

一种改进的地下水模型结构不确定性分析方法

孙晓卓, 曾献奎, 吴吉春, 孙媛媛

An improved method of groundwater model structural uncertainty analysis

SUN Xiaozhuo, ZENG Xiankui, WU Jichun, and SUN Yuanyuan

在线阅读 View online: <https://doi.org/10.16030/j.cnki.issn.1000-3665.202012061>

您可能感兴趣的其他文章

Articles you may be interested in

基于高斯过程回归的地下水模型结构不确定性分析与控制

Quantification and reduction of groundwater model structural uncertainty based on Gaussian process regression

钟乐乐, 曾献奎, 吴吉春 水文地质工程地质. 2019, 46(1): 1-1

基于改进稀疏网格替代模拟的地下水DNAPLs运移不确定性分析

Uncertainty analysis of groundwater DNAPLs migration based on improved sparse grids surrogate model

高鑫宇, 曾献奎, 吴吉春 水文地质工程地质. 2020, 47(1): 1-10

基于CFP的岩溶管道流溶质运移数值模拟研究

A study of the solute transport model for karst conduits based on CFP

杨杨, 赵良杰, 苏春田, 夏日元 水文地质工程地质. 2019, 46(4): 51-57

一种可增加海岛地下淡水资源储量的方法研究

A method for improving the fresh groundwater storage of oceanic islands

马婧, 鲁春辉, 吴吉春, 罗剑 水文地质工程地质. 2020, 47(3): 1-7

基于多点地质统计与集合平滑数据同化方法识别非高斯渗透系数场

Characterization of non-Gaussian hydraulic conductivity fields using multiple-point geostatistics and ensemble smoother with multiple data assimilation method

宗成元, 康学远, 施小清, 吴吉春 水文地质工程地质. 2020, 47(2): 1-8

一种基于弹性能释放率的岩石新型统计损伤本构模型

A statistical damage constitutive rock model based on elastic energy release rate

刘文博, 孙博一, 陈雷, 张树光 水文地质工程地质. 2021, 48(1): 88-95



关注微信公众号, 获得更多资讯信息

DOI: [10.16030/j.cnki.issn.1000-3665.202012061](https://doi.org/10.16030/j.cnki.issn.1000-3665.202012061)

一种改进的地下水模型结构不确定性分析方法

孙晓卓, 曾献奎, 吴吉春, 孙媛媛

(南京大学地球科学与工程学院/表生地球化学教育部重点实验室, 江苏 南京 210023)

摘要: 高斯过程回归(GPR)是一种基于贝叶斯理论的监督学习算法,在基于数据驱动(DDM)的模型结构不确定性分析中具有广泛应用。目前研究中通常假设物理参数和超参独立并进行联立识别,这会导致参数补偿。文章提出两步识别DDM量化模型结构误差,并通过2个地下水模型案例,分别在不考虑模型结构误差、考虑模型结构误差(联立识别DDM、两步识别DDM)的情况下,对比分析了参数识别和模型预测结果。结果表明,不考虑模型结构误差直接进行参数识别时,为补偿结构误差,物理参数会过度拟合,从而影响模型预测效果。基于DDM刻画模型结构偏差时,物理参数和超参的独立性假设会影响参数识别结果。提出的两步识别DDM法没有假设物理参数和超参独立,能够减少参数过度拟合效应,从而更准确刻画结构误差,有效提高了模型的预测性能。

关键词: 模型结构不确定性;数据驱动;高斯过程回归;马尔科夫链蒙特卡洛模拟;溶质运移

中图分类号: P641.2

文献标志码: A

文章编号: 1000-3665(2021)06-0024-10

An improved method of groundwater model structural uncertainty analysis

SUN Xiaozhuo, ZENG Xiankui, WU Jichun, SUN Yuanyuan

(Key Laboratory of Surficial Geochemistry of the Ministry of Education/School of Earth Sciences and Engineering, Nanjing University, Nanjing, Jiangsu 210023, China)

Abstract: Gaussian Process Regression (GPR) is a supervised learning algorithm based on Bayesian theory, which is widely used in model structural uncertainty analysis based on data-driven method (DDM). In this study, it is usually assumed that the physical parameters and hyperparameters are independent and identified jointly, which will lead to parameter compensation. In this paper, a two-stage based DDM method is proposed to quantify the model structural errors, and two case studies are used to compare and analyze the results of parameter identification and model prediction with considering the model structural errors (joint calibration based DDM and two-stage based DDM) and without considering the model structural errors. The results show that when the parameters are identified directly without considering the model structural errors, the parameters will be overfitted and compensate the model structural errors, thereby affecting the model prediction performance. When considering the model structure deviation based on DDM, the independence assumption of physical parameters and hyperparameters will affect the parameter estimation results. The proposed two-stage based DDM method does not assume that the physical parameters and hyperparameters are independent, and can reduce parameter overfitting caused by the independence assumption of physical parameters and hyperparameters, portraying more accurate structural errors and effectively improving the model prediction performance.

收稿日期: 2020-12-15; 修订日期: 2021-02-18

基金项目: 国家重点研发计划“场地土壤污染成因与治理技术”重点专项(2018YFC1800604); 国家自然科学基金项目(42072272)

第一作者: 孙晓卓(1998-), 女, 硕士研究生, 主要从事地下水数值模拟研究。E-mail: 1249879295@qq.com

通讯作者: 曾献奎(1985-), 男, 副教授, 主要从事地下水数值模拟研究。E-mail: zengxiankui@yeah.net

Keywords: model structure uncertainty; data-driven method; Gaussian process regression; MCMC; solute transport

地下水数值模型能够定量刻画地下水流和污染物的时空分布特征,已被广泛应用于地下水资源的科学管理和保护^[1-5]。由于实际地下水系统十分复杂,限于认知不足,通常需对其进行简化,这将导致建立的水文地质概念模型与实际水文地质条件之间存在差异,即模型存在结构误差,如含水层空间分布的刻画偏差、化学反应动力学模型存在偏差等。使用带有结构误差的模型进行预测时,即使进行参数识别,预测结果也可能存在偏差^[6-8]。因此,为提高模拟结果的可靠性,需要定量刻画模型的结构误差,从而可为地下水资源管理和决策提供更可靠的依据。

近十年来,众多研究者开始注重对模型结构不确定性的研究^[9-12]。贝叶斯模型平均(BMA)和基于数据驱动的结构误差统计学习方法是目前常用的处理方法。BMA在贝叶斯框架下建立一组可行的概念模型,根据每个模型的表现赋予相应的权重,从而对这些模型的预测结果进行加权平均^[13-18]。研究表明,对比于其他多模型方法或单一模型法,采用BMA方法能得到更准确和可靠的预测结果^[19-20]。尽管如此,BMA在实际应用过程中存在局限性,如模型空间(概念模型集合)的定义及先验权重具有主观性、模型间相关性未能准确量化等,从而影响BMA的预测效果^[20-21]。

数据驱动法(DDM)通过某种统计模型刻画模型结构误差,如支持向量机回归^[22-23]、人工神经网络^[24]、随机森林^[25]、高斯过程回归^[26-27]等机器学习方法。其中,高斯过程回归(GPR)是一种基于贝叶斯理论的监督学习算法,被广泛用于模型结构误差的学习。Xu等^[26-27]利用GPR刻画多个地下水流模型的结构误差,通过马尔科夫链蒙特卡洛模拟(MCMC)识别地下水模型参数(物理参数)和GPR参数(超参)。结果表明,GPR可以刻画模型结构误差的时空相关性,显著提高地下水模型的预测能力。Pan等^[28]通过GPR刻画地下水中重质非水相液体(DNAPL)运移模型的结构误差,并进行针对地下水DNAPL污染的人类健康风险评价。结果表明与不考虑模型结构误差相比,通过GPR刻画DNAPL运移模型结构误差,能够在关键区域获得更准确的人体健康风险评价。

尽管如此,当前基于GPR进行模型结构误差刻画

的研究中,通常假设物理参数和超参相互独立,且对这些参数进行联立识别。然而,在实际情况下物理参数和超参的独立性假设不够合理。同时,超参没有物理意义,进行物理参数和超参联立识别时,可能导致参数的过度识别,影响模型预测效果^[29-30]。

本文提出两步识别DDM方法进行参数识别和模型结构误差定量刻画,克服了传统联立识别DDM方法中的过度拟合问题,并通过一个地下水溶质运移案例和一个地下水水流案例,对提出方法的可靠性和优越性进行了验证和分析。

1 研究方法

1.1 高斯过程回归(GPR)

地下水模型中观测值 f 可以表示为:

$$f = M(\theta) + b(x, \varphi) + \eta \quad (1)$$

式中: M ——地下水模型;

θ ——地下水模型参数(物理参数);

b ——模型结构误差;

x ——模型输入;

φ ——超参;

η ——随机测量误差。

在GPR中,假定结构误差 $b(x, \varphi)$ 在空间各点服从均值为 μ ,协方差矩阵为 C 的联合多元高斯分布^[31-32]。本文均值函数 μ 为0,协方差矩阵 C 采用平方指数型:

$$C_{i,j} = p(x_i, x_j) = \sigma^2 \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{\lambda^2}\right) + \sigma_s^2 I \quad (2)$$

$i, j = 1, 2, \dots, n$

式中: λ ——特征长度;

σ^2 ——控制 $b(x, \varphi)$ 的边缘方差;

σ_s^2 ——测量误差;

n ——观测数据的个数;

I —— $n \times n$ 阶的单位矩阵。

测量误差 η 的先验为独立同分布 $N(0, \sigma_s^2)$ 。物理参数 θ 和超参 $\varphi = \{\lambda, \sigma^2, \sigma_s^2\}$ 被认为是相互独立并服从先验分布 $p(\theta, \varphi) = p(\theta)p(\varphi)$ 。根据贝叶斯原理,推导出 θ 与 φ 的后验分布为:

$$p(\theta, \varphi | f) \propto p(f | \theta, \varphi) p(\theta) p(\varphi) \quad (3)$$

式中: $p(\theta)$ —— θ 的先验函数;

$p(\varphi)$ —— φ 的先验函数。

似然函数 $p(f|\theta, \varphi)$ 表示为:

$$\lg(p(f|\theta, \varphi)) = -\frac{1}{2}(f - M - \mu)^T C^{-1}(f - M - \mu) - \frac{1}{2} \lg|C| - \frac{n}{2} \lg(2\pi) \quad (4)$$

由于参数后验 $p(\theta, \varphi|f)$ 很难通过解析法直接计算得到, 使用 MCMC 模拟获得 θ 和 φ 的后验样本, 从而得到其后验分布。

对结构误差进行模拟预测时, 由于 $f|M$ 与 b^* 的先验联合分布为多元正态分布, 所以 b^* 的后验分布为:

$$b^*|f - M, f, f^* \sim N(\bar{b}^*, C_{bb}(b^*)) \quad (5)$$

$$\bar{b}^* = \mu^* + C^{*T} C^{-1}(f - M - \mu) \quad (6)$$

$$C_{bb}(b^*) = C^{**} - C^{*T} C^{-1} C^* \quad (7)$$

式中: \bar{b}^* ——后验均值;

$C_{bb}(b^*)$ ——后验协方差矩阵;

μ^* ——先验均值, 即 $\mu(x^*, \varphi)$ 。

与 C 相同, 先验协方差矩阵 C^* , C^{**} 定义为 $C_{i,j}^* = p(x_i, x_j)$,

$C_{i,j}^{**} = p(x_i^*, x_j^*)$ 。

模型预测值表示为:

$$f^* = M^* + b^* + \eta^* \quad (8)$$

式中: f^* ——预测值;

M^* ——模型预测输出值;

b^* ——结构误差预测值;

η^* ——测量误差预测值。

1.2 基于联立识别的数据驱动方法

传统的数据驱动方法(DDM)中, 通常假设物理参数 θ 和超参 φ 独立分布, 通过 MCMC 模拟对其进行联立识别^[26, 28]。本次研究中采用改进的差分进化自适应算法 DREAMzs(DiffeREntial Evolution Adaptive Metropolis)^[33-34]进行 MCMC 模拟, 基于联立识别的 DDM 基本步骤如下:

(1) 定义物理参数 θ 和超参 φ 的先验分布;

(2) 通过 MCMC 模拟获得物理参数 θ 和超参 φ 的后验样本;

(3) 对于新的模型输入 x^* , 输入模型获得输出 M^* , 根据式(2)计算协方差矩阵 C 、 C^* 、 C^{**} ;

(4) 根据式(5)–(7), 使用多元正态抽样生成 b^* 、 η^* ;

(5) 通过式(8)计算预测值 f^* 。

1.3 基于两步识别的数据驱动方法

假设物理参数 θ 和超参 φ 不独立, 将其分开进行两步识别。首先, 在 φ 的全部先验分布空间内识别

θ 。根据贝叶斯理论, 物理参数 θ 后验概率 $p(\theta|f, M)$ 可以表示为:

$$p(\theta|f, M) = \frac{p(f|\theta, M)p(\theta)}{p(f|M)} \quad (9)$$

其中, $p(f|\theta, M)$ 为模型 $M(\theta)$ 的边缘似然值, 即 $M(\theta)$ 在 φ 的全部先验分布空间内似然函数的平均值。其表达式为:

$$p(f|\theta, M) = \int p(f|\varphi, \theta, M)p(\varphi) d\varphi \quad (10)$$

其中, $p(f|\varphi, \theta, M)$ 为物理参数 θ 和超参 φ 的联合似然函数, $p(\varphi)$ 为 φ 的先验分布。

获得 θ 的后验概率 $p(\theta|f, M)$ 之后, 超参后验概率 $p(\varphi|f, M)$ 可以表示为:

$$p(\varphi|f, M) = \frac{p(f|\varphi, M)p(\varphi)}{p(f|M)} \quad (11)$$

其中, $p(f|\varphi, M)$ 是模型 M 的边缘似然值, 即 $M(\varphi)$ 在 θ 的全部后验分布空间内似然函数的平均值。其表达式为:

$$p(f|\varphi, M) = \int p(f|\theta, \varphi, M)p(\theta|f, M) d\theta \quad (12)$$

边缘似然值即式(10)(12)表示似然函数在高维参数概率分布空间内的积分, 由于非线性似然函数的复杂性, 边缘似然值的解析表达式难以获得。本文采用算术平均估计方法 AME^[35-36] 计算式(10)(12), 计算公式为:

$$p(f|M) = \frac{1}{m} \sum_{i=1}^m p(f|\chi_i, M) \quad (13)$$

式中: χ_i ——模型 M 参数概率分布空间内的一个样本;

m —— χ_i 的样本数量。

通过 AME 计算边缘似然值通常需要大量的样本, m 数值一般为几十万至几千万, 一个样本代表运行物理模型一次。此外, 通过 MCMC 方法获得 θ 和 φ 后验分布(即式(9)(11))时需多次使用 AME。因此, 模型边缘似然值的计算具有严重的计算负荷。本文使用当前应用广泛的自适应稀疏网格(adaptive SG)技术^[37-41]构建边缘似然值的替代模型, 克服计算耗时困难。

通过式(9)–(13)获得 θ 和 φ 的后验分布之后, 进行模型预测。两步识别 DDM 法计算步骤为(图 1):

(1) 从 θ 的后验样本中, 选择一组在整个超参空间内表现最优的参数 θ_{\max} (即边缘似然值 $p(f|\theta, M)$ 最大), 并在物理模型中将其固定为常量;

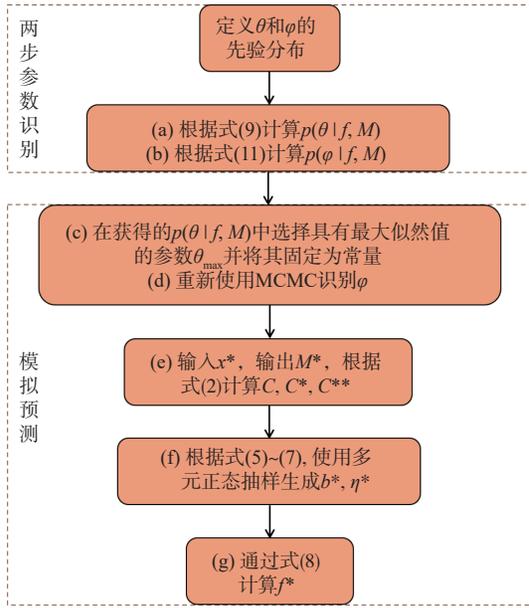


图 1 两步识别 DDM 法计算步骤图

Fig. 1 Flow chart of the two-stage based DDM method

- (2) 重新使用 MCMC 识别 ϕ 得到其后验分布;
- (3) 对于新的模型输入 x^* , 输入模型获得输出 M^* , 根据式(2)计算协方差矩阵 C, C^*, C^{**} ;
- (4) 根据式(5)–(7), 使用多元正态抽样生成 b^*, η^* ;
- (5) 通过式(8)计算预测值 f^* 。

2 地下水溶质运移案例分析

假设真实情况下的溶质运移过程为非平衡等温吸附模型, 实际使用线性平衡吸附模型, 因此存在模型结构误差。分别在不考虑模型结构误差、考虑模型结构误差(联立识别 DDM、两步识别 DDM)进行参数识别及模型预测。

2.1 溶质运移案例概况

该案例描述的是在定浓度通量边界条件下, 一维稳定流溶质运移模型。真实情况下的模型为非平衡等温吸附模型, 解析表达式见式(14)–(17)。其中, 输入浓度 C_0 为 50 mol/L, 水流速度 v 为 50 cm/d, 弥散系数 D 为 40 cm²/d, 注入污染物的时间 t_0 为 5 d, 系数 γ 为 0, 衰变常数 μ 为 0.1, 阻滞因数 R 为 1.3, 初始浓度 C_i 为 0 mol/L。

$$c(x, t) = \begin{cases} \gamma/\mu + (C_0 - \gamma/\mu)A(x, t) + \\ (C_i - \gamma/\mu)B(x, t) & 0 < t \leq t_0 \\ \gamma/\mu + (C_0 - \gamma/\mu)A(x, t) + \\ (C_i - \gamma/\mu)B(x, t) - C_0A(x, t - t_0) & t > t_0 \end{cases} \quad (14)$$

$$A(x, t) = \frac{1}{2} \exp\left[\frac{(v-u)x}{2D}\right] \operatorname{erfc}\left[\frac{Rx-ut}{2(DRt)^{1/2}}\right] + \frac{1}{2} \exp\left[\frac{(v+u)x}{2D}\right] \operatorname{erfc}\left[\frac{Rx+ut}{2(DRt)^{1/2}}\right] \quad (15)$$

$$B(x, t) = \exp\left(-\frac{\mu t}{R}\right) \times \left\{ \begin{aligned} & 1 - \frac{1}{2} \operatorname{erfc}\left[\frac{Rx-vt}{2(DRt)^{1/2}}\right] - \\ & \frac{1}{2} \exp\left(\frac{vx}{D}\right) \operatorname{erfc}\left[\frac{Rx+vt}{2(DRt)^{1/2}}\right] \end{aligned} \right\} \quad (16)$$

$$u = v(1 + 4\mu D/v^2)^{1/2} \quad (17)$$

将真实模型运行后得到浓度观测值, 同时考虑时间尺度($t: 0 \sim 10$ h)和空间尺度($x: 0 \sim 250$ cm), 如图 2 所示, 分别选取 21 个和 36 个观测数据用于模型识别和验证。

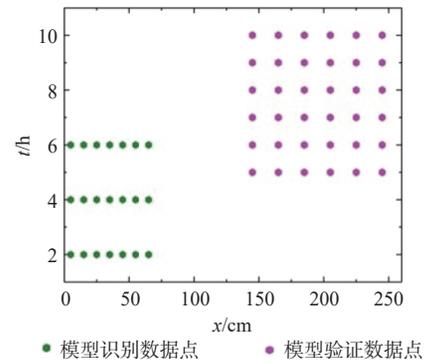


图 2 模型识别和验证数据点的位置

Fig. 2 Location of model observation wells during calibration and validation periods

实际使用的模型为线性平衡吸附模型, 解析表达式为:

$$c(x, t) = \begin{cases} (C_0 - C_i)A(x, t) & 0 < t \leq t_0 \\ (C_0 - C_i)A(x, t) - C_0A(x, t - t_0) & t > t_0 \end{cases} \quad (18)$$

$$A(x, t) = \frac{1}{2} \operatorname{erfc}\left[\frac{Rx-ut}{2(DRt)^{1/2}}\right] + \frac{1}{2} \exp\left[\frac{vx}{D}\right] \operatorname{erfc}\left[\frac{Rx+ut}{2(DRt)^{1/2}}\right] \quad (19)$$

基于该模型描述溶质运移过程时, 污染物吸附模型的偏差会导致模拟结果受到模型结构不确定性的影响。

2.2 案例分析

基于线性平衡吸附模型, 分别不考虑和考虑模型结构误差进行参数识别。假设输入浓度 C_0 和水流速度 v 为随机变量, 其余参数与真实模型相同。表 1 所示为物理参数与超参的先验分布。

采用 DREAMzs 算法进行参数识别, 设置 3 条平

表 1 模型参数的先验分布

Table 1 Prior distributions of model parameters

参数	先验分布
$C_0/(\text{mol}\cdot\text{L}^{-1})$	Uniform on [45.0,52.0]
$V/(\text{cm}\cdot\text{d}^{-1})$	Uniform on [45.0,52.0]
λ	Gamma, $k=5, \theta=0.2$, on [0.1,0.8]
σ	Exponential, $\mu=0.25$, on [3.0,10.0]
σ_δ	Uniform on [0.1,0.5]

行的马尔科夫链, 每条链的预热长度为 6 000, 预热后迭代长度为 6 000。对于不考虑和考虑模型结构误差情况下的参数识别, 均选取收敛后的 18 000 次样本统计边缘后验分布, 结果见图 3。

从图 3 可知, 3 种方法得到的物理参数后验分布差异明显, 考虑模型结构误差的情况下, 参数后验范

围明显变小, 呈正态分布的形式。对于参数 C_0 和 v , 联立识别 DDM 得到的参数后验分布范围均最小, 联立识别 DDM 和两步识别 DDM 具有相近的后验均值。两种 DDM 识别得到的超参边缘后验分布也存在显著差异。对比联立识别 DDM, 两步识别 DDM 得到的超参后验分布范围均较大。对于 λ , 两步识别法得到的参数后验分布集中在 0.4, 联立识别法的后验分布集中在 0.3。对于 σ , 两步识别法的后验分布趋势平缓, 联立识别法后验分布集中于 3.0。对于 σ_δ , 两种 DDM 得到的后验分布均集中在 0.1, 联立识别 DDM 具有较大的峰值概率密度。因此, 忽略模型结构偏差直接进行参数识别会导致参数补偿。基于 DDM 考虑模型结构偏差时, 物理参数和超参的独立性假设会影响参数识别结果。

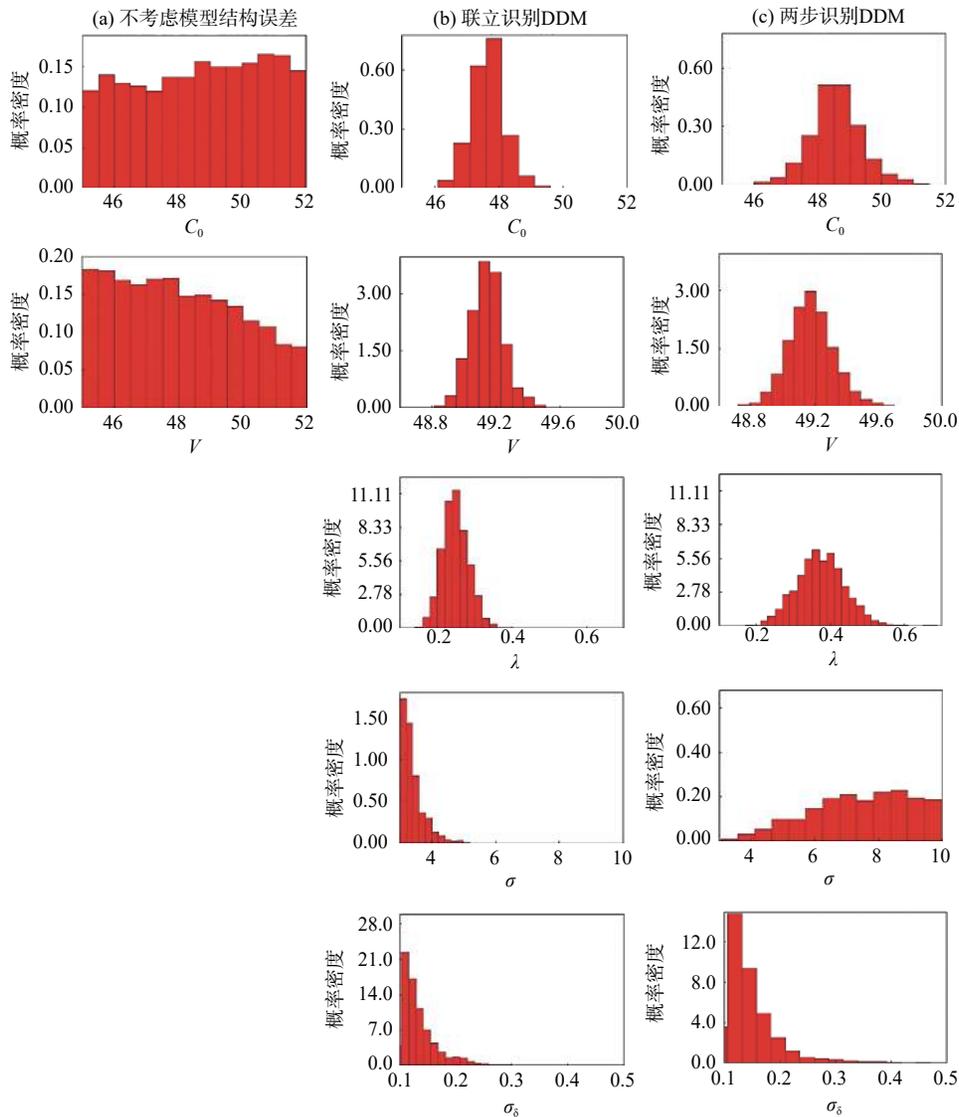


图 3 识别得到的模型参数和超参边缘后验分布

Fig. 3 Identified marginal posterior distributions of model parameters and hyperparameters

基于识别得到的模型参数后验分布, 获得模型在识别期和验证期的模拟值, 通过频率统计获得模拟值的置信区间和平均值等信息。本次研究采用 3 个指标评价模型在识别期和验证期的表现: 均方根误差 (RMSE)、预测误差均值 (MAE)、相对误差的均值 (MRE)。这 3 个指标值越小, 表明模拟值越接近观测值, 模型预测性能越好, 各指标统计结果见表 2。

表 2 模型预测性能指标统计结果

Table 2 Statistics of model prediction performance

	识别期			验证期		
	RMSE	MAE	MRE	RMSE	MAE	MRE
不考虑结构误差	4.605 6	4.105 6	0.197 4	8.958 3	6.769 9	0.227 9
联立识别DDM	5.316 5	4.952 1	0.222 5	8.235 6	6.388 9	0.224 0
两步识别DDM	4.650 0	4.291 6	0.209 4	7.777 0	5.625 5	0.206 8

从表 2 可以看出, 在模型识别期, 不考虑模型结构误差得到的 RMSE 值、MAE 值、MRE 值均最低, 识别效果最好; 在模型验证期, 不考虑模型结构误差得到的 RMSE 值、MAE 值、MRE 值均最高, 预测效果最差。结果证明, 不考虑结构误差直接进行参数识别会产生参数补偿问题, 即参数过度拟合补偿结构误差, 从而影响模型预测效果。对于两种 DDM, 两步识别 DDM 法在模型识别期和验证期均得到较低的 RMSE、MAE、MRE。因此, 通过两步识别 DDM 法量化模型结构误差, 能够减少物理参数和超参独立性假设导致的参数过度拟合, 从而提高模型的预测效果。

3 地下水水流案例分析

假设真实情况下地下水流模型为有越流过程的定流量非完整井流模型, 实际使用的模型没有考虑越流过程, 因此存在模型结构误差。分别在不考虑和考虑模型结构误差 (联立识别 DDM、两步识别 DDM) 条件下进行参数识别及模型预测。

3.1 水流案例概况

该案例刻画了在第一类越流系统中, 定流量非完整井流模型。真实情况下的模型为有越流过程的定流量非完整井流模型 (图 4)。抽水井以定流量 $Q=8\ 000\ \text{m}^3/\text{d}$ 向外抽水, 过滤管位于深度 $z=d$ 和 $z=l$ 之间 ($d=4\ \text{m}$, $l=8\ \text{m}$)。观测井水位表示含水层中 (r, z) 处的水头值, 降深 s 的解析表达式见式 (20)。其中, 观测时刻 $t=8\ \text{d}$, 承压含水层储水系数 μ 为 0.003, 垂向渗透系数 K_{zz} 为 $5\ \text{m}/\text{d}$ 。

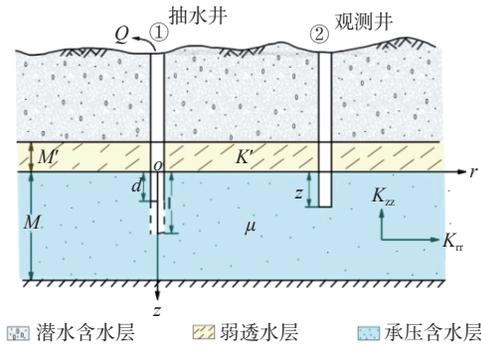


图 4 定流量非完整井流模型示意图

Fig. 4 Diagrammatic representation of partially penetrating well pumping at a constant rate

$s(r, z, t)$

$$= \frac{Q}{4\pi K_{rr} M} \left\{ W\left(u_r, \frac{r}{B_r}\right) + \frac{2M}{\pi(l-d)} \sum_{n=1}^{\infty} \frac{1}{n} \begin{bmatrix} \sin\left(\frac{n\pi l}{M}\right) \\ -\sin\left(\frac{n\pi d}{M}\right) \end{bmatrix} \right\} \times \cos\left(\frac{n\pi z}{M}\right) W\left[u_r, \sqrt{\left(\frac{r}{B_r}\right)^2 + \frac{K_{zz}}{K_{rr}} \left(\frac{n\pi r}{M}\right)^2}\right] \quad (20)$$

$$u_r = \frac{r^2}{4a_r t}, \quad a_r = \frac{K_{rr} M}{\mu}, \quad B_r^2 = \frac{K_{rr} M M}{K'}$$

将真实模型运行后得到降深观测值 (图 5), 考虑不同方向的空间尺度 ($r:0 \sim 20\ \text{m}$; $z:0 \sim 20\ \text{m}$), 时间固定在 $t=8\ \text{d}$, 分别选取 21, 58 个观测数据用于模型识别和验证。

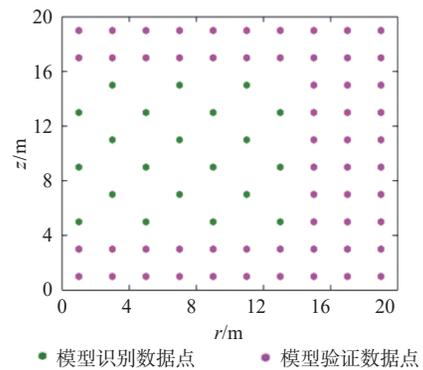


图 5 模型识别和验证数据点的位置

Fig. 5 Location of model observation wells during calibration and validation periods

实际使用的模型没有考虑越流过程, 即弱透水层的厚度 M' 为 $0\ \text{m}$, 渗透系数 K' 不存在, 其余参数设置和真实模型相同, 解析表达式见式 (20)。基于该模型描述定流量非完整井抽水引起的水位变化时, 由于对越流过程刻画的差异, 导致模型存在结构误差, 预测

结果受模型结构不确定性的影响。

3.2 案例分析

基于没有考虑越流过程的井流模型,分别不考虑和考虑模型结构误差进行参数识别。假设承压含水层水平渗透系数 K_{rr} 和厚度 M 为随机变量,其余参数与真实模型相同。表 3 为物理参数与超参的先验分布。

采用 DREAMzs 算法进行参数识别,设置 3 条平行的马尔科夫链,每条链预热 6 000 次,预热后迭代 6 000 次。不考虑和考虑模型结构误差(联立识别 DDM、

表 3 模型参数的先验分布

Table 3 Prior distributions of model parameters

参数	先验分布
$K_{rr}/(\text{m}\cdot\text{d}^{-1})$	Uniform on [8.0,14.0]
M/d	Uniform on [75.0,90.0]
λ	Gamma, $k=5, \theta=0.2$, on [0.2,0.6]
σ	Exponential, $\mu=0.25$, on [5.0,10.0]
σ_δ	Uniform on [0.05,0.5]

两步识别 DDM) 的参数边缘后验分布均取收敛后的 18 000 次样本,结果见图 6。

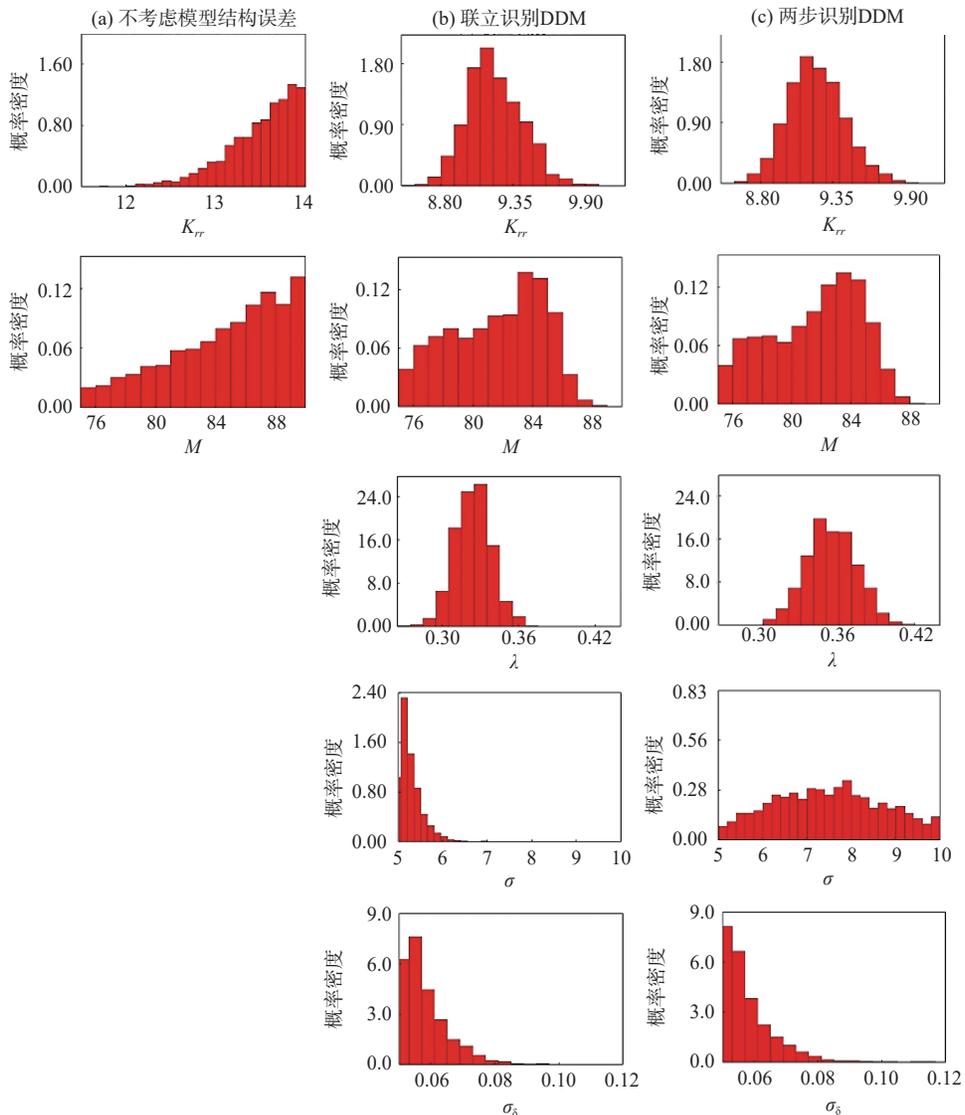


图 6 识别得到的模型参数和超参边缘后验分布

Fig. 6 Identified marginal posterior distributions of model parameters and hyperparameters

从图 6 可以看出,3 种方法得到的物理参数后验范围存在差异,且后验分布形态明显不同。考虑模型结构误差时,物理参数后验范围均较小,分布更集

中。对于 K_{rr} ,两种 DDM 的参数后验分布均集中在 9.2 附近,联立识别 DDM 具有较大的峰值概率密度。对于 M ,两种 DDM 的参数后验分布均集中在 84 附

近,两步识别 DDM 具有较大的概率密度。此外,两种 DDM 得到的超参边缘后验分布差异明显,两步识别 DDM 得到的超参后验分布范围均较大。对于 λ ,两步识别法的后验分布集中在 0.36, 联立识别法的后验分布集中在 0.33。对于 σ , 两步识别法的后验分布趋势平缓, 联立识别法的后验分布集中于 5.0。对于 σ_{δ} , 两种 DDM 得到的后验分布均集中在 0.05, 联立识别 DDM 具有较大的峰值概率密度。因此, 基于 DDM 刻画模型结构偏差时, 物理参数和超参的独立性假设会影响参数识别结果。

基于 DREAMzs 算法得到模型在识别期和验证期的模拟值, 通过频率统计获得模拟值的置信区间和平均值。模型预测性能指标统计见表 4。

表 4 模型预测性能指标统计结果
Table 4 Statistics of model prediction performance

	识别期			验证期		
	RMSE	MAE	MRE	RMSE	MAE	MRE
不考虑结构误差	5.208 4	3.559 1	0.222 6	3.576 4	3.172 8	0.448 8
联立识别DDM	8.138 0	7.984 3	0.657 6	1.891 2	1.524 9	0.225 3
两步识别DDM	8.039 0	7.782 1	0.631 5	1.718 4	1.358 9	0.195 4

从表 4 可以发现, 与地下水溶质运移案例结果类似(表 2), 在模型识别期, 不考虑模型结构误差得到的 RMSE 值、MAE 值、MRE 值均最低, 在模型验证期各指标值均最高。表明不考虑模型结构误差直接进行参数识别会影响模型预测效果。对于两种 DDM, 两步识别 DDM 法在模型识别期和验证期均得到较低的 RMSE、MAE、MRE。因此, 两步识别 DDM 能提高模型的预测精度。需要说明的是, 本次研究提出的两步识别 DDM 方法适用于任意形式的数学模型, 包括解析模型和数值模型, 因为对于实际复杂的水文地质条件, 建立的地下水模型均存在结构偏差。

4 结论

(1) 不考虑模型结构误差直接进行参数识别会产生参数补偿问题, 即通过参数过度拟合补偿模型结构误差, 从而影响模型的预测效果。

(2) 基于 DDM 刻画模型结构偏差时, 物理参数和超参的独立性假设会影响参数识别结果。相对于联立识别 DDM, 两步识别 DDM 得到模型预测的均方根误差(RMSE)、均值(MAE)、相对误差均值(MRE)均明显降低。通过两步识别 DDM 法量化模型结构误差, 能够减少物理参数和超参独立性假设而导致的参

数过度拟合, 从而能够更加准确地刻画结构误差, 预测结果更加可靠。

参考文献 (References) :

- [1] 薛禹群. 中国地下水数值模拟的现状与展望[J]. 高校地质学报, 2010, 16(1): 1 - 6. [XUE Yuqun. Present situation and prospect of groundwater numerical simulation in China[J]. *Geological Journal of China Universities*, 2010, 16(1): 1 - 6. (in Chinese with English abstract)]
- [2] DUMEDAH G, WALKER J P. Assessment of model behavior and acceptable forcing data uncertainty in the context of land surface soil moisture estimation[J]. *Advances in Water Resources*, 2017, 101: 23 - 36.
- [3] 周燕怡, 王旭升. 巴丹吉林沙漠潜水蒸发的数值模拟研究[J]. 水文地质工程地质, 2019, 46(5): 44 - 54. [ZHOU Yanyi, WANG Xusheng. Numerical simulation of groundwater evaporation in the Badain Jaran Desert of China[J]. *Hydrogeology & Engineering Geology*, 2019, 46(5): 44 - 54. (in Chinese with English abstract)]
- [4] 高焯, 梁收运, 王申宁, 等. 地下水数值模拟不确定性分析研究进展[J]. 地下水, 2020, 42(1): 28 - 31. [GAO Ye, LIANG Shouyun, WANG Shenning, et al. Research progress on uncertainty analysis of groundwater numerical simulation[J]. *Ground Water*, 2020, 42(1): 28 - 31. (in Chinese with English abstract)]
- [5] 陈梦佳, 吴剑锋, 孙晓敏, 等. 地下水典型非水相液体污染运移模拟的尺度提升研究[J]. 水文地质工程地质, 2020, 47(1): 11 - 18. [CHEN Mengjia, WU Jianfeng, SUN Xiaomin, et al. Upscaling of PCE transport modeling based on UTCHEM in heterogeneous porous media[J]. *Hydrogeology & Engineering Geology*, 2020, 47(1): 11 - 18. (in Chinese with English abstract)]
- [6] REFSGAARD J C, VAN DER SLUIJS J P, BROWN J, et al. A framework for dealing with uncertainty due to model structure error[J]. *Advances in Water Resources*, 2006, 29(11): 1586 - 1597.
- [7] WATSON T A, DOHERTY J E, CHRISTENSEN S. Parameter and predictive outcomes of model simplification[J]. *Water Resources Research*, 2013, 49(7): 3952 - 3977.
- [8] WU J C, ZENG X K. Review of the uncertainty analysis of groundwater numerical simulation[J]. *Chinese Science Bulletin*, 2013, 58(25): 3044 - 3052.
- [9] DOHERTY J, WELTER D. A short exploration of structural noise[J]. *Water Resources Research*, 2010,

- 46(5): W05525.
- [10] DOHERTY J, CHRISTENSEN S. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty[J]. *Water Resources Research*, 2011, 47(12): W12534.
- [11] ERDAL D, NEUWEILER I, HUISMAN J A. Estimating effective model parameters for heterogeneous unsaturated flow using error models for bias correction[J]. *Water Resources Research*, 2012, 48(6): W06530.
- [12] WHITE J T, DOHERTY J E, HUGHES J D. Quantifying the predictive consequences of model error with linear subspace analysis[J]. *Water Resources Research*, 2014, 50(2): 1152 – 1173.
- [13] DRAPER D. Assessment and propagation of model uncertainty[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1995, 57(1): 45 – 70.
- [14] HOETING J A, MADIGAN D, VOLINSKY R C T. Bayesian model averaging: a tutorial[J]. *Statistical Science*, 1999, 14(4): 382 – 401.
- [15] 杜新忠, 李叙勇, 王慧亮, 等. 基于贝叶斯模型平均的径流模拟及不确定性分析[J]. *水文*, 2014, 34(3): 6 – 10. [DU Xinzong, LI Xuyong, WANG Huiliang, et al. Multi-model ensemble runoff simulation based on Bayesian model averaging method and model structure uncertainty analysis[J]. *Journal of China Hydrology*, 2014, 34(3): 6 – 10. (in Chinese with English abstract)]
- [16] 王亮. 贝叶斯模型平均方法研究综述与展望[J]. *技术经济与管理研究*, 2016(3): 19 – 23. [WANG Liang. Overview and prospect of Bayesian model averaging[J]. *Technoeconomics & Management Research*, 2016(3): 19 – 23. (in Chinese with English abstract)]
- [17] 王倩, 师鹏飞, 宋培兵, 等. 基于贝叶斯模型平均法的洪水集合概率预报[J]. *水电能源科学*, 2016, 34(6): 64 – 66. [WANG Qian, SHI Pengfei, SONG Peibing, et al. Multi-model ensemble flood probability forecasting based on BMA[J]. *Water Resources and Power*, 2016, 34(6): 64 – 66. (in Chinese with English abstract)]
- [18] 江善虎, 任立良, 刘淑雅, 等. 基于贝叶斯模型平均的水文模型不确定性及集合模拟[J]. *中国农村水利水电*, 2017(1): 107 – 112. [JIANG Shanhu, REN Liliang, LIU Shuya, et al. An analysis of hydrological modeling and ensemble simulation uncertainty using the Bayesian model averaging[J]. *China Rural Water and Hydropower*, 2017(1): 107 – 112. (in Chinese with English abstract)]
- [19] ROJAS R, FEYEN L, DASSARGUES A. Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging[J]. *Water Resources Research*, 2008, 44(12): W12418.
- [20] LIU Z, MERWADE V. Separation and prioritization of uncertainty sources in a raster based flood inundation model using hierarchical Bayesian model averaging[J]. *Journal of Hydrology*, 2019, 578: 124100.
- [21] LU D, YE M, CURTIS G P. Maximum likelihood Bayesian model averaging and its predictive analysis for groundwater reactive transport models[J]. *Journal of Hydrology*, 2015, 529(3): 1859 – 1873.
- [22] CAO T T, ZENG X K, WU J C, et al. Integrating MT-DREAMzs and nested sampling algorithms to estimate marginal likelihood and comparison with several other methods[J]. *Journal of Hydrology*, 2018, 563: 750 – 765.
- [23] DEMISSIE Y K, VALOCCHI A J, MINSKER B S, et al. Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions[J]. *Journal of Hydrology*, 2009, 364(3/4): 257 – 271.
- [24] KHALIL A, ALMASRI M N, MCKEE M, et al. Applicability of statistical learning algorithms in groundwater quality modeling[J]. *Water Resources Research*, 2005, 41(5): W05010.
- [25] TESORIERO A J, GRONBERG J A, JUCKEM P F, et al. Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification[J]. *Water Resources Research*, 2017, 53(8): 7316 – 7331.
- [26] XU T F, VALOCCHI A J. A Bayesian approach to improved calibration and prediction of groundwater models with structural error[J]. *Water Resources Research*, 2015, 51(11): 9290 – 9311.
- [27] XU T F, VALOCCHI A J, YE M, et al. Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model[J]. *Water Resources Research*, 2017, 53(5): 4084 – 4105.
- [28] PAN Y, ZENG X K, XU H X, et al. Assessing human health risk of groundwater DNAPL contamination by quantifying the model structure uncertainty[J]. *Journal of Hydrology*, 2020, 584: 124690.
- [29] REICHERT P, SCHUWIRTH N. Linking statistical bias description to multiobjective model calibration[J]. *Water Resources Research*, 2012, 48(9): W09543.
- [30] BRYNJARSÓTTIR J, O'HAGAN A. Learning about physical parameters: the importance of model

- discrepancy[J]. *Inverse Problems*, 2014, 30(11): 114007.
- [31] KENNEDY M C, O'HAGAN A. Bayesian calibration of computer models[J]. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 2001, 63(3): 425 – 464.
- [32] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning[M]. Cambridge: MIT Press, 2006: 69-106.
- [33] VRUGT J A, BRAAK C J F T, DIKS C G H, et al. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling[J]. *International Journal of Nonlinear Sciences & Numerical Simulation*, 2009, 10(3): 273 – 290.
- [34] LALOY E, VRUGT J A. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing[J]. *Water Resources Research*, 2012, 50(3): 182 – 205.
- [35] KASS R E, RAFTERY A E. Bayesian factors[J]. *Journal of the American statistical association*, 1995, 90(430): 773 – 795.
- [36] LIU P G, ELSHALL A S, YE M, et al. Evaluating marginal likelihood with thermodynamic integration method and comparison with several other numerical methods[J]. *Water Resources Research*, 2016, 52(2): 734 – 758.
- [37] SMOLYAK S A. Quadrature and interpolation formulas for tensor products of certain classes of functions[J]. *Soviet Math Dokl*, 1963(4): 240 – 243.
- [38] MA X, ZABARAS N. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations[J]. *Journal of Computational Physics*, 2009, 228(8): 3084 – 3113.
- [39] ZENG X K, YE M, BURKARDT J, et al. Evaluating two sparse grid surrogates and two adaptation criteria for groundwater Bayesian uncertainty quantification[J]. *Journal of Hydrology*, 2016, 535: 120 – 134.
- [40] 侯泽宇, 卢文喜, 王宇. 基于替代模型的地下水 DNAPLs 污染源反演识别[J]. *中国环境科学*, 2019, 39(1): 188 – 195. [HOU Zeyu, LU Wenxi, WANG Yu. Surrogate-based source identification of DNAPLs-contaminated groundwater[J]. *China Environmental Science*, 2019, 39(1): 188 – 195. (in Chinese with English abstract)]
- [41] 高鑫宇, 曾献奎, 吴吉春. 基于改进稀疏网格替代模拟的地下水 DNAPLs 运移不确定性分析[J]. *水文地质工程地质*, 2020, 47(1): 1 – 10. [GAO Xinyu, ZENG Xiankui, WU Jichun. Uncertainty analysis of groundwater DNAPLs migration based on improved sparse grids surrogate model[J]. *Hydrogeology & Engineering Geology*, 2020, 47(1): 1 – 10. (in Chinese with English abstract)]

编辑: 张若琳