Contaminant source and aquifer characterization: An application of ES-MDA demonstrating the assimilation of geophysical data

Zi Chen[a,b], Leli Zong[a], J. Jaime Gómez-Hernández[d], Teng Xu[c,*], Yuehua Jiang[a,b], Quanping Zhou[a,b], Hai Yang[a,b], Zhengyang Jia[a,b], Shijia Mei[a,b]

[a] *Nanjing Center, China Geological Survey, Nanjing, China*
[b] *Key Laboratory of Watershed Eco-Geological Processes, Ministry of Natural Resources, Nanjing, China*
[c] *College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China*
[d] *Institute of Water and Environmental Engineering, Universitat Politècnica de València, Valencia, Spain*

Abstract

Contaminant source and aquifer characterization (CSAC) is critical in groundwater pollution evaluation and remediation. The ensemble smoother with multiple data assimilation (ES-MDA) is utilized to jointly identify contaminant source information and hydraulic conductivities by assimilating time-lapse electrical resistivity tomography (ERT) data. In a synthetic profile with a time-varying release history in a heterogenous aquifer, we verify the performance of the proposed data assimilation framework. The results show that the CSAC problem could be handled by the proposed approach. The time-varying release history and the high permeability area can be identified with adequate time-lapse ERT measurements. Further reproduction of the evolution of the plume after CSAC also shows consistency with the reference plume. The poor conditioning inversion caused by the filter inbreeding is analyzed by comparing four scenarios with different apparent resistivity measurements. Furthermore, we also evaluate the impact of uncertainties in the petrophysical properties and geophysical observations on our data assimilation framework. The results show that the proposed ES-MDA data assimilation framework could provide a convincing inversion of the time-varying release history and hydraulic conductivities.

*Corresponding author
Email address: teng.xu@hhu.edu.cn (Teng Xu)

Keywords: Coupled modeling, Release history, Hydraulic conductivity, Data assimilation, Inversion

1. Introduction

Groundwater is a fundamental component of the hydrologic cycle, crucial for water supply and with many dependent ecosystems. Unfortunately, they can be easily polluted by anthropogenic activities, such as landfill operations, industry leakages, urban sewage, and others. Groundwater contamination is an important issue that has drawn the attention of researchers in the past decades (Gómez-Hernández and Wen, 1994; Gómez-Hernández et al., 2003; Feyen et al., 2003; Li et al., 2011a,b; Megdal, 2018). A critical issue in groundwater contamination is the identification of the source of contamination together with the characterization of aquifer properties, mainly hydraulic conductivity. Inverse problems in hydrogeology have been the focus of many researchers, who have found it to be an ill-posed problem with different solutions (Carrera and Neuman, 1986; Capilla et al., 1998, 1999; Wen et al., 1999; Franssen and Gómez-Hernández, 2002; Bagtzoglou and Atmadja, 2005; Hanea et al., 2015).

To date, several methods have been developed for contaminant source identification and there are several good reviews available (Atmadja and Bagtzoglou, 2001; Michalak and Kitanidis, 2004; Gómez-Hernández and Xu, 2021). The methods summarized in these reviews could be classified into three main categories: optimization, probabilistic, and backward-in-time simulation methods. The optimization approaches build an objective function and attempt to minimize the discrepancies between simulated and observed measurements (Gorelick et al., 1983; Sun et al., 2006; Mirghani et al., 2009; Ayvaz, 2010; Li et al., 2012); the probabilistic approaches handle the problem with a stochastic framework and try to approximate the posterior probabilities of the simulated measurements conditioned on the observed ones (Woodbury and Ulrych, 1996; Zeng et al., 2012; Butera et al., 2013; Cupola et al., 2015; Pirot et al., 2019; Jiang et al., 2021); backward-in-time simulation methods solve the solute transport equations backward to identify the most likely contaminant release locations (Bagtzoglou et al., 1992; Neupauer and Wilson, 1999; Bagtzoglou and Atmadja, 2003; Ababou et al., 2010).

For the past few years, data assimilation methods have become increasingly prominent for their versatile and efficient features. The ensemble Kalman filter (EnKF) proposed by Evensen (2003) and the ensemble smoother with multiple data assimilation (ES-MDA) proposed by Emerick and Reynolds (2013) have been progressively employed for contaminant source identification. Xu and Gómez-Hernández (2016) first proposed the restart normal-score EnKF to handle the contaminant source identification problem and then extended this method to recognize contaminant source information and heterogeneous hydraulic conductivity jointly(Xu and Gómez-Hernández, 2018). Li et al. (2019) employed a mixed integer nonlinear programming optimization model together with Kalman filter to deduce contaminant source information. Panjehfouladgaran and Rajabi (2022) combined artificial neural networks and constrained the restart EnKF to characterize the pollutant source in a coastal aquifer and then moved one step further to identify aquifer heterogeneity in a tide-influenced coastal aquifer (Dodangeh et al., 2022). The ES-MDA method has been coupled with generative adversarial networks by Bao et al. (2020) to handle a channelized aquifer characterization problem. Todaro et al. (2021) applied the ES-MDA method to recognize pollutant source location and release history. Furthermore, Xu et al. (2022) handled the non-point source identification puzzle via ES-MDA.

The aforementioned research findings are proof of the capacity of data assimilation methods for the joint identification of pollutant sources and aquifer heterogeneity, however, despite a few applications in sandbox experiments, there is still a lack of verification in field cases. One main reason is that the measurements required are not available; at most, only a few sparse and discontinuous pollution data are accessible. One possible solution to the lack of contaminant data could be the use of geophysical surveys. Geophysical methods have been extensively utilized in groundwater contamination investigation, especially electrical resistivity tomography (ERT), which is a cost-efficient and non-intrusive method with a high sampling density (Nenna et al., 2011; Seferou et al., 2013; Binley et al., 2015; Mao et al., 2016; Shao et al., 2021; Xia et al., 2021). ERT could be the perfect data source for ensemble-based contaminant source identification problems. As far as we know, several works have already combined data assimilation methods with ERT data to study groundwater movement or contamination issues. Crestani et al. (2013) first compared the ensemble Kalman filter (EnKF) and the ensemble smoother (ES) capabilities in identifying hydraulic conductivity via a tracer test, and then directly using the ERT data to inverse the heterogenous hydraulic conductivity in both a synthetic and a real test case (Crestani et al., 2015). Bouzaglou et al. (2018) combined the EnKF method and the SUTRA model to update groundwater states and soil parameters by using ERT measurements in a seawater intrusion laboratory experiment. Kang et al. (2018) first developed an EnKF-based data assimilation algorithm to jointly recognize DNAPL saturation together with a hydraulic conductivity field by assimilating time-lapse ERT data and later employed the ensemble smoother-direct sampling method (ES-DS) to identify a non-Gaussian aquifer by using both time-lapse geophysical and geochemical datasets (Kang et al., 2019). Tso et al. (2020) utilised cross-borehole time-lapse ERT data to identify contaminant source information through an ensemble-based data assimilation framework. The aforementioned study substantiates the efficacy and dependability of the ERT technique when applied to the examination of groundwater contamination problems. However, the majority of these studies confine their focus solely to the identification of heterogenous hydraulic conductivity or contaminant source information, assuming the other factor is already known. These assumptions present significant impracticality when applied to real-world scenarios, owing to the inherent scarcity of subsurface information and the persistent lack of data to detect groundwater pollution.

In this paper, we establish a benchmark that employs the ensemble smoother with multiple data assimilations (ES-MDA) to jointly identify time-varying release history and aquifer heterogeneity by using time-lapse ERT data. Besides, we also evaluate the impact of different cross-hole configurations on the proposed data assimilation framework. To our knowledge, this is the first instance of time-lapse ERT measurements being used for the joint identification of contaminant source information and hydraulic conductivities. The rest of this paper is organized as follows: First, we outline the methodology of the proposed data assimilation framework in section 2, including the coupling of groundwater flow, solute transport, and geophysical modeling into the ES-MDA implementation; then, in sections 3, a benchmark case with a time-varying releasing history in a heterogenous aquifer property is built that will be used as the reference to test the proposed method; section 4 evaluates the proposed approach, and follows a discussion in section 5; finally, in section 6, we summarize the main findings of this work and propose some future works need to be done.

2. Methodology

2.1. Groundwater flow and solute transport model

Groundwater flow and solute transport in an aquifer can be described by the following partial differential equations (Bear, 1972; Zheng and Wang, 1999)

$$\nabla \cdot (K \nabla h) + w = 0, \tag{1}$$

$$\frac{\partial (\theta C)}{\partial t} = \nabla \cdot (\theta D \cdot \nabla C) - \nabla \cdot (\theta v C) - q_s C_s, \tag{2}$$

where K represents hydraulic conductivity $[LT^{-1}]$; h stands for hydraulic head $[L]$; $\nabla \cdot$ and ∇ are the divergence and gradient operators, respectively; w stands for additional sources and sinks $[T^{-1}]$; C denotes the dissolved contaminant concentration $[ML^{-3}]$; θ represents effective porosity of the medium [-]; t denotes time $[T]$; D is a hydrodynamic dispersion coefficient tensor $[L^2T^{-1}]$; v stands for the flow velocity of the groundwater $[LT^{-1}]$ obtained from the solution of groundwater flow model (Eq. (1)); q_s denotes the volumetric flow rate per unit volume of aquifer associated with a fluid source or sink $[T^{-1}]$, while C_s stands for the concentration of the source or sink $[ML^{-3}]$. The solution of both equations requires the specification of initial and boundary conditions. In this work, the groundwater flow equation is numerically solved through the finite difference MODFLOW program (Harbaugh, 2005) while the solute transport equation is handled by the finite difference MT3D program (Bedekar et al., 2016).

2.2. Geophysical model

The electrical potential field induced by a couple of electrodes can be characterized by the following equation

$$-\nabla \cdot \frac{1}{\rho}\nabla V = I(\delta(\mathbf{r}-\mathbf{r}_+) - \delta(\mathbf{r}-\mathbf{r}_-)), \tag{3}$$

where ρ is porous media resistivity; V denotes electrical potential field; I is the input current from a dipole; $\mathbf{r}_+$ and $\mathbf{r}_-$ are the locations of the positive and negative electrodes, respectively, and $\delta(\cdot)$ is the Dirac delta function.

Here, the porous media resistivity depends on several factors, such as porosity or pore water conductivity. The resistivity model proposed by Revil et al. (2018) is employed in this work

$$\frac{1}{\rho} = (S_w\phi)^m\sigma_w + (S_w\phi)^{m-1}\rho_s(B-\lambda)CEC, \tag{4}$$

where S_w is water saturation, which equals 1 in aquifers; ϕ is porosity, which in this study is 0.32 for both the coarse and fine sands (Power et al., 2013); m is a cementation exponent; ρ_s is grain density; B and λ are apparent mobility of the counterions responsible for surface conduction and polarization, respectively; CEC is the quantity of exchangeable cation on the surface of the silica grains, and can be calculated by the following equation (Revil, 2013; Revil et al., 2017)

$$CEC = 6\frac{Q_s}{\rho_s d}, \tag{5}$$

where Q_s equals 0.64 Cm^{-2}; d stands for the mean grain diameter of the sand. σ_w is the conductivity of the pore water, which is determined by the ionic concentration and temperature (Sen, 1992) according to the following expression

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

where T is temperature, which is assumed constant at 25 °C in this research; and C is ionic concentration.

In the context of resistivity field surveys, the apparent resistivity (ρ_a) is preferred over the resistivity as it serves as a proxy for the alteration in the electrical characteristics of the subsurface medium with a specific electrode array configuration, defined by

$$\rho_a = K\frac{\Delta U}{I}, \tag{7}$$

where K denotes a geometric factor which depends on the electrode array configuration; ΔU stands for the potential difference between the two potential electrodes M and N.

Specifically, the apparent resistivity can be computed from the resistivity values obtained from Eq. (4) by solving a geophysical forward problem. In this work, the geophysical forward problem is solved by the finite-element open-source software ResIPy (Blanchy et al., 2020).

With the contaminant concentration calculated from MODFLOW and MT3DS and several geophysical parameters, the resistivity values of the aquifer could be obtained from

Eq. (4),(5) and (6). Thus, with the geophysical forward model ResIPy, the relationship between apparent resistivity ρ_a and concentration C can be established.

2.3. Ensemble Smoother with Multiple Data Assimilation (ES-MDA)

The ES-MDA technology is employed to identify a contaminant source and aquifer heterogeneity from apparent resistivity data. A short description of the ES-MDA is provided next, for a more in-depth discussion the reader is referred to Emerick and Reynolds (2013):

1. Procedure

The first step of the ES-MDA technology is to generate a certain number N_e of realizations with unknown contaminant source and aquifer characterization (CSAC) parameters, which include time-varying release history and hydraulic conductivity spatial distribution in this work. Once the number of iterations N_a and the inflation factor α_j (explained in detail below) are determined, the method will go through two main steps: forecast step and analysis step.

In the forecast step, for each member of the realizations, the groundwater flow, solute transport and geophysical models (MODFLOW, MT3DS and ResIPy) are solved sequentially,

$$B^f_{i,j} = \psi[B_0, A_{i,j}], \tag{8}$$

where B^f denotes the vector of forecasted apparent resistivity, while B_0 represents the vector of initial apparent resistivity; i stands for the realization index, $i = 1, 2, \ldots, N_e$ and j is the iteration index of ES-MDA, $j = 1, 2, \ldots, N_a$; ψ denotes the forward numerical simulators, which are MODFLOW, MT3DS and ResIPy in this case; A stands for the vector of CSAC parameters, including time-varying release history and hydraulic conductivity distribution.

Then, the CSAC parameters are updated using a truncated singular value decomposition

(TSVD) algorithm in the analysis step, the detailed procedure can be expressed as following,

$$A_{i,j+1} = A_{i,j} + \Delta A_j (\Delta B_j^f)^T [\Delta B_j^f (\Delta B_j^f)^T + \alpha_j R]^{-1} [y_{obs} + \sqrt{\alpha_j}\varepsilon - B_{o,i,j}^f], \tag{9}$$

where i and j stand for the same meaning as in Eq. 8; α_j stands for the inflation factor; y_{obs} denotes an $N_o \cdot N_t \times 1$ vector of apparent resistivity observations (N_o denotes the amount of measurements in a single time step, while N_t stands for the amount of time steps with measurements); ε denotes the observation error, with a observation error covariance matrix R; $B_{o,i,j}^f$ represents the vector of forecasted apparent resistivity at observation locations; ΔA_j and ΔB_j are square root matrices defined as

$$\Delta A_j = \frac{1}{\sqrt{N_e - 1}} [A_{1,j} - \overline{A}_j, A_{2,j} - \overline{A}_j, \ldots, A_{N_e,j} - \overline{A}_j], \tag{10}$$

$$\Delta B_j^f = \frac{1}{\sqrt{N_e - 1}} [B_{1,j}^f - \overline{B^f}_j, B_{2,j}^f - \overline{B^f}_j, \ldots, B_{N_e,j}^f - \overline{B^f}_j], \tag{11}$$

where $\overline{A}_j$ and $\overline{B^f}_j$ are the ensemble means of the CSAC parameters subject to identification and of the forecasted apparent resistivity at the j_{th} iteration, respectively.

2. Inflation factor

The inflation factor α_j is employed to augment the covariance matrix associated with the measurements errors to damp the changes in the model parameters at early iterations (Emerick and Reynolds, 2013). It is influential to the functioning of the ES-MDA, therefore, several ways on how to compute them have been described in previous studies (Le et al., 2016; Rafiee and Reynolds, 2017; Evensen, 2018). In this work, based on our previous experience (Chen et al., 2022), we decide to apply the inflation factor scheme proposed by Rafiee and Reynolds (2017). According to their detailed procedures, the first step is to generate the initial inflation factor based on the following equation:

$$\alpha_1 = (\frac{1}{N}\sum_{i=1}^{N}\lambda_i)^2, \tag{12}$$

where α_1 denotes the initial inflation factor; N is the minimum between ensemble size N_e and the total amount of apparent resistivity measurements $N_o \cdot N_t$; λ_i denotes the singular values of the dimensionless sensitivity matrix D_j, which is defined as

$$D_j = R^{-\frac{1}{2}} \triangle B_j^f. \tag{13}$$

Then, the succeeding inflation factors are determined in a geometrical decreasing progression,

$$\alpha_j = \beta^{j-1}\alpha_1, \tag{14}$$

where β denotes the common ratio that ensures the summation of the reciprocal of the inflation factors equals to one. Its value can be obtained by

$$\frac{1-(1/\beta)^{N_a-1}}{1-1/\beta} = \alpha_1. \tag{15}$$

3. The normal-score transformation

Although ES-MDA is capable of handling non-linear models, its performance deteriorated obviously when the augmented state vector followed a non-Gaussian distribution (Zhou et al., 2014; Cao et al., 2018). To address this problem, several methods have been proposed, such as using iterative approaches, reparameterizations, Gaussian mixture models and normal-score transform (Hendricks Franssen and Kinzelbach, 2008; Zhou et al., 2011; Kumar and Srinivasan, 2019). In this paper, the normal-score transform algorithm is combined with ES-MDA to deal with non-Gaussianity. The main procedure of this method follows two steps: (i) transform the non-Gaussian augmented state vector into a marginally-Gaussian vector, and then perform ES-MDA in Gaussian space; (ii) back transform the ES-MDA

Table 1: A detailed description of the proposed data assimilation framework.

Framework: ES-MDA with coupled models

• Generate initial ensemble, A_0 (including K_0,C_0).
• Choose the number of ES-MDA iterations, N_a.
• For $j = 1$ *to* N_a
    ◦ Set $A^f_{i,j} = A^a_{i,j-1}$ for $i = 1, 2, ...N_e$.
    ◦ Execute the groundwater flow and solute transport simulators for each realization.
    ◦ Calculate ρ using (4) and (6).
    ◦ Execute the geophysical simulator, obtain apparent resistivity ρ_a.
    ◦ Calculate α_j using (12),(13),(14) and (15).
    ◦ Apply the normal-score transformation.
    ◦ Update model parameters $A^a_{i,j}$ based on (9).
    ◦ Apply the normal-score back transformation.
• Endfor

updates into its original space. One more thing need to point out is that the normal-score transform algorithm does not guarantee its higher-order moments will also follow a multi-Gaussian distribution, however, the outcome of normal-score ES-MDA outperforms ES-MDA for clearly non-Gaussian parameters (Jafarpour and Khodabakhshi, 2011; Kumar and Srinivasan, 2020).

2.4. Data assimilation workflow

Figure 1 illustrates the details of the proposed data assimilation workflow. Using this workflow, we are able to jointly update the non-Gaussian hydraulic conductivity and source release history by assimilating the apparent resistivity. Note that since MODFLOW and MT3DS are finite-difference numerical methods, while ResIPy is a finite-element method, an extra grid refinement procedure is needed before the geophysical model is run.

3. Application

3.1. Synthetic profile description

To test the proposed data assimilation framework, a benchmark case with a non-Gaussian confined aquifer and time-varying contaminant release is constructed. The contaminant movement in a two-dimensional synthetic model of $40 \times 1 \times 20 m^3$ is simulated, and then a quasi-3D geophysical model (infinite in y dimension) is used to capture the evolution in time of the pollutant concentration.

The profile model is discretized into 80 by 1 by 40 cells, each of which is 0.5 by 1 by 0.5 m. The model is filled with fine and coarse sand, and the detailed spatial distribution of these two materials are arranged based on a truncated Gaussian simulation (Journel and Isaaks, 1984) with a threshold of 25 percentage, as shown in Figure 1. Notice that, the hydraulic conductivity values of the fine and coarse sand are generated by normal distribution algorithm with a mean of $0.5, 15$ m/d and standard deviation of $0.06, 1$ m/d, respectively. The boundary conditions are defined as follows: the left boundary is a constant head boundary of $30m$; the right boundary is zero-flow, except for the top four cells that are time-varying income flow through which the contaminant enters the aquifer; the upper and lower boundaries are impermeable. Such contamination could mimic the release from a contaminated river or irrigation canal in reality. The release pattern follows the same bimodal pulse proposed by Skaggs and Kabala (1994) and used many times later by others (Figure 1) given by:

$$C(t) = 2 \cdot \exp(-\frac{(t-10)^2}{35}) + 0.6 \cdot \exp(-\frac{(t-25)^2}{80}) + \exp(-\frac{(t-45)^2}{40}) \qquad 0 \leq t \leq 100 \tag{16}$$

For the quasi-3D geophysical model, a fine triangular mesh is generated in the zone of investigation (the profile) to compute the voltage field. Moreover, the left, right and lower boundaries of the mesh have been extended to 200 m (5 times the profile length) away from the profile to mimic the infinite boundaries, and the elements gradually increase in

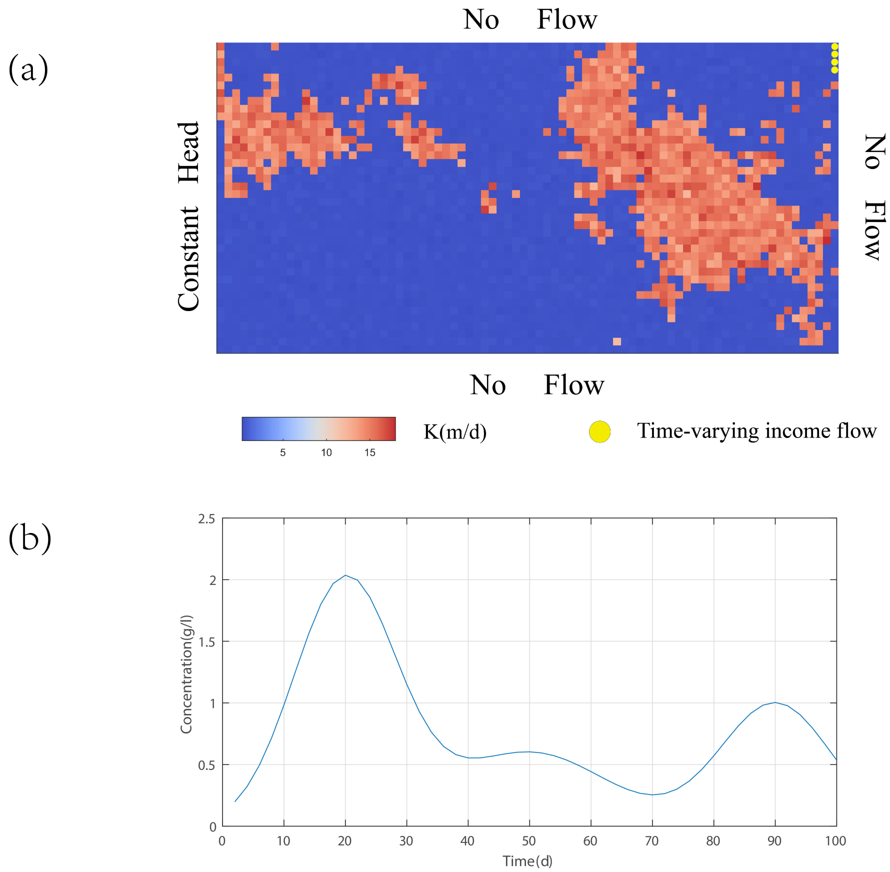

Figure 1: Schematic view of the groundwater flow and solute transport reference model. (a) Flow boundary conditions and reference hydraulic conductivity field. (b) Reference concentration release curve.

size laterally and vertically in this extension region. 30 electrodes are assigned in three vertical boreholes with an interval of 2 m as shown in Figure 2. A and B denote the current electrodes, while M and N stand for the potential electrodes. We adopt the bipole-bipole electrode array configuration based on previous research since the cross-hole AM-BN configuration yields greater flexibility in practice without any singularity problem in data acquisition (Zhou and Greenhalgh, 2000). And the measurement is performed by staying AM electrodes in one borehole and moving down the BN electrodes in the other. Once the potential electrode N reaches the bottom of the borehole, AM electrodes will move down one interval and then the second round of measurements get started. In this paper, the separation between the electrodes BN is kept the same as those in AM (AM=BN=a). Three vertical separation distance values are analyzed (a=6 m, 4 m, 2 m). Four different numbers of ρ_a measurements are considered (98, 128, 162, and 388). See Table 2 for the list of scenarios analyzed. In scenario S1, a sparse AM-BN scheme with an electrode spacing of 6 m is used (98 measurements per time step, 980 in total). This is a relatively small amount of measurements in a time-lapse ERT survey but could happen in reality when limited data processing capabilities are available (Binley and Kemna, 2005). We gradually increase the number of measurements in scenarios S2 and S3, with AM-BN schemes of 128 and 162 measurements, respectively. In scenario S4, all three electrode separation distances are used, resulting in a total of 388 measurements per time step.

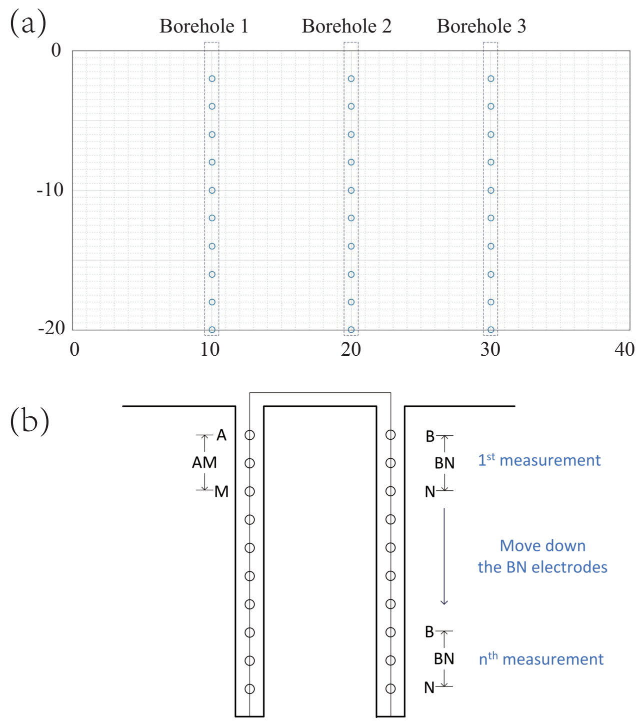

Figure 2: Schematic view of the geophysical synthetic model. (a) The distribution of the boreholes and electrodes. The black dotted box represents the borehole; the blue circle stands for a single electrode. (b) Configuration of the bipole-bipole electrodes array. A and B denote the current electrodes, while M and N stand for the potential electrodes.

The total simulation time of groundwater flow and transport solute is 100 days, and the models are run in 50 equal-sized time steps. As for the geophysical model, the measurements

Table 2: Definition of the synthetic scenarios

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Vertical separation distance, a (m)	6	4	2	2, 4 and 6
Number of ρ_a measurements	98	128	162	388

are acquired with a time interval of 10 days. A more detailed description of the parameters used in the MODFLOW, MT3DS and ResIPy models are listed in Table 3. The reference hydraulic head, a couple of snapshots of the reference contaminant plumes and their related reference resistivity are shown in Figure 3.

For the assimilation phase, and based on our previous work (Chen et al., 2018, 2021; Xu et al., 2021), the number of iterations (N_a) is chosen to be 4, the ensemble size (N_e) is taken as 500. The initial hydraulic conductivity distribution fields are generated from the same algorithm as the reference one, while the initial release history ensemble are generated based on uniform distribution with a range between [0.5, 1.5] g/l. In this work, a 1.20% relative error is added to the apparent resistivity (following a Gaussian distribution with a mean of 0 and standard deviation of 0.1 $\Omega \cdot m$), while the aforementioned forward models errors are neglected.

3.2. Evaluation Criterion

The root mean square error (RMSE) is one of the most effective criteria to evaluate the estimation accuracy of the ensemble-based methods and has been utilized in this work:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i^{ref} - \overline{A}_i)^2}, \tag{17}$$

where n denotes the number of segments in which the release history curve is discretized or the number of cells in which hydraulic conductivity must be identified, A_i^{ref} is the i^{th} component of reference CSAC parameters, while $\overline{A}_i$ represents the ensemble mean of the i^{th} component of the updated CSAC parameters. The smaller the values for RMSE, the better

Table 3: Groundwater flow, solute transport, geophysical model parameters

Parameters	Value
Model discretization	
model length along x (m)	40
model length along y (m)	1
model height along z (m)	20
grid size $\Delta x \times \Delta y \times \Delta z$ (m)	$0.5 \times 1 \times 0.5$
total simulation time (d)	100
time step length (d)	2
number of time steps	50
Groundwater flow model parameters	
mean of hydraulic conductivity, coarse sand (m/d)	15
mean of hydraulic conductivity, fine sand (m/d)	0.5
Std. of hydraulic conductivity, coarse sand (m/d)	1
Std. of hydraulic conductivity, fine sand (m/d)	0.06
Solute transport model parameters	
longitudinal dispersivity (m)	0.5
transverse dispersivity (m)	0.025
molecular diffusion coefficient	0
initial water concentration (g/l)	0.15
Geophysical model parameters	
porosity, coarse sand & fine sand	0.32[a]
diameter, coarse sand(mm)	2
diameter, fine sand(mm)	0.25
CEC, coarse sand(C/kg)	0.72[b]
CEC, fine sand(C/kg)	5.80[b]
Cementation exponent, coarse sand & fine sand	2.0[a]
$B(\mathrm{m^{-2}s^{-1}V^{-1}})$	4.1×10^{-9} [c]
$\lambda(\mathrm{m^{-2}s^{-1}V^{-1}})$	3.46×10^{-10} [c]

[a] Power et al. (2013).
[b] Revil (2013).
[c] Revil et al. (2018).

the estimation of the CSAC parameters.

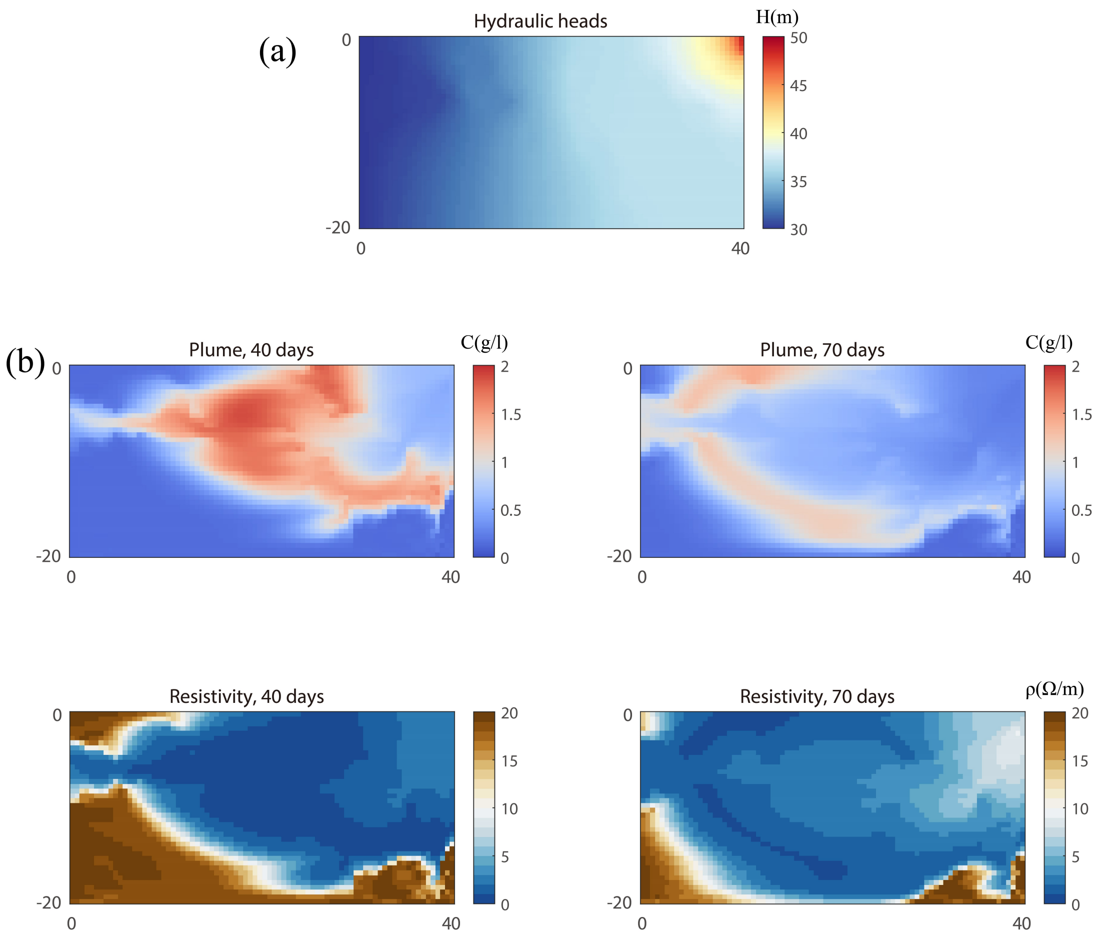

Figure 3: The properties of the reference models. (a) Hydraulic head distribution. (b) Reference contaminant plumes on day 40 and 70. (c) Reference resistivity distribution on day 40 and 70.

4. Results

4.1. Contaminant source and aquifer characterization

Figure 4 represents the recovered time-varying release histories for scenario S1 to S4. For each scenario, the blue dotted line illustrates the reference time-varying release history, while each gray line represents one recovered release history curve for one ensemble member, the red dotted line stands for the median, while the black dashed lines denote the 5 and 95 percentiles of the recovered release history curves. It can be clearly seen that the median of the recovered release history curves follows the true release in all scenarios, but with an excess of fluctuation. This noticeable fluctuation in the ensemble medians and individual curves is believed to be caused by the inherent ill-posedness of identification problem (Chen et al., 2022). For scenarios S1 to S3, the increase in the number of measurements seems able to improve the characterization of the pollutant release curve with a narrower width of the 90% confidence interval. However, in Scenario S4, with the largest number of observations with an AM-BN scheme with 388 measurements, the 90% confidence interval gets too narrow and in several instants does not contain the reference release curve. The calculated $RMSE_c$ for all scenarios is shown in Table 4 and reinforces the previous statements. For scenarios S1 to S2, the $RMSE_c$ declines slightly with the increasing of measurements, while in scenario S4, the $RMSE_c$ has a value second only to the initial ensemble. Despite the use of an adaptive covariance inflation in this work, the ensemble variance of the release history in scenario S4 is still underestimated, and this poorly conditioned inversion can be attribute to filter inbreeding (Hendricks Franssen and Kinzelbach, 2008). Another thing to pay attention to is the bad estimation of the release curve for the last time steps. This outcome can be explained in that there are not enough data for the ES-MDA method to estimate the release at the latest release times.

As for aquifer characterization, Figure 5 shows the ensemble means and variances of hydraulic conductivities for the initial ensemble and scenarios S1 to S4. The ensemble mean

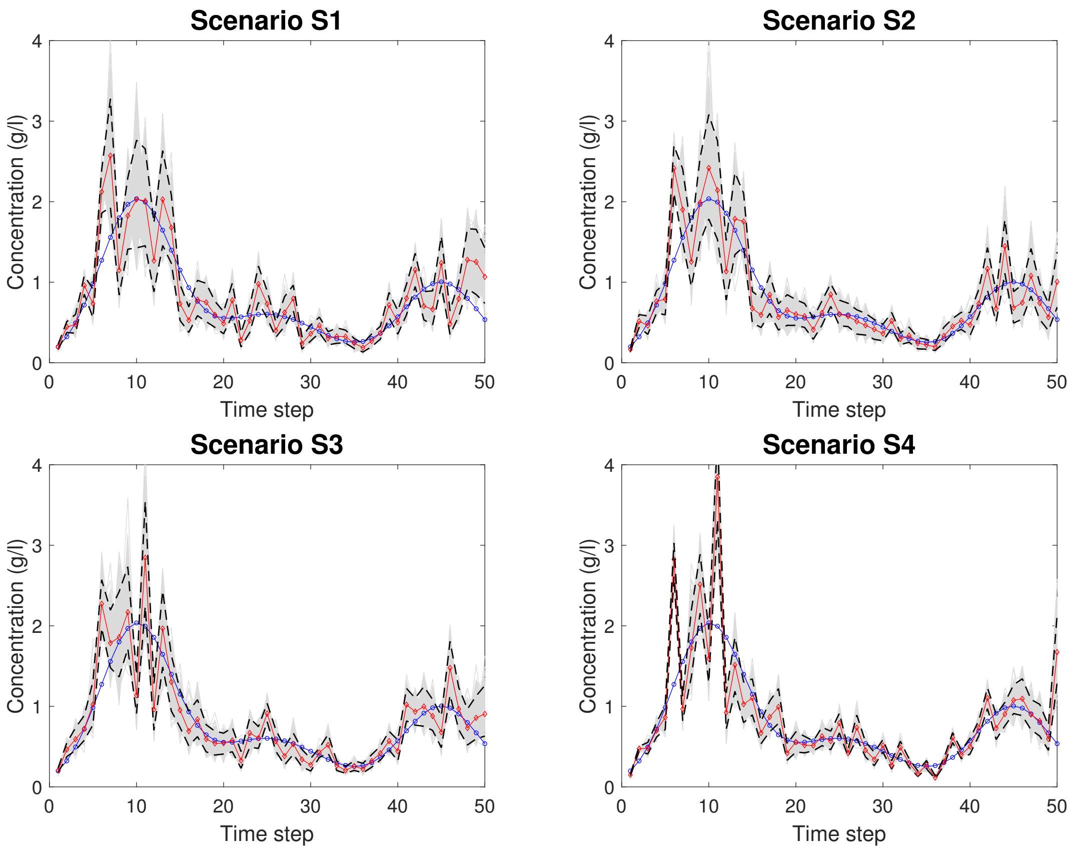

Figure 4: Recovered time-varying release histories for scenario S1-S4. The blue dotted line illustrates the reference time-varying release history, each gray line represents one recovered release history curve for one ensemble member, the red dotted line stands for the median, while the black dashed lines denote the 5 and 95 percentiles.

of the hydraulic conductivities in the initial ensemble is relatively homogenous, while the ensemble variance takes a large value. After assimilating the geophysical measurements, the ensemble mean of the updated hydraulic conductivities can delineate roughly the facies distribution in the aquifer with a substantial reduction of the ensemble variance in all scenarios. Further comparison between all scenarios demonstrate that S2 has the best aquifer characterization, with a clear identification of the high permeability area and a relatively small ensemble variance, while S1 and S3 could delineate the high hydraulic conductivity zone less precisely and with a larger variance. In Scenario S4, a similar outcome to S2 is obtained, but with a poorer description of the high permeability area. A quantitative evaluation of all

Table 4: Performance of the scenario S1 to S4

	Initial ensemble	Scenario 1	Scenario 2	Scenario 3	Scenario 4
$RMSE_C$	0.535	0.3236	0.2820	0.3057	0.4416
$RMSE_K$	7.435	6.6774	6.6243	6.7221	7.0766

scenarios is listed in Table 4. Once again, we could find out that S4 has an $RMSE_K$ value much greater than S1 to S3, this outcome is contrary to the general understanding (the more data the better the characterization). It can be the attribute to the fact that the poorly conditioned inversion of the release history in scenario S4 deteriorates the characterization of the heterogeneous hydraulic conductivities.

4.2. Contaminant plume reproduction

For a more intuitive representation of the inversion results, the updated CSAC parameters are utilized to simulate the contaminant plume evolution and representative snapshots are taken to check the performance of the proposed data assimilation framework. Figure 6-9 show the reference plume, and the ensemble mean plumes simulated with the initial set of CSAC parameters and with the updated CSAC values for all 4 scenarios in days 20, 40, 70 and 90. The ensemble mean plumes generated by the initial ensemble spread widely with very large uncertainty since no observed data have been assimilated yet. The comparison between the reference and simulated contaminant plumes speaks favorably about the proposed methodology since for all 4 scenarios main contaminant plume morphology is captured at the different time snapshots. A closer look, point to S2 as the scenario that performs best, especially in the high permeability area.

Figures 10-12 show the time evolution of the vertically-averaged concentration along the boreholes 1, 2, and 3 simulated with the updated CSAC parameters for all 4 scenarios. The blue curve stands for the concentration evolution in the reference, while each gray line represents one concentration curve for one ensemble member, and the red line denotes the ensemble median. The shape of the concentration curves is well reproduced in all scenarios. Again, a closer look shows that scenario S2 performs best, especially in borehole 2 and 3.

4.3. Uncertainty evaluation for electrical properties and apparent resistivity observation error

The aforementioned four scenarios (S1-S4) are carried out with two assumptions: (1) the electrical properties of fine and coarse sand (CEC values) are constant; (2) the observation error of apparent resistivity is relatively small. However, these two assumptions are somewhat idealistic. In a more realistic scenario, the uncertainties of the petrophysical properties and geophysical observations may deteriorate the performance of the proposed data assimilation framework (Troldborg et al., 2010; Laloy et al., 2012; Brunetti and Linde, 2018).

Hence, we have run two additional scenarios S5 and S6, which serve as repetitions of scenario S2 but with larger uncertainties in the electrical properties or apparent resistivity observation errors. More detailed, in scenario S5, the CEC values of fine and coarse sand are no longer constant, but follow a Gaussian distribution with a mean of $0.72, 5.80$ and standard deviation of $0.072, 0.580$ (10% relative error), respectively; in scenario S6, a more realistic noise (Jardani et al., 2013), which is following Gaussian distribution with a mean of 0 and standard deviation of $1.0\ \Omega \cdot m$ (12.0% relative error), is added to the apparent resistivity. The rest of the setup is exactly the same as in scenario S2, except for the electrical properties and apparent resistivity observation errors. We also show the same sets of figures shown for scenario S2. Figure 13 represents the recovered time-varying release histories for scenarios S5 and S6; Figure 14 shows the ensemble means and variances of the hydraulic conductivities for scenarios S5 and S6; Figure 16 illustrates the ensemble mean contaminant plumes computed with the updated CSAC values for scenarios S5-S6 at days 20, 40, 70 and 90.

A quick comparison between the figures for scenarios S2 and S5 shows that the outcomes are quite similar in both scenarios. The recovered time-varying release history, ensemble mean of hydraulic conductivities and the ensemble mean contaminant plumes computed with the updated CSAC values are all virtually the same in scenario S2 and S5, illustrating that the heterogeneity of electrical properties in this level hasn't deteriorate the performance of the proposed data assimilation framework. In contrast, by performing a comparative analysis of the inversion outcomes between scenarios S2 and S6, we can evaluate the effect of the observation uncertainties in the apparent resistivity. In scenario S5 , the median of the recovered release history curves has some deviations from the reference release curve (especially between time step 5 to 15) and the 90% confidence interval becomes much wider; the identified high permeability area is less accurate and the ensemble variance is still quite large; the ensemble mean plume struggles to recover the high plume concentration. In summary, we believe that both the uncertainties in the electrical properties and in the apparent resistivity observations both have some impact on our data assimilation framework, but at least in this case, the uncertainties in the apparent resistivity observations certainly deserve more attention.

In summary, we have again demonstrated the capacity of the proposed data assimilation framework by taking into account the uncertainties in the electrical properties and apparent resistivity observations.

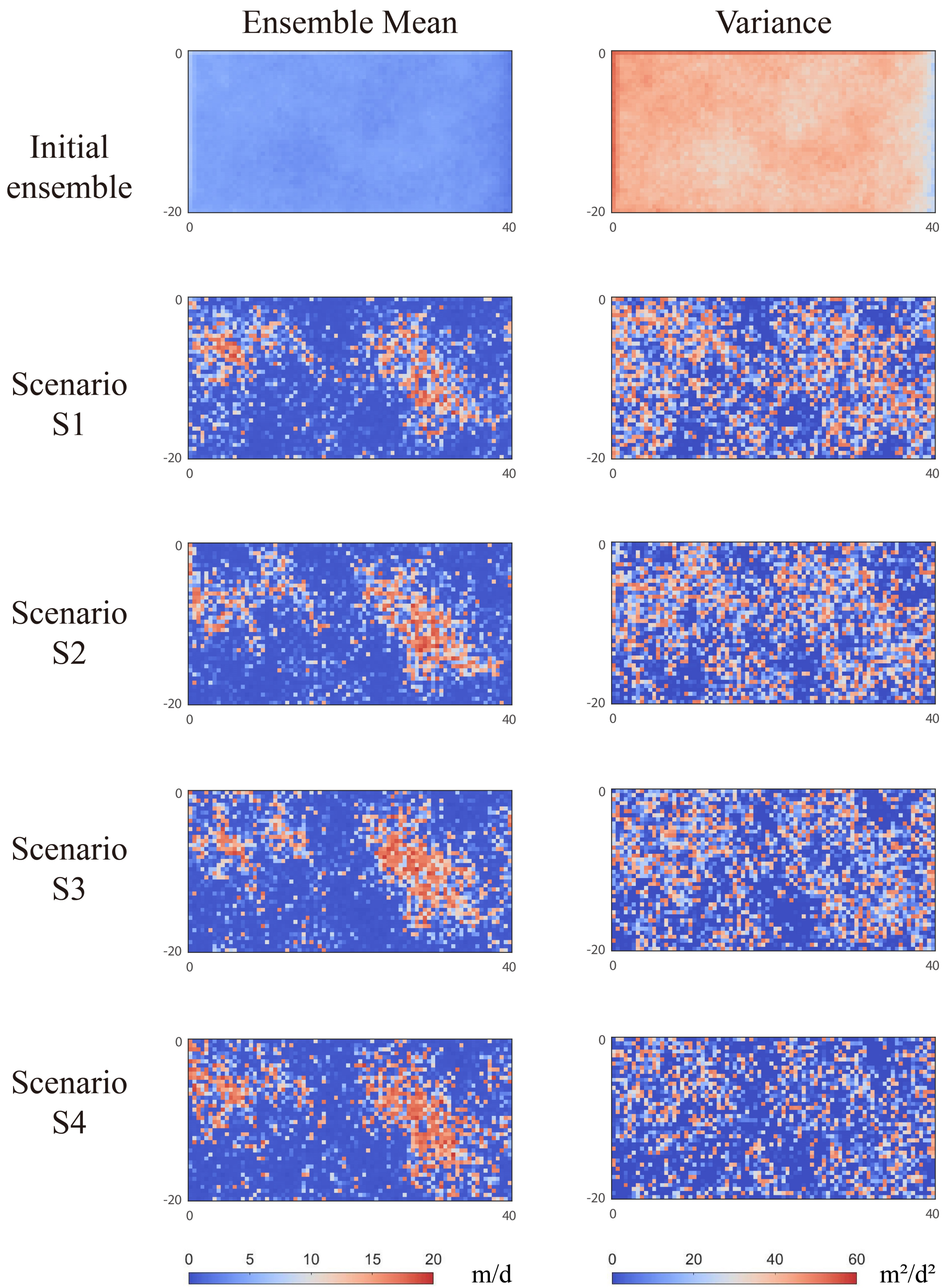

Figure 5: Ensemble means and variances of hydraulic conductivities for the initial ensemble and scenarios S1 to S4.

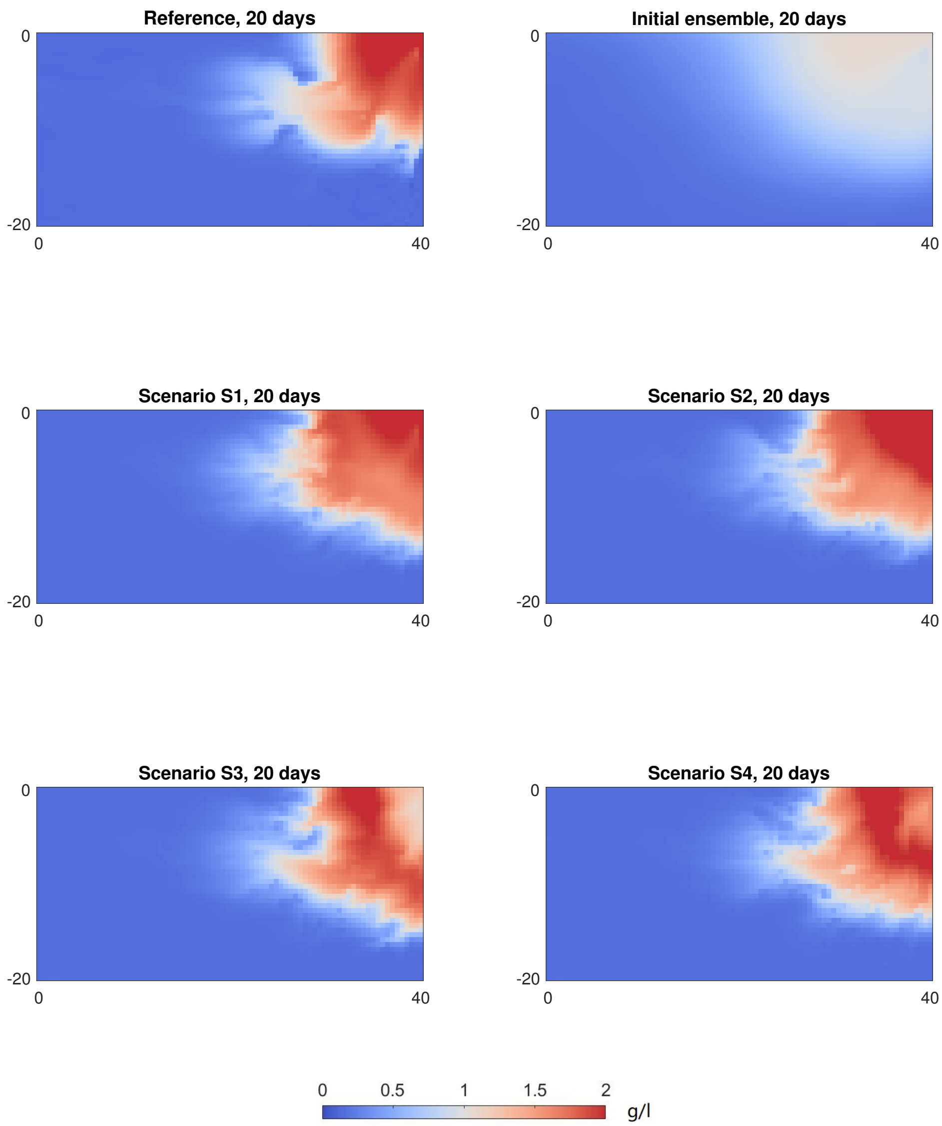

Figure 6: The reference contaminant plume and the ensemble mean contaminant plumes computed with the initial set of CSAC parameters and with the updated CSAC values for scenario S1-S4 on day 20.

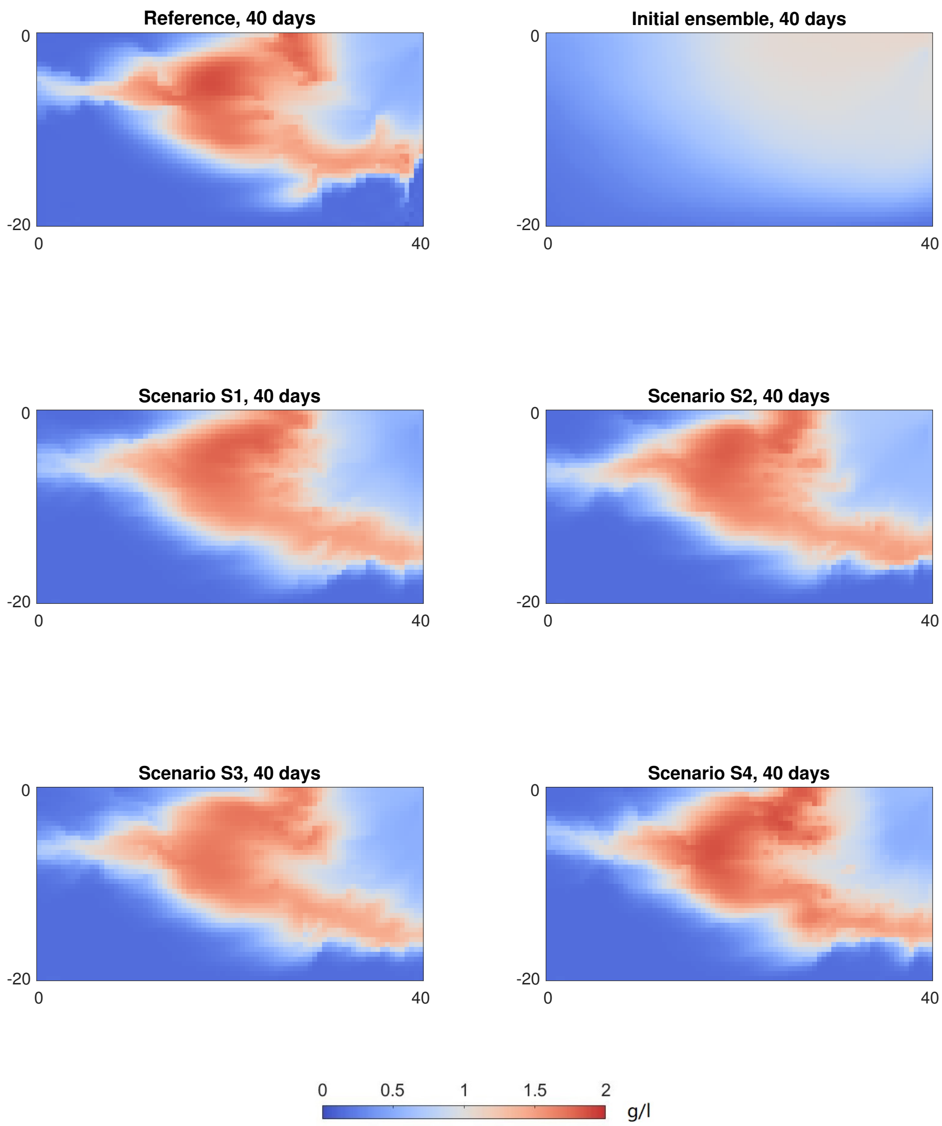

Figure 7: The reference contaminant plume and the ensemble mean contaminant plumes simulated with the initial set of CSAC parameters and with the updated CSAC values for scenario S1-S4 on day 40.

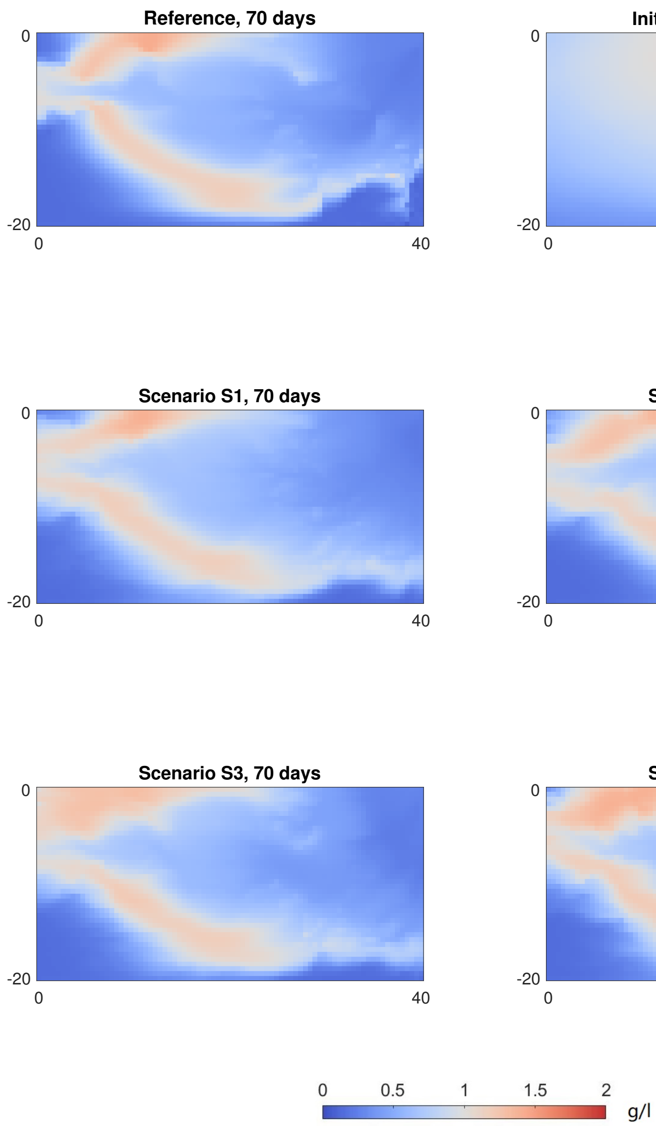

Figure 8: The reference contaminant plume and the ensemble mean contaminant plumes simulated with the initial set of CSAC parameters and with the updated CSAC values for scenario S1-S4 on day 70.

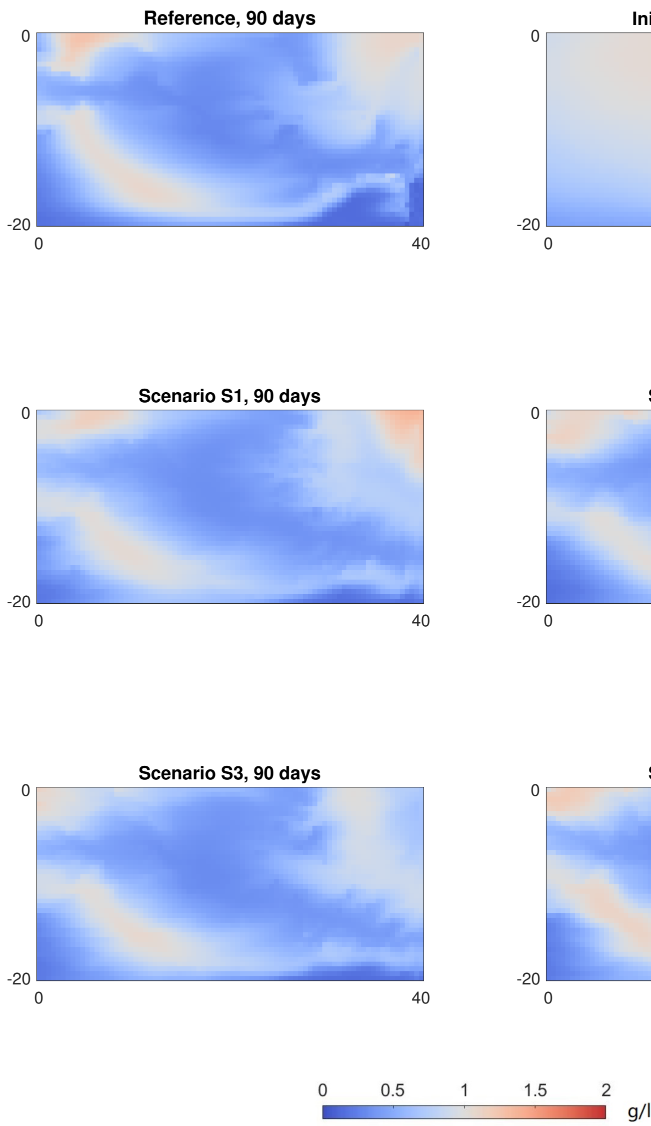

Figure 9: The reference contaminant plume and the ensemble mean contaminant plumes simulated with the initial set of CSAC parameters and with the updated CSAC values for scenario S1-S4 on day 90.

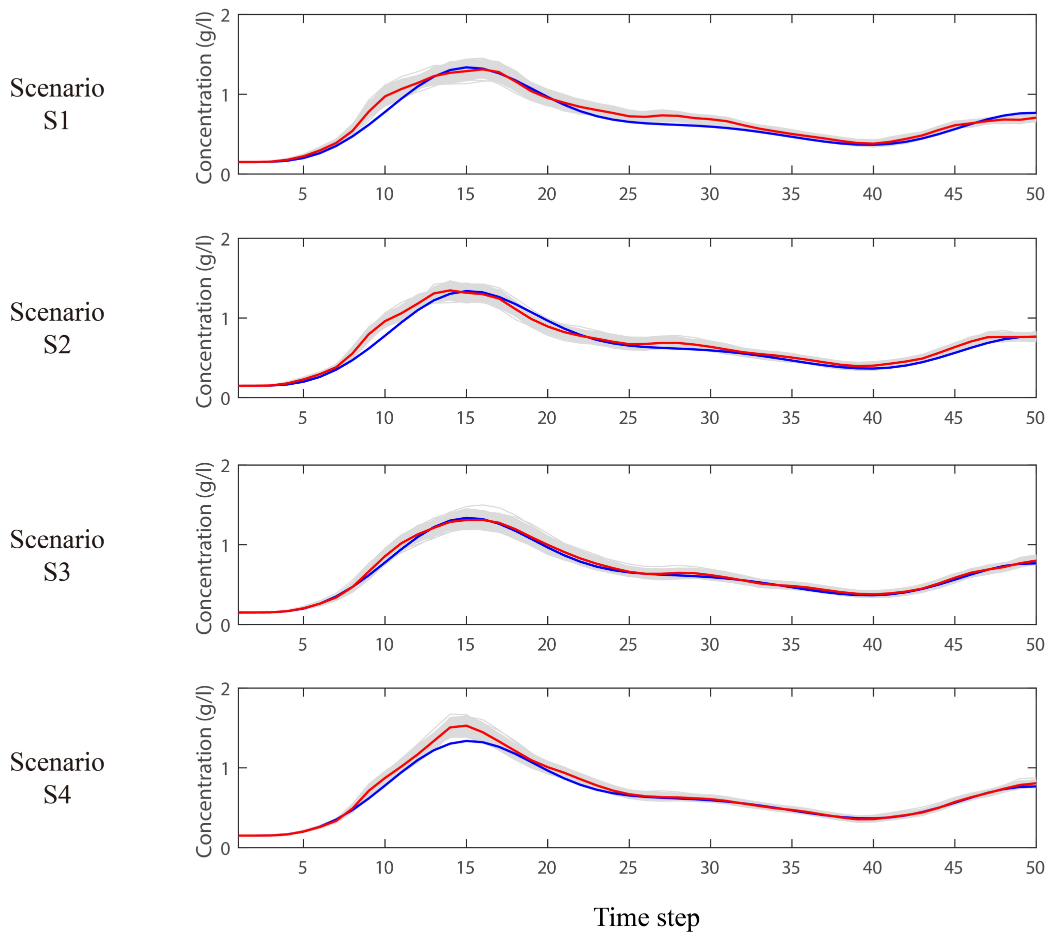

Figure 10: Time evolution of the vertically-averaged concentration of the recovered plume in borehole 1 for scenario S1-S4. The blue curve stands for the concentration evolution in the reference, while each gray line represents one concentration curve for one ensemble member, and the red line denotes the ensemble median.

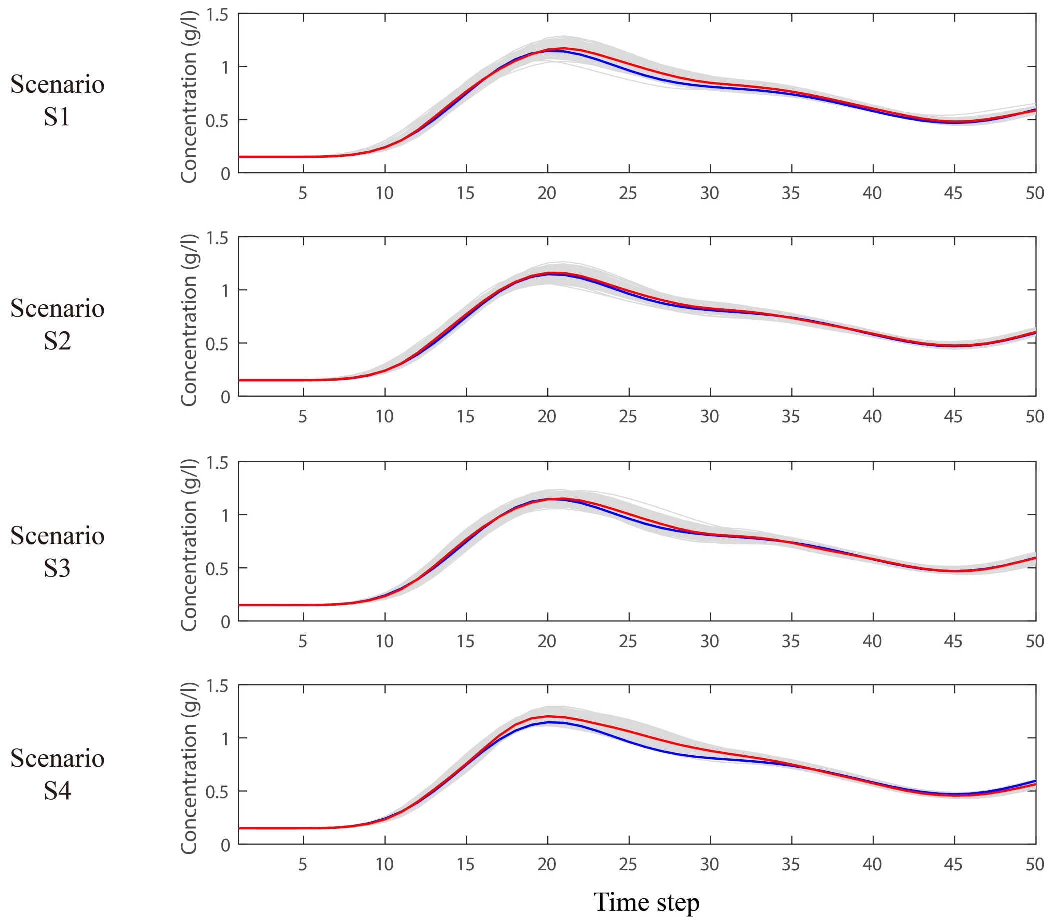

Figure 11: Time evolution of the vertically-averaged concentration of the recovered plume in borehole 2 for scenario S1-S4. The blue curve stands for the concentration evolution in the reference, while each gray line represents one concentration curve for one ensemble member, and the red line denotes the ensemble median.

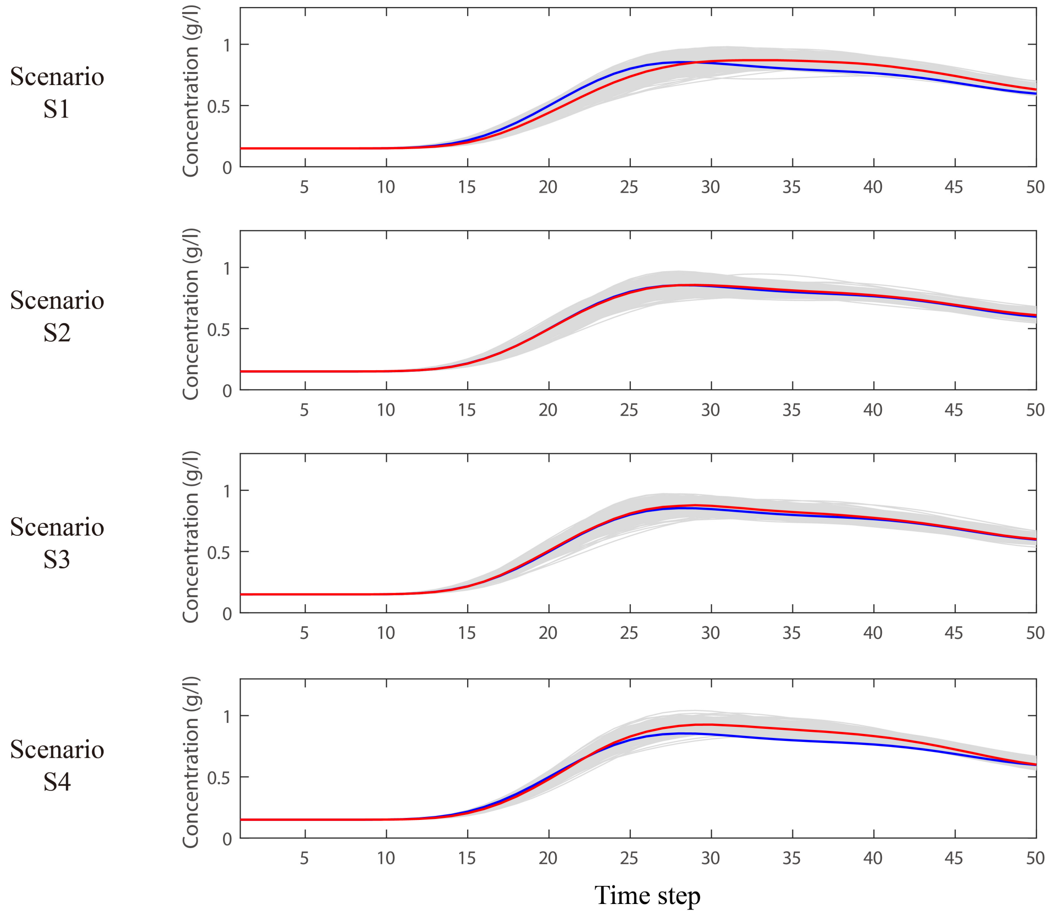

Figure 12: Time evolution of the vertically-averaged concentration of the recovered plume in borehole 3 for scenario S1-S4. The blue curve stands for the concentration evolution in the reference, while each gray line represents one concentration curve for one ensemble member, and the red line denotes the ensemble median.

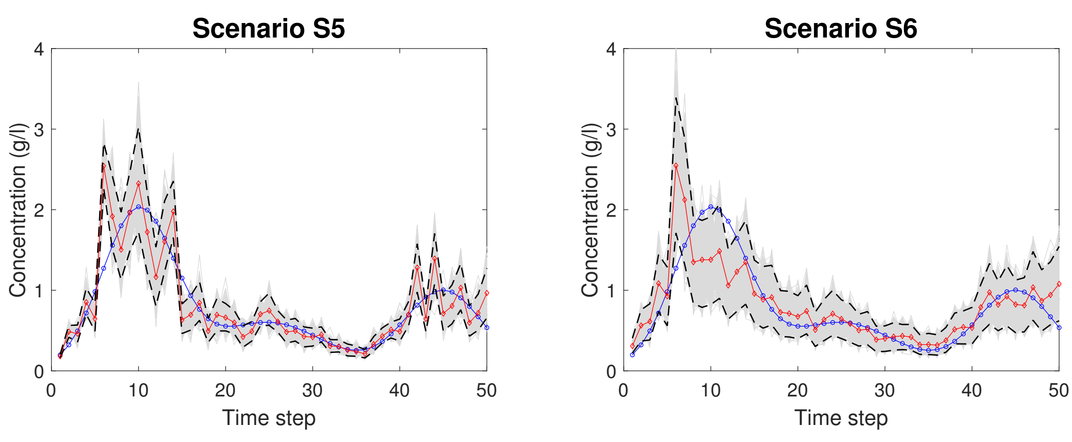

Figure 13: Recovered time-varying release histories for scenario S5-S6. The blue dotted line illustrates the reference time-varying release history, each gray line represents one recovered release history curve for one ensemble member, the red dotted line stands for the median, while the black dashed lines denote the 5 and 95 percentiles.

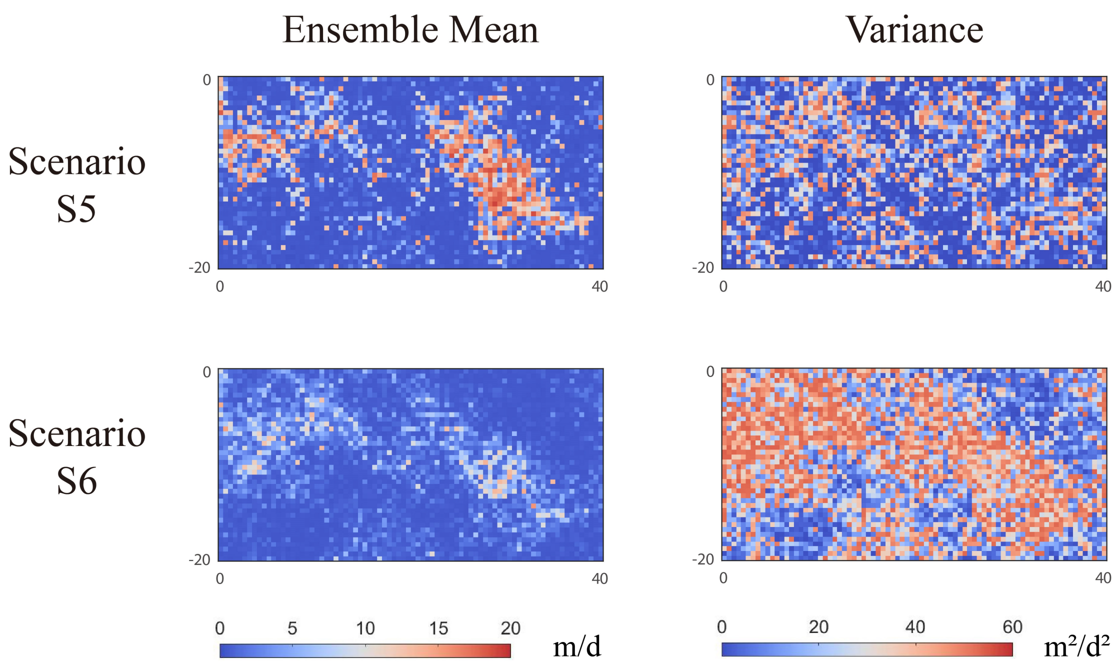

Figure 14: Ensemble means and variances of hydraulic conductivities for scenario S5-S6.

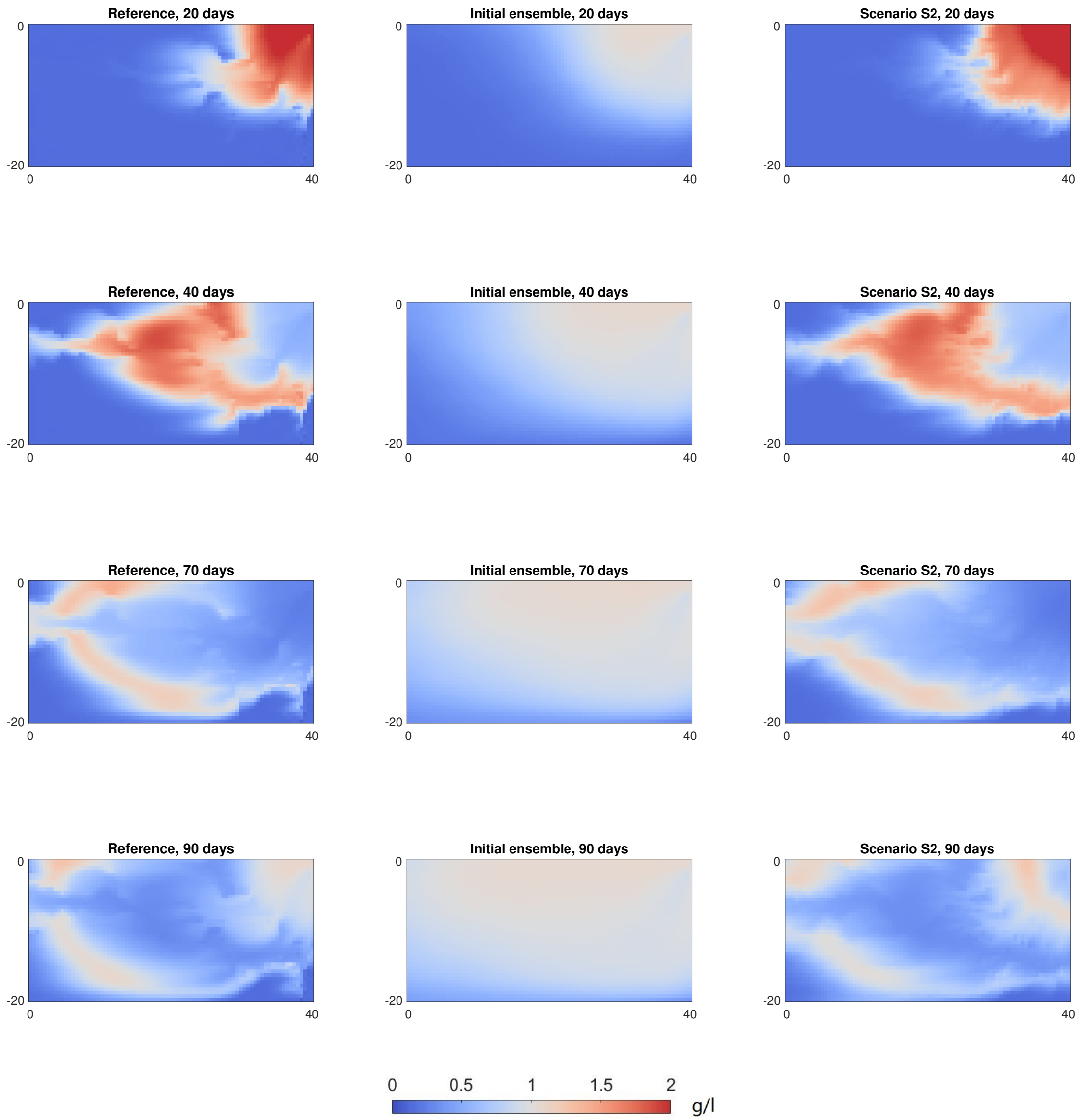

Figure 15: The ensemble mean contaminant plumes computed with the updated CSAC values for scenario S5-S6 on day 20, 40, 70 and 90.

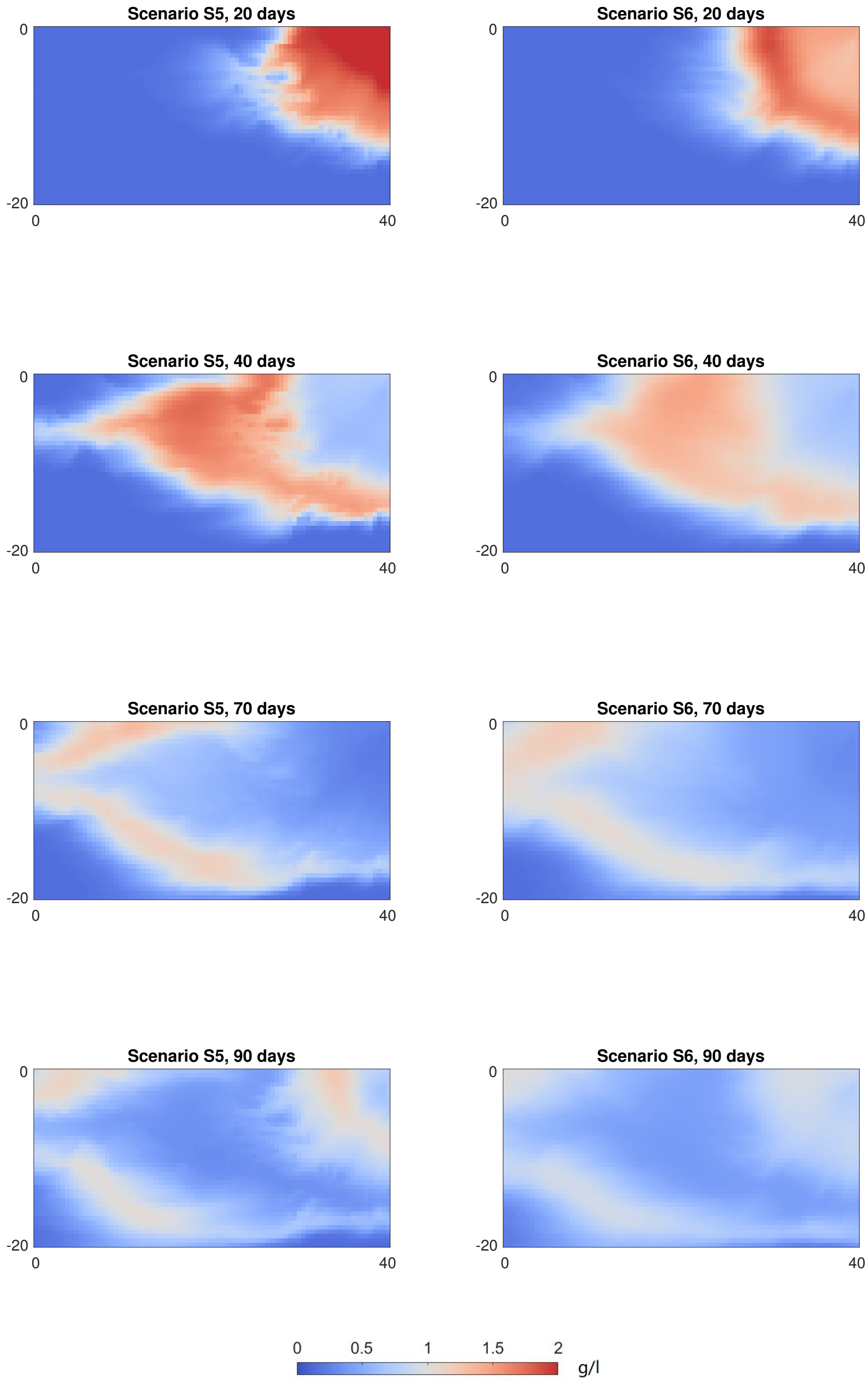

Figure 16: The ensemble mean contaminant plumes computed with the updated CSAC values for scenario S5-S6 on day 20, 40, 70 and 90.

5. Discussion

In this paper, we presented a performance evaluation of a proposed data assimilation framework for the groundwater contamination inverse problem in a benchmark case. The demonstration mainly presented through the following three aspects: (1) establish a coupled groundwater flow, transport solute and geophysical model; (2) build a data assimilation framework based on ES-MDA and ERT observation data; (3) joint identify contaminant source release history and hydraulic conductivities.

We show that the proposed data assimilation framework works well by using time-lapse ERT measurements with the proper electrode array configuration. Although the inversion of the release history may suffer from filter inbreeding, the reproduced contaminant plume via the updated CSAC parameters is still able to capture the dominant pattern of the reference contaminant plume. This result may have important implications for groundwater contamination modeling, as it suggests that an appropriate number of downstream ERT observations could help the researchers jointly identify the aquifer properties and unknown contaminant release history.

In this work, the electrical properties of the fine and coarse sands we used are assumed to be perfectly known. However, in field works, these parameters need to be ascertained according to dedicated field experiments before our proposed data assimilation framework is employed. While it may be argued that we have simplified the geophysical properties of the electrical model, the proposed data assimilation framework can be used as a starting point, leaving to the ES-MDA the task of evaluating the impact of uncertainties and heterogeneity around the electrical properties. In fact, this uncertainty evaluation procedure can be applied not only to the aquifer electrical properties but also to the apparent resistivity observation errors. This attribute is also indicative of the merits inherent in ensemble-based algorithms.

Besides, to characterize the contaminant source release history and hydraulic conductivities, the difference between the background ionic concentration and the contaminant release is also crucial. The temporal variation in apparent resistivity is an essential prerequisite for the proposed data assimilation framework. Hence, this data assimilation framework may suffer from larger uncertainty issues in cases with similar background ionic concentrations and contaminant releases.

In summary, we conclude that the proposed data assimilation framework has the capacity to identify contaminant source release history and hydraulic conductivities by using time-lapse ERT measurements with proper electrode array configuration. Moreover, it is refreshing to see that this data assimilation framework also has the capacity to handle a more complex geophysical model with large apparent resistivity observation errors.

6. Summary and conclusion

The main objective of this study is to joint identify a contaminant source release curve and hydraulic conductivities by using time-lapse ERT measurements. For this purpose, the study combines a coupled model of groundwater flow, solute transport and geophysics with the ES-MDA assimilation technique. In this data assimilation framework, only the apparent resistivity obtained from the time-lapse ERT measurements is utilized to recognize the hydraulic conductivities and contaminant release history. The proposed methodology is then validated in a synthetic benchmark with a time-varying contaminant source release in a heterogeneous aquifer. The results demonstrate that the CSAC problem could be handled by the proposed approach. The time-varying release history and the main patterns of high conductivity can be captured with proper time-lapse ERT measurements. The plume evolution computed with the updated parameters for both time-varying release curve and spatially-heterogeneous conductivity approximates well the plume computed in the reference field.

Besides, we also analyzed the influence of different AM-BN schemes in our data assimilation framework. Four scenarios with a different number of apparent resistivity measurements

(98, 128, 162 and 388) were designed. In scenario S4, the AM-BN scheme with a mixture of three different electrode spacing suffers from filter inbreeding producing poor results. This phenomenon can be primarily ascribed to the underestimation of ensemble variance of the release history, and needs additional attentions in further applications. We also evaluate the impact of uncertainties in the petrophysical properties and geophysical observations on our data assimilation framework, the outcomes show that the proposed ES-MDA data assimilation framework could provide convincing inversion of time-varying releasing history and hydraulic conductivities.

This study is significant since it is the first time that time-lapse ERT measurements are employed to identify contaminant source information together with hydraulic conductivities. And we believe this work also provides a way to assess the uncertainties from different sources when we face a more close to reality case. Insights from this work could provide a solid basis for more geophysical technologies applied in the future identification of contaminant source information and aquifer properties. But we also admit that a number of issues have not been considered, such as different electrode-array configurations, or more complex hydraulic systems. More research is needed in order to move forward and apply this approach to real problems.

7. Acknowledgments

Financial support to carry out this work was received from grants PID2019-109131RB-I00 and PRX17/00150 funded by MCIN/AEI/10.13039/501100011033, grant DD20190260 and DD20221728 funded by China Geological Survey and grant 2022YFC3705001 funded by National Key R&D Plan. Teng Xu acknowledges the National Natural Science Foundation of China (42377046 and 523024911). The authors acknowledge Graham Sander, Associate Editor and three anonymous Reviewers for their thoughtful and constructive comments.

References

Ababou R, Bagtzoglou AC, Mallet A. Anti-diffusion and source identification with the 'RAW' scheme: A particle-based censored random walk. Environmental Fluid Mechanics 2010;10(1):41–76. doi:10.1007/s10652-009-9153-4.

Atmadja J, Bagtzoglou AC. State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification. Environmental Forensics 2001;2(3):205–14. URL: http://www.sciencedirect.com/science/article/pii/S1527592201900552. doi:http://dx.doi.org/10.1006/enfo.2001.0055.

Ayvaz MT. A linked simulation-optimization model for solving the unknown groundwater pollution source identification problems. Journal of Contaminant Hydrology 2010;117(1-4):46–59. URL: http://dx.doi.org/10.1016/j.jconhyd.2010.06.004. doi:10.1016/j.jconhyd.2010.06.004.

Bagtzoglou AC, Atmadja J. Marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: Application to contaminant plume spatial distribution recovery. Water Resources Research 2003;39(2):1–14. doi:10.1029/2001WR001021.

Bagtzoglou AC, Atmadja J. Mathematical Methods for Hydrologic Inversion: The Case of Pollution Source Identification. Water Pollution 2005;5:65–96. URL: http://www.springerlink.com/index/10.1007/b11442. doi:10.1007/b11442.

Bagtzoglou AC, Dougherty DE, Tompson AFB. Application of particle methods to reliable identification of groundwater pollution sources. Water Resources Management 1992;6(1):15–23. doi:10.1007/BF00872184.

Bao J, Li L, Redoloza F. Coupling ensemble smoother and deep learning with generative adversarial networks to deal with non-Gaussianity in flow and transport data assimilation. Journal of Hydrology 2020;590(August):125443. URL: https://doi.org/10.1016/j.jhydrol.2020.125443. doi:10.1016/j.jhydrol.2020.125443.

Bear J. Dynamics of Fluids in Porous Media. American Elsevier, 1972.

Bedekar V, Morway ED, Langevin CD, Tonkin MJ. MT3D-USGS version 1: A US Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW. Technical Report; US Geological Survey; 2016.

Binley A, Hubbard SS, Huisman JA, Revil A, Robinson DA, Singha K, Slater LD. The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. Water Resources Research 2015;51(6):3837–66. URL: https://doi.org/10.1002/2015WR017016. doi:https://doi.org/10.1002/2015WR017016.

Binley A, Kemna A. DC resistivity and induced polarization methods. In: Hydrogeophysics. Springer; 2005. p. 129–56.

Blanchy G, Saneiyan S, Boyd J, McLachlan P, Binley A. ResIPy, an intuitive open source software for complex geoelectrical inversion/modeling. Computers and Geosciences 2020;137(February):104423. URL: https://doi.org/10.1016/j.cageo.2020.104423. doi:10.1016/j.cageo.2020.104423.

Bouzaglou V, Crestani E, Salandin P, Gloaguen E, Camporese M. Ensemble Kalman filter assimilation of ERT data for numerical modeling of seawater intrusion in a laboratory experiment. Water (Switzerland) 2018;10(4):1–26. doi:10.3390/w10040397.

Brunetti C, Linde N. Impact of petrophysical uncertainty on Bayesian hydrogeophysical inversion and model selection. Advances in Water Resources 2018;111(November 2017):346–59. URL: https://doi.org/10.1016/j.advwatres.2017.11.028. doi:10.1016/j.advwatres.2017.11.028.

Butera I, Tanda MG, Zanini A. Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach. Stochastic Environmental Research and Risk Assessment 2013;27(5):1269–80. doi:10.1007/s00477-012-0662-1.

Cao Z, Li L, Chen K. Bridging iterative Ensemble Smoother and multiple-point geostatistics for better flow and transport modeling. Journal of Hydrology 2018;565(August):411–21. URL: https://doi.org/10.1016/j.jhydrol.2018.08.023. doi:10.1016/j.jhydrol.2018.08.023.

Capilla JE, Gömez-Hernández JJ, Sahuquillo A. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric head data—3. application to the culebra formation at the waste isolation pilot plan (wipp), new mexico, usa. Journal of Hydrology 1998;207(3-4):254–69.

Capilla JE, Rodrigo J, Gómez-Hernández JJ. Simulation of non-gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. Mathematical Geology 1999;31(7):907–27.

Carrera J, Neuman SP. Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information. Water Resources Research 1986;22(2):199–210. doi:10.1029/WR022i002p00199.

Chen Z, Gómez-Hernández JJ, Xu T, Zanini A. Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter. Journal of Hydrology 2018;564:1074–84. doi:10.1016/j.jhydrol.2018.07.073.

Chen Z, Xu T, Gómez-Hernández JJ, Zanini A. Contaminant Spill in a Sandbox with Non-Gaussian Conductivities: Simultaneous Identification by the Restart Normal-Score

Ensemble Kalman Filter. Mathematical Geosciences 2021;53(7):1587–615. URL: https://doi.org/10.1007/s11004-021-09928-y. doi:10.1007/s11004-021-09928-y.

Chen Z, Xu T, Gómez-Hernández JJ, Zanini A, Zhou Q. Reconstructing the release history of a contaminant source with different precision via the ensemble smoother with multiple data assimilation. Journal of Contaminant Hydrology 2022;.

Crestani E, Camporese M, Baú D, Salandin P. Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. Hydrology and Earth System Sciences 2013;17(4):1517–31. doi:10.5194/hess-17-1517-2013.

Crestani E, Camporese M, Salandin P. Assessment of hydraulic conductivity distributions through assimilation of travel time data from ERT-monitored tracer tests. Advances in Water Resources 2015;84:23–36. URL: http://dx.doi.org/10.1016/j.advwatres.2015.07.022. doi:10.1016/j.advwatres.2015.07.022.

Cupola F, Tanda MG, Zanini A. Laboratory sandbox validation of pollutant source location methods. Stochastic Environmental Research and Risk Assessment 2015;29(1):169–82. doi:10.1007/s00477-014-0869-4.

Dodangeh A, Rajabi MM, Carrera J, Fahs M. Joint identification of contaminant source characteristics and hydraulic conductivity in a tide-influenced coastal aquifer. Journal of Contaminant Hydrology 2022;247(January):103980. URL: https://doi.org/10.1016/j.jconhyd.2022.103980. doi:10.1016/j.jconhyd.2022.103980.

Emerick AA, Reynolds AC. Ensemble smoother with multiple data assimilation. Computers and Geosciences 2013;55:3–15. URL: http://dx.doi.org/10.1016/j.cageo.2012.03.011. doi:10.1016/j.cageo.2012.03.011.

Evensen G. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. Ocean Dynamics 2003;53(4):343–67. doi:10.1007/s10236-003-0036-9.

Evensen G. Analysis of iterative ensemble smoothers for solving inverse problems. Computational Geosciences 2018;22(3):885–908. URL: http://link.springer.com/10.1007/s10596-018-9731-y. doi:10.1007/s10596-018-9731-y.

Feyen L, Gómez-Hernández J, Ribeiro Jr P, Beven KJ, De Smedt F. A bayesian approach to stochastic capture zone delineation incorporating tracer arrival times, conductivity measurements, and hydraulic head observations. Water resources research 2003;39(5).

Franssen H, Gómez-Hernández J. 3d inverse modelling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations. Stochastic Environmental Research and Risk Assessment 2002;16(2):155–74.

Gómez-Hernández J, Franssen HJH, Sahuquillo A. Stochastic conditional inverse modeling of subsurface mass transport: a brief review and the self-calibrating method. Stochastic Environmental Research and Risk Assessment 2003;17(5):319–28.

Gómez-Hernández J, Wen XH. Probabilistic assessment of travel times in groundwater modeling. Stochastic Hydrology and Hydraulics 1994;8(1):19–55.

Gómez-Hernández JJ, Xu T. Contaminant Source Identification in Aquifers: A Critical View. Mathematical Geosciences 2021;:1–22.

Gorelick SM, Evans B, Remson I. Identifying sources of groundwater pollution: An optimization approach. Water Resources Research 1983;19(3):779–90. doi:10.1029/WR019i003p00779.

Hanea R, Evensen G, Hustoft L, Ek T, Chitu A, Wilschut F. Reservoir management under geological uncertainty using fast model update. Society of Petroleum Engineers - SPE Reservoir Simulation Symposium 2015 2015;3:1912–23. doi:10.2118/173305-ms.

Harbaugh AW. MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process. volume 6. US Department of the Interior, US Geological Survey Reston, VA, USA, 2005.

Hendricks Franssen HJ, Kinzelbach W. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resources Research 2008;44(9):1–21. doi:10.1029/2007WR006505.

Jafarpour B, Khodabakhshi M. A Probability Conditioning Method (PCM) for Nonlinear Flow Data Integration into Multipoint Statistical Facies Simulation. Mathematical Geosciences 2011;43(2):133–64. URL: https://doi.org/10.1007/s11004-011-9316-y. doi:10.1007/s11004-011-9316-y.

Jardani A, Revil A, Dupont JP. Stochastic joint inversion of hydrogeophysical data for salt tracer test monitoring and hydraulic conductivity imaging. Advances in Water Resources 2013;52:62–77. URL: http://dx.doi.org/10.1016/j.advwatres.2012.08.005. doi:10.1016/j.advwatres.2012.08.005.

Jiang X, Ma R, Wang Y, Gu W, Lu W, Na J. Two-stage surrogate model-assisted Bayesian framework for groundwater contaminant source identification. Journal of Hydrology 2021;594(July 2020):125955. URL: https://doi.org/10.1016/j.jhydrol.2021.125955. doi:10.1016/j.jhydrol.2021.125955.

Journel AG, Isaaks EH. Conditional Indicator Simulation: Application to a Saskatchewan uranium deposit. Journal of the International Association for Mathematical Geology 1984;16(7):685–718. doi:10.1007/BF01033030.

Kang X, Shi X, Deng Y, Revil A, Xu H, Wu J. Coupled hydrogeophysical inversion of DNAPL source zone architecture and permeability field in a 3D heterogeneous sandbox by assimilation time-lapse cross-borehole electrical resistivity data via ensemble Kalman filtering. Journal of Hydrology 2018;567:149–64. URL: https://doi.org/10.1016/j.jhydrol.2018.10.019. doi:10.1016/j.jhydrol.2018.10.019.

Kang X, Shi X, Revil A, Cao Z, Li L, Lan T, Wu J. Coupled hydrogeophysical inversion to identify non-Gaussian hydraulic conductivity field by jointly assimilating geochemical and time-lapse geophysical data. Journal of Hydrology 2019;578(August):124092. URL: https://doi.org/10.1016/j.jhydrol.2019.124092. doi:10.1016/j.jhydrol.2019.124092.

Kumar D, Srinivasan S. Ensemble-Based Assimilation of Nonlinearly Related Dynamic Data in Reservoir Models Exhibiting Non-Gaussian Characteristics. Mathematical Geosciences 2019;51(1):75–107. URL: https://doi.org/10.1007/s11004-018-9762-x. doi:10.1007/s11004-018-9762-x.

Kumar D, Srinivasan S. Indicator-based data assimilation with multiple-point statistics for updating an ensemble of models with non-Gaussian parameter distributions. Advances in Water Resources 2020;141:103611. URL: http://www.sciencedirect.com/science/article/pii/S0309170819309297. doi:https://doi.org/10.1016/j.advwatres.2020.103611.

Laloy E, Linde N, Vrugt JA. Mass conservative three-dimensional water tracer distribution from Markov chain Monte Carlo inversion of time-lapse ground-penetrating radar data. Water Resources Research 2012;48(7):1–15. doi:10.1029/2011WR011238.

Le DH, Emerick AA, Reynolds AC. An Adaptive Ensemble Smoother With Multiple Data Assimilation for Assisted History Matching. SPE Journal 2016;21(06):2195–207. URL: https://doi.org/10.2118/173214-PA. doi:10.2118/173214-PA.

Li J, Lu W, Wang H, Fan Y. Identification of groundwater contamination sources using a statistical algorithm based on an improved Kalman filter and simulation optimization.

Hydrogeology Journal 2019;27(8):2919–31. URL: http://link.springer.com/10.1007/s10040-019-02030-y. doi:10.1007/s10040-019-02030-y.

Li L, Zhou H, Gómez-Hernández JJ. A comparative study of three-dimensional hydraulic conductivity upscaling at the macro-dispersion experiment (made) site, columbus air force base, mississippi (usa). Journal of Hydrology 2011a;404(3-4):278–93.

Li L, Zhou H, Gómez-Hernández JJ. Transport upscaling using multi-rate mass transfer in three-dimensional highly heterogeneous porous media. Advances in Water Resources 2011b;34(4):478–89.

Li L, Zhou H, Hendricks Franssen HJ, Gómez-Hernández JJ. Groundwater flow inverse modeling in non-MultiGaussian media: Performance assessment of the normal-score Ensemble Kalman Filter. Hydrology and Earth System Sciences 2012;16(2):573–90. doi:10.5194/hess-16-573-2012.

Mao D, Lu L, Revil A, Zuo Y, Hinton J, Ren ZJ. Geophysical Monitoring of Hydrocarbon-Contaminated Soils Remediated with a Bioelectrochemical System. Environmental Science and Technology 2016;50(15):8205–13. doi:10.1021/acs.est.6b00535.

Megdal SB. Invisible water: the importance of good groundwater governance and management. npj Clean Water 2018;1(1):1–5. URL: http://dx.doi.org/10.1038/s41545-018-0015-9. doi:10.1038/s41545-018-0015-9.

Michalak AM, Kitanidis PK. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. Water Resources Research 2004;40(8). doi:10.1029/2004WR003214.

Mirghani BY, Mahinthakumar KG, Tryby ME, Ranjithan RS, Zechman EM. A parallel evolutionary strategy based simulation-optimization approach for solving groundwater source identification problems. Advances in Water Resources 2009;32(9):1373–85. URL: http://dx.doi.org/10.1016/j.advwatres.2009.06.001. doi:10.1016/j.advwatres.2009.06.001.

Nenna V, Pidlisecky A, Knight R. Application of an extended Kalman filter approach to inversion of time-lapse electrical resistivity imaging data for monitoring recharge. Water Resources Research 2011;47(10):1–13. doi:10.1029/2010WR010120.

Neupauer RM, Wilson JL. Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resources Research 1999;35(11):3389–98. doi:10.1029/1999WR900190.

Panjehfouladgaran A, Rajabi MM. Contaminant source characterization in a coastal aquifer influenced by tidal forces and density-driven flow. Journal of Hydrology 2022;610(April):127807. URL: https://doi.org/10.1016/j.jhydrol.2022.127807. doi:10.1016/j.jhydrol.2022.127807.

Pirot G, Krityakierne T, Ginsbourger D, Renard P. Contaminant source localization via Bayesian global optimization. Hydrology and Earth System Sciences 2019;23(1):351–69. doi:10.5194/hess-23-351-2019.

Power C, Gerhard JI, Tsourlos P, Giannopoulos A. A new coupled model for simulating the mapping of dense nonaqueous phase liquids using electrical resistivity tomography. Geophysics 2013;78(4). doi:10.1190/GEO2012-0395.1.

Rafiee J, Reynolds AC. Theoretical and efficient practical procedures for the generation of inflation factors for ES-MDA. Inverse Problems 2017;33(11). doi:10.1088/1361-6420/aa8cb2.

Revil A. On charge accumulation in heterogeneous porous rocks under the influence of an external electric field. Geophysics 2013;78(4). doi:10.1190/GEO2012-0503.1.

Revil A, Qi Y, Ghorbani A, Soueid Ahmed A, Ricci T, Labazuy P. Electrical conductivity and induced polarization investigations at Krafla volcano, Iceland. Journal of Volcanology and Geothermal Research 2018;368:73–90. doi:10.1016/j.jvolgeores.2018.11.008.

Revil A, Sleevi MF, Mao D. Induced polarization response of porous media with metallic particles - part 5: Influence of the background polarization. Geophysics 2017;82(2):E77–96. URL: http://dx.doi.org/10.1190/geo2016-0388.1. doi:10.1190/GEO2016-0388.1.

Seferou P, Soupios P, Kourgialas NN, Dokou Z, Karatzas GP, Candasayar E, Papadopoulos N, Dimitriou V, Sarris A, Sauter M. Olive-oil mill wastewater transport under unsaturated and saturated laboratory conditions using the geoelectrical resistivity tomography method and the FEFLOW model. Hydrogeology Journal 2013;21(6):1219–34. doi:10.1007/s10040-013-0996-x.

Sen PN. Influence of temperature on electrical conductivity on shaly sands 1992;57(1):89–96.

Shao S, Guo X, Gao C, Liu H. Quantitative relationship between the resistivity distribution of the by-product plume and the hydrocarbon degradation in an aged hydrocarbon contaminated site. Journal of Hydrology 2021;596(February):126122. URL: https://doi.org/10.1016/j.jhydrol.2021.126122. doi:10.1016/j.jhydrol.2021.126122.

Skaggs TH, Kabala ZJ. Recovering the release history of a groundwater contaminant. Water Resources Research 1994;30(1):71–9. URL: http://doi.wiley.com/10.1029/93WR02656. doi:10.1029/93WR02656.

Sun AY, Painter SL, Wittmeyer GW. A constrained robust least squares approach for contaminant release history identification. Water Resources Research 2006;42(4):1–13. doi:10.1029/2005WR004312.

Todaro V, D'Oria M, Tanda MG, Gómez-Hernández JJ. Ensemble smoother with multiple data assimilation to simultaneously estimate the source location and the release history of a contaminant spill in an aquifer. Journal of Hydrology 2021;598(April). doi:10.1016/j.jhydrol.2021.126215.

Troldborg M, Nowak W, Tuxen N, Bjerg PL, Helmig R, Binning PJ. Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework. Water Resources Research 2010;46(1):1–19. doi:10.1029/2010WR009227.

Tso ChM, Johnson TC, Song X, Chen X, Kuras O, Wilkinson P, Uhlemann S, Chambers J, Binley A, Centre LE. Integrated hydrogeophysical modelling and data assimilation for geoelectrical leak detection. Journal of Contaminant Hydrology 2020;234(July):103679. URL: /doi.org/10.1016/j.jconhyd.2020.103679. doi:10.1016/j.jconhyd.2020.103679.

Wen XH, Capilla JE, Deutsch C, Gómez-Hernández J, Cullick A. A program to create permeability fields that honor single-phase flow rate and pressure data. Computers & Geosciences 1999;25(3):217–30.

Woodbury AD, Ulrych TJ. Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant. Water Resources Research 1996;32(9):2671–81.

Xia T, Dong Y, Mao D, Meng J. Delineation of LNAPL contaminant plumes at a former perfumery plant using electrical resistivity tomography. Hydrogeology Journal 2021;8(1):1189–201.

Xu T, Gómez-Hernández JJ. Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resources Research 2016;doi:10.1002/2014WR016618.Received.

Xu T, Gómez-Hernández JJ. Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter. Advances in Water Resources 2018;112(July 2017):106–23. URL: https://doi.org/10.1016/j.advwatres.2017.12.011. doi:10.1016/j.advwatres.2017.12.011.

Xu T, Gómez-Hernández JJ, Chen Z, Lu C. A comparison between ES-MDA and restart EnKF for the purpose of the simultaneous identification of a contaminant source and hydraulic conductivity. Journal of Hydrology 2021;595:125681. URL: https://www.sciencedirect.com/science/article/pii/S0022169420311422. doi:https://doi.org/10.1016/j.jhydrol.2020.125681.

Xu T, Zhang W, Gómez-Hernández JJ, Xie Y, Yang J, Chen Z, Lu C. Non-point contaminant source identification in an aquifer using the ensemble smoother with multiple data assimilation. Journal of Hydrology 2022;606(January):127405. doi:10.1016/j.jhydrol.2021.127405.

Zeng L, Shi L, Zhang D, Wu L. A sparse grid based Bayesian method for contaminant source identification. Advances in Water Resources 2012;37:1–9. URL: http://dx.doi.org/10.1016/j.advwatres.2011.09.011. doi:10.1016/j.advwatres.2011.09.011.

Zheng C, Wang PP. MT3DMS: A Modular Three-Dimensional Multispecies Transport Model 1999;(December):219.

Zhou B, Greenhalgh SA. Cross-hole resistivity tomography using different electrode configurations. Geophysical Prospecting 2000;48(5):887–912. doi:10.1046/j.1365-2478.2000.00220.x.

Zhou H, Gómez-Hernández JJ, Hendricks Franssen HJ, Li L. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Advances in Water Resources 2011;34(7):844–64. URL: http://dx.doi.org/10.1016/j.advwatres.2011.04.014. doi:10.1016/j.advwatres.2011.04.014.

Zhou H, Gómez-Hernández JJ, Li L. Inverse methods in hydrogeology: Evolution and recent trends. Advances in Water Resources 2014;63:22–37. URL: http://dx.doi.org/10.1016/j.advwatres.2013.10.014. doi:10.1016/j.advwatres.2013.10.014.