

基于相关性局域化迭代集合平滑反演渗透系数场

夏传安, 王浩, 简文彬

Estimation of conductivity fields by using a correlation-based localization scheme of iterative ensemble smoother

XIA Chuan'an, WANG Hao, and JIAN Wenbin

在线阅读 View online: <https://doi.org/10.16030/j.cnki.issn.1000-3665.202303033>

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

基于自组织神经网络的污染场地多监测指标相关性分析

Correlation analysis of multiple monitoring indicators of contaminated site based on self-organizing map

马春龙, 施小清, 许伟伟, 任静华, 王佩, 吴吉春 水文地质工程地质. 2021, 48(3): 191-202

基于数字图像相关方法的裂隙砂岩应变场演化规律及前兆识别

Evolution law of strain field and precursor identification of flawed sandstone based on Digital Image Correlation method

张科, 李娜 水文地质工程地质. 2021, 48(3): 150-156



关注微信公众号, 获得更多资讯信息

DOI: 10.16030/j.cnki.issn.1000-3665.202303033

夏传安, 王浩, 简文彬. 基于相关性局域化迭代集合平滑反演渗透系数场[J]. 水文地质工程地质, 2024, 51(1): 12-21.

XIA Chuan'an, WANG Hao, JIAN Wenbin. Estimation of conductivity fields by using a correlation-based localization scheme of iterative ensemble smoother[J]. Hydrogeology & Engineering Geology, 2024, 51(1): 12-21.

基于相关性局域化迭代集合平滑反演渗透系数场

夏传安, 王浩, 简文彬

(福州大学紫金地质与矿业学院岩土与地质工程系, 福建 福州 350108)

摘要: 在地下水流和溶质运移问题中, 有较多研究基于物理距离局域化集合同化方法反演水文地质参数。当反演参数与观测信息之间不存在物理距离时, 这种方法不适用。为了克服这个局限, 通过渗透系数与水头信息之间的相关性计算局域化方法的阻滞因子, 构建基于相关性的局域化迭代集合平滑方法。为了方便比较, 将该方法和一种基于物理距离的局域化迭代集合平滑一同用于同化水头信息反演二维孔隙承压含水层的渗透系数场。算例中考虑了不同集合大小、观测误差及观测数量等因子的组合, 便于分析其对渗透系数反演精度的影响。研究结果显示: (1) 在所有算例中新方法得到的渗透系数均方根误差范围为 [0.830 7, 0.959 0], 都小于基于物理距离方法的均方根误差, 范围为 [0.839 4, 1.000 0]; (2) 基于物理距离方法得到的渗透系数场空间上存在不连续性, 而新方法的结果不存在此现象。文章提出了一种新的基于相关性局域化迭代平滑方法, 该方法不需要依赖参数与观测信息之间的物理距离且参数反演精度高于基于物理距离的方法, 可作为参数反演的科学工具。

关键词: 数据同化; 相关性局域化; 迭代集合平滑; 物理距离局域化; 渗透系数场

中图分类号: P641.2

文献标志码: A

文章编号: 1000-3665(2024)01-0012-10

Estimation of conductivity fields by using a correlation-based localization scheme of iterative ensemble smoother

XIA Chuan'an, WANG Hao, JIAN Wenbin

(Department of Geotechnical and Geological Engineering, Zijin School of Geology and Mining, Fuzhou University, Fuzhou, Fujian 350108, China)

Abstract: In the studies of groundwater flow and solute transport, many efforts have been made to estimate hydrogeological parameters through physical-distance based localization schemes of ensemble assimilation approaches. However, these methods are unavailable when there are no physical distances between parameters and observations. To avoid this limitation, we calculate the tapering factors in terms of the correlation coefficients between parameters and observations and develop a novel correlation-based localization scheme of iterative ensemble smoother. For the purpose of comparison, a physical distance-based scheme of iterative ensemble smoother together with the new approach are used to assimilate hydraulic head information and estimate the hydraulic conductivity field of a 2D confined aquifer. Among the test cases, we consider different configurations of ensemble size, observation error and number of observations, and their impacts on the accuracy of conductivity

收稿日期: 2023-03-08; 修订日期: 2023-05-09

投稿网址: www.swdzgcdz.com

基金项目: 国家自然科学基金项目(42002247; U2005205; 41972268); 广东省基础与应用基础研究基金项目(2020A1515111054)

第一作者: 夏传安(1991—), 男, 博士, 讲师, 主要从事水文地质工程地质数值建模研究。E-mail: xiachuanan@163.com

通讯作者: 王浩(1978—), 男, 博士, 教授, 主要从事岩土工程与地质工程方面的研究。E-mail: h_wang@126.com

estimation can be well explored. The results show that (1) the root mean square error of hydraulic conductivity, R_{MSE} , obtained through the new approach for each test case of interest, is lower than its counterpart through the physical distance-based approach, with the R_{MSE} ranges of [0.830 7, 0.959 0] and [0.839 4, 1.000 0] for all test cases through the two approaches, respectively. (2) The estimated conductivity field has discontinuities when using the physical distance-based approach, but this does not happen when using the new approach. In this study, we develop a novel correlation-based localization scheme of iterative ensemble smoother, which is free of the definitions of physical distances between parameters and observations, yields higher accuracy of parameter estimation in comparison with the physical distance-based approach, and can be a useful tool for estimating hydrogeological parameters.

Keywords: data assimilation; correlation-based localization; iterative ensemble smoother; physical distance-based localization; conductivity field

集合数据同化方法,包括集合卡尔曼滤波、集合平滑及其迭代形式、粒子滤波等,已被应用于水文地质^[1-6]、工程地质^[7-8]、气象^[9]、石油工程^[10-11]等领域。在水文地质领域,集合数据同化方法常被用于反演水文地质参数,包括渗透系数、土壤水动力参数及其他参数。与三维/四维变分方法^[12]比较,集合数据同化方法的理论相对简单。由于受计算资源和时间限制,在应用数据同化方法时通常只考虑较小的集合,这会引入参数与观测之间的伪相关性。伪相关性会引起滤波近交甚至滤波发散,影响参数反演精度和观测信息的模拟精度^[13]。局域化方法通常被用来削弱伪相关性对参数反演的消极影响。目前,局域化方法按照原理可以分为基于物理距离^[14-16]和相关性^[10-12,17]2种。

基于物理距离的方法是依据参数与观测信息之间的物理距离估计2个变量之间的相关程度,距离越远越不相关,反之越相关。南统超等^[16]应用基于物理距离的局域化集合卡尔曼滤波同化水头信息反演一个二维理想孔隙含水层的渗透系数场,结果表明局域化方法能有效消除伪相关性的消极影响,提高渗透系数的反演精度。Tong等^[14]研究表明利用集合卡尔曼滤波反演渗透系数时只同化距离较近的水头观测信息有利于削弱伪相关性的消极影响。Xia等^[4]利用基于物理距离的局域化迭代集合卡尔曼滤波同化水头和盐度信息,反演海岸带含水层的渗透系数场。基于相关性的方法依据参数与观测信息之间的相关性估计2个变量之间的相关程度,数学定义比基于物理距离的方法更清晰。基于相关性的局域化集合数据同化方法常被用于解决气象预报^[18-19]和石油开采^[10-11,17]问题,专注于地下水问题的研究还比较少。Miyoshi^[19]利用基于相关性的局域化集合拓展卡尔曼滤波提高气象预报精度。Luo等^[11]和Soares等^[17]利用基于相

关性的局域化迭代集合平滑反演储油层的渗透率和孔隙度。与基于相关性的局域化比较,基于物理距离的方法存在一个不可避免的缺陷:参数与观测之间必须有明确的物理距离定义。在水文地质领域,有些观测信息和参数没有明确的坐标,比如多含水层的长滤管水头^[20]、土壤剖面的水动力参数^[21]、流域尺度土壤参数^[22]以及含水层尺度的渗透系数和孔隙度^[23]等,导致参数与观测信息之间没有明确的物理距离。综上所述,有必要加深对基于相关性局域化方法的研究。

本文提出了一种新的基于相关性的局域化方法并将其与迭代集合平滑方法耦合用于反演一个理想二维孔隙承压含水层的渗透系数场。在算例中,水头观测信息与渗透系数之间存在明确的物理距离。本文同时考虑了基于物理距离的局域化方法,便于分析2种局域化方法的表现。

1 理论与方法

1.1 地下水流方程

考虑理想二维孔隙承压含水层,非稳定流条件下水头(h)控制方程为^[24]:

$$S \frac{\partial h(\mathbf{x}, t)}{\partial t} = \nabla \cdot [K(\mathbf{x}) \nabla h(\mathbf{x}, t)] + f(\mathbf{x}, t) \quad (1)$$

式中: S ——储水系数;

t ——时间;

\mathbf{x} ——空间位置向量;

∇ ——梯度算子;

$\nabla \cdot$ ——散度算子;

K ——渗透系数;

f ——源汇项。

在本文中,考虑 S 为定值, K 为一个空间随机变量, f 为定值。通过局域化迭代集合平滑同化 h 观测

信息反演 K 。

1.2 局域化迭代集合平滑

参考 Luo 等^[25]的研究,局域化迭代集合平滑的第 i 次参数更新方程为:

$$\begin{cases} \mathbf{m}^{i+1} = \mathbf{m}^i + (\mathbf{R}^i \otimes \mathbf{K}_{\text{Gain}}^i) \Delta \mathbf{d}^i \\ \mathbf{K}_{\text{Gain}}^i = \mathbf{S}_m^i (\mathbf{S}_d^i)^\top (\mathbf{S}_d^i (\mathbf{S}_d^i)^\top + \gamma^i \mathbf{I})^{-1} \\ \Delta \mathbf{d}^i \equiv \mathbf{g}(\mathbf{m}^i) - \mathbf{d}_{\text{obs}} \\ \gamma^i = \xi^i \text{trace}(\mathbf{S}_d^i (\mathbf{S}_d^i)^\top) / O \end{cases} \quad (2)$$

式中: $\mathbf{m} = [m_1, m_2, \dots, m_p]^\top$ ——参数向量;

P ——参数总数;

\mathbf{R} ——阻滞矩阵($P \times O$);

$\mathbf{R} \otimes \mathbf{K}_{\text{Gain}}$ ——矩阵 \mathbf{R} 与 \mathbf{K}_{Gain} 的舒尔积;

\mathbf{S}_m 、 \mathbf{S}_d ——参数异常矩阵($P \times N$)、观测信息异常矩阵($O \times N$);

ξ^i ——Levenberg-Marquart 算法中自动调整的系数,调整方法与 Luo 等^[10] 相同;

$$r_{ks} = f_{\text{GC}}(z_{ks}) = \begin{cases} -\frac{1}{4}z_{ks}^5 + \frac{1}{2}z_{ks}^4 + \frac{5}{8}z_{ks}^3 - \frac{5}{3}z_{ks}^2 + 1 & 0 \leq z \leq 1 \\ \frac{1}{12}z_{ks}^5 - \frac{1}{2}z_{ks}^4 + \frac{5}{8}z_{ks}^3 + \frac{5}{3}z_{ks}^2 - 5z_{ks} + 4 - \frac{2}{3}z_{ks}^{-1} & 1 < z \leq 2 \\ 0 & z > 2 \end{cases} \quad (3)$$

Gaspari-Cohn 函数的输入 z_{ks} 不同得到的 r_{ks} 也不同。基于物理距离和相关性的局域化方法,由于其机理不同,它们对应的 z_{ks} 表达式理论上不相同。

1.3 基于物理距离的局域化

参考 Xia 等^[4]和 Nan 等^[15],基于物理距离的局域化方法考虑阻滞函数输入 z_{ks} 为:

$$\begin{cases} z_{ks} = 3 \left(\frac{d_1}{\beta_1} + \frac{d_2}{\beta_2} \right) \\ \beta_i = \delta_i (\sqrt{9 + 8N} - 5) / 4 \\ \delta_i = \lambda_i / 2 \end{cases} \quad (4)$$

式中: d_i ——两空间位置(这里指第 s 个观测信息与第 k 个参数的空间位置)在 x_i 方向的距离, $i = 1, 2$;

β_i ——在 x_i 方向的阻滞关键长度, $i = 1, 2$ (参见文献[4, 18]);

λ_i ——在 x_i 方向的渗透系数相关长度, $i = 1, 2$ 。

为了便于表达,基于物理距离的局域化迭代集合平滑被记为 DL_iES。

1.4 基于相关性的局域化

假定 z_{ks} 与相关系数(ρ_{ks})满足椭圆方程:

\mathbf{I} —— $O \times O$ 的单位矩阵;

$\mathbf{g}(\cdot)$ ——模型-观测算子;

$\mathbf{d}_{\text{obs}} = [d_{\text{obs}}^1, d_{\text{obs}}^2, \dots, d_{\text{obs}}^O]^\top$ ——观测向量;

O ——观测信息总数。

在本文中, \mathbf{d}_{obs} 包含了所有的水头观测信息, \mathbf{m} 包含了含水层不同位置的渗透系数。本文设置 2 点迭代集合平滑收敛的依据:

(1) 达到最大迭代次数;

(2) 连续 2 次迭代之间参数向量之差的二范小于设定的阈值。

不同局域化方法得到不同阻滞矩阵(\mathbf{R}), $\mathbf{R} \otimes \mathbf{K}_{\text{Gain}}$ 使得与参数相关度高的观测信息才能用于更新参数。 \mathbf{R} 中第 k 行 s 列的元素被称为阻滞因子(r_{ks}), 通过阻滞函数得到。与大多数研究^[14-15, 26] 相同, 本文中的阻滞函数为阻滞函数的输入与阻滞因子之间的映射关系。

本文选取 Gaspari-Cohn 函数为阻滞函数^[10, 26]:

$$\begin{cases} \frac{z_{ks}^2}{1/(1-w)^2} + \rho_{ks}^2 = 1, 0 \leq z_{ks} \\ \rho_{ks} = \frac{\sum_{i=1}^N (m_{k,i} - \bar{m}_k)(y_{s,i}^{\text{RS}} - \bar{y}_s^{\text{RS}})}{N \sigma_{m_k} \sigma_{y_s^{\text{RS}}}} \end{cases} \quad (5)$$

式中: \bar{m}_k ——第 k 个参数变量的集合均值;

\bar{y}_s^{RS} ——第 s 个观测变量的集合均值;

σ_{m_k} ——参数变量 m_k 的标准差;

$\sigma_{y_s^{\text{RS}}}$ ——观测变量 y_s^{RS} 的标准差。

由式(5)可得 z_{ks} 的表达式为:

$$z_{ks} = \varphi(\rho_{ks}) = \frac{\sqrt{1 - \rho_{ks}^2}}{1 - w} \quad (6)$$

式中: w ——观测水头与渗透系数不相关时相关系数的高斯白噪音阈值, $0 < w < 1$ 。

依据 Anderson^[27] 的研究, 本文定义 w 的表达式为:

$$w = \frac{\alpha}{\sqrt{N}} \quad (7)$$

式中: α ——相关系数白噪音标准差 ($1/\sqrt{N}$ ^[27]) 的倍数, $0 < \alpha < \sqrt{N}$ 。

依据高斯分布的概率密度函数, 通常取 1, 2 或 3 个标准差表示白噪音水平。本文将 z_{ks} 与 ρ_{ks} 之间映射

关系描述为椭圆方程。当 $|\rho_{ks}| = 1$ 时,认为参数与观测信息之间高度相关, $r_{ks} = f_{GC}(g(1)) = 1$ 。当 $|\rho_{ks}| < w$ 时,认为参数与观测信息之间相关性很弱,不同化第 s 个观测信息用于反演含水层的第 k 个渗透系数,即 $r_{ks} = 0$ 。综上,本文用于计算阻滞因子(r_{ks})的表达式为:

$$r_{ks} = \begin{cases} f_{GC}(\varphi(\rho_{ks})) & w \leq |\rho_{ks}| \\ 0 & |\rho_{ks}| < w \end{cases} \quad (8)$$

从式(8)可推得:

(1)当 N 较小时, z_{ks} 应该取较大的值(即椭圆的长轴半径 $a = 1/(1-w)$, $1 < a < \infty$)使得阻滞函数得到较小的 r_{ks} 。

(2)当 N 很大(比如 $N \rightarrow \infty$)时, $z_{ks} = \varphi(\rho_{ks}) = \sqrt{1 - \rho_{ks}^2}$ 不恒等于1。理论层面上,当 $N \rightarrow \infty$ 时,局域化方法应该满足 $r_{ks} \equiv 1$ 。但是,局域化方法只用于 N 较小的情况(通常 N 取值为50, 100, 300, 500等),对新方法的普适性影响较小。为了便于表达,将本文提出的基于相关性的局域化迭代集合平滑记为CL_iES。

2 算例

本文考虑二维承压孔隙含水层的渗透系数场反演作为算例。含水层大小为 80×80 (全文单位保持一致,本文剩余部分省略单位描述)。含水层左右2侧 $x_1 = 0, 80$ 处分别为给定水头边界 $h = 1, 0$,上下2侧 $x_2 = 0, 80$ 为隔水边界,中心处($x_1 = 40, x_2 = 40$)设置一口井(图1),抽水率恒定为2。含水层的初始水头为不

抽水情况下含水层的稳定水头。含水层储水系数为0.001。非稳定地下水流模型的总模拟时长为4,被均匀地分为20个应力期。数值模型采用标准有限元法求解(参见文献[20])。将含水层空间等距离散为 81×81 ($P = 6561$)的节点网格。渗透系数的对数($Y = \ln K$)为服从高斯分布的空间随机变量,用指数协方差函数刻画渗透系数空间变异性:

$$C_Y = \sigma_Y^2 \exp\left(-\frac{1}{2} \left[\frac{d_1}{\lambda_1} + \frac{d_2}{\lambda_2} \right]^2\right) \quad (9)$$

式中: σ_Y^2 ——渗透系数的方差。

算例只考虑渗透系数为随机变量,其余参数和边界条件设置为已知。使用均值0.0和方差1.0生成的 Y 参考场见图1(a)。此外,使用均值0.5和方差1.0生成初始 Y 场的集合。 Y 参考场与初始 Y 场的渗透系数具有相同的相关长度, $\lambda_1 = \lambda_2 = 8$ 。以 Y 参考场作为模型的输入获得的水头场被称为参考水头场。观测水头通过监测井的参考水头加上标准差为 σ_{obs} 的高斯白噪音得到。每个应力期的每口监测井都提供观测信息。

2种方法(即DL_iES和CL_iES)在迭代过程中使用的参数相同。具体包括:Levenberg-Marquart算法中初始系数(ξ^0)设置为20;迭代集合平滑的最大迭代次数为20,连续2次迭代之间参数向量之差的二范小于 1×10^{-6} 时迭代停止。

在算例中,除了集合大小(N)、水头观测误差的标准差(σ_{obs})、监测井数量(N_{mon})及相关系数白噪音标准差的倍数(α)的取值有变化,其余设置与以上描述相同。采用控制变量法研究 N 、 σ_{obs} 、 N_{mon} 及 α 的取值对渗透系数反演精度的影响。具体描述如下:第1组试验,设定 $\sigma_{\text{obs}} = 0.01$ 、 $N_{\text{mon}} = 48$ (即 $O = 960$)、 $\alpha = 2$, N 的取值为50, 100, 500;第2组,设定 $N = 100$ 、 $N_{\text{mon}} = 48$ 、 $\alpha = 2$, σ_{obs} 的取值为0.1, 0.01, 0.001;第3组,设定 $N = 100$ 、 $\sigma_{\text{obs}} = 0.01$ 、 $\alpha = 2$,考虑 $N_{\text{mon}} = 16$ (即 $O = 320$,图1b), 48(即 $O = 960$,图1c), 168(即 $O = 3360$,图1d);第4组,设定 $N = 100$ 、 $\sigma_{\text{obs}} = 0.01$ 、 $N_{\text{mon}} = 48$, α 的取值为1.0, 1.5, 2.0, 2.5, 3.0。

使用渗透系数的均方根误差(R_{MSE})和水头的绝对误差(E_h)定量评价2种方法的表现。 R_{MSE} 与 E_h 的表达式分别为:

$$R_{\text{MSE}} = \sqrt{\frac{1}{P} \sum_j^P \left(Y_j^{\text{ref}} - \langle Y_j \rangle^{\text{est}} \right)^2} \quad (10)$$

$$E_h = \frac{1}{P} \sum_j^P |h_j^{\text{ref}} - \langle h_j \rangle^{\text{up}}| \quad (11)$$

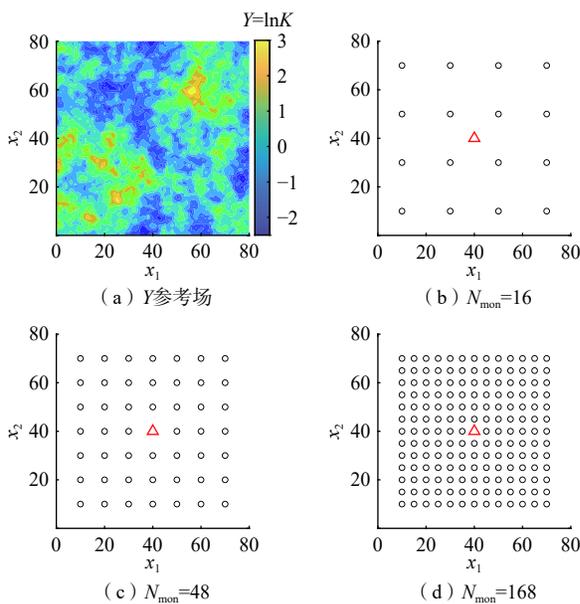


图1 Y 参考场和16, 48, 168口监测井空间分布

Fig. 1 Y reference field and spatial distribution of 16, 48 and 168 monitoring wells

式中： Y_j^{ref} 、 $\langle Y_j \rangle^{\text{est}}$ —— j 节点 Y 参考值、在该节点的 Y 估计值；

h_j^{ref} 、 $\langle h_j \rangle^{\text{up}}$ —— j 节点上的参考水头、更新水头。

Y 反演场的标准差 (S_Y)，常用于定量 Y 集合内部的差异性，其表达式为：

$$S_Y = \sqrt{\frac{1}{P} \sum_{j=1}^P (\sigma_{Y_j}^2)^{\text{est}}} \quad (12)$$

式中： $(\sigma_{Y_j}^2)^{\text{est}}$ ——第 j 个节点处 Y 方差的估计值。

3 结果

3.1 集合大小的影响(第 1 组)

图 2 展示 $N = 50, 100, 500$ 时由 DL_iES 和 CL_iES

得到 R_{MSE} (图 2a)、 S_Y (图 2b) 及 E_h (图 2c) 随迭代次数变化的曲线。当 N 由 50 增大到 100 时，DL_iES 得到的 R_{MSE} 、 S_Y 及 E_h 值的变化幅度比由 CL_iES 得到变化幅度小。随着 N 增大，DL_iES 和 CL_iES 得到的 R_{MSE} 、 S_Y 及 E_h 总体展现下降趋势。当集合大小相同时，CL_iES 得到的 R_{MSE} 和 E_h 总体都要低于 DL_iES 得到的值，特别是当集合较小(即 $N = 50$ 或 100)时。

与图 2 相对应，表 1 列出 DL_iES 和 CL_iES 第 20 次迭代得到的 R_{MSE} 、 S_Y 及 E_h 值。随着集合增大，DL_iES 和 CL_iES 对应的渗透系数反演误差和水头模拟误差均减小，前者对应的误差基本高于后者。综合 R_{MSE} 与 S_Y 结果，2 种方法都能维持 Y 集合内部的差异性和能有效减小伪相关性对反演渗透系数的消极影响。

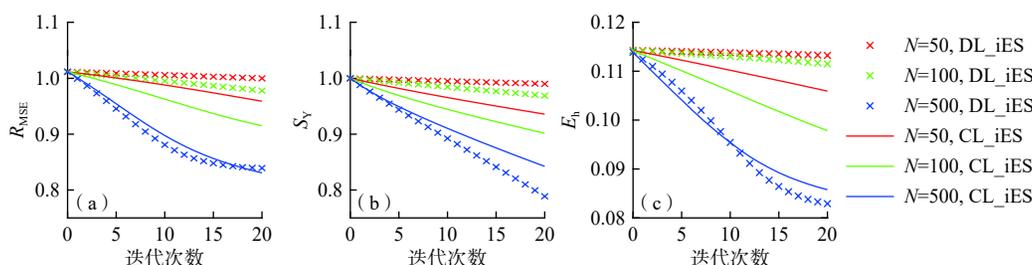


图 2 由 DL_iES 和 CL_iES 在 $N = 50, 100, 500$ 时得到的 R_{MSE} 、 S_Y 和 E_h 随迭代次数变化的曲线
Fig. 2 Changes of R_{MSE} , S_Y and E_h obtained through DL_iES and CL_iES with iteration times when $N = 50, 100, 500$

表 1 第 1 组第 20 次迭代后 DL_iES 和 CL_iES 得到的 R_{MSE} 、 S_Y 和 E_h 值

Table 1 Final values of R_{MSE} , S_Y and E_h through DL_iES and CL_iES for Group 1 at the 20th iteration

参数	模型	集合大小		
		50	100	500
R_{MSE}	DL_iES	1.000 0	0.977 9	0.839 4
	CL_iES	0.959 0	0.915 1	0.830 7
S_Y	DL_iES	0.990 1	0.968 9	0.788 7
	CL_iES	0.935 9	0.901 6	0.842 4
E_h	DL_iES	0.113 2	0.111 5	0.082 9
	CL_iES	0.105 9	0.097 8	0.085 8

图 3 绘制了第 1 组中 DL_iES 和 CL_iES 在第 20 次迭代后得到的 Y 估计场和对应的 Y 方差 (σ_Y^2) 估计场。从图 3 中可以看出：

(1) 当 $N = 50$ 或 100 时，DL_iES 得到的反演结果空间上出现不连续性，而 CL_iES 则无此现象，且 CL_iES 得到的结果更接近 Y 参考场，见图 1(a)。

(2) 当 $N = 500$ 时，2 种方法得到 Y 和 σ_Y^2 估计场都较为相似，接近参考场。DL_iES 得到的 σ_Y^2 估计场出

现规律的不连续性随着 N 增大而逐渐消失。当 N 相同时，DL_iES 和 CL_iES 得到的 σ_Y^2 估计场总体相近，均是介于下边界和抽水井的区域值较低，而介于上边界和抽水井的区域值较高。

综合以上结果，当考虑不同集合时 CL_iES 具有良好自适应性，且表现优于 DL_iES。

3.2 观测误差的影响(第 2 组)

第 2 组中 DL_iES 和 CL_iES 第 20 次迭代得到的 R_{MSE} 、 S_Y 及 E_h 值，见表 2。随着 σ_{obs} 由 0.001 增大到 0.1， R_{MSE} 和 S_Y 总体上变化不大，而 E_h 呈现明显的递增趋势。造成这一现象的原因是渗透系数反演过程存在多解性。这些结果说明局域化方法在应对观测误差变化时，也具有较好的适应能力。与 DL_iES 相比，CL_iES 得到的 R_{MSE} 和 E_h 更低。

综合以上结果，当考虑不同 σ_{obs} 时 CL_iES 具有良好的自适应性，且表现优于 DL_iES。

3.3 观测数量的影响(第 3 组)

第 3 组中 DL_iES 和 CL_iES 第 20 次迭代得到的 R_{MSE} 、 S_Y 及 E_h 值如表 3 所示。随着 N_{mon} 增大，渗透系

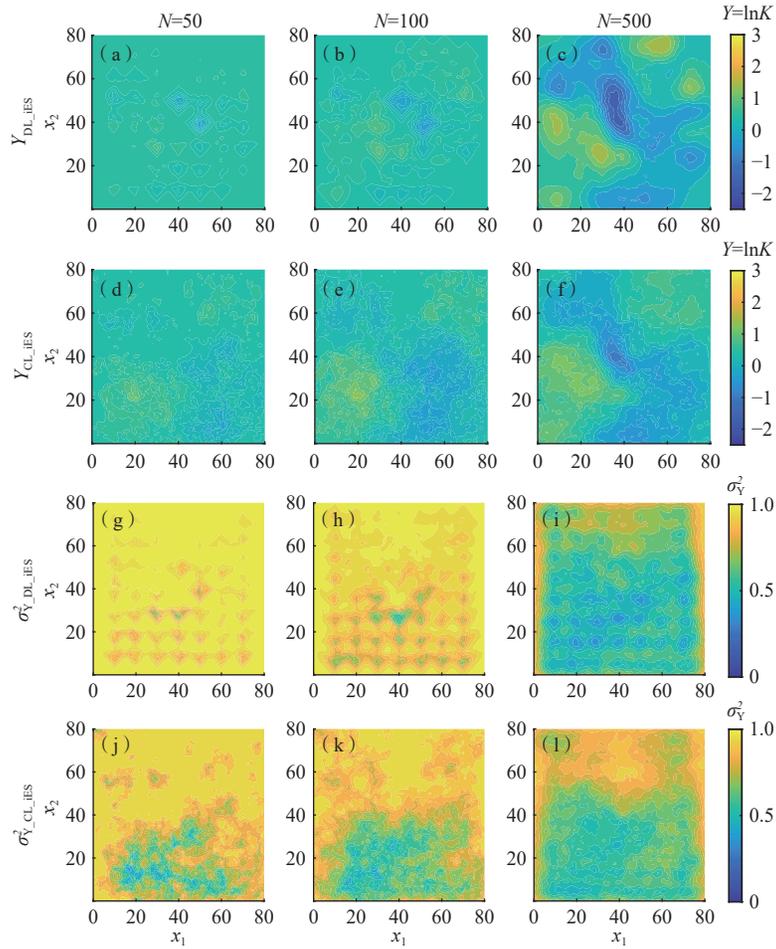


图 3 当 $N = 50, 100, 500$ 时, 通过 DL_iES 和 CL_iES 在第 20 次迭代后获得的 Y 估计场和 Y 方差 (σ_Y^2) 估计场

Fig. 3 Maps of Y fields and Y variance (σ_Y^2) estimated through DL_iES and CL_iES at the 20th iteration when $N = 50, 100, 500$

表 2 第 2 组第 20 次迭代后 DL_iES 和 CL_iES 得到的 R_{MSE} 、 S_Y 和 E_h 值

Table 2 Final values of R_{MSE} , S_Y and E_h through DL_iES and CL_iES for Group 2 at the 20th iteration

参数	模型	观测误差		
		0.1	0.01	0.001
R_{MSE}	DL_iES	0.974 4	0.977 9	0.980 1
	CL_iES	0.918 7	0.915 1	0.916 2
S_Y	DL_iES	0.969 1	0.968 9	0.969 0
	CL_iES	0.902 9	0.901 6	0.901 5
E_h	DL_iES	0.272 8	0.111 5	0.078 9
	CL_iES	0.267 7	0.097 8	0.057 1

数反演精度增高(即 R_{MSE} 减小)而水头模拟精度没有出现明显变化。造成这一现象的原因是渗透系数反演过程存在多解性。水头很可能不会对所有位置的渗透系数都敏感。

与图 3 相似, 图 4 绘制第 3 组中 DL_iES 和 CL_iES 第 20 次迭代得到的 Y 和 σ_Y^2 估计场。当 $N_{mon} = 16$ 或 48 时, DL_iES 得到的 Y 和 σ_Y^2 估计场在空间上呈现不

表 3 第 3 组第 20 次迭代后 DL_iES 和 CL_iES 得到的 R_{MSE} 、 S_Y 和 E_h 值

Table 3 Final values of R_{MSE} , S_Y and E_h through DL_iES and CL_iES for Group 3 at the 20th iteration

参数	模型	观测数量		
		16	48	168
R_{MSE}	DL_iES	0.999 9	0.977 9	0.957 5
	CL_iES	0.953 7	0.915 1	0.897 4
S_Y	DL_iES	0.986 2	0.968 9	0.954 3
	CL_iES	0.928 1	0.901 6	0.885 0
E_h	DL_iES	0.102 3	0.111 5	0.109 0
	CL_iES	0.094 8	0.097 8	0.093 2

连续性; 而当 $N_{mon} = 168$ 时, 这种不连续性有所减弱。CL_iES 得到的 Y 和 σ_Y^2 估计场无此现象。随着 N_{mon} 增大, CL_iES 得到的 Y 估计场逐渐逼近 Y 参考场。综合以上结果表明当考虑不同 N_{mon} 时 CL_iES 具有良好自适应性, 且表现优于 DL_iES。

3.4 相关系数高斯白噪音阈值的影响(第 4 组)

第 4 组中 CL_iES 第 20 次迭代得到的 R_{MSE} 、 S_Y 及

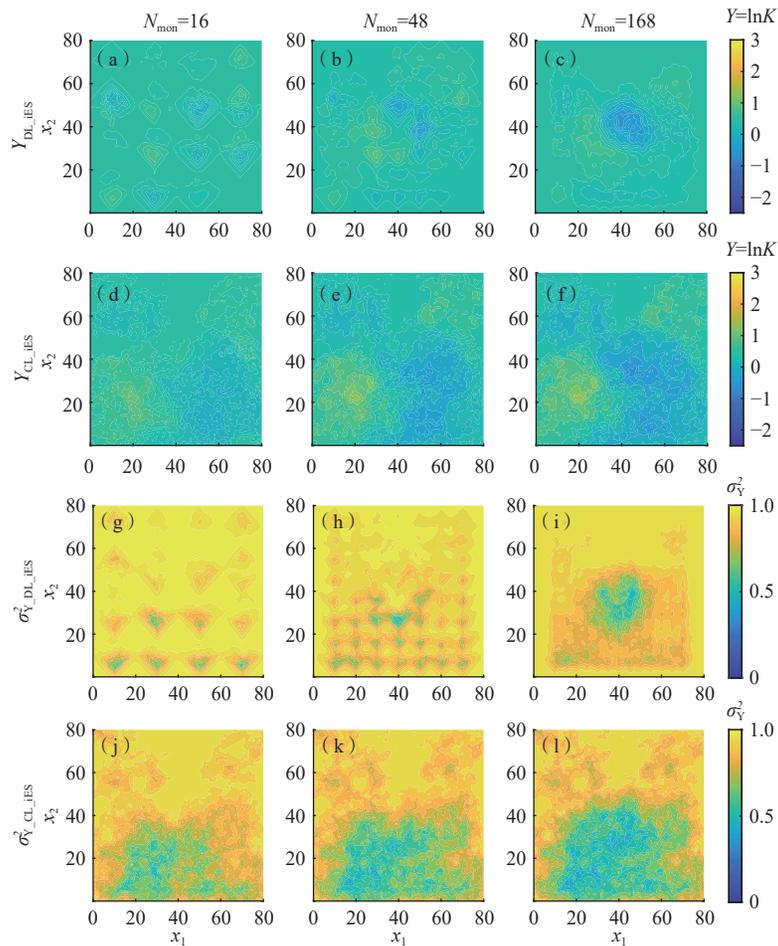


图 4 在 $N_{\text{mon}} = 16, 48, 168$ 时, 通过 DL_iES 和 CL_iES 在第 20 次迭代后获得的 Y 估计场和 Y 方差 (σ_Y^2) 估计场

Fig. 4 Maps of Y fields and Y variance (σ_Y^2) estimated through DL_iES and CL_iES at the 20th iteration when $N_{\text{mon}} = 16, 48, 168$

E_h 值如表 4 所示。随着相关系数白噪音阈值的增大 (即 α 由 1.0 增大到 3.0), R_{MSE} 、 S_Y 及 E_h 值都呈现明显的增大趋势。使用 CL_iES 时, 设置 1 个较大白噪音阈值可以削弱更多相关系数噪音对渗透系数反演的影响, 但同时也增大算法整体阻滞强度, 所以 R_{MSE} 、 S_Y 及 E_h 值都呈现明显的增大趋势。这些结果指示 CL_iES 在应对相关系数高斯白噪音的不同阈值时能做出合理的响应。

表 4 第 4 组第 20 次迭代后 CL_iES 得到的 R_{MSE} 、 S_Y 和 E_h 值
Table 4 Final values of R_{MSE} , S_Y and E_h through CL_iES for Group 4 at the 20th iteration

参数	相关系数高斯白噪音标准差的倍数				
	1.0	1.5	2.0	2.5	3.0
R_{MSE}	0.878 2	0.892 0	0.915 1	0.944 3	0.970 6
S_Y	0.836 7	0.871 5	0.901 6	0.927 3	0.950 0
E_h	0.090 1	0.093 6	0.097 8	0.102 7	0.107 1

4 讨论

DL_iES 和 CL_iES 在应对不同集合大小、观测信

息误差和监测井数目时都表现良好。与 DL_iES 相比, CL_iES 得到渗透系数反演和水头模拟精度都要更高。这是因为 CL_iES 的局域化基于渗透系数与观测水头之间的相关系数, 而 DL_iES 基于渗透系数与观测水头之间的物理距离。CL_iES 在进行局域化的过程中考虑了渗透系数与观测水头之间的相关性。这样可以避免基于物理距离局域化方法 (包括 DL_iES) 的重要不足: 在计算阻滞因子时假定渗透系数与观测水头之间相关关系与物理距离成正比, 且物理距离远近能够完美地反映相关性大小。在计算过程中将最靠近点 (0, 0) 的监测井, 在第 1 个时间步长的观测水头设为第 1 个观测信息。该水头与 Y 的相关系数为 ρ_{k1} ($k = 1, 2, \dots, 6561$)。图 5 展示 $N = 50, 100, 500, 100\ 000$ 时 ρ_{k1} 的绝对值 $|\rho_{k1}|$ 分布和通过 DL_iES 和 CL_iES 获得的对应的阻滞绝对值 $r_{k1}|\rho_{k1}|$ ($k = 1, 2, \dots, 6561$) 的分布。Xia 等^[20] 研究显示当 $N = 10\ 000$ 时伪相关性基本消失。基于此, $N = 100\ 000$ 对应的相关系数场噪音很小, 本研究将其设定为真实的相关系数空间分

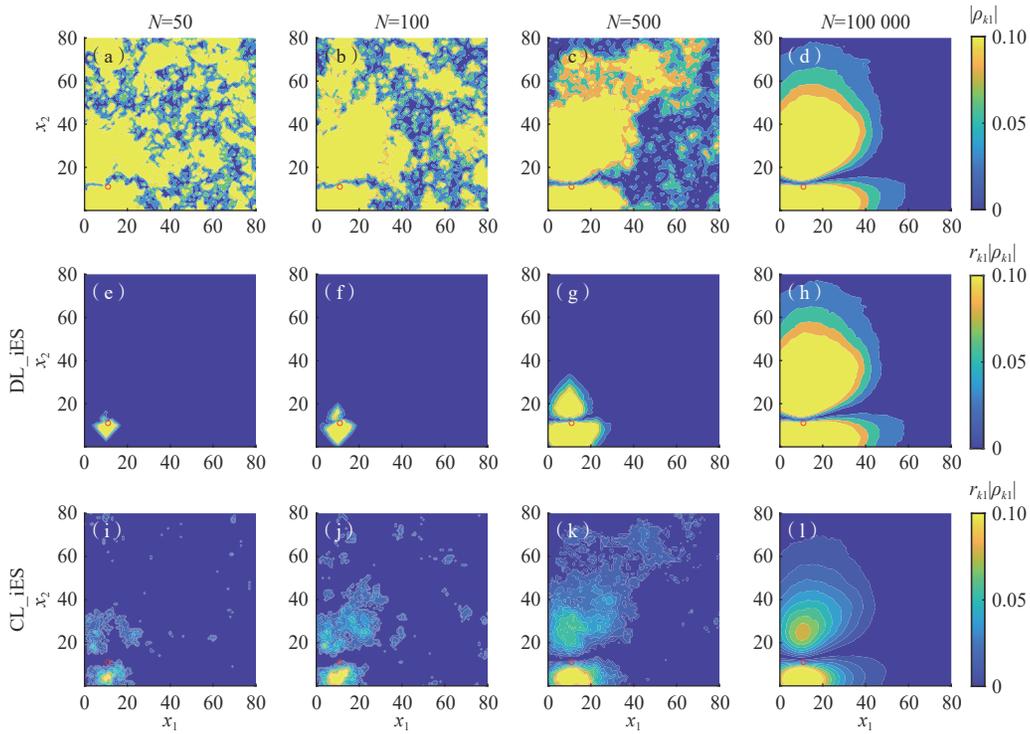


图5 当 $N = 50, 100, 500$ 和 $100\ 000$ 的 $|\rho_{k1}|$ 及 $r_{k1}|\rho_{k1}|$ 分布图

Fig. 5 $|\rho_{k1}|$ and $r_{k1}|\rho_{k1}|$ maps considering $N = 50, 100, 500, 100\ 000$

注:图中红色圆圈代表最靠近点(0,0)的监测井位置为观测水头的固定点。

布。图5(a)—(d)展示:(1) $|\rho_{k1}|$ 在空间上总体越靠近监测井值越大,随着 N 增大到100 000,这种现象更加明显;(2)有相当多的位置距离监测井近但 $|\rho_{k1}|$ 值低,同时也有较多的位置离监测井相对较远但渗透系数与水头之间存在相关性。图5(d)—(f)展示 DL_iES 获得的 $r_{k1}|\rho_{k1}|$ 分布图。随着 N 增大,以监测井为中心 $r_{k1}|\rho_{k1}|$ 不为 0 的范围也增大,当 N 增大到100 000时, $r_{k1}|\rho_{k1}|$ 接近 $|\rho_{k1}|$, 这是合理的。依据式(4),图5(f)对应的 r_{k1} 取值范围为 $[0.712\ 3, 0.992\ 6]$ 。图5(i)—(l)展示 CL_iES 获得的 $r_{k1}|\rho_{k1}|$ 分布图。与 DL_iES 相比, CL_iES 对应的 $r_{k1}|\rho_{k1}|$ 的空间分布相对不规则,但不为 0 的值总体也集中在监测井附近。CL_iES 能够考虑到渗透系数与水头之间物理距离大但又具有相关性的情况,也能考虑物理距离小但相关性微弱甚至没有的情况。基于相关性局域化方法在数学理论上比基于物理距离的局域化方法更合理,数理意义更加清晰。

由图5可知 CL_iES 比较明显的缺陷有 2 个:

(1)当 N 从50增大到500,有些零星的伪相关性难以被去除。

(2)当 $N = 100\ 000$ 时,因为 r_{k1} 值总体不接近于1, $r_{k1}|\rho_{k1}|$ 的空间分布差异明显,见图5(l)与图5(d)。

这两点都指示式(8)并不能完美地刻画 r_{k1} 与 ρ_{k1} 之

间的映射关系。依据 R_{MSE} 与 S_Y 的结果(表1—3),缺陷(1)并没有很大程度削弱 Y 集合内部的差异性和 Y 的反演精度。此外,通过增大白噪音阈值(w),理论上可适当削弱这些零星的伪相关性。缺陷(2)与1.4节的阐述相同,实际应用过程中受计算成本限制,一般只考虑较小的 N 。所以缺陷(2)对 CL_iES 方法的普适性影响较小。

5 结论

本文提出了一种基于相关性的局域化迭代集合平滑方法(记为 CL_iES)。将 CL_iES 与一种基于物理距离的局域化迭代集合平滑(记为 DL_iES)用于反演二维孔隙承压含水层的渗透系数场。分析了不同集合大小、观测误差、观测数量及相关系数白噪音阈值条件下 CL_iES 与 DL_iES 在反演渗透系数场方面的表现。基于理论推演与算例结果,对 CL_iES 方法的优劣进行了分析和讨论。

(1)CL_iES 是一种有效可靠的反演方法。在不同集合大小、水头观测误差和观测数目等条件下, CL_iES 均得到合理的渗透系数反演结果。随着集合增大, CL_iES 得到的渗透系数反演精度逐步提高。

(2)由 CL_iES 反演得到的渗透系数场的精度高

于 DL_iES 得到的结果。在不同集合大小、水头观测误差和观测数目等条件下, CL_iES 得到渗透系数反演和水头模拟精度都要高于 DL_iES。此外, 由 DL_iES 得到渗透系数及其方差在较小集合和观测数目条件下均出现了空间上斑状不连续现象, 而 DL_iES 不存在此现象。

(3) 基于物理距离的局域化迭代集合平滑方法依据参数与观测信息的物理距离定量二者之间的相关程度, 局域化的效果受阻滞函数和阻滞关键长度影响。与基于物理距离的方法相比, 基于相关性的局域化方法依据参数与观测信息之间的相关性定量二者之间的相关程度, 数学意义更清晰, 且不需要二者之间存在物理距离定义。在未来研究中, 值得更加关注利用基于相关性的局域化集合数据同化方法反演水文地质参数。

参考文献 (References) :

- [1] 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)]
- [2] 宗成元, 康学远, 施小清, 等. 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 [J]. 水文地质工程地质, 2020, 47(2): 1–8. [ZONG Chengyuan, KANG Xueyuan, SHI Xiaoqing, et al. Characterization of non-Gaussian hydraulic conductivity fields using multiple-point geostatistics and ensemble smoother with multiple data assimilation method[J]. Hydrogeology & Engineering Geology, 2020, 47(2): 1–8. (in Chinese with English abstract)]
- [3] MO Shaoxing, ZABARAS N, SHI Xiaoqing, et al. Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification[J]. *Water Resources Research*, 2019, 55(5): 3856–3881.
- [4] XIA Chuanan, HU B X, TONG Juxiu, et al. Data assimilation in density-dependent subsurface flows via localized iterative ensemble Kalman filter[J]. *Water Resources Research*, 2018, 54(9): 6259–6281.
- [5] ZHANG Jiangjiang, LIN Guang, LI Weixuan, et al. An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions[J]. *Water Resources Research*, 2018, 54(3): 1716–1733.
- [6] CHEN Zi, GÓMEZ-HERNÁNDEZ J J, XU Teng, et al. Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter[J]. *Journal of Hydrology*, 2018, 564: 1074–1084.
- [7] 李丽敏, 温宗周, 董勋凯, 等. 基于矩阵奇异值分解约束型无迹粒子滤波的滑坡位移预测模型研究 [J]. 水土保持通报, 2019, 39(1): 132–136. [LI Limin, WEN Zongzhou, DONG Xunkai, et al. Landslide displacement prediction model based on singular value decomposition constrained unscented particle filter[J]. *Bulletin of Soil and Water Conservation*, 2019, 39(1): 132–136. (in Chinese with English abstract)]
- [8] 薛长虎. 基于改进粒子滤波的大型滑坡数据同化方法研究 [D]. 武汉: 武汉大学, 2019 [XUE Changhu. Research on data assimilation method of large landslide based on improved particle filter[D]. Wuhan: Wuhan University, 2019. (in Chinese with English abstract)]
- [9] ZHANG Hongqin, TIAN Xiangjun. A multigrid nonlinear least squares four-dimensional variational data assimilation scheme with the advanced research weather research and forecasting model[J]. *Journal of Geophysical Research: Atmospheres*, 2018, 123(10): 5116–5129.
- [10] LUO Xiaodong, BHAKTA T. Automatic and adaptive localization for ensemble-based history matching[J]. *Journal of Petroleum Science and Engineering*, 2020, 184: 106559.
- [11] LUO Xiaodong, LORENTZEN R, VALESTRAND R, et al. Correlation-based adaptive localization for ensemble-based history matching: applied to the norne field case study[Z]. *Spe Norway One Day Seminar*, Norway, 2018.
- [12] BISHOP C H, HODYSS D. Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation[J]. *Monthly Weather Review*, 2011, 139(4): 1241–1255.
- [13] 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.
- [14] TONG Juxiu, HU B X, YANG Jinzhong. Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field[J]. *Stochastic Environmental Research and Risk Assessment*, 2012, 26(3): 467–478.

- [15] NAN Tongchao, WU Jichun. Groundwater parameter estimation using the ensemble Kalman filter with localization[J]. *Hydrogeology Journal*, 2011, 19(3): 547–561.
- [16] 南统超, 吴吉春. 集合卡尔曼滤波估计水文地质参数的局域化修正 [J]. 水科学进展, 2010, 21(5): 613–621. [NAN Tongchao, WU Jichun. Localization corrections for the estimation of hydrogeological parameters using ensemble Kalman filter[J]. *Advances in Water Science*, 2010, 21(5): 613–621. (in Chinese with English abstract)]
- [17] SOARES R V, MASCHIO C, SCHIOZER D J. A novel localization scheme for scalar uncertainties in ensemble-based data assimilation methods[J]. *Journal of Petroleum Exploration and Production Technology*, 2019, 9(4): 2497–2510.
- [18] FURRER R, BENGTTSSON T. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants[J]. *Journal of Multivariate Analysis*, 2007, 98(2): 227–255.
- [19] MIYOSHI T. An adaptive covariance localization method with the LETKF[C]//14th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS). Atlanta: American Meteorological Society, 2010.
- [20] XIA Chuanan, LUO Xiaodong, HU B X, et al. Data assimilation with multiple types of observation boreholes via the ensemble Kalman filter embedded within stochastic moment equations[J]. *Hydrology and Earth System Sciences*, 2021, 25(4): 1689–1709.
- [21] SONG Xuehang, SHI Liangsheng, YE Ming, et al. Numerical comparison of iterative ensemble Kalman filters for unsaturated flow inverse modeling[J]. *Vadose Zone Journal*, 2014, 13(2): 1–12.
- [22] 周研来, 郭生练, 郭家力, 等. VIC 模型参数的地区分布规律及在无资料流域的移用 [J]. 水资源研究, 2012, 1(3): 56–63. [ZHOU Yanlai, GUO Shenglian, GUO Jiali, et al. Regional distribution of the VIC model parameters and application in ungauged basins [J]. *Journal of Water Resources Research*, 2012, 1(3): 56–63 (in Chinese with English Abstract)]
- [23] 石鸿蕾, 郝奇琛, 邵景力, 等. 基于多源数据的弱透土层水文地质参数反演研究——以呼和浩特盆地某淤泥层为例 [J]. 水文地质工程地质, 2021, 48(2): 1–7. [SHI Honglei, HAO Qichen, SHAO Jingli, et al. Research on hydrogeological parameter inversion of an aquitard based on multi-source data: A case study of a silt layer in the Hohhot Basin[J]. *Hydrogeology & Engineering Geology*, 2021, 48(2): 1–7. (in Chinese with English abstract)]
- [24] BEAR J. *Hydraulics of groundwater*[M]. New York: McGraw-Hill Book Co, 1979.
- [25] LUO Xiaodong, STORDAL A S, LORENTZEN R J, et al. Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory and applications[J]. *SPE Journal*, 2015, 20(5): 962–982.
- [26] ZHANG Hongqin, TIAN Xiangjun. An efficient local correlation matrix decomposition approach for the localization implementation of ensemble-based assimilation methods[J]. *Journal of Geophysical Research: Atmospheres*, 2018, 123(7): 3556–3573.
- [27] ANDERSON T W. *An introduction to multivariate statistical analysis*[M]. 3rd ed. Hoboken: Wiley-Interscience, 2003.

编辑: 张若琳