to the best groundwater flow model.

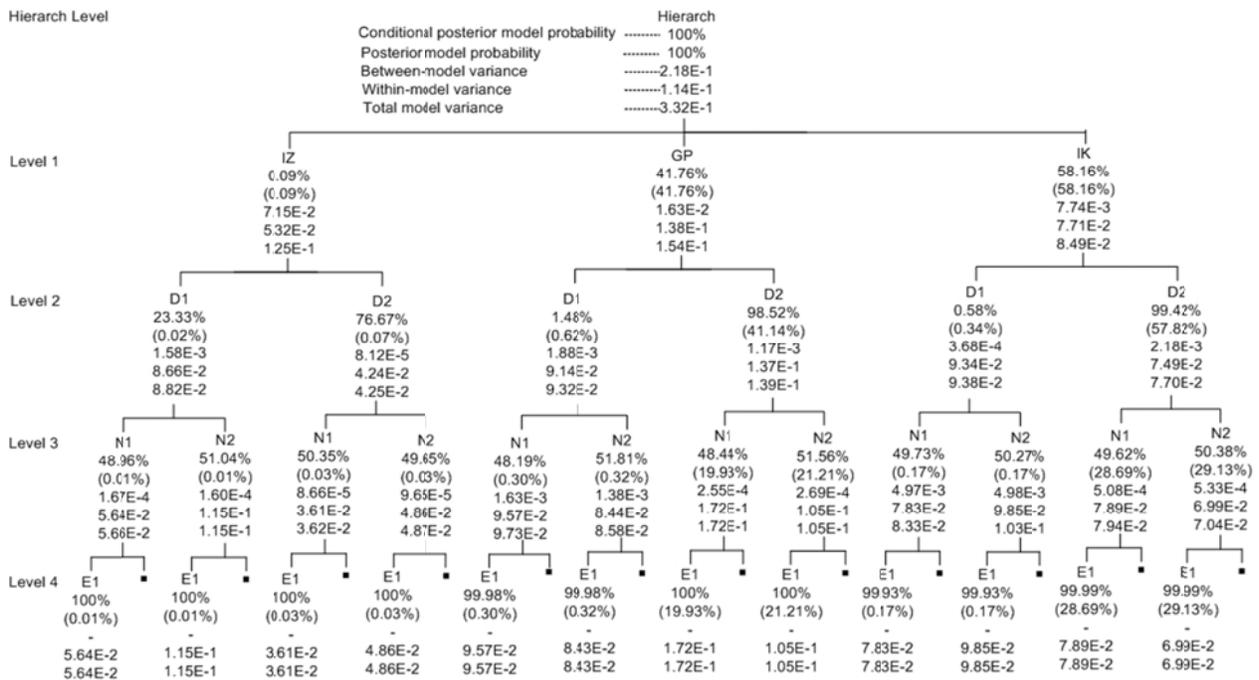


Figure 41 BMA tree of the posterior model probabilities and prediction variances of the four uncertain model components: hydrofacies architecture reconstruction method (IZ, GP and IK), formation dip (D1 and D2), northern boundary conditions (N1 and N2) and eastern boundary condition in the middle domain (E1 and E2). Models that have posterior model probabilities less than 0.01% are not shown.

7.2.4 Uncertainty propagation and prioritization

The second feature is that the hierarchical BMA facilitates the depiction of the uncertainty propagation. For the illustration purpose, averaged prediction variances at all the head observation locations for all the time steps are calculated as shown in Figure 41. Hierarchical BMA separates the uncertainty contribution of each source of uncertainty through providing the between-model variance for each source of uncertainty. 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 Figure 41.

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 Figure 41, 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, and thus the within-model variance is the total model variance.

Tracing the uncertainty propagation starts with the BMA models at level 3 of the BMA tree in Figure 41. Although E1 and E2 propositions produce very different estimation as previously shown in Figure 40, yet since posterior model probability of the E1 proposition is substantially higher than the E2 proposition, thus the between-model variance is minimal. At level 2, even though the N1 and N2 propositions have similar conditional posterior model probabilities, still the between-model variance is minimal due to the similar prediction of N1 and N2 propositions. Level 1 shows that the D1 and D2 propositions introduce small within-model variance for the IZ and GP branches. The hierarch level shows that the hydrofacies architectures 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 Figure 41 shows that generally the total variance increases by adding more uncertain components, yet this is not necessarily the case. For example, IZD1N2 model has higher posterior model probability and higher within-model variance than IZD1N1 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 Figure 41, 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 11 that lists the variance contributions of individual sources of uncertainty to the total variance using equation (34). The hydrofacies reconstruction method contributes the most variance, followed by model parameters.

Table 11 Between-model variance (BMA), 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	-	1.00×10^{-1}	1.00×10^{-1}
3	Eastern boundary condition	4.35×10^{-4}	1.00×10^{-1}	1.01×10^{-1}
2	Northern boundary condition	1.75×10^{-3}	1.01×10^{-1}	1.03×10^{-1}
1	Formation dip	1.14×10^{-2}	1.03×10^{-1}	1.14×10^{-1}
Hierarch	Hydrofacies reconstruction method	2.18×10^{-1}	1.14×10^{-1}	3.32×10^{-1}

7.2.5 Temporal and spatial distribution of head prediction and variance

The study further illustrates 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 the Hierarchy, 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 Figure 42. Figure 42(a) shows very similar head predictions at different levels for EB-90 that has many observations. However, due to only one observation data at EB-878B, the head prediction at this well changes at the hierarch level as shown in Figure 42(b). Since conditioning on head observations reduces the between-model variance, thus 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.

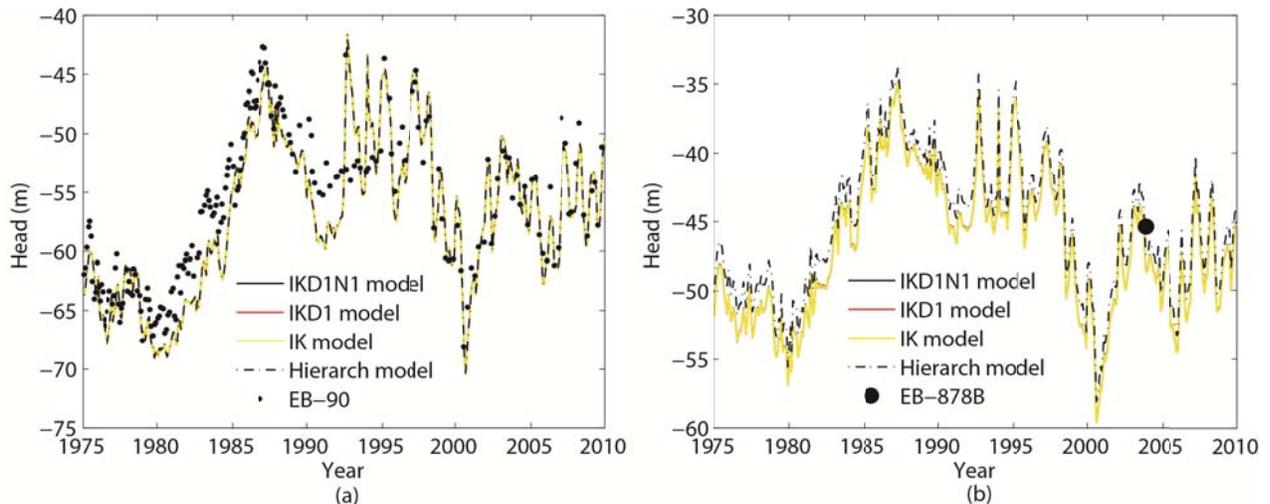


Figure 42 BMA head predictions for the best branch of the BMA tree for observation wells (a) EB-90 and (b) EB-878B.

The study shows spatial distribution of head prediction and variance for the selected layer in Figure 38 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 Figure 38(h).

The head predictions in the middle domain in Figure 43(a)-(c) are similar, and they are different from Figure 43(d). This confirms the previous remark that significant head prediction

changes are only due to the hydrofacies reconstruction method. The reason for this is that given a relatively similar within-model variance for all the base model, the between-model prediction variance is a factor of two things. Different posterior model probabilities with very different head prediction will result in small between-model prediction variance. Similar posterior model probabilities with similar head prediction will result also in small between-model prediction variance. Alternatively, similar posterior model probabilities with different head prediction will result in large between-model prediction variance.

The between-model prediction variance as depicted in Figure 44(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 Figure 44 (a) and Figure 44 (c), respectively, are minimal. The variance contribution from the northern boundary condition is large in the north domain and minimal in the middle and south 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 Figure 44 (d).

The within-model variance and total model variance as shown in Figure 44 (e)-(h) and 10(i)-(l), respectively, show the construction of uncertainty. Figure 44 (e) and (i) at level 3 are similar, and Figure 44 (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 Figure 44 (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, Figure 44 (i)-(l) introduce more uncertain regions. However, variance magnitude can decrease as shown in the north domain in Figure 44 (k) and Figure 44(l).

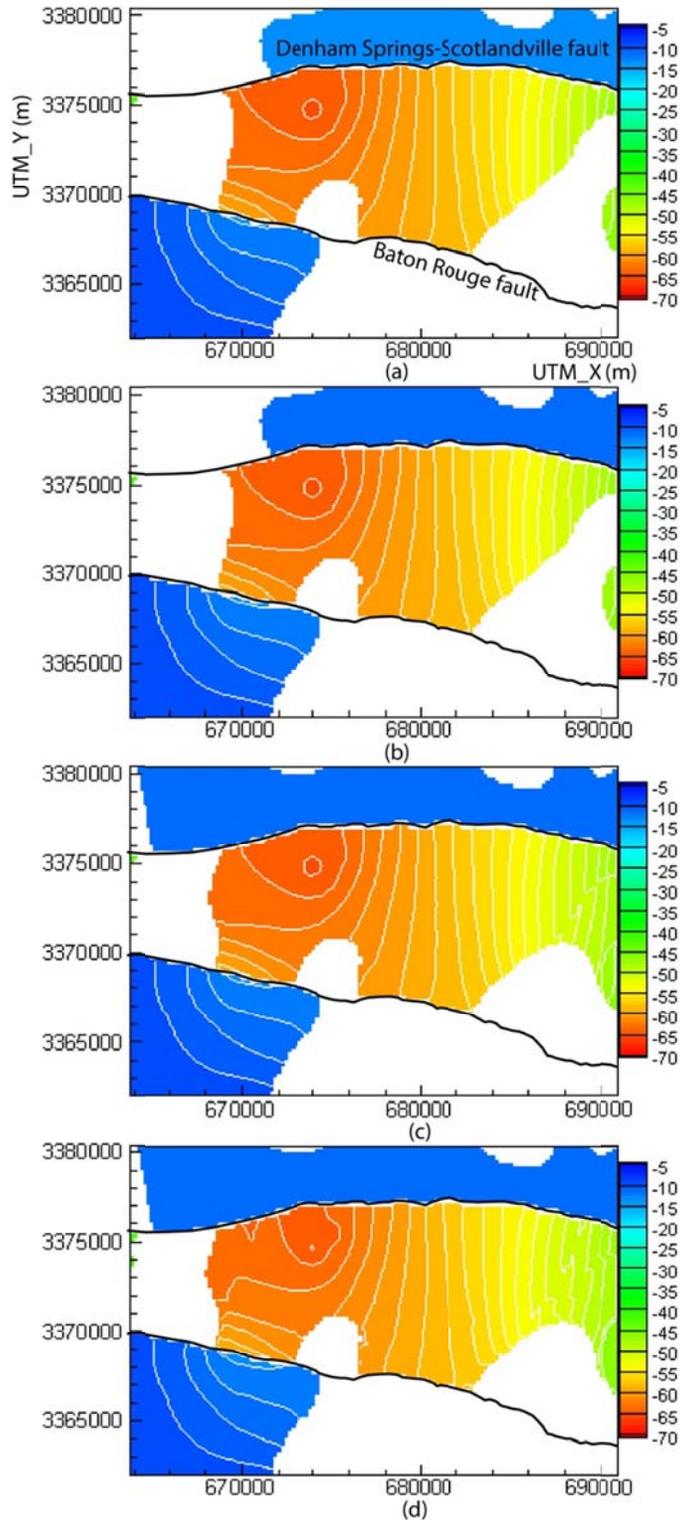


Figure 43 BMA head predictions (meters) for the selected layer in Figure 38 selected layer that has a top elevation of -556 m NGVD29 at northeast corner and top elevation of -667 m NGVD29 at the southwest corner for the best branch: (a) IKD1N1 model, (b) IKD1 model, (c) IK model and (d) Hierarch model. The location of the Baton Rouge fault and the Denham Springs-Scotlandville fault are marked.

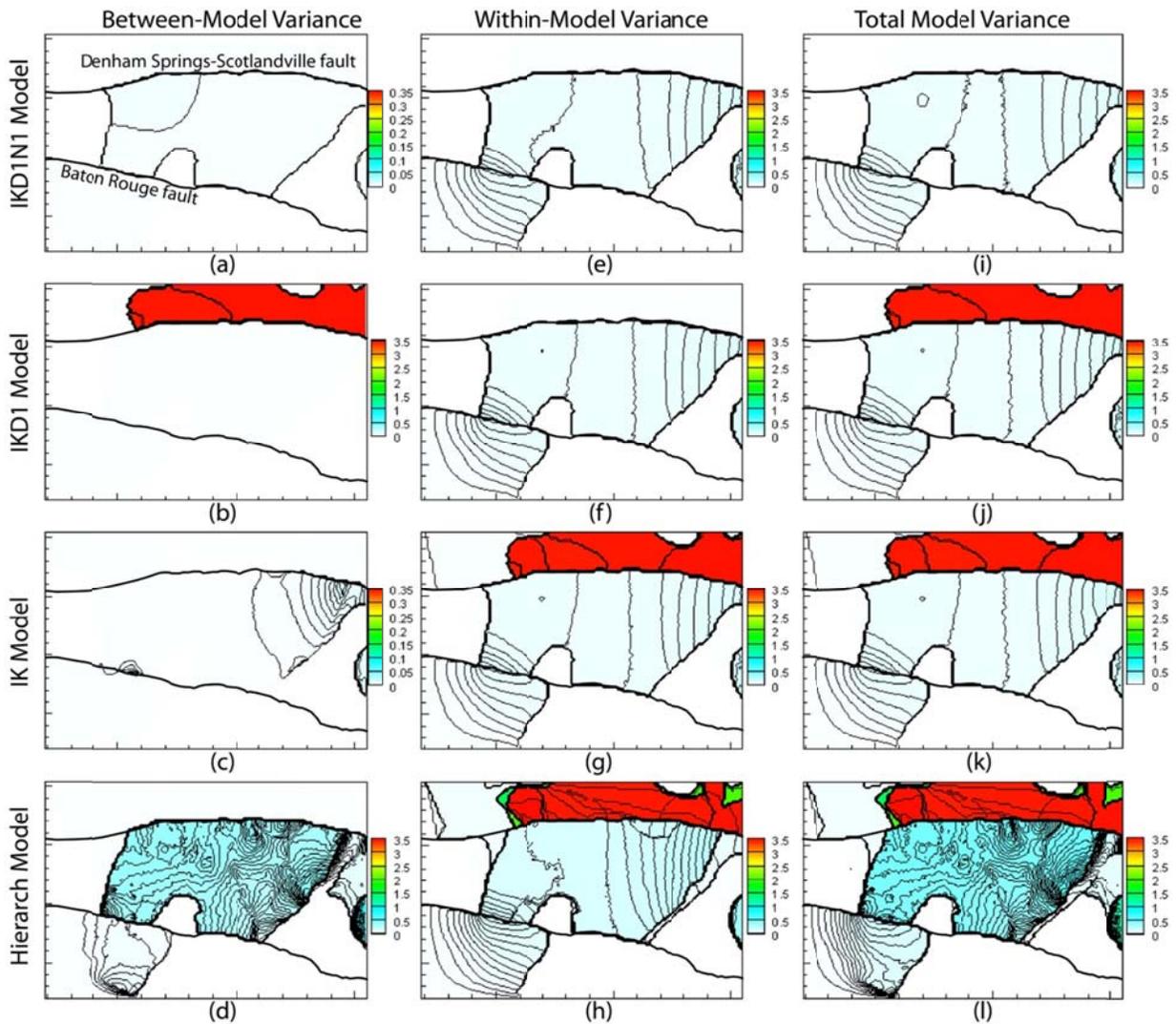


Figure 44 Between-model variance (a)-(d), within-model variance (e)-(h) and total model variance (i)-(l) for the best branch that contains IKD1N1 model, IKD1 model, IK model and the hierarch model for the selected layer in Figure 38 and the last time step. The location of the Baton Rouge fault and the Denham Springs-Scotlandville fault are marked.

7.2.6 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 facilitates 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 E1 proposition 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 E1 proposition are directly connected to the eastern boundary condition. Thus, this level of uncertainty can be dropped.

7.2.7 Critical issues in implementing hierarchical BMA

There are several theoretical and practical challenges in implementing hierarchical BMA. First, quantifying the posterior model probabilities still requires extensive treatment. A major practical 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].

Second, even if the considered uncertain model propositions are exhaustive, adding new unknown model parameters in the calibration process would definitely results in new posterior model probabilities.

Third, a more critical issue is obviously the selection of statistical functions and statistical inference methods, and even more broadly would be a “general hierarchical system of metrics that covers the dimensions of space, time, state/process, and application” [Gupta et al., 2012]. In addition, statistical inference methods do not necessarily need to be confined to Bayesian statistics, but can extend to modern mathematical theories such as evidence theory and imprecise

probability. 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.”

However, the aforementioned concerns imply the plausibility of redirecting our understanding of the model solution from an ontological understanding that is modeling nature per se to an epistemic understanding that is modeling nature relative to our knowledge [Jaynes, 2003; Christakos, 2004; Williamson, 2005]. The term knowledge is not merely limited to our knowledge about the different propositions of the model data, structure, parameters and processes, but also extends to the statistical matrices that shall facilitates the discrimination among these different propositions.

7.3 Conclusions

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 case 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 appears more robust 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 different levels change at observation wells that have limited observation data. 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 “2,000-foot” sand flow model is subject to revision shall new knowledge become available.

The study 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, 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 case study shows that one proposition about eastern boundary condition appears substantially robust, 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 calibration misfit, which is a function of the capability of the calibration algorithm to reach a precise solution in a rugged and noisy search landscape. Yet more importantly is the between-model variance, which is a measure of the uncertainty resulting from candidate knowledge propositions about the natural system. The study shows that between-model variance contribution to the overall uncertainty is additive. This implies that the more we know by testing more propositions, the more the overall model uncertainty will increase, which appears counter

intuitive a priori. How 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 is a topic that requires thorough analysis. For such analysis the hierarchical BMA would be a useful tool.

8 What do we mean by groundwater model uncertainty?

The results of the parameter uncertainty quantification in Section 6.2.4 and Section 7.2.1 provide insights on the meaning of groundwater model uncertainty. The retrieved variance for the “2,000-foot” sand model, which is quantified based on a precise covariance matrix, is very small in comparison to the fitting error. This shows that the quantified variance is only due to parameter estimation error. Thus, it is a measure of the precision of the solution, regardless of the adequacy of the solution. The parameter uncertainty is thus trivialized, since it just represents the estimation error of the calibration algorithm.

Some research groups [e.g. Refsgaard et al., 2006; Rojas et al., 2008] take a step further by noting that by including a calibration step the errors in the conceptual models will be compensated by biased parameter estimates and the calibration result will be at the risk of being biased toward unobserved variables. This study agrees with the idea that the estimated parameters are biased by the data and the model structure, yet suggests that a calibration step is still needed. For a given data and model structure, a global solution for the model parameter will always exist. Regardless of the model adequacy, it is valid to estimate maximum likelihood parameters and quantify the variance related to the parameter estimation error. That would basically be a measure of how far you are from that global solution.

Yet an immediate question arises; how can we then retrieve the model variance with respect to the natural system? That can be done through quantifying model structure variance. Yet unlike the model parameter variance, model structure variance is not the deviation from the “true model” because there is no true model. Being under the impression that the model structure variance is a physical and mind independent feature is to fall in what Jaynes [2003] coined as the mental projection fallacy:

“Common language- or at least, the English language - has an almost universal tendency to disguise epistemological statements by putting them into a grammatical form which suggests to the unwary an ontological statement. A major source of error in current probability theory arises from an unthinking failure to perceive this. To interpret the first kind of statement in the ontological sense is to assert that one's own private thoughts and sensations are realities existing externally in Nature. We call this the “Mind Projection Fallacy”, and note the trouble it causes many times in what follows. But this trouble is hardly confined to probability theory; as soon as it is pointed out, it becomes evident that much of the discourse of philosophers and Gestalt psychologists, and the attempts of physicists to explain quantum theory, are reduced to nonsense by the author falling repeatedly into the Mind Projection Fallacy.”

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 accommodating different candidate model propositions in a constructive epistemic framework that is guided by scientific reasoning as shown in Section 4 and Section 7. In that case, data and model structure variances are retrieved through considering the between model variance of the various candidate model propositions.

Yet still, what do we mean by uncertainty? From the aforesaid perspective, variance is the uncertainty of our current state of knowledge. This uncertainty can increase or decrease by testing new candidate propositions and ousting inadequate propositions.

9 Conclusions

This study addresses the characterization and uncertainty analysis of groundwater systems. The study aims at answering specific question about the hydrogeological settings of the Baton Rouge aquifer-fault system. In addition, the study aims at answering general question with respect the use of indicator geostatistics for hydrofacies architecture reconstruction, CMA-ES for solving the inverse groundwater problem and hierarchical BMA for constructive epistemic modeling.

With respect to the characterization of the Baton Rouge fault-aquifer system, the study revealed the following key points. The study reconstructs the Baton Rouge aquifer-fault system architecture for a Miocene-Pliocene depth interval that consists of the “1,200-foot” sand to the “2,000-foot” sand that are crosscut by the Baton Rouge fault system. First with respect to the aquifer units, the study reveals the following information. There is strong hydraulic connection between the “1,200-foot” sand and the “1,500-foot” sand. Merger of the sand units indicates groundwater recharge from the “1,200-foot” sand to the “1,500-foot” sand. There are four sand deposits that compose the “1,500-foot” sand and the “1,700-foot” sand. There is large amount of missing sand in “1,500-foot” sand in the industrial district and in West Baton Rouge Parish. A distinct clay confining layer separates the “2,000-foot” sand from the “1,700-foot” sand. The sand proportion for the considered depth interval is around 34%.

Second with respect to the Baton Rouge fault system, the study reveals the following information. The Baton Rouge fault has higher sand displacement than the Denham Springs-Scotlandville fault. Displacement increases over depth for both faults. The Denham Springs-Scotlandville fault causes significant sand displacement, and hydraulic continuity occurs due to connection of offset sands. Groundwater model calibration results suggest that at the “2,000-

foot” sand the Denham-Springs Scotlandville fault has much lower permeability in comparison to the Baton Rouge fault. Detailed binary fault architecture and groundwater model calibration implies that the Baton Rouge fault acts as a leaky barrier providing various leaky areas for saltwater to intrude the fresh water aquifers.

Third with respect to the characterization the Baton Rouge aquifer-fault system, the formation dip is the most critical factor. For example, the narrow connection in the “2,000-foot” sand at the east, which allows major leakage from the south, disappears at a step dip, given the available data.

With respect to using indicator geostatistics for hydrofacies architecture reconstruction, the study provides the following contributions. First with respect to hydrofacies architecture reconstruction, the following is concluded. Hydrofacies architecture reconstruction facilitates the detailed analysis of the aquifer-fault system hydrogeological settings, by providing detailed distribution of thickness, lateral extent and depth of different aquifer units. The calibration of hydrofacies architecture models can be less computationally expensive than flow models allowing for finer discretization and extended uncertainty analysis.

Second with respect to the variogram based indicator geostatistics, the following can be concluded. For the depositional environment scale of characterization, traditional variogram-based geostatistics is still a robust choice over the multiple-point training images geostatistics when there are no predefined patterns of the shapes of the aquifer units in practice. While traditional variogram-based geostatistics are robust for handling strongly bimodal heterogeneity, multiple-point training images geostatistics can then be used at smaller scales of characterization. For example, to improve the “2,000-foot” sand groundwater model, it is recommended to further

characterize the sand hydrofacies to several sand types using multiple-point training images geostatistics.

Third with respect to the use of hydrofacies architecture in groundwater modeling, the following can be inferred. By accounting for the geometry and locations flow pathways across the faults and the interconnections of different aquifer units, the hydrofacies architecture makes the geological structure of the groundwater model consistent with the real geology of the aquifer and thus improves model adequacy. Not to mention that hydrofacies data is greatly abundant than flow data. In addition, decoupling geological model structure and parameter estimation alleviates the non-uniqueness of inverse groundwater modeling. Moreover, hydrofacies architecture reduces the complex hydraulic conductivity field to only few hydrofacies that have similar hydraulic characteristics, and thus significantly reduces the groundwater flow model calibration effort.

With respect to using CMA-ES algorithm to solve the inverse groundwater problem, the study showed the following points. First, the CMA-ES is very promising tool for solving the inverse groundwater problem. The elaborate search mechanism of CMA-ES algorithms prove to be more robust in terms of reaching a near-optimal solution for a rugged, nonseparable and noisy function. In addition, the CMA-ES has only one parameter to tune, exhibits solution consistency for repeated runs, shows favorable scaling with increasing the number of processors for parallel run, and has several established invariance properties. Moreover, parallel CMA-ES significantly reduces the computation cost of the inverse groundwater problem, which encourages the development of realistic groundwater model using hydrofacies architectures. In addition, the empirically estimated covariance matrix is precise and can be used for Monte Carlo sampling to quantify parameter related uncertainty.

With respect to using hierarchical BMA for constructive epistemic modeling, the following can be concluded. Using hierarchical Bayesian model averaging (BMA), the study contributes to the debate on the uncertainty of groundwater models by introducing the idea of constructive epistemic modeling that proposes 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. Systemic model dissection through hierarchical BMA permits the understanding of the individual contribution of each uncertain model component and the evaluation of the candidate propositions of each uncertain model component. The study provides two case studies on hydrofacies architecture modeling and groundwater flow modeling. The study shows through developing multiple model the 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 as some final teleological state.

The results of the parameter uncertainty quantification provided some insights on the meaning of groundwater model uncertainty. The retrieved within-model variance for the “2,000-foot” sand model, which is quantified based on a precise covariance matrix, is very small in comparison to the fitting error. This shows that the quantified variance is only due to parameter estimation error, which is a measure of the precision of the solution, regardless of the adequacy of the model. Yet unlike the model parameter variance, model structure variance is not the deviation from the “true model” because there is no true model. Accommodating different candidate model propositions in a constructive epistemic framework is one mean to quantify the model structure variance. Yet can the model structure variance assist in assessing the model adequacy? That is an open question.

Finally, the practical application of this study is to use the groundwater flow model to develop a saltwater intrusion model for the Baton Rouge aquifer-fault system in southeastern Louisiana. The saltwater intrusion model can be used to predict the migration of the saltwater plume and for saltwater intrusion remediation designs.

10 References

- Agarwal, H., J. E. Renaud, E. L. Preston, and D. Padmanabhan (2004), Uncertainty quantification using evidence theory in multidisciplinary design optimization, *Reliability Engineering & System Safety*, 85(1-3), 281-294.
- Akimoto, Y., Y. Nagata, I. Ono, and S. Kobayashi (2012), Theoretical foundation for CMA-ES from information geometry perspective, *Algorithmica*, 64(4), 698-716.
- Aly, A. H., and R. C. Peralta (1999), Optimal design of aquifer cleanup systems under uncertainty using a neural network and a genetic algorithm, *Water Resources Research*, 35(8), 2523-2532.
- Anderson, C. E. (2012), Sources of salinization of the Baton Rouge aquifer system: Southeastern Louisiana, Louisiana State University, MSc Thesis 75pp.
- Andrieu, C., and J. Thoms (2008), A tutorial on adaptive MCMC, *Statistics and Computing*, 18(4), 343-373.
- Auger, A., and N. Hansen (2012), Tutorial CMA-ES: evolution strategies and covariance matrix adaptation, In: *Proceedings of the Fourteenth International Conference on Genetic and Evolutionary Computation Conference*, Philadelphia, Pennsylvania, USA, T. Soule (ed.), 827-848.
- Babu, B. V., and R. Angira (2006), Modified differential evolution (mDE) for optimization of non-linear chemical processes, *Computers & Chemical Engineering*, 30(6-7), 989-1002.
- Bardenet, R., and B. Kégl (2009), Sampling-based optimization with mixtures, In: *OPT 2009 2nd NIPS Workshop on Optimization for Machine Learning*.
- Bastani, M., M. Kholghi, and G. R. Rakhshandehroo (2010), Inverse modeling of variable-density groundwater flow in a semi-arid area in Iran using a genetic algorithm, *Hydrogeology Journal*, 18(5), 1191-1203.
- Baudrit, C., D. Guyonnet, and D. Dubois (2007), Joint propagation of variability and imprecision in assessing the risk of groundwater contamination, *Journal of Contaminant Hydrology*, 93(1-4), 72-84.
- Bayer, P., and M. Finkel (2004), Evolutionary algorithms for the optimization of advective control of contaminated aquifer zones, *Water Resources Research*, 40(6), W06506.
- Bense, V. F., and M. A. Person (2006), Faults as conduit-barrier systems to fluid flow in siliciclastic sedimentary aquifers. *Water Resources Research* 42(5), W05421.
- Bersezio, R., M. Giudici and M. Mele (2007), Combining sedimentological and geophysical data for high-resolution 3-D mapping of fluvial architectural elements in the Quaternary Po plain (Italy). *Sedimentary Geology* 202: 230-248.

- Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrogeological modeling, *Advances in Water Resources*, 16(1), 41-51.
- Beven, K. (2000), On model uncertainty, risk and decision making, *Hydrological Processes*, 14(14), 2605-2606.
- Beven, K. (2005), On the concept of model structural error, *Water Science and Technology*, 52(6), 167-175.
- Beven, K. (2006), On undermining the science?, *Hydrological Processes*, 20(14), 3141-3146.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Processes*, 6(3), 279-298.
- Beven, K., and P. Young (2013), A guide to good practice in modeling semantics for authors and referees, *Water Resources Research*, 49(8), 5092–5098.
- Blasone, R. S., H. Madsen, and D. Rosbjerg (2007), Parameter estimation in distributed hydrological modelling: comparison of global and local optimisation techniques, *Nordic Hydrology*, 38(4-5), 451-476.
- Bledsoe, K. C., J. A. Favorite, and T. Aldemir (2011), A comparison of the covariance matrix adaptation evolution strategy and the Levenberg-Marquardt method for solving multidimensional inverse transport problems, *Annals of Nuclear Energy*, 38(4), 897-904.
- Bredehoeft, J. D. (1997), Fault permeability near Yucca mountain, *Water Resources Research*, 33(11), 2459-2463.
- Buono, A. (1983), The Southern Hills regional aquifer system of southeastern Louisiana and southwestern Mississippi. U.S. Geological Survey, *Water-Resources Investigations Report*: 83-4189, 43pp.
- Caers, J. (2001), Geostatistical reservoir modelling using statistical pattern recognition. *Journal of Petroleum Science and Engineering* 29: 177-188.
- Cardiff, M., and P. K. Kitanidis (2009), Bayesian inversion for facies detection: An extensible level set framework, *Water Resources Research*, 45(10), W10416.
- Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms, *Water Resources Research*, 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, 66p.
- Chen, J. S., Y. R. Rubin (2003), An effective Bayesian model for lithofacies estimation using geophysical data. *Water Resources Research* 39(5), 1118.

- Chester, F. M., J. P. Evans, and R. L. Biegel (1993), Internal structure and weakening mechanisms of the San Andreas fault, *Journal of Geophysical Research-Solid Earth*, 98(B1), 771-786.
- Chilès, J. P., and P. Delfiner (1999), *Geostatistics: modeling spatial uncertainty*, Wiley, New York.
- Christakos, G. (2004), The cognitive basis of physical modelling, In: *Computational Methods in Water Resources (Vol.1)*, *Developments in Water Science* 55, T. M. Miller, M. W. Farthing, W. G. Gray, and G. F. Pinder (eds.), p661-669, Elsevier, Amsterdam, The Netherlands.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resources Research*, 47(9), W09301.
- Comunian, A., P. Renard, J. Straubhaar, and P. Bayer (2011) Three-dimensional high resolution fluvio-glacial aquifer analog - Part 2: Geostatistical modeling. *Journal of Hydrology* 405: 10-23.
- Cui, T., C. Fox, and M. J. O'Sullivan (2011), Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm, *Water Resources Research*, 47(10), W10521.
- 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, *Journal of Hydrology*, 364(3-4), 257-271.
- Desbarats, A.J., and S. Bachu (1994) Geostatistical analysis of the aquifer heterogeneity from the core scale to the basin scale- a case study. *Water Resources Research* 30(3), 673-684.
- 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 Resources Research*, 47(12), W12534.
- Draper, D. (1995), Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 45-97.
- Durham, C.O., and F.M. Peeples (1956), Pleistocene fault zone in southeastern Louisiana: *Transactions of the Gulf Coast Association of Geological Societies* 65 pp.
- ElHarrouni, K., D. Ouazar, G. A. Walters, and A. H. D. Cheng (1996), Groundwater optimization and parameter estimation by genetic algorithm and dual reciprocity boundary element method, *Engineering Analysis with Boundary Elements*, 18(4), 287-296.
- Ellison, A. M. (2004), Bayesian inference in ecology, *Ecology Letters*, 7(6), 509-520.

Elshall, A. S., and F. T.-C. Tsai, Constructive epistemic modeling of groundwater flow with geological structure and boundary condition uncertainty under Bayesian paradigm, *Water Resources Research* (submitted).

Elshall, A. S., F. T.-C. Tsai, and J. S. Hanor (2013), Indicator geostatistics for reconstructing Baton Rouge aquifer-fault hydrostratigraphy, Louisiana, USA, *Hydrogeology Journal*, doi:10.1007/s10040-013-1037-5.

Elshall, A. S., H. Pham, L. Yan and F. T.-C. Tsai, Parallel groundwater model calibration and uncertainty quantification with covariance matrix adaptation, *Advances in Water Resources*, (submitted).

Engdahl, N. B., G. S. Weissmann, and N. D. Bonal (2010), An integrated approach to shallow aquifer characterization: combining geophysics and geostatistics, *Computational Geosciences*, 14(2), 217-229.

Ezzedine S., Y. Rubin, and J.S. Chen (1999), Bayesian method for hydrogeological site characterization using borehole and geophysical survey data: Theory and application to the Lawrence Livermore National Laboratory Superfund site. *Water Resources Research*, 35(9), 2671-2683.

Fairley, J., J. Heffner, and J. Hinds (2003), Geostatistical evaluation of permeability in an active fault zone, *Geophysical Research Letters*, 30(18), 1962.

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, *Hydrogeology Journal*, 15(5), 861-876.

Faunt C.C., K. Belitz, and R.T. Hanson (2010) Development of a three-dimensional model of sedimentary texture in valley-fill deposits of Central Valley, California, USA. *Hydrogeology Journal* 18: 625-649.

Feyen, L., and J. Caers (2006), Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations, *Advances in Water Resources*, 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 Resources Research*, 49(1), 260-282.

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 Resources Research*, 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, *Advances in Water Resources*, 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, *Advances in Water Resources*, 31(8), 1074-1086.
- Griffith, J. M. (2003), Hydrogeologic framework of southeastern Louisiana, Louisiana Department of Transportation and Development Water Resources Technical Report 72.
- Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resources Research*, 48(8), W08301.
- Haario, H., E. Saksman, and J. Tamminen (1999), Adaptive proposal distribution for random walk Metropolis algorithm, *Computational Statistics*, 14(3), 375-395.
- Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7(2), 223-242.
- Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: Efficient adaptive MCMC, *Statistics and Computing*, 16(4), 339-354.
- Hanor, J., E.L. Chamberlain, and F.T.-C. Tsai (2011), Evolution of the permeability architecture of the Baton Rouge fault zone, Louisiana Gulf Coastal plain, American Geophysical Union Fall Meeting December 5-9, 2011, San Francisco.
- Hansen, N. (2006), The CMA-evolution strategy: A comparing review. In: Lozano JA, Larrañaga P, Inza I, Bengoetxea E (eds) *Towards a new evolutionary computation*, Advances in estimation of distribution algorithms pp 75-102, J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea (eds.), Springer, Berlin, Germany.
- Hansen, N., and A. Ostermeier (2001), Completely derandomized self-adaptation in evolution strategies, *Evolutionary Computation*, 9(2), 159-195.
- Hansen, N., and S. Kern (2004), Evaluating the CMA evolution strategy on multimodal test functions, In: *Parallel Problem Solving from Nature- PPSN VIII*, X. Yao, E. K. Burke, J.A. Lozano, J. Smith, J. J. Merelo-Guervós, J. A. Bullinaria, J. E. Rowe, P. Tiño, A. Kabán and H.-P. Schwefel (eds.) pp 282-291, Springer, Berlin, Germany.
- Hansen, N., S. D. Müller, and P. Koumoutsakos (2003), Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evolutionary Computation*, 11(1), 1-18.
- Harbaugh, A. W. (2005), MODFLOW-2005, The U.S. Geological Survey modular groundwater model—the Ground-Water Flow Process: U.S. Geological Survey Techniques and Methods 6-A16, U.S. Geological Survey, Reston, Virginia.
- Haupt, R.L., and S.E. Haupt (2004), *Practical genetic algorithms*, Second Edition, John Wiley and Sons, New York, USA.

- He, L., G. H. Huang, and H. W. Lu (2008), A simulation-based fuzzy chance-constrained programming model for optimal groundwater remediation under uncertainty, *Advances in Water Resources*, 31(12), 1622-1635.
- Hjort, N. L., and G. Claeskens (2003), Frequentist model average estimators, *Journal of the American Statistical Association*, 98(464), 879-899.
- Hjort, N. L., and G. Claeskens (2006), Focused information criteria and model averaging for the Cox hazard regression model, *Journal of the American Statistical Association*, 101(476), 1449-1464.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Statistical Science*, 14(4), 382-401.
- Hooke, R., and T. A. Jeeves (1961), Direct search solution of numerical and statistical problems, *Journal of the Association for Computing Machinery*, ACM 8 (2), 212–229.
- Huntzinger, T. L., C. D. Whiteman, Jr., and D. D. Knochenmus (1985), Simulation of groundwater movement in the “1,500-and 1,700 foot” aquifer of the Baton Rouge area, Louisiana, Louisiana Department of Transportation and Development, Office of Public Works, Water Resources Technical Report 34, 52pp.
- Inoue, M., J. Simunek, S. Shiozawa, and J. W. Hopmans (2000), Simultaneous estimation of soil hydraulic and solute transport parameters from transient infiltration experiments, *Advances in Water Resources*, 23(7), 677-688.
- Irving, J., and K. Singha (2010), Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *Water Resources Research*, 46(11). W11514.
- Iwasaki, N., and K. Yasuda (2005), Adaptive particle swarm optimization using velocity feedback, *International Journal of Innovative Computing Information and Control*, 1(3), 369-380.
- Iwasaki, N., K. Yasuda, and G. Ueno (2006), Dynamic parameter tuning of particle swarm optimization, *IEEJ Transactions on Electrical and Electronic Engineering*, 1(4), v-vi.
- Jacquin, A. P., and A. Y. Shamseldin (2007), Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, *Water Resources Research*, 43(4), W04425.
- Jaynes, E. T. (1990), Probability Theory as Logic, In: Maximum entropy and Bayesian methods, P. F. Fougère (ed.), Kluwer Academic Publishers, Dartmouth.
- Jaynes, E. T. (2003), Probability theory: The logic of science, Cambridge University Press, Cambridge, UK.

- Jiang, Y., C. Liu, C. Huang, and X. Wu (2010), Improved particle swarm algorithm for hydrological parameter optimization, *Applied Mathematics and Computation*, 217(7), 3207-3215.
- Johnson, N.M. (1995), Characterization of alluvial hydrostratigraphy with indicator semivariograms. *Water Resources Research* 31(12), 3217-3227.
- Johnson NM, Dreiss SJ (1989) Hydrostratigraphic interpretation using indicator geostatistics. *Water Resources Research* 25(12), 2501-2510.
- Johnson, N. M. (1995), Characterization of alluvial hydrostratigraphy with indicator semivariograms, *Water Resources Research*, 31(12), 3217-3227.
- Johnson, N. M., and S. J. Dreiss (1989), Hydrostratigraphic interpretation using indicator geostatistics, *Water Resources Research*, 25(12), 2501-2510.
- Journel, A.G. (1983), Nonparametric-estimation of spatial distributions. *Journal of the International Association for Mathematical Geology* 15: 445-468.
- Karpouzou, D. K., F. Delay, K. L. Katsifarakis, and G. de Marsily (2001), A multipopulation genetic algorithm to solve the inverse problem in hydrogeology, *Water Resources Research*, 37(9), 2291-2302.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resources Research*, 42(3), W03407.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis, *Journal of Hydrology*, 320(1-2), 187-201.
- Kennedy, J., and R. Eberhart (1995), Particle swarm optimization, In: *IEEE International Conference on Neural Networks*, November 1995, Perth, Washington, USA, Vol.4, 1942-1948.
- Kitanidis, P. K. (1986), Parameter uncertainty in estimation of spatial functions – Bayesian-analysis, *Water Resources Research*, 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 Resources Research* 32(9), 2617-2658.
- Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resources Research*, 32(9), 2617-2658.
- Li, L., H. Zhou, H. J. H. Franssen, and J. J. Gomez-Hernandez (2012), Groundwater flow inverse modeling in non-MultiGaussian media: performance assessment of the normal-score Ensemble Kalman Filter, *Hydrology and Earth System Sciences*, 16(2), 573-590.

- Li, X., and F. T.- C. Tsai (2009), Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod, *Water Resources Research*, 45(9), W09403.
- Li, L., Zhou, H., Hendricks Franssen, H.J., and Gómez-Hernández, J.J. (2012a), Groundwater flow inverse modeling in non-multiGaussian media: performance assessment of the normal-score Ensemble Kalman Filter. *Hydrology and Earth System Sciences*, 16, 573–590, doi:10.5194/hess-16-573-2012.
- Li, L., Zhou, H., Hendricks Franssen, H.J., and Gómez-Hernández, J.J. (2012b), Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via Ensemble Kalman Filter. *Journal of Hydrology*, vol. 428-429, 152–169.
- Linde, N., A. Binley, A. Tryggvason, L.B. Pedersen, and A. Revil (2006), Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltimes data. *Water Resources Research* 42(12), W12404.
- Lloyd, S.P. (1982), Least-squares quantization in PCM. *IEEE Transactions on Information Theory* 28: 129-137.
- Lovelace, J.K. (2007), Chloride concentrations in ground water in East and West Baton Rouge Parishes, Louisiana, 2004-05 Scientific Investigation Report U.S. Geological Survey.
- Lovelace, J.K. (2009), Ground water levels and salt water encroachment in major aquifers in Louisiana, USGS Ground Water Summit, February 4, 2009.
- Marquardt, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431-441.
- Matott, L. S., and A. J. Rabideau (2008a), Calibration of subsurface batch and reactive-transport models involving complex biogeochemical processes, *Advances in Water Resources*, 31(2), 269-286.
- Matott, L. S., and A. J. Rabideau (2008b), Calibration of complex subsurface reaction models using a surrogate-model approach, *Advances in Water Resources*, 31(12), 1697-1707.
- McCulloh, R. P., and P. V. Heinrich (2012), Surface faults of the south Louisiana growth-fault province. In: *Recent Advances In North American Paleoseismology and Neotectonics East of the Rockies and Use of the Data in Risk Assessment and Policy*, R.T. Cox, M. Tuttle, O. Boyd, and J. Locat (eds.), Geological Society of America, Boulder, Colorado.
- Meyer R.R., J.R. Rollo (1965), Saltwater encroachment, Baton Rouge area, Louisiana Department of Conservation, Louisiana Geological Survey and Louisiana Department of Public Works Water Resources, pp. 9.
- Meyer, R.R., and A.N. Turcan Jr. (1955), Geology and ground-water resources of the Baton Rouge area, Louisiana: U.S. Geological Survey Water-Supply Paper 1296, 138 p.

- Meyer, R. R., and A. N. Turcan, Jr. (1955), Geology and ground-water resources of the Baton Rouge area, Louisiana, U.S. Geological Survey, Water Supply Paper 1296, 138pp.
- Miller, R.B., J.W. Castle, and T.J. Temples (2000), Deterministic and stochastic modeling of aquifer stratigraphy, South Carolina. *Ground Water* 38: 284-295
- 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 Environmental Research and Risk Assessment*, 24(6), 863-880.
- Morgan C.O., and M.D. Winner Jr. (1964), Salt-water encroachment in aquifers of the Baton Rouge area— preliminary report and proposal Louisiana Department of Public Works, pp. 37.
- Mugunthan, P., and C. A. Shoemaker (2006), Assessing the impacts of parameter uncertainty for computationally expensive groundwater models, *Water Resources Research*, 42(10), W10428.
- Murray, G. E. (1961), *Geology of the Atlantic and Gulf coastal province of North America*, Harper & Brothers, New York.
- Müller, C. L. (2010), Exploring the common concepts of adaptive MCMC and Covariance Matrix Adaptation schemes, In: *Theory of evolutionary algorithms*, A. Auger, J. L. Shapiro, L.D. Whitley and C. Witt (eds.), 10361, Dagstuhl Seminar Proceedings, Dagstuhl, Germany.
- Müller, C., and I. F. Sbalzarini (2010), Gaussian adaptation as a unifying framework for continuous black-box optimization and adaptive Monte Carlo sampling, In: *2010 IEEE Congress on Evolutionary Computation (CEC)*, Barcelona, Spain, doi: 10.1109/ CEC.2010. 5586491.
- Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environmental Research and Risk Assessment*, 17(5), 291-305.
- Neuman, S. P., and P. J. Wierenga (2003), A comprehensive strategy of hydrogeologic modelling and uncertainty analysis for nuclear facilities and sites, NUREG/CR-6805, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 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, *Hydrogeology Journal*, 17(7), 1699-1725.
- Nowak, W., and O. A. Cirpka (2004), A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inverting, *Advances in Water Resources*, 27(7), 737-750.
- 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 Resources Research*, 46(3), W03535.

- Nützmann, G., M. Thiele, S. Maciejewski, and K. Joswig (1998), Inverse modelling techniques for determining hydraulic properties of coarse-textured porous media by transient outflow methods, *Advances in Water Resources*, 22(3), 273-284.
- Pham, V. H. and F. T.-C. Tsai (2013), Conversion of highly complex faulted hydrostratigraphic architectures into MODFLOW grid for groundwater modeling, 2013 American Geophysical Union Fall Meeting, abstract, San Francisco, California, 9-13 December 2013.
- Poeter, E., and D. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Ground Water*, 43(4), 597-605.
- Popper, K. (1959), The propensity interpretation of probability, *British Journal for the Philosophy of Science*, 10, 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: *Logic, Epistemology, and the Unity of Science*, Vol. 10, S. Rahman and J. Symons (eds.), Springer, Dordrecht, The Netherlands.
- Proce, C.J., R.W. Ritzi, D.F. Dominic, Z.X. Dai (2004), Modeling multiscale heterogeneity and aquifer interconnectivity. *Ground Water* 42: 658-670.
- Qi, Y., and T. P. Minka (2002), Hessian-based Markov chain Monte-Carlo algorithms, In: *First Cape Cod Workshop on Monte Carlo Methods*, Cape Cod, Massachusetts, USA.
- Raftery, A. E. (1995), Bayesian model selection in social research, *Sociological Methodology*, 25, 111-163.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* 25, 111-163.
- Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem (2007), Uncertainty in the environmental modelling process: A framework and guidance, *Environmental Modelling & Software*, 22(11), 1543-1556.
- 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, *Advances in Water Resources*, 29(11), 1586-1597.
- Refsgaard, J. C., S. Christensen, T. O. Sonnenborg, D. Seifert, A. L. Højberg, and L. Troldborg (2012), Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, *Advances in Water Resources*, 36, 36-50.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, 46(5), W05521.

- Riegler, A. (2012), Constructivism, In: Paradigms, in Theory Construction, L. L'Abate (ed.). Springer, New York, pp. 235–256.
- Rojas, R., L. Feyen, and A. Dassargues (2008), Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging, *Water Resources Research*, 44(12), W12418.
- 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, *Hydrological Processes*, 23(8), 1131-1146.
- 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 Resources Research*, 46(8), W08520.
- Rojas, R., O. Batelaan, L. Feyen, and A. Dassargues (2010a), Assessment of conceptual model uncertainty for the regional aquifer Pampa del Tamarugal - North Chile, *Hydrology and Earth System Sciences*, 14(2), 171-192.
- 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, *Journal of Hydrology*, 394(3-4), 416-435.
- Rollo, J. R. (1969), Salt-water encroachment in aquifers of the Baton Rouge area, Louisiana, Louisiana Department of Conservation, Louisiana Geological Survey, and Louisiana Department of Public Works, *Water Resources Bulletin No. 13*, 45pp.
- Rubin, Y. (2003), *Applied stochastic hydrogeology*, Oxford University Press, New York.
- Sakaki, T., C. C. Fripiat, M. Komatsu, and T. H. Illangasekare (2009), On the value of lithofacies data for improving groundwater flow model accuracy in a three-dimensional laboratory-scale synthetic aquifer, *Water Resources Research*, 45(11), W11404.
- Salve, R., and C.M. Oldenburg (2001), Water flow within a fault in altered nonwelded tuff, *Water Resources Research* 37(12), 3043-3056.
- Salve, R., and C. M. Oldenburg (2001), Water flow within a fault in altered nonwelded tuff, *Water Resources Research*, 37(12), 3043-3056.
- Sargent, B.P. (2012), *Water use in Louisiana, 2010: Louisiana Department of Transportation and Development Water Resources Special Report no. 17 (Revised)*, 135pp.
- Scharling P.B., E.S. Rasmussen, T.O. Sonnenborg, P. Engesgaard, and K. Hinsby (2009), Three-dimensional regional-scale hydrostratigraphic modeling based on sequence stratigraphic methods: a case study of the Miocene succession in Denmark. *Hydrogeology Journal* 17: 1913-1933.

- Scheerlinck, K., V. R. N. Pauwels, H. Vernieuwe, and B. De Baets (2009), Calibration of a water and energy balance model: Recursive parameter estimation versus particle swarm optimization, *Water Resources Research*, 45(10), W10422.
- Schulmeister, M.K., J.J. Butler, J.M. Healey, L. Zheng, D.A. Wysocki, and G.W. McCall (2003), Direct-push electrical conductivity logging for high-resolution hydrostratigraphic characterization. *Ground Water Monitoring and Remediation* 23: 52-62
- Schwarz, G. E. (1978), Estimating the dimension of a model. *Annals of Statistics* 6 (2): 461–464.
- Seifert, D., T. O. Sonnenborg, J. C. Refsgaard, A. L. Højberg, and L. Troldborg (2012), Assessment of hydrological model predictive ability given multiple conceptual geological models, *Water Resources Research*, 48(6), W06503.
- Sibson, R. (1980), Vector identity for Dirichlet tessellation. *Mathematical Proceedings of the Cambridge Philosophical Society* 87: 151-155.
- Singh, A., S. Mishra, and G. Ruskauff (2010), Model averaging techniques for quantifying conceptual model uncertainty, *Ground Water*, 48(5), 701-715.
- Singha, K., D. W. Hyndman, and F. D. Day-Lewis (2007), Introduction, In: *Subsurface Hydrology: Data Integration for Properties and Processes*, Geophysical Monograph Series, Vol. 171, D. W. Hyndman, F. D. Day-Lewis, and K. Singha (eds.), pp. 1–5, American Geophysical Union, Washington, D.C.
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resources Research*, 44(12), W00B05.
- Socha, K., and M. Dorigo (2008), Ant colony optimization for continuous domains, *European Journal of Operational Research*, 185(3), 1155-1173.
- Solomatine, D. P., Y. B. Dibike, and N. Kukuric (1999), Automatic calibration of groundwater models using global optimization techniques, *Hydrological Sciences Journal*, 44(6), 879-894.
- Storn, R., and K. Price (1997), Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, 11(4), 341-359.
- Strebelle, S. (2002), Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology* 34: 1-21.
- Sun, N. Z., and W. W. G. Yeh (1990), Coupled inverse problems in groundwater modeling: 1. sensitivity analysis and parameter-identification, *Water Resources Research*, 26(10), 2507-2525.
- Sun, N.-Z. (1999), Inverse problems in groundwater modeling, In: *Theory and applications of transport in porous media*, J. Bear (ed.), Kluwer Academic Publishers, Dordrecht, Netherlands.

- Tang, G., E. F. D'Azevedo, F. Zhang, J. C. Parker, D. B. Watson, and P. M. Jardine (2010), Application of a hybrid MPI/OpenMP approach for parallel groundwater model calibration using multi-core computers, *Computers & Geosciences*, 36(11), 1451-1460.
- Tang, Y., P. M. Reed, and J. B. Kollat (2007), Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications, *Advances in Water Resources*, 30(3), 335-353.
- Tartakovsky, A.M., D. Bolster, D.M. Tartakovsky (2008), Hydrogeophysical approach for identification of layered structures of the vadose zone from electrical resistivity data. *Vadose Zone Journal* 7: 1207-1214.
- 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.
- Torak L.J., and C.D. Whiteman Jr (1982), Applications of digital modeling for evaluating the ground-water resources of the "2,000-Foot" sand of the Baton Rouge area, Louisiana, Louisiana Department of Transportation and Development, Office of Public Works Water Resources Technical Report no. 27 , 87pp.
- 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, Louisiana Department of Transportation and Development, Office of Public Works, Water Resources Technical Report No.27, 87pp.
- Trevisani, S., and P. Fabbri (2010), Geostatistical modeling of a heterogeneous site bordering the Venice Lagoon, Italy, *Ground Water*, 48(4), 614-623.
- Troldborg, M., W. Nowak, N. Tuxen, P. L. Bjerg, R. Helmig, and P. J. Binning (2010), Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework, *Water Resources Research*, 46(12), W12552.
- Tsai, F. T.-C. (2006), Enhancing random heterogeneity representation by mixing the kriging method with the zonation structure. *Water Resources Research* 42(8), W08428.
- Tsai, F. T.-C. (2010), Bayesian model averaging assessment on groundwater management under model structure uncertainty. *Stochastic Environmental Research and Risk Assessment* 24: 845-861.
- Tsai, F. T.-C., X., Li (2008), Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window. *Water Resources Research* 44(9), W09434.
- Tsai, F. T.-C., and W.W.G. Yeh (2004), Characterization and identification of aquifer heterogeneity with generalized parameterization and Bayesian estimation. *Water Resources Research* 40(10), W10102.

- Tsai, F. T.-C. (2010), Bayesian model averaging assessment on groundwater management under model structure uncertainty, *Stochastic Environmental Research and Risk Assessment*, 24, 845-861.
- Tsai, F. T.-C., N. Z. Sun, and W. W. G. Yeh (2003a), Global-local optimization for parameter structure identification in three-dimensional groundwater modeling, *Water Resources Research*, 39(2), 1043.
- Tsai, F. T. -C., N. Z. Sun, and W. W. G. Yeh (2003b), A combinatorial optimization scheme for parameter structure identification in ground water modeling, *Ground Water*, 41(2), 156-169.
- Tsai, F. T.-C. (2006), Enhancing random heterogeneity representation by mixing the kriging method with the zonation structure, *Water Resources Research*, 42(8), W08428.
- Tsai, F. T.-C. (2009), Indicator generalized parameterization for interpolation point selection in groundwater inverse modeling, *Journal of Hydrologic Engineering*, 14(3), 233-242.
- Tsai, F. T.-C. (2010), Bayesian model averaging assessment on groundwater management under model structure uncertainty, *Stochastic Environmental Research and Risk Assessment*, 24(6), 845-861.
- Tsai, F. T.-C., and A. S., Elshall (2013) Hierarchical Bayesian model averaging for hydrostratigraphic modeling: Uncertainty segregation and comparative evaluation, *Water Resources Research*, 49(9), 5520–5536.
- Tsai, F. T.-C., and W. W.-G. Yeh (2004), Characterization and identification of aquifer heterogeneity with generalized parameterization and Bayesian estimation, *Water Resources Research*, 40(10), W10102.
- Tsai, F. T.-C., and X. Li (2008a), Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window, *Water Resources Research*, 44(9), W09434.
- Tsai, F. T.-C., and X. Li (2008b), Multiple parameterization for hydraulic conductivity identification, *Ground Water*, 46(6), 851-564.
- 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 Resources Research*, 46(2), W02802.
- Tsai, F. T.-C., N.-Z. Sun, and W. W.-G. Yeh (2003), A combinatorial optimization scheme for parameter structure identification in ground water modeling, *Ground Water*, 41(2) 156-169.
- Usunoff, E., J. Carrera, and S. F. Mousavi (1992), An approach to the design of experiments for discriminating among alternative conceptual models, *Advances in Water Resources*, 15(3), 199-214.

von Bertalanffy, L. (1968), *General system theory: Essays on its foundation and development*, Revised Edition, George Braziller, New York.

von Mises, R. (1964), *Mathematical theory of probability and statistics*, Academic Press, New York.

Vrugt, J. A., B. O Nuallain, B. A. Robinson, W. Bouten, S. C. Dekker, and P. M. A. Sloot (2006), Application of parallel computing to stochastic parameter estimation in environmental models, *Computers & Geosciences*, 32(8), 1139-1155.

Vrugt, J. A., P. H. Stauffer, T. Woehling, B. A. Robinson, and V. V. Vesselinov (2008), Inverse modeling of subsurface flow and transport properties: A review with new developments, *Vadose Zone Journal*, 7(2), 843-864.

Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochastic Environmental Research and Risk Assessment*, 19(6), 378-387.

Weiss, R., and L. Smith (1998), Parameter space methods in joint parameter estimation for groundwater flow models, *Water Resources Research*, 34(4), 647-661.

Weissmann, G.S., S.F. Carle SF, and G.E. Fogg (1999), Three dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics. *Water Resources Research* 35(6), 1761-1770.

Weissmann, G.S., and G.E. Fogg (1999), Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. *Journal of Hydrology* 226: 48-65.

Wiederhold H, H.M., Rumpel, E. Auken, B. Siemon, W. Scheer, and R. Kirsch (2008), Geophysical methods for investigation and characterization of groundwater resources in buried valleys. *Grundwasser* 13: 68-77.

Williamson, J. (2005), *Bayesian nets and causality: philosophical and computational foundations*, Oxford University Press, Oxford, UK.

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 Resources Research*, 44(12), W12432.

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 Frank T.-C. Tsai and Xiaobao Li, *Water Resources Research* 46(2), W02801.

- Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohll, and D. M. Reeves (2010b), A Model-Averaging Method for Assessing Groundwater Conceptual Model Uncertainty, *Ground Water*, 48(5), 716-728.
- Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resources Research*, 40(5), W05113.
- 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 Resources Research*, 41(12), W12429.
- Yeh, W. W.-G. (1986), Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resources Research*, 22(2), 95–108.
- Yoon, J. H., and C. A. Shoemaker (1999), Comparison of optimization methods for groundwater bioremediation, *Journal of Water Resources Planning and Management-ASCE*, 125(1), 54-63.
- Zappa G, R. Bersezio, F. Felletti, and M. Giudici (2006), Modeling heterogeneity of gravel-sand, braided stream, alluvial aquifers at the facies scale. *Journal of Hydrology* 325: 134-153
- Zhang, Y., and C. Sutton (2011), Quasi-Newton Methods for Markov Chain Monte Carlo, In: *Advances in neural information processing systems* 24, J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. C. N. Pereira and K. Q. Weinberger (eds.), pp. 2393-2401.
- Zhang, Y., D. Gallipoli, and C. E. Augarde (2009), Simulation-based calibration of geotechnical parameters using parallel hybrid moving boundary particle swarm optimization, *Computers and Geotechnics*, 36(4), 604-615.
- Zhou, H., Li, L., Hendricks Franssen, H. J., and Gómez-Hernández, J. J. (2011) Pattern recognition in a bimodal aquifer by normal score Ensemble Kalman Filter. *Mathematical Geosciences*, 44 (2), 169-185.
- Zidane, A., A. Younes, P. Huggenberger, and E. Zechner (2012), The Henry semianalytical solution for saltwater intrusion with reduced dispersion, *Water Resources Research*, 48(6), W06533.

Appendix: Open access permissions

Open access permission for the Water Resources Research manuscript

10/31/13	Rightslink Printable License
JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS	
Oct 31, 2013	
<hr/> <hr/>	
<p>This is a License Agreement between Ahmed S. Elshall ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.</p>	
<p>All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.</p>	
License Number	3259571303569
License date	Oct 31, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Water Resources Research
Licensed content title	Hierarchical Bayesian model averaging for hydrostratigraphic modeling: Uncertainty segregation and comparative evaluation
Licensed copyright line	©2013. American Geophysical Union. All Rights Reserved.
Licensed content author	Frank T.-C. Tsai,Ahmed S. Elshall
Licensed content date	Sep 9, 2013
Start page	5520
End page	5536
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Total	0.00 USD
Terms and Conditions	
TERMS AND CONDITIONS	
<p>This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at</p>	
https://s100.copyright.com/AppDispatchServlet	1/7

<http://myaccount.copyright.com>).

Terms and Conditions

1. The materials you have requested permission to reproduce (the "Materials") are protected by copyright.

2. You are hereby granted a personal, non-exclusive, non-sublicensable, non-transferable, worldwide, limited license to reproduce the Materials for the purpose specified in the licensing process. This license is for a one-time use only with a maximum distribution equal to the number that you identified in the licensing process. Any form of republication granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before may be distributed thereafter). The Materials shall not be used in any other manner or for any other purpose. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Material. Any third party material is expressly excluded from this permission.

3. With respect to the Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Materials, or any of the rights granted to you hereunder to any other person.

4. The Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc or one of its related companies (WILEY) or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

5. NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT

AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

6. WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

7. You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

8. IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

9. Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

10. The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

11. This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

12. Any fee required for this permission shall be non-refundable after thirty (30) days from receipt

13. These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

14. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

15. WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

16. This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

17. This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

Wiley Open Access Terms and Conditions

Wiley publishes Open Access articles in both its Wiley Open Access Journals program [<http://www.wileyopenaccess.com/view/index.html>] and as Online Open articles in its subscription journals. The majority of Wiley Open Access Journals have adopted the [Creative Commons Attribution License](#) (CC BY) which permits the unrestricted use, distribution, reproduction, adaptation and commercial exploitation of the article in any medium. No permission is required to use the article in this way provided that the article is properly cited and other license terms are observed. A small number of Wiley Open Access journals have retained the [Creative Commons Attribution Non Commercial License](#) (CC BY-NC), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Online Open articles - Authors selecting Online Open are, unless particular exceptions apply, offered a choice of Creative Commons licenses. They may therefore select from the CC BY, the CC BY-NC and the [Attribution-NoDerivatives](#) (CC BY-NC-ND). The CC BY-NC-ND is more restrictive than the CC BY-NC as it does not permit adaptations or modifications without rights holder consent.

Wiley Open Access articles are protected by copyright and are posted to repositories and websites in accordance with the terms of the applicable Creative Commons license referenced on the article. At the time of deposit, Wiley Open Access articles include all changes made during peer review, copyediting, and publishing. Repositories and websites that host the article are responsible for incorporating any publisher-supplied amendments or retractions issued subsequently.

Wiley Open Access articles are also available without charge on Wiley's publishing platform, **Wiley Online Library** or any successor sites.

Conditions applicable to all Wiley Open Access articles:

- The authors' moral rights must not be compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be damaged).
- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.
- If article content is copied, downloaded or otherwise reused for research and other purposes as permitted, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on Wiley Online Library) should be maintained. Copyright notices and disclaimers must not be deleted.
 - Creative Commons licenses are copyright licenses and do not confer any other rights, including but not limited to trademark or patent rights.
- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

Conditions applicable to non-commercial licenses (CC BY-NC and CC BY-NC-ND)

For non-commercial and non-promotional purposes individual non-commercial users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles. In addition, articles adopting the CC BY-NC may be adapted, translated, and text- and data-mined subject to the conditions above.

Use by commercial "for-profit" organizations

Use of non-commercial Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:

- Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;
- Copying, downloading or posting by a site or service that incorporates advertising with such content;
- The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)

- Use of article content (other than normal quotations with appropriate citation) by for-profit organizations for promotional purposes
- Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;
- Use for the purposes of monetary reward by means of sale, resale, license, loan, transfer or other form of commercial exploitation such as marketing products
- Print reprints of Wiley Open Access articles can be purchased from corporatesales@wiley.com

The modification or adaptation for any purpose of an article referencing the CC BY-NC-ND License requires consent which can be requested from RightsLink@wiley.com.

Other Terms and Conditions:

BY CLICKING ON THE "I AGREE..." BOX, YOU ACKNOWLEDGE THAT YOU HAVE READ AND FULLY UNDERSTAND EACH OF THE SECTIONS OF AND PROVISIONS SET FORTH IN THIS AGREEMENT AND THAT YOU ARE IN AGREEMENT WITH AND ARE WILLING TO ACCEPT ALL OF YOUR OBLIGATIONS AS SET FORTH IN THIS AGREEMENT.

v1.8

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501148950. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

**SPRINGER LICENSE
TERMS AND CONDITIONS**

Nov 06, 2013

This is a License Agreement between Ahmed S. Elshall ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3263120549787
License date	Nov 06, 2013
Licensed content publisher	Springer
Licensed content publication	Hydrogeology Journal
Licensed content title	Indicator geostatistics for reconstructing Baton Rouge aquifer-fault hydrostratigraphy, Louisiana, USA
Licensed content author	Ahmed S. Elshall
Licensed content date	Jan 1, 2013
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	
Title of your thesis / dissertation	Characterization and uncertainty analysis of siliciclastic aquifer-fault system
Expected completion date	Nov 2013
Estimated size(pages)	172
Total	0,00 USD

Terms and Conditions

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

Limited License

With reference to your request to reprint in your thesis material on which Springer Science and

Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: <http://www.sherpa.ac.uk/romeo/>). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding: with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers

established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms

Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration). **OR:**

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501152918. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:

Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support:
customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-
2777.

Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable
license for your reference. No payment is required.

Vita

Ahmed Elshall was born in Cairo, Egypt in December, 1978. He received a bachelor's degree in Construction Engineering from the American University in Cairo in March 2003. Then he worked in construction engineering and environmental management. In September 2007, he moved to Tübingen, Germany from where he obtained a master's degree in Applied Environmental Geoscience in September 2009. He continued working in Tübingen until August 2010 when he moved to Baton Rouge to start a PhD in Civil Engineering. He is expected to receive the degree in December 2013.