Journal of Hydrology 601 (2021) 126682

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Accounting for uncertainty in complex alluvial aquifer modeling by Bayesian multi-model approach

Jina Yin [a,b], Frank T.-C. Tsai [c,*], Shih-Chieh Kao [d]

[a] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China
[b] Yangtze Institute for Conservation and Development, Hohai University, Nanjing, China
[c] Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
[d] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Zoi Dokou, Associate Editor

Keywords:
Groundwater
Modeling
Alluvial aquifer
Parameter uncertainty
Bayesian model averaging

ABSTRACT

Alluvial aquifers by nature are complex caused by varied depositional environments. Developing a reliable groundwater model to represent an alluvial aquifer is non-trivial. Also, relying on a single best calibrated model may not be sufficient because of an inadequate choice of model parameter values. To better understand groundwater dynamics and improve model prediction reliability, this study presents a Bayesian multi-model uncertainty quantification (BMMUQ) framework to account for model parameter uncertainty in complex alluvial groundwater modeling. The methodology was applied to the agriculturally intensive Mississippi River alluvial aquifer (MRAA), Northeast Louisiana. An aquifer architecture was first constructed using 7,259 well logs in the MRAA area which covers three fluvial deposits (alluvium, braided-stream terrace, and braided-stream terrace-loess). A 12-layer MODFLOW model was then developed to address the alluvial aquifer complexity and well calibrated through a genetic algorithm. This study quantified model parameter uncertainty in hydraulic conductivity and specific storage of sand facies. Bayesian model averaging (BMA) with the Expectation Maximization (EM) algorithm was adopted to derive posterior model weights and head variances of 50 alternative conceptual groundwater flow models, and thereby obtains BMA ensemble model predictions instead of only relying on the best calibrated conceptual model. Results show that an estimated around 950 million m^3 of groundwater storage loss occurs in 2015 with respect to the beginning of 2004, due to high groundwater demand for irrigation in the MRAA area. Explicitly quantifying model uncertainty can produce more reliable groundwater level predictions from BMA ensemble model. The presented groundwater modeling framework improves our understanding of the MRAA and provides a valuable tool to assist agricultural water management.

1. Introduction

Groundwater serves as a critical source to meet water demands for different sectors, especially for agriculture in the USA (Dieter et al., 2018). In Louisiana, an estimated 6.62 million m^3/day of groundwater was withdrawn in 2015, 41 percent of which was consumed for crop irrigation (Collier and Sargent, 2018). Excessive groundwater pumping can lead to various detrimental problems, such as groundwater level decline, groundwater storage reduction and saltwater intrusion (Clark et al., 2013; Pauloo et al., 2020; Smith et al., 2017; Yin and Tsai, 2018). Accurate prediction in groundwater levels matters to decision support in sustainable water resources management, especially in an agriculture-dominated region. The prediction reliability of groundwater flow models is strongly influenced by different sources of uncertainty (Mustafa et al., 2018; Tsai and Elshall, 2013). To ensure reliable predictions and decision support, it is important to assess conceptual model uncertainty related to the complexity of an aquifer system.

More than eleven (11) freshwater aquifer systems exist in Louisiana (Stuart et al., 1994). Among them, the Mississippi River alluvial aquifer (hereafter referred to as MRAA) is the second most pumped aquifer

Abbreviations: BMMUQ, Bayesian Multi-model Uncertainty Quantification; BMA, Bayesian Model Averaging; M, Expectation Maximization; GHB, General Head Boundary; LDNR, Louisiana Department of Natural Resources; MAP, Mississippi Alluvial Plain; MRAA, Mississippi River Alluvial Aquifer; NGVD, National Geodetic Vertical Datum; NHD, National Hydrography Dataset; NHDPlus, National Hydrography Dataset Plus; NSE, Nash-Sutcliffe Efficiency; RMSE, Root Mean Square Error; USGS, United States Geological Survey; VIC, Variable Infiltration Capacity; WBD, Watershed Boundary Dataset.

* Corresponding author.
E-mail addresses: jnyin@hhu.edu.cn (J. Yin), ftsai@lsu.edu (F. T.-C. Tsai), kaos@ornl.gov (S.-C. Kao).

https://doi.org/10.1016/j.jhydrol.2021.126682
Received 24 January 2021; Received in revised form 20 May 2021; Accepted 9 July 2021
Available online 15 July 2021

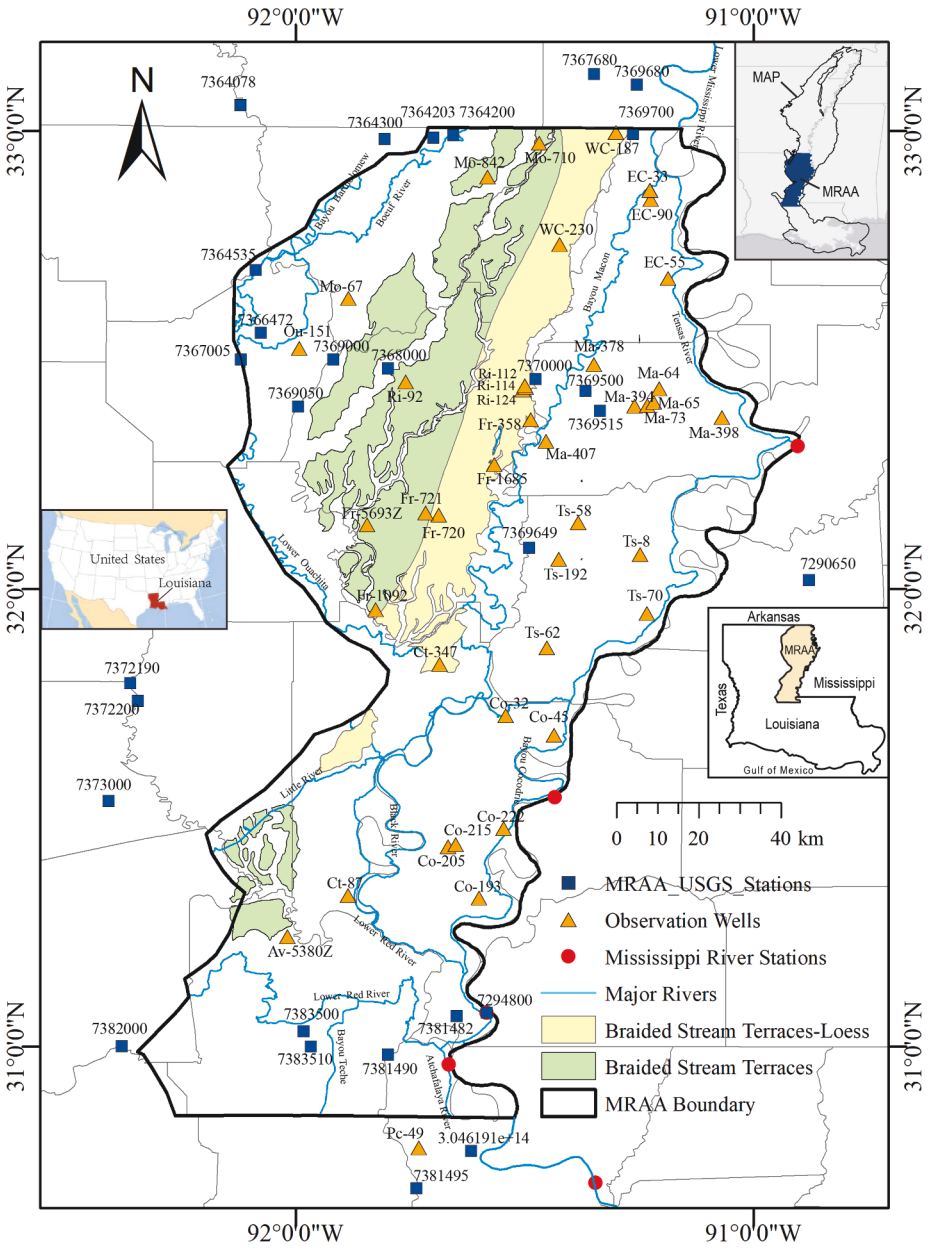

Fig. 1. Map of the study area including USGS surface-water stations, Mississippi River stations, USGS groundwater observation wells and major rivers. Different geological formation zones are indicated with different colors.

(Collier and Sargent, 2018), which underlies the Mississippi Alluvial Plain (MAP), from the north border between Arkansas and Louisiana to the south-central part of Louisiana (Carlson, 2006; Stuart et al., 1994; Whitfield, 1975). As the uppermost aquifer in this area, withdrawals of groundwater from the MRAA are primarily used for irrigation of rice, soybeans, and other crops. However, heavy pumping has led to substantial, widespread water-level decline in the northeastern Louisiana (Yasarer et al., 2020). Groundwater recharge is exceeded by groundwater pumping in some areas, and has resulted in extensive cones of depression. Since the MRAA supplies a large quantity of water for agriculture, understanding groundwater dynamics in the MRAA is imperative to support sustainable agriculture. The MRAA has been known to be highly complex in geological architecture in the Lower Mississippi Valley (Krinitzsky and Wire, 1964). The complexity refers to highly non-uniform sand thickness, interbedded clays, and pinch-outs, which are revealed by a large quantity of well log data. However, the MRAA was overly simplified in some region-scale groundwater models, such as the Mississippi Embayment aquifer system (Clark et al., 2013). Also, the existing groundwater studies related to the Mississippi River Valley alluvial aquifer (MRVA) were mostly limited to Arkansas State and Mississippi State, which include studies on (i) groundwater modeling studies (Arthur, 2001; Gillip and Czarnecki, 2009; Reed, 2003), (ii) groundwater sustainability (Czarnecki et al., 2003), (iii) water quality (Sharif et al., 2008; Sharif et al., 2011), and (iv) recharge zone evaluation (Dyer et al., 2015). There is a lack of high-fidelity groundwater models for the MRAA, which can assist proper agriculture water management in the northeastern Louisiana. Nevertheless, the development of a high-fidelity MRAA groundwater model is not trivial given the complicated geological architecture of the alluvial aquifer (Bowling et al., 2005; Vahdat-Aboueshagh and Tsai, 2021). Well logs are the essential data to understand geology of the MRAA. Developing realistic lithofacies architecture of the alluvial aquifer using a large quantity of well logs is of great challenge (Pham and Tsai, 2017). This study demonstrates the complexity of the alluvial aquifer by dealing

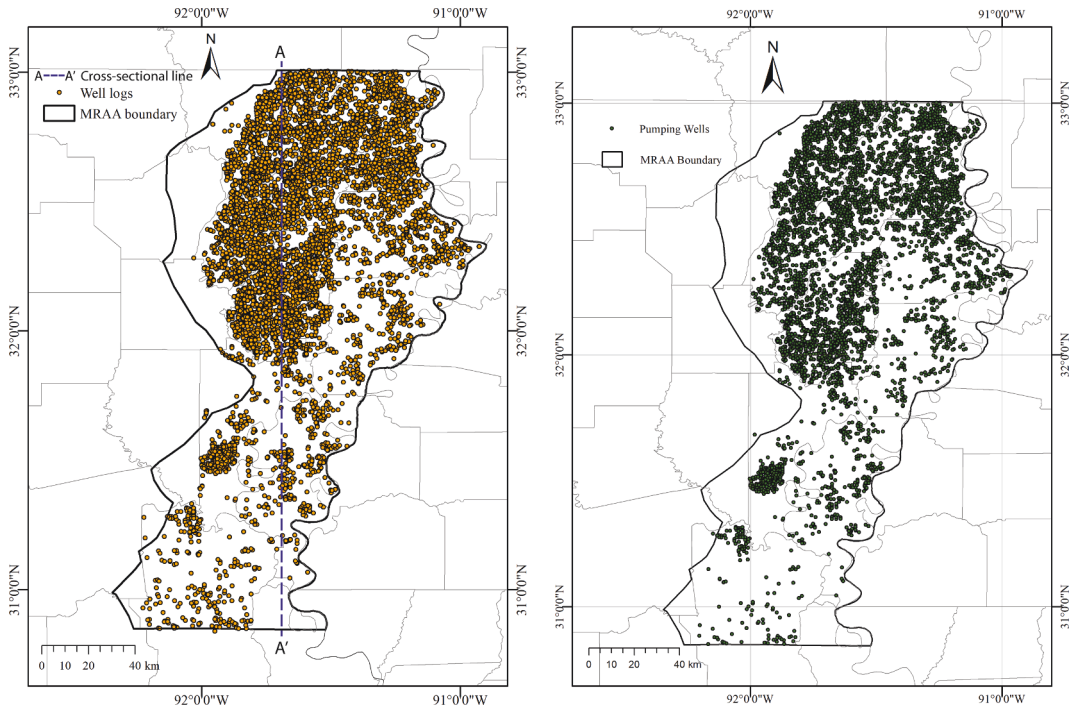

Fig. 2. (a) Well logs used to construct the lithofacies architecture of the MRAA, and (b) map of the pumping wells in the MRAA.

with a huge number of well logs. Besides, the MRAA is hydraulically connected to many streams, especially the Mississippi River, resulting in considerable volumes of water being contributed to or taken from these surface-water bodies. Pumping from the aquifer would induce groundwater recharge from those streams and rivers. Understanding interactions between the Mississippi River and the alluvial aquifer remains in its infancy. Therefore, this study aims to fill in the gap and better understand groundwater dynamics in the alluvial aquifer. This study introduces an uncertainty-based modeling approach to develop a complex three-dimensional groundwater model for the MRAA using a huge amount of data sets including a large volume of well logs.

Considering one conceptual groundwater model may not adequately sample the relevant space of plausible models (Ajami et al., 2007; Rojas et al., 2008). Single model techniques are unable to account for errors in model output resulting from an inadequate choice of model parameter values (Liu and Merwade, 2019; Neuman, 2003). It has been recognized that singling out the best concept model is not sufficient because the best calibrated model does not necessarily guarantee better prediction. As a consequence, a well-calibrated model does not always accurately predict the behavior of the dynamic system. Choosing a single model out of plausible alternative models may contribute to either type I (reject true model) or type II (fail to reject false model) model errors (Li and Tsai, 2009; Neuman, 2003). Hence, different alternative conceptual models of various parameter values for a hydrogeological system have been suggested (Mustafa et al., 2018; Refsgaard et al., 2007; Troldborg et al., 2007). Multi-model approaches can be used to estimate a broader uncertainty band so that it is more likely to include unknown true predicted values. To fill in the gap, this study adopts the Bayesian model averaging (BMA) approach (Hoeting et al., 1999) as a model averaging technique to combine multiple model predictions and account for model uncertainty.

The Bayesian model averaging method (Draper, 1995; Hoeting et al., 1999; Tsai and Elshall, 2013) derives predictions from a set of alternative conceptual models to construct a predictive uncertainty distribution using probabilistic techniques. The weights in the BMA methods are assessed based on the relative performance of each model to reproduce system behavior during the observation period (Draper, 1995; Vrugt and Robinson, 2007). Recently, BMA has received attention of researchers in diverse fields because of its more reliable and accurate predictions than other existing model averaging methods, such as the Generalized Likelihood Uncertainty Estimation (Singh et al., 2010) and information criterion averaging (Liu and Merwade, 2018; Yin and Tsai, 2020; Zhang et al., 2009). The advantage of the BMA approach over other multi-model combing methods is that BMA not only provides a deterministic model weighted average prediction of the interested variable, but also produces the forecast distribution which reflects the uncertainty associated with the deterministic prediction (Mustafa et al., 2018; Raftery et al., 2005). Such a multi-model approach is more likely to include unknown true predicted values. An important challenge in implementing Bayesian model averaging is to estimate posterior model weights and prediction variances. There are different methods including the analytical techniques (Schöniger et al., 2014), the information-theoretic criteria (Schöniger et al., 2014), and the Laplace approximation (Singh et al., 2010). However, the analytical techniques strongly depend on assumptions, and the information-theoretic criteria may provide contradictory results in model weights (Poeter and Anderson, 2005; Ye et al., 2010). In this study, a maximum-likelihood Bayesian model averaging approach was applied to analyze predictive distribution of alternative conceptual models because of its statistical robustness and numerical efficiency. The study introduces a Bayesian multi-model uncertainty quantification (BMMUQ) framework to account for model parameter uncertainty in developing conceptual groundwater flow models.

This work is structured as follows. The study first introduces the MRAA model area. Second, data and procedures for the MRAA groundwater model development are described. Third, the BMMUQ framework is introduced. Fourth, results and discussions are described with respect to the method demonstration to the MRAA area. Lastly, conclusions and limitations of the study are drawn.

2. Study area

The Mississippi River alluvial aquifer (Fig. 1) is the uppermost aquifer underneath the Mississippi Alluvial Plain that stretches from southern Illinois to the Gulf of Mexico. The alluvium of the Mississippi River and its tributaries formed the MRAA. As shown in Fig. 1, the study

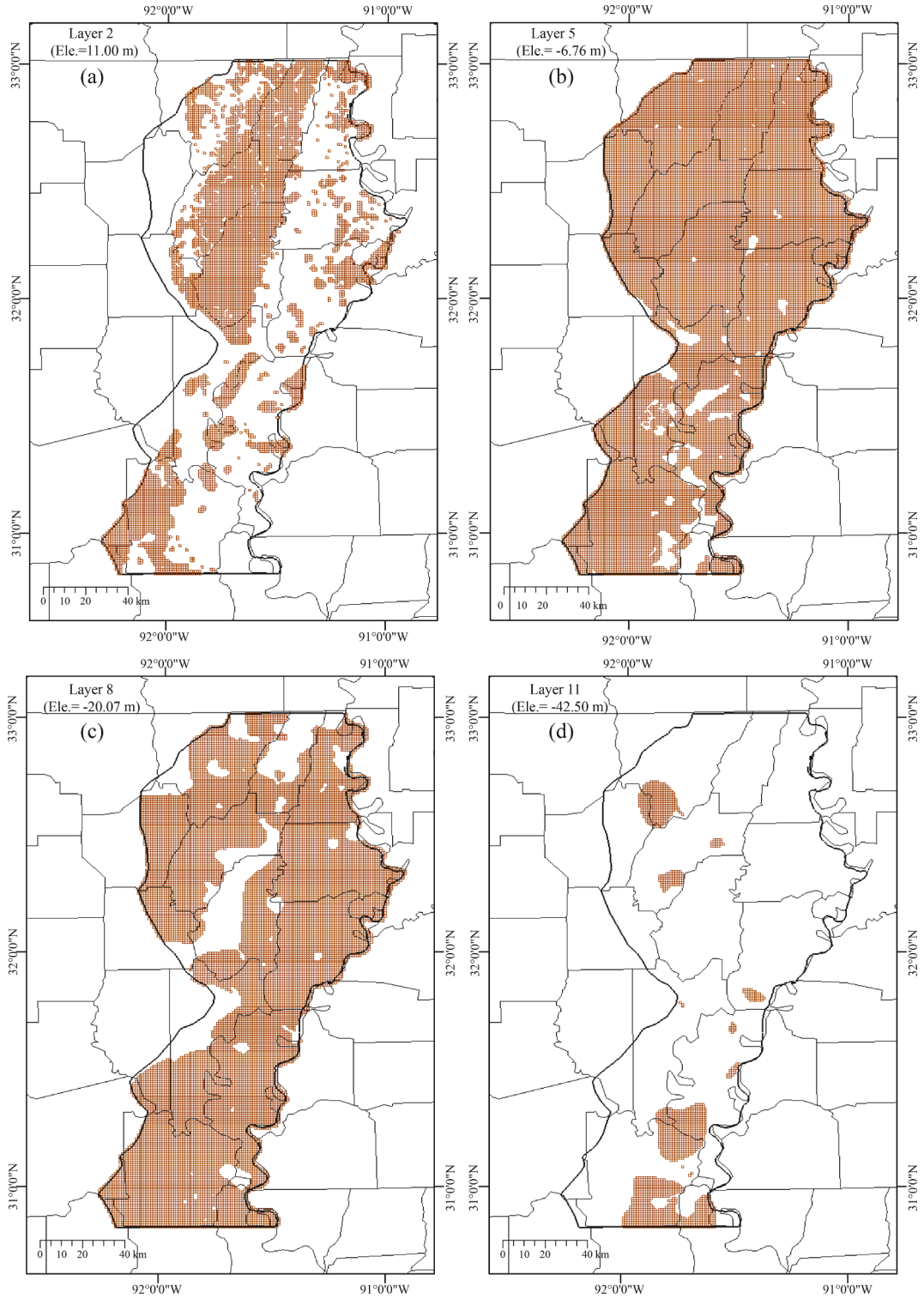

Fig. 3. Sand facies distributions in the MRAA model: (a) layer 2, (b) layer 5, (c) layer 8, and (d) layer 11. The cells represent sands. The elevation represents average bottom elevation (m), NGVD 1929.

area covers approximately 18,000 km^2 and comprises the entire or majority of 16 parishes in Louisiana. The freshwater extent of the MRAA was delineated by the USGS (Smoot, 1986). The topography varies from around 30 m (100 ft) in the north to 7 m (23 ft) above NGVD 29 (National Geodetic Vertical Datum of 1929) in the south. The surficial geology includes Holocene Alluvium, Pleistocene Braided-stream Terraces, and Pleistocene Braided-stream Terraces-Loess. The terraces and terraces-loess in the north-central area belong to Macon Ridge that has relatively high elevation compared to the Alluvium. Sand and gravel of Pleistocene underlies fine sediments: silt and clay of Holocene that generally confine the aquifer (Whitfield, 1975). The MRAA consists of fining upward sequences from surface to bottom: clay, fine sand, medium sand, coarse sand and gravel in a sequence (Carlson, 2006).

The model domain is hydraulically connected with its major streams and with the Mississippi River as the eastern boundary, and is adjacent to Quaternary-Tertiary deposits at the west. Major rivers and bayous are

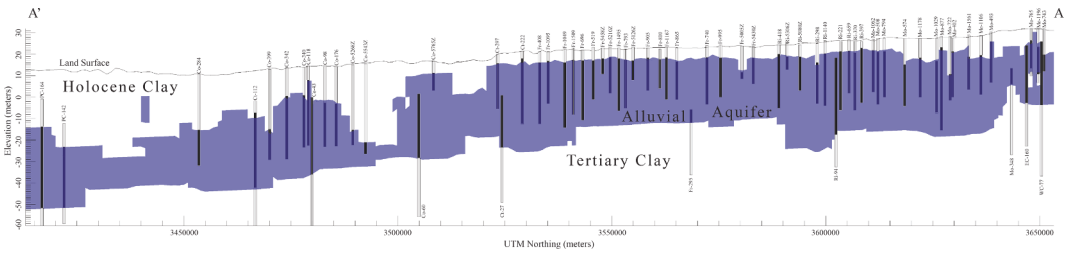

Fig. 4. Modeled sand and clay facies distribution in AA' cross section (see Fig. 2a for location) along with nearby well logs.

included in the developed groundwater model, such as the Mississippi River, Atchafalaya River, Red River, Black River, etc. As a primary water source to agriculture, the MRAA provides about 795,000 m^3/day of groundwater (Collier and Sargent, 2018) for crops and has significant contributions to Louisiana's economy. The study investigates the impact of irrigation pumping on groundwater dynamics through groundwater modeling.

3. MRAA groundwater model development

3.1. Well logs

Well logs are the essential source to understand the geology of the MRAA and construct lithofacies architecture of the MRAA. To faithfully represent the geological architecture of MRAA, a total of 7,000 good-quality drillers' logs in the study area were collected from the well registration record in the Louisiana Department of Natural Resources (LDNR). The large amount of well logs was attributed to extensive drilling of irrigation wells over the past decades. As shown in Fig. 2(a), the distribution of well logs indicates extensive irrigation activities in the central and north of the MRAA. Fig. 2(b) shows the 5,565 pumping wells during 2004–2015. Comparing to irrigation wells that are normally shallow and do not penetrate the alluvial aquifer, drillers' logs contain valuable lithological information from land surface to the deep bottom at the drilling sites. As processing lithological descriptions in drillers' logs is time-consuming and subjected to drillers' interpretations, this study follows methods from Elshall et al. (2013) to categorize lithological descriptions into sand facies and clay facies.

Besides the drillers' logs, 259 wireline electrical logs were also collected from the LDNR. The electrical logs generally penetrate the alluvial aquifer and provide deeper records. For the sediments in Louisiana, short-normal electrical resistivity of 20 ohm-m is a good threshold to identify sand facies (Elshall et al., 2013). Spontaneous potential was used to identify sand facies when saline water is present. The electrical logs were hence used to estimate the aquifer thickness. The maximum depth of the drillers' logs can reach around 60 m, while some electrical logs can reach more than 300 m.

3.2. Lithofacies modeling and MODFLOW grid generation

The study area was discretized into 18,045 horizontal cells with cell size of 1 km by 1 km. According to the topography, a south dip angle of 0.0049° was applied to developing a lithofacies model. This study adopted the indicator modeling technique by Pham and Tsai (2017) to construct an MRAA lithofacies model by interpolating the 7,259 well log data using the natural neighbor interpolation method (Tsai et al., 2005). The full details of the lithofacies modeling technique can be referred to Pham and Tsai (2017). The MODFLOW (Harbaugh, 2005) computing grids with 12 model layers were then generated to represent the complex sand facies structures of MRAA (shown in Fig. 3). Clay facies formed inactive cells in MODFLOW. The bottom elevation of the model in the south was around −58 m (-190 ft) such that the entire MRAA is included. The MRAA structural complexity can also be illustrated in a north–south cross section shown in Fig. 4. MRAA thickness is highly non-uniform. Holocene clay is the relatively thinner in the north and thicker in the south.

The total number of computational cells is 104,971. The layer thickness varies between 1 m and 13 m. Sand distributions in all the 12 model layers can be found in supporting Fig. S1. The MRAA is generally confined by clay deposits of varying thickness and extent at the land surface and at the base. The major outcrops are located in the west-central and northern areas. Sand thickness ranges from 18 m to 46 m with an average thickness around 30 m. The averaged top clay thickness is around 17 m. The existence of the bottom clay layer shows the separation of the MRAA from other aquifers below. The finer clay, silt and sand sediments generally occur in the upper part of the alluvium, but these sediments can occur to varying degrees throughout the entire thickness of the alluvium. The outcrops provide potential rainfall and river recharge corridors to the MRAA. Based on the bathymetric data of Mississippi River at several locations that are available from the geological investigations in the Lower Mississippi Valley (Saucier, 1967), the lithofacies model shows strong hydraulic connection between the MRAA and Mississippi River at the east boundary of the model.

3.3. Groundwater use data compilation

According to the well registration record from the LDNR, the 5,565 pumping wells (see Fig. 2b) during 2004–2015 in the study area includes 5,536 active irrigation wells and 29 public and industrial wells. Monthly irrigation pumping data were provided by the Louisiana State University, Agricultural Center. The irrigation season was from May to September with peak pumpage in August. Monthly pumping data for public and industrial wells were provided by the USGS Lower Mississippi-Gulf Water Science Center. Monthly pumping data were compiled to support MRAA groundwater modeling.

3.4. Groundwater recharge estimation

Groundwater recharge was estimated as a fraction of baseflow from a hydrologic model, the Variable Infiltration Capacity (VIC) model, which was set up by Oubeidillah et al. (2014) and Naz et al. (2016) that covers the MRAA study area. The VIC model is a semi-distributed macroscale hydrological model which solves the full water and energy balances (i.e., evapotranspiration, snowpack, runoff, and baseflow) independently at each specified grid cell within a watershed (Liang et al., 1994). The VIC model output used in this study was driven by the 1980–2015 Daymet meteorologic forcings (Thornton et al., 1997) at 1/24° (~4 km) grid resolution. As reported by Oubeidillah et al. (2014), VIC model has overall satisfactory performance in the MRAA region with the Nash-Sutcliffe efficiency (NSE) of monthly total runoff greater than 0.5 in most of the hydrologic subbasins. Further modeling details can be referred to Oubeidillah et al. (2014).

Since the VIC model does not explicitly model the process of groundwater recharge, the amount of water flux that may contribute to surficial recharge should still be a part of its baseflow component. This study hence assumed that a portion of the baseflow may contribute to the surficial recharge at the MRAA outcrop sand cells. The ratio between baseflow and surficial recharge was further estimated during model

calibration. This study assumed a constant ratio for the simplicity purpose and to reduce overparameterization during model calibration.

3.5. Mississippi River stage for eastern boundary

Mississippi River is located at the eastern boundary of the MRAA study area. The boundary shape was determined by the Mississippi River flowline taken from the National Hydrography Dataset Plus (NHDPlus) Dataset (Moore et al., 2019). NHDPlus integrates the National Hydrography Dataset (NHD), National Elevation Dataset (NED), National Watershed Boundary Dataset (WBD) and other national geospatial datasets to provide in-depth information of river channels and network (such as connectivity, channel elevation, slope, cumulative drainage area, etc.). Mississippi River was treated as a general head boundary (GHB) condition within the MODFLOW model, where river stage and hydraulic conductance were needed for sand cells that contact the river. Readers are referred to Harbaugh (2005) for the definition of hydraulic conductance in the GHB condition. Mississippi River stages in 2004–2015 at six (6) stream gauges were collected from the U.S. Army Corp of Engineers river gage network. Four (4) stations (Vicksburg, Natchez, Knox Landing, and Red River Landing) are in the study area (Fig. 1) while two (2) stations (Greenville and St. Francisville-South) are outside of the study area. Monthly river stages for all cells on the Mississippi River flowline were derived by interpolating the gage data (see supporting Fig. S2). The river stage is normally above flood stage at Vicksburg between April and June and at Red River Landing between May and June. The hydraulic conductance of the Mississippi River was determined through the model calibration process.

3.6. Input of major streams

Ten (10) major streams including Atchafalaya River, Tensas River, Red River, Black, Ouachita River, Little River, Boeuf Rivers, Bayou Macon, Bayou Bonne Ide and Bayou Bartholomew were involved in the developed groundwater model. The NHDPlus flowlines of these major streams are shown in Fig. 1. The streams were considered as a head-dependent flux boundary condition through the River Package in the MODFLOW model. USGS streamflow stations (Fig. 1) were used to estimate stream stages for these rivers. The bottom elevations of these rivers were determined from the available geologic maps and cross sections (Fleetwood, 1969; Saucier, 1967). The hydraulic conductance of the streams was estimated through the model calibration process. Readers are referred to Harbaugh (2005) for the definition of river hydraulic conductance in the River Package.

3.7. Initial head and Non-River boundary conditions

The initial groundwater head distribution on January 1, 2004 was estimated by interpolating groundwater level data from the USGS observation wells (Fig. 1) that are in or around the study area and screened at the MRAA. Vertical variation of groundwater head is neglected for the initial head distribution. It is understood that the MRAA extends northward and southward beyond the modeling domain by the Mississippi River changing course and contacts Quaternary-Tertiary aquifers in the west. Therefore, the northern, southern and western non-river boundaries of the MRAA model are considered as time-variant specified-head boundaries. Monthly boundary heads in 2004–2015 were derived by interpolating 826 groundwater level time series data from 25 USGS observation wells.

3.8. Model calibration

In this study, model parameters to be calibrated were hydraulic conductivity, specific storage, specific yield, hydraulic conductance of the Mississippi River, hydraulic conductance of the major streams, and a fraction of the VIC-modeled baseflow for groundwater recharge. A genetic algorithm (Cao and Wu, 1999) was adopted to estimate the model parameters by minimizing the root mean square error (RMSE) between the calculated and observed groundwater heads:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(h_i - h_i^{obs}\right)^2} \tag{1}$$

where N is the number of groundwater head data (N = 826 in this study). h_i is the simulated groundwater head and h_i^{obs} is the observed groundwater head.

4. Bayesian Multi-model uncertainty quantification

Subjected to uncertainty arising from model parameters, singling out the best concept model is not sufficient to account for errors in model outputs because the best calibrated model does not necessarily guarantee better prediction. Instead of relying on a standalone groundwater flow model, extracting information from an ensemble of groundwater models is more reliable. The study presents a Bayesian multi-model uncertainty quantification (BMMUQ) framework to account for model parameter uncertainty and produce reliable groundwater level predictions in the MRAA. The BMA predictive mean ($\bar{h}$) is expressed as follows:

$$\bar{h} = \sum_{k=1}^{K}\omega_k h_k \tag{2}$$

where h_k is the mean groundwater head simulated using the k^{th} groundwater flow model. K is the number of groundwater flow models in the ensemble members (K = 50 in this study); and ω_k is the model weight for the k^{th} model.

4.1. Bayesian model averaging

Bayesian model averaging (BMA) is a probabilistic scheme for combining predictions from multiple models (Hoeting et al., 1999) to provide a more reliable description of total prediction uncertainty. The BMA predictive probability density function (PDF) (Raftery et al., 2005) of a variable, such as groundwater head in this study, is the weighted average of the PDFs associated with each model member forecast in the ensemble. In case of a groundwater level h to be predicted on the basis of training data h_T using K groundwater models, the BMA predictive PDF is:

$$p(h|h_T) = \sum_{k=1}^{K} p(M_k|h_T)\, g_k(h|M_k) \tag{3}$$

where M_k is the k^{th} groundwater flow model used to simulate groundwater head; $p(h|h_T)$ is the BMA posterior probability to predict the quantity h on the basis of training data h_T using all groundwater models $(M_1,\ldots,M_K)$; $g_k(h|M_k)$ is the conditional PDF which can be interpreted as the conditional PDF of h conditioned on M_k, given that M_k is the best forecast model in the ensemble; and $p(M_k|h_T)$ is the posterior probability of model M_k given the training data h_T, which is based on model performance in the training period (T). In the BMA model, the posterior model probability $p(M_k|h_T)$ represents the weight for each model member in the ensemble, which is ω_k in equation (2). All weights add up to one, i.e., $\sum_k p(M_k|h_T) = \sum_k \omega_k = 1$. Estimation of the weight $p(M_k|h_T)$ is described in the next subsection. Therefore, the BMA predictive mean is:

$$\bar{h} = \sum_{k=1}^{K} p(M_k|h_T)\, h_k \tag{4}$$

where $\bar{h}$ is the BMA predictive mean of the groundwater head. The BMA

method considers the uncertainty of each model's forecast and is used to develop a predictive distribution rather than only a weighted average. So, the BMA method provides an average forecast along with an associated posterior distribution. The forecast distribution can be used for developing predictive confidence intervals. Based on the law of total variance, the BMA predictive variance $\mathrm{var}(h|h_T)$ of groundwater head is as follows:

$$\mathrm{var}(h|h_T) = \sum_{k=1}^{K} \mathrm{var}(h_k|h_T)p(M_k|h_T) + \sum_{k=1}^{K} (h-h_k)^2 p(M_k|h_T) \quad (5)$$

where $var(h_k|h_T)$ represents the variance of h_k conditioned on training data h_T under the conceptual groundwater flow models, $M_1, \ldots, M_K$. The first term in the right-hand side of Eq. (5) represents the within-model variance, while the second term represents the between-model variance.

4.2. Expectation Maximization (EM) method

Successful implementation of the BMA method requires estimates of the model weights, and prediction variances, of the conditional PDFs of the ensemble members. The values are estimated by maximum likelihood on the basis of training dataset. Typically, it is assumed that the PDF of predicted quantity approximates a Gaussian distribution. In case of non-Gaussian data, Box-Cox power transformation is used to map the variables from their original space into Gaussian space (Box and Cox, 1964). In this study, the observed and simulated groundwater head data were normalized using Eq. (6).

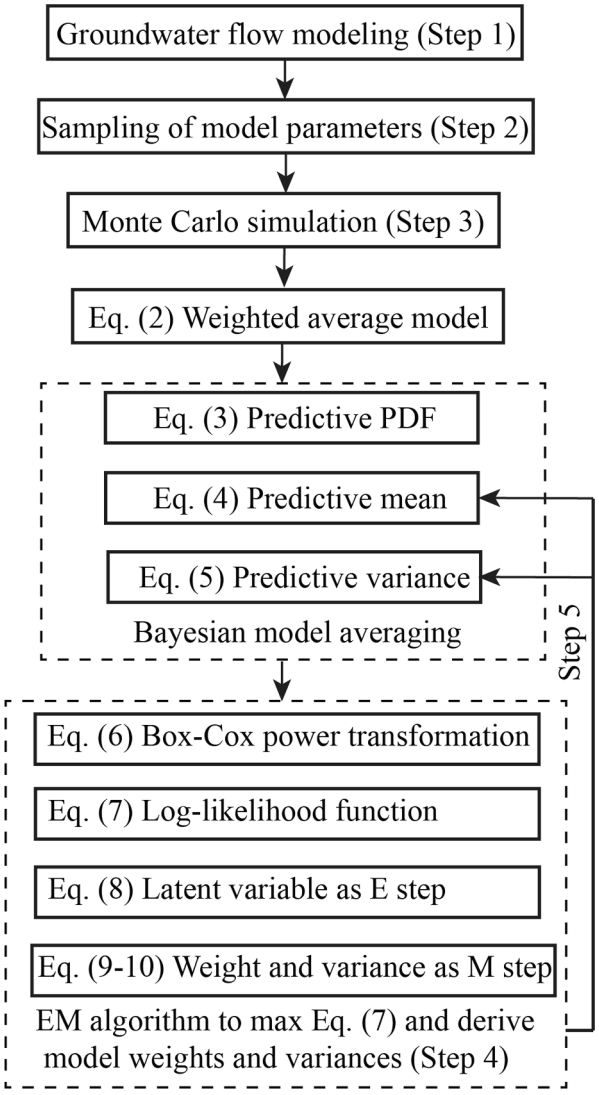

Fig. 5. The flowchart of Bayesian multi-model uncertainty quantification (BMMUQ) framework.

$$h(\lambda) = \begin{cases} \dfrac{h^{\lambda}-1}{\lambda}, \lambda \neq 0 \\ \ln(h), \lambda = 0 \end{cases} \quad (6)$$

where λ is the power that used to transform the data to a normal distribution, which varies from −5 to 5. Different powers were tested and the best power that yielded the closest normal distribution of transformed groundwater head data was found and applied. In order to estimate model weights and groundwater head standard deviations, on the basis of training data, the log-likelihood function (7) below is maximized.

$$L(\boldsymbol{\omega}, \boldsymbol{\sigma}^2) = \sum_{l,t} \ln\left(\sum_{k=1}^{K} \omega_k g_k(h_{lt}|M_{k,lt})\right) \quad (7)$$

where the summation is over locations (l) and time (t) in the training data set. The weight $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_K\}$, the variance $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \ldots, \sigma_K^2\}$. Generally, the independence assumption is unlikely to hold, but is not expected to have significant impact on estimates because Eq. (7) calculates the conditional distribution for a scalar observation given forecasts, rather than for several observations simultaneously (Raftery et al., 2005). There are no closed-form solutions that maximize Eq. (7) analytically (Vrugt and Robinson, 2007). Expectation-maximization (EM) algorithm is used to find the maximum likelihood for BMA model training. The EM method is iterative and alternates between two steps, the E (or expectation) step and the M (or maximization) step by using a latent variable z. In the E step, z is estimated given the current values of the BMA weights and variances. The E step is:

$$\hat{z}_{k,lt}^{(j)} = \frac{\omega_k^{(j-1)} N(h_{lt}|E_{k,lt}, \sigma_k^{(j-1)})}{\sum_{k=1}^{K} \omega_k^{(j-1)} N(h_{lt}|E_{k,lt}, \sigma_k^{(j-1)})} \quad (8)$$

where the superscript j refers to the j^{th} iteration of the EM algorithm, and the mean value $E_{k,lt}$ is the corresponding prediction from k^{th} groundwater flow model, which is same with h_k in equation (2). The initial value ω_k^0 is 1/K, and the initial value σ_k^0 is $\sqrt{h_o}$, where h_o is the transformed groundwater head observation. $N(h_{lt}|E_{k,lt}, \sigma_k^{(j-1)})$ represents a normal density with mean $E_{k,lt}$ and standard deviation $\sigma_k^{(j-1)}$. The M step then consists of estimating the weight ω_k and standard deviation σ_k with the current estimation of $\hat{z}_{k,lt}$. Thus,

$$\omega_k^{(j)} = \frac{1}{n} \sum_{l,t} \hat{z}_{k,lt}^{(j)} \quad (9)$$

$$\sigma_k^{2(j)} = \frac{\sum_{l,t} \hat{z}_{k,lt}^{(j)} (h_{lt} - E_{k,lt})^2}{\sum_{l,t} \hat{z}_{k,lt}^{(j)}}, \quad (10)$$

where n is the number of observations in the training data set. h_{lt} is an observed value. By alternating between the E step and the M step, convergence is achieved when the changes of the latent variables $\hat{z} = \{\hat{z}_{k,lt}\}$, the weights $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_K\}$, the variances $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \ldots, \sigma_K^2\}$, and the log-likelihood are smaller than a tolerance. The study adopted a tolerance of 10^{-6}. The log likelihood is guaranteed to increase at each EM iteration (Wu 1983) and reaches the maximum of the likelihood.

4.3. Multi-model generation and data description

Selection of alternative groundwater models is an important aspect of Bayesian model averaging. We understand that many sources, such as

initial conditions, boundary conditions, geological architectures and model parameters (Li and Tsai, 2009; Mustafa et al., 2018), can contribute uncertainty to groundwater models. Nevertheless, this study only considered hydraulic conductivity and specific storage of sand facies for methodology demonstration. A pre-specified fifty (50) alternative groundwater flow models were generated through combination of model parameters which follow multivariate Gaussian distribution. The mean values of model parameters (hydraulic conductivity and specific storage of sand facies) were from the calibrated model based on maximum-likelihood estimation. This study adopted the linear statistical approach (Bard, 1974) to determine the covariance matrix of model parameters as follows:

$$\mathrm{Cov} = \frac{1}{N-F}\sum_{i=1}^{N}\left[h_i - h_i^{obs}\right]^2 [J^T J]^{-1}, \quad (11)$$

where Cov is the covariance matrix of model parameters. J is the N × F Jacobian matrix of sensitivities of N data points with respect to F model parameters.

The BMMUQ framework is implemented as the following steps:

1. Constructing a lithofacies model in the MRAA area which covers three deposits (alluvium, braided-stream terrace, and braided-stream terrace-loess). A 12-layer MODFLOW model was then developed and well calibrated through a genetic algorithm.

2. Generating 50 samples of model parameters from multivariate normal distribution with mean vector and a covariance matrix. The mean values of hydraulic conductivity and specific storage of sands were from the calibrated MODFLOW model. The covariance matrix was obtained by calculating the Jacobian matrix in equation (11) (Bard, 1974).

3. Conducting Monte Carlo simulations (MCS). 50 sets of simulated groundwater level outputs were obtained through running MODFLOW-2005 models using the sampled parameter values in Step 1.

4. Deriving model weights and variances. The posterior model weights and groundwater head variances of each ensemble model member were calculated after convergence using the EM algorithm based on training data.

5. Obtaining an average forecast along with an associated forecast distribution. Finally, multi-model predictions were obtained by assessing the BMA predictive mean and variance using Eq. (4) and Eq. (5).

The BMMUQ framework is illustrated in Fig. 5.

Through conducting Monte Carlo simulations, 50 sets of simulated groundwater level outputs were obtained through running groundwater models using the sampled parameter values. Groundwater level data from the USGS observation wells along with simulated groundwater levels from the 50 alternative models comprise the training and testing datasets. 544 groundwater level data (~75% of total) in this study were used for training, and independent 189 groundwater level data (~25%) at four USGS observation wells (Co-215, Ma-64, Fr-1092 and Ri-124) were used for testing. Training data were used in the study to calculate model weights and groundwater head variances of 50 model members. Independent testing data were used to validate model performance.

5. Results and discussions

5.1. Model parameters and calibration

The simulation time of the MRAA groundwater model was 12 years from 2004 to 2015 with 144 monthly stress periods. The USGS groundwater level data from USGS observation wells exhibit two distinct patterns of groundwater levels. Groundwater levels in the Alluvium shows much greater variation compared to those in Braided-stream Terraces, indicating that groundwater in the Alluvium response much fast to nearby rivers. The USGS groundwater data also show that groundwater level behavior in the Braided-stream Terraces-Loess is different from that in the Braided-stream Terraces although groundwater level variations are low in both formations. Therefore, hydraulic conductivity, specific storage, specific yield of the MRAA were parameterized into three zones: Alluvium, Braided-stream Terraces, and Braided-stream Terraces-Loess. Other parameters are considered lumped values. Totally, there are 12 parameters to be estimated (Table 1).

Table 1
Calibrated model parameters for the Mississippi River Alluvial Aquifer.

Parameter	Range[a]	Estimate
Hydraulic conductivity for Alluvium (m/day)	5–300	279.5
Hydraulic conductivity for Terrace (m/day)	5–300	76.9
Hydraulic conductivity for Terrace-Loess (m/day)	5–300	95.5
Specific storage for Zone Alluvium (1/m)	10^{-5}–10^{-3}	9.605×10^{-5}
Specific storage for Zone Terrace (1/m)	10^{-5}–10^{-3}	8.735×10^{-5}
Specific storage for Zone Terrace-Loess (1/m)	10^{-5}–10^{-3}	3.033×10^{-5}
Specific yield for Zone Alluvium	0.15–0.45	0.4379
Specific yield for Zone Terrace	0.15–0.45	0.3796
Specific yield for Zone Terrace-Loess	0.15–0.45	0.4457
Mississippi River conductance (m^2/day)	–	883
Major stream conductance per unit length (m^2/day/m)	–	3.564
Surficial recharge fraction	–	0.0532

[a] Range refers to Freeze and Cherry (1979).

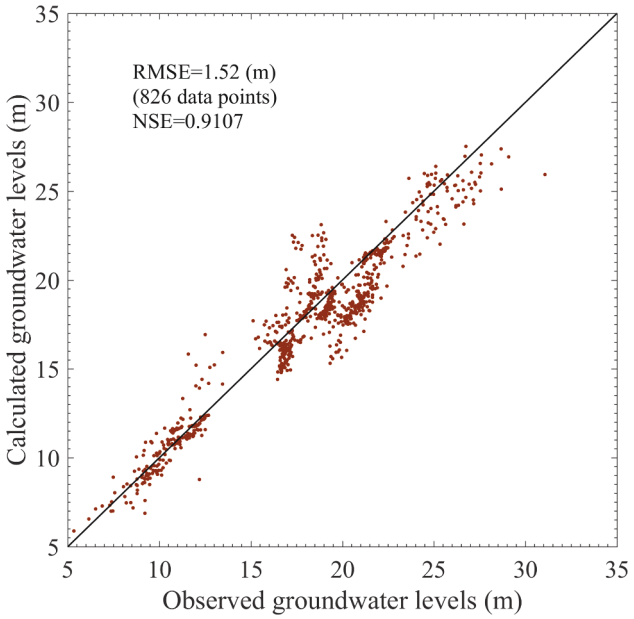

Fig. 6. Scatter plot for observed groundwater level vs. simulated groundwater level.

A genetic algorithm (Cao and Wu, 1999) was adopted to estimate the 12 model parameters by minimizing the root mean square error (RMSE) of simulated 826 groundwater level at USGS observation wells from January 2004 through December 2015. The estimated parameters are listed in Table 1. Fig. 6 shows the scatter plot of calculated vs. observed groundwater heads. The calculated RMSE is 1.52 m and the overall NSE is 0.91, which is considered a good calibration result to the observed USGS groundwater levels. Hydraulic conductivity of Alluvium was estimated around three to four times higher than Braided-stream Terraces and Braided-stream Terraces-Loess, as shown in Table 1. The study estimated that about 5.32% of the VIC-modeled baseflow may recharge MRAA. Rainfall-based groundwater recharge was only given to the outcrop sand cells. As a result, surficial recharge rates were estimated up to 0.55 mm/day (see supporting Fig. S3). Surficial recharge is insignificant to the MRAA as a source of water to the aquifer. In addition, water budget analysis through ZONEBUDGET (Harbaugh, 1990) shows that the aquifer started to lose groundwater since August 2005, with an estimate of 950 million m^3 of groundwater storage loss in 2015 relative to

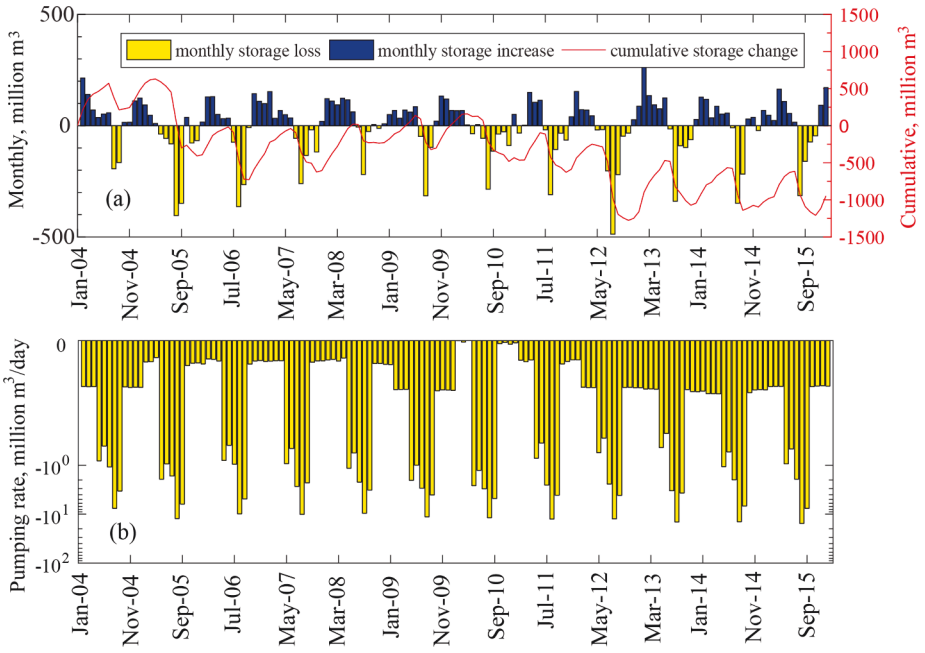

Fig. 7. (a) Estimated monthly storage change (million m^3) and cumulative groundwater storage change (million m^3) with respect to the beginning groundwater storage of 2004, and (b) estimated monthly pumping rate (million m^3/day).

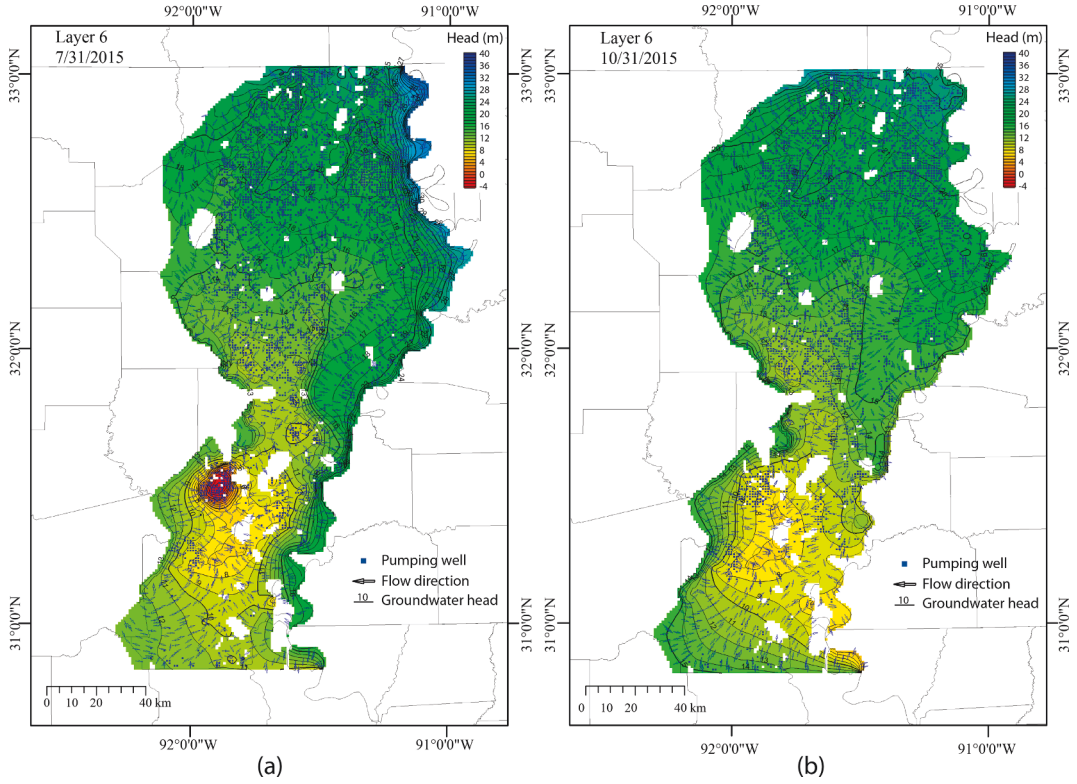

Fig. 8. Simulated groundwater level distribution and velocity field in the MRAA area: (a) July 31, 2015 (irrigation season), and (b) October 31, 2015 (post-irrigation season).

the beginning of 2004, as shown in Fig. 7. The storage variation exhibits a periodic pattern that is strongly corresponding to seasonal groundwater pumping, where groundwater loss mainly occurred from May to September each year. This is mainly attributed to high groundwater demand for crop irrigation in the MRAA area. The groundwater head distribution and velocity field of the MRAA area for the irrigation season (July 2015) and post-irrigation season (October 2015) are shown in Fig. 8. The model found a noticeable cone of depression occurs in the southwestern area during the irrigation season, where dense wells are heavily pumped for irrigation. The aquifer is then replenished in the fall.

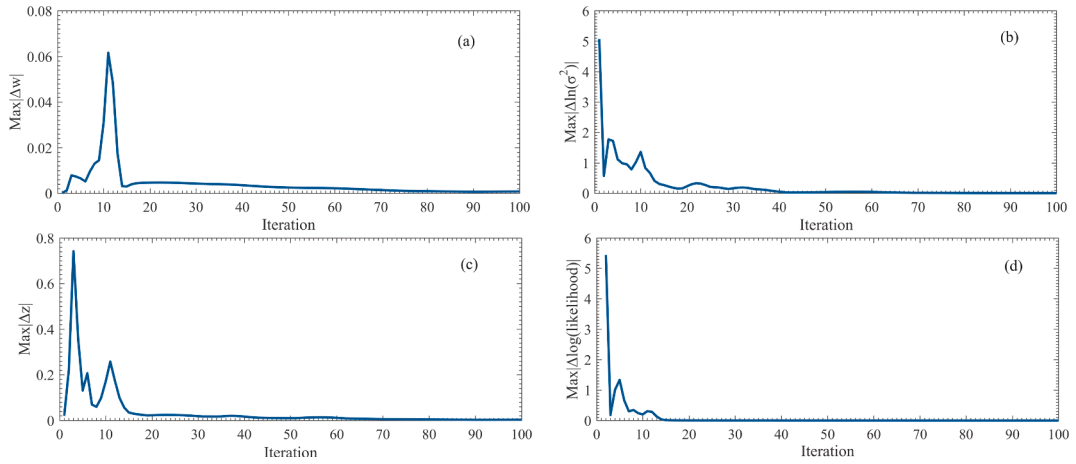

Fig. 9. Convergence of four indicators after EM iterations during the training period: (a) maximum absolute change of weight ($|\Delta\omega|$), (b) maximum absolute change of logarithmic variance ($|\Delta ln(\sigma^2)|$), (c) maximum absolute change of latent variable ($|\Delta z|$), and (d) maximum absolute change of logarithmic likelihood ($|\Delta log(likelihood)|$).

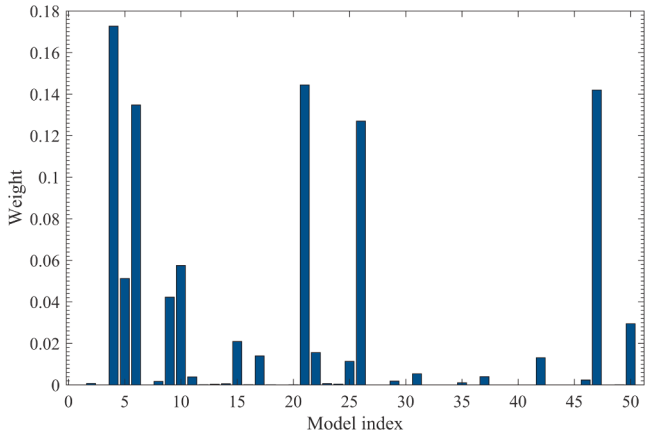

Fig. 10. BMA model weights for each model configuration at convergence.

5.2. Convergence of Expectation Maximization algorithm and BMA weights

After model calibration, 50 models were generated by sampling model parameters. The Expectation Maximization algorithm alternated between the E step in equation (8) and the M step in equations (9) and (10) and iteratively estimated model weights and prediction variances. As shown in Fig. 9, the four indicators converged after 100 iterations, including the maximum absolute change of model weight $|\Delta\omega|$ in Fig. 9a, maximum absolute change of logarithmic variance $|\Delta ln(\sigma^2)|$ in Fig. 9b where variance measures how far each groundwater level in the set is from the mean, maximum absolute change of latent variable $|\Delta z|$ in Fig. 9c, and maximum absolute change of logarithmic likelihood $|\Delta ln(likelihood)|$in Fig. 9d where likelihood represents the probablity that predictions are close to observations. During the process, 544 groundwater level data from 15 USGS observation wells were used to calculate posterior model weights and variances. Fig. 10 shows model weights of the fifty (50) alternative groundwater flow models at convergence obtained using equations (8)-(10). It shows that ten groundwater models stand out and produce their best predictions with relatively higher weights (greater than 0.02, the equal weight of 50 model members), while some model members have lower weights due to relatively poor model performance.

5.3. Validation of BMA prediction performance and evaluation of uncertainty

In this study, independent groundwater level data from four observation wells (Co-215, Ma-64, Fr-1092 and Ri-124) that are well distributed in the domain were used for model performance validation. The performance difference in these models is shown in Fig. 11 by groundwater level RMSE. The BMA model predict better than ensemble mean, except at well Co-215. The BMMUQ framework has decreased the RMSE of head prediction compared to most model members and ensemble mean predictions. Besides, BMA predictions are generally found closer to those best model predictions whereas ensemble mean predictions are generally positioned in the middle of model members, as is shown in Fig. 11 Additionally, the difference of RMSEs among these predictions varies from one well to another. Overall, all the models perform better at wells Co-215, Fr-1092 and Ri-124, with the RMSE less than 0.8 m, while the RMSE is slightly worse at well Ma-64. This is because there is a wider range of training data from well Ma-64 compared to other well sites. From the perspective, BMA model is more reliable and trustworthy when used for predictions at locations other than the four validation well sites because BMA models perform consistently well.

In addition, the BMA method provides a forecast distribution that can be used for constructing confidence intervals using the BMA predictive mean in equation (4) and BMA predictive variance in equation (5). Fig. 12 shows the 95% confidence interval (CI) of BMA groundwater level predictions. It can be seen that the 95% CI of the BMA predictions covers most of the observations. This is an indication of the improved model predictions and accuracy of the uncertainty bounds. The bandwidth (space between upper interval and lower interval) reflects uncertainty and it is dynamic across all time periods. The results from Fig. 12 reveal that an explicit consideration of conceptual model uncertainty is necessary to improve accuracy of uncertainty bounds.

6. Conclusions

The study presents a Bayesian multi-model uncertainty quantification framework to explicitly account for uncertainty originating from errors in model parameter values of a distributed 12-layer MODFLOW model for the Mississippi River alluvial aquifer, Northeast Louisiana. The study focuses on two objectives: groundwater modeling and uncertainty quantification.

A facies model was firstly constructed based on a large number of well logs and confirmed complexity of the MRAA, which encompasses

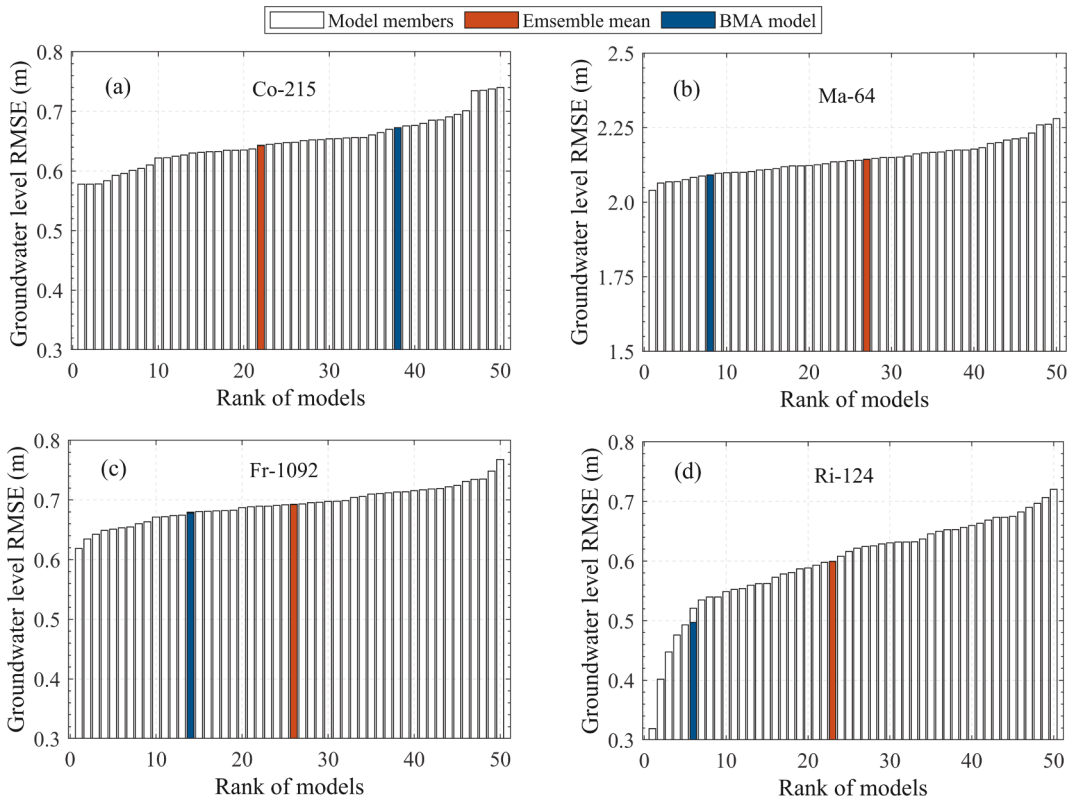

Fig. 11. Rank of RMSEs of groundwater levels from model members, BMA model and ensemble mean with respect to observed groundwater levels at four USGS observation wells for validation.

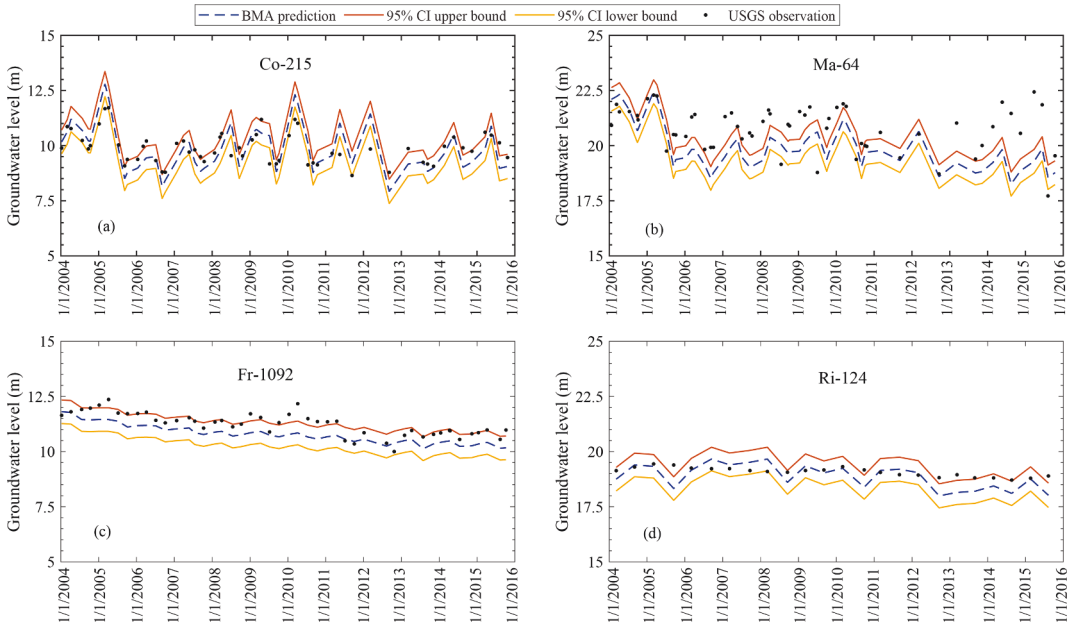

Fig. 12. Prediction uncertainty of groundwater level at each observation well under 95% confidence interval.

Pleistocene terraced braided-stream deposits and Holocene alluvium. A distinct confining layer was found underneath the base of MRAA, indicating that the alluvial aquifer has limited connection with its underlying aquifers. To better understand groundwater dynamics in the agricultural area, a high-fidelity conceptual MODFLOW model was then developed and well calibrated using a genetic algorithm. The study found that parameterizing the model domain into the three zones was an effective approach. The groundwater model was able to demonstrate low groundwater level variability in the terraces and high groundwater level variability in the alluvium. An estimated 950 million m^3 of groundwater storage loss occurred in 2015 relative to the beginning of 2004 due to high groundwater demand for irrigation in the MRAA area.

This study further proposed a multi-model approach to quantify model prediction uncertainty through Bayesian model averaging. Posterior model weights of fifty alternative groundwater flow models were derived to construct a BMA weighed ensemble model for groundwater

level prediction. The result confirms that the BMA ensemble model has decreased the RMSE of head prediction compared to most model members. The BMMUQ framework is a useful tool to have better and more reliable groundwater level predictions, which is important for decision support applications. The framework is highly flexible to be implemented as there is no limitation to the number and complexity of alternative conceptual models. Also, future studies can consider other sources of uncertainty beyond the current scope of the study. However, the number and complexity of alternative conceptual models should be considered based on the modelling objective during implementation of the integrated uncertainty assessment approach to avoid expensive computational burden.

CRediT authorship contribution statement

Jina Yin: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft. **Frank T.-C. Tsai:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - original draft. **Shih-Chieh Kao:** Methodology, Resources, Supervision, Validation, Visualization, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported in part by the U.S. Geological Survey under Grant/Cooperative Agreement No. G16AP00056 and the U.S. National Science Foundation (Award No. 2019561). The authors thank the USGS for providing the industrial and public supply pumping data and Krishna Paudel of LSU Agricultural Center for providing agricultural pumping data. Louisiana Department of Natural Resources is acknowledged for granting access to well registration records.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2021.126682.

References

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Research 43 (1). https://doi.org/10.1029/2005WR004745.

Arthur, J.K., 2001. Hydrogeology, model description, and flow analysis of the Mississippi River alluvial aquifer in northwestern Mississippi. Water-Resources Investigations Report 2001–4035. https://doi.org/10.3133/wri014035.

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press, New York.

Bowling, J.C., Rodriguez, A.B., Harry, D.L., Zheng, C., 2005. Delineating alluvial aquifer heterogeneity using resistivity and GPR data. Groundwater 43 (6), 890–903. https://doi.org/10.1111/j.1745-6584.2005.00103.x.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological) 26 (2), 211–243. https://doi.org/10.1111/rssb:1964.26.issue-210.1111/j.2517-6161.1964.tb00553.x.

Cao, Y.J., Wu, Q.H., 1999. Teaching Genetic Algorithm Using Matlab. The International Journal of Electrical Engineering & Education 36 (2), 139–153. https://doi.org/10.7227/IJEEE.36.2.4.

Carlson, D., 2006. Systematic variability of hydraulic conductivity within the Mississippi River alluvial aquifer in northeastern Louisiana. Transactions - Gulf Coast Association of Geological Societies 56, 121–136.

Clark, B. R., D. A. Westerman, and D. T. Fugitt., 2013. Enhancements to the Mississippi Embayment Regional Aquifer Study (MERAS) groundwater-flow model and simulations of sustainable water-level scenarios. U.S. Geological Survey Scientific Investigations Report: 2013–5161, 29 p. https://pubs.usgs.gov/sir/2013/5161/.

Collier, A., Sargent, B. P., 2018. Water use in Louisiana, 2015. Water Resources Special Report 18. Louisiana Department of Transportation and Development, Baton Rouge, LA, p. 138.

Czarnecki, J.B., Clark, B.R., Stanton, G.P., 2003. Conjunctive-use optimization model of the Mississippi River Valley alluvial aquifer of Southeastern Arkansas. U.S. Geological Survey Water-Resources Investigations Report 03–4230. https://doi.org/10.3133/wri034230.

Dieter, C. A., M. A. Maupin, R. R. Caldwell, M. A. Harris, T. I. Ivahnenko, J. K. Lovelace, N. L. Barber, and K. S. Linsey., 2018. Estimated use of water in the United States in 2015: U.S. Geological Survey Circular 1441, 65 p. https://doi.org/10.3133/cir1441.

Draper, D., 1995. Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society: Series B (Methodological) 57 (1), 45–70. https://doi.org/10.1111/rssb:1995.57.issue-110.1111/j.2517-6161.1995.tb02015.x.

Dyer, J., Mercer, A., Rigby, J.R., Grimes, A., 2015. Identification of recharge zones in the Lower Mississippi River alluvial aquifer using high-resolution precipitation estimates. Journal of Hydrology 531, 360–369. https://doi.org/10.1016/j.jhydrol.2015.07.016.

Elshall, A.S., Tsai, F.T.-C., Hanor, J.S., 2013. Indicator geostatistics for reconstructing Baton Rouge aquifer-fault hydrostratigraphy, Louisiana, USA. Hydrogeology Journal 21 (8), 1731–1747. https://doi.org/10.1007/s10040-013-1037-5.

Fleetwood, A.R., 1969. Geological investigation of the Ouachita River area, Lower Mississippi Valley. U.S. Army Corps of Engineers. Waterways Experiment Station, Technical Report S-69-2, 24p.

Freeze, R.A., Cherry, J.A., 1979. Groundwater. Prentice-Hall, Englewood, Cliffs, New Jersey, p. 604.

Gillip, J. A., and J. B. Czarnecki., 2009. Validation of a ground-water flow model of the Mississippi River Valley alluvial aquifer using water-level and water-use data for 1998-2005 and evaluation of water-use scenarios: U.S. Geological Survey Scientific Investigations Report 2009-5040, 22 p. https://doi.org/10.3133/sir20095040.

Harbaugh, A.W., 1990. A computer program for calculating subregional water budgets using results from the U.S. Geological Survey modular three-dimensional groundwater flow model. U.S. Geological Survey Open-File Report 90–392, 46p. https://doi.org/10.3133/ofr90392.

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. https://doi.org/10.3133/tm6A16.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Statistical science 382–401. https://doi.org/10.1214/ss/1009212519.

Krinitzsky, E.L., Wire, J.C., 1964. Groundwater in alluvium of the Lower Mississippi Valley (upper and central areas): Vicksburg. U.S. Army Engineer Waterways Experiment Station, U.S. Corps of Engineers, Mississippi.

Li, X., Tsai, F.T.-C., 2009. Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod. Water Resources Research 45 (9). https://doi.org/10.1029/2008WR007488.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research: Atmospheres 99 (D7), 14415–14428. https://doi.org/10.1029/94JD00483.

Liu, Z., Merwade, V., 2018. Accounting for model structure, parameter and input forcing uncertainty in flood inundation modeling using Bayesian model averaging. Journal of Hydrology 565, 138–149. https://doi.org/10.1016/j.jhydrol.2018.08.009.

Liu, Z., Merwade, V., 2019. Separation and prioritization of uncertainty sources in a raster based flood inundation model using hierarchical Bayesian model averaging. Journal of Hydrology 578 (11), 124100. https://doi.org/10.1016/j.jhydrol.2019.124100.

Moore, R.B., McKay, L.D., Rea, A.H., Bondelid, T.R., Price, C.V., Dewald, T.G., Johnston, C.M., 2019. User's guide for the national hydrography dataset plus (NHDPlus) high resolution. U.S. Geological Survey Open-File Report 2019–1096. https://doi.org/10.3133/ofr20191096.

Mustafa, S.M.T., Nossent, J., Ghysels, G., Huysmans, M., 2018. Estimation and impact assessment of input and parameter uncertainty in predicting groundwater flow with a fully distributed model. Water Resources Research 54 (9), 6585–6608. https://doi.org/10.1029/2017WR021857.

Naz, B.S., Kao, S.-C., Ashfaq, M., Rastogi, D., Mei, R., Bowling, L.C., 2016. Regional Hydrologic Response to Climate Change in the Conterminous United States Using High-resolution Hydroclimate Simulations, Global Planet. Change 143, 100–117. https://doi.org/10.1016/j.gloplacha.2016.06.003.

Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stochastic Environmental Research and Risk Assessment 17 (5), 291–305. https://doi.org/10.1007/s00477-003-0151-7.

Oubeidillah, A.A., Kao, S.-C., Ashfaq, M., Naz, B.S., Tootle, G., 2014. A Large-Scale, High-Resolution Hydrological Model Parameter Data Set for Climate Change Impact Assessment for the Conterminous US. Hydrol. Earth Syst. Sci. 18, 67–84. https://doi.org/10.5194/hess-18-67-2014.

Pauloo, R.A., Escriva-Bou, A., Dahlke, H., Fencl, A., Guillon, H., Fogg, G.E., 2020. Domestic well vulnerability to drought duration and unsustainable groundwater management in California's Central Valley. Environmental Research Letters 15 (4), 044010. https://doi.org/10.1088/1748-9326/ab6f10.

Pham, H.V., Tsai, F.T.-C., 2017. Modeling complex aquifer systems: a case study in Baton Rouge, Louisiana (USA). Hydrogeology Journal 25 (3), 601–615. https://doi.org/10.1007/s10040-016-1532-6.

Poeter, E., Anderson, D., 2005. Multimodel ranking and inference in ground water modeling. Groundwater 43 (4), 597–605. https://doi.org/10.1111/gwat.2005.43.issue-410.1111/j.1745-6584.2005.0061.x.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Monthly weather review 133 (5), 1155–1174. https://doi.org/10.1175/MWR2906.1.

Reed, T.B., 2003. Recalibration of a Groundwater Flow Model of the Mississippi River Valley Alluvial Aquifer of Northeastern Arkansas, 1918–1998, with Simulations of Water Levels Caused by Projected Groundwater withdrawls through 2049. U.S. Geological Survey Water-Resources Investigations Report 03–4109. https://doi.org/10.3133/wri034109.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process–a framework and guidance. Environmental Modelling & Software 22 (11), 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004.

Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. Water Resources Research 44 (12). https://doi.org/10.1029/2008WR006908.

Saucier, R.T., 1967. Geological investigation of the Boeuf-Tensas Basin. Lower Mississippi Valley, Waterways Experiment Station, United States Mississippi River Commission.

Schöniger, A., Wöhling, T., Samaniego, L., Nowak, W., 2014. Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence. Water resources research 50 (12), 9484–9513. https://doi.org/10.1002/2014WR016062.

Sharif, M., Davis, R., Steele, K., Kim, B., Kresse, T., Fazio, J., 2008. Inverse geochemical modeling of groundwater evolution with emphasis on arsenic in the Mississippi River Valley alluvial aquifer, Arkansas (USA). Journal of Hydrology 350 (1/2), 41–55. https://doi.org/10.1016/j.jhydrol.2007.11.027.

Sharif, M.S.U., Davis, R.K., Steele, K.F., Kim, B., Hays, P.D., Kresse, T.M., Fazio, J.A., 2011. Surface complexation modeling for predicting solid phase arsenic concentrations in the sediments of the Mississippi River Valley alluvial aquifer, Arkansas, USA. Applied Geochemistry 26 (4), 496–504. https://doi.org/10.1016/j.apgeochem.2011.01.008.

Singh, A., Mishra, S., Ruskauff, G., 2010. Model averaging techniques for quantifying conceptual model uncertainty. Groundwater 48 (5), 701–715. https://doi.org/10.1111/j.1745-6584.2009.00642.x.

Smith, R.G., Knight, R., Chen, J., Reeves, J.A., Zebker, H.A., Farr, T., Liu, Z., 2017. Estimating the permanent loss of groundwater storage in the southern S an J oaquin V alley. California. Water Resources Research 53 (3), 2133–2148. https://doi.org/10.1002/2016WR019861.

Smoot, C. W., 1986. Louisiana hydrologic atlas map no. 2: Areal extent of freshwater in major aquifers of Louisiana: U.S. Geological Survey Water-Resources Investigations Report 86-4150, 1 sheet. https://doi.org/10.3133/wri864150.

Stuart, C. G., D. D. Knochenmus, and B. D. McGee., 1994. Guide to Louisiana's groundwater resources. Water-Resources Investigations Report: 94-4085. U.S. Geological Survey, USGS Earth Science Information Center, Open-File Reports Section. https://dx.doi.org/10.3133/wri944085.

Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology 190 (3-4), 214–251. https://doi.org/10.1016/S0022-1694(96)03128-9.

Troldborg, L., Refsgaard, J.C., Jensen, K.H., Engesgaard, P., 2007. Uncertainty in the environmental modelling process–a framework and guidance. Environmental Modelling & Software 22 (11), 1543–1556. https://doi.org/10.1007/s10040-007-0192-y.

Tsai, F.T.-C., Elshall, A.S., 2013. Hierarchical Bayesian model averaging for hydrostratigraphic modeling: Uncertainty segregation and comparative evaluation. Water Resources Research 49 (9), 5520–5536. https://doi.org/10.1002/wrcr.20428.

Tsai, F.T.-C., Sun, N.-Z., Yeh, W.-W.-G., 2005. Geophysical parameterization and parameter structure identification using natural neighbors in groundwater inverse problems. Journal of Hydrology 308 (1), 269–283. https://doi.org/10.1016/j.jhydrol.2004.11.004.

Vahdat-Aboueshagh, H., Tsai, F.T.-C., 2021. Constructing large-scale complex aquifer systems with big well log data: Louisiana model. Computers & Geosciences 148, 104687. https://doi.org/10.1016/j.cageo.2021.104687.

Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research 43 (1), W01411. https://doi.org/10.1029/2005WR004838.

Whitfield, M.S., 1975. Geohydrology and water quality of the Mississippi River alluvial aquifer, northeastern Louisiana. Water Resources Technical Report, No. 10. Louisiana Dept. of Public Works, Baton Rouge, Louisiana.

Wu, C.J., 1983. On the convergence properties of the EM algorithm. The Annals of statistics 11 (01), 95–103.

Yasarer, L.M., Taylor, J.M., Rigby, J.R., Locke, M.A., 2020. Trends in land use, irrigation, and streamflow alteration in the Mississippi River Alluvial Plain. Frontiers in Environmental Science. https://doi.org/10.3389/fenvs.2020.00066.

Ye, M., Pohlmann, K.F., Chapman, J.B., Pohll, G.M., Reeves, D.M., 2010. A model-averaging method for assessing groundwater conceptual model uncertainty. Groundwater 48 (5), 716–728. https://doi.org/10.1111/j.1745-6584.2009.00633.x.

Yin, J., Tsai, F.T.-C., 2018. Saltwater scavenging optimization under surrogate uncertainty for a multi-aquifer system. Journal of Hydrology 565 (10), 698–710. https://doi.org/10.1016/j.jhydrol.2018.08.021.

Yin, J., Tsai, F.T.-C., 2020. Bayesian set pair analysis and machine learning based ensemble surrogates for optimal multi-aquifer system remediation design. Journal of Hydrology 580 (1), 124280. https://doi.org/10.1016/j.jhydrol.2019.124280.

Zhang, X., Srinivasan, R., Bosch, D., 2009. Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. Journal of Hydrology 374 (3), 307–317. https://doi.org/10.1016/j.jhydrol.2009.06.023.