AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2013WR013755

Assessment of parametric uncertainty for groundwater reactive transport modeling

Xiaoqing Shi[1,2], Ming Ye[2], Gary P. Curtis[3], Geoffery L. Miller[2], Philip D. Meyer[4], Matthias Kohler[3], Steve Yabusaki[4], and Jichun Wu[1]

[1]Key Laboratory of Surficial Geochemistry, Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China, [2]Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA, [3]U.S. Geological Survey, Menlo Park, California, USA, [4]Pacific Northwest National Laboratory, Richland, Washington, USA

Key Points:
- Complexity of response surface due to model nonlinearity and calibration data
- Non-Gaussian residuals require non-Gaussian likelihood function
- Formal generalized likelihood function gives better uncertainty quantification

Correspondence to:
M. Ye,
mye@fsu.edu

Citation:
Shi, X., M. Ye, G. P. Curtis, G. L. Miller, P. D. Meyer, M. Kohler, S. Yabusaki, and J. Wu (2014), Assessment of parametric uncertainty for groundwater reactive transport modeling, *Water Resour. Res.*, *50*, doi:10.1002/2013WR013755.

Received 28 FEB 2013
Accepted 6 MAY 2014
Accepted article online 10 MAY 2014

Abstract The validity of using Gaussian assumptions for model residuals in uncertainty quantification of a groundwater reactive transport model was evaluated in this study. Least squares regression methods explicitly assume Gaussian residuals, and the assumption leads to Gaussian likelihood functions, model parameters, and model predictions. While the Bayesian methods do not explicitly require the Gaussian assumption, Gaussian residuals are widely used. This paper shows that the residuals of the reactive transport model are non-Gaussian, heteroscedastic, and correlated in time; characterizing them requires using a generalized likelihood function such as the formal generalized likelihood function developed by Schoups and Vrugt (2010). For the surface complexation model considered in this study for simulating uranium reactive transport in groundwater, parametric uncertainty is quantified using the least squares regression methods and Bayesian methods with both Gaussian and formal generalized likelihood functions. While the least squares methods and Bayesian methods with Gaussian likelihood function produce similar Gaussian parameter distributions, the parameter distributions of Bayesian uncertainty quantification using the formal generalized likelihood function are non-Gaussian. In addition, predictive performance of formal generalized likelihood function is superior to that of least squares regression and Bayesian methods with Gaussian likelihood function. The Bayesian uncertainty quantification is conducted using the differential evolution adaptive metropolis ($DREAM_{(ZS)}$) algorithm; as a Markov chain Monte Carlo (MCMC) method, it is a robust tool for quantifying uncertainty in groundwater reactive transport models. For the surface complexation model, the regression-based local sensitivity analysis and Morris- and $DREAM_{(ZS)}$-based global sensitivity analysis yield almost identical ranking of parameter importance. The uncertainty analysis may help select appropriate likelihood functions, improve model calibration, and reduce predictive uncertainty in other groundwater reactive transport and environmental modeling.

1. Introduction

Groundwater reactive transport modeling is a vital tool for analyzing and managing the subsurface environment [*Davis et al*., 2004a; *Steefel et al*., 2005; *Scheibe et al*., 2008]. Because model predictions are inherently uncertain, accurate quantification of predictive uncertainty is necessary to avoid poorly managed and overly costly remediation and monitoring. Predictive uncertainty is caused by propagation of various sources of uncertainty in model structures, parameters, driving forces, and data used in modeling [*Neuman*, 2003; *Wagener and Gupta*, 2005; *Srinivasan et al*., 2007; *Meyer et al*., 2007; *Clark et al*., 2011; *Refsgaard et al*., 2012; *Gupta et al*., 2012; *Tartakovsky*, 2013]. Parametric uncertainty is due to imperfect knowledge of model parameters, initial and boundary conditions, and/or driving forces. Model structure uncertainty may be manifested by different plausible conceptualizations and/or mathematical descriptions of the transport processes (e.g., the traditional advection-dispersion model versus other alternative models, as discussed in *Srinivasan et al*. [2007] and *Tang et al*. [2009]) and the geochemical reaction processes (e.g., different formulations of surface complexation models developed for simulating uranium adsorption in batch experiments [*Davis et al*., 2004b], column experiments [*Kohler et al*., 1996], and tracer experiments [*Davis et al*., 2004b; *Curtis and Davis*, 2006]). Quantification of model uncertainty can be pursued using either model selection methods [e.g., *Kohler et al*., 1996; *Davis et al*., 2004b; *Matott and Rabideau*, 2008a] or model averaging methods [e.g., *Neuman*, 2003; *Poeter and Anderson*, 2005; *Ye et al*., 2004, 2008, 2010a, 2010b; *Lu et al.,* 2011]. These methods require quantifying parametric uncertainty so that model selection or model averaging is

conducted with consideration of parametric uncertainty, instead of relying on a single parameter value or a small set of discrete parameter values.

This study was focused on quantification of parametric uncertainty in groundwater reactive transport modeling, with particular attention paid to uranium reactive transport modeling using surface complexation models. In comparison with parametric uncertainty analysis in groundwater flow and nonreactive transport modeling (see review articles by *Neuman* [1973], *Neuman and Yakowitz* [1979], *Yeh and Yoon* [1981], *Yeh* [1986], *Kitanidis* [1986], *Beck* [1987], *Carrera* [1993], *Liu and Gupta* [2007], and *Matott et al*. [2009]), quantification of parametric uncertainty in reactive transport modeling has its own challenges [*Leavitt et al*., 2011]. In particular, reactive transport models are nonlinear with respect to the parameters due to nonlinear reaction equations and coupling of advective, dispersive, and reactive processes [*Cabaniss*, 1999; *Mugunthan et al*., 2005]. The nonlinearities lead to complex surfaces of objective functions (e.g., least squares or likelihood) with multiple local minima, as shown in *Matott and Rabideau* [2008b]. For such objective function surfaces, gradient-based parameter estimation algorithms (e.g., the Gauss-Levenberg-Marquardt method implemented in UCODE_2005 [*Poeter et al*., 2005] and PEST [*Doherty*, 2010]) may perform poorly, for example, by terminating at a local minimum. In addition, due to the nonlinearities, the probability distributions of model parameters and outputs may be non-Gaussian with multiple modes and/or long tails [*Cabaniss*, 1997; *Denison and Carnier-Laplace*, 2005; *Leavitt et al*., 2011]. This puts question in the accuracy of uncertainty quantification methods that either explicitly [*Srinivasan et al*., 2007; *Liu et al*., 2008; *Tartakovsky et al*., 2009] or implicitly [*Dai and Samper*, 2004] assume Gaussian parameters of groundwater reactive transport models. A literature review by *Lu et al*. [2012] showed that regression confidence intervals based on the Gaussian assumption of model parameters may differ from the Bayesian credible intervals that do not require the Gaussian assumption. However, the numerical studies of *Lu et al*. [2012] and *Shi et al*. [2012] for flow models showed that discrepancy between the two kinds of intervals is small. For groundwater reactive transport models, it is unknown to what extent the Gaussian assumption affects accuracy of uncertainty quantification, and this is one of motivating questions in this study.

Markov Chain Monte Carlo (MCMC) methods may be a solution to the challenges of model nonlinearity and parameter non-Gaussianity [e.g., *Marshall et al*., 2004; *Gallagher and Doherty*, 2007; *Smith and Marshall*, 2008; *Vrugt et al*., 2008a, 2009a, 2009b]. As a Bayesian technique, MCMC can be directly applied to nonlinear models without any approximation of model linearization, and it does not require any assumptions on the form of the parameter probability distributions. Instead, it estimates the posterior parameter distributions and takes into account model nonlinearity in the estimation. *Smith and Marshall* [2008] demonstrated that MCMC techniques are able to identify multimodal parameter distributions of hydrologic models. Building on the Differential Evolution-Markov Chain (DE-MC) method of *Ter Braak* [2006], *Vrugt et al*. [2008a, 2009a] developed the Differential Evolution Adaptive Metropolis (DREAM) algorithm; its efficiency in estimating multimodal non-Gaussian parameter distributions was demonstrated in *Vrugt et al*. [2009b] for simple synthetic test problems. While DREAM appears to be a promising tool, it is unknown whether DREAM can address the above challenges in groundwater reactive transport modeling, because it has not been used in such problems. The application of DREAM to groundwater flow and nonreactive transport modeling in *Keating et al*. [2010] suggested that DREAM may not be practical for high-dimensional and computationally expensive models, because it requires tens to hundreds of thousands of model runs. While this problem was resolved in *Laloy and Vrugt* [2012] using the multitry $\text{DREAM}_{(ZS)}$ algorithm, applications of DREAM and $\text{DREAM}_{(ZS)}$ to groundwater modeling, especially reactive transport modeling, are still limited.

Another challenge to quantification of parametric uncertainty is the selection of likelihood functions of parameters or, equivalently, probability distributions of residuals (difference between observations and corresponding model simulations). Informal and formal likelihood functions have been used in Bayesian uncertainty quantification or as the objective functions in maximum likelihood and regression theories. Informal likelihood functions are used when exact expressions of likelihood functions are unknown. *Beven et al*. [2008] advocated using informal generalized likelihood function (or likelihood measures) that are not necessarily related to the residual probabilities. *Rubin et al*. [2010] and *Over et al*. [2013] used a nonparametric likelihood function in the method of anchored distributions developed to quantify model and parametric uncertainty. *Zhang et al*. [2013] used informal exponential likelihood function in their Bayesian uncertainty analysis for groundwater reactive transport modeling. Formal likelihood functions are used more widely than informal ones, and the Gaussian likelihood function is the most popular one. However, it may not be

always the most appropriate way to characterize the residual probability, and other likelihood functions are needed. For example, *Chen et al*. [2010] used the *t* distribution because its heavier tails are more robust than Gaussian distribution to fit residuals of surface seismic refraction data. *Schoups and Vrugt* [2010] developed a formal generalized likelihood function that uses the skew exponential power (SEP) distribution, and it is probably the most inclusive formal likelihood function to describe residuals that are correlated, heteroscedastic, and non-Gaussian. While the formal generalized likelihood function of *Schoups and Vrugt* [2010] has been used for hydrologic models, it is unknown whether it is appropriate for characterizing residual distributions of groundwater reactive transport models, which is explored in this study.

To address the above unresolved questions in parametric uncertainty quantification for groundwater reactive transport modeling, a numerical study was conducted for a surface complexation model (SCM) developed by *Kohler et al*. [1996] to simulate uranium reactive transport in column experiments. The SCM is ideal for this numerical exploration and comparative study for the following two reasons: (1) the SCM has only four unknown parameters and is thus not high-dimensional, and (2) the SCM is not computationally expensive because each forward model run takes about 3–5 min using a single processor. The following three kinds of parametric uncertainty analysis were performed:

1. Regression-based uncertainty quantification implemented in UCODE_2005. The quantification uses approximate linearization of nonlinear models and assumes Gaussian residuals; the Gaussian assumption leads to asymptotic Gaussian distributions of model parameters and predictions.

2. Bayesian uncertainty quantification implemented in DREAM$_{(ZS)}$. The quantification does not require either linearizing nonlinear models or Gaussian parameters, but uses Gaussian likelihood function.

3. The same Bayesian uncertainty quantification of (2) but uses the formal generalized likelihood function of *Schoups and Vrugt* [2010] (not the Gaussian likelihood function).

The results of the three kinds of uncertainty analysis were compared to investigate:

1. Whether the parameter distributions of the three methods are different. The theoretical Gaussian parameter distributions of the regression methods are compared with the parameter histograms obtained from the DREAM$_{(ZS)}$ simulations using the Gaussian and formal generalized likelihood functions.

2. Whether the confidence intervals of regression methods and the credible intervals of the Bayesian methods (using the two different likelihood functions) are different. A cross-validation study was conducted to evaluate predictive performance of the confidence and credible intervals.

The rest of the paper is organized as follows. Section 2 provides necessary details about the theories of nonlinear regression and Bayesian uncertainty analysis. Section 3 starts with a brief introduction of the SCM and then presents the results of parameter estimation, uncertainty quantification, and predictive analysis obtained using UCODE_2005 and DREAM$_{(ZS)}$. In addition, the MCMC results are used to understand sensitivity of simulated uranium concentration to the SCM parameters. Major findings of this study are summarized in section 4.

2. Regression and Bayesian Methods for Parametric Uncertainty Analysis

A brief overview of the regression and Bayesian techniques is given in this section; detailed discussions of the methods are provided by *Draper and Smith* [1981], *Hill and Tiedeman* [2007], *Box and Tiao* [1992], *Gelman and Rubin* [1992], and *Draper* [2007]. A thorough comparison (theoretical and numerical) of the confidence intervals of nonlinear regression methods and credible intervals of Bayesian approaches in the context of groundwater flow modeling can be found in *Lu et al*. [2012].

2.1. Nonlinear Regression Methods

The statistical model used for regression-based nonlinear model calibration and uncertainty quantification can be expressed as

$$\mathbf{y}=f(\boldsymbol{\beta})+\boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is a vector of n observations, f is a nonlinear model with respect to its parameters, $\boldsymbol{\beta}$ (a vector of p model parameters), and $\boldsymbol{\varepsilon}$ is a vector of n errors. The errors are random and assumed to follow a multivariate

normal distribution, $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \mathbf{C}_\varepsilon)$. The covariance matrix, $\mathbf{C}_\varepsilon$, is often represented as $\mathbf{C}_\varepsilon = \sigma^2 \boldsymbol{\omega}^{-1}$, where $\boldsymbol{\omega}$ is a $n \times n$ known weight matrix and σ^2 is a scalar, possibly unknown but can be estimated. A common practice in nonlinear regression is to linearize the model by expanding it in a Taylor series and retaining the first two terms, i.e., $f(\hat{\mathbf{b}}) \approx f(\boldsymbol{\beta}) + \mathbf{X}_{\boldsymbol{\beta}}(\hat{\mathbf{b}} - \boldsymbol{\beta})$, where $\hat{\mathbf{b}}$ is the estimate of $\boldsymbol{\beta}$ obtained by minimizing the generalized least squares objective function

$$S(\mathbf{b}) = [\mathbf{y} - f(\mathbf{b})]^T \boldsymbol{\omega} [\mathbf{y} - f(\mathbf{b})], \tag{2}$$

and $\mathbf{X}_{\boldsymbol{\beta}} = \left[\frac{\partial f}{\partial \mathbf{b}}\right]_{\mathbf{b}=\boldsymbol{\beta}}$ is a $n \times p$ sensitivity matrix [*Seber and Wild*, 2003, pp. 23–24]. The objective function is also referred to as sum of squared weighted residuals (*SSWR*), where residuals are defined as the difference between observed and corresponding simulated values. The parameter estimates, $\hat{\mathbf{b}}$, follow asymptotically the multivariate normal distribution [*Seber and Wild*, 2003, p. 24; *Hill and Tiedeman*, 2007, p. 398]

$$\hat{\mathbf{b}} \sim N_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}_{\boldsymbol{\beta}}^T \boldsymbol{\omega} \mathbf{X}_{\boldsymbol{\beta}})^{-1}), \tag{3}$$

because of the normality assumption of the errors. If the nonlinear prediction function $g(\hat{\mathbf{b}})$ is also approximated to the first order by $g(\hat{\mathbf{b}}) \approx g(\boldsymbol{\beta}) + \mathbf{Z}_{\boldsymbol{\beta}}^T(\hat{\mathbf{b}} - \boldsymbol{\beta})$, then $g(\hat{\mathbf{b}})$ follows asymptotically the normal distribution

$$g(\hat{\mathbf{b}}) \sim N(g(\boldsymbol{\beta}), \sigma^2 \mathbf{Z}_{\boldsymbol{\beta}}^T (\mathbf{X}_{\boldsymbol{\beta}}^T \boldsymbol{\omega} \mathbf{X}_{\boldsymbol{\beta}})^{-1} \mathbf{Z}_{\boldsymbol{\beta}}), \tag{4}$$

where $\mathbf{Z}_{\boldsymbol{\beta}}$ is the prediction sensitivity vector $\mathbf{Z}_{\boldsymbol{\beta}} = \left[\frac{\partial g}{\partial \boldsymbol{\beta}}\right]_{\boldsymbol{\beta}=\hat{\mathbf{b}}}$ [*Seber and Wild*, 2003, p. 192]. Thus, the $(1-\alpha)\times$ 100% *linear* confidence interval of $g(\boldsymbol{\beta})$ is

$$g(\hat{\mathbf{b}}) \pm t_{1-\alpha/2}(n-p) [s^2 \mathbf{Z}_{\boldsymbol{\beta}}^T (\mathbf{X}_{\boldsymbol{\beta}}^T \boldsymbol{\omega} \mathbf{X}_{\boldsymbol{\beta}})^{-1} \mathbf{Z}_{\boldsymbol{\beta}}]^{1/2}, \tag{5}$$

where $t_{1-\alpha/2}(n-p)$ is a t statistic with significance level α and degrees of freedom $n-p$, and

$$s^2 = (\mathbf{y} - f(\hat{\mathbf{b}}))^T \boldsymbol{\omega} (\mathbf{y} - f(\hat{\mathbf{b}}))/(n-p) \tag{6}$$

is the estimate of σ^2. In practice, the sensitivity matrices $\mathbf{X}_{\boldsymbol{\beta}} = \left[\frac{\partial f}{\partial \mathbf{b}}\right]_{\mathbf{b}=\boldsymbol{\beta}}$ and $\mathbf{Z}_{\boldsymbol{\beta}} = \left[\frac{\partial g}{\partial \mathbf{b}}\right]_{\mathbf{b}=\boldsymbol{\beta}}$ are approximated by $\hat{\mathbf{X}}_{\hat{\mathbf{b}}} = \left[\frac{\partial f}{\partial \mathbf{b}}\right]_{\mathbf{b}=\hat{\mathbf{b}}}$ and $\hat{\mathbf{Z}}_{\hat{\mathbf{b}}} = \left[\frac{\partial g}{\partial \mathbf{b}}\right]_{\mathbf{b}=\hat{\mathbf{b}}}$, respectively, with $\boldsymbol{\beta}$ replaced by its estimate, $\hat{\mathbf{b}}$, obtained from the nonlinear regression [*Seber and Wild*, 2003, p. 191].

Estimating the *nonlinear* confidence intervals does not require linearizing the model. The intervals are determined as the minimum and maximum of model predictions intersecting with a confidence region of model parameters [*Vecchia and Cooley*, 1987]. The region is defined as the set of parameter values whose corresponding objective function values satisfy [*Christensen and Cooley*, 1999; *Cooley*, 2004]

$$S(\mathbf{b}) \leq S(\hat{\mathbf{b}}) \left[\frac{1}{n-p} t_{\alpha/2}^2 (n-p) + 1\right]. \tag{7}$$

This parameter region contains the true model parameter with approximate probability of $(1-\alpha)\times 100\%$ for the errors, $\boldsymbol{\varepsilon}$, defined in equation (1). Estimating the nonlinear confidence interval requires the following assumptions: (1) the model accurately represents the system, (2) model predictions, g($\boldsymbol{\beta}$), are sufficiently monotonic, (3) there is a single minimum in the objective function, and (4) the residuals are multivariate normal distributed to obtain a valid critical value, and (5) model intrinsic nonlinearity is small [*Cooley and Naff*, 1990; *Hill and Tiedeman*, 2007; *Lu et al.*, 2012]. If the assumptions are not satisfied, equation (7) may not define the objective function value associated with the designated $1-\alpha$ confidence level and the nonlinear intervals may be inaccurate.

Calculation of the linear and nonlinear confidence intervals relies on $\hat{\mathbf{b}}$ obtained by minimizing equation (2). The optimization is conventionally conducted in groundwater modeling using the Gauss-Marquardt-

Levenberg (GML) method, which has been implemented in several popular codes of automated calibration including PEST [*Doherty*, 2005], UCODE_2005 [*Poeter et al*., 2005], iTOUGH2 [*Finsterle*, 2007; *Finsterle and Zhang*, 2011]. However, the gradient-based GML method may not be able to find the global minimum of the objective function; it may be terminated in regions of local minima [*Chen et al*., 2008]. Whether the UCODE-2005 parameter estimates are global optima was investigated in this study using global optimization methods such as DREAM$_{(ZS)}$ and other techniques discussed in section 3.

2.2. Bayesian Method

Different from the nonlinear regression methods that treat model parameters as deterministic variables but parameter estimates as random variables, the Bayesian methods treat model parameters themselves as random variables and characterize parametric uncertainty using their posterior distribution obtained from the Bayes' theorem

$$p(\boldsymbol{\beta}|\mathbf{y}) = \frac{p(\mathbf{y}|\beta)p(\boldsymbol{\beta})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\beta)p(\boldsymbol{\beta})}{\int p(\mathbf{y}|\beta)p(\boldsymbol{\beta})d\boldsymbol{\beta}}, \tag{8}$$

where $p(\boldsymbol{\beta}|\mathbf{y})$ is the posterior parameter distribution conditioned on data $\mathbf{y}$, $p(\boldsymbol{\beta})$ is the prior distribution, and $p(\mathbf{y}|\boldsymbol{\beta})$ is the likelihood function that will be discussed in detail in section 2.3. The posterior distributions are always obtained numerically by various methods, and one of them is the MCMC techniques, which do not require linearizing the models and are able to infer non-Gaussian posterior distributions. After the posterior parameter distributions are obtained, the predictive uncertainty is quantified by first drawing parameter samples from the distribution and then executing the prediction model with the samples. The $(1-\alpha)\times 100$ % credible interval is determined via

$$\int_l^u p(g(\boldsymbol{\beta})|\mathbf{y})\mathrm{d}g(\boldsymbol{\beta}) = 1-\alpha, \tag{9}$$

where $p(g(\boldsymbol{\beta})|\mathbf{y})$ is the posterior distribution of $g(\boldsymbol{\beta})$ conditioned on data $\mathbf{y}$. In this study, the credible interval limits, l and u, are determined using the equal-tailed method as [*Casella and Berger*, 2002]

$$p(g(\boldsymbol{\beta}) \le l|\mathbf{y}) = p(g(\boldsymbol{\beta}) \ge u|\mathbf{y}) = \alpha/2. \tag{10}$$

For example, when $\alpha = 0.05$, the 95% credible interval is determined by the 2.5% and 97.5% percentiles of model predictions. Other methods of estimating the credible intervals (e.g., highest posterior density interval) are also available [*Box and Tiao*, 1992; *Chen and Shao*, 1999; *Casella and Berger*, 2002], but were not used in this work.

Efficiently and accurately estimating $p(\boldsymbol{\beta}|\mathbf{y})$ is a daunting challenge to MCMC techniques, especially when $p(\boldsymbol{\beta}|\mathbf{y})$ is multimodal. Recently, *Ter Braak and Vrugt* [2008] and *Laloy and Vrugt* [2012] developed the DREAM$_{(ZS)}$ algorithm to generate samples of the posterior distribution. DREAM$_{(ZS)}$ uses Differential Evolution Adaptive Metropolis (DREAM) [*Vrugt et al*., 2008a, 2009a, 2009b] as its main building block, which runs multiple Markov chains from different starting points in parallel for global search in the parameter space to make it possible to search multiple parameter regions corresponding to multiple modes. This is explained quantitatively by the equation below to generate a candidate parameter point [*Vrugt et al*., 2009b]

$$\mathbf{z}^i = \mathbf{x}_{t-1}^i + \gamma(\mathbf{x}_{t-1}^{r_1} - \mathbf{x}_{t-1}^{r_2}) + \mathbf{e} \;\; r_1 \ne r_2 \ne i, \tag{11}$$

where $\mathbf{z}^i$ is the candidate point, $\mathbf{x}^i$, $\mathbf{x}^{r_1}$, and $\mathbf{x}^{r_2}$ are different Markov chains, γ is a scale factor, and $\mathbf{e}$ is random error following a normal distribution with small variance. Using equation (11) allows for direct jumps between different modes. In addition, the scale (γ) and orientation of the proposal distribution $(\mathbf{x}_{t-1}^{r_1} - \mathbf{x}_{t-1}^{r_2})$ are automatically tuned during the evolution to the posterior distribution. While DREAM requires at least N (the number of parameters) chains to be run in parallel, DREAM$_{(ZS)}$ generates candidate points from an archive of past states rather than the current locations of the N chains [*Laloy and Vrugt*, 2012]. This enables

use of a smaller N value to generate the jumps in equation (11) from the past states of the different chains and acceleration of convergence to a limiting distribution. More details of the DREAM$_{(ZS)}$ algorithm are provided by *Ter Braak and Vrugt* [2008] and *Laloy and Vrugt* [2012].

2.3. Gaussian and Formal Generalized Likelihood Functions

The Bayesian methods can incorporate any likelihood function, and a widely used one is the Gaussian likelihood function

$$p(\mathbf{y}|\boldsymbol{\beta}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|} \exp\left[-\frac{1}{2}(\mathbf{y}-f(\boldsymbol{\beta}))^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-f(\boldsymbol{\beta}))\right], \tag{12}$$

where $\boldsymbol{\Sigma}$ is the covariance function of residuals, $\mathbf{r}=\mathbf{y}-f(\boldsymbol{\beta})$. Analytical expressions of the posterior distribution are in general not available except for special cases [e.g., *Woodbury and Ulrich*, 2000; *Hou and Rubin*, 2005]. For the Gaussian likelihood, *Box and Tiao* [1992, p. 428] showed that, when using noninformative prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, the posterior parameter distribution follows:

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto [S(\boldsymbol{\beta})]^{-n/2}, \tag{13}$$

where n is number of observations, and

$$S(\boldsymbol{\beta}) = (\mathbf{y}-f(\boldsymbol{\beta}))^T(\mathbf{y}-f(\boldsymbol{\beta})) \tag{14}$$

is the sum of squared residuals (SSR), the objective function of ordinary least squares methods. Equations (13) and (14) are the connection between the shape of posterior distribution and the surface of SSR objective function. For example, multiple minima on a least square objective function may correspond to multiple modes of the parameter density function. This is illustrated in Figure 1 for a single parameter following a bimodal distribution (Figure 1a). As shown in Figure 1b, parameter samples corresponding to the high peak of the distribution may produce model simulations $f(\boldsymbol{\beta})$ close to observation, $\mathbf{y}$; model simulations produced by samples corresponding to the low peak deviate more from the observation; model simulations corresponding to other parts of the distribution deviate even more from the observation. As a result, Figure 1c shows that the objective function has two minima, the global minimum corresponding to the high peak and the local minimum to the low peak. In line with this, if the probability density function of the parameter has multiple modes, the objective function will have multiple minima. The relations between the modes of parameter distributions and minima of objective functions are proved by *Box and Tiao* [1992, p. 489] for linear models. Although such a proof does not exist for nonlinear models, if multiple minima are observed during regression-based model calibration, it is an indicator that model parameters may be non-Gaussian.

The formal generalized likelihood function of *Schoups and Vrugt* [2010] was considered in this study, because it explicitly considers heteroscedasticity, temporal correlation, and non-Gaussianity in residuals that were found in this study (see section 3.3). The formal generalized likelihood function characterizes the residual, r_t, using the following model [*Schoups and Vrugt*, 2010]

$$\Phi_p(B)r_t = \sigma_t a_t \ \text{ with } \ a_t \sim SEP(0, 1, \xi, \theta), \tag{15}$$

where $\Phi_p(B) = 1-\sum_{i=1}^{p}\phi_i B^i$ is an autoregressive polynomial with p autoregressive parameters, ϕ_i, B is the backshift operator, $B^i r_t = r_{t-i}$, σ_t is standard deviation at time t, and a_t is an i.i.d. random error with zero mean and unit standard deviation, described by a skew exponential power (SEP) density defined below with parameters ξ and θ to account for non-Gaussianity. Following *Schoups and Vrugt* [2010], heteroscedasticity was explicitly accounted for using the linear model

$$\sigma_t = \sigma_0 + \sigma_1 E_t, \tag{16}$$

where E_t is the mean simulation, and the coefficients, σ_0 and σ_1, are determined as nuisance parameters during the MCMC simulation (the nuisance parameters are not of immediate interest but are necessary to

estimate the reactive transport parameters of this study). Residual non-Gaussianity was accounted for using the SEP(0,1,ξ,θ) density function [*Schoups and Vrugt*, 2010]

$$p(a_t|\xi,\theta)=\frac{2\sigma_\xi}{\xi+\xi^{-1}}\omega_\xi \exp\left[-c_\theta|a_{\xi,t}|^{2/(1+\theta)}\right], \tag{17}$$

where $a_{\xi,t}=\xi^{-sign(\mu_\xi+\sigma_\xi a_t)}(\mu_\xi+\sigma_\xi a_t)$, and values for the nuisance parameters, μ_ξ, σ_ξ, c_θ, and ω_θ, are computed as a function of skewness parameter, ξ, and kurtosis parameter, θ, as described in *Schoups and Vrugt* [2010]. These nuisance parameters are also estimated during the MCMC simulation. The MCMC simulation was conducted using the DREAM$_{(ZS)}$ code developed by *Vrugt* [2009] that includes the formal generalized likelihood function as a choice.

3. Surface Complexation Model, Numerical Experiments, and Result Analyses

This section starts by introducing the surface complexation model (SCM) and its uncertainty and nonlinearity in section 3.1. The numerical experiments conducted in this study include parameter estimation and residual analysis conducted using UCODE_2005 (section 3.2) and DREAM$_{(ZS)}$ simulation based on the Gaussian and formal generalized likelihood functions (section 3.3). The analysis of parametric uncertainty and predictive analysis based on UCODE_2005 and DREAM$_{(ZS)}$ are given in sections 3.4 and 3.5. The reasons that the formal generalized likelihood function outperforms the Gaussian likelihood function are discussed in section 3.6. The parameter sensitivity analysis using UCODE_2005 and DREAM$_{(ZS)}$ is discussed in section 3.7.

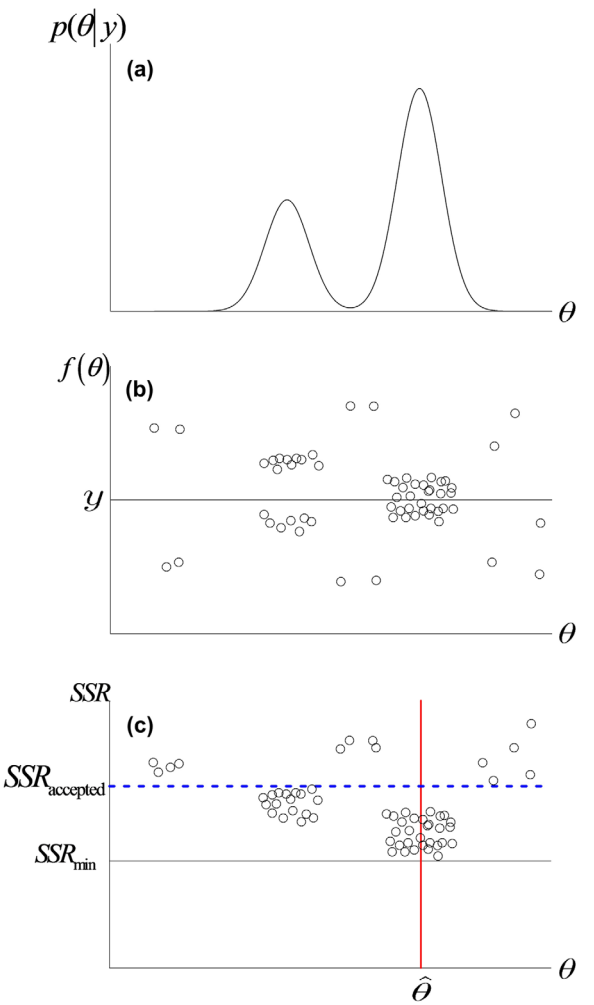

Figure 1. (a) Bimodal distribution of a single parameter θ, (b) model simulations (y) corresponding to parameter samples from the distribution, where y is an observation, and (c) *SSR* (sum of squared residuals) as objective function, where SSR_{min} and $SSR_{accepted}$ are minimum and accepted *SSR*. This figure illustrates that multiple minima of the objective function are caused by multiple modes of the parameter density function.

3.1. Surface Complexation Model, Parametric Uncertainty, and Model Nonlinearity

In order to study uranium transport and test potential applicability of surface complexation modeling, *Kohler et al*. [1996] conducted eight column experiments in a well-characterized U(VI)-quartz-fluoride column system. The breakthrough curves of U(VI) exiting the column over the course of several pore volumes of water showed retardation due to uranium adsorption on the quartz surface. The uranium adsorption was simulated in *Kohler et al*. [1996] using SCMs, in which uranium is absorbed on surface hydroxyl functional groups according to chemical reactions, for example,

$$S_jOH + UO_2^{2+} + H_2O = S_jOUO_2OH + 2H^+, \quad (18)$$

where S_jOH represents a surface hydroxyl functional group. The formation constant, K, of (18) is defined as

$$K_j = \frac{(S_jOUO_2OH)(H^+)^2}{(S_jOH)(UO_2^{2+})}, \quad (19)$$

where the quantities in parenthesis denote the activity of each species (the activity coefficients of the surface species and the activity of water are assumed to be equal to one). A SCM may involve several functional groups with different adsorption affinity measured by K. The number of functional groups that participate in the adsorption reaction and the adsorption affinity of these groups are unknown because of a wide range in bonding environments on the mineral surface. This is a significant contribution to model uncertainty for the SCMs. The distribution in bonding affinity is often approximated by two or three discrete site types [*Davis and Kent*, 1990]. *Kohler et al*. [1996] postulated seven alternative SCMs with different numbers of surface functional groups and different reaction stoichiometry. For each of the seven models, the formation constants of each reaction and the fractions of the reaction surface of each functional group were unknown, and these comprise parametric uncertainty for each model. This study is focused on quantification of parametric uncertainty of one of the seven models. In a future study, the parametric uncertainty quantification will be extended to the other six models, and the model uncertainty will be quantified.

The assessment of parametric uncertainty was conducted for Model C4 selected by *Kohler et al*. [1996] as the best model based on a qualitative analysis and the principle of parsimony. This model has two functional groups, S_1OH and S_2OH, hereinafter referred to as weak site and strong site, respectively. The weak site is associated with one reaction

$$S_1OH + UO_2^{2+} + H_2O = S_1OUO_2OH + 2H^+, \quad (20)$$

and the strong site with two reactions

$$S_2OH + UO_2^{2+} + H_2O = S_2OUO_2OH + 2H^+, \quad (21)$$

$$S_2OH + UO_2^{2+} = S_2OUO_2 + H^+. \quad (22)$$

Model calibration and uncertainty analysis was performed for the base 10 logarithms of the formation constants of the three reactions (denoted as logK1, logK2, and logK3) and the base 10 logarithm of the fraction of the strong site (denoted as logSite). The fraction of the weak site is not considered explicitly, because the summation of the two site fractions is one. The four calibrated parameters played different roles in determining the shape of the simulated breakthrough curves. logK1, the formation constant for the equilibrium reaction associated with the weak site, strongly influences the climbing limb of the breakthrough curve and the center of mass. The decreasing limb of the breakthrough curve is most sensitive to logK2 and logK3, the formation constants of the equilibrium reactions associated with the strong site; surface fraction of the strong site is measured by logSite. The parameters were subject to the physical constraints that logK1 $<$ logK2 and logK3 $>$ logK1 + 3.9. These bounds were used when sampling parameter realizations for MCMC simulations. Nonlinearity of this model was evaluated using the total nonlinearity and intrinsic model nonlinearity measures, which are defined in *Hill and Tiedeman* [2007, pp. 144–145] and calculated by UCODE_2005. According to *Hill and Tiedeman* [2007, pp. 142–145], a model is effectively linear when these measures are less than 0.09, and highly nonlinear when they are greater than 1.0. For this SCM, the total nonlinearity and intrinsic nonlinearity measures were 0.5 and 0.3, indicating that the model is moderately nonlinear.

3.2. Parameter Estimation and Residual Analysis

Following *Kohler et al*. [1996], the regression-based model calibration was conducted by matching simulated uranium concentrations to a total of 120 observations from three experiments denoted as Experiments 1, 2, and 8 which covered a range of *p*H and uranium concentrations. By taking $\mathbf{C}_\varepsilon = \sigma^2\boldsymbol{\omega}^{-1}$ and assuming that the measurement error of each observation is independent, the objective function was

Table 1. Parameter Estimates Obtained Using UCODE_2005, DAKOTA, SCE-UA, and MCMC Simulation (With Gaussian and Formal Generalized Likelihood Functions)[a]

Parameter Name	UCODE Initial Values	UCODE Estimates	LB	UB	DAKOTA	SCE-UA	MCMC (Gaussian)	MCMC (Generalized)
logK1	−5.0	−4.866	−6.5	−3.5	−4.865	−4.864	−4.864	−4.946
logK2	−3.0	−3.128	−5.2	−2.0	−3.121	−3.116	−3.116	−3.122
logK3	1.0	1.220	0.0	2.0	1.225	1.230	1.230	1.621
logSite	−2.0	−1.964	−3.0	−1.0	−1.970	−1.975	−1.975	−2.279

[a]The lower (LB) and upper (UB) bounds of the parameter ranges are also listed.

$$SSWR = \sum_{j=1}^{Ngroup} W_j SSWR_j = \sum_{j=1}^{Ngroup} \left[\frac{N}{N_{obs,j}} \sum_{i=1}^{Nobs,j} \frac{\omega_{ji} \left(Y_{ji} - \hat{Y}_{ji} \right)^2}{\sigma^2} \right], \quad (23)$$

where $Ngroup = 3$ is the number of experiments, $SSWR_j$ is the objective function of each experiment, $W_j = N/N_{obs,j}$ is used to balance different numbers of observations in the individual experiments ($N = 120$ being the total number of observations and $N_{obs,j}$ being the number of observations for the jth experiment, 39, 32, and 49 for Experiments 1, 2, and 8, respectively), ω_{ji} is the inverse of the variance of measurement error for the ith observation of the jth experiment, σ^2 is the error variance that can be estimated via equation (6), and Y_{ji} and $\hat{Y}_{ji}$ are observed and simulated concentration, respectively. The SCM described above is implemented in the reactive transport computer code RATEQ [*Curtis*, 2005] for the forward model simulation. The forward model execution was the same as that of *Kohler et al*. [1996] except that new thermodynamic data of *Guillamont et al*. [2003] were used in this study. A complete simulation of the three experiments took 3–5 min on the supercomputer of the Florida State University.

UCODE_2005 was used for the local optimization to minimize the *SSWR* of equation (23). The initial parameter values are listed in Table 1, and they were close to those obtained by *Kohler et al*. [1996] who used a now outdated set of thermodynamic data. The parameter ranges needed for the local optimization were determined empirically around the initial values based on a preliminary geochemical calculation, which is not included here. The parameters estimates using UCODE_2005 are listed in Table 1, and they are almost identical to those obtained using the global optimization methods of Shuffled complex evolution (SCE-UA) [*Duan et al*., 1992, 1993, 1994] and Multistart Broyden-Fletcher-Goldfarb-Shanno (BFGS) method implemented in DAKOTA [*Adams et al*., 2010] (Table 1). Details of the global optimization are not included here. As a result, the parameter set corresponding to local-UCODE is defined as "global optimization parameter" and used hereinafter in the regression-based uncertainty quantification below to evaluate the linear and nonlinear confidence intervals.

It should be noted that finding the global optimum by UCODE_2005 depends on the initial values (Table 1). Figure 2 plots the $-\log_{10}SSWR$ surface with logK1 and logSite. The $-\log_{10}SSWR$ values are based on 180,000 model runs conducted by varying the three most influential parameters (logK1, logSite, and logK2) identified in the sensitivity analysis discussed below (the fourth parameter, logK3, was fixed at its optimum value). The figure shows a mixture of closely grouped peaks and valleys. For a gradient-based algorithm, if the initial parameter guesses are far from the global optimum, they are likely trapped at local optima, as observed in the results of the multistart BFGS (not shown). The complex response surface also poses challenges to global optimization methods. Without fine tuning of certain parameters of SCE-UA and multistart BFGS, they cannot find the global optima listed in Table 1.

Figure 3 plots the observed and simulated breakthrough curves for the three experiments (Figures 3a1–3a3) and the error-weighted observations and simulations (Figure 3b). While the two figures suggest that the model-fit of UCODE_2005 is acceptable, a Gaussian likelihood function may not accurately characterize the residuals. Figure 3c shows that the residuals are on the order of 10^{-1} and significantly larger than the measurement errors that are on the order of 10^{-3}, indicating that model error dominates over measurement error. The residual variance in general increases with the simulated concentrations, a sign of heteroscedasticity. The residual histogram (Figure 3d) has a sharp peak that cannot be characterized by the assumed Gaussian density function. In addition, the residuals are correlated in time, because the partial autocorrelation function (PACF) at lags 1–2 is beyond the 95% confidence interval, as shown in Figure 3e for Experiment 1 as an example (the temporal correlation was observed for all the three experiments). The

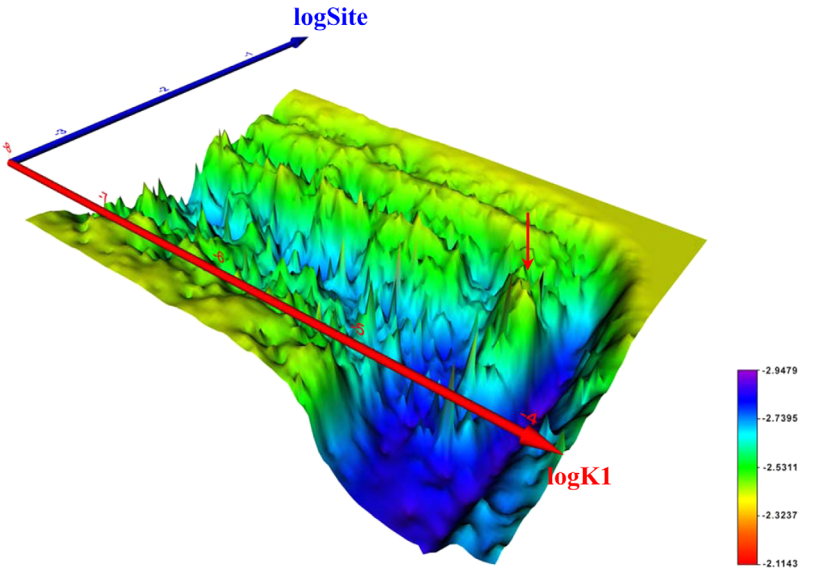

Figure 2. Two-dimensional surface of $-\log_{10}$SSWR when logK1, logSite, and logK2 vary simultaneously (the least important parameter, logK3, is fixed at its global optimum). The red arrow points to the global optima of logK1, logSite, and logK2.

formal generalized likelihood function is sufficient designed to characterize these three characteristics of heteroscedasticity, peaked density function, and temporal correlation.

3.3. MCMC Simulation Using Gaussian and Formal Generalized Likelihood Functions

The MCMC simulations for the Gaussian and formal generalized likelihood functions were conducted by running three chains in parallel. Uniform distributions were used as the prior parameter distributions, and the lower and upper bounds of the four model parameters are listed in Table 1. Convergence of the MCMC simulation was monitored using the scale reduction factor ($\hat{R}$) defined by *Gelman et al*. [2004, pp. 331–333], with the rule of thumb that convergence is attained after $\hat{R}$ is less than 1.2. Figure 4 shows that the MCMC simulation using the Gaussian likelihood reached convergence after 3000 MCMC runs. This is confirmed by the well mixing of the three chains plotted in Figure 5. A total of 15,000 model executions were conducted, and the first 5000 samples were discarded as the burn-in period. The remaining samples were used for the uncertainty quantification discussed below.

For the MCMC simulation using the formal generalized likelihood function, it is more difficult to reach convergence. In this simulation, in addition to the four parameters of the SCM, the parameter set includes the two variance parameters (σ_0 and σ_1 in equation (16)), two shape parameter (ξ and θ in equation (17)), and two autocorrelation coefficients (ϕ_1 and ϕ_2 in equation (15)). The order of autoregressive model, AR(p), was determined using the method described in *Lu et al*. [2013]. A total of 75,000 MCMC runs were conducted. Figure 6 shows that all the $\hat{R}$ values became smaller than 1.2 after 43,000 MCMC samples, but they remained systematically lower only after 60,000 samples. Therefore, the first 60,000 samples were discarded, and the remaining 15,000 samples were used for the uncertainty quantification discussed below. The larger burn-in period may be attributed to the augmented parameter set, and more investigation is warranted in a future study. The evolution of the chains in Figure 7 shows the difference of burning the first 43,000 and 60,000 samples. When the burn-in period is 43,000 samples, logK3 and logSite can take values of 0.8 and -1.4, respectively, and these values are modes on parameter histograms plotted using the remaining 32,000 samples. However, if the burn-in period is 60,000, these values were not observed in the last 15,000 samples. The uncertainty quantification discussed below was based on the more reasonable burn-in period of 60,000 samples. Residual analysis of the DREAM$_{(ZS)}$ realizations is given in section 3.6.

3.4. Comparison of Parametric Uncertainty

The parameter distributions obtained from the regression and Bayesian methods are compared for the four model parameters (logK1, logK2, logK3, and logSite). Figure 8 plots the Gaussian density function obtained using the UCODE_2005 results, i.e., the means and variances of the density functions are respectively the

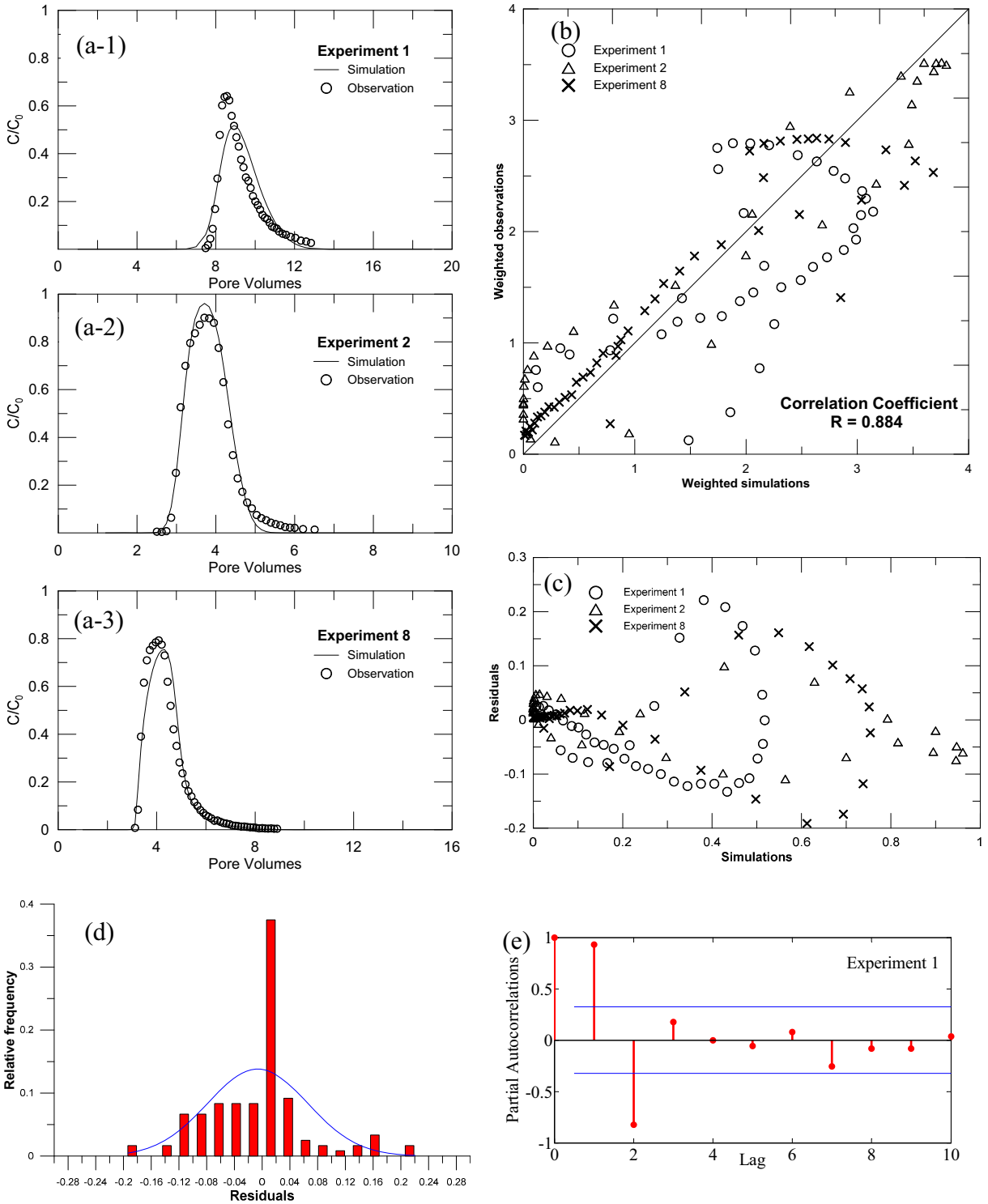

Figure 3. Results of model calibration using UCODE_2005: (a1)–(a3) comparison of observed and simulated breakthrough curves for experiments 1, 2, and 8, (b) graph of weighted simulations and weighted observations, (c) graph of residuals and simulations, (d) actual histogram of residuals and fit to a Gaussian density function, and (e) partial autocorrelation coefficient of the residuals of Experiment 1 with 95% significance levels (blue lines).

parameter estimates and estimation variances evaluated using the regression theories. Figure 8 also plots the scaled relative frequency (i.e., relative frequency divided by bin width) histograms of the MCMC samples obtained using the two likelihood functions. The histograms of the MCMC samples using the Gaussian likelihood function also have the Gaussian shape. The modes of the histograms are almost identical to the

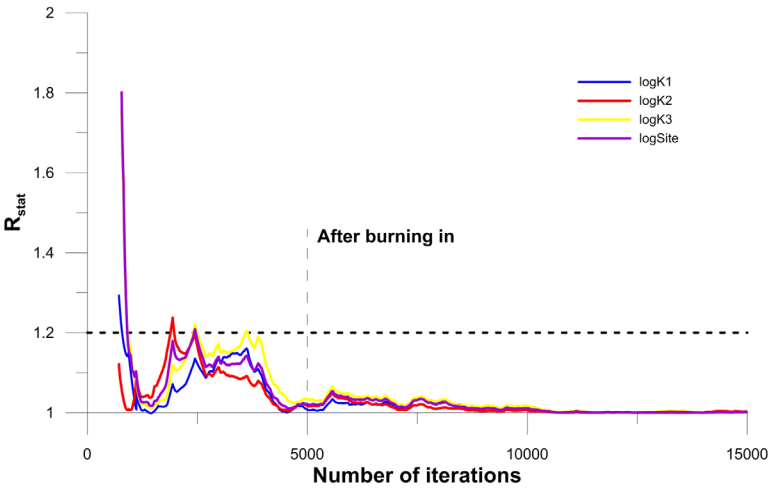

Figure 4. Convergence behaviors of the Gelman-Rubin Statistics for DREAMZS realizations using Gaussian likelihood function.

means of the Gaussian density functions, i.e., the UCODE_2005 parameter estimates (Table 1). The standard deviations of the MCMC samples are larger than those of the UCODE_2005 results; this explains why the histograms are wider than the Gaussian density plots. The similarity indicates that if Gaussian likelihood functions are used, the posterior parameter distributions of Bayesian methods are similar to the Gaussian parameter distributions of the regression methods. While this was mathematically proved for linear models

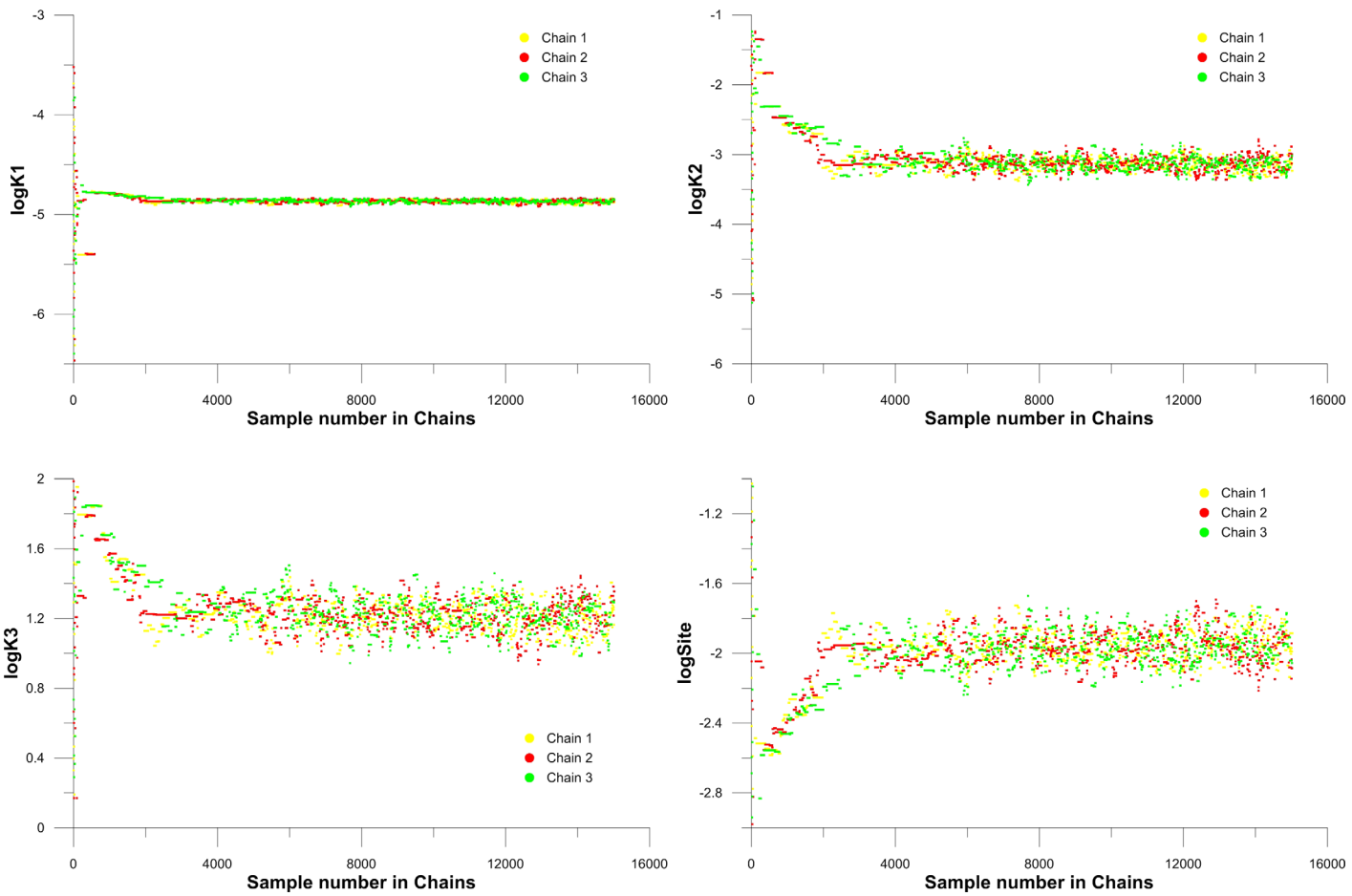

Figure 5. Evolution of three MCMC chains for the four parameters when using Gaussian likelihood function.

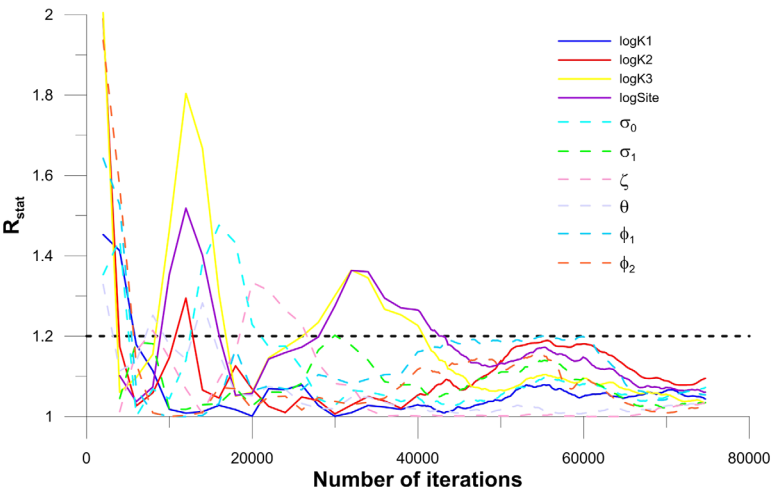

Figure 6. Convergence behaviors of the Gelman-Rubin Statistics for DREAMzs realizations using formal generalized likelihood function.

(equation (13) adapted from *Box and Tiao* [1992]), it is interesting to observe that the same is true for the moderately nonlinear SCM in this study. However, it should be noted that since the residuals are non-Gaussian (as shown in Figure 3 and discussed in section 3.3), the Gaussian likelihood is invalid. Therefore, the resulting parameter distributions are inappropriate for quantifying predictive uncertainty, as shown in section 3.6 below.

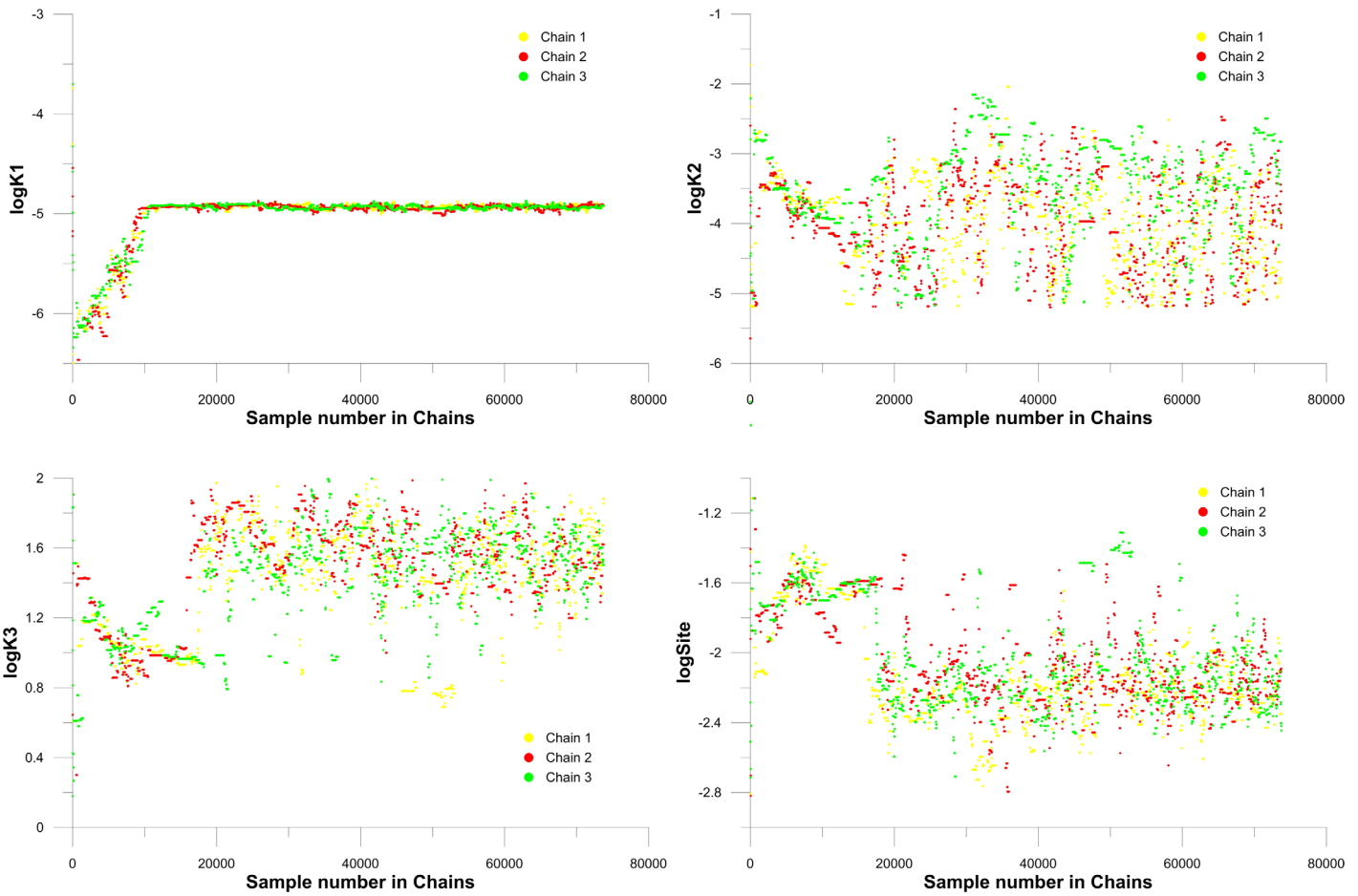

Figure 7. Evolution of three MCMC chains for the four parameters when using formal generalized likelihood function.

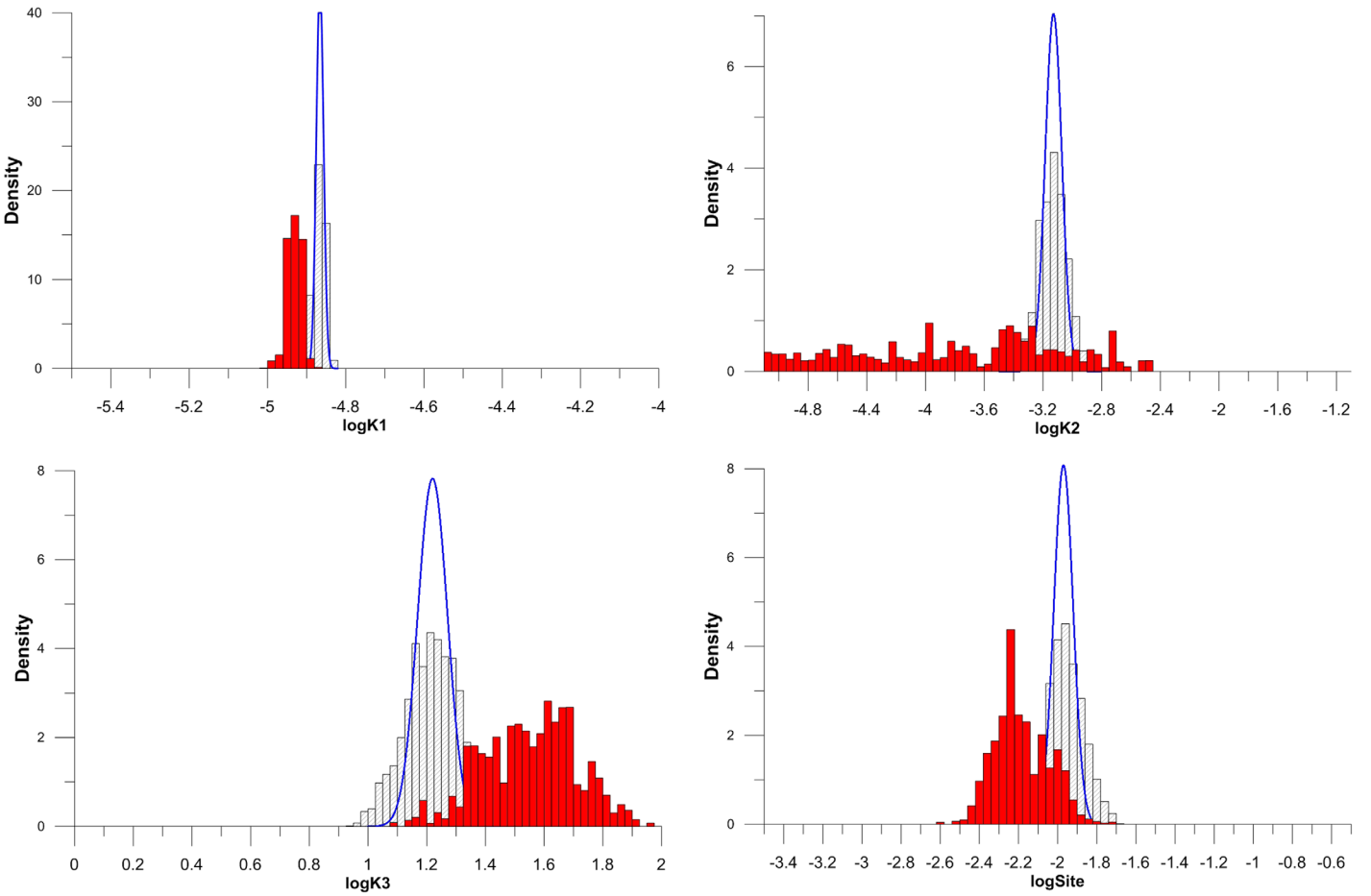

Figure 8. Scaled relative frequency (i.e., relative frequency divided by bin width) histograms of four model parameters from DREAM(zs) samples based on Gaussian (black) and formal generalized (red) likelihood functions. The blue lines plot the normal distributions based on UCODE_2005 results, i.e., using the parameter estimates as the means and the estimation variances as the variances.

Using the formal generalized likelihood function substantially changes the global parameter optima and posterior parameter distributions. The global optima corresponding to the formal generalized likelihood function are listed in Table 1, and they are substantially different from those obtained using UCODE_2005 and DREAM$_{(ZS)}$ with the Gaussian likelihood function. Figure 8 shows that the histograms of the MCMC samples based on the Gaussian likelihood are dramatically different from those based on the formal generalized likelihood. The histogram of logK1 is unimodal, but the mode is different from that based on the Gaussian likelihood function. The histogram of logK2 based on the formal generalized likelihood appears to be a uniform distribution, spreading significantly wider than that based on the Gaussian likelihood. The histogram of logK3 based on the formal generalized likelihood is somewhat bimodal; the two modes are close but substantially different from that of the histogram based on the Gaussian likelihood. The same is true for the histograms of logSite. These results indicate that parametric uncertainty is larger when the formal generalized likelihood function is used. The same was reported in *Schoups and Vrugt* [2010], and they attributed this to the following three reasons: (1) the heteroscedastic error model results in larger uncertainty; (2) considering residual correlation reduces information content in the data and leads to larger uncertainty; and (3) the SEP distribution has heavier tails than the Gaussian distribution. There may be other reasons specific for the SCM, investigation of which is warranted in a future study.

Using the formal generalized likelihood function also substantially changes the parameter correlation. Comparing the two-parameter scatter plots in Figure 9 based on the Gaussian likelihood function with those in Figure 10 based on the formal generalized likelihood function shows that, except the correlation between logK3 and logSite, the correlation between any two model parameters decreases when the formal

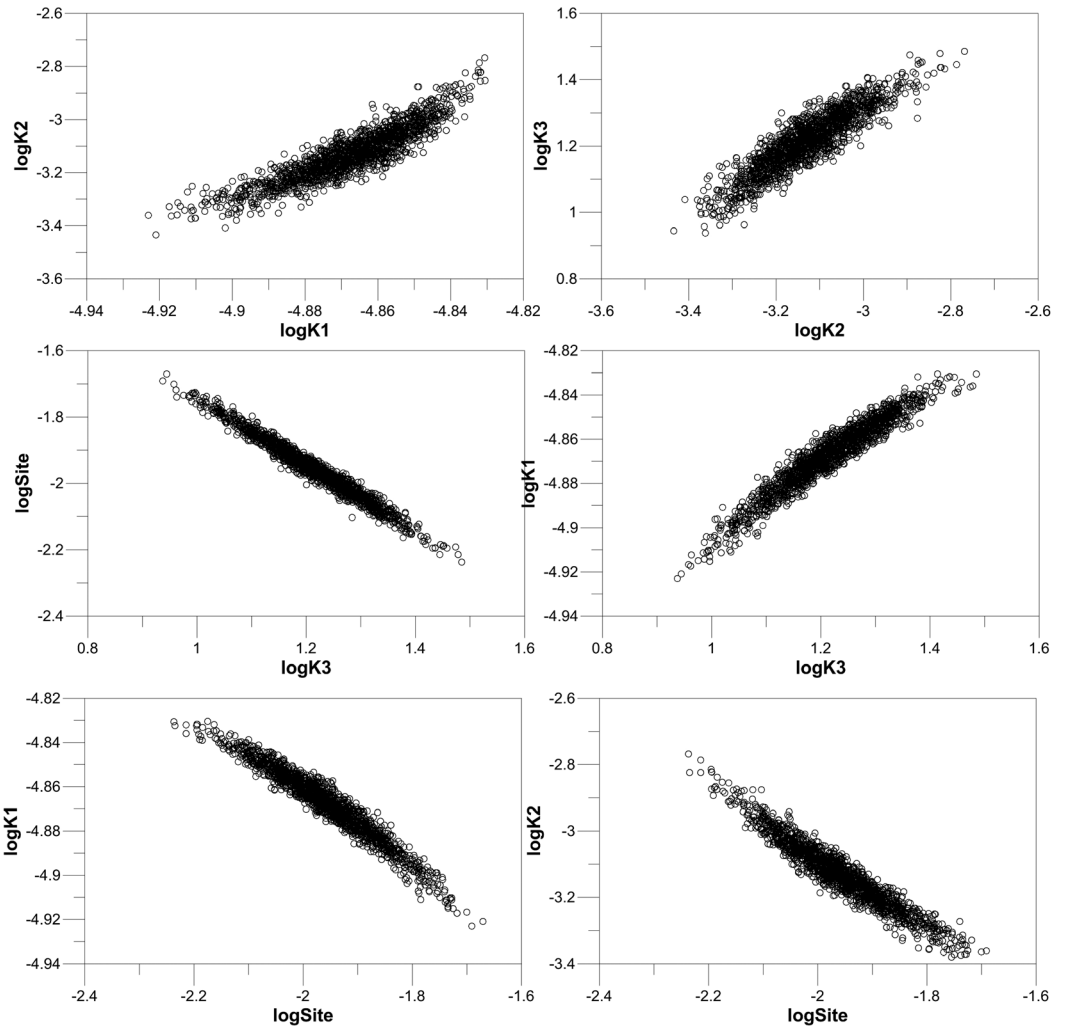

Figure 9. Scatter plots of parameters using the MCMC samples based on the Gaussian likelihood.

generalized likelihood function is used. The weaker correlation may be attributed to the interaction between the SCM parameters and the parameters of the generalized likelihood function.

3.5. Comparison of Predictive Uncertainty

In a cross-validation manner, the SCM was used to predict Experiment 5 with 41 observations and Experiment 7 with 170 observations. The two experiments were not used for the model calibration; their chemical conditions are different from those of Experiments 1, 2, and 8 used for model calibration, and so are the breakthrough curves. Predictive uncertainty was quantified using the 95% confidence intervals (linear and nonlinear) estimated using UCODE_2005 and the 95% credible intervals estimated using the DREAM$_{(ZS)}$ samples based on the Gaussian and formal generalized likelihood functions.

The MCMC results using the formal generalized likelihood function are the best for predicting the data of Experiments 5 and 7. Figures 11a and 12a show that the 95% linear and nonlinear confidence intervals are too narrow to include the observed breakthrough data. In addition, the mean predictions (the center of the intervals, not shown in the figures) are biased toward delayed breakthrough. Figures 11b and 12b show that although the 95% credible intervals based on the Gaussian likelihood function are wider than the linear and nonlinear confidence intervals, the mean predictions are still biased. The mean predictions of the MCMC simulation based on the formal generalized likelihood function are close to the observations, and the corresponding 95% credible intervals cover significantly more observations than the other three sets of

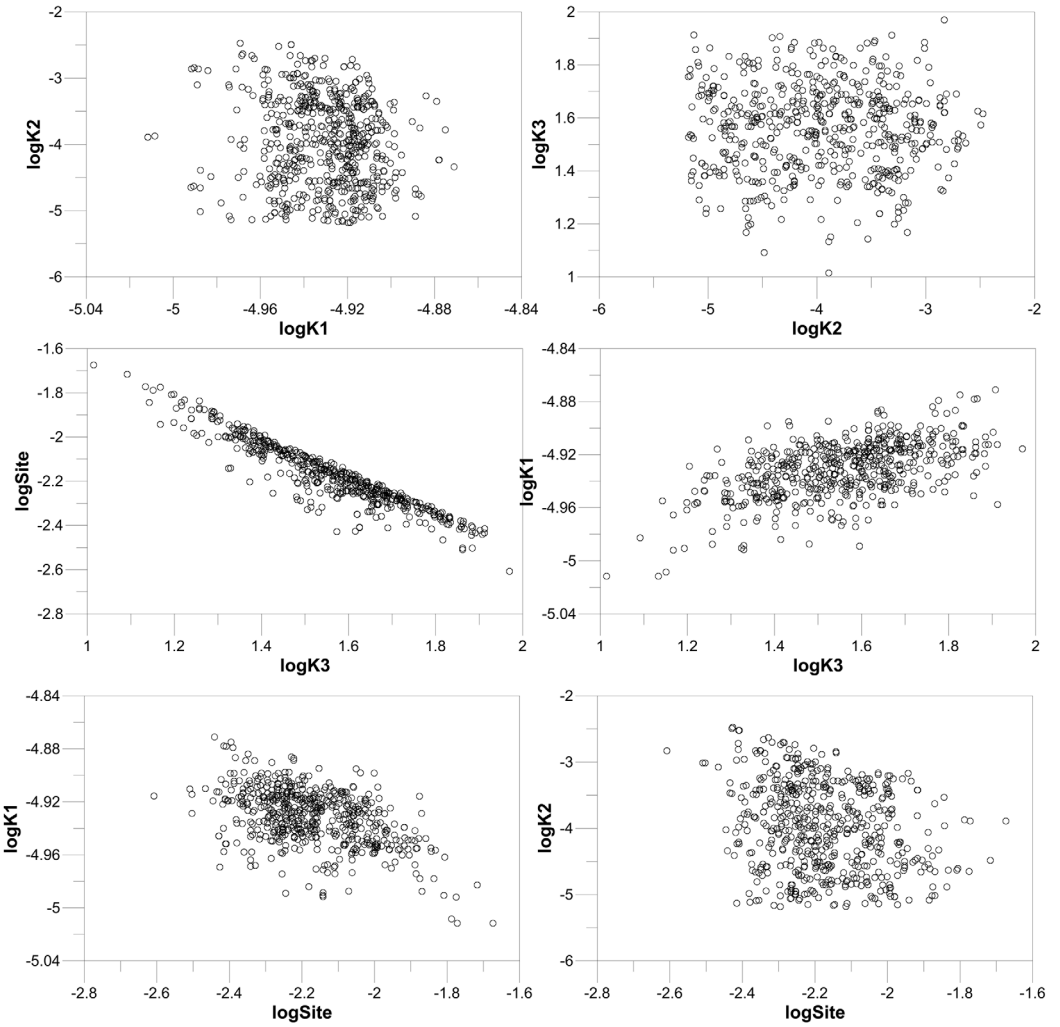

Figure 10. Scatter plots of parameter correlation using the MCMC samples based on the formal generalized likelihood function.

intervals (Figures 11b and 12b). These results indicate that using the formal generalized likelihood function greatly improves model prediction and uncertainty quantification. Therefore, the likelihood function should be used for quantifying predictive uncertainty of groundwater reactive transport modeling, even though its corresponding MCMC simulation is most computationally expensive.

3.6. Uncertainty of Formal Generalized Likelihood Parameters

To understand the reasons that the formal generalized likelihood function outperforms the Gaussian likelihood function, Figure 13 plots the posterior histograms of the six error model parameters: the two variance parameters (σ_0 and σ_1 in equation (16)), the two shape parameter (ξ and θ in equation (17)), and the two autocorrelation coefficients (ϕ_1 and ϕ_2 in equation (15)). The figure shows that while σ_0 is close to zero, σ_1 ranges between 0.1 and 0.4 with the peak around 0.2. According to equation (16), these values implying that the error standard deviation is heteroscedasticity and can be as large as 40% of the mean simulation. The histogram of ξ suggests that the values of ξ are larger than 0.5 and with the majority of the data between 0.7 and 0.9. The histogram of θ indicates that the θ values are larger than 0.65 with the peak around 0.85. Considering these values and *Schoups and Vrugt* [2010, Figure 1], the SEP density functions of the MCMC realizations are non-Gaussian, with sharp peak and right skewness. Since the ϕ_2 values are close to zero, the AR(1) model should be sufficient to characterize the temporal correlation. To better understand how SEP characterizes residuals, the residuals a_t (not e_t) of equation (15) corresponding to the MCMC realization with the largest likelihood function value are analyzed; the six error model parameter values of this realization are $\sigma_0 = 0.0026$, $\sigma_1 = 0.1995$, $\xi = 0.9873$,

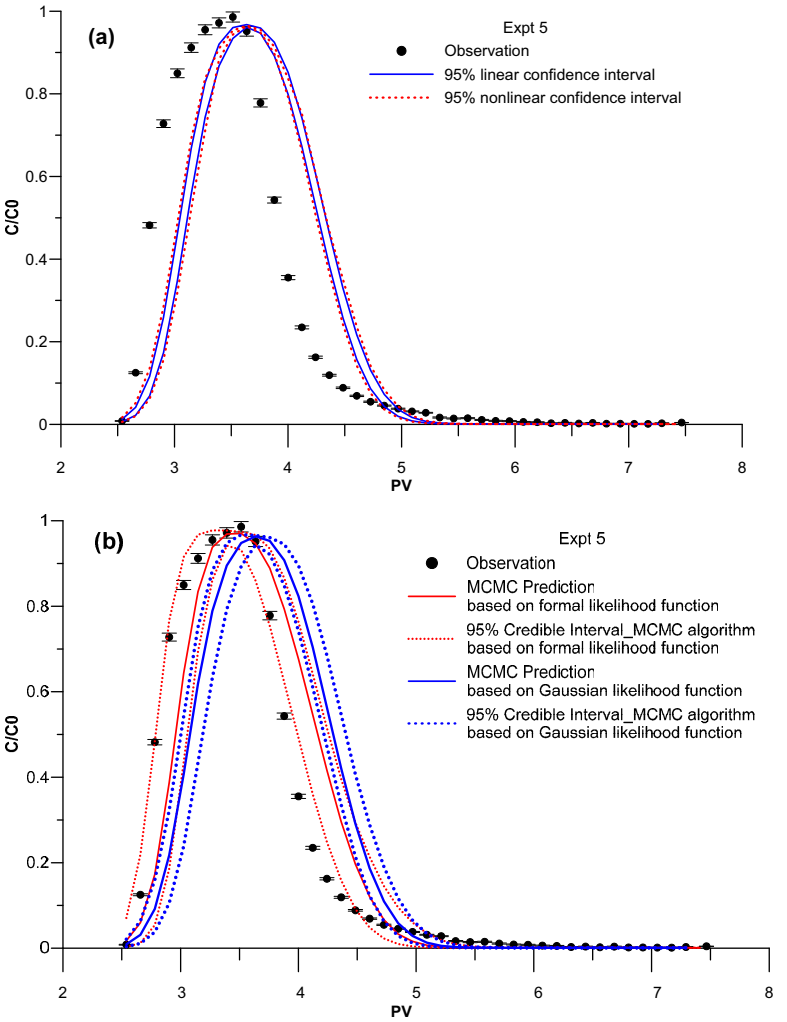

Figure 11. (a) 95% linear and nonlinear confidence intervals evaluated using UCODE_ 2005 and (b) 95% nonlinear credible intervals evaluated using the Gaussian and formal generalized likelihood functions for Experiment 5.

θ=0.8354, ϕ_1=0.8965, and ϕ_2=0.0034. Figure 14a qualitatively shows that the error model can remove residual heteroscedasticity in that the residual variance appears to be constant for simulated concentrations. Figure 14b shows that the SEP can well characterize the residual distribution, which cannot be achieved using the Gaussian likelihood function. Figure 14c manifests that AR(1) can remove the first-order correlation, with some minor correlation remained in a_t. Therefore, it is concluded that the formal generalized likelihood function is suitable to characterize the residuals of the SCM model of uranium reactive transport.

3.7. Parameter Sensitivity Analysis

The MCMC results were used to investigate parameter sensitivity using the reduction of the Shannon entropy of the prior to the posterior distributions. The Shannon entropy is defined as $-\sum_{i=1}^{N} p_i \log p_i$ [*Papoulis*, 1991, pp. 533–537], where p is parameter probability mass function and N is number of bins on a histogram. Large entropy corresponds to large uncertainty. Table 2 lists reduction of the entropy from the prior to the posterior parameter distributions. Considering that uncertainty reduction is the largest for the most sensitive parameters, the parameter sensitivity is in the order of logK1 > logSite > logK2 > logK3 for the Gaussian likelihood function but logK1 > logSite > logK3 > logK2 for the formal generalized likelihood function. In other words, logK1 (formation constant of the reaction in equation (20) associated with the weak site) and logSite (surface fraction of the strong site) are the most and second most important parameters. The two least important parameters are logK2 and logK3 of the reactions in equations (21) and (22) associated with the strong site. This order of importance is physically reasonable because the multisite

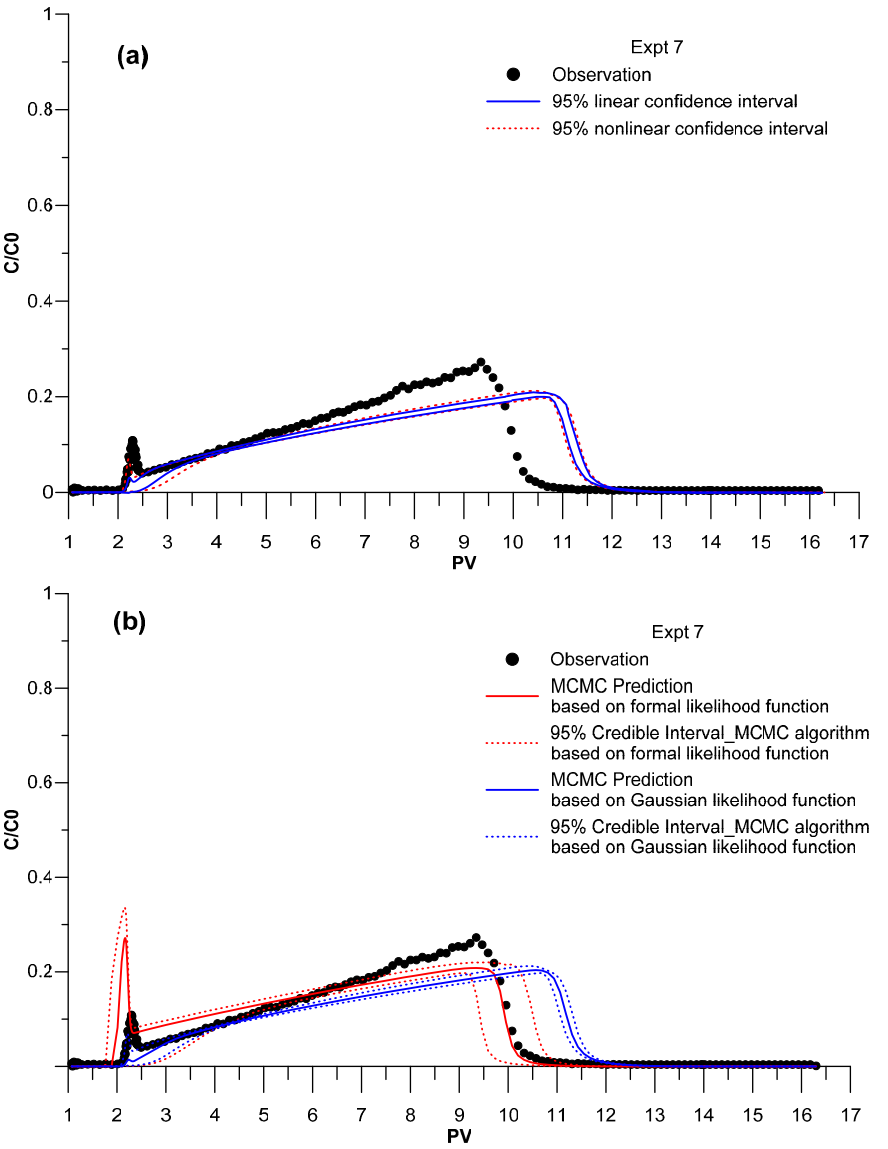

Figure 12. (a) 95% linear and nonlinear confidence intervals evaluated using UCODE_ 2005, (b) 95% nonlinear credible intervals evaluated using the Gaussian and formal generalized likelihood functions for Experiment 7.

adsorption model causes a self-sharpening front. For the SCM, logK1 strongly influences the initial and mean arrival of uranium; in contrast, logK2 and logK3 determine adsorption to the strong site which controls the extent of tailing shown by low concentrations on the breakthrough curve. These low concentrations and the competitive adsorption onto a single site lead to a smaller sensitivity to the two strong site parameters. It is reasonable that when the formal generalized likelihood function was used, logK2 became the least influential parameter, because its histogram is uniform for the formal generalized likelihood function but Gaussian-like for the Gaussian likelihood function.

The sensitivity ranking is similar to that based on the local and hybrid of local and global techniques. The local sensitivity was conducted using UCODE_2005 to calculate, for each of the four model parameters, the composite scaled sensitivity (CSS) defined in *Hill and Tiedeman* [2007]. Smaller CSS values indicate smaller parameter sensitivity. The hybrid of local and global sensitivity was conducted using the Morris One-At-a-Time (MOAT) method implemented in DAKOTA software [*Adams et al.*, 2010]. The method evaluates elementary effect (a local sensitivity) at a number of discrete points in parameter space. The mean elementary effect, μ, measures importance of a parameter on the model output; a high mean value indicates large overall importance. A high standard deviation, σ, of the elementary effect suggests that the parameter is either

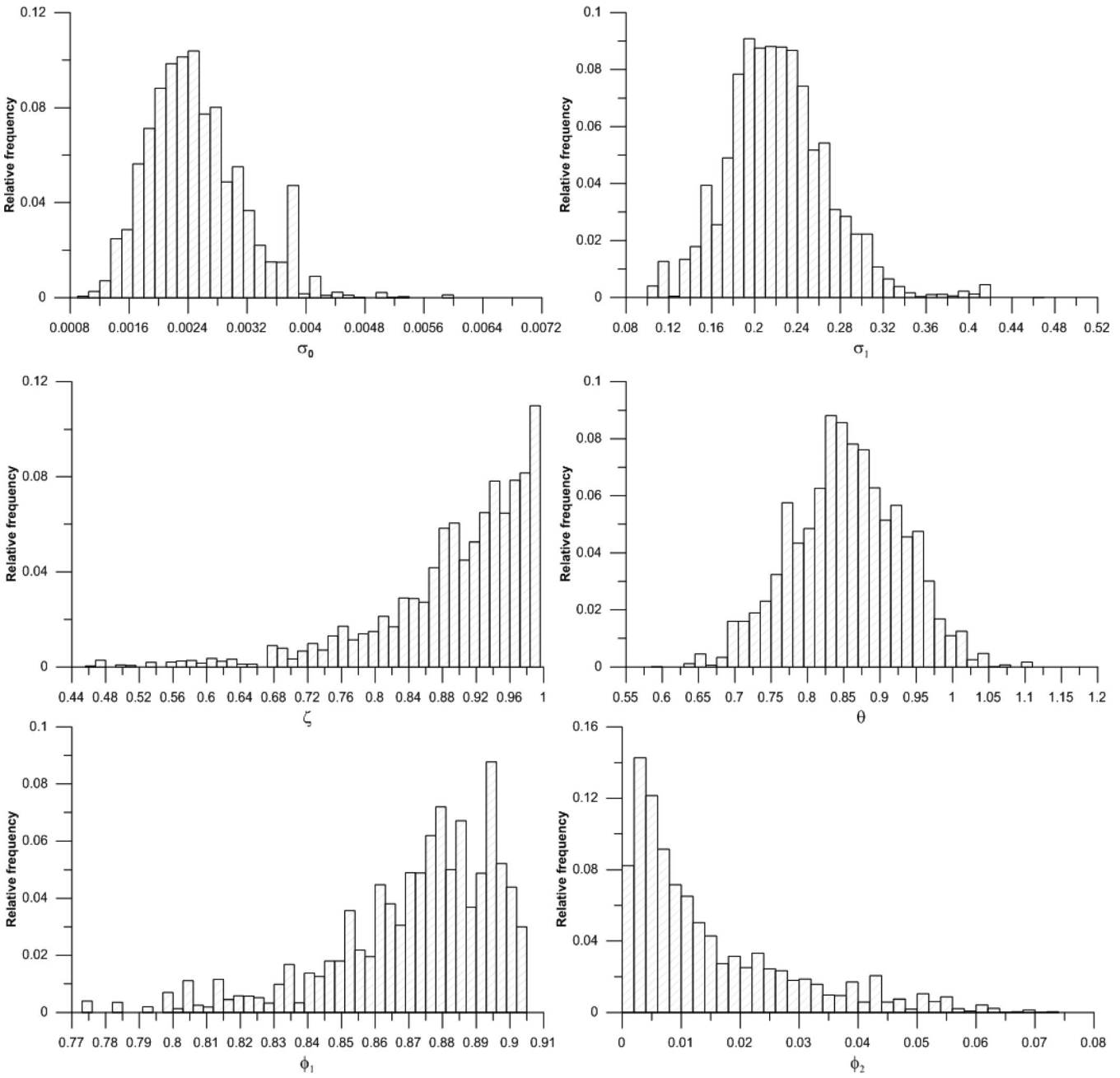

Figure 13. Posterior histograms of six error model parameters of the generalized formal likelihood function.

interacting with other parameters or has a high nonlinear effect on the output. Figure 15 shows that the order of sensitivity based on CSS and MOAT is the same as that based on the Gaussian likelihood function (Table 2). If this sensitivity ordering is generally consistent for the different methods, it is computationally efficient to conduct the local and MOAT sensitivity analysis first to select the most influential parameters for computationally expensive MCMC simulations.

4. Conclusions

The following is a summary of the key findings of the uncertainty quantification based on the regression and local optimization methods implemented in UCODE_2005 and Bayesian methods using MCMC simulation implemented in DREAM$_{(ZS)}$ with the Gaussian and formal generalized likelihood functions:

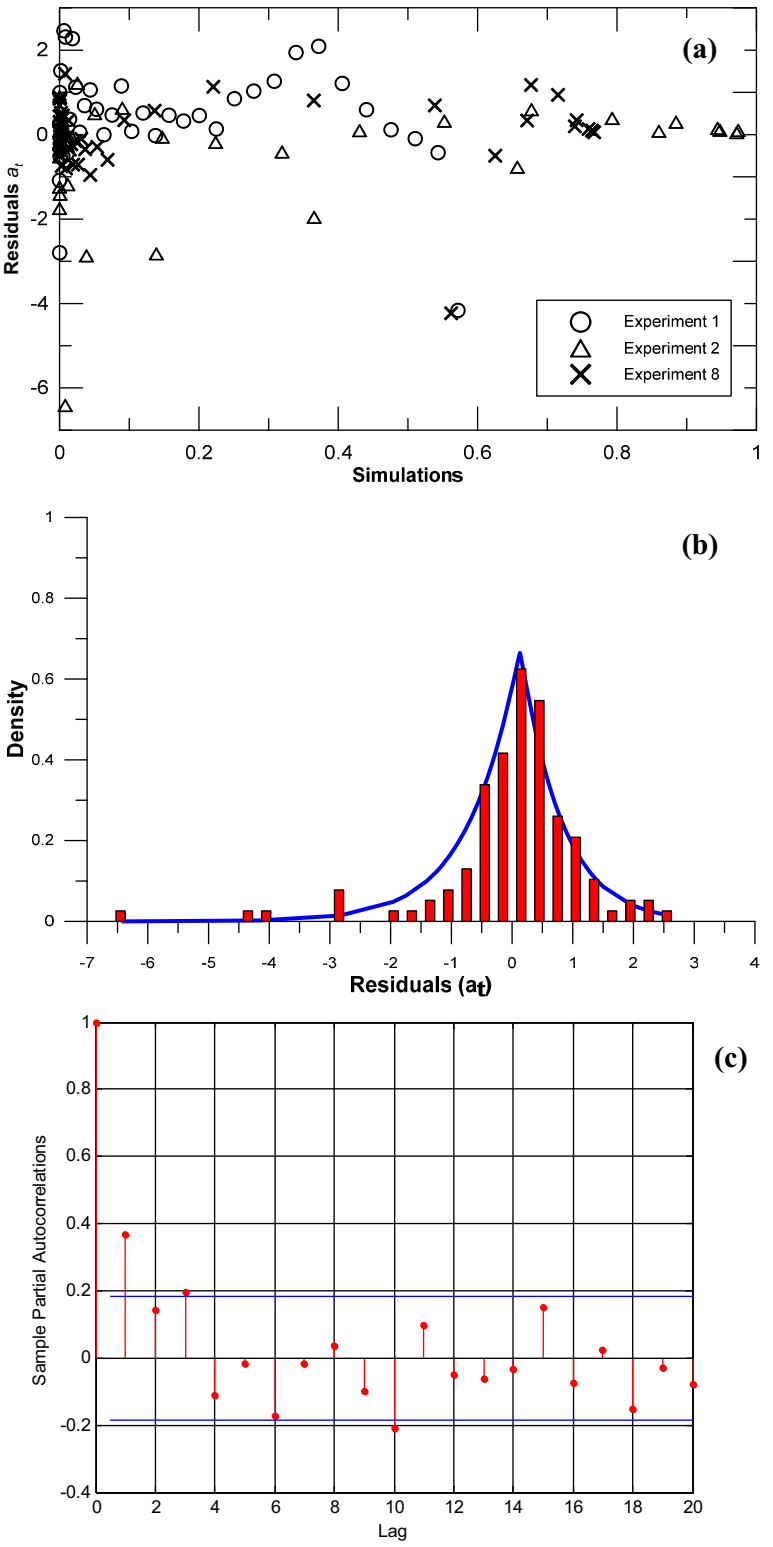

Figure 14. (a) Residuals at as a function of simulated concentration, (b) estimated (blue solid line) skewed exponential power distribution and actual histogram (red bar) of residuals at, and (c) partial autocorrelation coefficient of the residuals at with 95% significance levels (blue lines).

Table 2. Entropy Values of Prior and Posterior Parameter Distributions Based on the Gaussian and Formal Generalized Likelihood Function

		Gaussian Likelihood			Formal Generalized Likelihood		
Parameter Name	Prior Entropy	Posterior Entropy	Entropy Reduction	Rank of Reduction	Posterior Entropy	Entropy Reduction	Rank of Reduction
logK1	1.9031	0.5238	1.3793	1	0.5940	1.3091	1
logK2	1.9031	1.0008	0.9023	3	1.6552	0.2479	4
logK3	1.9031	1.1675	0.7356	4	1.3936	0.5095	3
logSite	1.9031	0.9703	0.9328	2	1.1312	0.7719	2

1. The response surface of least squares objective function is complex, with mixture of closely grouped peaks and valleys, due to nonlinearity of groundwater reactive transport models. This poses great challenges to both local and global optimization methods. DREAM$_{(ZS)}$ is a robust method to find the global parameter optima.

2. In the residuals of UCODE_2005 model calibration, model errors are significantly larger than measurement errors, indicating that the Gaussian likelihood function explicitly used in UCODE_2005 may not adequately

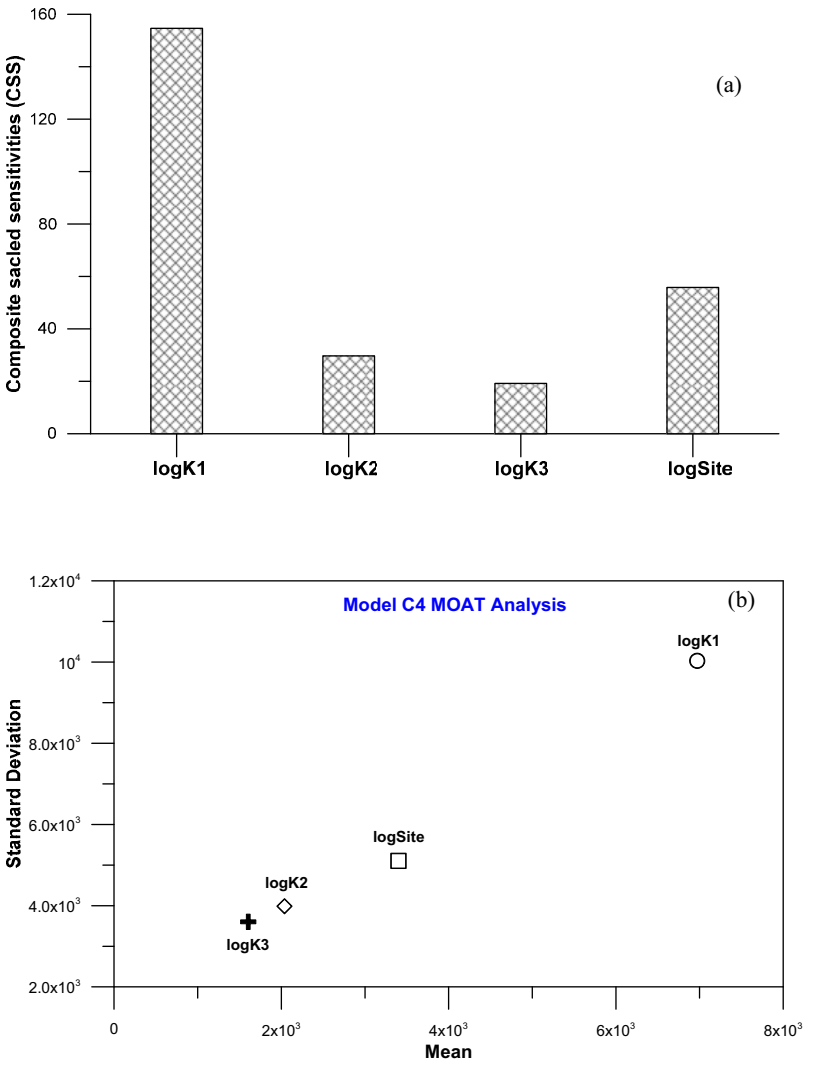

Figure 15. (a) Local sensitivity estimated using UCODE_2005 for the global parameter optima and (b) global sensitivity estimated using the MOAT method. The order of parameter is consistent that the most and second most important parameters are logK1 and logSite, and the two least important parameters are logK2 and logK3.

characterize the residuals. The formal generalized likelihood function of *Schoups and Vrugt* [2010] is needed to characterize heteroscedasticity, non-Gaussian density function, and temporal correlation of the residuals.

3. When the Gaussian likelihood function is used in the DREAM$_{(ZS)}$ simulation, the posterior parameter distributions and credible intervals are similar to those of UCODE_2005 based on the Gaussian assumption of likelihood function, model parameters, and model predictions.

4. When the formal generalized likelihood function is used in the DREAM$_{(ZS)}$ simulation, the posterior parameter distributions are non-Gaussian with multiple modes. The parametric uncertainty (measured by the Shannon entropy) is larger than that of UCODE_2005 and DREAM$_{(ZS)}$ with the Gaussian likelihood function.

5. The credible intervals based on the formal generalized likelihood function are less biased and have larger predictive coverage than the credible intervals based on Gaussian likelihood function and the linear and nonlinear confidence intervals of regression methods. However, the credible interval based on the generalized likelihood function cannot cover all the cross-validation data, suggesting that, in addition to considering parametric uncertainty, model uncertainty should be considered to further improve model predictions.

6. DREAM$_{(ZS)}$, local, and global sensitivity methods gave the similar order of parameter importance (from most to least importance): logK1 > logSite > logK3 > logK2 or logK1 > logSite > logK2 > logK3. When DREAM$_{(ZS)}$ is not computationally feasible, sensitivity analysis may be conducted using computationally frugal methods such as MOAT.

The computational cost of DREAM$_{(ZS)}$ may be reduced by using computationally efficient surrogate models [*Razavi et al*., 2012], which have be implemented using polynomial chaos expansion [*Laloy et al*., 2013] and sparse grid collocation [*Zhang et al*., 2013] within the DREAM framework. An alternative is not to use DREAM$_{(ZS)}$ but to use an appropriate transformation (e.g., the Box-Cox transformation) to yield Gaussian residuals and then to use generalized least squares methods for parameter estimation and uncertainty quantification [*Carroll and Ruppert*, 1988]. This is warranted in a future study. Although this study was conducted as a numerical experiment for the surface complexation model of uranium reactive transport, it is expected that many conclusions of this study are applicable to other groundwater reactive transport models and environmental models.

Acknowledgments
This work was supported in part by National Natural Science Foundation of China grants 51328902, 41172206, and 41172207, NSF-EAR grant 0911074, DOE-SBR grants DE-SC0002687 and DE-SC0000801, DOE Early Career Award DE-SC0008272, and ORAU/ORNL High Performance Computing Grant. The first author was employed by the Florida State University when conducting part of this research. The authors thank Jasper Vrugt for providing the DREAM$_{(ZS)}$ code, and Bruno Mendes, and David Draper for discussion on comparison of confidence and credible intervals. We also thank G.-H. Crystal Ng and the anonymous reviewers for improving the manuscript.

References

Adams, B. M., K. R. Dalbey, M. S. Eldred, D. M. Gay, L. P. Swiler, W. J. Bohnhoff, J. P. Eddy, K. Haskell, and P. D. Hough (2010), DAKOTA, A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis, Sandia Natl. Lab., Albuquerque, N. M.

Beck, B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*(8), 1393–1442.

Beven, K., P. J. Smith, and J. E. Freer (2008), So just why would a modeler choose to be incoherent, *J. Hydrol.*, *354*, 15–32.

Box, E. P., and G. C. Tiao (1992), *Bayesian Inference in Statistical Analysis*, 558 pp., John Wiley, N. Y.

Cabaniss, S. E. (1997), Propagation of uncertainty in aqueous equilibrium calculations: Non-Gaussian output distributions, *Anal Chem*, *69*(18),3658-3664.

Cabaniss, S. E. (1999), Uncertainty propagation in geochemical calculation: Nonlinearity in solubility equilibria, *Appl. Geochem.*, *14*, 255–262.

Carrera, J. (1993), An overview of uncertainties in modelling groundwater solute transport, *J. Contam. Hydrol.*, *13*, 23–48.

Carroll, R. J., and D. Ruppert (1988), *Transformation and Weighting in Regression*, Chapman and Hall, N. Y.

Casella, G., and R. L. Berger (2002), *Statistical Inference*, Duxbury, Pacific Grove, Calif.

Chen, J., A. Kemna, and S. S. Hubbard (2008), A comparison between Gauss-Newton and Markov chain Monte Carlo based methods for inverting spectral induced polarization data for Cole-Cole parameters, *Geophysics*, *73*(6), F247–F259.

Chen, J., S. S. Hubbard, D. Gaines, V. Korneev, G. Baker, and D. Watson (2010), Stochastic estimation of aquifer geometry using seismic refraction data with borehole depth constraints, *Water Resour. Res.*, *46*, W11539, doi:10.1029/2009WR008715.

Christensen, S., and R. L. Cooley (1999), Evaluation of confidence intervals for a steady-state leaky aquifer model, *Adv. Water Resour.*, *22*(8), 807–817, doi: 10.1016/S0309-1708(98)00055-4.

Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, *47*, W09301, doi:10.1029/2010WR009827.

Cooley, R. L., and R. L. Naff (1990), Regression modeling of ground-water flow, in U.S. Geological Survey Techniques of Water-Resources Investigations, Book 3, Chap. B4, 232 pp., U.S. Geol. Surv., Washington, D. C.

Cooley, R. L. (2004), A theory for modeling groundwater flow in heterogeneous media, *U.S. Geol. Surv. Prof. Pap.*, *1679*.

Curtis, G. P. (2005), Documentation and applications of the reactive geochemical transport model RATEQ, *Rep. NUREG CR-6871*, 97 p., U.S. Nucl. Regul. Comm., Rockville, Md.

Curtis, G. P., and J. A. Davis (2006), Tests of uranium(VI) adsorption models in a field setting, *Rep. NUREG CR-6911*, 99 p., U.S. Nucl. Regul. Comm., Rockville, Md.

Dai, Z., and J. Samper (2004), Inverse problem of multicomponent reactive chemical transport in porous media: Formulation and applications, *Water Resour. Res.*, *40*, W07407, doi:10.1029/2004WR003248.

Davis, J. A., and D. B. Kent (1990), Surface complexation modeling in aqueous geochemistry: in Mineral-Water Interface Geochemistry, edited by M.F. Hochella and A.F. White, Reviews in Mineralogy, v. 23, p. 177–260.
Davis, J. A., S. B. Yabusaki, C. I. Steefel, J. M. Zachara, G. P. Curtis, G. D. Redden, L. J. Criscenti, and B. D. Honeyman (2004a), Assessing conceptual models for subsurface reactive transport of inorganic contaminants, *Eos Trans. AGU*, *85*(4), 449–455.
Davis, J. A., D. E. Meece, M. Kohler, and G. P. Curtis (2004b), Approaches to surface complexation modeling of uranium(VI) adsorption on aquifer sediments, *Geochim. Cosmochim. Acta*, *68*, 3621–3641.
Denison, F. H., and J. Garnier-Laplace (2005), The effects of database parameter uncertainty in uranium (VI) equilibrium calculations, *Geochim. Cosmochim. Acta*, *69*(9), 2183–2191.
Doherty, J. (2010), *PEST: Model-Independent Parameter Estimation, User Manual*, 5th ed., Watermark Numer. Comput., Brisbane, Australia.
Draper, D. (2007), Bayeisian multilevel analysis and MCMC, in *Handbook of Multilevel Analysis*, edited by J. de Leeuw and E. Meijer, pp. 77–139, Springer, N. Y.
Draper, N. R., and H. Smith (1981), *Applied Regression Analysis*, 2nd ed., John Wiley, N. Y.
Duan, Q., V. K. Gupta, and S. Sorooshian (1993), Shuffled complex evolution approach for effective and efficient global minimization, *J. Opt. Theory Appl.*, *76*, 501–521.
Duan, Q. Y., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.
Duan, Q. Y., S. Sorooshian, and V. K. Gupta (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*(3–4), 265–284.
Finsterle, S. (2007), iTOUGH2 command reference, *Rep. LBNL-40041*, Lawrence Berkeley Natl. Lab., Berkeley, Calif.
Finsterle, S., and Y. Zhang (2011), Solving iTOUGH2 simulation and optimization on problems using the PEST protocol, *Environ. Modell. Software*, *26*, 959–968, doi:10.1016/j.envsoft.2011.02.008.
Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty, analysis for a watershed model, *Environ. Model. Softw.*, *22*, 1000–1020, doi: 10.1016/j.envsoft.2006.06.007.
Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, *7*, 457–511.
Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, Chapman and Hall, Boca Raton, Fla.
Guillamont, R., T. Fanghanel, I. Grenthe, V. Neck, D. Palmer, and M. H. Rand (2003), Update on the chemical thermodynamics of uranium, neptunium, plutonium, americium and technetium, OECD-NEA, in *Chemical Thermodynamics*, vol. 5, pp. 182–187, Elsevier, Amsterdam.
Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, *48*, W08301, doi:10.1029/2011WR011044.
Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*, John Wiley & Sons, Inc., Hoboken, N. J.
Hou, Z., and Y. Rubin (2005), On minimum relative entropy concepts and prior compatibility issues in vadose zone inverse and forward modeling, *Water Resour. Res.*, *41*, W12425, doi:10.1029/2005WR004082.
Keating, E. H., J. Doherty, J. A. Vrugt, and Q. Kang (2010), Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality, *Water Resour. Res.*, *46*, W10517, doi:10.1029/2009WR008584.
Kitanidis, P. K. (1986), Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, *22*(4), 499–507.
Kohler, M., G. Curtis, D. Kent, and J. Davis (1996), Experimental investigation and modeling of uranium(VI) transport under variable chemical conditions, *Water Resour. Res.*, *32*(12), 3539–3551.
Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, *Water Resour. Res.*, *48*, W01526, doi:10.1029/2011WR010608.
Laloy, E., B. Rogiers, J. A. Vrugt, D. Mallants, and D. Jacques (2013), Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov Chain Monte Carlo simulation and polynomial chaos expansion, *Water Resour. Res.*, *49*, 2664–2682, doi:10.1002/wrcr.20226.
Leavitt, J. J., K. J. Howe, and S. E. Cabaniss (2011), Equilibrium modeling of U(VI) speciation in high carbonate groundwaters: Model error and propagation of uncertainty, *Appl. Geochem.*, *26*, 2019–2026, doi:10.1016/j.apgeochem.2011.06.031.
Liu, C., J. M. Zachara, N. P. Qafoku, and Z. Wang (2008), Scale-dependent desorption of uranium from contaminated subsurface sediments, *Water Resour. Res.*, *44*, W08413, doi:10.1029/2007WR006478.
Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, *43*, W07401, doi:10.1029/2006WR005756.
Lu, D., M. Ye, and S. P. Neuman (2011), Dependence of Bayesian model selection criteria and Fisher information matrix on sample size, *Math. Geosci.*, *43*(8), 971–993, doi:10.1007/s11004-011-9359-0.
Lu, D., M. Ye, and M. C. Hill (2012), Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification, *Water Resour. Res.*, *48*, W09521, doi:10.1029/2011WR011289.
Lu, D., M. Ye, P. D. Meyer, G. P. Curtis, X. Shi, X.-F. Niu, and S. B. Yabusaki (2013), Effects of error covariance structure on estimation of model averaging weights and predictive performance, *Water Resour. Res.*, *49*, 6029–6047, doi:10.1002/wrcr.20441.
Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, *40*, W02501, doi:10.1029/2003WR002378.
Matott, L. S., and A. J. Rabideau (2008a), Calibration of complex subsurface reaction models using a surrogate-model approach, *Adv. Water Resour.*, *31*(2), 1697–1707, doi:10.1016/j.advwatres.2008.08.006.
Matott, L. S., and A. J. Rabideau (2008b), Calibration of subsurface batch and reactive-transport models involving complex biogeochemical processes, *Adv. Water Resour.*, *31*(2), 269–286.
Matott, L. S., J. E. Babendreier, and S. T. Purucker (2009), Evaluating uncertainty in integrated environmental models: A review of concepts and tools, *Water Resour. Res.*, *45*, W06421, doi:10.1029/2008WR007301.
Meyer, P. D., M. Ye, M. L. Rockhold, S. P. Neuman, and K. J. Cantrell (2007), Combined estimation of hydrogeologic conceptual model, parameter, and scenario uncertainty, *NUREG/CR-6940, PNNL-16396*, U.S. Nucl. Regul. Comm., Off. of Nucl. Regul. Res., Rockville, Md.
Mugunthan, P., C. A. Shoemaker, and R. G. Regis (2005), Comparison of function approximation, heuristic, and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models, *Water Resour. Res.*, *41*, W11427, doi:10.1029/2005WR004134.
Neuman, S. P. (1973), Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty, *Water Resour. Res.*, *9*(4), 1006–1021, doi:10.1029/WR009i004p01006.
Neuman, S. P., and S. Yakowitz (1979), Statistical approach to the inverse problem of aquifer hydrology. 1. Theory, *Water Resour. Res.*, *15*(4), 845–860, doi:10.1029/WR015i004p00845.

Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of alternative, conceptual-mathematical models, *Stochastic Environ. Res. Risk Assess.*, *17*(5), 291–305, doi: 10.1007/s00477-003-0151-7.
Over, M. W., Y. Yang, X. Chen, and Y. Rubin (2013), A strategy for improved computational efficiency of the method of anchored distributions, *Water Resour. Res.*, *49*, 3257–3275, doi:10.1002/wrcr.20182.
Papoulis, A. (1991), *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, N. Y.
Poeter, E. P., and D. A. Anderson (2005), Multimodel ranking and inference in ground water modeling, *Ground Water*, *43*(4), 597–605, doi: 10.1111/j.1745-6584.2005.0061.x.
Poeter, E., M. Hill, E. Banta, S. Mehl, and S. Christensen (2005), UCODE–2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation, U.S. Geol. Surv., Reston, Va.
Razavi, S., B. A. Tolson, and D. H. Burn (2012), Review of surrogate modeling in water resources, *Water Resour. Res.*, *48*, W07401, doi: 10.1029/2011WR011527.
Refsgaard, J. C., S. Christensen, T. O. Sonnenborg, D. Seifert, A. L. Hojberg, and L. Troldborg (2012), Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, *Adv. Water Resour.*, *36*, 36–50.
Rubin, Y., X. Chen, H. Murakami, and M. Hahn (2010), A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, *Water Resour. Res.*, *46*, W10523, doi:10.1029/2009WR008799.
Scheibe, T. D., P. Meakin, P. C. Lichtner, and D. Zachmann (2008), Science for problems under the surface, *SciDAC Rev.*, *7*, 32–41.
Schoups, G., and J. A. Vrugt (2010), A formal generalized likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, *46*, W10531, doi:10.1029/2009WR008933.
Seber, G. A. F., and C. J. Wild (2003), Nonlinear Regression, John Wiley & Sons, Hoboken, N. J.
Shi, X., M. Ye, S. Finsterle, and J. Wu (2012), Comparing nonlinear regression and Markov chain Monte Carlo methods for assessment of predictive uncertainty in vadose zone modeling, *Vadose Zone J.*, *11*(4), doi:10.2136/vzj2011.0147.
Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, *44*, W00B05, doi:10.1029/2007WR006705.
Srinivasan, G., D. M. Tartakovsky, B. A. Robinson, and A. B. Aceves (2007), Quantification of uncertainty in geochemical reactions, *Water Resour. Res.*, *43*, W12415, doi:10.1029/2007WR006003.
Steefel, C. I., D. J. DePaolo, and P. C. Lichtner (2005), Reactive transport modeling: An essential tool and a new research approach for the earth sciences, *Earth Planet. Sci. Lett.*, *240*, 539–558.
Tang, G., M. A. Mayes, J. C. Parker, X. L. Yin, D. B. Watson, and P. M. Jardine (2009), Improving parameter estimation for column experiments by multi-model evaluation and comparison, *J. Hydrol.*, *376*, 567–578, doi:10.1016/j.jhydrol.2009.07.063.
Tartakovsky, D. M., M. Dentz, and P. C. Lichtner (2009), Probability density functions for advective-reactive transport with uncertain reaction rates, *Water Resour. Res.*, *45*, W07414, doi:10.1029/2008WR007383.
Tartakovsky, D. M. (2013), Assessment and management of risk in subsurface hydrology: A review and perspective, *Adv. Water Resour.*, *51*, 247-260, doi: 10.1016/j.advwater.2012.04.007.
Ter Braak, C. J. F. (2006), A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces, *Stat. Comput.*, *16*(3), 239–249.
Ter Braak, C. J. F., and J. A. Vrugt (2008), Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, *18*, 435–446, doi:10.1007/s11222-008-9104-9.
Vecchia, A. V., and R. L. Cooley (1987), Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model, *Water Resour. Res.*, *23*(7), 1237–1250, doi: 10.1029/WR023i007p01237.
Vrugt, J. A., P. H. Stauffer, T. Wohling, B. A. Robinson, and V. V. Vesselinov (2008a), Inverse modeling of subsurface flow and transport properties: A review with new developments, *Vadose Zone J.*, *7*(2), 843–864, doi:10.2136/vzj2007.0078.
Vrugt, J. A., B. A. Robinson, and J. M. Hyman (2009a), Self-adaptive multimethod search for global optimization in real-parameter spaces, *IEEE Trans. Evol. Comput.*, *13*(2), 243–259, doi:10.1109/TEVC.2008.924428.
Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009b), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlinear Sci. Numer. Simul.*, *10*(3), 273–290.
Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stoch. Environ. Res. Risk Assess.*, *19*, 378–387.
Woodbury, A. D., and T. J. Ulrych (2000). A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Res. Res.*, *36*(8), 2081–2093.
Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, *40*, W05113, doi:10.1029/2003WR002557.
Ye, M., P. D. Meyer, and S. P. Neuman (2008), On model selection criteria in multimodel analysis, *Water Resour. Res.*, *44*, W03428, doi: 10.1029/2008WR006803.
Ye, M., D. Lu, S. P. Neuman, and P. D. Meyer (2010a), Comment on "Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window" by Frank T.-C. Tsai and Xiaobao Li, *Water Resour. Res.*, *46*, W02801, doi:10.1029/2009WR008501.
Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohll, and D. M. Reeves (2010b), A model-averaging method for assessing groundwater conceptual model uncertainty, *Ground Water*, *48*, 716–728, doi:10.1111/j.1745-6584.2009.00633.x.
Yeh, W. W.-G. (1986), Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*(2), 95–108, doi:10.1029/WR022i002p00095.
Yeh, W. W.-G., and Y. S. Yoon (1981), Aquifer parameter identification with optimum dimension in parameterization, *Water Resour. Res.*, *17*(3), 664–672, doi:10.1029/WR017i003p00664.
Zhang, G., D. Lu, M. Ye, M. Gunzburger, and C. Webster (2013), An adaptive sparse-grid high-order stochastic collocation method for Bayesian inference in groundwater reactive transport modeling, *Water Resour. Res.*, *49*, 6871–6892, doi:10.1002/wrcr.20467.