

DOI:10.16030/j.cnki.issn.1000-3665.201903084

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

高鑫宇, 曾献奎, 吴吉春

(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: 由于实际的水文地质条件具有复杂性和变异性, 地下水 DNAPLs 运移数值模拟的不确定性不可避免。为解决不确定性分析中因多次调用 DNAPLs 运移模型导致的计算耗时问题, 本次研究在传统稀疏网格 (SG) 替代模型的基础上, 提出了一种将局部自适应 (LA) 和维数自适应 (DA) 耦合的改进替代模型 DA-LA-SG。通过两个解析案例和一个 PCE 运移室内实验, 验证了 DA-LA-SG 的替代效率和精度, 并将其用于室内砂箱 PCE 运移模拟不确定性分析。研究表明: 在替代模型构建初期, LA-SG 的替代效率优于或接近于 DA-SG 和 DA-LA-SG, 随着插值节点数量的增多, DA-SG 和 DA-LA-SG 的替代效率逐渐优于 LA-SG, DA-LA-SG 具有最高的替代效率。DA-LA-SG 能够高效、高精度地建立 PCE 运移模型的似然函数替代模型, 且被用于 PCE 运移模拟参数不确定性分析。结果表明: 背景介质渗透率 k_1 、拟合系数 n_1 和透镜体渗透率 k_2 的可识别性较强, 背景介质孔隙度 μ_1 、透镜体孔隙度 μ_2 和拟合系数 n_2 的识别效果较差, 这些参数对饱和度和观测数据不敏感。

关键词: DNAPLs 运移; 稀疏网格; 替代模型; 不确定性分析; MCMC

中图分类号: P641.69; X523

文献标识码: A

文章编号: 1000-3665(2020)01-0001-10

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

GAO Xinyu, ZENG Xiankui, WU Jichun

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

Abstract: Uncertainty in the numerical simulation of groundwater DNAPLs migration is inevitable due to the complex and unknown hydrogeological conditions in the actual world. In order to overcome the computational burden caused by repetitive calls of the DNAPLs migration model in uncertainty analysis, this study is carried out, based on the traditional sparse grid (SG) surrogate model. In this paper, an improved SG method DA-LA-SG is proposed, which couples the local adaptive (LA) and dimensional adaptive (DA). The surrogate efficiency and accuracy of the proposed DA-LA-SG are verified through two analytical cases and a PCE migration laboratory experiment, and the proposed model is applied to the uncertainty analysis of indoor sandbox PCE migration simulation. The results show that in the early stage of establishing the surrogate model, the surrogate efficiency of LA-SG is better than or close to that of DA-SG and DA-LA-SG, but the efficiency of DA-SG and DA-LA-SG is gradually better than that of LA-SG with the increasing number of the interpolation

收稿日期: 2019-03-14; 修订日期: 2019-08-16

基金项目: 国家重点研发计划“水资源高效开发利用”重点专项(2016YFC0402802); 国家自然科学基金(41672233, 41571017); 中央高校基本科研业务费专项资金项目(020614380085)

第一作者: 高鑫宇(1995-), 男, 硕士研究生, 主要从事地下水数值模拟的研究。E-mail: xygaoweilan13@163.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

nodes. DA-LA-SG performs the best. DA-LA-SG can be used to efficiently and accurately establish the likelihood function surrogate model of the PCE migration model, and analyze the uncertainty of parameters in PCE migration simulation. The posterior distribution of the model parameters is obtained, and it is found that the background medium permeability k_1 , the fitting coefficient n_1 and the weak lens permeability k_2 can be identified clearly, but the background medium porosity μ_1 , the weak lens porosity μ_2 and the fitting coefficient n_2 are basically evenly distributed, which are poorly recognized and insensitive to saturation observations.

Keywords: DNAPLs migration; sparse grids; surrogate model; uncertainty analysis; MCMC

近年来,随着石油化工产品和农药的广泛使用,大量有机污染物进入地下水中,造成地下水资源的污染。非水相流体(Non-Aqueous Phase Liquids, NAPLs)有机污染物是当前地下水污染研究的重要对象^[1],根据其密度相对水的密度大小,分为轻非水相流体(LNAPLs)和重非水相流体(DNAPLs)。DNAPLs 难溶于水,且密度比水大,可能附着于含水层而成为地下水的长期污染源^[2-3]。DNAPLs 在地下水存在水相、油相和气相,它们之间相互转化构成多相流体系统。由于 DNAPLs 运移过程复杂、处理费用高昂,通过数值模拟技术定量描述 DNAPLs 的分布和迁移转化,成为当前 DNAPLs 场地污染修复与风险评价的重要手段^[4-6]。

由于地质环境的复杂性(如含水介质的非均质性、地层结构的分布等)、有限的观测数据以及 DNAPLs 的复杂运移规律,导致地下水 DNAPLs 运移模型的参数、结构难以进行有效识别,从而不能精确刻画其迁移转化过程,模拟结果具有不确定性。地下水模型的不确定性分析成为 DNAPLs 运移定量刻画的重要前提,基于贝叶斯理论的 MCMC (Markov Chain Monte Carlo)模拟是当前主要的不确定性定量分析方法。Liu^[7]应用 MCMC 对 DNAPLs 溶解及运移的半解析模型进行了参数估计,并用于 DNAPLs 污染修复模型的污染源识别^[8]。不确定性分析通常需要重复调用 DNAPLs 运移模型几千至几万次,模型运行时间成为当前 DNAPLs 运移模拟不确定性分析的重要问题。替代模型是当前克服不确定性分析计算耗时问题的有效途径,即通过建立与原始模型具有相同或接近模拟精度的模型代替原始模型。DNAPLs 模拟中常见的替代模型技术有支持向量机^[9-10]、克里金插值^[11-12]、径向基函数人工神经网络^[9]和核极限学习机^[13]等。

通过 MCMC 进行不确定性分析时,需要不断运行模型计算似然函数,MCMC 结果对似然函数高度敏感。因此,MCMC 对替代模型的精度要求很高,当前研究中尚未见基于似然函数替代模拟的 DNAPLs 运移不确定

性分析。本次研究拟采用稀疏网格(sparse grid, SG)替代模型技术,该方法适用于构建具有复杂响应关系的替代模型。SG 替代模型构建过程中,为提高构建效率、减少参数空间网格冗余,一般有两种自适应方式,局部自适应(local adaptive, LA)和维数自适应(dimension adaptive, DA)。如 Zhang 等^[14]通过一种高阶 LA-SG 技术建立了铀素反应性溶质运移替代模型,并通过 MCMC 识别得到模型参数后验分布。Zeng 等^[15]基于 DA-SG 技术建立了溶质运移替代模型,并通过 MCMC 进行了地下水污染源识别。替代模型的构建成本(即需运行一定数量的原始模型)与替代对象的复杂度密切相关,且随着模型参数维数的增加而显著增加。Zeng^[15]等利用 DA-SG 建立 5 维溶质运移替代模型需运行原模型 1 545 次。Zhang 等^[14]通过高阶 LA-SG 建立 6 维反应性溶质运移替代模型的成本是运行原模型 3 765 次。Zeng 等^[16]使用 LA-SG 及约 1 万次原模型运行建立了 10 维非均质地下水流替代模型。因此,高效率、高精度的替代模型技术对于复杂、高维 DNAPLs 运移模拟的不确定性分析具有重要的理论和实际意义。

本次研究拟基于 SG 技术,将替代模型参数空间的两种网格自适应生成方式(LA、DA)相结合,进一步提高替代模型的构建效率和替代精度。此外,通过两个解析案例(5 维 rastrign 函数、20 维 banana 函数)和一个 DNAPLs 运移室内模拟模型,验证所提出的维数、局部自适应稀疏网格替代模型(DA-LA-SG)的效率和精度。

1 研究方法

1.1 稀疏网格替代模型

1.1.1 基本原理

稀疏网格(SG)是一种基于 Smolyak^[17]规则的叠层拉格朗日插值方法,具有方法原理简单、插值精度高等优点,在水资源与水环境模拟中应用广泛。对于一维非线性函数 $f(x)$,其叠层拉格朗日插值项 U 定义为:

$$U^L(f(x)) = \sum_{i=0}^L \Delta U^i(f(x))$$

$$\Delta U^i(f(x)) = \sum_{j=0}^{m_i} c_j^i \varphi_j^i(x), i = 0, 1, \dots, L \quad (1)$$

$$m_0 = 1, m_1 = 2,$$

$$m_i = 2^{i-1} \quad \text{if } i \geq 2 \quad (2)$$

SG 插值点的生成规则有 Clenshaw-Curits 规则、Gauss-Patterson 规则、Uniform 规则等^[18], 本研究中使用 Uniform 规则, 在 $[0, 1]$ 间的坐标表达式为:

$$x_1^0 = 0.5, x_1^1 = 0, x_1^2 = 1,$$

$$x_j^i = (2j - 1) / \left(\sum_{k=0}^i m_k - 1 \right) \quad \text{if } i \geq 2 \quad (3)$$

$$c_1^0 = \Delta U^0(f(x_1^0)) = U^0(f(x_1^0)) = f(x_1^0), i = 0$$

$$c_j^i = \Delta U^i(f(x_j^i)) = U^i(f(x_j^i)) - U^{i-1}(f(x_j^i)) =$$

$$f(x_j^i) - U^{i-1}(f(x_j^i)), i \geq 1 \quad (4)$$

式中: L ——最大插值级数;

m_i ——对应级数 i 的稀疏网格插值点的数量;

$\phi(x)$ ——插值基函数, 本次研究使用三次多项式基函数^[18];

c ——插值系数, 也叫基函数的叠层盈余, 表示插值公式与目标函数在 x 处的差值。

当函数 $f(x)$ 相对平滑时, 随着插值级数 i 增加, 叠层盈余将接近于零, 所以 c 可以作为局部自适应的误差标准。

基于一维叠层插值方法, 多维稀疏网格插值为:

$$V^{L,D}[f(x)] = \sum_{\lambda(i) \leq L} \Delta V^{i,D}[f(x)], \lambda(i) = i_1 + \dots + i_D$$

$$i = (i_1, \dots, i_D) \quad (5)$$

式中: D ——模型参数 $x = (x^1, \dots, x^D)$ 的维度;

i —— $\Delta V^{i,D}[f(x)]$ 的多维插值级数向量;

$i_k (k = 1, \dots, D)$ ——第 k 维插值级数。

$\Delta V^{i,D}[f(x)]$ 定义为:

$$\Delta V^{i,D}[f(x)] = \Delta U^{i_1} \otimes \dots \otimes \Delta U^{i_D}[f(x)] =$$

$$\sum_j c_j^i \varphi_j^i(x) j = (j_1, \dots, j_D)$$

$$\varphi_j^i(x) = \prod_{n=1}^D \varphi_{j_n}^{i_n}(x_n) \quad (6)$$

式中: j ——向量 ($j_l = 1, \dots, m_{i_l}, l = 1, \dots, D$);

$\varphi_j^i(x)$ ——多维叠层基函数;

$\varphi_{j_n}^{i_n}(x_n)$ ——一维插值基函数。

当 $L = 0$ 时, $c_1^0 = f(x_1^0)$, 当 $L \geq 1$ 时, $c_j^i = V^{L,D}[f(x_j^i)] - V^{L-1,D}[f(x_j^i)] = f(x_j^i) - V^{L-1,D}[f(x_j^i)]$ 。

1.1.2 自适应稀疏网格 (Adaptive SG)

普通 SG 方法的插值级数向量的集合为: $\beta =$

$\{i \in N^D : \lambda(i) \leq L\}$ 。相对于全张量网格节点, 缓解了“维数诅咒”效应, 但对于高维参数空间, 替代模型的成本仍不可忽视。根据目标函数参数和输出之间的响应关系, 通过网格自适应的生成方式, 减少模型参数空间的冗余网格节点, 可以进一步降低 SG 替代模型的成本。

(1) 维数自适应稀疏网格 (DA-SG)

DA-SG 的基本思路^[19]可以总结为: ① 从级数向量 $(0, 0, \dots, 0)$ 开始, 在当前级数向量 i 的前邻域 $\{i + e_j : 1 \leq j \leq D\}$ (e_j 表示第 j 维为 1、其他为 0 的单位向量) 中搜索候选级数向量 A , 候选级数向量条件为 $i - e_j \in O$ for $1 \leq j \leq D, i_j > 1$, O 为已有级数向量集合; ② 在 A 中找到使测试点的误差项趋于最大的级数向量 k , 并将其加入已有级数向量集合 O 中; ③ 将 k 作为当前已有级数向量, 重复步骤①和②, 直到误差低于预定义的误差标准。这种方法结合了贪婪算法的思想, 可在构建稀疏网格的同时确定对目标函数变化有最大影响的级数向量, 提高了构建替代模型的效率。

图 1 所示为 DA-SG 插值节点的自适应生成过程。图 1(a) ~ (c) 所示为当前插值级数向量, 图 1(d) ~ (f) 表示相应级数向量对应的网格点分布图。图 1(a) 中候选级数向量 A 包含 $(4, 0)$ 、 $(1, 1)$ 和 $(0, 3)$ 。首先, 将 A 中误差项最大的级数向量 i 即 $(1, 1)$ 加入已有级数向量集合 O 中。然后, 对该级数向量 i 的所有前邻域级数向量 (即 $(2, 1)$ 、 $(1, 2)$) 进行候选条件测试, 结果发现均满足条件, 将其加入集合 A 中, 此时 A 包含 4 个级数向量。最后, 重复上述过程, 寻找误差最大项, 一直循环计算直至误差达到标准。

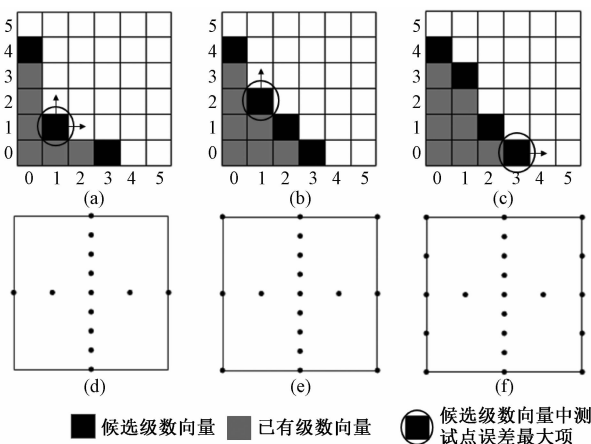


图 1 二维维数自适应稀疏网格 (DA-SG) 示意图

Fig. 1 Diagrammatic sketch of two dimensional DA-SG

(2) 局部自适应稀疏网格 (LA-SG)

LA-SG 中插值节点生成以插值系数 c_j^i 作为自适应

标准^[20-21]。至上向下逐级生成插值节点时,当某节点处的 c_j^i 小于预先设置的误差标准 ε 时,认为替代模型在该处的插值精度已满足需求,不在下级增加插值节点,反之则继续向下增加插值节点。图 2 表示了二维 LA-SG 示意图,普通 SG (level = 3) 需要插值网格节点 29 个。通过局部自适应,若插值点的 c_j^i 已达到误差标准,则下级的插值点则被删除。若插值点的 c_j^i 未达到误差标准,则其下级节点保留(如①、②点)。通过局部自适应,只需要插值节点 23 个。LA-SG 方法通过在插值对象非线性程度低的区域减少节点,提高了替代模型的构建效率。

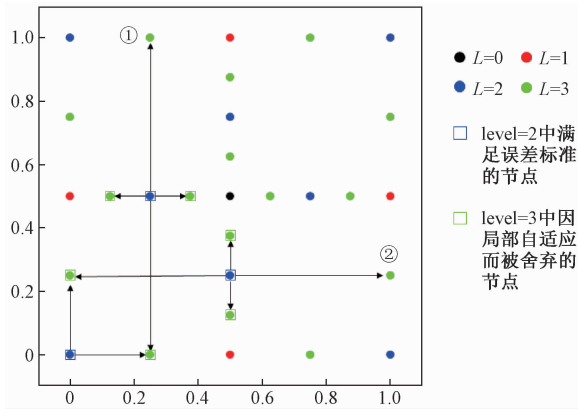


图 2 二维局部自适应稀疏网格 (LA-SG) 示意图^[18]

Fig. 2 Diagrammatic sketch of two dimensional LA-SG

(3) 维数-局部自适应稀疏网格 (DA-LA-SG)

DA-SG 方法通过识别对目标函数变化有显著影响的级数向量,提高了 SG 构建效率,但当目标函数变化区域集中在参数空间较小区域时,DA-SG 方法的效率较低。与 DA-SG 相比,LA-SG 方法通过在非线性程度低的区域减少插值节点,克服了 DA-SG 的不足,但 LA-SG 在所有级数向量空间均构造插值节点。因此,提出将 DA、LA 结合的改进稀疏网格替代模型方法 DA-LA-SG,能够进一步提到 SG 替代模型的构建效率。如图 3 所示,首先对目标函数进行维数自适应,寻找误差最大项;其次,在计算插值系数时进行局部自适应,判断局部自适应的标准,舍弃已满足误差标准节点的下级节点;最后,重复进行上述过程,直至维数自适应误差达到预设的标准,结束替代模型运行。

1.2 贝叶斯不确定性分析

贝叶斯不确定性分析方法的基础是贝叶斯公式,可表示^[22]为:

$$p(\theta | y) = \frac{p(\theta)p(y | \theta)}{\int p(\theta)p(y | \theta) d\theta} \quad (7)$$

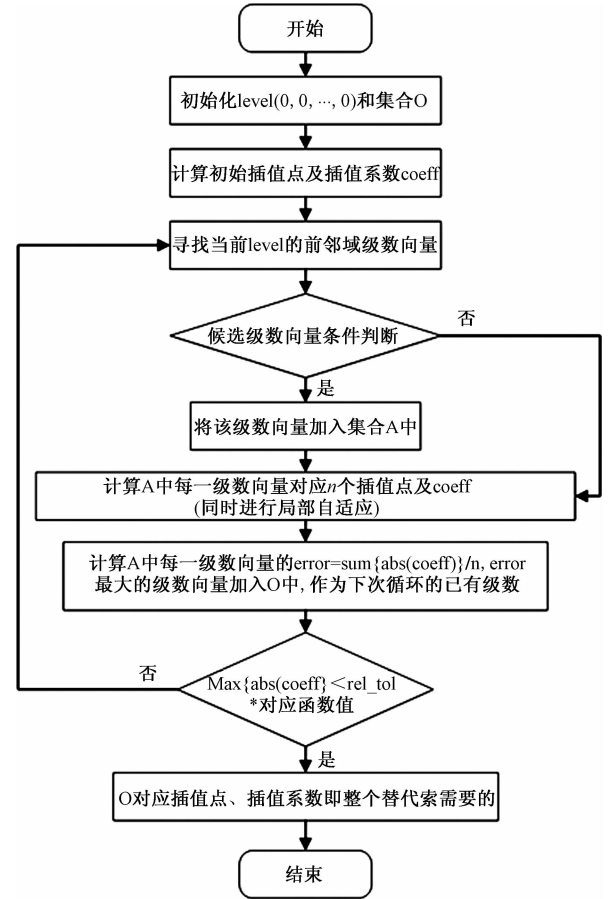


图 3 DA-LA-SG 算法流程图

Fig. 3 Algorithm flow chart of DA-LA-SG

式中: $p(\theta | y)$ —— 参数 θ 的后验概率分布;

$p(\theta)$ —— θ 的先验概率分布;

$p(y | \theta)$ —— θ 的似然函数;

$\int p(\theta)p(y | \theta) d\theta$ —— 归一化常数。

常用的似然函数为高斯函数,其自然对数表达式为:

$$\ln(p(y | \theta)) = -\frac{1}{2}(y - f(\theta))^T \Sigma^{-1}(y - f(\theta)) - \frac{1}{2} \ln |\Sigma| - \frac{n}{2} \ln(2\pi) \quad (8)$$

式中: Σ —— 观测误差的协方差矩阵;

$|\Sigma|$ —— 行列式。

由于 $p(\theta | y)$ 难以直接计算,一般采用统计抽样的方式间接获取,如 MCMC 模拟。MCMC 方法利用构建马尔科夫链来反演模型参数 θ 的后验分布,通过在 θ 的概率空间内持续搜索,逐渐收敛于一个稳定的后验分布 $\pi(\theta)$ 。本次研究中采用最新的抽样算法 DREAMzs^[23] 进行参数不确定性研究,基本步骤如下:

①基于先验信息,设置模型参数的先验分布和马尔科夫链的起点 $[\theta^i, i = 1, \dots, N]$, N 表示平行马尔科夫链的数量;②将式(8)作为似然函数,利用观测数据 y ,通过 DREAMzs 抽样得到参数后验样本;③对采样得到的结果(模型参数和输出的后验分布)进行统计分析,定量分析不确定性大小。

1.3 替代模型的精度评估

采用目标函数和替代模型间的相关系数 R^2 和相对均方根误差 $NRMSE$ ^[24] 作为替代模型精度的评价指标。

(1) 相关系数 (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (9)$$

式中: n ——样本个数;

y_i ——目标函数的输出值;

$\hat{y}_i$ ——替代模型的输出值;

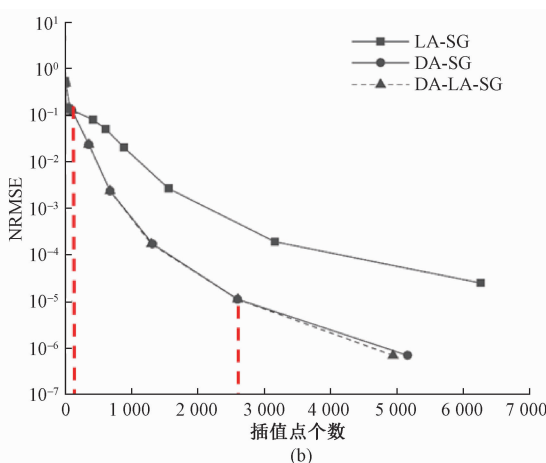
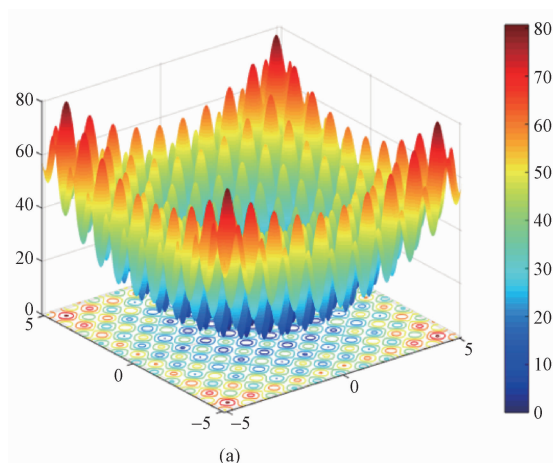


图 4 (a) Rastrigrn 函数在 2 维参数空间上的响应面; (b) 各替代模型 $NRMSE$ 随样本数量的变化

Fig. 4 (a) Response surface of the Rastrigrn function in 2-D space; (b) The $NRMSE$ of surrogates vary with the number of samples

图 4(a) 为目标函数在二维空间的响应面示意图。图 4(b) 显示 DA-SG、LA-SG 和 DA-LA-SG 等 3 个替代模型随着插值节点数量的增多,替代误差($NRMSE$)均逐渐降低。从图中可以看出,在替代模型建立的初期(插值节点少于 100),LA-SG 的替代效率(达到一定替代精度所需的构建成本)高于 DA-SG 和 DA-LA-SG,但是插值节点数量的增多,DA-SG、DA-LA-SG 的替代效率逐渐高于 LA-SG。这是由于在替代模型的构建初期,插值节点分布较少,LA-SG 能够更有效识别目标函数的局部特征,而随着插值节点数量增多,DA-SG 方法能够有效识别对目标函数变化有显著影响的级数向

$\bar{y}_i$ ——目标函数输出的平均值。

R^2 越接近于 1,替代模型效果越好。

(2) 相对均方根误差 ($NRMSE$)

$$NRMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} / |f|_{\max} \quad (10)$$

式中: $|f|_{\max}$ ——目标函数输出最大绝对值。

2 算例分析

2.1 解析案例

2.1.1 5 维 rastrigrn 函数

选择 5 维 rastrigrn 函数作为本次替代模型的目标函数,用于测试 DA-SG、LA-SG 和 DA-LA-SG 等 3 个替代模型的效率和精度。Rastrigrn 函数的解析表达式为:

$$f(x) = An + \sum_{i=1}^n [x_i^2 - A \cos(2\pi x_i)]$$

其中, $A = 10$, $-5.12 \leq x_i \leq 5.12$, $n = 5$ 。

量,进行插值空间的降维处理。当插值节点数量低于 2 600 时,DA-SG 与 DA-LA-SG 替代模型表现基本相同,这是由于替代模型前期插值系数(误差)较大,达不到局部自适应误差标准。当插值节点数量超过 2 600,DA-LA-SG 表现优于 DA-SG,DA-LA-SG 中的部分插值节点因局部自适应而被舍弃。因此,在替代成本相同的条件下,DA-LA-SG 的精度高于 DA-SG。

2.1.2 20 维 banana 函数

为了进一步测试高维情况下 DA-SG、LA-SG 和 DA-LA-SG 等 3 个替代模型的效率和精度,选择 20 维 banana 函数作为本次替代模型的目标函数,banana 函

数的解析表达式为:

$$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2] \\ - 10 \leq x_i \leq 10, 1 \leq i \leq n$$

图 5(a)为目标函数在二维参数空间的响应面示意图。图 5(b)所示为 DA-SG、LA-SG 和 DA-LA-SG 等 3 个替代模型随着插值节点数量的增多,替代误差(NRMSE)逐渐降低的过程。从图中可以看出,在替代模型的建立初期(插值节点少于 1 000),LA-SG 的替代效率比 DA-SG 和 DA-LA-SG 高,但随着插值节点数量的增多,DA-SG、DA-LA-SG 的替代效率逐渐比 LA-

SG 高。这是由于在替代模型的构建初期,插值节点较少,LA-SG 通过识别目标函数的局部特征进行插值节点的自适应,而随着插值节点的增多,DA-SG 通过识别重要的级数向量进行降维处理,其替代效率逐渐升高。插值节点数量低于 2 000 时,DA-SG 与 DA-LA-SG 表现基本相同,这是由于替代模型前期的插值误差较大,局部自适应过程很少发生。当插值节点数量超过 2 000,DA-LA-SG 表现优于 DA-SG,这是由于局部自适应 DA-LA-SG 舍弃了部分冗余插值节点,从而提高了替代效率。

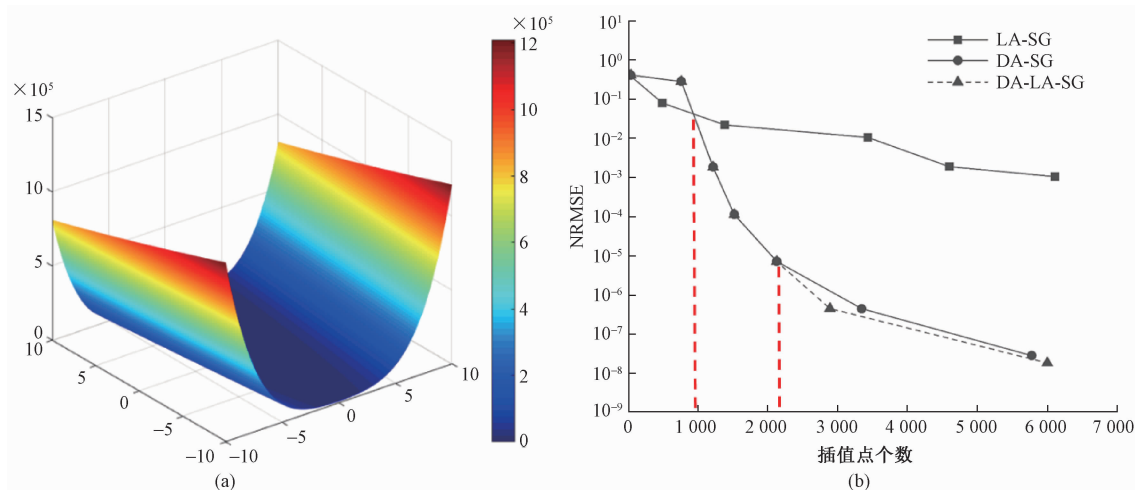


图 5 (a) Banana 函数在 2 维空间上的响应面; (b) 各替代模型 NRMSE 随样本数量的变化

Fig.5 (a) Response surface of the banana function in 2-D space; (b) The NRMSE of surrogates vary with the number of samples

2.2 二维砂箱 DNAPLs 运移室内实验

本次实验选用四氯乙烯(PCE)作为典型 DNAPLs 污染物,通过向砂箱中注入 PCE 的方式模拟污染物的泄露。图 6(a)为二维砂箱多孔介质中 PCE 的运移模拟室内实验示意图。砂箱的长、宽、厚分别为 0.6 m × 0.45 m × 0.016 m。砂箱左侧为入流边界,流速保持 $7.2 \times 10^{-3} \text{ m}^3/\text{d}$,由蠕动泵控制。砂箱背景介质为均质多孔介质,内部分布 5 个低渗透介质作为透镜体模拟介质非均质性对 PCE 运移的影响。砂箱实验中,PCE 注入点位于 $x = 0.3 \text{ m}, y = 0.4 \text{ m}, z = 0.008 \text{ m}$ 处。通过光透法^[25](LTM)实时监测砂箱中的 PCE 饱和度,且用作本次实验的观测数据。此外,本次案例研究采用 T2VOC 程序^[26]对污染源注入过程的 PCE 运移进行数值模拟,PCE 注入时间共持续 80 min。图 6(b)显示 PCE 停止注入时刻(即 $t = 80 \text{ min}$),砂箱中 PCE 的饱和度分布。

在实际情况下,由于有限的观测信息,一些水文地

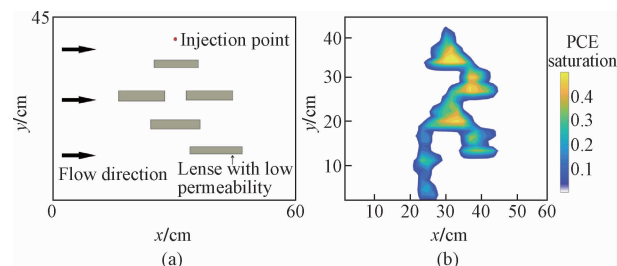


图 6 (a) 二维砂箱 PCE 运移实验示意图;

(b) 停止注入时刻 PCE 饱和度分布图

Fig.6 (a) Diagrammatic sketch of the 2-D sandbox PCE experiment; (b) Saturation distribution of PCE after stopping injection

质参数难以准确给定,例如渗透率、孔隙度等。对于 PCE 运移模拟模型,PCE 的浓度分布对含水层的渗透性十分敏感^[27]。因此,本次研究中选择两个区域(背景介质和低渗透体)的渗透率(k_1, k_2)和孔隙度(μ_1, μ_2)作为未知参数,同时将 T2VOC 程序中相对渗透率

函数中的拟合系数(n_1, n_2)作为未知参数。通过观测数据(PCE 的饱和度分布),对本次 PCE 运移模拟的 6 个未知参数进行识别。

将本次砂箱实验处理为二维 PCE 运移问题,模拟区域在 x, y, z 方向上离散为 $35 \times 38 \times 1$ 个网格单元。PCE 运移数值模型中,采用 Parker 本构模型。此外,对 PCE 运移模型进行参数不确定性分析时,未知模型参数的先验分布均假定为均匀分布,范围见表 1。

表 1 PCE 运移模型中未知参数的取值范围
Table 1 Ranges of the unknown parameters in the PCE migration model

参数	取值范围
背景介质渗透率 k_1/m^2	$[0.8 \times 10^{-10}, 5.0 \times 10^{-10}]$
背景介质孔隙度 μ_1	$[0.15, 0.25]$
拟合系数 n_1	$[2.0, 4.0]$
弱透镜体渗透率 k_2/m^2	$[0.05 \times 10^{-11}, 5.00 \times 10^{-11}]$
弱透镜体孔隙度 μ_2	$[0.10, 0.40]$
拟合系数 n_2	$[2.0, 4.0]$

2.2.1 建立似然函数的替代模型

贝叶斯不确定性分析中,需重复运行 PCE 运移模拟模型,计算似然函数 $p(y|\theta)$ 。为克服重复调用 T2VOC 程序所导致的计算负荷,采用稀疏网格方法耦合 T2VOC 程序,建立对数似然函数 $\lg p(y|\theta)$ 的替代模型。替代模型与 T2VOC 模拟器的耦合过程可以总结为如下步骤:(1) 根据参数先验区间和 SG 格点规则生成 PCE 运移模型的参数样本点 θ ;(2) 将样本点输入 T2VOC 程序,运行得到 PCE 的饱和度分布模拟值;(3) 通过饱和度观测值和模拟值计算对数似然函数 $\lg p(y|\theta)$,并将其作为 SG 的目标函数 $f(x)$;(4) 计算 SG 中各样本点 θ 的插值系数 c_j^i ;(5) 根据式(6)建立 $\lg p(y|\theta)$ 的替代模型。

本次案例分析中,分别采用 LA-SG、DA-SG 和 DA-LA-SG 等 3 种方法建立 $\lg p(y|\theta)$ 的替代模型,并对这三种替代方法的效率和精度进行对比分析。其中,替代效率表示建立替代模型的成本,替代精度为替代模型与原始 T2VOC 结果的接近程度。

图 7 为砂箱 PCE 运移模拟对数似然函数的替代模型(DA-SG、LA-SG 和 DA-LA-SG)随着插值节点数量的增多,替代精度逐渐升高的过程。从图中可以看出,在替代模型的构建初期(插值节点数量小于等于 100),3 个替代模型具有相接近的替代效率,替代模型的 NRMSE 随插值点数量的衰减速度相似。随着样本数量的增多,替代模型的替代精度均逐渐升高,但 NRMSE 的衰减速度与替代模型初期相比,均有所降

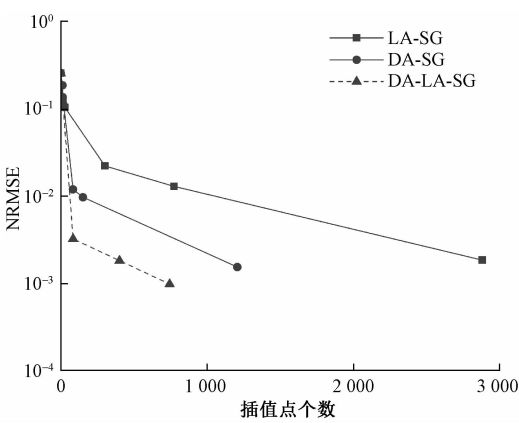


图 7 PCE 运移模拟对数似然函数替代模型的 NRMSE
Fig.7 NRMSE of surrogates for the logarithmic likelihood function

低。DA-LA-SG 具有最高的替代效率,即在相同的替代成本条件下,DA-LA-SG 具有最高的替代精度,其次为 DA-SG 和 LA-SG。在替代模型的构建初期,各参数维度上的插值节点数量较少,进行维数自适应时,所舍弃的插值节点并不明显。随着插值水平(Level)的增加,各参数维度所需的插值节点数量迅速增加,进行维数自适应能够显著减少插值节点的数量,提高替代模型的效率。

图 8 显示了本次 PCE 运移模型的对数似然函数 $\lg p(y|\theta)$ 的响应面及其替代模型的效果。图 8(a)~(c)分别表示 $\lg p(y|\theta)$ 在不同 2 维参数空间上的响应面,(d)表示采用 DA-LA-SG 方法建立的 $\lg p(y|\theta)$ 替代模型的效果。结果显示,建立该替代模型共需要运行原始 PCE 运移模型 745 次,原始模型与替代模型间的相关系数 R^2 为 0.999 7,替代模型的 RMSE 为 3.63,表明通过 DA-LA-SG 方法建立的 $\lg p(y|\theta)$ 替代模型具有较高的替代效率和精度。本次室内砂箱实验 PCE 运移模拟参数不确定性分析中,采用 DA-LA-SG 建立的似然函数替代模型,而无需调用原始的 PCE 运移模型。

2.2.2 基于 DA-LA-SG 替代模型的 PCE 运移模拟参数不确定性分析

采用基于 DREAMzs 算法的 MCMC 模拟识别 6 个未知模型参数的后验分布,参数先验分布见表 1。为了保证搜索效率,使用 3 条平行马尔科夫链,每条链长度为 30 000,预热期长度为 10 000,剩下的样本用于生成参数的后验分布。在 MCMC 运行过程中,调用基于 DA-LA-SG 建立的似然函数替代模型不需运行原始模型。图 9 为识别得到的 6 个模型参数的后验概率分

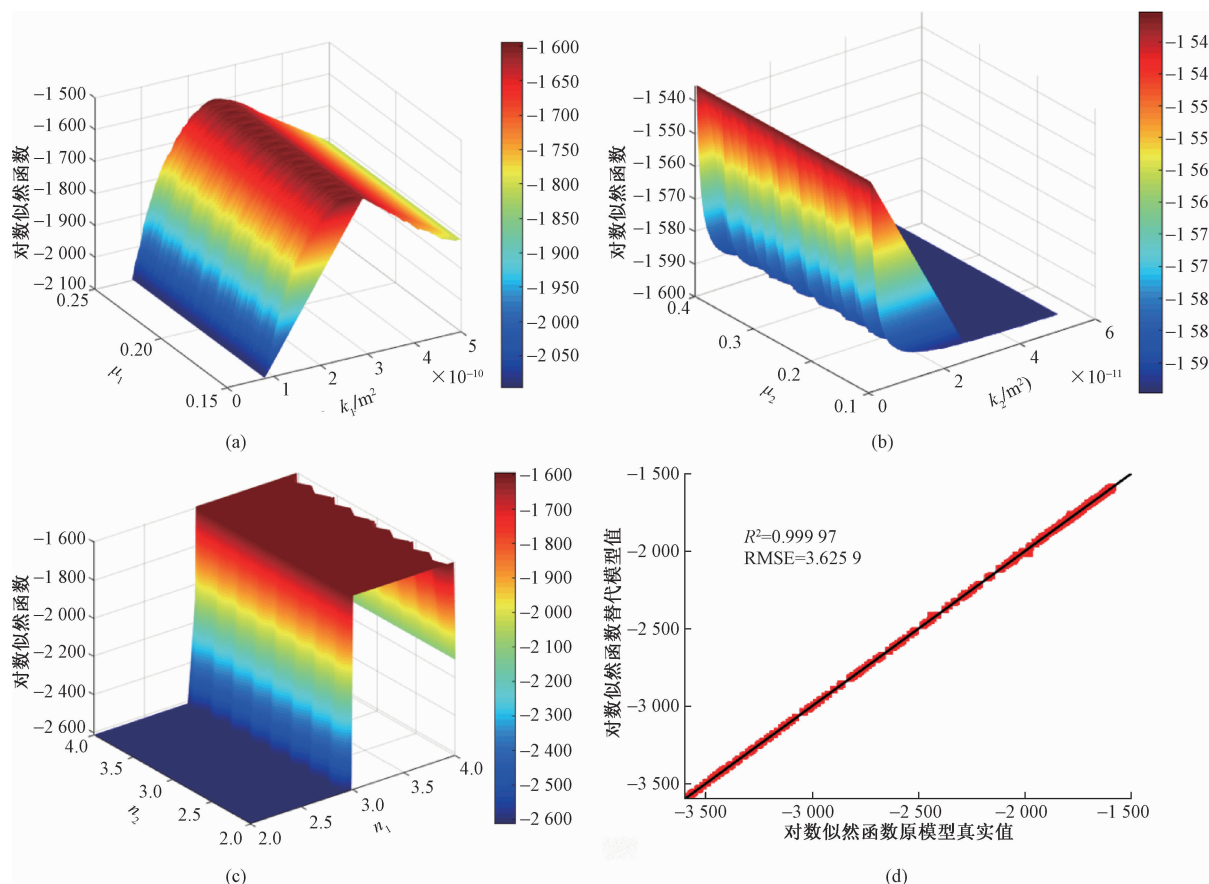


图 8 (a) ~ (c) 对数似然函数在二维参数空间上的响应面; (d) DA-LA-SG 替代模型效果图

Fig. 8 (a) ~ (c) Response surfaces of the 2-D logarithmic likelihood function; (d) Performance of DA-LA-SG

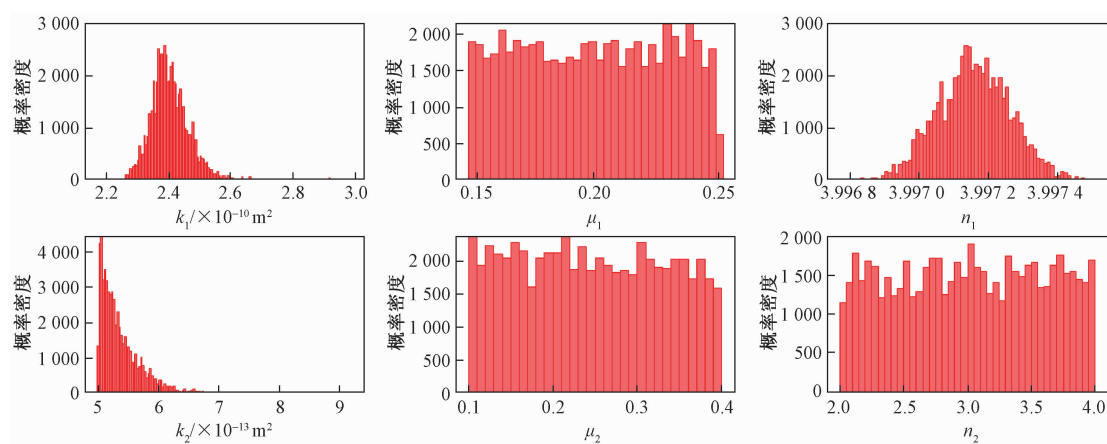


图 9 PCE 运移模型参数的后验概率分布

Fig. 9 Posterior probability density of the PCE migration parameters

布,坐标横轴区间表示相应参数的先验分布范围。经过 MCMC 参数识别, k_1 、 k_2 和 n_1 的后验概率密度呈正态分布,相对于先验分布,参数的后验分布范围明显缩小,说明这 3 个参数对观测数据(PCE 的饱和度)比较敏感。相比之下,模型参数 μ_1 、 μ_2 和 n_2 后验概率分布大致呈均匀分布,且与先验分布范围相同,因此这 3 个

参数对观测数据不够敏感,参数识别效果较差。本次 DNAPL 运移模拟不确定性分析结果表明,砂箱实验中 DNAPL 的运移主要受介质的渗透性影响^[28-29],DNAPL 在运移过程中受到弱透镜体的阻碍,由于无法克服毛细压力而在其上方聚积并侧向运移,孔隙度对 DNAPL 运移的影响较小。因此,对于实际 DNAPL 污

染场地的评估与修复,场地含水层的渗透性质特别是弱透水介质分布的精确刻画,是建立准确、可靠 DNAPLs 运移模拟模型的重要条件。

3 结论

为克服 DNAPLs 运移模拟不确定性分析中的计算耗时问题,本次研究在传统稀疏网格(SG)替代模型的基础上,提出了一种将局部自适应(LA)和维数自适应(DA)两种网格生成方式耦合的改进替代模型 DA-LA-SG。通过两个解析案例和一个 DNAPL 运移室内实验,对比分析了 DA-SG、LA-SG 和 DA-LA-SG 三种替代模型技术的表现,并将 DA-LA-SG 替代模型用于 DNAPL 运移数值模拟不确定性分析中,得到以下结论:

(1)在替代模型的构建初期,LA-SG 的替代效率(一定替代精度条件下的替代成本)优于或接近于 DA-SG 和 DA-LA-SG,随着插值节点数量的增多,DA-SG 和 DA-LA-SG 的替代效率逐渐优于 LA-SG,且 DA-LA-SG 具有最高的替代效率。

(2)使用改进的维数、局部自适应稀疏网格技术(DA-LA-SG)能够高效、高精度的建立 PCE 运移模型的似然函数替代模型。构建替代模型需运行原始模型 745 次,原始模型与替代模型间的相关系数 R^2 达到 0.999 7。

(3)将建立的似然函数替代模型用于 PCE 运移模拟参数不确定性分析,克服了因重复调用 PCE 运移模型导致的计算耗时问题,通过 MCMC 模拟获得了模型参数的后验分布。结果发现,背景介质渗透率 k_1 、拟合系数 n_1 和透镜体渗透率 k_2 的可识别性较强,背景介质孔隙度 μ_1 、透镜体孔隙度 μ_2 和拟合系数 n_2 的识别效果较差,这些参数对饱和度观测数据不敏感。

参考文献(References):

- [1] SABA T, ILLANGASEKARE T H. Effect of groundwater flow dimensionality on mass transfer from entrapped nonaqueous phase liquid contaminants[J]. Water Resources Research, 2000, 36(4):971-979.
- [2] HE L, HUANG G H, LU H W, et al. Optimization of Surfactant-Enhanced Aquifer Remediation for a Laboratory BTEX System under Parameter Uncertainty[J]. Environmental Science & Technology, 2008, 42(6):2009-2014.
- [3] YEH C K, HSU C Y, CHIU C H, et al. Reaction efficiencies and rate constants for the goethite-catalyzed Fenton-like reaction of NAPL-form aromatic hydrocarbons and chloroethylenes[J]. Journal of Hazardous Materials, 2008, 151(2):562-569.
- [4] SAENTON S, ILLANGASEKARE T H. Upscaling of mass transfer rate coefficient for the numerical simulation of dense nonaqueous phase liquid dissolution in heterogeneous aquifers[J]. Water Resources Research, 2007, 43(2):2428.
- [5] WEST M R, KUEPER B H. Numerical simulation of DNAPL source zone remediation with in situ chemical oxidation (ISCO)[J]. Advances in Water Resources, 2012, 44(5):126-139.
- [6] WU M, WU J, WU J. Simulation of DNAPL migration in heterogeneous translucent porous media based on estimation of representative elementary volume[J]. Journal of Hydrology, 2017, 553:276-288.
- [7] LIU X, CARDIFF M A, KITANIDIS P K. Parameter estimation in nonlinear environmental problems[J]. Stochastic Environmental Research and Risk Assessment (SERRA), 2010, 24(7):1003-1022.
- [8] LIU X, LEE J, KITANIDIS P K, et al. Value of information as a context-specific measure of uncertainty in groundwater remediation[J]. Water Resources Management, 2012, 26(6):1513-1535.
- [9] HOU Z, LU W, XUE H, et al. A comparative research of different ensemble surrogate models based on set pair analysis for the DNAPL-contaminated aquifer remediation strategy optimization[J]. Journal of Contaminant Hydrology, 2017, 203:28-37.
- [10] Q OU Y, LU W X, LIN J, et al. Conservative strategy-based ensemble surrogate model for optimal groundwater remediation design at DNAPLs-contaminated sites[J]. Journal of Contaminant Hydrology, 2017, 203:1-8.
- [11] XUE J, WENXI L, JIN N, et al. A stochastic optimization model based on adaptive feedback correction process and surrogate model uncertainty for DNAPL-contaminated groundwater remediation design[J]. Stochastic Environmental Research and Risk Assessment, 2018, 32:3195-3206.
- [12] 王涵, 卢文喜, 李久辉, 等. 地下水 DNAPLs 污染多相流的随机模拟及其不确定性分析[J]. 中国环境科学, 2018, 38(7):2572-2579. [WANG H, LU W XI, LI J H, et al. Stochastic simulation and uncertainty analysis of multi-phase flow of groundwater polluted by DNAPLs. China Environmental Science, 2018, 38(7):2572-2579. (in

- Chinese)]
- [13] 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 Z Y, LU W X, WANG Y. Surrogate-based source identification of DNAPLs-contaminated groundwater [J]. China Environmental Science, 2019, 39(1): 188 – 195. (in Chinese)]
- [14] ZHANG G, DAN L, MING Y, et al. An adaptive sparse-grid high-order stochastic collocation method for Bayesian inference in groundwater reactive transport modeling[J]. Water Resources Research, 2013, 49(10):6871 – 6892.
- [15] ZENG L Z, SHI L S, ZHANG D X, et al. A sparse grid based Bayesian method for contaminant source identification [J]. Advances in Water Resources, 2012, 37:1 – 9.
- [16] ZENG X K, YE M, WU J, et al. Improved nested sampling and surrogate - enabled comparison with other marginal likelihood estimators [J]. Water Resources Research, 2018, 54:797 – 826.
- [17] SMOLYAK S. Quadrature and interpolation formulas for tensor products of certain classes of functions[J]. Doklady Akademii nauk SSSR, 1963, 4(5):240 – 243.
- [18] 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.
- [19] GERSTNER T, GRIEBEL M. Dimension-adaptive tensor-product quadrature[J]. Computing, 2003, 71(1):65 – 87.
- [20] 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.
- [21] MA X, ZABARAS N. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations[J]. Journal of Computational Physics, 2010, 229(10): 3884 – 3915.
- [22] 陆乐, 吴吉春, 陈景雅. 基于贝叶斯方法的水文地质参数识别 [J]. 水文地质工程地质, 2008, 35(5): 58 – 63. [LU L, WU J C, CHEN J Y. Identification of hydrogeological parameters based on the Bayesian method [J]. Hydrogeology & Engineering Geology, 2008, 35(5): 58 – 63. (in Chinese)]
- [23] 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.
- [24] MO S, DAN L, SHI X, et al. A Taylor expansion - based adaptive design strategy for global surrogate modeling with applications in groundwater modeling [J]. Water Resources Research, 2017, 53(5): 10802 – 10823.
- [25] 乔文静, 叶淑君, 吴吉春. 非均质孔隙介质中两相流的光透法应用研究 [J]. 水文地质工程地质, 2015, 42(2): 112 – 119. [QIAO W J, YE S J, WU J C. A study of the two-phase flow in heterogeneous porous media with the light transmission method[J]. Hydrogeology & Engineering Geology, 2015, 42(2): 112 – 119. (in Chinese)]
- [26] FALTA R W, PRUESS K, FINSTERLE S, et al. T2VOC user' s guide [R]. Berkeley: Lawrence Berkeley Lab, Report No. LBL-36400, 1995.
- [27] 施小清, 吴吉春, 刘德朋. 饱和介质中重非水相液体运移的数值模拟及敏感性分析[J]. 南京大学学报(自然科学版), 2011, 47(3):299 – 307. [SHI X Q, WU J C, LIU D P, et al. Numerical simulation of transportation of dense nonaqueous phase liquids in the subsurface environment[J]. Journal of Nanjing University (Natural Sciences), 2011, 47(3):299 – 307. (in Chinese)]
- [28] 郑菲, 高燕维, 徐红霞, et al. 非均质性对 DNAPL 污染源区结构特征影响的实验研究 [J]. 水文地质工程地质, 2016, 43(5):140 – 148. [ZHENG F, GAO Y W, XU H X, et al. An experimental study of the influence of heterogeneity on the DNAPL source-zone architecture [J]. Hydrogeology & Engineering Geology, 2016, 43(5):140 – 148. (in Chinese)]
- [29] 邓亚平, 郑菲, 施小清, 等. 多孔介质中 DNAPLs 运移行为研究进展 [J]. 南京大学学报(自然科学), 2016, 52(3):409 – 420. [DENG Y P, ZHENG F, SHI X Q, et al. Review on the transport of dense non-aqueous phase liquids in porous media [J]. Journal of Nanjing University (Natural Sciences), 2016, 52(3):409 – 420. (in Chinese)]

编辑:张若琳