

一种新的估计非高斯分布含水层渗透系数场的方法

孙 猛, 骆乾坤, 孔志伟, 郭 明, 刘明力, 钱家忠

A novel approach for estimating hydraulic conductivity of non-Gaussian aquifer

SUN Meng, LUO Qiankun, KONG Zhiwei, GUO Ming, LIU Mingli, and QIAN Jiazhong

在线阅读 View online: <https://doi.org/10.16030/j.cnki.issn.1000-3665.202308022>

您可能感兴趣的其他文章

Articles you may be interested in

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

考虑预报偏差的迭代式集合卡尔曼滤波在地下水水流数据同化中的应用

Application of the bias aware Ensemble Kalman Filter with Confirming Option (Bias-CEnKF) in groundwater flow data assimilation

杨运, 吴吉春, 骆乾坤, 钱家忠 水文地质工程地质. 2022, 49(6): 13-23

基于相关指标的裂隙岩体渗透系数估算模型研究

A model for estimating hydraulic conductivity of fractured rock mass based on correlation indexes

王玮, 钱家忠, 马雷, 王德健, 马海春, 赵卫东 水文地质工程地质. 2021, 48(3): 82-89

基于遗传算法-支持向量机的滑坡渗透系数反演

Estimation of hydraulic conductivity of landslides based on support vector machine method optimized with genetic algorithm

胡鹏, 文章, 胡新丽, 张玉明 水文地质工程地质. 2021, 48(4): 160-168

基于广义径向流模型的非均质孔隙含水层井流试验分析

Pumping tests analyses of a heterogeneous pore aquifer based on the Generalized Radial Flow model

邓祺文, 陈刚, 郑可, 施雯, 胡成 水文地质工程地质. 2022, 49(2): 17-23

基于多源数据的弱透水层水文地质参数反演研究

Research on hydrogeological parameter inversion of an aquitard based on multi-source data: A case study of a silt layer in the Hohhot Basin

石鸿蕾, 郝奇琛, 邵景力, 崔亚莉, 张秋兰 水文地质工程地质. 2021, 48(2): 1-7



关注微信公众号, 获得更多资讯信息

DOI: 10.16030/j.cnki.issn.1000-3665.202308022

孙猛, 骆乾坤, 孔志伟, 等. 一种新的估计非高斯分布含水层渗透系数场的方法 [J]. 水文地质工程地质, 2024, 51(3): 23-33.
SUN Meng, LUO Qiankun, KONG Zhiwei, et al. A novel approach for estimating hydraulic conductivity of non-Gaussian aquifer[J]. Hydrogeology & Engineering Geology, 2024, 51(3): 23-33.

一种新的估计非高斯分布含水层渗透系数场的方法

孙 猛, 骆乾坤, 孔志伟, 郭 明, 刘明力, 钱家忠
(合肥工业大学资源与环境工程学院, 安徽 合肥 230009)

摘要: 集合卡尔曼滤波(ensemble Kalman filter, EnKF)是最流行的数据同化方法之一。然而,在处理非高斯问题时,EnKF 存在局限性。为了解决非高斯问题并准确描述含水介质连通性,将正态分数变换(normal-score transformation, NST)与多重数据同化集合平滑器(ensemble smoother with multiple data assimilation, ES-MDA)相结合,提出 NS-ES-MDA 方法。通过对比实验,验证了 NS-ES-MDA 方法估计非高斯分布含水层渗透系数场的有效性。相较于重启正态分数集合卡尔曼滤波器(restart normal-score ensemble Kalman filter, rNS-EnKF)方法,NS-ES-MDA 在吸收相同数据后,参数估计精度提升约 34%,计算效率提升约 35%。此外,NS-ES-MDA 方法受“异参同效”现象的影响较小,具有较强的更新能力,能够保障得到较准确的参数估计值。研究可为非高斯分布含水层参数估计提供一种有效的求解方法。

关键词: 数据同化;非高斯场;参数估计;集合平滑器;正态分数变换;渗透系数

中图分类号: P641.2

文献标志码: A

文章编号: 1000-3665(2024)03-0023-11

A novel approach for estimating hydraulic conductivity of non-Gaussian aquifer

SUN Meng, LUO Qiankun, KONG Zhiwei, GUO Ming, LIU Mingli, QIAN Jiazhong
(School of Resources and Environmental Engineering, Hefei University of Technology,
Hefei, Anhui 230009, China)

Abstract: The ensemble Kalman filter (EnKF) is one of the most widely used data assimilation methods. However, it exhibits limitations in handling non-Gaussian problems. To effectively address such issues and accurately describe the connectivity of aquifers, a novel approach named NS-ES-MDA is developed in this study. The proposed NS-ES-MDA synergistically combines the normal-score transformation (NST) with ensemble smoother with multiple data assimilation (ES-MDA). Through comparative experiments, the efficacy of NS-ES-MDA in estimating the hydraulic conductivity of non-Gaussian distributed aquifers is demonstrated. By assimilating the same dataset, NS-ES-MDA exhibits approximately 34% improvement in parameter estimation accuracy and about 35% enhancement in computational efficiency compared to the restart normal-score ensemble Kalman filter (rNS-EnKF). Furthermore, the NS-ES-MDA shows case robustness against the “equifinality” and displays remarkable updating capabilities, which leads to more precise parameter estimates. This study provides an effective solution for parameter estimation in non-Gaussian distributed aquifers.

Keywords: data assimilation; non-Gaussian fields; parameter estimation; ensemble smoother with multiple

收稿日期: 2023-08-04; 修订日期: 2023-10-29

投稿网址: www.swdgcgz.com

基金项目: 国家重点研发计划项目(2022YFC3702200); 安徽省自然科学基金项目(JZ2022AKZR0451)

第一作者: 孙猛(1997—), 男, 硕士研究生, 主要从事地下水数值模拟研究。E-mail: sunmeng_alvin@163.com

通讯作者: 骆乾坤(1984—), 女, 博士, 副研究员, 硕士生导师, 主要从事地下水模拟和水资源优化管理研究。E-mail: QKLuo@hfut.edu.cn

data assimilation; normal-score transformation; hydraulic conductivity

准确识别含水层的关键参数,如渗透系数、贮水率和孔隙度等,对于高精度地下水数值模拟、地下水资源管理以及地下水污染修复等工作具有重大意义。然而,由于含水系统的复杂性和观测数据的有限性,参数估计面临许多挑战^[1-2]。传统的参数估计方法因受限于模型误差、参数不确定性和观测误差等因素,估计结果的精度和可靠性较差^[3]。为了克服这些问题,数据同化(data assimilation, DA)方法应运而生。数据同化方法可以将模型模拟值和观测数据进行有效融合,以提高参数估计的准确性和可靠性^[4-5]。集合卡尔曼滤波(ensemble Kalman filter, EnKF)和它的改进方法是最流行的数据同化方法之一,在海洋学、大气科学、水文学、石油工程等诸多领域得到了广泛应用^[6-9]。EnKF 及其改进方法实现参数估计包括 2 个主要步骤:对系统演变进行预测;根据预测值与观测值的差异对所描述系统的参数进行更新(或修正),更新以线性组合的形式进行,权重使用协方差函数计算^[10]。在每个时间步,EnKF 都会按照这个顺序对系统参数进行预测和更新,以逐步融合模型预测和观测数据。然而,这种以线性组合形式更新系统参数的方法,会导致 EnKF 面临许多非高斯问题的挑战。尤其对于含水层分布高度渠道化,渗透系数明显不符合高斯分布的情况,EnKF 难以还原在地下水流和溶质运移中起关键作用的含水介质的连通性^[11]。因此,寻找可以解决非高斯问题同时可以准确刻画含水介质连通性的方法至关重要。

目前,解决非高斯问题的方法大致可以分为 4 类:其他滤波方法、深度学习类方法、与地质统计方法相结合的方法和转换类方法^[12-18]。Moradkhani 等^[12]采用粒子滤波解决水文模型中的参数估计问题,通过对一组粒子赋不同权重以逼近参数的后验分布,其最显著的优点是不再局限于高斯假设,但是该方法通常比其他滤波方法需要更多的样本,并且样本量会随着状态变量个数的增加呈指数增长,面对高维问题时计算量让人难以接受。Zhang 等^[13]借助深度学习方法挖掘创新向量到更新向量之间的非线性映射关系代替基于协方差矩阵的线性更新步骤。但是深度学习模型的选择较为主观,同时在深度学习模型中存在大量超参数,计算量巨大且难以保证计算精度。Cao 等^[14]提出在场地中选取向导点,用基于卡尔曼的方法更新向导点处的渗透系数并使用多点地质统计(multiple

point statistics, MPS)插值非高斯场,但是插值结果依赖于训练图像的质量。Zhou 等^[15]提出正态分数集合卡尔曼滤波(normal-score ensemble Kalman filter, NS-EnKF)方法,将正态分数变换(normal-score transformation, NST)引入 EnKF。NS-EnKF 在更新之前将非高斯分布的变量转换为高斯分布,更新后再反向转换为原始分布^[15-17]。Xu 等^[18]基于 NS-EnKF 发展了重启正态分数集合卡尔曼滤波器(restart normal-score ensemble Kalman filter, rNS-EnKF),每次更新后将后验参数代入模型从初始时刻重新计算得到当前时刻的状态以防止 EnKF 可能存在的非物理更新,并成功用于非高斯含水层中渗透系数和污染源参数联合识别。

通常,当目的只是估计参数时,一次性使用所有观测数据进行全局更新比顺序更新更高效,集合平滑器(ensemble smoother, ES)就可以用远小于 EnKF 的计算成本得到与其相似的结果^[19-20]。当系统强非线性时,应用 ES 方法的迭代形式(ensemble smoother with multiple data assimilation, ES-MDA)才能较准确求解参数估计问题^[20]。因此,基于 ES-MDA 估计参数的高效性和 NST 处理非高斯变量的强大能力,本文将 NST 与 ES-MDA 相结合,提出正态分数多重数据同化集合平滑器(normal-score ensemble smoother with multiple data assimilation, NS-ES-MDA)方法求解非高斯分布含水层中渗透系数场估计问题。同时,通过二维算例验证新方法的有效性。

1 研究方法

1.1 标准 ES-MDA

标准 ES-MDA 分为初始化阶段和迭代计算阶段。其中,迭代计算阶段包含预测和更新 2 个主要步骤。实现过程如下:

(1) 初始化

在执行数据同化之前,需要设置一些超参数,通常根据经验选择或多次尝试确定。初始集合的大小(N_c)是一个关键参数,通常在数量级为 10^2 范围内选择,使用地质统计方法生成 N_c 个参数场的实现,每个实现中包含 N_p 个未知参数,即初始集合 $\mathbf{X}_0 \in \mathbf{R}^{N_p \times N_c}$ 。除初始集合之外,还有 2 个重要参数:迭代次数 N 和每次迭代需要的膨胀系数($\alpha_i, i=1, 2, \dots, N$)。本文遵循 Evensen^[21]提出的方案,膨胀系数满足:

$$\sum_{i=1}^N \frac{1}{\alpha_i} = 1 \quad (1)$$

$$\alpha_i = \alpha'_i \left(\sum_{i=1}^N \frac{1}{\alpha'_i} \right) \quad (2)$$

$$\alpha'_{i+1} = \frac{\alpha'_i}{\alpha_{\text{geo}}} \quad (3)$$

式中: α_i ——第 i 次迭代需要的膨胀系数;

α'_i ——膨胀前的系数, 其中 α'_1 不能为 0;

α_{geo} ——常数, 代表 α_i 从一次迭代到下一次迭代的变化程度。

(2) 预测

对参数集合中的第 j 个实现 $\mathbf{X}_{ji} \in \mathbb{R}^{N_p}$, 进行预测模拟, 得到初始时刻到当前时刻的所有观测值。表达式为:

$$\mathbf{Y}_{ji} = y_1, y_2, \dots, y_k = f(\mathbf{X}_{ji}) \quad (4)$$

式中: $\mathbf{X}_{ji}$ ——第 i 次迭代参数集合第 j 个实现;

y_k —— k 时刻的观测值;

$\mathbf{Y}_{ji}$ ——第 i 次迭代的预测集合中第 j 个实现, 包括初始时刻到 k 时刻所有的观测值;

$f(\cdot)$ ——预报算子。

(3) 更新参数

根据观测值与预测值之间的差距对先验参数 ($\mathbf{X}_{ji}$) 进行修正:

$$\mathbf{X}_{ji+1} = \mathbf{X}_{ji} + \mathbf{C}_{\mathbf{X}, \mathbf{Y}}^i (\mathbf{C}_{\mathbf{Y}, \mathbf{Y}}^i + \alpha_i \mathbf{R})^{-1} (\mathbf{D} + \sqrt{\alpha_i} \boldsymbol{\varepsilon}_j - \mathbf{Y}_{ji}) \quad (5)$$

式中: $\mathbf{C}_{\mathbf{X}, \mathbf{Y}}^i$ ——第 i 次迭代中先验参数与预测值之间的互协方差矩阵;

$\mathbf{C}_{\mathbf{Y}, \mathbf{Y}}^i$ ——预测值的自协方差矩阵;

$\mathbf{R}$ ——观测误差的协方差矩阵;

$\mathbf{D}$ ——初始时刻到当前时刻 k 所有观测值;

$\boldsymbol{\varepsilon}_j$ ——均值为 0、协方差矩阵为 $\mathbf{R}$ 的观测误差。

在第 i 次迭代中, $\mathbf{C}_{\mathbf{X}, \mathbf{Y}}^i$ 和 $\mathbf{C}_{\mathbf{Y}, \mathbf{Y}}^i$ 按如下方法计算:

$$\mathbf{C}_{\mathbf{X}, \mathbf{Y}}^i = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{X}_{ji} - \bar{\mathbf{X}}_i)(\mathbf{Y}_{ji} - \bar{\mathbf{Y}}_i)^T \quad (6)$$

$$\mathbf{C}_{\mathbf{Y}, \mathbf{Y}}^i = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{Y}_{ji} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ji} - \bar{\mathbf{Y}}_i)^T \quad (7)$$

式中: $\bar{\mathbf{X}}_i$ ——第 i 次迭代参数集合的均值;

$\bar{\mathbf{Y}}_i$ ——第 i 次迭代预测集合的均值。

(4) 迭代计算

返回步骤(2), 直到达到指定的迭代次数。

1.2 正态分数变换

为了增强 ES-MDA 公式处理高斯参数识别问题的能力, 更新前对先验参数集合执行 NST 操作, 在高斯空间中更新后再反向转换为其原始分布。相比于标准 ES-MDA, 主要变化如下:

(1) 在预测步骤之后更新之前对先验参数集合执行 NST 操作:

$$\mathbf{G}_i = \phi(\mathbf{X}_i) \quad (8)$$

式中: $\mathbf{X}_i$ ——第 i 次迭代的先验参数集合;

$\mathbf{G}_i$ ——正态分数变换后的先验参数集合;

$\phi(\cdot)$ ——正态分数变换算子。

用 $\mathbf{G}_{ji}$ 代替 $\mathbf{X}_{ji}$ 进入更新步骤, 即式(5)修改为:

$$\mathbf{G}_{ji+1} = \mathbf{G}_{ji} + \mathbf{C}_{\mathbf{G}, \mathbf{Y}}^i (\mathbf{C}_{\mathbf{Y}, \mathbf{Y}}^i + \alpha_i \mathbf{R})^{-1} (\mathbf{D} + \sqrt{\alpha_i} \boldsymbol{\varepsilon}_j - \mathbf{Y}_{ji}) \quad (9)$$

式中: $\mathbf{C}_{\mathbf{G}, \mathbf{Y}}$ ——正态分数变换后的先验参数集合与预测集合的互协方差矩阵。

(2) 在更新步骤之后对后验参数集合进行反向变换:

$$\mathbf{X}_{i+1} = \phi^{-1}(\mathbf{G}_{i+1}) \quad (10)$$

1.3 协方差局域化

基于集合方法的一个问题是欠采样问题。当集合规模很小, 不具有统计代表性时, 可能会导致远距离伪相关和滤波器发散。协方差局域化(covariance localization, CL)可以解决这个问题^[22-24]。协方差局域化通常通过对协方差矩阵执行逐元素乘法从而修正协方差矩阵:

$$\tilde{\mathbf{C}}_{\mathbf{G}, \mathbf{Y}}^i = \boldsymbol{\rho}_{\mathbf{G}, \mathbf{Y}} \circ \mathbf{C}_{\mathbf{G}, \mathbf{Y}}^i \quad (11)$$

$$\tilde{\mathbf{C}}_{\mathbf{Y}, \mathbf{Y}}^i = \boldsymbol{\rho}_{\mathbf{Y}, \mathbf{Y}} \circ \mathbf{C}_{\mathbf{Y}, \mathbf{Y}}^i \quad (12)$$

式中: $\boldsymbol{\rho}_{\mathbf{G}, \mathbf{Y}}$ ——参数点与观测点的距离相关矩阵;

$\boldsymbol{\rho}_{\mathbf{Y}, \mathbf{Y}}$ ——观测点与观测点的距离相关矩阵。

本文采用距离相关矩阵进行协方差局域化, 其中距离相关矩阵根据 Gaspari 等^[25]提出的五阶相关函数进行计算。该函数可以有效减小空间点之间随着距离增加而产生的相关性, 并且可以限制超过特定距离的远程相关性。具体计算过程如下:

$$\rho = \begin{cases} -\frac{1}{4}\left(\frac{\delta}{b}\right)^5 + \frac{1}{2}\left(\frac{\delta}{b}\right)^4 + \frac{5}{8}\left(\frac{\delta}{b}\right)^3 - \frac{5}{3}\left(\frac{\delta}{b}\right)^2 + 1 & 0 \leq \delta \leq b \\ \frac{1}{12}\left(\frac{\delta}{b}\right)^5 - \frac{1}{2}\left(\frac{\delta}{b}\right)^4 + \frac{5}{8}\left(\frac{\delta}{b}\right)^3 + \frac{5}{3}\left(\frac{\delta}{b}\right)^2 - 5\left(\frac{\delta}{b}\right) + 4 - \frac{2}{3}\left(\frac{\delta}{b}\right)^{-1} & b < \delta \leq 2b \\ 0 & \delta > 2b \end{cases} \quad (13)$$

式中: b ——局域化半径;

δ ——参数点-观测点或观测点-观测点的空间距离, 如果距离超过 $2b$, 则变量之间的相关性消失。

用修正后的协方差矩阵代替原来的协方差矩阵进行更新, 即式(9)修改为:

$$G_{j,i+1} = G_{ji} + \tilde{C}_{G,Y}^i (\tilde{C}_{Y,Y}^i + \alpha_i R)^{-1} (D + \sqrt{\alpha_i} \epsilon_j - Y_{ji}) \quad (14)$$

2 算例应用

2.1 算例设置

为验证 NS-ES-MDA 方法识别非高斯分布含水层参数的能力, 以含水层非高斯分布的渗透系数场估计为例构造一个二维理想算例。如图 1(a) 所示, 研究区 x 、 y 方向长度均为 800 m, 均匀离散成 80×80 网格, 网

格大小为 $10 \text{ m} \times 10 \text{ m}$, 南北为隔水边界, 西边界为给定水头边界, 水头值为 0 m, 东边界为出流边界, 流量为 $-20 \text{ m}^3/\text{d}$, 水流场稳定后停止抽水, 场地内水位逐渐恢复。采用二维承压含水层非稳定流模型模拟含水层水位逐渐恢复的过程。研究区共设置 64 个水位观测点、3 个水头控制点(#1、#2、#3)和 3 个浓度控制点(#4、#5、#6), 位置如图 1(a) 所示。含水层由两相组成, 渗透系数较大的砂土以及渗透系数较小的黏土, 参数设置见表 1, 对数渗透系数参考场如图 1(b) 所示。从场地对数渗透系数($\ln K$)直方图(图 1(c))可以看出, 场地的 $\ln K$ 明显不服从高斯分布。模拟期总时长为 5.0 d, 均匀离散为 100 个时间步, 前 20 个时间步的数据用于数据同化, 剩余时间步的数据用于检验数据同化效果。使用 FloPy 求解地下水流数值模拟模型^[26]。

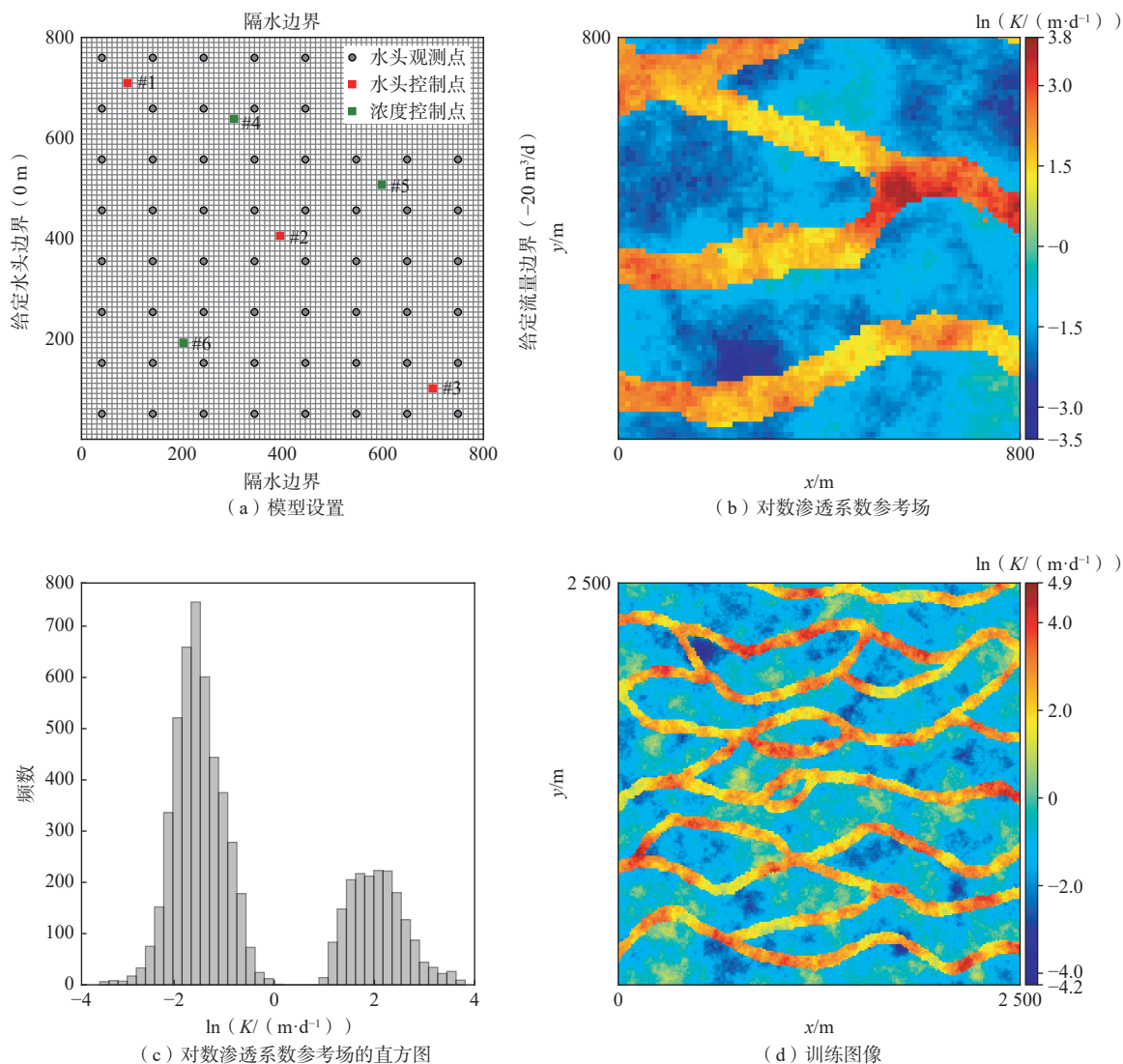


图 1 参考场的设置以及场地的训练图像

Fig. 1 Modeling setting and training image of the site

表 1 二维算例对数渗透系数参考场参数设置

Table 1 Parameters of the reference $\ln K$

岩性	变异函数 类型	均值 /($\text{m} \cdot \text{d}^{-1}$)	标准差 /($\text{m} \cdot \text{d}^{-1}$)	x方向 变程/m	y方向 变程/m
砂土	指数型	2.0	0.5	200	200
黏土	指数型	-1.5	0.5	200	200

以便于对比分析,设置不同的情景,见表 2。其中,使用 rNS-EnKF 进行参数估计的情景命名为 S_0 ,使用 NS-ES-MDA 进行参数估计的情景根据迭代次数分别命名为 S_1 、 S_2 、 S_3 、 S_4 、 S_5 、 S_6 。

表 2 二维算例含水层渗透系数场估计的情景设置及估计结果
Table 2 Estimation of the two-dimensional aquifer hydraulic conductivity field in different scenarios

情景	方法 (迭代次数)	$I_{\text{RMSE}}(\ln K)$ /($\text{m} \cdot \text{d}^{-1}$)	$I_{\text{ES}}(\ln K)$ /($\text{m} \cdot \text{d}^{-1}$)	运行 时间/s	与 S_0 的 比值/%
S_0	rNS-EnKF()	1.38	1.19	1 680	100%
S_1	NS-ES-MDA(1)	1.40	1.31	254	15%
S_2	NS-ES-MDA(2)	1.28	1.18	354	21%
S_3	NS-ES-MDA(4)	1.05	0.95	628	37%
S_4	NS-ES-MDA(6)	0.92	0.82	901	54%
S_5	NS-ES-MDA(8)	0.91	0.76	1 084	65%
S_6	NS-ES-MDA(10)	0.92	0.74	1 285	76%

注: rNS-EnKF 不是迭代算法,没有迭代次数; $I_{\text{RMSE}}(\ln K)$ 代表集合均值与真实 $\ln K$ 值的均方根误差; $I_{\text{ES}}(\ln K)$ 是后验参数集合的集合扩散; 运行时间指在 CPU 主频为 3.80 GHz 的计算机上运行所用时间; 与 S_0 的比值为不同情景运行时间与 S_0 运行时间的比值。

2.2 数据同化效果评价指标

采用均方根误差 (root mean square error, RMSE)、集合扩散 (ensemble spread, ES) 和纳什效率系数 (Nash-Sutcliffe efficiency coefficient, NSE) 定量评价 NS-ES-MDA 和 rNS-EnKF 的参数估计效果。均方根误差是评价数据同化效果的常用指标,均方根误差越小,参数估计结果与真实值越接近,其数学表达式为:

$$I_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{i,\text{ref}} - \bar{x}_i)^2} \quad (15)$$

式中: $\bar{x}_i$ ——网格 i 处的集合均值;

$x_{i,\text{ref}}$ ——网格 i 处的真实值;

n ——网格数量。

集合扩散用于量化与系统参数表征相关的不确定性,集合扩散越小,对系统认识的不确定性越小^[27]。集合扩散计算公式为:

$$I_{\text{ES}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \sigma_{x_i}^2} \quad (16)$$

式中: $\sigma_{x_i}^2$ ——网格 i 处的集合方差。

纳什效率系数是用来衡量模型预测结果与观测值之间相对误差的指标,它通过计算模型预测值与观

测值之间的方差比例来衡量拟合程度^[28]。纳什效率系数的取值范围从负无穷到 1,越接近 1 说明模拟值对观测值的拟合效果越好。纳什效率系数计算公式为:

$$I_{\text{NSE}} = 1 - \frac{\sum_{t=1}^T (O_t - M_t)^2}{\sum_{t=1}^T (O_t - \bar{O})^2} \quad (17)$$

式中: I_{NSE} ——纳什效率系数值;

O_t ——模型中某一点在 t 时刻的观测值;

M_t —— t 时刻的模拟值;

$\bar{O}$ ——所有时刻观测值的均值;

T ——总模拟时长。

3 结果

本文中所有情景均使用同一个初始集合,该集合大小为 500,集合中的每一个实现均使用快速采样法 (quick sampling, QS) 从训练图像中生成^[29],见图 1(d)。快速采样法是一种基于像素的多点地质统计方法,能够处理连续型变量。多点地质统计的理念非常简单:从 2D(或 3D)训练图像中复制模式,同时尊重场地数据。训练图像是预期空间结构的概念性描述,展示与模拟区域相关的地质非均质性,通常基于先验信息构建,Strebel^[30]介绍了建立训练图像的一般方法。在本文中,假设可以确定一个一般的先验特征,即场地内存在高渗通道,但结构的精确类型(通道的方向和尺寸)是不确定的,所以使用一个大于场地尺寸的训练图像以尽量涵盖在模拟区域中可能找到的所有模式。水头观测值服从均值为 0,标准差为 0.01 的高斯分布^[31-32]。NS-ES-MDA 的所有情景中 α_{geo} 均设置为 3。

3.1 数据同化效果

图 2 展示了初始集合和所有情景中后验参数集合的均值场和方差场。观察图 2 可知, S_0 同化水头数据后仅还原了部分通道特征且不确定性较大。在使用 NS-ES-MDA 的情景中,随迭代次数增加,数据同化效果逐渐改善, S_2 的结果与 S_0 非常接近,但 S_2 的运行时间仅为 S_0 的 21%(表 2)。 S_4 在恢复场地大部分通道特征的同时显著降低了方差。然而,两相交界处方差比相内部更大,表明非高斯场内部的曲线结构给参数估计带来了挑战。与 S_4 相比, S_5 、 S_6 提升不大,这表明虽然通过增加迭代次数可提升参数估计效果,但超过一定次数后提升变得困难。

从表 2 中可明显看出:在使用 NS-ES-MDA 的情

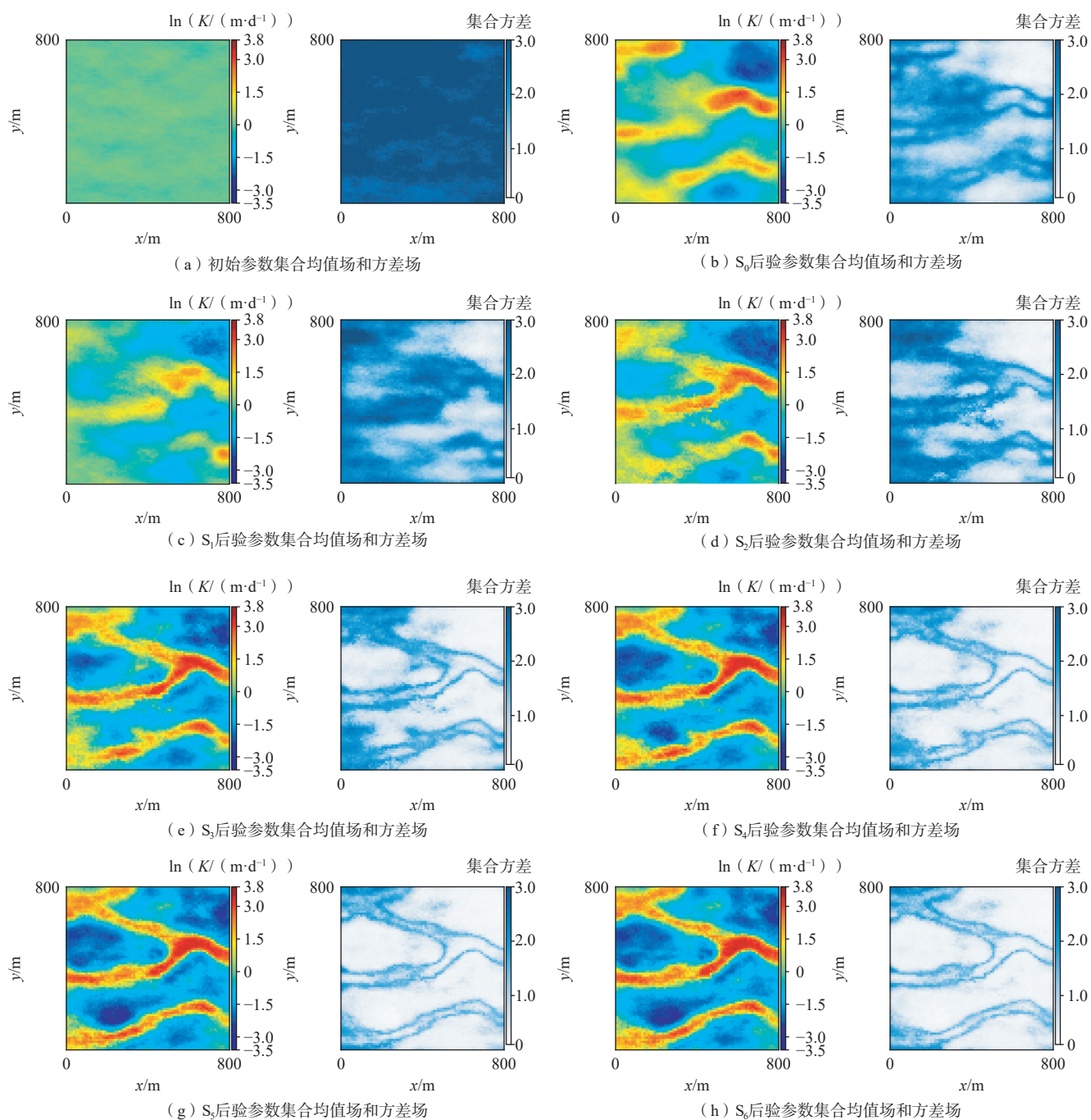


图 2 初始参数集合以及所有情景中后验参数集合的均值场和方差场

Fig. 2 Mean and variance fields of the initial ensemble and posterior ensemble for all scenarios

景中, 随迭代次数增加, $\ln K$ 均方根误差逐渐减小, 并最终稳定在 0.92 左右, 情景 S_5 表现最佳, 其 $\ln K$ 均方根误差为 0.91。将 S_5 作为 NS-ES-MDA 的最终结果, 相较于 S_0 , 其 $\ln K$ 均方根误差减小了约 34%, 集合扩散减小了约 36%, 同时运行时间减少了约 35%。

3.2 水头预测

将所有情景的参数估计结果重新代入水流模型, 得到水头控制点处的预测值, 并与观测值进行比较,

结果如图 3 所示。从图 3 可以看出, 随着迭代次数的增加 NS-ES-MDA 的水头预测效果明显得到提升, 且优于 S_0 的预测效果。其中, S_4 在 3 个水头控制点处的纳什效率系数均达到 0.99 以上(表 3)。 S_5 、 S_6 与 S_4 相比水头预测效果提升不大, 即在当前观测数据条件下超过一定迭代次数后预测效果提升变得困难。

值得关注的是, S_0 的 $\ln K$ 均方根误差略小于 S_1 , 大于其他情景(表 2), 即 S_0 的参数估计结果仅略好于 S_1 ,

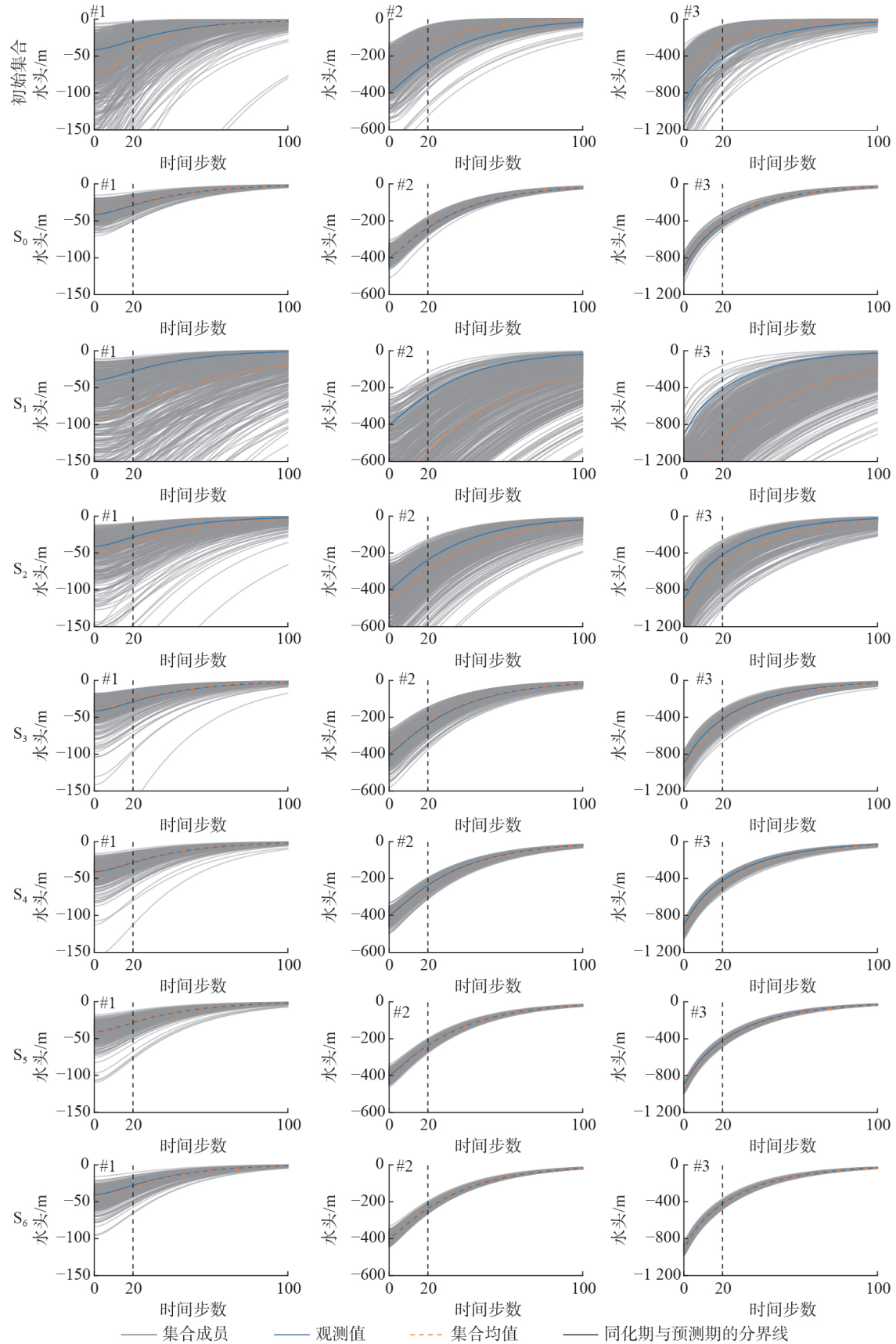


图 3 初始参数集合以及所有情景的后验参数集合对应的控制点处水头预测情况

Fig. 3 Head evolution at control points for the initial ensemble and all scenarios

表 3 水头控制点处评价指标
Table 3 Metrics at head control points

评价指标	初始集合	S_0	S_1	S_2	S_3	S_4	S_5	S_6
$I_{RMSE}(\text{Head \#1})/\text{m}$	12.10	1.07	38.00	5.82	1.03	0.70	0.24	1.22
$I_{ES}(\text{Head \#1})/\text{m}$	27.88	4.47	50.17	18.28	8.53	6.27	5.51	5.03
$I_{NSE}(\text{Head \#1})$	0.03	0.99	-8.59	0.78	0.99	1.00	1.00	0.99
$I_{RMSE}(\text{Head \#2})/\text{m}$	57.96	3.14	230.50	41.99	5.30	6.14	3.00	1.42
$I_{ES}(\text{Head \#2})/\text{m}$	50.92	12.54	202.68	76.99	26.19	14.53	9.68	8.84
$I_{NSE}(\text{Head \#2})$	0.72	1.00	-3.40	0.85	1.00	1.00	1.00	1.00
$I_{RMSE}(\text{Head \#3})/\text{m}$	123.40	9.81	416.68	72.18	19.15	24.69	11.03	3.89
$I_{ES}(\text{Head \#3})/\text{m}$	100.99	17.04	255.07	98.46	38.17	21.65	13.06	12.09
$I_{NSE}(\text{Head \#3})$	0.70	1.00	-2.42	0.90	0.99	0.99	1.00	1.00

但其水头预测明显优于 S_1 、 S_2 、 S_3 ，与 S_4 、 S_5 、 S_6 接近。根据表 3 中评价指标， S_0 与 S_5 的水头预测结果最为接近，除了 $I_{RMSE}(\text{Head \#1})$ 值相差较大（相对误差 77.6%），其他评价指标的相对误差均在 25% 以内，甚至在某些指标上 S_0 优于 S_5 。

3.3 溶质运移预测

为了验证数据同化后参数在溶质运移预测中的表现，在水流模型的基础上进一步开展了溶质运移试验。模拟过程被分为 2 个应力期：第 1 个应力期持续 200 d，在场地整个西边界线状污染源以恒定速率（300 g/d）释放非反应性溶质，纵向和横向弥散度分别为 40，4 m；第 2 个应力期持续 300 d，污染物停止释放，在整个模拟期内东边界以恒定速率（-20 m³/d）抽水，水流为稳定流，其余设置与表 1 相同。通过计算浓度控制点处的归一化浓度穿透曲线，并与初始集合、参考场的穿透曲线对比，评价参数估计效果。

从图 4 可以明显观察到初始集合中浓度的不确定性最大，拟合效果最差， S_5 对穿透曲线的拟合效果明显优于初始集合和 S_0 。表 4 为情景 S_0 和 S_5 下不同观测点处溶质运移预测结果评价指标统计值。 S_5 在所有评价指标上表现均明显优于 S_0 ，其在 #4、#5、#6 处的浓度均方根误差相比 S_0 分别提升约 78%、91%、62%；浓度集合扩散相比 S_0 下降约 71%、77%、29%；浓度纳什效率系数均达到 0.96 以上，拟合较好。

4 讨论

4.1 数据同化算法对比与分析

对水头数据的数据同化结果表明，在估计非高斯渗透系数场时，NS-ES-MDA 可以用较少的时间得到比 rNS-EnKF 更准确的结果，对于本文中的算例，即使 NS-ES-MDA 迭代 10 次运行时间仍然小于 rNS-EnKF。rNS-EnKF 的优势在于是一种在线的数据同化方法，可以

实现对参数的实时更新，而 NS-ES-MDA 需要等待收集所有的观测数据才能进行更新。NS-ES-MDA 迭代超过一定次数后参数估计效果提升困难，表明水头观测数据中包含的系统参数信息有限，如果想进一步提升参数估计效果，可能需要同化其他类型的数据，如浓度、温度等。

4.2 异参同效性的影响

在水头预测方面，尽管 S_0 的参数估计结果较差，其水头预测效果仍与 S_5 接近，甚至在某些指标上优于 S_5 。这说明 rNS-EnKF 方法更容易受到“异参同效性”的影响。然而，一旦场地的边界条件发生改变或不再进行水头预测，预测效果会大打折扣，溶质运移预测证实了这一结论。

4.3 空间连通性对溶质运移的影响

在溶质运移预测中，对比初始集合与 S_0 中浓度控制点处穿透曲线与真实值的不匹配可以发现，#4 和 #5 的预测值比观测值更晚到达峰值，而 #6 的预测值比观测值更早到达峰值（图 4）。这是因为 #4 和 #5 位于高渗通道中，初始集合与 S_0 均未明显识别出这些通道特征，导致溶质无法借助通道快速到达观测位置；而 #6 位于低渗区，这 2 种情景均高估了低渗区的渗透系数，导致预测时溶质更快到达 #6。

5 结论

(1) NS-ES-MDA 方法可以较为准确识别渗透系数场和含水介质的连通性，吸收相同数据后， S_5 与 S_0 对比，NS-ES-MDA 方法的参数估计精度比 rNS-EnKF 方法提升了约 34%，同时计算效率较 rNS-EnKF 提升了约 35%。

(2) NS-ES-MDA 方法相较于 rNS-EnKF 方法不易受到“异参同效”现象的影响，具有较强的更新能力，以保障同化得到较为准确的参数场估计值。

此外，本研究只用 NS-ES-MDA 同化水位数据识

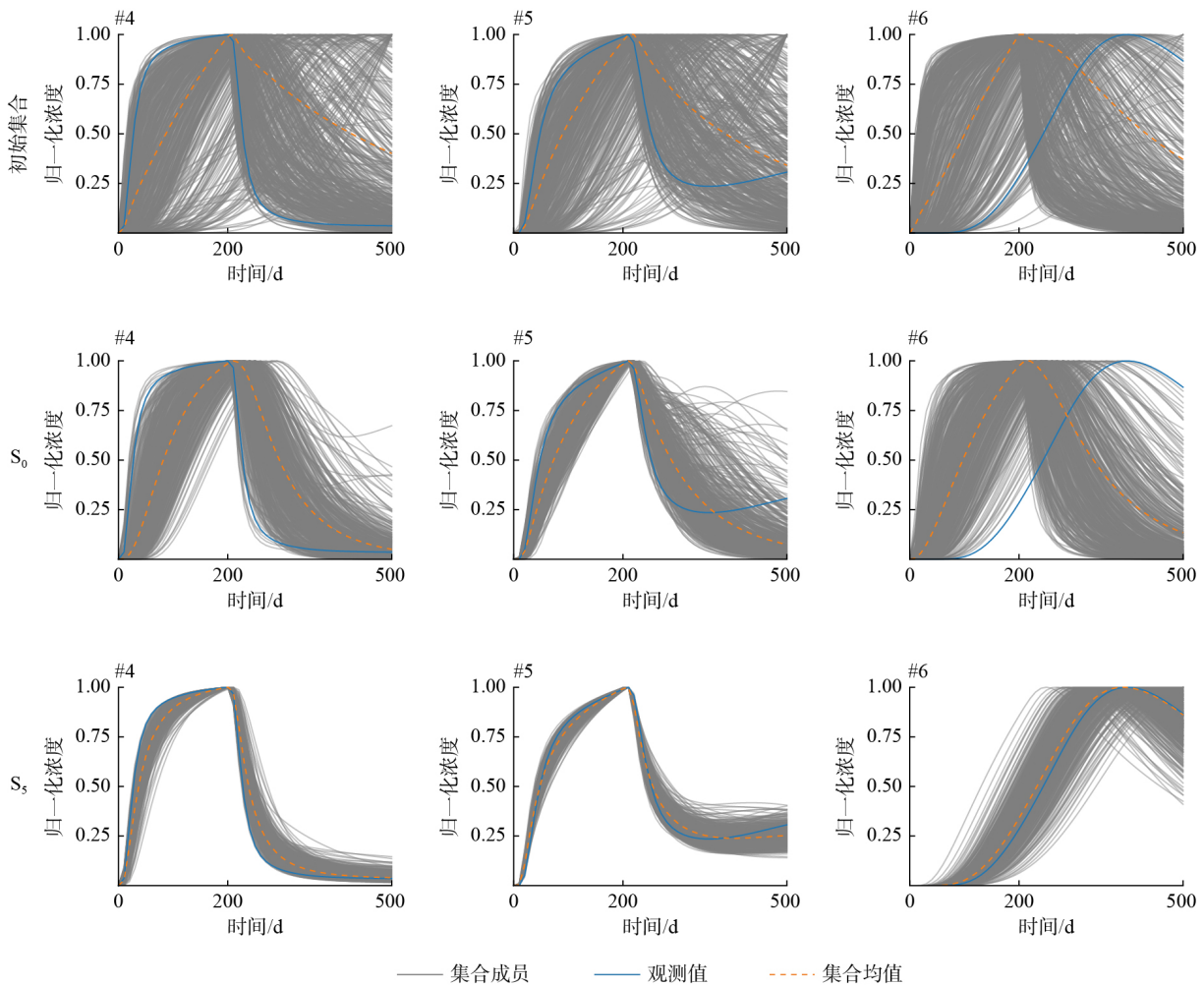


图 4 初始集合以及 S_0 、 S_5 在浓度控制点处的归一化浓度穿透曲线

Fig. 4 Normalized concentration breakthrough curves at concentration control points for the initial ensemble and S_0 , S_5

表 4 浓度控制点处的评价指标

Table 4 Metrics at concentration control points

评价指标	初始集合	S_0	S_1	S_2	S_3	S_4	S_5	S_6
$I_{\text{RMSE}}(\text{Conc \#4})/(\text{g}\cdot\text{m}^{-3})$	4.57	3.80	3.28	2.41	1.25	0.89	0.84	0.71
$I_{\text{ES}}(\text{Conc \#4})/(\text{g}\cdot\text{m}^{-3})$	6.58	2.69	4.72	3.19	1.29	0.84	0.79	0.75
$I_{\text{NSE}}(\text{Conc \#4})$	-0.21	0.42	0.57	0.75	0.92	0.97	0.96	0.97
$I_{\text{RMSE}}(\text{Conc \#5})/(\text{g}\cdot\text{m}^{-3})$	1.86	2.74	3.09	2.54	0.80	0.28	0.24	0.25
$I_{\text{ES}}(\text{Conc \#5})/(\text{g}\cdot\text{m}^{-3})$	2.61	1.64	2.18	1.84	0.99	0.55	0.38	0.33
$I_{\text{NSE}}(\text{Conc \#5})$	0.34	0.79	0.67	0.72	0.95	0.99	0.99	0.99
$I_{\text{RMSE}}(\text{Conc \#6})/(\text{g}\cdot\text{m}^{-3})$	25.12	26.63	28.40	28.07	27.79	18.57	9.99	6.84
$I_{\text{ES}}(\text{Conc \#6})/(\text{g}\cdot\text{m}^{-3})$	13.19	8.82	9.77	8.61	7.96	6.76	6.22	5.58
$I_{\text{NSE}}(\text{Conc \#6})$	-0.21	-0.83	-1.06	-0.98	-1.06	0.57	0.99	1.00

别了渗透系数场, 后续可以推广至同化多源数据(比如同时吸收水头、浓度等观测值)对多种参数的联合识别(比如同时识别渗透系数和污染源参数等)。

参考文献 (References) :

[1] 张江江. 地下水污染源解析的贝叶斯监测设计与

参数反演方法 [D]. 杭州: 浙江大学, 2017. [ZHANG Jiangjiang. Bayesian monitoring design and parameter inversion for groundwater contaminant source identification[D]. Hangzhou: Zhejiang University, 2017. (in Chinese with English abstract)]

[2] 薛佩佩, 文章, 梁杏. 地质统计学在含水层参数空间变异研究中的应用进展与发展趋势 [J]. 地质科技通

- 报, 2022, 41(1): 209 – 222. [XUE Peipei, WEN Zhang, LIANG Xing. Application and development trend of geostatistics in the research of spatial variation of aquifer parameters[J]. Bulletin of Geological Science and Technology, 2022, 41(1): 209 – 222. (in Chinese with English abstract)]
- [3] 曾献奎. 地下水流数值模拟不确定性分析与评价 [D]. 南京: 南京大学, 2012. [ZENG Xiankui. The uncertainty analysis and assessment of groundwater flow numerical simulation[D]. Nanjing: Nanjing University, 2012. (in Chinese with English abstract)]
- [4] CHEN Yan, ZHANG Dongxiao. Data assimilation for transient flow in geologic formations via ensemble Kalman filter[J]. *Advances in Water Resources*, 2006, 29(8): 1107 – 1122.
- [5] HUANG Chunlin, HU B X, LI Xin, et al. Using data assimilation method to calibrate a heterogeneous conductivity field and improve solute transport prediction with an unknown contamination source[J]. *Stochastic Environmental Research and Risk Assessment*, 2009, 23(8): 1155 – 1167.
- [6] 胡旭冉. 基于集合卡尔曼滤波融合的西北太平洋海表面温度时空分布特征研究 [D]. 上海: 上海海洋大学, 2018. [HU Xuran. Spatial and temporal distribution characteristics of sea surface temperature based on ensemble Kalman filtering fusion in the northwest Pacific Ocean[D]. Shanghai: Shanghai Ocean University, 2018. (in Chinese with English abstract)]
- [7] 马超群. 基于集合卡尔曼滤波的大气化学资料同化技术研究 [D]. 南京: 南京大学, 2020. [MA Chaoqun. Study of atmospheric chemistry data assimilation based on ensemble Kalman filter[D]. Nanjing: Nanjing University, 2020. (in Chinese with English abstract)]
- [8] 杨运, 吴吉春, 骆乾坤, 等. 考虑预报偏差的迭代式集合卡尔曼滤波在地下水水流数据同化中的应用 [J]. 水文地质工程地质, 2022, 49(6): 13 – 23. [YANG Yun, WU Jichun, LUO Qiankun, et al. Application of the bias aware Ensemble Kalman Filter with Confirming Option(Bias-CEnKF) in groundwater flow data assimilation[J]. Hydrogeology & Engineering Geology, 2022, 49(6): 13 – 23. (in Chinese with English abstract)]
- [9] 高善露. 基于集合卡尔曼滤波的油藏参数反演方法研究 [D]. 合肥: 合肥工业大学, 2019. [GAO Shanlu. Research of the ensemble Kalman filter for reservoir parameter inversion method[D]. Hefei: Hefei University of Technology, 2019. (in Chinese with English abstract)]
- [10] GODOY V A, NAPA-GARCÍA G F, GÓMEZ-HERNÁNDEZ J J. Ensemble random forest filter: An alternative to the ensemble Kalman filter for inverse modeling[J]. *Journal of Hydrology*, 2022, 615: 128642.
- [11] ZOVI F, CAMPORESE M, HENDRICKS FRANSSEN H J, et al. Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter[J]. *Journal of Hydrology*, 2017, 548: 208 – 224.
- [12] MORADKHANI H, HSU K L, GUPTA H, et al. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter[J]. *Water Resources Research*, 2005, 41(5): W05012.
- [13] ZHANG Jiangjiang, ZHENG Qiang, WU Laosheng, et al. Using deep learning to improve ensemble smoother: Applications to subsurface characterization[J]. *Water Resources Research*, 2020, 56(12): e2020WR027399.
- [14] CAO Zhendan, LI Liangping, CHEN Kang. Bridging iterative Ensemble Smoother and multiple-point geostatistics for better flow and transport modeling[J]. *Journal of Hydrology*, 2018, 565: 411 – 421.
- [15] ZHOU Haiyan, GÓMEZ-HERNÁNDEZ J J, HENDRICKS FRANSSEN H J, et al. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering[J]. *Advances in Water Resources*, 2011, 34(7): 844 – 864.
- [16] XU Teng, JAIME GÓMEZ-HERNÁNDEZ J, ZHOU Haiyan, et al. The power of transient piezometric head data in inverse modeling: An application of the localized normal-score EnKF with covariance inflation in a heterogenous bimodal hydraulic conductivity field[J]. *Advances in Water Resources*, 2013, 54: 100 – 118.
- [17] TANG Q, KURTZ W, BRUNNER P, et al. Characterisation of river-aquifer exchange fluxes: The role of spatial patterns of riverbed hydraulic conductivities[J]. *Journal of Hydrology*, 2015, 531: 111 – 123.
- [18] XU Teng, GÓMEZ-HERNÁNDEZ J J. Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter[J]. *Advances in Water Resources*, 2018, 112: 106 – 123.
- [19] SKJERVHEIM J A, EVENSEN G, HOVE J, et al. An Ensemble Smoother for assisted History Matching[C]//

- SPE Reservoir Simulation Symposium All Days. Woodlands: Society of Petroleum Engineers, 2011: SPE-141929-MS.
- [20] EMERICK A A, REYNOLDS A C. Ensemble smoother with multiple data assimilation[J]. *Computers & Geosciences*, 2013, 55: 3 – 15.
- [21] EVENSEN G. Analysis of iterative ensemble smoothers for solving inverse problems[J]. *Computational Geosciences*, 2018, 22(3): 885 – 908.
- [22] HAMILL T M, WHITAKER J S, SNYDER C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter[J]. *Monthly Weather Review*, 2001, 129(11): 2776 – 2790.
- [23] ANDERSON J L. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter[J]. *Physica D: Nonlinear Phenomena*, 2007, 230(1/2): 99 – 111.
- [24] CHEN Yan, OLIVER D S. Cross-covariances and localization for EnKF in multiphase flow data assimilation[J]. *Computational Geosciences*, 2010, 14(4): 579 – 601.
- [25] GASPARI G, COHN S E. Construction of correlation functions in two and three dimensions[J]. *Quarterly Journal of the Royal Meteorological Society*, 1999, 125(554): 723 – 757.
- [26] BAKKER M, POST V, LANGEVIN C D, et al. Scripting MODFLOW model development using python and FloPy[J]. *Ground Water*, 2016, 54(5): 733 – 739.
- [27] CHEN Zi, XU Teng, GÓMEZ-HERNÁNDEZ J J, et al. Contaminant spill in a sandbox with non-gaussian conductivities: Simultaneous identification by the restart normal-score ensemble Kalman filter[J]. *Mathematical Geosciences*, 2021, 53(7): 1587 – 1615.
- [28] NASH J E, SUTCLIFFE J V. River flow forecasting through conceptual models part I—a discussion of principles[J]. *Journal of Hydrology*, 1970, 10(3): 282 – 290.
- [29] GRAVEY M, MARIETHOZ G. QuickSampling v1.0: A robust and simplified pixel-based multiple-point simulation approach[J]. *Geoscientific Model Development*, 2020, 13(6): 2611 – 2630.
- [30] STREBELLE S. Multiple-point statistics simulation models: Pretty pictures or decision-making tools?[J]. *Mathematical Geosciences*, 2021, 53(2): 267 – 278.
- [31] 崔凯鹏. 集合卡尔曼滤波在地下水流及溶质运移数据同化中的应用探讨 [D]. 南京: 南京大学, 2013. [CUI Kaipeng. The application and discussion of ensemble Kalman filter method in groundwater flow and solute transport data assimilation[D]. Nanjing: Nanjing University, 2013. (in Chinese with English abstract)]
- [32] WEN Xianhuan, GÓMEZ-HERNÁNDEZ J J. Numerical modeling of macrodispersion in heterogeneous media: A comparison of multi-Gaussian and non-multi-Gaussian models[J]. *Journal of Contaminant Hydrology*, 1998, 30(1/2): 129 – 156.

编辑: 张若琳