

Evaluating two sparse grid surrogates and two adaptation criteria for groundwater Bayesian uncertainty quantification



Xiankui Zeng^{a,b}, Ming Ye^b, John Burkardt^b, Jichun Wu^{a,*}, Dong Wang^a, Xiaobin Zhu^a

^a Key Laboratory of Surficial Geochemistry, Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

^b Department of Scientific Computing, Florida State University, Tallahassee, FL, USA

ARTICLE INFO

Article history:

Received 28 October 2015

Received in revised form 31 December 2015

Accepted 22 January 2016

Available online 1 February 2016

This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Felipe de Barros, Associate Editor

Keywords:

Log-likelihood surrogate

State-variable surrogate

Adaptive sparse grid

Markov chain Monte Carlo

Bayesian inference

Groundwater modeling

SUMMARY

Sparse grid (SG) stochastic collocation methods have been recently used to build accurate but cheap-to-run surrogates for groundwater models to reduce the computational burden of Bayesian uncertainty analysis. The surrogates can be built for either a log-likelihood function or state variables such as hydraulic head and solute concentration. Using a synthetic groundwater flow model, this study evaluates the log-likelihood and head surrogates in terms of the computational cost of building them, the accuracy of the surrogates, and the accuracy of the distributions of model parameters and predictions obtained using the surrogates. The head surrogates outperform the log-likelihood surrogates for the following four reasons: (1) the shape of the head response surface is smoother than that of the log-likelihood response surface in parameter space, (2) the head variation is smaller than the log-likelihood variation in parameter space, (3) the interpolation error of the head surrogates does not accumulate to be larger than the interpolation error of the log-likelihood surrogates, and (4) the model simulations needed for building one head surrogate can be recycled for building others. For both log-likelihood and head surrogates, adaptive sparse grids are built using two indicators: absolute error and relative error. The adaptive head surrogates are insensitive to the error indicators, because the ratio between the two indicators is hydraulic head, which has small variation in the parameter space. The adaptive log-likelihood surrogates based on the relative error indicators outperform those based on the absolute error indicators, because adaptation based on the relative error indicator puts more sparse-grid nodes in the areas in the parameter space where the log-likelihood is high. While our numerical study suggests building state-variable surrogates and using the relative error indicator for building log-likelihood surrogates, selecting appropriate type of surrogates and error indicators depends on the shapes of response surfaces. The shapes should be approximated and examined before building sparse grid surrogates.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Uncertainty analysis has become a common practice in groundwater modeling in the last several decades for evaluating model predictive performance, improving model structures, and supporting science-informed decision-making (Gupta et al., 2012; Matott et al., 2009; Tartakovsky, 2013). Among various methods developed for uncertainty analysis, Bayesian approaches are one of the most popular methods. However, in comparison with other methods of uncertainty analysis that are computationally frugal (Hill et al., 2015), Bayesian approaches are computationally expensive, because they always involve Markov chain Monte Carlo (MCMC) simulations, in which tens to hundreds of thousands of model

executions are necessary for estimating the probability distributions of model parameters and predictions. To alleviate the computing burden, one solution is to replace a model by its surrogate that is sufficiently accurate but computationally cheap, and a review article of surrogate modeling is given by Razavi et al. (2012). Among various methods of building surrogates, the sparse grid (SG) stochastic collocation methods are used in this study. Although the SG methods have become popular, using them for Bayesian uncertainty quantification has been reported only in a limited number of groundwater studies (Zeng et al., 2012; Zhang et al., 2013, 2015). In other uses of SG methods (e.g., Lin and Tartakovsky, 2009, 2010; Lin et al., 2010; Shi and Yang, 2009; Zhang et al., 2010; Dai and Ye, 2015), SG methods are used to estimate the distributions or moments (e.g., mean and covariance) of groundwater state variables (e.g., hydraulic head and solute concentration). These studies assumed known parameter

* Corresponding author.

E-mail address: jcwu@nju.edu.cn (J. Wu).

distributions, and did not estimate the distributions using Bayesian approaches.

This study investigates an important problem for SG-based Bayesian uncertainty quantification, i.e., how to evaluate the likelihood function used in Bayesian inference. Consider a Bayesian inference problem for a nonlinear model, f , used to simulate state variables (e.g., hydraulic head and solute concentration),

$$\mathbf{d} = f(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{d} is a vector dataset of state variable, $\boldsymbol{\theta}$ is a vector of model parameters, and $\boldsymbol{\varepsilon}$ is a vector of residuals that may include errors in data, model parameters, and model structures. The goal of Bayesian inference is to estimate the posterior distributions, $p(\boldsymbol{\theta}|\mathbf{d})$, of model parameters, $\boldsymbol{\theta}$, given data, \mathbf{d} , using Bayes' theorem (Box and Tiao, 1992)

$$p(\boldsymbol{\theta}|\mathbf{d}) = \frac{L(\boldsymbol{\theta}|\mathbf{d})p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\mathbf{d})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (2)$$

where $p(\boldsymbol{\theta})$ is the prior distribution and $L(\boldsymbol{\theta}|\mathbf{d})$ is the likelihood function to measure goodness-of-fit between model simulations, $f(\boldsymbol{\theta})$, and data, \mathbf{d} . The prior distribution can be specified using data from previous studies or expert judgment. When prior information is lacking, a common practice is to assume uniform distributions with relatively large parameter ranges so that the prior distributions do not affect the estimation of posterior distributions. Defining a likelihood function appropriate to a specific problem is still an open question, and it has been shown that the likelihood function has substantial impacts on the results of Bayesian inference (Evin et al., 2014; Lu et al., 2013; Schoups and Vrugt, 2010; Shi et al., 2014; Smith et al., 2010). While SG methods can work with various likelihood functions (Zhang et al., 2013), this study uses the commonly used Gaussian likelihood function,

$$L(\boldsymbol{\theta}|\mathbf{d}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{d} - f(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}^{-1}(\mathbf{d} - f(\boldsymbol{\theta}))\right), \quad (3)$$

where N is the number of data (i.e., the dimension of \mathbf{d}), and $\boldsymbol{\Sigma}$ is the covariance matrix of the residuals, $\boldsymbol{\varepsilon}$. Because analytical expressions for $p(\boldsymbol{\theta}|\mathbf{d})$ are unavailable for nonlinear models, Markov chain Monte Carlo (MCMC) methods are often used for estimating $p(\boldsymbol{\theta}|\mathbf{d})$. In MCMC, a large number (tens to hundreds of thousands) of parameter samples are drawn; for each sample, the nonlinear function, $f(\boldsymbol{\theta})$, and the likelihood function, $L(\boldsymbol{\theta}|\mathbf{d})$, are evaluated. If the nonlinear function is computationally expensive, the computational cost for the Bayesian inference may be unaffordable. This necessitates the use of SG surrogates.

In SG applications for Bayesian inference, two kinds of SG surrogates have been used. One is for the logarithm of the likelihood function used to directly replace $L(\boldsymbol{\theta}|\mathbf{d})$ during Bayesian inference; the other is for the state variables used to replace $f(\boldsymbol{\theta})$ for evaluating the likelihood. Building the state variable surrogates is common in the literature of not only SG collocation (Ma and Zabarar, 2009b; Zeng et al., 2012; Zhang et al., 2015) but also other stochastic collocation methods of Bayesian inference (Marzouk et al., 2007; Marzouk and Xiu, 2009; Liao and Zhang, 2013; Laloy et al., 2013). While building log-likelihood surrogates is less common (Zhang et al., 2013), it is theoretically superior to building state-variable surrogates for two reasons. First, only one log-likelihood surrogate is needed regardless of the number of observations, whereas one state-variable surrogate is needed for each observation. When the number of observations is large, the computational cost of building multiple state-variable surrogates can be significantly higher than that of building a single log-likelihood surrogate. In addition, each state-variable surrogate has its SG interpolation error, and the error may accumulate and become

large when the surrogates are used for evaluating the likelihood function. However, building the log-likelihood surrogates has its own disadvantages as discussed in the numerical example below. It is therefore necessary to evaluate the two kinds of surrogates to determine which kind of surrogate is more appropriate for Bayesian inference.

To the best of our knowledge, there has been no reported reference on comparing the state-variable surrogates and the log-likelihood surrogates. The study of Petvipusit et al. (2014) is the only reference related to the comparison that we are aware of. The study compared two surrogates used for optimization of CO₂ sequestration. One surrogate was built for a break-even tax credit function, and the other for the moments (i.e., mean and variance) of the function. The comparative study of Petvipusit et al. (2014) showed that building the moment surrogate is computationally more efficient than building the function surrogate. However, their study is irrelevant to comparison between the log-likelihood and state-variable surrogates. The two kinds of surrogates are compared in this study in terms of accuracy and efficiency. The accuracy is evaluated by comparing the posterior distributions obtained using the two kinds of surrogates with the reference distributions obtained using the original model without any surrogates. The computational efficiency is evaluated by directly comparing the number of model executions needed for building the log-likelihood and state-variable surrogates. The comparative evaluation is done by conducting a numerical study for a synthetic groundwater flow model. The conclusions drawn from the synthetic study through the quantitative and comprehensive evaluations are expected to be applicable to other groundwater studies, given that the complexity of the synthetic model is representative for groundwater modeling.

This study also addresses another important issue for building adaptive SG, i.e., whether absolute or relative error should be used as the indicator for adaptation. Building adaptive SG is common for saving computational cost by adding SG nodes only in the areas where SG interpolation error is larger than a user-specified tolerance value (Barthelmann et al., 2000; Klimke, 2006; Ma and Zabarar, 2009a; Pfluger, 2010; Zhang et al., 2013). The absolute error (difference between a model simulation and its surrogate) is the interpolation error itself, and has been used widely (Ma and Zabarar, 2009a, 2009b; Stoyanov, 2013a, 2013b; Webster et al., 2014; Zeng et al., 2012; Zhang et al., 2013), because it directly controls SG accuracy. However, it should be noted that having an accurate SG surrogate is insufficient to having an accurate Bayesian inference, i.e., obtaining accurate posterior parameter distributions. For example, adding adaptive SG points in low likelihood regions to reduce SG error is useless to Bayesian inference, because only parameter samples generated from high likelihood regions are accepted during MCMC simulation; this is demonstrated below using the numerical examples based on the synthetic groundwater model. The key question is where to add adaptive SG nodes in Bayesian inference, and this problem is resolved empirically in this study by using relative error, i.e., absolute error divided by the model simulation. We explore whether the relative error outperforms the absolute error by using both absolute and relative error indicators to build adaptive log-likelihood and state-variable surrogates. As discussed below in Section 4, the two error indicators lead to significantly different SG node locations when building the log-likelihood surrogates, but not the case when building the state-variable surrogates. As a result, the two error indicators have substantial impacts on the accuracy of estimating the posterior distributions of model parameters and predictions.

It should be noted that this study is focused on using SG for Bayesian uncertainty quantification; other uses of SG are beyond

the scope of this study. For example, this study does not consider the use of SG for polynomial chaos expansion methods (Xiu and Hesthaven, 2005). The remainder of this paper is organized as follows. In section 2, we provide a brief description of developing adaptive sparse grids. The synthetic groundwater model, its random parameters and predictions, and the procedure for building the log-likelihood and state-variable surrogates are described in Section 3. The results of evaluating the two kinds of surrogates using the absolute and relative error indicators are analyzed in Section 4. Conclusions of this study are given in Section 5.

2. Methodology

This section briefly describes the sparse grid methods for building surrogates of nonlinear functions (e.g., likelihood function and hydraulic head) in Section 2.1 to make this paper self-contained. The techniques of building adaptive SG surrogates using the absolute and relative error indicator as adaptation criteria are discussed in Section 2.2. Section 2.3 defines the criteria used to evaluate accuracy and efficiency of the distributions of model parameters and predictions obtained using the different SG surrogates.

2.1. Sparse grid stochastic collocation methods

Since sparse grid (SG) surrogates are constructed from a series of one-dimensional (1-D) hierarchical Lagrange interpolants (Bungartz and Griebel, 2004; Klimke, 2006; Ma and Zabarar, 2009b; Zhang et al., 2013), we start this section with 1-D interpolation (Bungartz and Griebel, 2004; Klimke, 2006). Consider a nonlinear function $\eta(\theta)$. For convenience, we will assume θ is contained in the range $[0, 1]$, but simple translation and dilation allows these results to be extended to any bounded interval. The function $\eta(\theta)$ can represent the original nonlinear model or a related function such as the likelihood function. The 1-D hierarchical Lagrange interpolation formula U is defined as

$$U^L(\eta)(\theta) = \sum_{i=0}^L \Delta U^i(\eta)(\theta), \tag{4}$$

where L is the resolution level. $U^L(\eta)(\theta)$ is a simplified expression of $U^L(\eta(\theta))(\theta)$, implying that $U^L(\eta)$ is first built for function $\eta(\theta)$ and then applied to any values of parameter θ of function η . This meaning of the notation is also applied to the incremental interpolation function, $\Delta U^i(\eta)$. The summation over levels up to the resolution level exhibits the hierarchical structure of U^L . The incremental interpolation operator $\Delta U^i(\eta)(\theta)$ is given as,

$$\Delta U^i(\eta)(\theta) = \sum_{j=0}^{m_i} c_j^i \phi_j^i(\theta) \quad i = 0, \dots, L, \tag{5}$$

where the integer m_i is the number of 1-D interpolation points (or SG nodes) for level i , defined by

$$\begin{cases} m_0 = 1 \\ m_1 = 2 \\ m_i = 2^{i-1} \quad \text{if } i \geq 2 \end{cases} \tag{6}$$

and $\phi_j^i(\theta)$ are the interpolation basis functions, and c_j^i are the corresponding interpolation coefficients.

Among the various quadrature rules (e.g., Clenshaw-Curits rule, Fejer rule, Gauss–Legendre rule, Gauss–Patterson rule, and uniform rule) developed for generating the SG nodes (Bungartz and Griebel, 2004; Klimke, 2006; Pfluger, 2010), the uniform rule is used in this study. The abscissas, θ_j^i , of a standard 1-D uniform grid are given by (Bungartz and Griebel, 2004)

$$\begin{cases} \theta_1^0 = 0.5 & \text{if } i = 0 \\ \theta_1^1 = 0, \theta_2^1 = 1 & \text{if } i = 1 \\ \theta_j^i = (2j - 1) / \left(\sum_{k=0}^i m_k - 1 \right) & \text{if } i \geq 2 \end{cases} \tag{7}$$

While linear, quadratic, or cubic hierarchical basis function can be used for hierarchical Lagrange interpolation (Bungartz and Griebel, 2004; Klimke, 2006; Pfluger, 2010), Zhang et al. (2013) demonstrated that the cubic hierarchical basis outperforms the linear and quadratic bases with a significant reduction in the number of required model executions. Therefore, the cubic hierarchical basis is used, and it is given by (Bungartz and Griebel, 2004)

$$\phi_1^0(\theta) = 1 \quad 0 \leq \theta \leq 1 \quad \text{if } i = 0 \tag{8}$$

$$\begin{cases} \phi_1^1(\theta) = \frac{\theta - \theta_1^1 - d\theta}{-d\theta} \cdot \frac{\theta - \theta_1^1 - 2d\theta}{-2d\theta} & 0 \leq \theta \leq 1 \quad \text{if } i = 1 \\ \phi_2^1(\theta) = \frac{\theta - \theta_2^1 + d\theta}{-d\theta} \cdot \frac{\theta - \theta_2^1 + 2d\theta}{-2d\theta} & 0 \leq \theta \leq 1 \quad \text{if } i = 1 \end{cases} \tag{9}$$

$$\begin{cases} \phi_j^i(\theta) = \prod_{k=1}^3 \frac{\theta - \theta_j^i - (2k - 3)d\theta}{-(2k - 3)d\theta} & \theta \in \Pi_j^i, \quad \text{if } i \geq 2, \text{ and } j \text{ is odd} \\ \phi_j^i(\theta) = \prod_{k=1}^3 \frac{\theta - \theta_j^i - (2k - 5)d\theta}{-(2k - 5)d\theta} & \theta \in \Pi_j^i, \quad \text{if } i \geq 2, \text{ and } j \text{ is even} \\ \phi_j^i(\theta) = 0 & \text{otherwise} \end{cases} \tag{10}$$

where $d\theta = \frac{1}{2}$ and $\Pi_j^i = [\theta_j^i - d\theta, \theta_j^i + d\theta]$. Graphic illustrations of this and other basis functions can be found in literature (e.g., Bungartz and Griebel, 2004; Klimke, 2006; Ma and Zabarar, 2009b; Zhang et al., 2013). The interpolation coefficient, c_j^i , is iteratively derived as follows:

$$c_1^0 = \Delta U^0(\eta)(\theta_1^0) = U^0(\eta)(\theta_1^0) = \eta(\theta_1^0) \quad i = 0 \tag{11}$$

$$\begin{aligned} c_j^i &= \Delta U^i(\eta)(\theta_j^i) = U^i(\eta)(\theta_j^i) - U^{i-1}(\eta)(\theta_j^i) \\ &= \eta(\theta_j^i) - U^{i-1}(\eta)(\theta_j^i) \quad i \geq 1 \end{aligned} \tag{12}$$

Equation (12) shows that, the coefficient, also called the hierarchical surplus of the basis function $\phi_j^i(\theta)$, is the difference between the interpolation function $\eta(\theta)$ and the interpolant $U^{i-1}(\eta)$ at θ_j^i . When the function $\eta(\theta)$ is smooth with respect to θ , the magnitude of the surplus c_j^i approaches zero as the resolution level i increases (Klimke, 2006; Ma and Zabarar, 2009a, 2009b). Therefore, the surplus can be used as an error indicator to guide the sparse grid refinement, i.e., adding SG nodes of the next level. It is also used in the numerical exercise below to compare the surrogates for hydraulic head and log-likelihood based on the absolute error and relative error indicators.

Based on the 1-D hierarchical interpolation above, the multi-dimensional hierarchical interpolation for a multivariate function is built as

$$V^{L,D}(\eta)(\theta) = \sum_{\lambda(i) \leq L} \Delta V^{i,D}(\eta)(\theta), \tag{13}$$

where D is the dimensions of model parameter $\theta = (\theta^1, \dots, \theta^D)$. For a full tensor-product grid (not a sparse grid), the index λ of SG nodes at level i is given by

$$\lambda(i) = \max(i_1, \dots, i_D). \tag{14}$$

The total number of points of the full-tensor product grid is $(\sum_{i=1}^L m_i)^D$, which grows exponentially with the model parameter

dimension, D . For a sparse grid isotropic for all the parameters, $\lambda(i)$ is given by

$$\lambda(i) = i_1 + \dots + i_D. \tag{15}$$

where $\mathbf{i} = (i_1, \dots, i_D)$ is a multi-index of the resolution level of $\Delta V^{i,D}(\eta)(\theta)$. As a function of level, the sparse grid used far fewer points, as prescribed by the Smolyak rule (Eq(15)). Fig. 1 illustrates the full tensor-product grid and the sparse grid built for a two-dimensional problem ($D = 2$) with the same maximum resolution level of $L = 3$. The number of SG grid nodes (i.e., the number of model executions) is reduced from 81 for the full grid to 29 for the sparse grid.

The incremental interpolation operator $\Delta V^{i,D}(\eta)(\theta)$ is given by

$$\Delta V^{i,D}(\eta)(\theta) = \Delta U^{i_1} \otimes \dots \otimes \Delta U^{i_D} = \sum_j c_j^i \phi_j^i(\theta), \tag{16}$$

where $\mathbf{j} = (j_1, \dots, j_D)$ and each j is a set ($j_l = 1, \dots, m_{i_l}, l = 1, \dots, D$). The multi-dimensional hierarchical basis function, ϕ_j^i , is defined by

$$\phi_j^i(\theta) = \prod_{n=1}^D \phi_{j_n}^{i_n}(\theta_n), \tag{17}$$

where $\phi_{j_n}^{i_n}(\theta_n)$ is the 1-D hierarchical basis function given above. The multi-dimensional interpolation coefficient, c_j^i , is also derived iteratively by

$$c_1^0 = \Delta V^{0,D}(\eta)(\theta_1^0) = V^{0,D}(\eta)(\theta_1^0) = \eta(\theta_1^0) \quad L = 0 \tag{18}$$

$$\begin{aligned} c_j^i &= \Delta V^{i,D}(\eta)(\theta_j^i) = V^{L,D}(\eta)(\theta_j^i) - V^{L-1,D}(\eta)(\theta_j^i) \\ &= \eta(\theta_j^i) - V^{L-1,D}(\eta)(\theta_j^i) \quad L \geq 1 \end{aligned} \tag{19}$$

2.2. Absolute and relative error indicators for building adaptive sparse grids

For high dimensional problems, using adaptive SG is necessary to further control the growth in SG nodes, i.e., the computational cost of building the SG surrogate. Following Ma and Zabarar (2009a) and Zhang et al. (2013) and using a 1-D SG as an example, each SG node has two children at the next level. For example, the

j -th SG node θ_j^i at level i has two child nodes, θ_{2j-1}^{i+1} and θ_{2j}^{i+1} , at level $i + 1$. Before adding the two child nodes, the error indicator at θ_j^i (i.e., the corresponding interpolation coefficient c_j^i) is compared with a user-prescribed error tolerance, α . If the error indicator is larger than α , then the two children are included in the new SG nodes of level $i + 1$; otherwise, the grid refinement is terminated at node θ_j^i . An example of the one-dimensional adaptive sparse grid is referred to Fig. 4 of Zhang et al. (2013). Evaluating the error indicator is important to the adaptation, because the SG only refines for the nodes whose error indicator are larger than the error tolerance, α . In this study, we consider two error indicators: the absolute error (AE),

$$AE = |\eta(\theta) - V^{L,D}(\eta)(\theta)|, \tag{20}$$

and the relative error (RE),

$$RE = \left| \frac{\eta(\theta) - V^{L,D}(\eta)(\theta)}{\eta(\theta)} \right|. \tag{21}$$

The impact of the two error indicators on accuracy and efficiency of SG-based Bayesian inference will be evaluated in the numerical example below.

For a multi-dimensional problem, developing adaptive SG (isotropic for all the parameters) is more complicated, because the child nodes of each SG node have a tree-like structure with two children in each direction in parameter space. For a SG node $\theta_j^i(\theta_{j_1}^i, \dots, \theta_{j_D}^i)$ at level i , the two children, $C_1^n(\theta_j^i)$ and $C_2^n(\theta_j^i)$, in the n -th dimension are

$$\begin{aligned} C_1^n(\theta_j^i) &= (\theta_{j_1}^i, \dots, \theta_{j_{n-1}}^i, \theta_{2j_{n-1}+1}^{i+1}, \theta_{j_{n+1}}^i, \dots, \theta_{j_D}^i) \\ C_2^n(\theta_j^i) &= (\theta_{j_1}^i, \dots, \theta_{j_{n-1}}^i, \theta_{2j_n}^{i+1}, \theta_{j_{n+1}}^i, \dots, \theta_{j_D}^i) \end{aligned} \tag{22}$$

Because a node in the multi-dimensional SG has multiple parents in each direction, the adaptation procedure becomes more complicated, although the adaptation procedure is still similar to that of the 1-D case by comparing the error indicator with an error tolerance. For example, $C_1^n(\theta_j^i)$ is not only a child of $\theta_j^i(\theta_{j_1}^i, \dots, \theta_{j_D}^i)$ but also

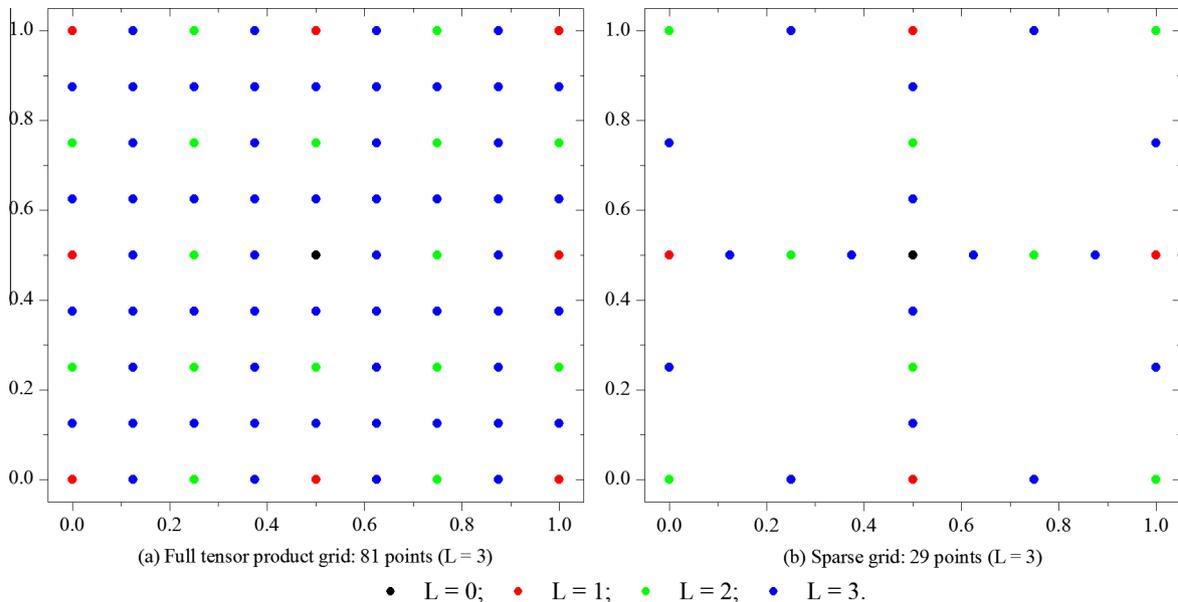


Fig. 1. Illustration of (a) full tensor-product grid and (b) sparse grid built for a case of two parameters and with the maximum resolution level of 3. Sparse grid nodes of different levels are marked in different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a child of $\theta_j^i(\theta_{j_1}^{i-1}, \dots, \theta_{j_{n-1}}^{i-1}, \theta_{2j_{n-1}}^{i+1}, \theta_{j_{n+1}}^{i+1}, \dots, \theta_{j_D}^i)$, $\theta_j^i(\theta_{j_1}^i, \theta_{j_2}^{i-1}, \dots, \theta_{j_{n-1}}^{i-1}, \theta_{2j_{n-1}}^{i+1}, \theta_{j_{n+1}}^{i+1}, \dots, \theta_{j_D}^i)$, and other SG nodes, where

$$\begin{cases} j_1' = (j_1 + 1)/2 & \text{if } j_1 \text{ is odd} \\ j_1' = j_1/2 & \text{if } j_1 \text{ is even} \end{cases} \quad (23)$$

As a result, a child node that is not needed by a parent node at one adaptation step may be needed by another parent node in the same or different adaptation step. This is illustrated in Fig. 2. While the blue nodes with square frames at level 3 are not needed by the green nodes with square frames at level 2, the nodes with labels (1) and (2) are needed by the green nodes at the southeast and northwest corners. This adaptive SG may correspond to the scenario that SG node refinement is not needed at the third quadrant but needed in other quadrants, if the function is smooth in the third quadrant where a smaller number of SG nodes are needed to build the SG surrogates. In other words, the adaptation strategy described above considers local smoothness of the nonlinear functions in different quadrants. The same adaptation procedure applies when different basis functions and quadrature rules are used, although the locations of SG nodes become different.

Special attention should be paid when using the relative error defined in Eq. (21), especially when it is used for the log-likelihood function, since the function depends on multiple factors, including function forms (e.g., Gaussian and non-Gaussian), number of observations, \mathbf{d} , and error statistics (e.g., the covariance matrix used in Gaussian likelihood). We suggest examining the shape and magnitude of the log-likelihood to ensure that the relative error is useful to improve the accuracy of adaptive SG. The definition in Eq. (21) meets this goal for the Gaussian likelihood defined in Eq. (3), because, when the number of observation is not small and when the fit of model simulations to observation is not perfect, the log-Gaussian likelihood is negative with small absolute values in the area at the vicinity of the true parameter

values. The relative error is thus large in this area, and more adaptation SG nodes are added in this area to improve the adaptive SG and the surrogate-based Bayesian inference. This is illustrated in the numerical example below. When the relative error is used for building state-variable surrogates, if the state-variable variation is small in parameter space, using the relative error should yield similar results to those obtained using the absolute error. This is also illustrated in the numerical example below. When the relative error is used for building the state-variable surrogates, more nodes are added in the area where the state-variable simulations are small. This may be a disadvantage, because it makes difficult to directly control SG accuracy. However, the advantages and disadvantages are unknown, and more research is warranted in a future study, especially for groundwater reactive transport modeling.

2.3. Evaluation of accuracy and computational efficiency

For the four SG surrogates (log-likelihood and state-variable surrogates with the absolute and relative error indicators), computational efficiency is evaluated by comparing the number of model executions needed to build the surrogate. The computational time of using the surrogates for estimating the distributions of model parameters and predictions is ignored, because it is negligible in comparison with the computational time of building the surrogates. For the issue of accuracy, since the SG interpolant accuracy is directly measured by the surplus of the basis functions (see Eq(19)), we focus on the two kinds of posterior distributions: (1) the distributions of model parameters estimated using the MCMC simulation with the adaptive SG surrogates and (2) the distributions of model predictions based on the parameter samples obtained from the MCMC simulations. The accuracy is measured by comparing the distributions obtained using the surrogates with the distributions obtained using the original model without any surrogates; the latter distributions are also referred to as reference distributions. The measure used in this study is the relative entropy, $D(p||q)$, defined as (Cover and Thomas, 2006)

$$D(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx, \quad (24)$$

where $p(x)$ is the reference distribution, $q(x)$ is the surrogate-based distribution. Variable x can be either model parameters or predictions in this study. In addition, the relative entropy is evaluated for both the joint distribution of all model parameters/predictions and for the marginal distributions of individual model parameters/distributions. In practice, the distributions, $p(x)$ and $q(x)$, are approximated by the empirical cumulative distribution functions based on the MCMC samples of model parameters and predictions (Lee and Park, 2006; Perez-Cruz and Fernando, 2008). Smaller values of $D(p||q)$ indicate more accurate surrogate-based distributions that are closer to the reference distributions. Generally speaking, $q(x)$ is considered to be sufficiently close to $p(x)$, if the relative entropy is in the order of magnitude of 10^{-4} (Ma and Zabarar, 2009b).

3. Synthetic groundwater model, random parameters, and surrogate building

The synthetic model, revised after that of Rojas et al. (2008), considers steady-state groundwater flow in a three-dimensional domain with the dimension of 5000 m in length, 3000 m in width, and 60 m in depth. As shown in Fig. 3, the domain has three geological layers, layer 1 being unconfined, layer 2 being a confining layer, and layer 3 being confined. The thicknesses of Layers 1–3 are 35 m, 5 m and 25 m, respectively. The north and south lateral boundaries are impermeable. The east boundary is a river located

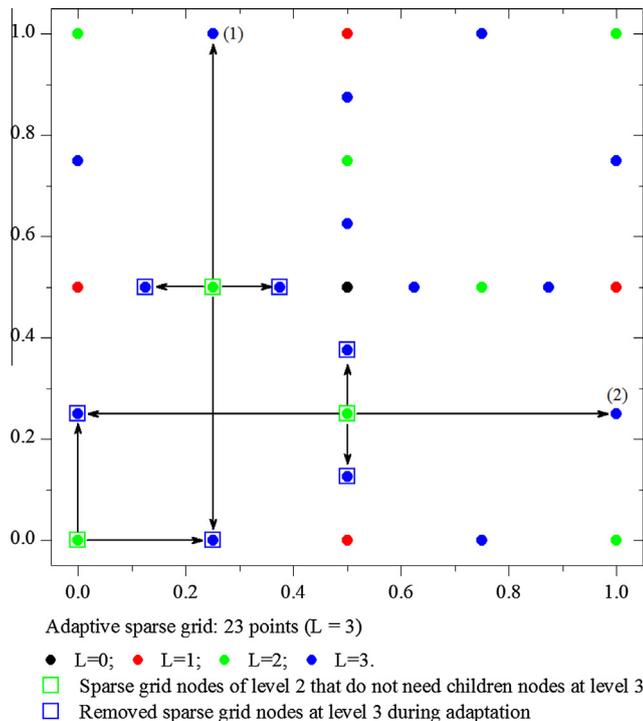


Fig. 2. The standard uniform sparse grid (a) and adaptive sparse grid (b) built for 2-dimensional case. The maximum resolution level is 3. The end point of an arrow line represents the SG node (level 3) deleted by the grid refinement termination of the start point (level 2).

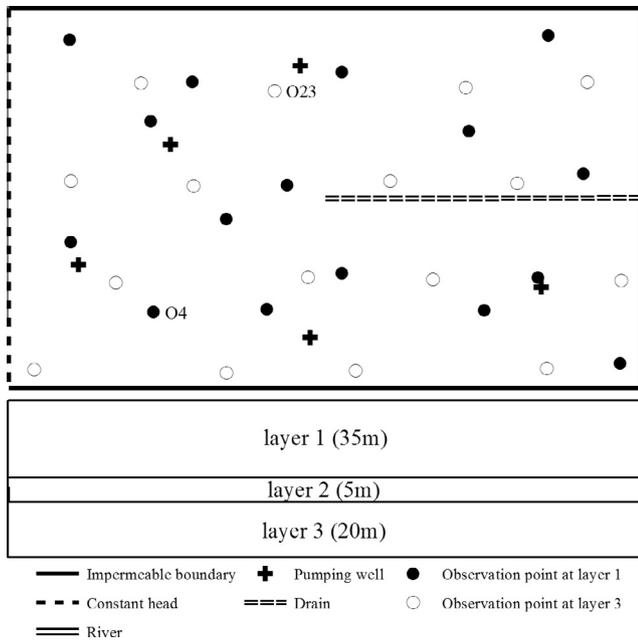


Fig. 3. Sketch map of the synthetic model domain, boundary conditions, and locations of drain and head observations. Observations O4 and O23 marked in the figure are selected to evaluate the head surrogates using absolute and relative error indicators.

in layer 1; the river width is 10 m, and the river stage elevation is 35 m. The elevation of the riverbed bottom is 30 m, and the riverbed thickness is 5 m; the riverbed conductance is $20 \text{ m}^2/\text{d}$. The east boundary beneath the river is impermeable. The west boundary has a constant head of 56 m for all the three model layers. The bottom boundary is impermeable. At the domain surface, uniform precipitation with the rate of $9.0 \times 10^{-4} \text{ m/d}$ is applied. The recharge rate to the aquifer is estimated by multiplying the precipitation rate to a recharge ratio, which is set as 0.15 in the synthetic model. A drain is located at the center of the right half of the domain; the drain bottom elevation is 45 m, and the drain conductance is $20 \text{ m}^2/\text{d}$. A total of five pumping wells are located across the domain, and a pumping rate of $250 \text{ m}^3/\text{d}$ per well is applied to the confined layer (Layer 3). Heterogeneous fields of log hydraulic conductivity are generated for each of the three layers using the sequential Gaussian simulator (SGSIM) of the Geostatistical Library (GSLIB) (Deutsch and Journel, 1998). For all the three layers, an isotropic exponential covariance function is used with the correlation length of 200 m, and the variance of log hydraulic conductivity is set as 1.0. The mean values are 1.0 m/d, 0.1 m/d, and 5.0 m/d for layers 1–3, respectively. The generated hydraulic conductivity is used for the horizontal direction, and the vertical hydraulic conductivity is calculated as 1/10 of the horizontal one. The flow model is solved numerically using MODFLOW-2005 (Harbaugh, 2005). The quantities of prediction are the groundwater discharge to the river, the recharge from the constant head boundary to the aquifer, and the groundwater discharge to the drain.

A total of five random parameters are considered in this study, and they are the recharge ratio, the boundary constant head, the riverbed conductance, and the variance and correlation length of the log hydraulic conductivity field of layer 1. These random parameters represent different uncertainty in the three important factors of the groundwater system: driving force (recharge ratio), boundary conditions (boundary constant head and riverbed conductance), and hydraulic conductivity (variance and correlation length). For the latter two parameters, they are not used to generate multiple realizations of hydraulic conductivity. Instead, they are used in the manner of kriging to generate a mean field of

hydraulic conductivity, and the mean field is used for the head simulation, as done in the stochastic inverse modeling of Hernandez et al. (2006). Uniform priors are assumed for the parameters, and their ranges are [0.08, 0.28] for the recharge ratio, [46, 62] for the boundary constant head, [1, 50] for the riverbed conductance, [0.1, 10] for the variance, and [20, 1000] for the correlation length. Although the number of random parameters is relatively small and cannot demonstrate the capability of the SG methods to handle relatively high-dimensional problems, it is expected that the conclusions of this study about the two kinds of surrogates and the two kinds of error tolerance are applicable to the relatively high-dimensional problems. Extending this study to a relatively high-dimensional problem is warranted in a future study.

The data used for estimating the posterior distributions include 120 measurements of hydraulic conductivity (K) (corrupted by 3% white noise of measured values) and 32 observations of hydraulic head (16 from layer 1 and 16 from layer 3) (Fig. 3). Measurement errors of white noise (with mean of zero and variance of 0.01 m^2) are also added to the head data. The corrupted data are used in the MCMC simulations. While the measurements of hydraulic conductivity are used as conditioning data of SGSIM to generate heterogeneous hydraulic conductivity, only measurements of hydraulic head are used to evaluate the likelihood function. The Differential Evolution Adaptive Metropolis Approach (DREAM) developed by Vrugt et al. (2008, 2009) is used to estimate the posterior distributions of model parameters, and the most recent DREAM_(ZS) code (Laloy and Vrugt, 2012; Vrugt and Ter Braak, 2011) is used in this study. When the original model is used to estimate the reference parameter distributions, three parallel chains are used, and the length of each chain is 30,000. After the burn-in period, the last 20,000 samples of each chain are used. Therefore, the computational cost without using surrogates requires a total of 90,000 model executions, which are significantly larger than that needed for building the surrogates, as shown below. It should be noted that the conclusions drawn from the numerical study are independent of the MCMC approaches, and other approaches (e.g., the delayed rejection adaptive Metropolis Hastings algorithm developed by Haario et al. (2006)) can also be used.

When building the adaptive SG surrogates for this five-dimensional problem, the adaptation occurs at level 5, because the number of SG nodes is only 180 at level 4 and the number of SG nodes increases substantially at level 5. The maximum level is set to 8, and the results below show that it is sufficient to have accurate estimates of distributions of model parameters and predictions. Given that the log-likelihood varies between negative one million and ten, when building the log-likelihood surrogate (SG_L_AE) using the absolute error indicator, the error tolerance is set as 0.1. When building the log-likelihood surrogate (SG_L_RE) using the relative error indicator, the error tolerance is set as 0.01, i.e., 1% of the absolute log-likelihood. The head surrogates are built for each head observation using the absolute error (SG_H_AE) and relative error (SG_H_RE). The maximum resolution level is also set to 8, and the adaptation starts at level 5. Since hydraulic head varies between 40 and 60 m and the standard deviation of measurement error is assumed to be 0.1 m, the absolute and relative error tolerances are set to 0.001 and $2.0\text{E}-5$, respectively, about two orders of magnitude smaller than the measure error and the relative error (measurement error divided by the mean head). Because the error tolerances are strict, a large number of model evaluations are needed when building the SG surrogates. The computational cost can be reduced dramatically by increasing the error tolerances but at the cost of moderately decreasing the accuracy of the results of Bayesian inference (results not shown).

It should be noted that building head surrogates for multiple head observations is not necessarily computationally more

expensive than building a likelihood surrogate for a single likelihood function. Since forward model runs provide head simulations for all head observations, head simulations used for building one head surrogate can be used for building another head surrogate without increasing the computational cost. In an extreme case, if all head surrogates use the same SG nodes in the parameters space, the computational cost of building the multiple head surrogates is the same as that of building one head surrogate.

When estimating the distributions of model predictions, a common practice is to first build the surrogates for the prediction variables and then use the surrogates to generate model predictions based on the MCMC samples (i.e., the posterior samples) obtained from the Bayesian inference. This however is not done in this study, because we need to evaluate the accuracy of the model predictions based on the MCMC samples obtained using the four surrogates. In other words, we need to exclude the error that may be caused by using the surrogates of model predictions. Therefore, the original model is used to estimate the distributions of the model predictions, i.e., running MC simulations using the original model with the MCMC samples.

4. Numerical results and evaluation of the surrogates

This section starts in Section 4.1 to present the distributions of model parameters and predictions obtained using the original model without using any surrogates. These distributions are used as the reference for evaluating the accuracy of the corresponding distributions obtained using the surrogates. In Section 4.2, the distributions obtained using the log-likelihood surrogates are presented. The results indicate that, the log-likelihood surrogates do not give satisfactory results, although using the relative error indicator is better than using the absolute error indicator. Section 4.3 shows the distributions obtained using the head surrogates are better than those obtained using the log-likelihood surrogates. The section also shows that using absolute or relative error indicator has minimal impacts on the distributions of model parameters

and predictions. The reasons that the head surrogates outperform the log-likelihood surrogates are given in Section 4.4.

4.1. Reference distributions obtained without using surrogates

Fig. 4(a)–(e) plot the marginal posterior probability density functions (converted from histograms) of the five model parameters based on the samples obtained from MCMC simulations using the original model without any surrogates; these histograms are labeled as reference. Fig. 4(f)–(h) does the same for the three model predictions. These plots indicate that the MCMC results are satisfactory for quantifying parametric and predictive uncertainty in that the true parameter values and predictions are at the modes or close to the modes of the distributions. In addition, except for the riverbed conductance, the posterior parameter distributions of the other four parameters are narrow and significantly concentrated relative to the uniform prior distributions (results not shown). The difference between the prior and posterior parameter distributions indicates that the available data of hydraulic conductivity and head observations are sufficient for reducing parametric uncertainty. Fig. 4(f)–(h) shows that the predictive uncertainty is still significant, and quantifying its uncertainty is necessary.

4.2. Evaluating log-likelihood surrogates built using absolute and relative errors

Fig. 4(a)–(e) plot the marginal posterior probability density functions for the five model parameters obtained using the SG surrogates built for the log-likelihood function using the absolute error indicator (denoted as SG_L_AE) and the relative error indicator (denoted as SG_L_RE). These plots show that the surrogate-based density functions are different from the reference density functions, especially in the areas near the true values that have high density and are important for uncertainty analysis and risk assessment. The relative entropies listed in Table 1 for the marginal and joint distributions are larger than $1E-4$, suggesting that the SG-based density functions are significantly different from the

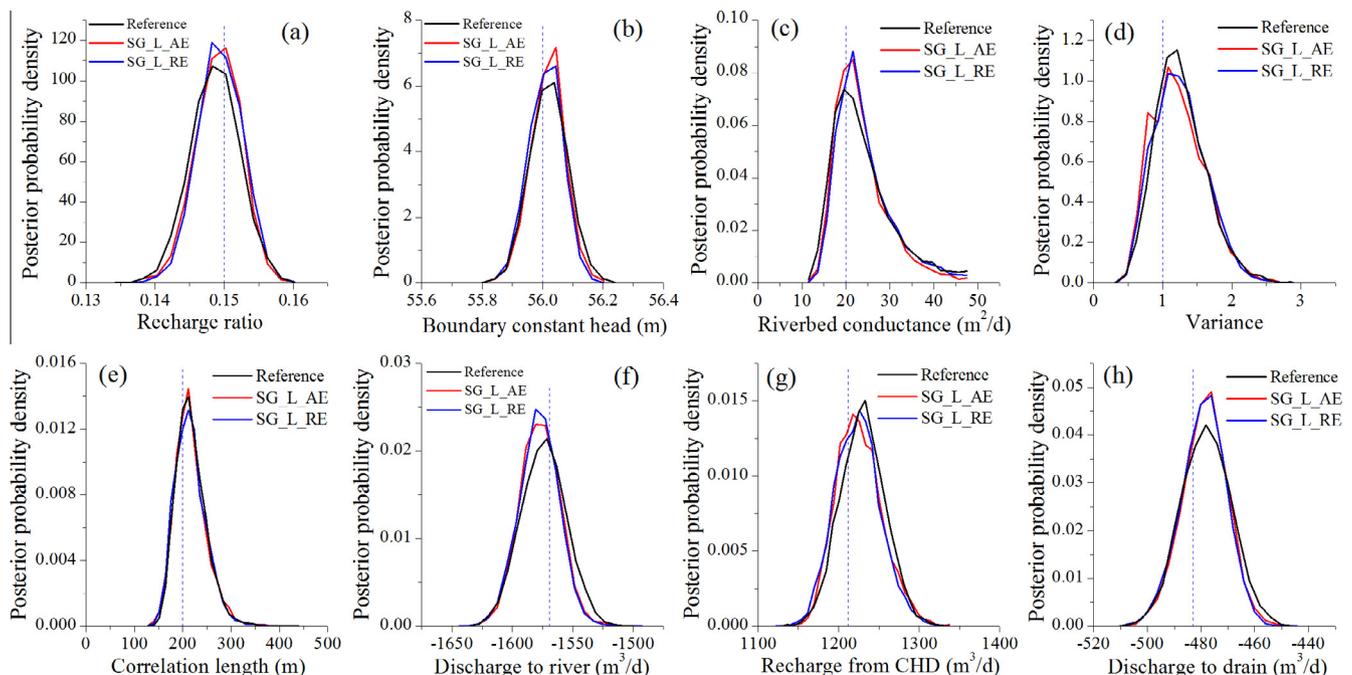


Fig. 4. Posterior probability density functions of (a–e) five random parameters and (f–h) three model predictions based on the MCMC simulations using the original model (reference) and sparse grid surrogates built for log-likelihood, SG_L_AE, and SG_L_RE. The vertical blue lines represent the true parameter values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Relative entropies (REn) between reference posterior distributions (obtained without using surrogates) and those obtained using the log-likelihood surrogate (SG_L_AE) with absolute error and the log-likelihood surrogate (SG_L_RE) with relative error. REn is evaluated for the marginal distributions of individual model parameters/predictions and for the joint distribution of all the parameters/predictions.

Model parameters	SG_L_AE		SG_L_RE	
	Marginal REn	Joint REn	Marginal REn	Joint REn
Recharge ratio	1.58E-2	1.01E-1	3.80E-2	9.22E-2
Boundary constant head	8.75E-3		3.25E-4	
Riverbed conductance	2.99E-2		8.63E-3	
Variance	4.56E-3		2.39E-4	
Correlation length	1.39E-3		1.38E-3	
Model predictions				
Discharge to river	3.08E-2	7.09E-1	4.78E-2	5.68E-1
Recharge from boundary constant head	2.38E-2		1.03E-2	
Discharge to drain	2.09E-2		1.24E-2	

reference distributions. In Fig. 4(a)–(e), the density functions based on SG_L_RE appear to be slightly closer to the reference distributions than those based on SG_L_AE. This is confirmed by the relative entropy values listed in Table 1. Except for the parameter of recharge ratio, the relative entropy values of the marginal distribution are smaller for SG_L_RE than for SG_L_AE. The exception for recharge ratio may not be critical, because the distributions based on the two surrogates are visually similar, as shown in Fig. 4(a), except at the peaks.

Fig. 4(f)–(h) plot the marginal posterior probability density functions for the three predictions obtained using the original model (without surrogate) and the two surrogates of SG_L_AE and SG_L_RE. These plots show again that the two surrogate-based density functions are different from the reference density functions, especially in the areas near the true values with high density. The relative entropy listed in Table 1 for the joint distribution is also in the order of magnitude of 1.0E-1, indicating a significant difference from the reference distributions. Table 1 shows that, overall, the relative entropy values of the marginal distributions are larger for SG_L_AE than for SG_L_RE.

The reason that the distributions based on the log-likelihood surrogates are not satisfactory is attributed to the accuracy of the log-likelihood surrogates. Fig. 5 plots the variation of SG interpolation error (averaged over all SG nodes) with the number of SG nodes. Each dot in the figure corresponds to a resolution level. The SG adaptation starts at level 5, and the SG error reaches its minimum of 0.2 for SG_L_AE and 0.3 for SG_L_RE at level 8, after which the SG error does not decrease. The larger interpolation error for SG_L_RE is reasonable, because the relative error tolerance is set as 1% of the log-likelihood values that are large in magnitude. Although the SG interpolation error can be further reduced by increasing the maximum resolution level of 8, it is not pursued in this study, because the numbers of SG nodes (25,980 and 13,964 for SG_L_AE and SG_L_RE, respectively) are already significantly larger than those for building the head surrogates discussed below.

The number of SG nodes and their locations in the parameter space is determined by not only the error indicators but also the shape and magnitude of the log-likelihood function. This can be understood by examining Fig. 6, which plots the log-likelihood function and the node locations of the two surrogates for the individual parameters (the other parameters are fixed at their true value). The figure shows that the log-likelihood function ranges between -60,000 and 10. Given the large absolute values (i.e., the negative values with large magnitude), when building the likelihood surrogate, the absolute error is seldom smaller than the absolute error tolerance of 0.1. Therefore, more nodes are needed, and the added nodes are always in the areas where the log-

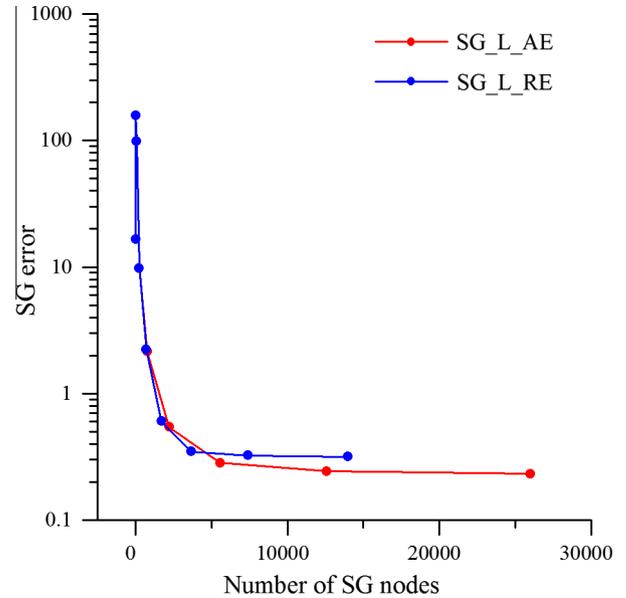


Fig. 5. Variation of average SG interpolation error with number of SG nodes at each level (dot) for surrogates SG_L_AE and SG_L_RE.

likelihood function has large absolute value (i.e., the log-likelihood is negative but with large magnitude). Such nodes however are useless for estimating posterior parameter distribution, because of the low log-likelihood values.

This problem can be avoided by using the relative error indicator, because the relative error indicator can quickly become smaller than the relative error tolerance of 0.01. In addition, since larger relative errors always occur near the peak of the log-likelihood where the absolute value (magnitude) of the log-likelihood function is smaller, more nodes of SG_L_RE are added to the areas where the log-likelihood function has smaller absolute value (magnitude). As a result, SG_L_RE has fewer nodes than SG_L_AE in the entire parameter space, but has more nodes than SG_L_AE near the peak of the log-likelihood function. This is a useful feature for accurately estimating the posterior parameter. It should be noted that the usefulness of the relative error depends on the shape of the log-likelihood. The discussion above is for the log-likelihood that is negative with small absolute values near the peak of the log-likelihood, which is always true for the Gaussian likelihood function in practice.

To further examine the accuracy of SG_L_AE and SG_L_RE, the two surrogates are used to evaluate the log-likelihood function at certain points in the parameter space. Fig. 7 shows the true log-likelihood function values and the errors of the two surrogates at 100 uniform points along the range of prior distributions of the five individual parameters (the other four parameters are fixed at their true values). The root mean squared error (RMSE) of the 100 points is also evaluated for the two surrogates, and they are given in Fig. 7. Although SG_L_RE has larger RMSE than SG_L_AE for all the five parameters, the error of SG_L_RE is smaller than that of SG_L_AE at the peaks of the log-likelihood, i.e., the high probability region on the posterior parameter distributions. Nonetheless, the errors of the two surrogates are all larger than their error tolerance, and this is believed to be the reason that the log-likelihood surrogates do not produce accurate distributions of the model parameters and predictions.

4.3. Evaluating head surrogates built using absolute and relative errors

Fig. 8(a)–(e) plots the marginal posterior probability density functions for the five model parameters obtained using the original

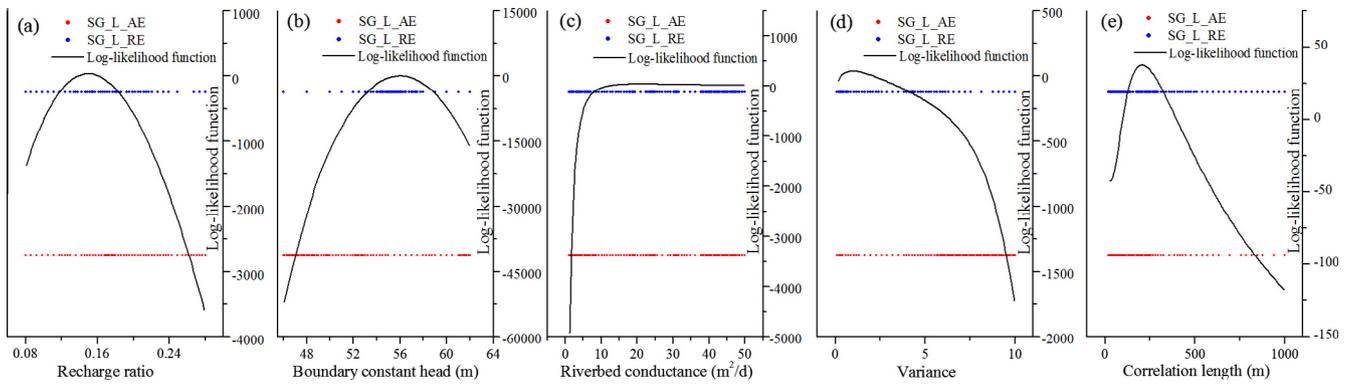


Fig. 6. Log-likelihood function and SG nodes in the space of individual parameters (the other four parameters are fixed at their true values). Each dot denotes a SG node.

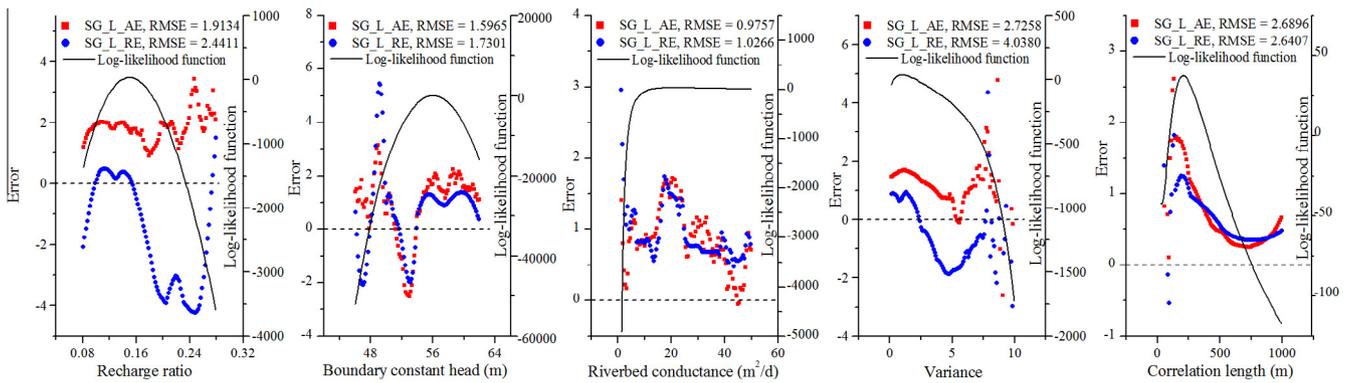


Fig. 7. Errors of SG_L_AE and SG_L_RE for evaluating log-likelihood function at 100 uniform points in the space of the individual parameters (the other four parameters are fixed at their true values). RMSE is the root mean squared errors averaged over the 100 points.

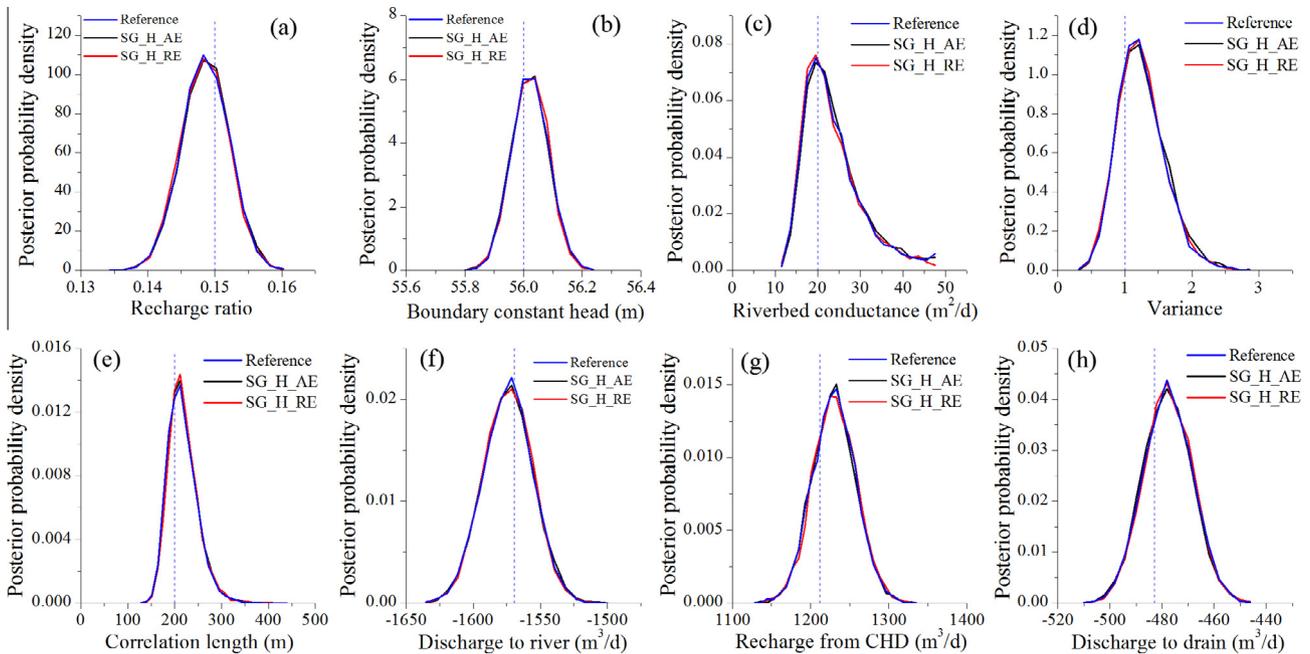


Fig. 8. Posterior probability density functions of (a–e) five random parameters and (f–h) three model predictions based on the MCMC simulations without using surrogate (reference) and two sparse grid surrogates built for head, SG_H_AE and SG_H_RE. The vertical blue lines represent the true parameter values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

model (without any surrogates), the sparse grid surrogates built for head using the absolute error (denoted as SG_H_AE) and relative error (denoted as SG_H_RE). The numbers of SG nodes are 8162

and 8205 for SG_H_AE and SG_H_RE, respectively, significantly smaller than the number of 51,713 for the full SG without adaptation. The numbers of SG nodes of the two head surrogates are also

significantly smaller than those for SG_L_AE (25,980) and SG_L_RE (13,964). Therefore, building the head surrogates is computationally more efficient than building the log-likelihood surrogates.

Comparing Fig. 8(a)–(e) with Fig. 4(a)–(e) shows that the density functions based on head surrogates are significantly closer to the reference than those based on log-likelihood surrogates. This is confirmed by the relative entropy values listed in Table 2 for the head surrogates, because these values are about one order of magnitude smaller than those listed in Table 1 for the log-likelihood surrogates. The small relative entropy values listed in Table 2 suggest that the distributions based on the head surrogates are sufficiently close to the reference. Nonetheless, the distributions based on SG_H_RE are better than those based on SG_H_AE. For example, the relative entropy of the joint distribution for SG_H_RE is smaller than that for SG_H_AE. Considering the approximation error of evaluating the relative entropy (Wang et al., 2009), the difference of relative entropies for SG_H_AE and SG_H_RE in Table 2 is insignificant.

Fig. 8(f)–(h) plot the marginal posterior probability density functions for the three predictions with and without using surrogates. Comparing these figure with Fig. 4(f)–(h) shows that the distributions based on the head surrogates are more accurate than the distributions based on the log-likelihood surrogates. The same conclusion can be drawn by comparing the values of relative entropies listed in Table 1 (for the log-likelihood surrogates) and Table 2 (for the head surrogates). For the distributions of SG_H_AE and SG_H_RE shown in Fig. 8(f)–(h), they are visually similar, although the relative entropy values listed in Table 2 indicate that the results of SG_H_RE are slightly better. However, the difference is insignificant to conclude that SG_H_RE outperforms SG_H_AE in the synthetic study.

The reasons for the similar results for SG_H_AE and SH_H_RE are investigated by examining accuracy of the two surrogates. Similar to Fig. 5, Fig. 9 plots the SG errors averaged over all SG nodes for the two head surrogates built for two head observations at location O4 and O23 marked in Fig. 3 (results for the other observations are similar and thus not shown). The error decreases sharply when the level increases, and the errors of the two head surrogates are similar in terms of their magnitudes and variation trends at all the resolution levels, despite the fact that the error for SG_H_RE is slightly smaller than that for SG_H_AE (at O4, the final errors are $3.5E-4$ and $3.3E-4$ for SG_H_AE and SG_H_RE, respectively; at O23, the final errors are $5.4E-4$ and $5.2E-4$ for SG_H_AE and SG_H_RE, respectively). This explains the reasons why the posterior parameter distributions based on the two surrogates are

similar but the distributions based on SG_H_RE are slightly closer to the reference distributions than those based on SG_H_AE.

The impacts of the absolute and relative error indicators on determining SG nodes are also examined by plotting the nodes of the two surrogates in the prior space of the individual parameters (the other four parameters are fixed at their true value). Fig. 10 includes such plots for O4 and O23, and it shows that the spatial patterns of SG nodes in the parameter space are similar for the two surrogates, indicating that the two error indicators have minimal impacts on building adaptive sparse grids. Fig. 10 also plots the head values to analyze the relation between the SG nodes and hydraulic head in parameter space. This figure shows that, unlike what was shown in Fig. 6, the spatial patterns of the SG nodes are similar for the two surrogates. This is attributed to the small variation (0.4–4 m) of hydraulic head in the parameter space. For example, assuming that two SG nodes have absolute interpolation error of $2.0E-3$, since it is larger than the tolerance of $1.0E-3$, child nodes will be added to the two nodes. Further assume that the head values at the two nodes are 44 and 46 m, respectively. The relative errors at the two nodes are $4.5E-5$ and $4.3E-5$, respectively. Since these errors are larger than the tolerance of $2E-5$, child nodes will be also added to the two nodes. Therefore, due to the small variation of head in the parameter space, the absolute error and relative error have an almost constant ratio (the head value). As a result, the two head surrogates have similar numbers and spatial patterns of the SG nodes. However, this may not be true for other groundwater variables that are highly nonlinear functions of model parameters. For example, in groundwater reactive transport modeling, concentration simulations may vary substantially in parameter space, and the relative performance of SG surrogates based on the absolute and relative errors should be carefully examined.

Similar to Fig. 7, Fig. 11 plots the true head values at observations O4 and O23 evaluated at 100 uniform points. It also plots the errors when the head values are estimated using the two surrogates. The RMSE of the 100 points is in the order of $1.0E-3$ – $1.0E-4$, which are similar to average errors at the SG nodes (Fig. 9). It indicates that the head surrogates are accurate to reproduce the head values. Because of the high accuracy of the head surrogates, they produce accurate estimates of the log-likelihood. Similar to Fig. 7, Fig. 12 plots the log-likelihood at the 100 uniform points in parameter space as well as the errors corresponding to the two head surrogates. Since the log-likelihood is based on heads at the 32 observation locations, Fig. 12 is directly comparable with Fig. 7. Comparing the two figures shows that the RMSE of the head surrogates is about two orders of magnitude smaller than that of the log-likelihood surrogates, when the log-likelihood function is evaluated by the two kinds of surrogates. Although the error of the head surrogates built for each head observation accumulates when the head surrogates are used to evaluate the log-likelihood function, the head surrogates still give dramatically more accurate estimation of the log-likelihood than the likelihood surrogates.

Table 2

Relative entropies (REn) between reference posterior distributions (obtained without using surrogates) and those obtained using the head surrogate (SG_H_AE) with absolute error and the head surrogate (SG_H_RE) with relative error. REn is evaluated for the marginal distributions of individual model parameters/predictions and for the joint distribution of all the parameters/predictions.

Model parameters	SG_H_AE		SG_H_RE	
	Marginal REn	Joint REn	Marginal REn	Joint REn
Recharge ratio	$1.28E-3$	$1.45E-2$	$2.96E-4$	$7.40E-3$
Boundary constant head	$5.38E-4$		$5.81E-4$	
Riverbed conductance	$5.98E-3$		$7.75E-4$	
Variance	$2.27E-3$		$3.04E-3$	
Correlation length	$9.28E-4$		$8.68E-4$	
Model predictions				
Discharge to river	$5.40E-4$	$6.20E-3$	$6.50E-5$	$1.89E-3$
Recharge from boundary constant head	$9.15E-4$		$7.10E-5$	
Discharge to drain	$1.61E-3$		$6.20E-4$	

4.4. Reasons that head surrogates outperforms log-likelihood surrogates

As discussed in Sections 4.2 and 4.3, the posterior distributions of model parameters and predictions based on the head surrogates are significantly more accurate than those based on the log-likelihood surrogates. In addition, the computational cost for building the head surrogates is also significantly lower than that for building the log-likelihood surrogates. The reasons are explained in this section. The analysis below is focused on SG_H_RE and SG_L_RE, because SG_H_RE slightly outperforms SG_H_AE and SG_L_RE significantly outperforms SG_L_AE.

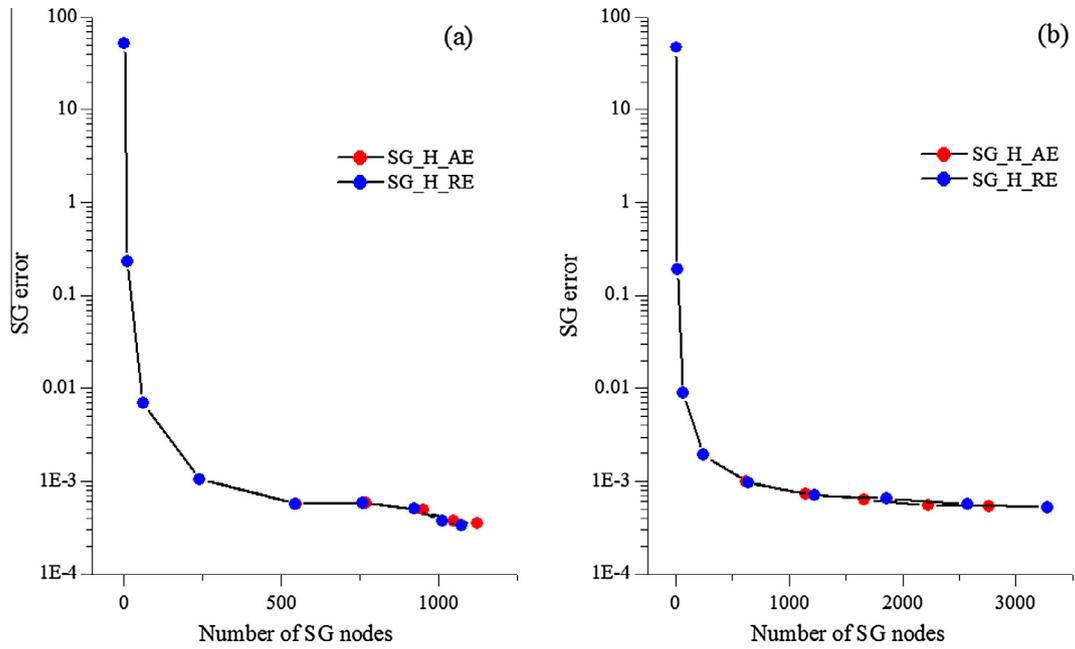


Fig. 9. Variation of average SG interpolation error with number of SG nodes at each level (dot) for surrogates SG_H_AE and SG_H_RE built for (a) observation 4 (O4) and (b) observation 23 (O23) marked in Fig. 3.

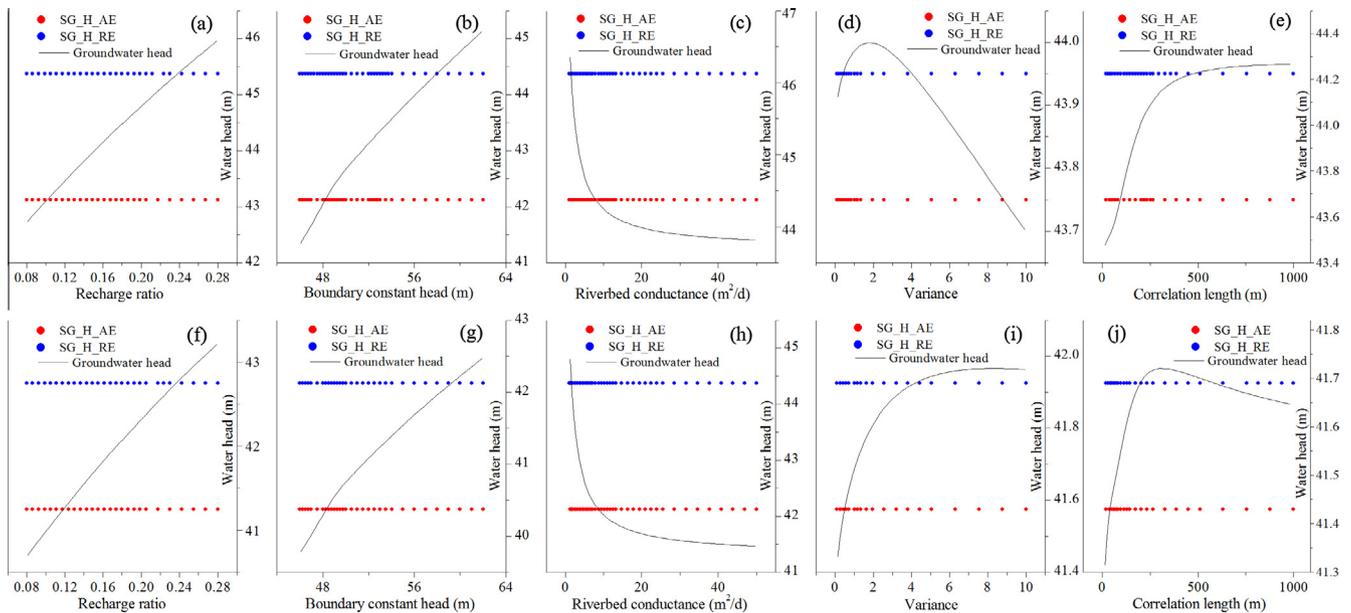


Fig. 10. Hydraulic head and SG nodes in the space of individual parameters (the other four parameters are fixed at their true values). Each dot denotes a SG node. The plots in the top and bottom rows are for observations O4 and O23 (marked in Fig. 3), respectively.

Fig. 13 plots the response surfaces of the likelihood, log-likelihood, and hydraulic head at observation location O4 in the parameter spaces of two parameter combinations. The left column of the Fig. 13 is for boundary constant head and river conductance, and the right column for variance and correlation length of hydraulic conductivity. The surfaces are obtained using brute-force MC simulations with 10,000 (100 points for each dimension) model executions. The response surfaces for other parameter combinations are similar to those shown in Fig. 13, and thus not shown. In Fig. 13(a) and (b), the surfaces are flat in most of the parameter

space but have sharp peaks. When the peak locations are unknown, it is likely that the peaks cannot be captured by SG surrogates, especially when the problem is high dimensional and/or the peak is sharp.

Fig. 13(c) and (d) shows that the log-likelihood surfaces are smoother than the likelihood surfaces (Fig. 13(a) and (b)) in the parameter space, and that there is no sharp peak on the log-likelihood surfaces. Therefore, building log-likelihood surrogates is preferred over building likelihood surrogates. However, building accurate log-likelihood surrogates is not trivial, because the

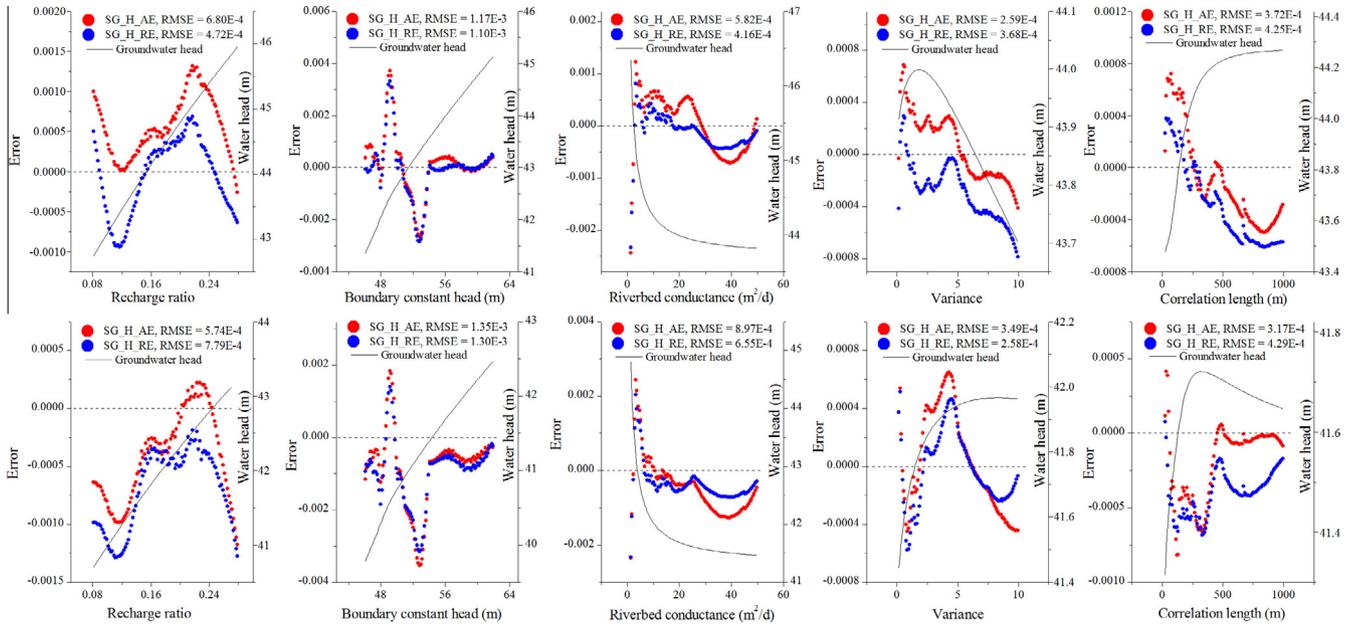


Fig. 11. Errors of SG_H_AE and SG_H_RE for evaluating hydraulic head at 100 uniform points in the space of the individual parameters (the other four parameters are fixed at their true values). RMSE is the root mean squared errors averaged over the 100 points. The plots in the top and borrow rows are for observations O4 and O23 (marked in Fig. 3), respectively.

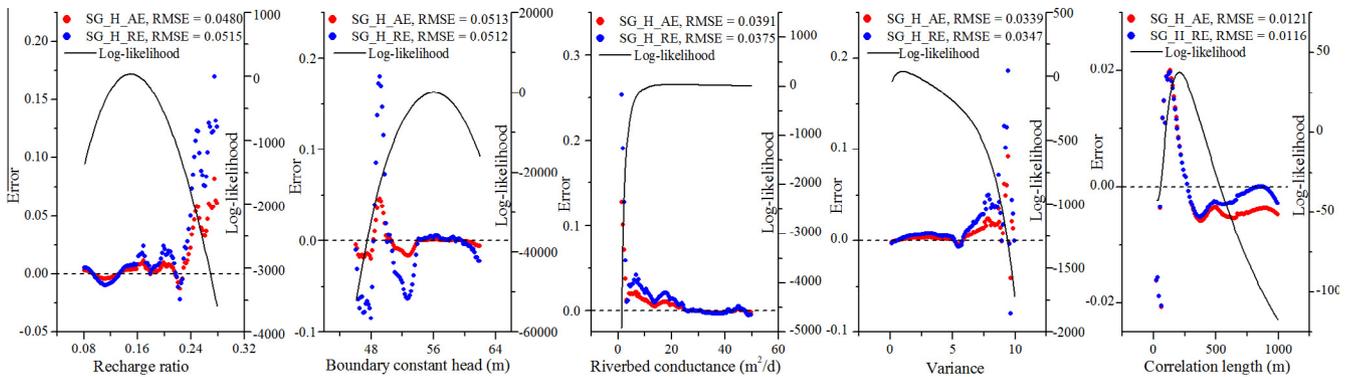


Fig. 12. Errors of SG_H_AE and SG_H_RE for evaluating log-likelihood at 100 uniform points in the space of the individual parameters (the other four parameters are fixed at their true values). RMSE is the root mean squared errors averaged over the 100 points. The plots in the top and borrow rows are for observations O4 and O23 (marked in Fig. 3), respectively.

log-likelihood surfaces vary substantially at the edges of the two parameter spaces. For example, the log-likelihood function still has a large range varying from 1.0E1 to -1.0E5. The drastic changes (especially at the edges of parameter spaces) make it difficult to develop accurate log-likelihood surrogates, especially when the absolute error is used for adaptation, as discussed in Section 4.2.

Fig. 13(e) and (f) for the response surface of hydraulic head do not show any drastic change in the head values. Instead, the surfaces are smooth, and the head variation is less than 15 m (Fig. 13(e)). Therefore, the head surrogates are more accurate than the log-likelihood surrogate. Comparing Figs. 5 and 9 show that the SG error for SG_H_RE is about 2–3 orders of magnitudes smaller than that for SG_L_RE. In addition, building the head surrogates is more computationally efficient, because the number of nodes for building SG_H_RE is 62.42% of that for building SG_L_RE and the results of SG_H_RE are more accurate than those of SG_L_RE. Therefore, building head surrogate is preferred over building likelihood surrogates.

Since examining response surfaces is vital to selecting the appropriate type of surrogate and error indicator, the response surfaces should be estimated before building SG surrogates. Direct estimates of a response surface can be done by parallel computing using state-of-the-art computing software such as PFIOTRAN (Hammond and Lichtner, 2010), if the needed computational resources are available. If computational resources for conducting parallel computing are limited, a rough approximation of the response surface is still needed. One may use global optimization methods (e.g., the dynamically dimensioned search algorithm of Tolson and Shoemaker (2007)) to allocate parameter optima and then build a rough response surface near the optima in parameter space. The approximated response surfaces should be sufficient for examining their sign, magnitude, and variation pattern in the parameter space. The computational cost of obtaining approximated response surfaces should be smaller than that of building SG surrogates. In addition, some of the model outputs for approximating the response surfaces can be recycled for building the surrogates.

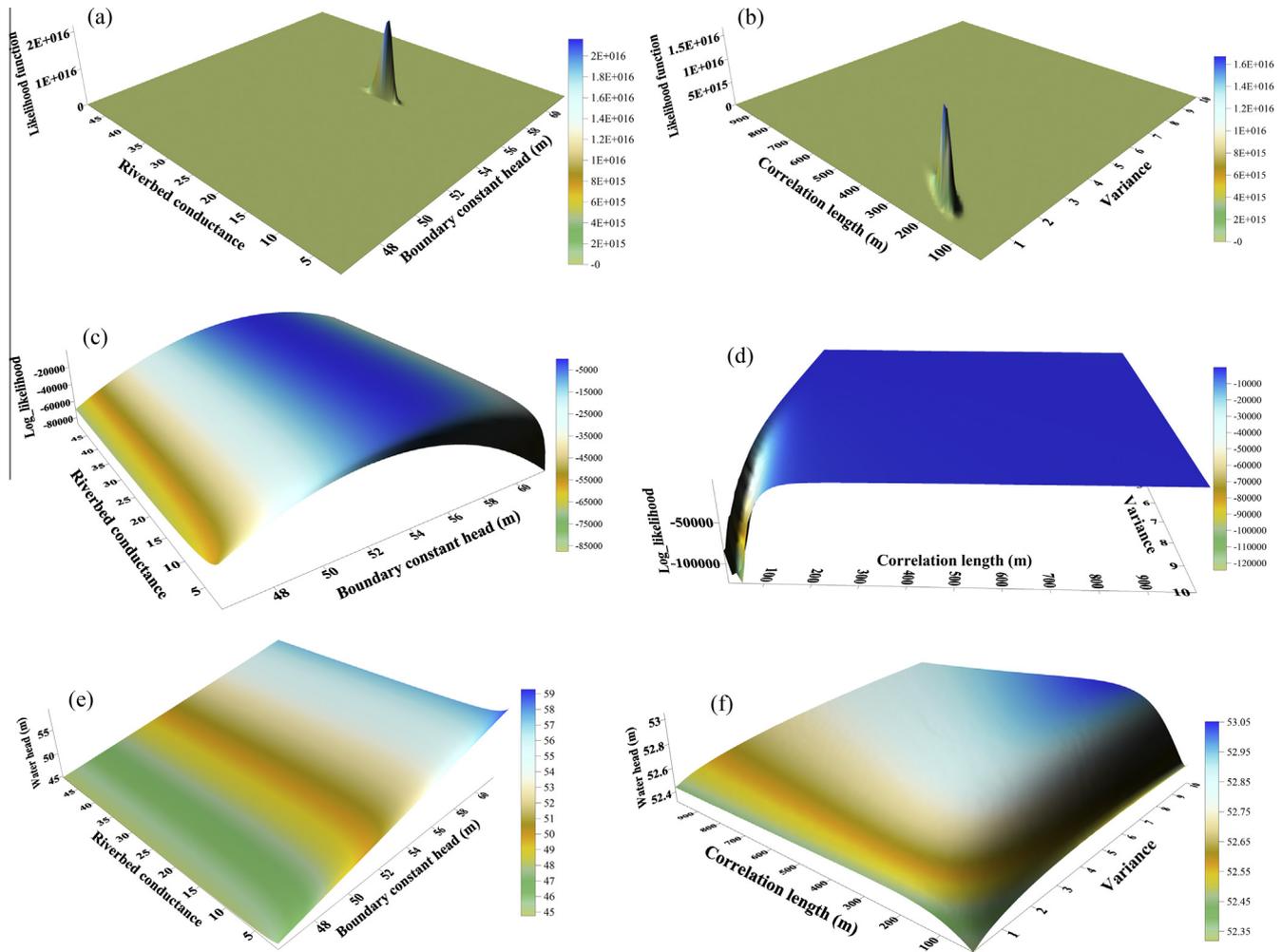


Fig. 13. Response surfaces of (a–b) likelihood function, (c–d) log-likelihood function, and (e–f) hydraulic head at observation 4 (O4 marked in Fig. 3) in the two-dimensional parameter space of (left column) boundary constant head and river conductance and (right column) variance and correlation length of hydraulic conductivity.

5. Conclusions and discussion

This paper addresses two important issues for applying SG methods to Bayesian uncertainty analysis. The first issue is to compare the two kinds of SG surrogates built for log-likelihood and state variables (hydraulic head in this study). The second issue is to compare two adaptive surrogates built using absolute and relative error indicators. A total of four surrogates are built; they are the log-likelihood surrogates based on absolute error indicator (SG_L_AE) and relative error indicator (SG_L_RE), and head surrogates based on absolute error indicator (SG_H_AE) and relative error indicator (SG_H_RE). The four surrogates are evaluated in terms of their accuracy and computational efficiency. The accuracy is evaluated by comparing the surrogate-based distributions of model parameters and predictions with those obtained using the original model without any surrogate. The efficiency is evaluated by comparing the number of model executions needed for building the surrogates.

Although the log-likelihood surrogates yield satisfactory distributions of model parameters and predictions, the head surrogates outperform the log-likelihood surrogates in terms of accuracy and efficiency for three reasons. The first reason is that the response surface of hydraulic head is significantly smoother than the response surface of log-likelihood in the parameter space. The SG surrogates of smooth functions are more accurate and easier to be built than those of non-smooth functions. The second reason is that the head variation is significantly smaller than the

log-likelihood variation in the parameter space. The small head variation is physically reasonable, because groundwater flow is governed by first principles regardless of the variation of parameter values. This is not the case for the log-likelihood, because there are no physical bounds for this statistical variable. The last reason is that the interpolation error of the individual head surrogates does not accumulate to be larger than the interpolation error of the log-likelihood surrogate, when the head surrogates are used to evaluate the log-likelihood functions. This is not surprising because the interpolation error of the individual head surrogate is three orders of magnitude smaller than the error of the log-likelihood function. The three reasons may change under different situations, and more research is warranted to further evaluate the log-likelihood and state-variable surrogates for different situations with more observations and more nonlinear models.

The log-likelihood surrogate based on the relative error indicator is significantly more accurate and computationally more efficient than the surrogate based on the absolute error indicator. The SG nodes of the two log-likelihood surrogates are significantly different, because of the large magnitude variation of the log-likelihood. When the absolute error is used, new SG nodes are placed in the areas where the absolute values of log-likelihood are large. While this reduces the SG interpolation error, the new SG nodes are useless for improving the accuracy of parameter distributions, especially for identifying high-probability regions on the parameter distributions, because the areas with large absolute values of log-likelihood do not correspond to the high-probability

regions. This problem is resolved by using the relative error. Because large relative errors occur in the areas where the absolute log-likelihood value is small, new SG nodes are added to such areas. Given that the regions with small absolute values of log-likelihood correspond to high probability regions on the parameter distributions, using the relative error indicator gives more accurate parameter distributions than using the absolute error indicator.

The head surrogates are insensitive to the error indicator, because the two head surrogates have similar number of SG nodes and the node locations in parameter space are also similar. This is attributed to the small magnitude variation of the hydraulic head. Recalling that the ratio between absolute error and relative error is hydraulic head and that the hydraulic head has small variation in the parameter space, the ratio is roughly a constant. In other words, the absolute and relative errors play the same roles for determining the nodes of adaptive SG. However, this conclusion may change if the state variable is not head but another variable such as flow rate or solute concentration which may change dramatically in parameter space. In this case, it is necessary to evaluate the response surface of the state variables to determine whether using the relative error indicator can help build more accurate state-variable surrogates. For example, using the relative error indicator defined in this study, more adaptive SG nodes are added to improve SG accuracy for low values of the state variables. It is necessary to evaluate whether this affects the accuracy of the posterior distributions of model parameters and predictions.

Although not discussed in this paper, building state-variable surrogates has another advantage over building log-likelihood surrogates, because the state-variable surrogates are more flexible for Bayesian updating. For example, when more data becomes available, the log-likelihood may change dramatically, and the log-likelihood surrogate has to be re-built. This is particularly true for data-worth analysis in which new data are determined gradually (e.g., Lu et al. (2012), Neuman et al. (2012), Zhang et al. (2015)). For the state-variable surrogates, it is only necessary to build new surrogates for the new data. In addition, once the state-variable surrogates are built, they can be used with any kind of likelihood functions. This is a critical feature for exploring appropriate likelihood functions as shown in the study of Shi et al. (2014) and for varying a given kind of likelihood function as shown in the study of Liu et al. (2016) for thermodynamic integration to evaluate Bayesian evidence.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (NSFC) – Xinjiang project (U1503282), NSFC projects (41302181, 41172207, 41571017, 51190091, and 51409161), Science and Technology Program of Jiangsu Province (No. BK20140080), and Program for New Century Excellent Talents in University of China (NCET-12-0262). The second author was supported by the DOE Early Career Award, DE-SC0008272. Part of the research was conducted when the first author was a visiting scholar at the Florida State University.

References

Barthelmann, V., Novak, E., Ritter, K., 2000. High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.* 12 (4), 273–288.

Box, G.E., Tiao, G.C., 1992. *Bayesian Inference in Statistical Analysis*, vol. 40. John Wiley & Sons.

Bungartz, H.-J., Griebel, M., 2004. Sparse grids. *Acta Numer.* 13, 147–269.

Cover, T.M., Thomas, J.A., 2006. *Elements of Information Theory*. John Wiley & Sons Inc.

Dai, H., Ye, M., 2015. Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation. *J. Hydrol.* 528, 286–300. <http://dx.doi.org/10.1016/j.jhydrol.2015.06034>.

Deutsch, C., Journel, A., 1998. *GSLIB Geostatistical Software Library and User's Guide*. Oxford Univ. Press, New York.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.* 50 (3), 2350–2375.

Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G., Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resour. Res.* 48.

Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. *Stat. Comput.* 16 (4), 339–354.

Hammond, G.E., Lichtner, P.C., 2010. Field-scale model for the natural attenuation of uranium at the Hanford 300 Area using high-performance computing. *Water Resour. Res.*, 46.

Harbaugh, A.W., 2005. The U.S. Geological Survey modular ground-water model—the Ground-Water Flow Process: U.S. Geological Survey Techniques and Methods 6–A16.

Hernandez, A.F., Neuman, S.P., Guadagnini, A., Carrera, J., 2006. Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media. *Water Resour. Res.* 42, W05425. <http://dx.doi.org/10.1029/2005WR004449>.

Hill, M.C., Kavetski, D., Clark, M., Ye, M., Arabi, M., Lu, D., Foglia, L., Mehl, S., 2015. Practical use of computationally frugal model analysis methods. *Groundwater*.

Klimke, W.A., 2006. *Uncertainty Modeling using Fuzzy Arithmetic and Sparse Grids*. University Stuttgart, Stuttgart, 136 pp.

Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resour. Res.* 48.

Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage MCMC simulation and polynomial chaos expansion. *Water Resour. Res.* 49, 2664–2682. <http://dx.doi.org/10.1002/wrcr.20226>.

Lee, Y.K., Park, B.U., 2006. Estimation of Kullback–Leibler divergence by local likelihood. *Ann. Inst. Stat. Math.* 58 (2), 327–340.

Liao, Q., Zhang, D., 2013. Probabilistic collocation method for strongly nonlinear problems: 1. Transform by location. *Water Resour. Res.* 49, 7911–7928. <http://dx.doi.org/10.1002/2013WR014055>.

Lin, G., Tartakovsky, A.M., 2009. An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. *Adv. Water Resour.* 32 (5), 712–722.

Lin, G., Tartakovsky, A.M., 2010. Numerical studies of three-dimensional stochastic Darcy's equation and stochastic advection-diffusion-dispersion equation. *J. Sci. Comput.* 43 (1), 92–117.

Lin, G., Tartakovsky, A.M., Tartakovsky, D.M., 2010. Uncertainty quantification via random domain decomposition and probabilistic collocation on sparse grids. *J. Comput. Phys.* 229 (19), 6995–7012.

Liu, P.G., Elshali, A.S., Ye, M., Beerli, P., Zeng, X.K., Lu, D., Tao, Y.Z., 2016. Evaluating marginal likelihood with thermodynamic integration method and comparison with several other numerical methods. *Water Resour. Res.* (in press).

Lu, D., Ye, M., Meyer, P.D., Curtis, G.P., Shi, X.Q., Niu, X.F., Yabusaki, S.B., 2013. Effects of error covariance structure on estimation of model averaging weights and predictive performance. *Water Resour. Res.* 49 (9), 6029–6047.

Lu, D., Ye, M., Neuman, S.P., Xue, L., 2012. Multimodel Bayesian analysis of data-worth applied to unsaturated fractured tuffs. *Adv. Water Resour.* 35, 69–82.

Ma, X., Zabarab, N., 2009a. An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. *Inverse Probl.* 25 (3).

Ma, X., Zabarab, N., 2009b. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comput. Phys.* 228 (8), 3084–3113.

Marzouk, Y.M., Najm, H.B., Rahn, R.A., 2007. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.* 224 (2), 560–586.

Marzouk, Y.M., Xiu, D., 2009. A stochastic collocation approach to bayesian inference in inverse problems. *Commun. Comput. Phys.* 6, 826–847.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. *Water Resour. Res.*, 45.

Neuman, S.P., Xue, L., Ye, M., Lu, D., 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. *Adv. Water Resour.* 36, 75–85.

Perez-Cruz, Fernando, 2008. Kullback–Leibler divergence estimation of continuous distributions. In: 2008 IEEE International Symposium on Information Theory Proceedings, 1–6, 1666–1670 pp.

Petvipusit, K.R., Elsheikh, A.H., Laforce, T.C., King, P.R., Blunt, M.J., 2014. Robust optimisation of CO₂ sequestration strategies under geological uncertainty using adaptive sparse grid surrogates. *Comput. Geosci.* 18 (5), 763–778.

Pfluger, D., 2010. *Spatially Adaptive Sparse Grids for High-Dimensional Problems*. Technische Universität München, München, 194 pp.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. *Water Resour. Res.*, 48.

Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resour. Res.* 44 (12), W12418.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.*, 46.

- Shi, L.S., Yang, J.Z., 2009. Qualification of uncertainty for simulating solute transport in the heterogeneous media with sparse grid collocation method. *J. Hydrodyn* 21 (6), 779–789.
- Shi, X.Q., Ye, M., Curtis, G.P., Miller, G.L., Meyer, P.D., Kohler, M., Yabusaki, S., Wu, J. C., 2014. Assessment of parametric uncertainty for groundwater reactive transport modeling. *Water Resour. Res.* 50 (5), 4416–4439.
- Smith, T., Sharma, A., Marshall, L., Mehrotra, R., Sisson, S., 2010. Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. *Water Resour. Res.*, 46.
- Stoyanov, M., 2013a. Hierarchy-Direction Selective Approach for Locally Adaptive Sparse Grids. Oak Ridge National Laboratory, Tennessee.
- Stoyanov, M., 2013b. User Manual: TASMANIAN Sparse Grids. Oak Ridge National Laboratory, Tennessee.
- Tartakovsky, D.M., 2013. Assessment and management of risk in subsurface hydrology: a review and perspective. *Adv. Water Resour.* 51, 247–260.
- Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.* 43 (1), Artn W01413.
- Vrugt, J.A., Ter Braak, C.J.F., 2011. DREAM((D)): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrol. Earth Syst. Sci.* 15 (12), 3701–3713.
- Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.*, 44.
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Int. J. Nonlinear Sci. Numer.* 10 (3), 273–290.
- Wang, Q., Kulkarni, S.R., Verdu, S., 2009. Divergence estimation for multidimensional densities via k-nearest-neighbor distances. *IEEE Trans. Inform. Theory* 55 (5), 2392–2405.
- Webster, C.G., Zhang, G.N., Gunzburger, M., 2014. An adaptive sparse-grid iterative ensemble Kalman filter approach for parameter field estimation. *Int. J. Comput. Math.* 91 (4), 798–817.
- Xiu, D., Hesthaven, J.S., 2005. High-order collocation methods for differential equations with random inputs. *SIAM. J. Sci. Comput.* 27 (3), 1118–1139.
- Zeng, L.Z., Shi, L.S., Zhang, D.X., Wu, L.S., 2012. A sparse grid based Bayesian method for contaminant source identification. *Adv. Water Resour.* 37, 1–9.
- Zhang, D.X., Shi, L.S., Chang, H.B., Yang, J.Z., 2010. A comparative study of numerical approaches to risk assessment of contaminant transport. *Stoch. Environ. Res. Risk A* 24 (7), 971–984.
- Zhang, G.N., Lu, D., Ye, M., Gunzburger, M., Webster, C., 2013. An adaptive sparse-grid high-order stochastic collocation method for Bayesian inference in groundwater reactive transport modeling. *Water Resour. Res.* 49 (10), 6871–6892.
- Zhang, J., Zeng, L., Chen, C., Chen, D., Wu, L., 2015. Efficient Bayesian experimental design for contaminant source identification. *Water Resour. Res.* 51 (1), 576–598.