

DOI:10.13476/j.cnki.nsbdtk.2022.0048

周念清,张瑞城,江思珉,等. ES-MDA 算法融合 ERT 数据联合反演地下水污染源与含水层参数[J]. 南水北调与水利科技(中英文),2022,20(3):478-486. ZHOU N Q,ZHANG R C,JIANG S M,et al. Joint inversion of contaminant source and aquifer parameters by assimilating ERT data with the ES-MDA algorithm[J]. South-to-North Water Transfers and Water Science & Technology,2022,20(3):478-486. (in Chinese)

ES-MDA 算法融合 ERT 数据联合反演 地下水污染源与含水层参数

周念清¹,张瑞城¹,江思珉¹,夏学敏²

(1. 同济大学土木工程学院,上海 200092;2. 上海理工大学环境与建筑学院,上海 200093)

摘要:针对未知的污染场地,为了准确估计污染物运移模型的参数,提出一种基于多重数据同化集合平滑器(ensemble smoother with multiple data assimilation,ES-MDA)算法的地下水模型参数反演方法,通过融合由高密度电阻率(electrical resistance tomography,ERT)法采集的 ERT 观测数据,实现对污染源强度和渗透系数场的联合反演。以此为基础设计 3 组数值算例,比较不同类型观测数据对反演精度的影响。研究表明:融合 ERT 数据的 ES-MDA 算法对模型参数的反演精度更高,并且将 ERT 数据和传统的质量浓度与水头观测数据相结合,能进一步优化反演结果。

关键词:数据同化;集合平滑;地球物理;高密度电阻率法;渗透系数场

中图分类号:TV213;X523

文献标志码:A

开放科学(资源服务)标志码(OSID):



地下水资源由于具有分布广泛、水质良好、不易受污染等特点,在人类日常生活和工农业生产中被广泛利用。然而,随着城市化进程的加快和现代工业的迅速发展,越来越多的污染物因泄漏或排放不当而进入地下水体中,导致地下水污染问题日趋严重。为了有效地治理和修复地下水污染,首先需要对污染源信息(包括污染源强度和位置等)进行准确识别。传统的地下水监测采样方法由于成本较高且获取的观测数据较少,往往难以直接获取污染源信息以及场地的水文地质参数^[1-2]。现阶段的研究多通过构建地下水数值模拟模型并用求解逆问题的方式对污染源进行参数反演^[3-4]。

参数反演法是通过分析多源观测信息,并与已知的参照值进行比较,不断迭代更新模型参数,使正演模型输出的观测值与参照值不断趋近,从而得到

未知模型参数的近似估计,将其应用于地下水污染溯源问题中,可实现对未知污染源参数的模拟预测^[5]。近年来,贝叶斯推断已被广泛地应用于地下水参数反演问题之中^[6-8]。其中,基于马尔科夫链蒙特卡罗(Markov chain Monte Carlo,MCMC)方法和基于集合的数据同化方法(ensemble-based data assimilation)是贝叶斯推断的两种常见应用^[9-11]。MCMC 方法通过构造合适的马尔科夫链进行抽样并使用蒙特卡洛方法进行积分计算,最终收敛到平稳分布的参数后验分布,但使用 MCMC 方法通常需要大量调用系统模型才能达到收敛,并且模型的调用次数将随着参数维度的增加而大幅增加,给计算带来沉重的负担^[12]。基于集合的数据同化算法在求解参数反演问题时,需要调用系统模型的次数相对较少,因而比 MCMC 算法的计算

收稿日期:2022-01-28 修回日期:2022-04-06 网络出版时间:2022-04-14

网络出版地址:https://kns.cnki.net/kcms/detail/13.1430.TV.20220413.1136.004.html

基金项目:国家自然科学基金项目(42077176)

作者简介:周念清(1964—),男(土家族),湖南石门人,教授,博士生导师,主要从事水资源调度与优化管理、地下水资源评价、地下水污染治理、水生态建设等方面研究。E-mail:nq.zhou@tongji.edu.cn

效率更高,其中集合卡尔曼滤波^[13](ensemble Kalman filter, EnKF)和集合平滑器^[14](ensemble smoother, ES)是其典型代表,二者的区别在于 EnKF 需要同时更新输入参数和状态变量,而 ES 仅进行参数更新,因此可以有效避免 EnKF 中参数和状态的不一致问题^[15]。为了提高 ES 方法的计算效率和对高维问题的适用性,Emerick 等^[16]提出了一种多重数据同化集合平滑器(ensemble smoother with multiple data assimilation, ES-MDA),利用添加扰动后的观测误差协方差矩阵对观测值进行多次数据同化,提高了算法在求解高维参数反演问题时的性能。

在利用数值模型对地下水污染物迁移转化过程进行模拟预测的过程中,模型参数的准确性是影响预测精度的关键。然而,由于模型参数存在空间变异性特点,对于非均质性很大的污染场地,难以利用稀疏的观测数据来准确反演估计污染场地的模型参数^[17]。为了解决传统的勘测方法成本较高且获取的观测数据较为稀疏的问题,近年来有学者^[18-19]将地球物理方法(如高密度电阻率法、地质雷达法等)引入到污染场地调查中,能够以较低的成本获取大量的连续观测数据,有效提高了参数反演方法在应对高维参数空间时的适用性。

本文提出一种基于 ES-MDA 算法求解地下水污染溯源问题的数据同化方法,并融合由高密度电阻率(electrical resistivity tomography, ERT)法采集的 ERT 观测数据进行参数反演,然后将该反演结果与传统的以浓度和水头为观测值方法而非采用地球物理方法得到的结果进行比较,验证该方法在求解高维参数反演问题时的有效性。

1 研究方法

1.1 地下水污染迁移模型

利用 MODFLOW 程序^[20]与 MT3DMS 程序^[21]构建地下水污染迁移的数值模拟模型,其中,MODFLOW 程序可以求解饱和和多孔介质中的地下水流问题,用基本微分方程表示为

$$\frac{\partial}{\partial x} \left(K_x \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_y \frac{\partial H}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial H}{\partial z} \right) + W = S_s \frac{\partial H}{\partial t} \quad (1)$$

式中: K_x 、 K_y 、 K_z 分别为渗透系数在 X、Y、Z 方向上的分量, m/d; H 为水头, m; W 为单位时间从单位体积含水层中流入或流出的水量, d^{-1} ; S_s 为贮水率, m^{-1} ; t 为时间, d。

在查明了地下水流动信息的基础上,再结合 MT3DMS 程序,对地下水污染物的溶质运移过程进行模拟,其基本控制方程为

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x_i} \left(D_{ij} \frac{\partial C}{\partial x_j} \right) - \frac{\partial}{\partial x_i} (v_i C) + q_s C_s + \sum R_n \quad (2)$$

式中: C 为质量浓度, mg/L; t 为时间, d; D_{ij} 为水力弥散系数, m^2/d ; v_i 为含水介质中的实际水流速度, m/d; q_s 为单位体积含水层的源汇项流量, d^{-1} ; C_s 为源汇项中溶质质量浓度, mg/L; $\sum R_n$ 为化学反应项总和, mg/(L·d)。

1.2 多重数据同化集合平滑器(ES-MDA)

ES-MDA 算法是地下水污染溯源问题中一种常用的数据同化方法,其基本原理与 ES 算法类似,区别在于 ES-MDA 算法在运算过程中利用观测误差的协方差矩阵对观测值进行多次数据同化,实现了对参数样本的迭代更新,以更好地应对高维参数空间反演问题^[22-23]。ES-MDA 算法的基本步骤如下:

第一步确定数据同化的迭代次数(N_a)和每次迭代对应的膨胀系数 a_i (其中, $i=1, 2, \dots, N_a$)。

第二步从模型参数的先验分布中选取 N_e 个样本,组成初始样本集合, $M_1 = [m_{1,1}, m_{2,1}, \dots, m_{N_e,1}]$ 。

第三步从 $i=1$ 开始执行多次迭代计算,在第 i 次迭代中,对样本集合 $M_i = [m_{1,i}, m_{2,i}, \dots, m_{N_e,i}]$ 中的每一个样本运行正演模型以获取该样本对应的预测值 $d_{j,i}$, 正演模型的表达式如下:

$$d_{j,i} = f(m_{j,i}), j=1, 2, \dots, N_e \quad (3)$$

第四步根据公式(4)对样本集合进行更新:

$$m_{j,i+1} = m_{j,i} + C_{MD} (C_{DD} + a_i C_D)^{-1} (d_{obs_j} - d_{j,i}), j=1, 2, \dots, N_e \quad (4)$$

式中: C_{MD} 为模型参数 m_i 与预测值 d_i 之间的协方差矩阵; C_{DD} 为预测值的自协方差矩阵; C_D 为观测误差的协方差矩阵; d_{obs_j} 为根据 $a_i C_D$ 添加扰动后的观测值。

第五步重复执行第 3 步和第 4 步直到完成第 N_a 次迭代,得到最终的参数样本集合 $M_{N_a} = [m_{1,N_a}, m_{2,N_a}, \dots, m_{N_e,N_a}]$,并从中获得对参数后验的反演估计。

1.3 耦合地下水-地球物理模型的数据同化方法

根据岩石地球物理关系构建联结地球物理信息与水文地质信息的数值模型,利用 Archie 公式^[24]描述地层电阻率、溶液电阻率与含水饱和度之间关系,其公式基本形式如下:

$$\sigma_t = \frac{\sigma_w}{\varphi^m S_w^n} \quad (5)$$

式中： σ_t 为地层总电阻率， $\Omega \cdot m$ ； σ_w 为溶液电阻率， $\Omega \cdot m$ ； φ 为有效孔隙度； S_w 为含水饱和度； m 为黏结指数； n 为饱和度指数。

溶液电阻率与温度和溶液质量浓度等因素有关，Sen^[17]综合考虑了这 2 种因素对溶液电阻率的影响，提出了如下模型公式：

$$\frac{1}{\sigma_w} = (5.6 + 0.27T - 1.5 \times 10^{-4} T^2) C - (2.36 + 0.099T) C^{\frac{3}{2}} / (1.0 + 0.214C) \quad (6)$$

式中： C 为溶液的质量浓度，mg/L； T 为溶液温度，文中假定恒为 25 °C。将式(5)与式(6)联合，便可构建地层电阻率与溶液质量浓度的数值关系。

高密度电阻率法(ERT 法)作为一种常用的地

球物理方法，可以通过供电电极(A、B)向地下空间输入稳定的电流 I ，然后根据测量电极(M、N)处的电位值 V 得到视电阻率的观测数据 ρ_a ，再据此推算出电阻率的空间分布。利用开源的 Pygimli 程序^[25]构建地球物理模型并求解 ERT 正演问题。ERT 法的正演模型可表示为

$$\nabla \cdot \left(\frac{1}{\rho} \nabla V \right) = -I \delta(r) \quad (7)$$

式中： ρ 为地层电阻率的空间分布， $\Omega \cdot m$ ； V 为电位值，V； I 为电流值，A； $\delta(r)$ 为狄拉克 δ 函数。

以上述方法为基础，提出基于 ES-MDA 算法融合地球物理数据的数据同化方法，通过耦合地下水-地球物理模型，将质量浓度数据转换为 ERT 观测值，然后利用 ES-MDA 算法融合 ERT 观测数据对未知污染源参数进行反演估计。该方法的流程图见图 1。

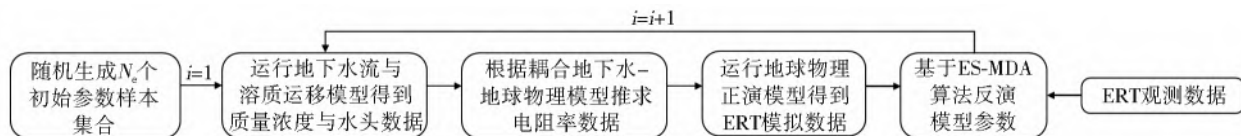


图 1 基于 ES-MDA 算法的数据同化方法流程

Fig. 1 The flowchart of the data assimilation method based on ES-MDA algorithm

2 数值算例

假定污染场地为非均质各向异性的承压含水层，将其概化为二维 XZ 剖面(100 m×40 m)，用边长为 2 m×2 m 的正方形网格将区域剖分为 20 行 50 列有限差分网格，见图 2。假设含水层水流运动为稳定流，左边界为定水头边界(水头为 50 m)，右边界为定水头边界(水头为 49 m)，下边界为隔水边

界，上边界无源汇项。含水介质孔隙度为 0.30，纵向弥散度为 20 m，垂直横向弥散度为 3 m。初始时刻，研究区无污染物。含水层的渗透系数 K 符合对数正态分布，参照场对数均值($\ln K$)为 3.0，方差($\sigma_{\ln K}^2$)为 1.0，X、Z 方向的相关长度分别为 60 m 和 24 m。模拟总时长为 200 d，并平均划分为 10 个应力期，每个应力期为 20 d。

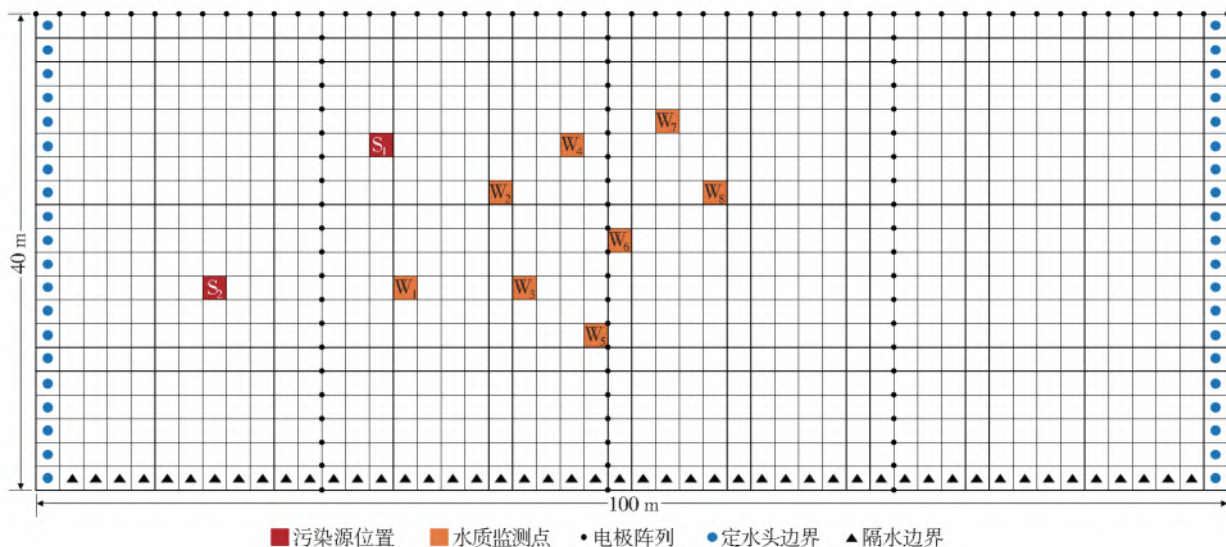


图 2 污染场地示意图

Fig. 2 The sketch map of contaminated site

如图 2 所示,污染场地存在 2 个点源污染,污染物成分相同,根据场地调查情况确定位置为 S_1 和 S_2 ,假定在模拟时间内其质量浓度保持不变,点源 S_1 处的污染物质量浓度为 60 mg/L,点源 S_2 处的污染物质量浓度为 30 mg/L。场地内共布设 8 个水文监测点(W_1 至 W_8),用于测量质量浓度与水头数据,并在模拟期 t 为 40、60、80、100、120、140、160 d 时进行数据采集,共获得 8 个固定水头数据与 56 个质量浓度数据。同时,在地表布设 1 条水平测线(51 个电极,间距 2 m)与 3 条垂直方向的测线(位于 $x=24、48、72$ m 处),每条测线包含 20 个间距为 1 m 的电极,共布设 111 个电极,并采用 dipole-dipole 的方式在相同的时间点测量电位数据(共有 7 次电流注入,电流为 1 A),共获得 763 个 ERT 观测数据。考虑到观测误差对参数反演结果的影响,设置质量浓度与水头观测值相对误差为 2%,ERT 观测值相对误差为 5%。

针对不同类型的观测数据,设计了 3 组算例来对比其参数反演效果。3 组算例均采用 ES-MDA 算法作为数据同化方法:算例 1 使用质量浓度与水头数据作为观测数据;算例 2 仅使用 ERT 数据作为观测数据;算例 3 使用质量浓度、水头和 ERT 数据作为观测数据。3 组算例观测数据的采集时间相同,具体设置见表 1。

表 1 算例设置

Tab. 1 The setting of case studies

算例编号	观测数据	数据采集时间/d
算例 1	质量浓度、水头	40、60、80、100、120、140、160
算例 2	ERT	40、60、80、100、120、140、160
算例 3	质量浓度、水头、ERT	40、60、80、100、120、140、160

考虑到含水层的空间各向异性,在对渗透系数场进行参数反演时计算代价往往较高。本算例中的渗透系数场根据模型的网格剖分维度为 1 000 维,属于维度较高的反演问题,如果直接进行求解效率会很低。因此,采用 Karhunen-Loeve 展开^[6](KL 展开)的方法对渗透系数场进行降维。

$$\ln K(x, y) \approx \overline{\ln K}(x, y) + \sum_{i=1}^{N_{KL}} \zeta_i \sqrt{\lambda_i} s_i(x, y) \quad (8)$$

式中: ζ_i 为独立标准高斯分布的随机数; λ_i 和 $s_i(x, y)$ 为特征值和特征向量; N_{KL} 为 KL 展开保留的项数,本算例中取 $N_{KL} = 60$,可以保留对数渗透系数场 95% 以上的变异性。

利用 ES-MDA 算法进行参数反演的估计精度可以使用均方根误差(E_{RMS})来量化,该值反映了模型参数的估计值与真实值之间的差距, E_{RMS} 的值越

小(趋近于 0),参数反演的精度越高。 E_{RMS} 的计算公式为

$$E_{RMS} = \sqrt{\frac{1}{N_e} \sum_{n=1}^{N_e} (\hat{y}_n - y_n)^2} \quad (9)$$

式中: \hat{y}_n 为反演的模型参数估计值; y_n 为模型参数真实值; N_e 为 ES-MDA 算法中样本集合的大小。

3 结果与讨论

在 3 组算例中,潜在污染点源 S_1 和 S_2 的质量浓度在模拟期内保持不变,共有 2 个未知质量浓度参数需要识别。渗透系数场采用 KL 展开的方式对其降维,其中 KL 展开项数 $N_{KL} = 60$,因此对高维参数场(1 000 维)的识别降维成对 60 个高斯分布的 KL 展开项的反演。由此得出,对于 3 组算例,总共需要进行数据同化的未知模型参数数量均为 62 个。根据以往设计数值算例的经验,将 ES-MDA 算法的基础参数设置为:样本集合数 $N_e = 1\ 000$;迭代次数 $N_a = 7$;膨胀系数 $a_i = 14、12、10、8、6、4、2$ (其中, $i = 1, 2, \dots, N_a$)。

3.1 污染源源强的反演识别

图 3(a)至 3(c)分别为 3 组算例对污染源源强的反演结果。图 3(a)是将质量浓度与水头数据作为观测值得到的源强反演结果,经过 ES-MDA 算法 7 次迭代后,污染源 S_1 处的源强已基本收敛,污染源 S_2 处的源强与真值仍有较大偏差,其主要原因是观测井数量有限,导致采集到的观测数据较少。图 3(b)是将 ERT 数据作为观测值得到的源强反演结果, S_1 与 S_2 处的源强参数均已基本收敛,且与算例 1 相比,反演结果更加趋近于真值,说明利用地球物理方法获取的大量 ERT 观测数据能有效改善传统观测方法获取数据稀疏的不足。图 3(c)是将质量浓度与水头数据和 ERT 数据结合,共同作为观测值得到的源强反演结果,经过 7 次迭代后 S_1 与 S_2 处的源强参数明显收敛,且其与真值的拟合程度均优于算例 1 和算例 2。

由此可得,融合 ERT 数据的 ES-MDA 算法可以更加精确地反演污染源源强,并且在此基础上添加质量浓度与水头观测数据,将使反演结果得到进一步优化,说明观测信息的数量会直接影响参数反演的效果。利用地球物理方法获取的 ERT 数据虽然测量精度不如质量浓度与水头数据,但可以便捷地获取大量观测信息,本案例中采集到的 ERT 观测数据量(763 个)远多于算例 1 中采集到的质量浓度与水头数据量(64 个)。并且,在对污染源源强的

反演识别上,通过融合多源观测数据能够有效提升参数反演的精度。然后,通过量化的方式进一步比较使用不同类型的观测数据反演污染源源强的效果,将 3 组算例污染源源强反演结果的 E_{RMS} 值列于

表 2 中。由表 2 可知,3 组算例是针对任一个污染源的 E_{RMS} 值均呈现逐渐缩小的趋势,证明了融合多源观测数据的算例 3 对污染源强度的反演精度要优于算例 1 和算例 2。

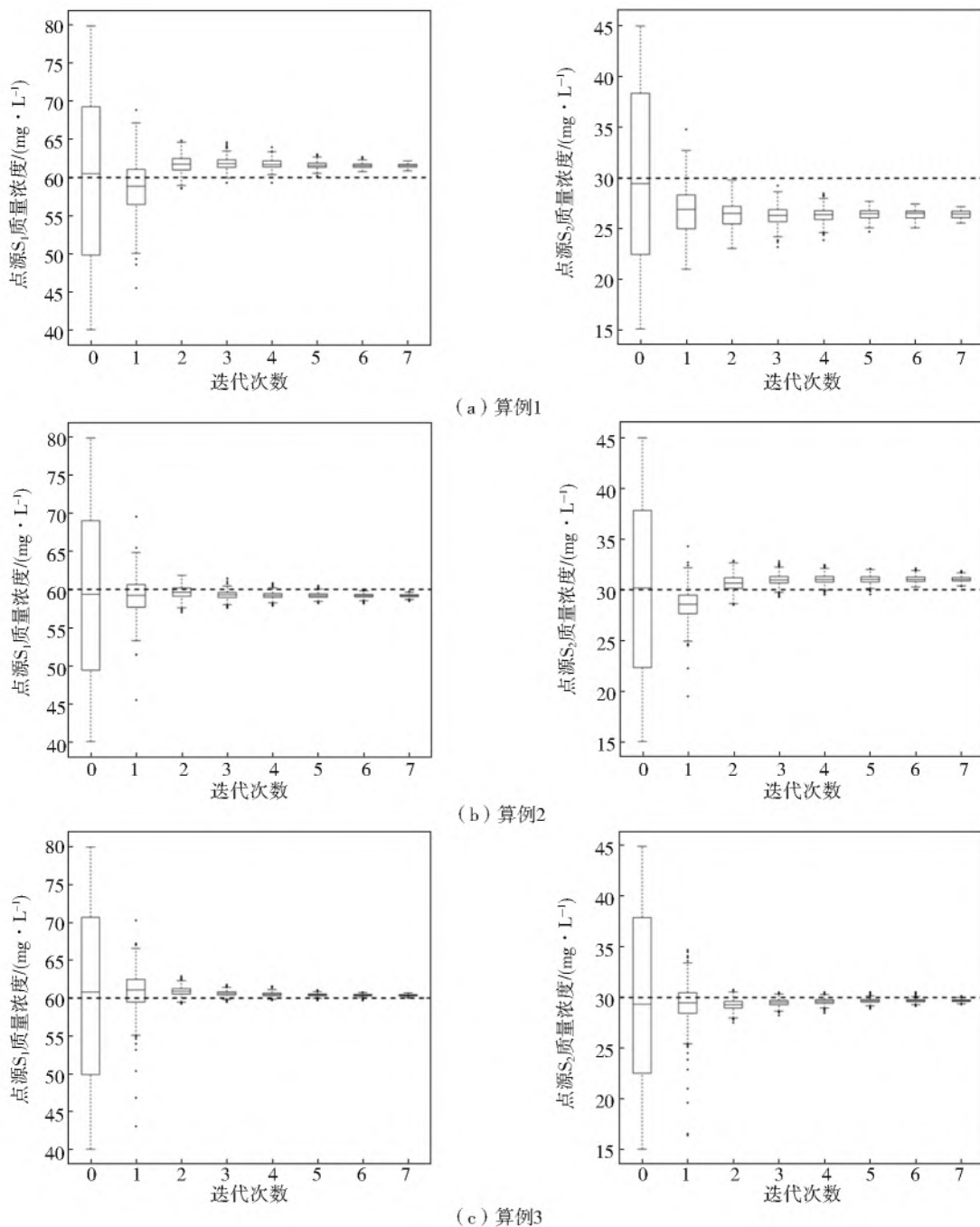


图 3 污染源源强的反演结果

Fig. 3 The inversion results of contaminant source strength

表 2 污染源参数反演结果的均方根误差

Tab. 2 Root-mean-square error of Case 1, Case 2 and Case 3

污染源	S ₁			S ₂		
	算例 1	算例 2	算例 3	算例 1	算例 2	算例 3
E_{RMS}	1.572 0	0.874 2	0.009 5	3.623 0	1.072 3	0.008 9

3.2 渗透系数场的反演识别

图 4、图 5 和图 6 分别为 3 组算例对渗透系数

场的反演结果。作为参照场的 $\ln K$ 真实分布如图 4(a)、图 5(a)和图 6(a);在反演开始阶段生成的初始随机场如图 4(b)、图 5(b)和图 6(b),为了对比 3 组算例的反演效果,采用相同的初始随机场。对渗透系数场进行参数反演得到的后验均值场如图 4(c)、图 5(c)和图 6(c);反映集合离散程度的方差场如图 4(d)、图 5(d)和图 6(d)。

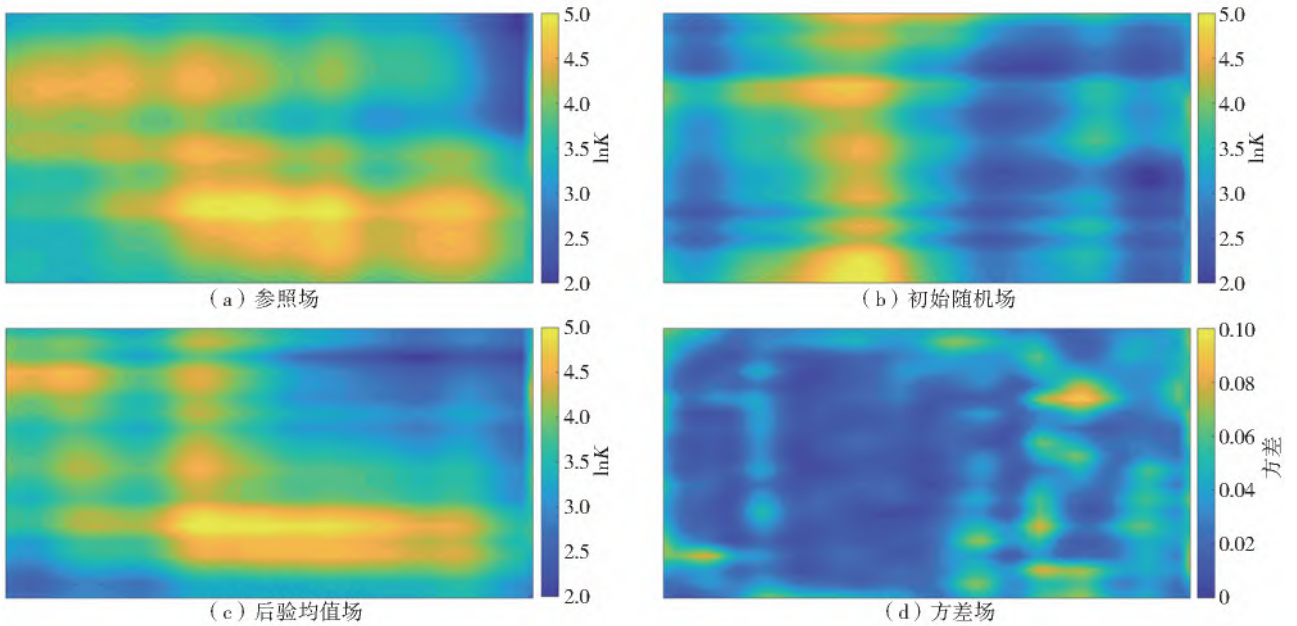


图 4 算例 1 渗透系数场的反演结果

Fig. 4 The inversion results of hydraulic conductivity in Case 1

如图 4 所示,在使用质量浓度与水头数据作为观测值的算例 1 中,其后验均值场基本反映出 $\ln K$ 场的高、低值分布情况,但是对 $\ln K$ 场主体形态的描述还不够精细,其反演结果与参照场仍存在不小的差距。在使用 ERT 数据作为观测值的算例 2(图 5)中,其后验均值场不仅能反映出 $\ln K$ 场的高、低值分布区域,还能较好地刻画渗透系数场的主体形态,因此,算例 2 与参照场的拟合程度要优于算例

1,但是在对渗透系数场分布细节的刻画上仍然不够精细。由图 5(d)可以得出算例 2 的估计方差已趋近于 0,说明继续进行迭代反演对结果的提升空间有限。在算例 3 中,结合 ERT 数据和质量浓度与水头数据作为观测值,得出的结果见图 6。从图 6 可以明显地看出后验均值场在主体形态与细节上均较为精细地刻画出了 $\ln K$ 场的空间分布情况,其与参照场的拟合程度要明显优于算例 1 和算例 2。

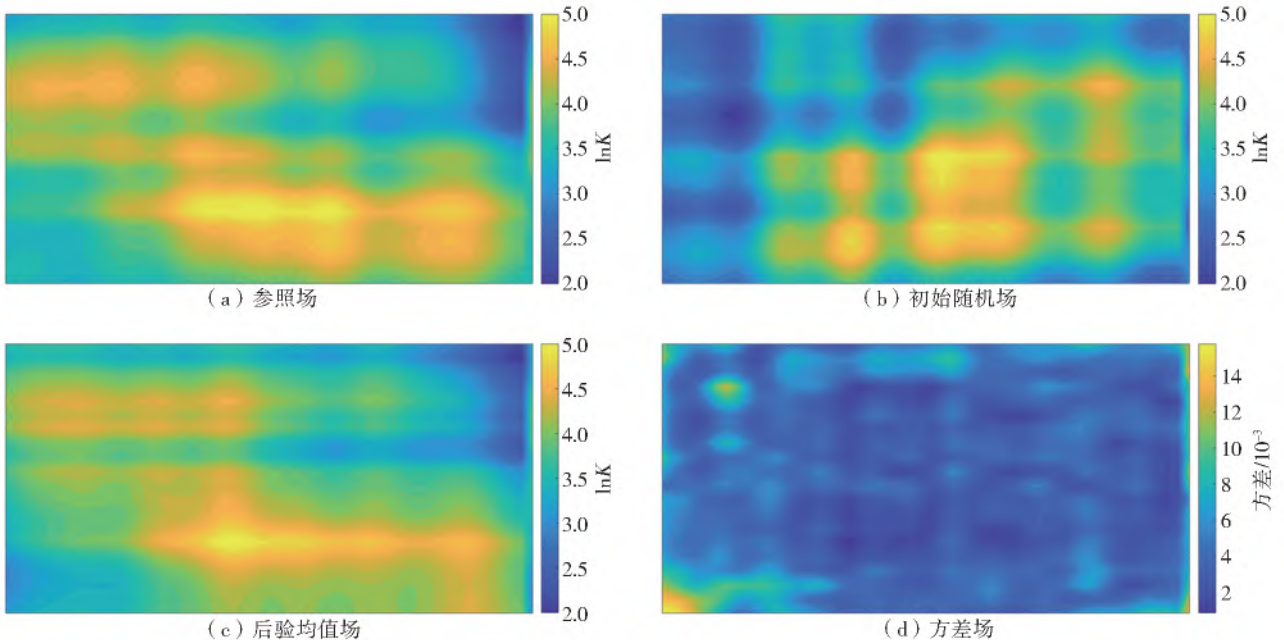


图 5 算例 2 渗透系数场的反演结果

Fig. 5 The inversion results of hydraulic conductivity in Case 2

通过上述对比分析,在对高维渗透系数场反演识别上,融合 ERT 数据的 ES-MDA 算法可以得到更加精确的结果,并且在此基础上融合质量浓度与

水头数据作为观测值,可以有效地提升渗透系数场的反演精度,说明利用多源观测数据在求解参数反演问题上具有明显的优势。

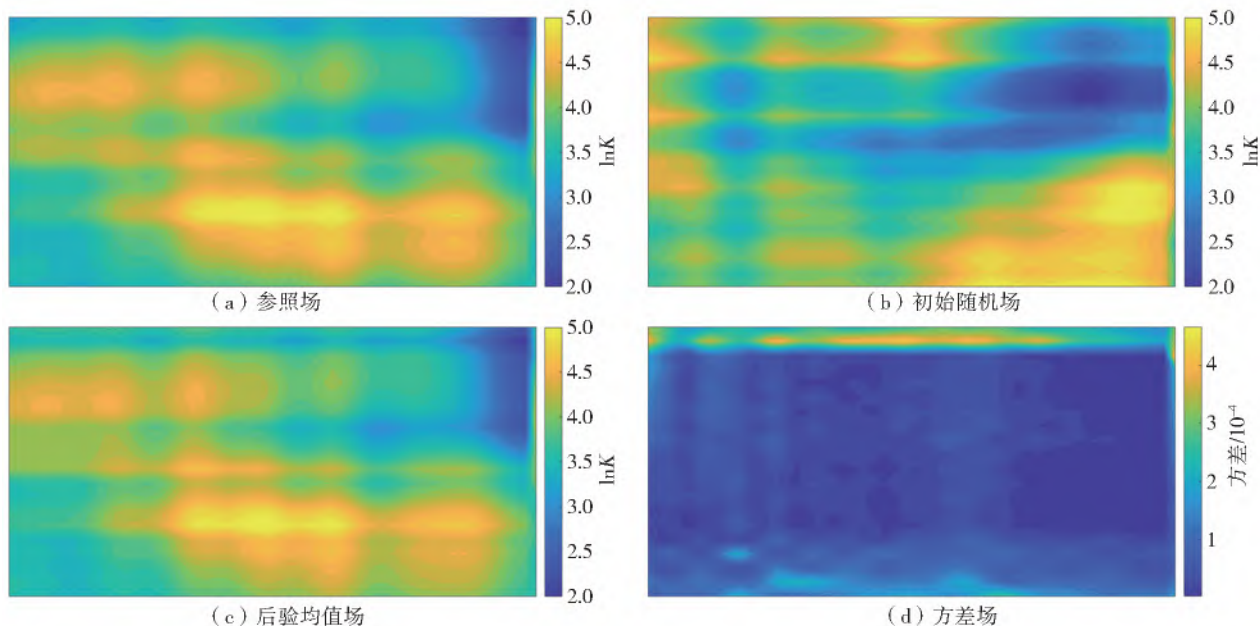


图 6 算例 3 渗透系数场的反演结果

Fig. 6 The inversion results of hydraulic conductivity in Case 3

4 结论与展望

提出了基于 ES-MDA 算法融合 ERT 数据的地下水参数反演方法,并通过数值算例研究,得出以下结论:

利用 ES-MDA 算法融合质量浓度与水头数据或融合 ERT 数据,均可以实现对污染源强度和渗透系数场的联合反演,并且随着迭代次数的增加,反演结果将在一定程度上更加趋近于真实值,证明了该数据同化算法在求解地下水参数反演问题上的有效性。

通过 3 组数值算例来比较融合不同类型观测数据的 ES-MDA 算法的参数反演效果。算例 1 至算例 3 在反演污染源源强时,对任一污染源的 E_{RMS} 值逐渐递减,说明反演精度从优到劣依次为算例 3、算例 2、算例 1。算例 1 至算例 3 在反演渗透系数场时融合了多源观测数据(质量浓度、水头与 ERT 数据)的算例 3 对于 $\ln K$ 场主体形态和局部细节的表征要明显优于算例 1 和算例 2;算例 2 对 $\ln K$ 场主体形态的表征优于算例 1,但对 $\ln K$ 场细节的刻画程度与算例 3 仍存在差距:证明了融合多源观测数据的 ES-MDA 算法在求解参数反演问题上的优越性,能够有效提升参数反演的精度。

算例 2 中的反演结果要明显地优于算例 1,其主要原因是算例 2 中的 ERT 观测数据的数量远多于算例 1 中的质量浓度与水头数据,即使精度略低也能够为参数反演提供更多的信息。高密度电阻率法能够快速而便捷地获取大量具有空间连续性的

ERT 观测数据,而传统观测方法只能在有限的观测井处获取少量不连续的观测数据。因此,在实际场地调查中 ERT 方法将更具优势。后续研究考虑将数据同化方法与其他地球物理方法相结合,探索更为精确高效的算法。

ES-MDA 算法仅适用于对符合高斯分布的模型参数的反演求解,而对于非均质性更强、参数维度更高的非高斯场景,可以考虑结合机器学习、深度学习和多点地质统计方法,进一步探求算法的适用性和改进策略。

参考文献(References):

- [1] 康学远,施小清,邓亚平. 基于 EnKF 融合地球物理数据刻画含水层非均质性[J]. 水科学进展,2018,29(1): 40-49. (KANG X Y, SHI X Q, DENG Y P, et al. Assimilation of hydrogeophysical data for the characterization of subsurface heterogeneity using Ensemble Kalman Filter(EnKF)[J]. Advances in Water Science, 2018,29(1): 40-49. (in Chinese)) DOI: 10.14042/j.cnki.32.1309.2018.01.005.
- [2] 郑哪,江思珉,刘金炳. 基于自组织映射(SOM)算法的地下水污染源反演[J]. 安全与环境工程,2021,28(3): 220-227. (ZHENG N, JIANG S M, LIU J B, et al. Groundwater contaminant source identification based on self-organizing map algorithm[J]. Safety and Environmental Engineering, 2021,28(3): 220-227. (in Chinese)) DOI: 10.13578/j.cnki.issn.1671-1556.20201196.
- [3] LI L, PUZEL R, DAVIS A. Data assimilation in groundwater modelling: ensemble Kalman filter versus ensemble

- ble smoothers [J]. *Hydrological Processes*, 2018, 32 (13): 2020-2029. DOI: 10. 1002/hyp. 13127.
- [4] 张双圣,刘汉湖,强静,等. 基于贝叶斯公式的地下水污染源及含水层参数同步反演[J]. *中国环境科学*, 2019, 39(7): 2902-2912. (ZHANG S S, LIU H H, QIANG J, et al. Synchronous inversion of groundwater pollution source and aquifer parameters based on Bayesian formula[J]. *China Environmental Science*, 2019, 39(7): 2902-2912. (in Chinese)) DOI: 10. 19674/j. cnki. issn1000-6923. 2019. 0343.
- [5] 郝辰宇. 基于自适应替代模型的地下水数据同化研究[D]. 杭州: 浙江大学, 2021. (HAO C Y. An adaptive surrogate-based method for groundwater data assimilation[D]. Hangzhou: Zhejiang University, 2021. (in Chinese)) DOI: 10. 27461/d. cnki. gjzdx. 2021. 000628.
- [6] ZHOU J, SU X, CUI G. An adaptive Kriging surrogate method for efficient joint estimation of hydraulic and biochemical parameters in reactive transport modeling [J]. *Journal of Contaminant Hydrology*, 2018, 216: 50-57. DOI: 10. 1016/j. jconhyd. 2018. 08. 005.
- [7] CHANDRA R, AZAM D, MÜLLER R D, et al. Bayeslands: A Bayesian inference approach for parameter uncertainty quantification in Badlands [J]. *Computers & Geosciences*, 2019, 131: 89-101. DOI: 10. 1016/j. cageo. 2019. 06. 012.
- [8] EMERY J M, GRIGORIU M D, FIELD JR. R V. Bayesian methods for characterizing unknown parameters of material models [J]. *Applied Mathematical Modelling*, 2016, 40(13-14): 6395-6411. DOI: 10. 1016/j. apm. 2016. 01. 046.
- [9] 莫绍星. 基于深度学习的地下水模拟高维不确定性分析和反演[D]. 南京: 南京大学, 2019. (MO S X. Towards efficient high-dimensional uncertainty quantification and inverse analysis in groundwater modeling using deep learning[D]. Nanjing: Nanjing University, 2019. (in Chinese)) DOI: 10. 27235/d. cnki. gnjiu. 2019. 000108.
- [10] JU L, ZHANG J, MENG L, et al. An adaptive Gaussian process-based iterative ensemble smoother for data assimilation [J]. *Advances in Water Resources*, 2018, 115: 125-135. DOI: 10. 1016/j. advwatres. 2018. 03. 010.
- [11] XING Z, QU R, ZHAO Y, et al. Identifying the release history of a groundwater contaminant source based on an ensemble surrogate model [J]. *Journal of Hydrology*, 2019, 572: 501-516. DOI: 10. 1016/j. jhydrol. 2019. 03. 020.
- [12] ZHANG R, ZHOU N, XIA X, et al. Joint estimation of hydraulic and biochemical parameters for reactive transport modelling with a modified ILUES algorithm [J]. *Water*. 2020, 12 (8): 2161. DOI: 10. 3390/w12082161.
- [13] EVENSEN G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics [J]. *Journal of Geophysical Research: Oceans*, 1994, 99 (C5): 10143-10162. DOI: 10. 1029/94JC00572.
- [14] VAN LEEUWEN PJ, EVENSEN G. Data assimilation and inverse methods in terms of a probabilistic formulation [J]. *Monthly Weather Review*, 1996, 124 (12): 2898-2913. DOI: 10. 1175/1520-0493(1996)124 <2898: DAAIMI>2. 0. CO; 2.
- [15] ZHANG J, LIN G, LI W, 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. DOI: 10. 1002/2017WR020906.
- [16] EMERICK A A, REYNOLDS A C. Ensemble smoother with multiple data assimilation [J]. *Computers & Geosciences*, 2013, 55: 3-15. DOI: 10. 1016/j. cageo. 2012. 03. 011.
- [17] KANG X, SHI X, REVIL A, et al. Coupled hydrogeophysical inversion to identify non-Gaussian hydraulic conductivity field by jointly assimilating geochemical and time-lapse geophysical data [J]. *Journal of Hydrology*, 2019, 578: 124092. DOI: 10. 1016/j. jhydrol. 2019. 124092.
- [18] CAMPORESE M, CASSIANI G, DEIANA R, et al. Coupled and uncoupled hydrogeophysical inversions using ensemble Kalman filter assimilation of ERT-monitored tracer test data [J]. *Water Resources Research*, 2015, 51(5): 3277-3291. DOI: 10. 1002/2014WR016017.
- [19] CUI F, BAO J, CAO Z, et al. Soil hydraulic parameters estimation using ground penetrating radar data via ensemble smoother with multiple data assimilation [J]. *Journal of Hydrology*, 2020, 583: 124552. DOI: 10. 1016/j. jhydrol. 2020. 124552.
- [20] HARBAUGH A W, BANTA E, HILL M, et al. MODFLOW-2000, the U. S. geological survey modular ground-water flow model—User guide to modularization concepts and the ground-water flow process [R]. Reston, Virginia: U. S. Geological Survey, 2000.
- [21] ZHENG C, WANG P P. MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; Documentation and user's guide [R]. Washington, DC: U. S. Army Corps of Engineers, 1999: 1-219. DOI: 10. 2214/ajr. 169. 4. 9308495.

- [22] EMERICK A A. Deterministic ensemble smoother with multiple data assimilation as an alternative for history-matching seismic data [J]. *Computational Geosciences*, 2018, 22(5): 1175-1186. DOI: 10. 1007/s10596-018-9745-5.
- [23] GUO Q, SHI X, KANG X, et al. Evaluation of the benefits of improved permeability estimation on high-resolution characterization of DNAPL distribution in aquifers with low-permeability lenses [J]. *Journal of Hydrology*, 2021, 603: 126955. DOI: 10. 1016/j. jhydrol. 2021. 126955.
- [24] 张洁, 罗健, 夏瑜, 等. 阿尔奇公式的适用性分析及其拓展 [J]. *地球物理学报*, 2018, 61 (1): 311-322. (ZHANG J, LUO J, XIA Y, et al. Limitation analysis and modification of the Archie equation [J]. *Chinese Journal of Geophysics*, 2018, 61(1): 311-322. (in Chinese)) DOI: 10. 6038/cjg2018J0719.
- [25] RÜCKER C, GÜNTHER T, WAGNER F M. py-GIMLi: An open-source library for modelling and inversion in geophysics [J]. *Computers and Geosciences*, 2017, 109: 106-123. DOI: 10. 1016/j. cageo. 2017. 07. 011.

Joint inversion of contaminant source and aquifer parameters by assimilating ERT data with the ES-MDA algorithm

ZHOU Nianqing¹, ZHANG Ruicheng¹, JIANG Simin¹, XIA Xuemin²

(1. *Department of Hydraulic Engineering, Tongji University, Shanghai 200092, China*; 2. *School of Environment and Architecture, University of Shanghai for Science and Technology, Shanghai 200093, China*)

Abstract: Predicting the migration and transformation process of contaminants by utilizing numerical models is a commonly used method. The accuracy of model parameters could seriously affect the simulation results. While in practical site investigation, model parameters are often unavailable to obtain directly, thus it is needed to be estimated by solving the inverse problem of groundwater. In recent years, data assimilation has become a prevalent method for parametric uncertainty quantification. It can provide an estimate of the unknown model parameters by combining the observation data with the underlying dynamical principles governing the system. When solving high dimensional inversion problems, conventional observation methods can only provide sparse information (such as mass concentration and hydraulic head), so it can hinder the accuracy of the results. To deal with this issue, geophysical methods (such as electrical resistance tomography) are introduced to combine with the data assimilation because they can provide a large amount of continuous observation data.

The groundwater flow and solute transport model was constructed with the MODFLOW and MT3DMS programs and KL expansion was introduced to implement the dimensionality reduction of the hydraulic conductivity field. An inversion method for groundwater model parameters based on the ensemble smoother with multiple data assimilation (ES-MDA) was proposed. The joint inversion of contaminant source strength and hydraulic conductivity field was realized by integrating conventional observation data (mass concentration and hydraulic head) and geophysical data (ERT data) collected through the electrical resistance tomography (ERT). Three numerical cases were designed based on the former ideas to compare the inversion accuracy with different types of observation data. All three cases were the ES-MDA algorithm as the data assimilation method, while Case 1 was integrated the mass concentration and hydraulic head as the observation data; Case 2 was integrated the ERT data as the observation data and Case 3 was integrated the three types of data simultaneously. The root-mean-square error (E_{RMS}) was used to quantify the accuracy of the inversion results of the three cases.

The results showed that the E_{RMS} value for estimating the contaminant source strength got smaller (closer to zero) from Case 1 to Case 3, which exhibited that Case 3 obtained more accurate results by integrating multi-source observation data. As for the characterization of the hydraulic conductivity field, the posterior mean estimate of the log-conductivity field of Case 3 depicted the spatial distribution of the $\ln K$ field more accurately and its goodness of fit with the reference field was better than that of Case 1 and Case 2, which also demonstrated the advantage of using multi-source data as the observation.

ES-MDA algorithm could be utilized to realize the joint inversion of contaminant source strength and hydraulic conductivity field by assimilating both conventional observation data (mass concentration and hydraulic head) and ERT data. The inversion results by assimilating ERT data showed better accuracy than the case with conventional observation data, which demonstrated that a large amount of continuous geophysical data could provide more effective information for the inversion. The inversion results could be further optimized by combing the ERT data with the conventional observation data of mass concentration and head. It also showed the importance of using multi-source observation data when dealing with inversion problems.

Key words: data assimilation; ensemble smoother (ES); geophysics; electrical resistivity tomography (ERT); hydraulic conductivity field