Joint Estimation of Hydraulic and Biochemical Parameters for Reactive Transport Modelling with a Modified ILUES Algorithm

Ruicheng Zhang [1], Nianqing Zhou [1], Xuemin Xia [1], Guoxian Zhao [2] and Simin Jiang [1,3,*]

[1] Department of Hydraulic Engineering, Tongji University, Shanghai 200092, China; 1810038@tongji.edu.cn (R.Z.); nq.zhou@tongji.edu.cn (N.Z.); 1510256@tongji.edu.cn (X.X.)

[2] School of Materials Science and Engineering, Xi'an Shiyou University, Xi'an 710065, China; gxzhao@xsyu.edu.cn

[3] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

[*] Correspondence: jiangsimin@tongji.edu.cn

Abstract: Multicomponent reactive transport modeling is a powerful tool for the comprehensive analysis of coupled hydraulic and biochemical processes. The performance of the simulation model depends on the accuracy of related model parameters whose values are usually difficult to determine from direct measurements. In this situation, estimates of these uncertain parameters can be obtained by solving inverse problems. In this study, an efficient data assimilation method, the iterative local updating ensemble smoother (ILUES), is employed for the joint estimation of hydraulic parameters, biochemical parameters and contaminant source characteristics in the sequential biodegradation process of tetrachloroethene (PCE). In the framework of the ILUES algorithm, parameter estimation is realized by updating local ensemble with the iterative ensemble smoother (IES). To better explore the parameter space, the original ILUES algorithm is modified by determining the local ensemble partly with a linear ranking selection scheme. Numerical case studies based on the sequential biodegradation of PCE are then used to evaluate the performance of the ILUES algorithm. The results show that the ILUES algorithm is able to achieve an accurate joint estimation of related model parameters in the reactive transport model.

Keywords: reactive transport model; parameter estimation; iterative local updating ensemble smoother; data assimilation

1. Introduction

Tetrachloroethene (PCE), a common chlorinated solvent, has been widely used as a dry-cleaning solvent and a metal degreaser. It is among the most widespread and persistent organic contaminants in groundwater resources [1]. PCE is also known as a dense non-aqueous phase liquid (DNAPL) due to its high density and low solubility in groundwater. As a result of these properties, traditional remediation approaches, such as pump-and-treat and vapor stripping, are very expensive and sometimes inapplicable for PCE contaminants in subsurface environments [2]. Therefore, in-situ bioremediation is considered to be a promising method for site remediation of chlorinated solvents, because it requires less intervention and is more cost-effective [3–6].

Multicomponent reactive transport modeling is an important tool for predicting the fate and transport of reactive contaminants in groundwater by integrating simulations of water flow, biochemical reaction and solute transport [7,8]. However, the uncertainty in related parameters of the reactive transport model poses a major challenge to reliable simulations [9]. In practice, the spatially

variable aquifer properties such as hydraulic conductivity and porosity have significant effects on groundwater flow and solute transport modeling [10–12]. Additionally, the reactive transport model that characterizes the reaction and migration of contaminants in subsurface aquifers is quite sensitive to biochemical parameters (e.g., rate constant) and contaminant source characteristics (e.g., source strength, source location) [13,14]. Therefore, the simultaneous estimation of hydraulic parameters, biochemical parameters and contaminant source characteristics are of vital importance to achieve satisfactory accuracy of the forward model.

Direct determination of specific model parameters is usually infeasible in heterogeneous subsurface environments. To solve this problem, information obtained from indirect measurements (e.g., contaminant concentration and hydraulic head) is used to estimate uncertain model parameters by solving inverse problems [15]. Many researchers have focused on the calibration of key parameters of reactive transport models using inverse methodologies in the past few years. Bailey and Baù [14] employed an ensemble Kalman filter (EnKF) method to estimate the spatial distribution of the first-order rate constant of a reactive transport model by assimilating the concentration data. Bailey et al. [16] estimated the spatially variable denitrification rates in an agricultural groundwater system by assimilating the concentration and mass data of nitrate, which is implemented with the help of the Ensemble Smoother (ES). Dai and Samper [17] investigated the ability of a general methodology for solving inverse problems in a column experiment and found it useful in the estimation of model parameters. Carniato et al. [18] compared the performance of weighted least squares (WLS), weighted least squares with weight estimation (WLS(we)) and a multivariate (MV) approach on parameter estimation of reactive transport models and found that residual correlation did not evidently affect the predictive uncertainty. Nevertheless, the joint estimation of various parameters of reactive transport models is still limited to a few studies [19–21].

In recent years, Bayesian inference has become a prevalent method for parametric uncertainty quantification [22–24]. According to Bayes' theorem, the unknown parameters are treated as random variables whose posterior probability distributions can be calculated by updating the prior probability with information provided by the observations [25]. For highly nonlinear systems, the posterior distributions of unknown variables are difficult and in most cases impossible to evaluate analytically [26]. To address this issue, Markov Chain Monte Carlo (MCMC) methods [27,28] have been widely used to approximate the posterior distribution of unknown parameters by generating random samples from the probability distribution. Cao et al. [29] integrated the multi-try differential evolution adaptive Metropolis (MT-DREAMzs) algorithm with the nested sampling estimator (NSE) to improve the performance of marginal likelihood estimation. Shi et al. [30] tested the validity of Gaussian assumptions for the parametric uncertainty quantification in a reactive transport model. Vrugt et al. [31] proposed the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) to inverse the hydrologic model parameters and found it more efficient than traditional Metropolis-Hastings samplers. Yan et al. [32] developed a Bayesian-based integrated approach, in which a Kriging surrogate model (KSM) is used to improve the computational efficiency to identify the groundwater contamination sources. Nevertheless, MCMC methods are considered to be computationally expensive because a large number of model evaluations are required for high-dimensional parameter space exploration [33]. Consequently, the applications of MCMC methods to inverse modeling of reactive transport are still limited [34,35].

The ensemble Kalman filter (EnKF), which is a Monte Carlo implementation of the Kalman filter [36], is a computationally efficient alternative to the MCMC methods. The EnKF algorithm proposed by Evensen [37] is able to update model parameters and state variables through sequential data assimilation of measurements. Recently, the EnKF algorithm has been widely used for high-dimensional nonlinear data assimilation in geophysical [38], atmospheric [39] and hydrological [40–44] modeling. However, the application of EnKF to numerical models that involve multiple processes would be inconvenient because the implementation of EnKF requires the modification of restart files and the update of model parameters simultaneously at each assimilation step [45]. The ensemble smoother

(ES) proposed by van Leeuwen and Evensen [46] is a more efficient alternative to EnKF when dealing with high-dimensional parameter estimation problems. Instead of implementing the sequential data assimilation scheme as EnKF, ES computes a global update of parameters by assimilating all observations simultaneously [45]. Skjervheim and Evensen [47] compared the performance of ES and EnKF for solving history matching problems of reservoirs and found that ES could provide similar results to EnKF. The difference is that ES works more efficiently than EnKF because it can be implemented without restarting the simulation model multiple times.

However, both the standard ES and standard EnKF algorithm cannot provide satisfactory data matches when dealing with highly nonlinear problems [48–50]. To further improve the performance and efficiency of ES, the iterative forms of ES have been developed recently. Chen and Oliver [51] proposed an iterative ensemble smoother (IES) algorithm based on the Levenberg–Marquardt (LM) method, in which the ensemble randomized maximum likelihood (EnRML) was used as the smoother. Emerick and Reynolds [52] developed an ensemble smoother with multiple data assimilation (ES-MDA) to solve the reservoir history-matching problem. Li and Davis [53] compared the performance of ES and IES in groundwater modeling and their results showed that IES works much better than the standard ES by continuously updating parameters with the available measurements for all time steps. Ma et al. [54] developed an adaptive IES method (ES-LM) based on the Levenberg-Marquardt algorithm which can reduce the computational complexity of ES. Zhang et al. [55] proposed an iterative local updating ensemble smoother (ILUES) for parameter estimation in hydrologic models. In the framework of ILUES, an iterative form of ES is used to update the local ensemble of each sample, which is defined based on a comprehensive measure of distance to the sample and the measurements. The performance of the ILUES algorithm was evaluated with several numerical examples and the results showed that ILUES could provide an accurate estimation of model parameters with multimodal distributions. Thus, the ILUES algorithm is of great potential to solve highly nonlinear problems such as the inverse modeling of multicomponent reactive transport.

In this study, the ILUES algorithm is employed to estimate the key parameters of a multicomponent reactive transport model. The numerical model, which characterizes the sequential biodegradation of tetrachloroethene (PCE), describes the coupled simulation of groundwater flow, biochemical reaction and solute transport. As the subsurface heterogeneity has significant effects on the biodegradation process, the key model parameters, including the hydraulic paramters, biochemical parameters and contaminant source characteristics, are jointly estimated using the ILUES algorithm [55]. The major objective of this study is to verify the effectiveness and efficiency of the ILUES algorithm for solving the inverse problem of the multicomponent reactive transport model.

To the best of our knowledge, this is the first study that focuses on the application of the ILUES algorithm to joint estimation of hydraulic parameters, geochemical parameters and contaminant source characteristics in a sequential biodegradation process. Although model parameter estimation with inverse methods has been investigated recently, related studies seldom focus on the reactive solute transport process, which is difficult to analyze due to the high dimensionality. Furthermore, the ILUES algorithm in this study is modified by determining the local ensemble partly with a linear ranking selection scheme [56] which is able to more extensively explore the parameter space. The performance of the ILUES algorithm is then evaluated through three different numerical case studies. A comparison between ILUES and ES-MDA is also made to demonstrate the advantage of ILUES on estimation accuracy.

The organization of the paper is as follows: in Section 2, the basic theory of the sequential biodegradation of PCE is explained. Then the general framework and mathematical formulation of the ILUES algorithm is introduced. In Section 3, three different numerical case studies based on the biodegradation of PCE are utilized to test the performance of the ILUES algorithm. The results and discussion of corresponding case studies are also presented in this section. The conclusions of the current study are drawn in Section 4.

2. Methodology

2.1. Sequential Anaerobic Biodegradation of PCE

In this study, PCE contaminants undergo a sequential reductive dechlorination reaction whose degradation kinetics are assumed to be first order in nature. In the process of dechlorination, microorganisms can utilize PCE as the electron acceptor, sequentially removing chlorine atoms from PCE to form trichloroethene (TCE). TCE is degraded to dichlorethene (DCE) and then vinyl chloride (VC) in turn. Finally, non-toxic ethylene (ETH) is produced by the degradation of VC. The typical sequential reductive dechlorination framework of PCE under anaerobic conditions can be written as:

$$PCE \overset{k_1}{\Rightarrow} TCE \overset{k_2}{\Rightarrow} DCE \overset{k_3}{\Rightarrow} VC \overset{k_4}{\Rightarrow} ETH, \tag{1}$$

where k_i denotes the first-order rate constants of PCE and its daughter products. The governing equations which characterize the transformation and transport of PCE and its daughter products are represented by the following partial differential equations:

$$R_P \frac{\partial[PCE]}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial[PCE]}{\partial x_j}\right) - \frac{\partial(v_i[PCE])}{\partial x_i} + \frac{q_s}{\phi}[PCE]_S - K_1[PCE], \tag{2}$$

$$R_T \frac{\partial[TCE]}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial[TCE]}{\partial x_j}\right) - \frac{\partial(v_i[TCE])}{\partial x_i} + \frac{q_s}{\phi}[TCE]_S + Y_{T/P}K_1[PCE] - K_2[TCE], \tag{3}$$

$$R_D \frac{\partial[DCE]}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial[DCE]}{\partial x_j}\right) - \frac{\partial(v_i[DCE])}{\partial x_i} + \frac{q_s}{\phi}[DCE]_S + Y_{D/T}K_2[TCE] - K_3[DCE], \tag{4}$$

$$R_V \frac{\partial[VC]}{\partial t} = \frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial[VC]}{\partial x_j}\right) - \frac{\partial(v_i[VC])}{\partial x_i} + \frac{q_s}{\phi}[VC]_S + Y_{V/D}K_3[DCE] - K_4[VC], \tag{5}$$

where R_i is the retardation factor; D_{ij} denotes the hydrodynamic dispersion coefficient $\left[L^2T^{-1}\right]$; $[PCE]$, $[TCE]$, $[DCE]$, $[VC]$ represent concentrations of corresponding contaminant in the dechlorination reaction $\left[ML^{-3}\right]$; v_i is the pore velocity $\left[LT^{-1}\right]$; q_s is the volumetric flux of water per unit volume of aquifer $\left[T^{-1}\right]$; ϕ is the soil porosity; $[PCE]_S$, $[TCE]_S$, $[DCE]_S$, $[VC]_S$ are concentrations of corresponding contaminants in the dechlorination reaction at source point $\left[ML^{-3}\right]$; K_i is the first-order anaerobic degradation rate $\left[T^{-1}\right]$; and $Y_{T/P}$, $Y_{D/T}$, $Y_{V/D}$ are chlorinated compound yield values under anaerobic reductive dechlorination conditions. In this study, groundwater flow and reactive transport models are simulated using MODFLOW-2000 [57] and RT3D [58,59], respectively.

2.2. Parameter Estimation

Parameter estimation of reactive transport models aim to obtain the accurate estimations of key model parameters of groundwater flow and solute transport by solving inverse problems. Bayesian inference is an important method in inverse modeling and has been widely used in hydrologic science.

Considering the Bayesian inference for a nonlinear model, the relationship between model parameters and measurements can be expressed in the following form:

$$d = f(m) + \varepsilon, \tag{6}$$

where d is the vector for measurements, m is the vector for model parameters, $f(m)$ is the prediction from the forward model, and ε is the vector for Gaussian-distributed measurement errors.

According to Bayes' theorem, posterior distribution is proportional to likelihood times prior distribution, which can be written as:

$$p(m|d) \propto p(d|m)p(m), \tag{7}$$

where $p(m|d)$ is the posterior probability distribution, $p(d|m)$ is the prior distribution and $p(m)$ is the likelihood function that measures the goodness of fit of the model to observations of the unknown parameters.

In this study, the analytical evaluation of the posterior distribution of unknown parameters is unavailable due to the high nonlinearity of the model. For parameter estimation problems in nonlinear models, ensemble smoother (ES) is a prevalent method because of its effectiveness and efficiency. However, when dealing with highly nonlinear problems, the standard ES method is unable to obtain satisfactory estimations. In this paper, an iterative local updating ensemble smoother (ILUES) is employed, which is proposed by Zhang et al. [55], to jointly estimate various parameters of the reactive transport model. In comparison with ES, the ILUES algorithm updates the local ensemble of each sample with the iterative form of ES, instead of globally assimilating parameter realizations in the ensemble. In this way, ILUES has a better performance on highly nonlinear problems such as the inverse modeling of reactive transport.

2.3. Iterative Local Updating Ensemble Smoother

The basic updating equation of ES can be written as:

$$m_j^a = m_j^f + C_{MD}^f (C_{DD}^f + C_D)^{-1} \left[d_j - f\left(m_j^f\right) \right], \tag{8}$$

for $j = 1, 2, \cdots, N_e$, N_e denotes the size of ensemble members.

According to Equations (1) and (8), $M^f = \left(m_1^f, \cdots, m_j^f \right)$ is an ensemble of samples randomly drawn from the prior distribution, $M^a = \left(m_1^a, \cdots, m_j^a \right)$ is the updated ensemble after taking measurements d into account, $D^f = (f(m_1^f), \cdots, f(m_j^f))$ is the vector for predictions from the forward model, C_{MD}^f is the cross-variance matrix between M^f and D^f, C_{DD}^f is the auto-covariance matrix of D^f, C_D is the covariance matrix of the measurement errors, d_j represents the jth measurement of model parameter m with error of ε_j.

On the basis of Bayesian inference and the ES method, the implementation of ILUES can be summarized in the following steps and its framework is represented by the flowchart in Figure 1.

Step1: Initialization.

To begin with, set the iteration counter i to 0. N_e samples are randomly drawn from the prior distribution of model parameters as the initial input ensemble. Therefore, the corresponding output ensemble can be obtained by evaluating the forward model.

Step 2: Determination of local ensemble.

The local ensemble of the sample $m_j^f (j = 1, 2, \cdots, N_e)$ is defined based on a comprehensive measure of the distance (J) from aspects of the model responses (J_1) and the model parameters (J_2).

$$J(m) = J_1(m)/J_1^{max} + J_2(m)/J_2^{max}, \tag{9}$$

$$J_1(m) = [f(m) - d]^T C_D^{-1} [f(m) - d], \tag{10}$$

$$J_2(m) = \left[m - m_j^f \right]^T C_{MM}^{-1} \left[m - m_j^f \right], \tag{11}$$

where $J_1(m)$ denotes the distance between the measurements d and the model responses $f(m)$, $J_2(m)$ denotes the distance between the selected sample m_j^f and the model parameters m, J_1^{max} and J_2^{max} are the maximum values of $J_1(m)$ and $J_2(m)$, C_{MM} is the auto-covariance matrix of the model parameters.
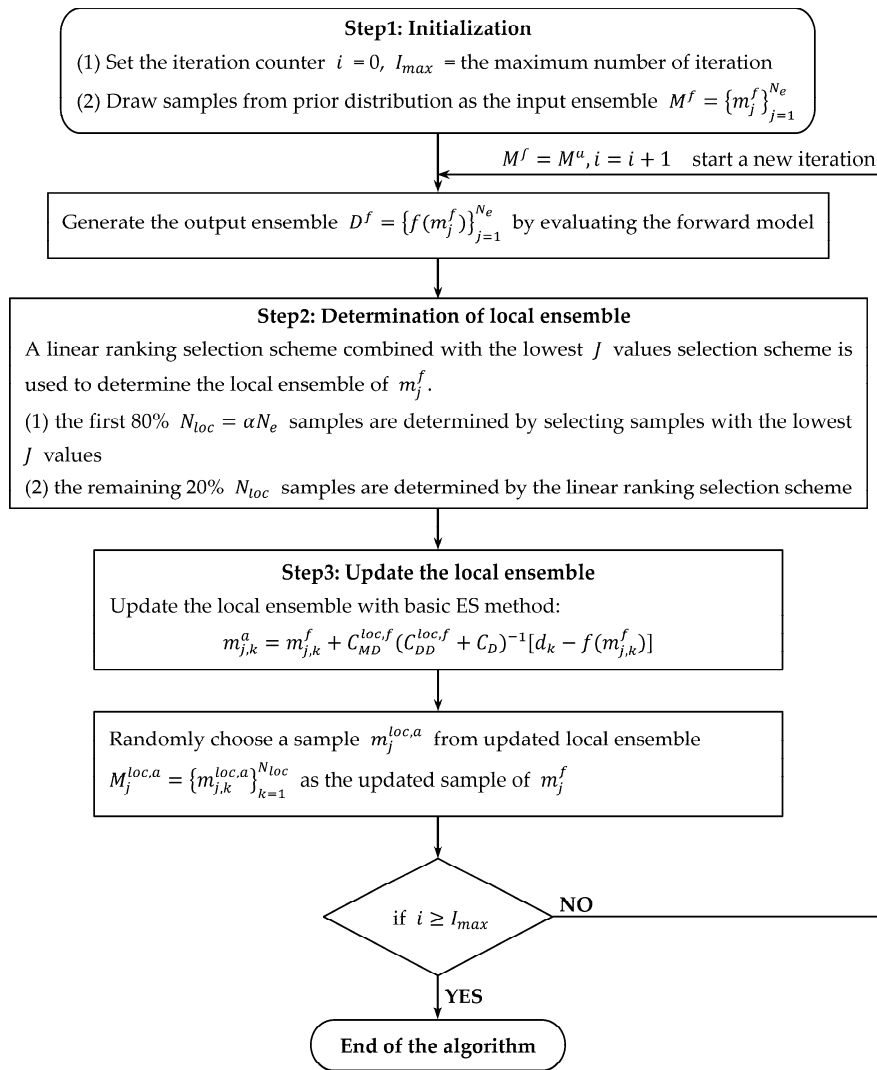
Figure 1. The framework of the iterative local updating ensemble smoother (ILUES) algorithm.

According to the original ILUES algorithm proposed by Zhang et al. [55], the local ensemble of m_j^f is determined by selecting $N_{loc} = \alpha N_e$ ($\alpha \in (0,1]$) samples with the lowest J values, expressed as $M_j^{loc,f} = \left(m_{j,1}^f, \cdots, m_{j,N_{loc}}^f\right)$. In this paper, a linear ranking selection scheme [56] is combined to determine the local ensemble according to the J values. The selection probability of each sample is expressed as:

$$P_i = P_{min} + (P_{max} - P_{min})\frac{rank(\theta_i) - 1}{N_e - 1}, \tag{12}$$

$$\theta_i = \rho_i / \sum_{j=1}^{N_e} \rho_j, \tag{13}$$

$$\rho_i = 1/J(m_i), \tag{14}$$

for $i = 1, \cdots, N_e$

Where $P_{min} = 0.1$, $P_{max} = 0.9$, $rank(\theta_i)$ represents the rank of each individual after sorted by the fitness value of θ_i.

To better explore the parameter space, this rank-based selection operator is used together with the original selection scheme in ILUES, which means that the first 80% of N_{loc} samples are determined by

selecting samples with the lowest *J* values and the remaining 20% of N_{loc} samples are determined by the linear ranking selection scheme.

　　Step 3: Update the local ensemble.

　　Update the local ensemble with the basic ES method, written as:

$$m_{j,k}^a = m_{j,k}^f + C_{MD}^{loc,f} \left(C_{DD}^{loc,f} + C_D \right)^{-1} \left[d_k - f\left(m_{j,k}^f \right) \right],$$

(15)

for $k = 1, \cdots, N_{loc}$.

　　In the above equation, $C_{MD}^{loc,f}$ is the cross-covariance matrix between $M_j^{loc,f}$ and $D_j^{loc,f}$, $C_{DD}^{loc,f}$ is the auto-covariance matrix of measurement $D_j^{loc,f}$, d_k represents the *k*th measurement of model parameters *m* with error of ε_k.

　　Then choose a random sample $m_j^{loc,a}$ from the updated local ensemble $M_j^{loc,a} = \left(m_{j,1}^a, \cdots, m_{j,N_{loc}}^a \right)$ as the updated sample of m_j^f ($j = 1, 2, \cdots, N_e$). The updated global ensemble $M^a = \left(m_1^{loc,a}, m_2^{loc,a}, \cdots, m_{N_e}^{loc,a} \right)$ can be obtained after this procedure.

　　The framework of the ILUES algorithm can be generalized into the following flowchart.

3. Illustrative Case Studies

　　In this section, three numerical case studies for joint parameter estimation of the reactive transport model are introduced to evaluate the performance of the modified ILUES algorithm. These studies are based on the sequential reductive dechlorination of tetrachloroethene (PCE) under anaerobic conditions which follows a series of chain reactions. The dechlorination reaction and solute transport can be simulated by a multicomponent reactive transport model simultaneously. A one-dimensional case is firstly used to demonstrate the validity of the ILUES algorithm for estimating key parameters of the reactive transport model. Then, the performance of the ILUES algorithm is further tested by jointly estimating the source strength and rate constants of the model in a two-dimensional heterogeneous aquifer. The effects of ensemble size and factor α on the prediction accuracy of ILUES are also investigated in Case 2. Finally, the complexity of the problem is increased by adding various unknown parameters, including the hydraulic conductivity and source location. The numerical model constructed in this case is based on the two-dimensional aquifer in Case 2. The third case focuses on the joint estimation of hydraulic parameters, biochemical parameters and contaminant source characteristics in the sequential biodegradation process of PCE.

3.1. Case 1: A One-Dimensional Case

　　The first case aims to test the effectiveness of the ILUES algorithm for estimating key parameters of the reactive transport model. In this case, the sequential biodegradation process of PCE and the transport of reactive solutes are considered in a one-dimensional confined aquifer with steady-state groundwater flow. As shown in Figure 2, the aquifer is assumed to be 100 m in length and 5 m in width. The model domain is discretized into equal size grids of 5 m × 5 m and has constant heads of 100 m and 99.8 m at the right and left boundaries, respectively. No flow condition is prescribed at the upper and lower boundaries. The PCE contaminant with unknown strength is released at the location of (87.5 m, 2.5m) and an observation well is set at the location of (27.5 m, 2.5 m). The simulation time is 1000 days with 10 equal time steps, so each time step lasts 100 days. More details of related model parameters are listed in Table 1.
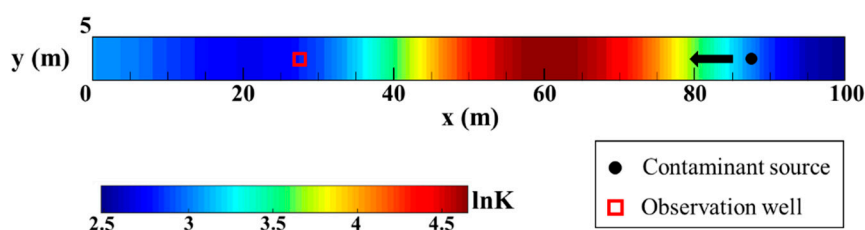
Figure 2. The conceptual model of the one-dimensional confined aquifer. The distribution of the reference log-conductivity field in Case 1 is presented in the background of the model. The black dot denotes the contaminant source and the red square denotes the observation well. The black arrow indicates the direction of groundwater flow.

Table 1. Basic parameters of the reactive transport model in Case 1.

Parameters	Value
Model length (m)	100
Model width (m)	5
Grid spacing (m)	5×5
Simulation time (days)	1000
Time steps	10
Porosity	0.3
Hydraulic gradient	0.002
Longitudinal dispersivity (m)	10

The major objective of this case study is to obtain the joint estimation of hydraulic conductivity (K), contaminant source strength (S_S) of PCE and degradation rate constants (k_1, k_2, k_3, k_4) of the chain reaction using the ILUES algorithm. Measurements of the concentrations of PCE, TCE, DCE and VC are collected from the observation well at $t = (2, \ 4, \ 6, \ 8, \ 10) * 100$ days to provide information for the estimation. The measurement errors considered in this case follow a normal (Gaussian) distribution with mean $\mu = 0$ and variance $\sigma^2 = 0.005^2$. The prior distributions of the contaminant source strength (S_S) and the first-order rate constants (k_1, k_2, k_3, k_4) are assumed to follow uniform distributions. Their ranges and true values are listed in Table 2.

Table 2. The prior distributions and true values of the model parameters estimated in Case 1. U stands for the uniform distribution.

Parameters	Prior Distribution	True Value
S_S	$U(0, 500)$	200
$log_{10}(k_1)$	$U(-5, -2)$	-2.3010
$log_{10}(k_2)$	$U(-5, -2)$	-2.5230
$log_{10}(k_3)$	$U(-5, -2)$	-2.6990
$log_{10}(k_4)$	$U(-5, -2)$	-3.0460

Considering the spatial heterogeneity of the aquifer, the log-conductivity field $Y = \ln(K)$ is assumed to follow a Gaussian distribution, with mean $\mu_Y = 3$ and variance $\sigma_Y^2 = 1$. The identification of hydraulic conductivity is computationally expensive because the number of unknown parameters increases a lot. To reduce the dimensionality of the heterogeneous hydraulic conductivity field and efficiently obtain the converged inversion solutions, the Karhunen-Loève (KL) expansion [60] is utilized to parameterize the log-conductivity field (Y). The log hydraulic conductivity $Y = \ln(K)$ at two

arbitrary spatial locations (x_1, y_1) and (x_2, y_2) in the model are assumed to be exponentially correlated. Their relationship can be expressed by the following covariance function:

$$C_Y(x_1, y_1; x_2, y_2) = \sigma_Y{}^2 \exp\left(-\frac{|x_1 - x_2|}{\lambda_x} - \frac{|y_1 - y_2|}{\lambda_y}\right), \tag{16}$$

where $\sigma_Y{}^2 = 1$ is the variance, $\lambda_x = 5\,m$ and $\lambda_y = 50\,m$ are the correlation length along x and y directions, respectively. The KL expansion of the log-conductivity field $Y = \ln(K)$ can be expressed as:

$$Y(x) \approx \overline{Y}(x) + \sum_{i=1}^{N_{KL}} \xi_i \sqrt{\tau_i} f_i(x), \tag{17}$$

where ξ_i are independent standard Gaussian random variables, and τ_i and $f_i(x)$ are eigenvalues and eigenfunctions of the covariance function C_Y, respectively. In this case, approximately 92% of the conductivity field variance can be preserved by keeping the first five KL terms (i.e., $N_{KL} = 5$), which means that an approximative distribution of the conductivity field can be calculated using Equation (17). Therefore, the identification of the whole hydraulic conductivity field is then simplified to the estimation of only five parameters. To sum up, there are total 10 unknown parameters to be estimated in this case.

For joint estimation of the unknown parameters, the ILUES algorithm is implemented with the ensemble size $N_e = 500$ and the iteration number $N_{Iter} = 7$. The factor α is set to 0.1, which is in the range of $[0, 1]$. The contaminant source strength (S_S) and the rate constants of the chain reaction (k_1, k_2, k_3, k_4) can be accurately estimated within seven iterations, as shown in Figure 3.
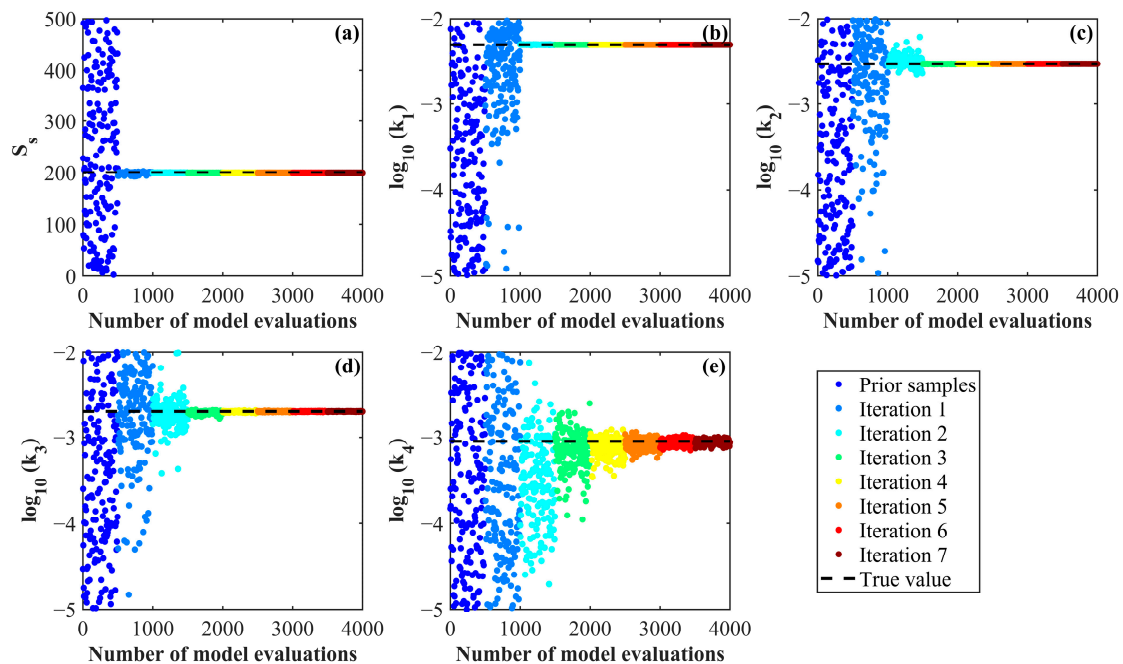

Figure 3. Trace plots of unknown model parameters estimated by the ILUES algorithm in Case 1. (**a**) represents the trace plot of the source strength (S_S). (**b**–**e**) represents the trace plot of the four rate constants (k_1, k_2, k_3, k_4), respectively. The black dashed lines denote the reference true values of corresponding model parameters.

Meanwhile, with the five estimated KL terms, the identification of the hydraulic conductivity field is realized by calculating the KL expansion equation. To evaluate the performance of the ILUES algorithm on the conductivity field identification, the mean and the standard deviation of the

log-conductivity field at selected iterations (Iteration 1, 3, 5 and 7) are plotted in Figure 4. The true values and randomly generated and initial values are also presented in Figure 4 as the reference. As illustrated in Figure 4a, the mean of the log-conductivity field gets closer to the true values as the number of iterations increases and it greatly approximates the true values after seven iterations. In Figure 4b, the standard deviation of the log-conductivity field is relatively high at the early iterations. However, it decreases dramatically with the increase in the iteration number and gets close to zero after seven iterations. The results shown in Figure 4 verify the effectiveness of the ILUES algorithm for the conductivity field identification.
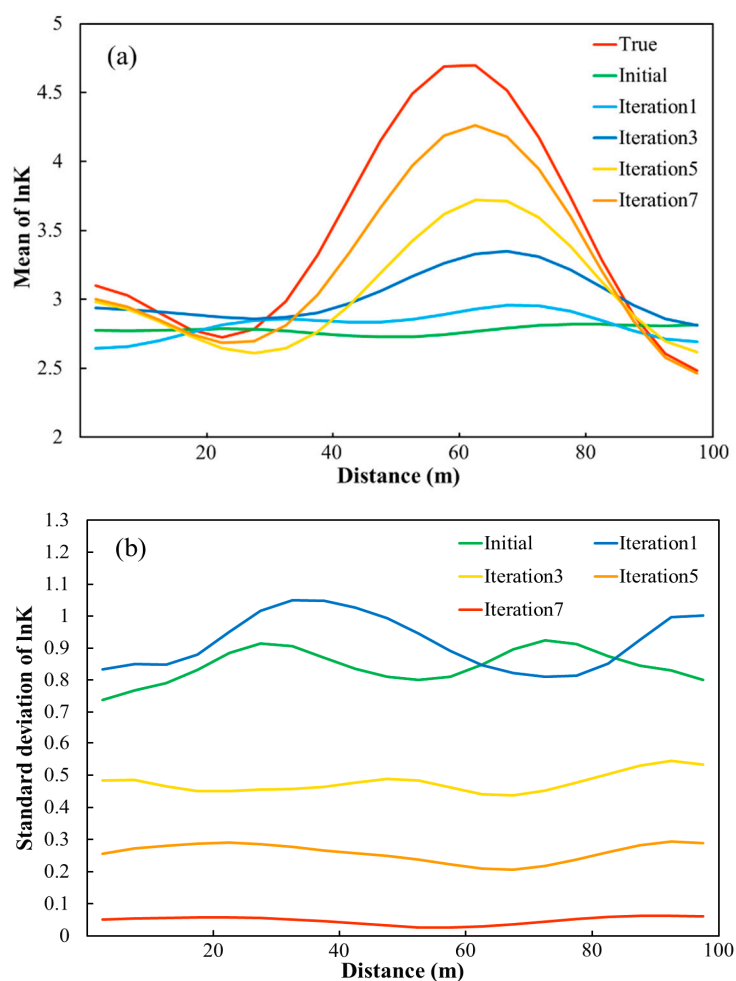
Figure 4. (**a**) is the mean of the log-conductivity field at selected iterations. The red line in (**a**) denotes the true values of the log-conductivity field. (**b**) shows the standard deviation of the log-conductivity field at selected iterations. The green line denotes the mean and the standard deviation of the log-conductivity field of the initial input ensemble in Figure 4a,b, respectively.

To further demonstrate the estimation accuracy of the ILUES algorithm, the performance of data match is also evaluated using the measurements (i.e., contaminant concentrations at the observation well). The concentration data of the initial and the last ensemble are presented in Figure 5. The true values of the concentration data simulated by the forward model are also shown in Figure 5 as the reference. As illustrated in Figure 5, the concentration data of the last ensemble approximates the true values of the measurements after seven iterations. Therefore, the ILUES algorithm can obtain a satisfactory data match on this problem.
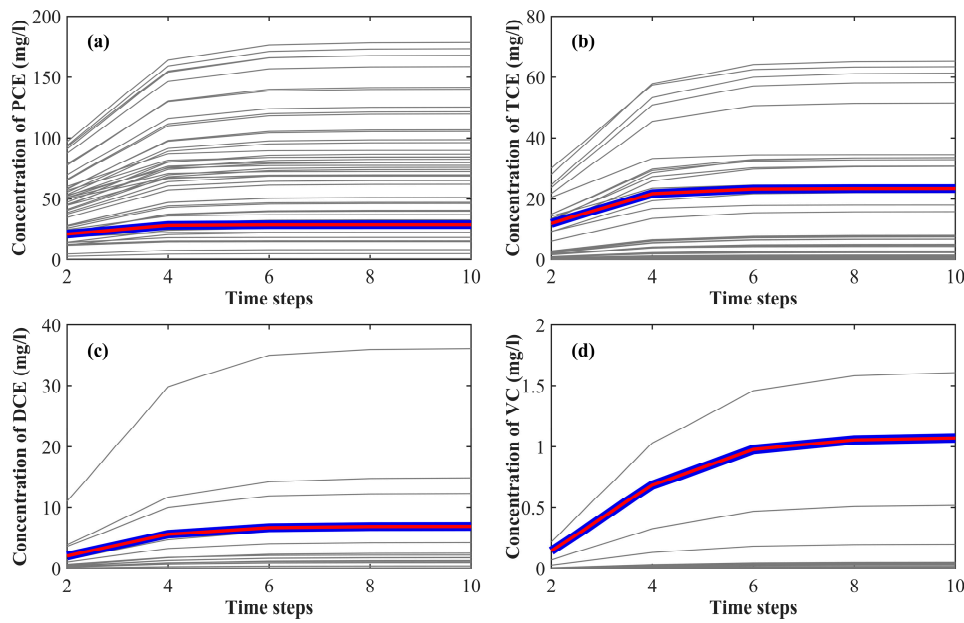
Figure 5. The performance of data match. (**a**–**d**) shows the data match of the concentration of tetrachloroethene (PCE), trichloroethene (TCE), dichlorethene (DCE) and vinyl chloride (VC) at the observation well, respectively. The grey lines denote the concentration data which is drawn from the initial ensemble every five samples. The blue lines denote the concentration data of the last ensemble. The red line denotes the true values of the measurements.

The estimation accuracy of ILUES can also be evaluated by the root-mean-square error (RMSE), which is a value that measures the differences between values predicted by an estimator and the values observed. In general, the dataset with lower RMSE (closer to 0) indicates better prediction accuracy. The RMSE of parameters estimated in this case is defined by the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{N_e} \sum_{n=1}^{N_e} (\hat{y}_n - y_n)^2},$$ (18)

where $\hat{y}_n$ denotes the predicted value of measurements, y_n denotes the reference true value of measurements and N_e is the ensemble size. As shown in Figure 6, the RMSE of model parameters estimated in this case decreases evidently as the number of iterations increases and all RMSE values approach 0 after seven iterations, which demonstrate the accuracy of ILUES for parameter estimation. In summary, the ILUES algorithm is an effective method for the joint estimation of unknown parameters of the multicomponent reactive transport model.
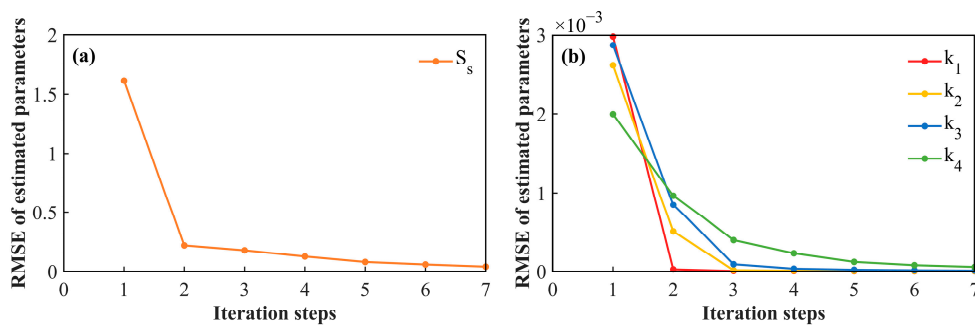
Figure 6. The RMSE of model parameters estimated in Case 1. (**a**) represents the RMSE of source strength at different iteration steps. (**b**) represents the RMSE of rate constants of the reactive transport model at different iteration steps. Different coloured lines denote the RMSE of each parameter at different iteration steps.

3.2. Case 2: A Two-Dimensional Case with Heterogeneous Hydraulic Conductivity Field

In this case, the validity of the ILUES algorithm is tested in a two-dimensional model. The coupled simulation of groundwater flow, biodegradation reaction and solute transport based on the sequential reductive dechlorination of PCE is considered in a two-dimensional confined aquifer with steady state groundwater flow and heterogeneous hydraulic conductivity. As shown in Figure 7, the aquifer is assumed to be 500 m in length and 300 m in width, with constant-head boundaries of 100 m and 99 m on the left and right side, respectively. No-flow boundaries are specified along the upper and lower side of the model domain. The model domain is discretized into uniform grids of 10 m $\times$ 10 m. The PCE contaminant is released from a point source located at (95 m, 145 m) and an observation well is set at the location of (295 m, 145 m). The total simulation time is 1200 days and is divided into 15 equal time steps. Other related parameters used in the simulation model are listed in Table 3.
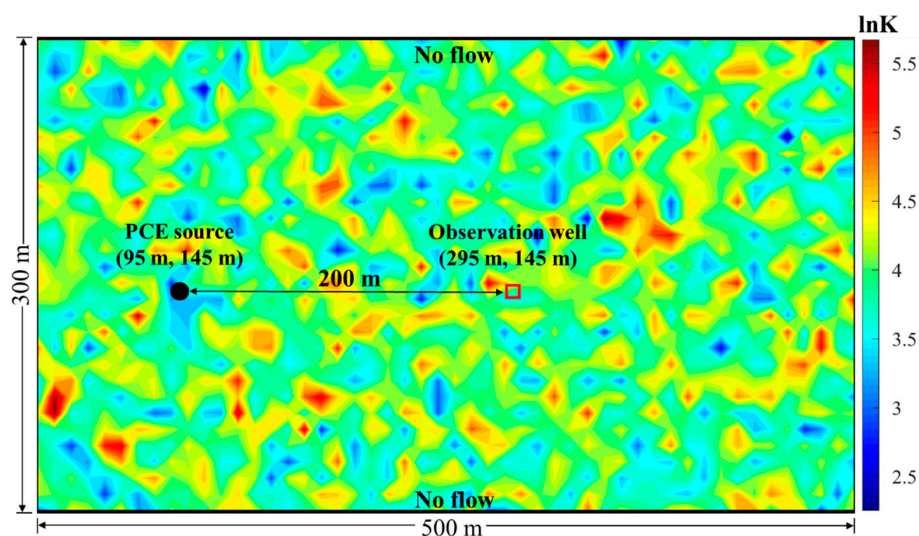
Figure 7. The conceptual model of the two-dimensional confined aquifer considered in Case 2. The background of the model depicts the spatial distribution of the log-conductivity field. The black dot and red square denotes the contaminant source and the observation well, respectively.

Table 3. Basic parameters of the reactive transport model in Case 2.

Parameters	Value
Model length (m)	500
Model width (m)	300
Grid spacing (m)	10×10
Total simulation time (days)	1200
Time steps	15
Porosity	0.3
Hydraulic gradient	0.002
Longitudinal dispersivity (m)	10
Horizontal transverse dispersivity (m)	3

The second case study focus on the joint estimation of unknown model parameters, including the contaminant source strength (S_S) and the first-order rate constants (k_1, k_2, k_3, k_4) of the degradation reaction. As the information required for parameter estimation, the contaminant concentrations of PCE, TCE, DCE and VC are obtained from the observation well at t $= [6, 8, 10, 12, 14] * 80$ days. The measurement errors in this case are assumed to follow a Gaussian distribution with zero mean and variance $\sigma^2 = 0.005^2$. To increase the complexity of the model, the aquifer is assumed to be heterogeneous and the log-conductivity field $Y = \ln(K)$ of the model is randomly generated from a Gaussian distribution with mean $\mu_Y = 4$ and variance $\sigma_Y^2 = 0.25$. The prior distributions of

the contaminant source strength (S_S) and the first-order rate constants (k_1, k_2, k_3, k_4) are uniform distributions. Their ranges and reference true values are listed in Table 4. In total, there are five unknown parameters to be estimated in Case 2.

Table 4. The prior distributions and reference true values of the model parameters estimated in Case 2. U stands for the uniform distribution.

Parameters	Prior Distribution	True Value
S_S	$U(0, 1000)$	200
$log_{10}(k_1)$	$U(-5, -2)$	-2.3010
$log_{10}(k_2)$	$U(-5, -2)$	-2.5230
$log_{10}(k_3)$	$U(-5, -2)$	-2.6990
$log_{10}(k_4)$	$U(-5, -2)$	-3.0460

During the procedure of data assimilation, determination of the local ensemble is a crucial step that can evidently affect the inversion results. According to the equation $N_{loc} = \alpha N_e$ mentioned in Section 2, the size of the local ensemble is affected by two factors, α and N_e, where N_e denotes the size of global ensemble and α denotes the ratio of local ensemble over the global ensemble. To investigate the impacts of the two factors on the accuracy of estimation, different combinations of α and N_e are evaluated by the RMSE of the measurements.

To facilitate the comparison between different combinations of α and N_e, the log-transformed RMSE is introduced in Figure 8.
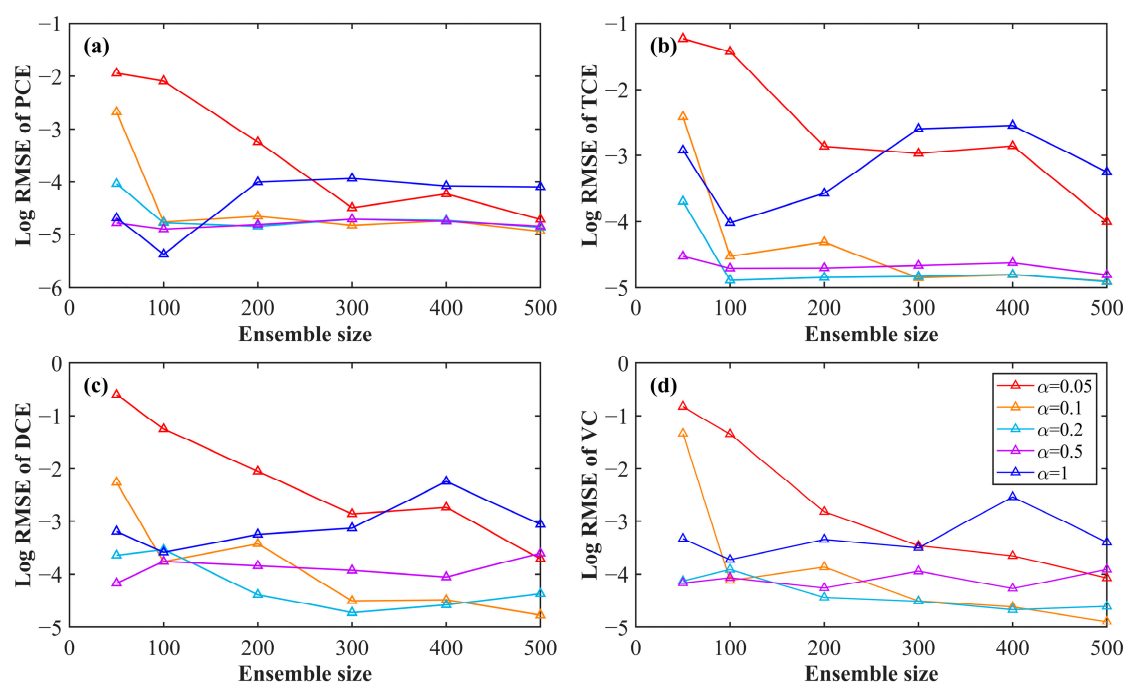

Figure 8. (**a**–**d**) represents the log-transformed root-mean-square error (RMSE) of concentration measurements of PCE, TCE, DCE and VC at the third iteration, respectively. Different triangles represent the log-RMSE for different combinations of α and N_e. Different colored lines denote the variation of log-RMSE with the increase of N_e when the value of α is fixed. The number of iterations is set to 10 for the ILUES algorithm.

As shown in Figure 8, when α is very small or very large (e.g., $\alpha = 0.05$ or $\alpha = 1$), the log-transformed RMSE are obviously larger than datasets with other α values, no matter what the value of N_e is. This indicates that the ILUES algorithm is unable to obtain satisfactory estimations in both situations. If α is too small, the local ensemble is so small that it may miss the samples that are

close to the observations, as pointed out in Section 2. Meanwhile, if α is too large, the size of the local ensemble will be close to the global ensemble, so the estimation scheme will be similar to the ensemble smoother (ES) without local updating. Furthermore, the datasets with $\alpha = 0.1$ and $\alpha = 0.2$ usually give relatively lower log-RMSE values of the measurements, regardless of the value of N_e. Therefore, in this case, rather than using a constant α value in the original ILUES algorithm, α that randomly generated in the range of [0.1, 0.2] is selected at each iteration to determine the local ensemble.

As for the factor N_e, when the ensemble size N_e is too small (e.g., $N_e = 50$), limited ensemble updating may underestimate the uncertainty in model parameters and give biased inversion results. In Figure 8, when the value of α is fixed, the datasets with larger N_e values usually give relatively lower log-RMSE values, which also indicates better prediction accuracy. However, the larger ensemble size could result in a higher computational burden and the involvement of more measurement errors in the model. To balance the computational efficiency and the estimation accuracy, a combination of α, which is randomly generated in the range of [0.1, 0.2], and $N_e = 300$ is selected as the optimal setting of the ILUES algorithm in the subsequent parameter estimation for this case study.

After determining the optimal setting of α and N_e, the model parameter estimation is implemented with the ILUES algorithm. The number of iterations is set to 10 in this case. As shown in Figure 9, accurate joint estimations of the contaminant source strength and the degradation rate constants can be obtained within 10 iterations, which demonstrates the ability of ILUES to solve inverse problems. To further evaluate the estimation accuracy of ILUES, the RMSE of model parameters estimated in this case study is introduced. As shown in Figure 10, the RMSE values of all parameters decrease and gradually approximate 0 with the increase of the iteration number, which means that the estimation error gets lower and lower with the increase of the iteration number. Therefore, model parameter estimation can achieve sufficient accuracy when the number of iterations is large enough. To summarize, for high-dimensional inverse problems, the iteration number and ensemble size of ILUES should be enlarged appropriately. The investigation into the impact of α on estimation accuracy provides a reference for the subsequent case study.
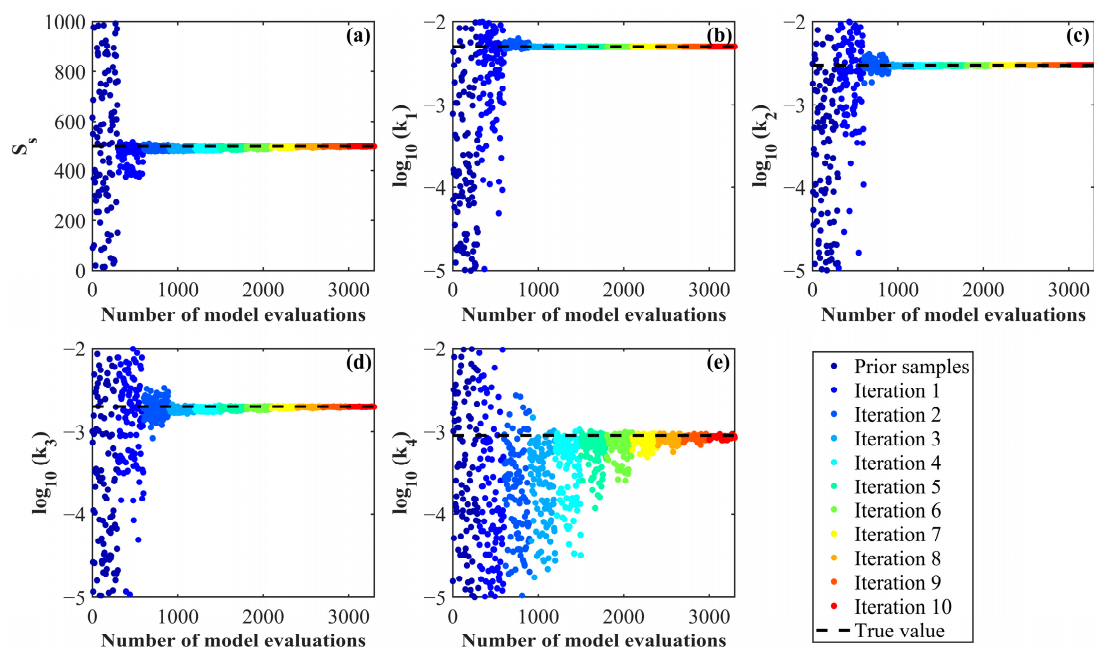
Figure 9. Trace plots of unknown model parameters estimated by the ILUES algorithm in Case 2. (**a**) represents the trace plot of the source strength (S_S). (**b**–**e**) represents the trace plot of the four rate constants (k_1, k_2, k_3, k_4), respectively. The black dashed lines denote the reference true values of corresponding model parameters.
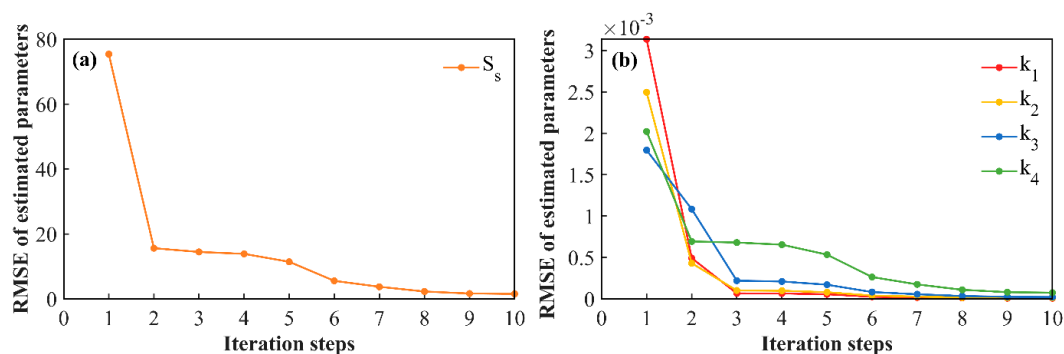
Figure 10. The RMSE of model parameters estimated in Case 2. (**a**) represents the RMSE of source strength at different iteration steps. (**b**) represents the RMSE of rate constants of the reactive transport model at different iteration steps. Different coloured lines denote the RMSE of each parameter at different iteration steps.

3.3. Case 3: A Two-Dimensional Case with Unknown Contaminant Source and Hydraulic Conductivity Field

In Case 3, a parameter estimation problem with a large number of unknown parameters is studied to evaluate the performance of the ILUES algorithm on joint estimation of model parameters. The sequential reductive dechlorination of PCE is considered in a two-dimensional confined aquifer with steady state groundwater flow, whose numerical model has the same model domain and boundary conditions as Case 2, as shown in Figure 11. The total simulation time is still 1200 days with 15 equal time steps. However, in addition to contaminant source strength and degradation rate constants, the contaminant source location and the hydraulic conductivity are assumed to be uncertain in this case, which significantly increases the complexity of the model. The PCE contaminant is released from a potential area depicted by the white dashed rectangle in Figure 11 and the source strength is assumed to be an unknown constant. More details of related model parameters are listed in Table 5.
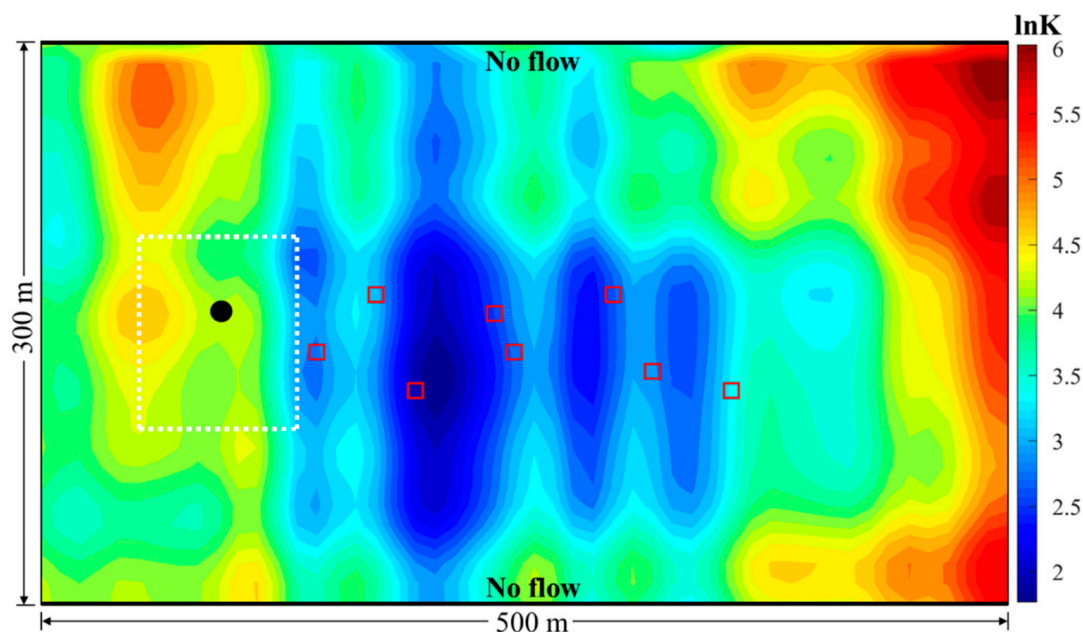
Figure 11. The conceptual model of the two-dimensional confined aquifer considered in Case 3. The background of the model depicts the spatial distribution of the log-conductivity field. The black dot denotes the contaminant source and the red squares denote the observation wells. The white dashed rectangle denotes the potential area of the contaminant source.

Table 5. Basic parameters of the reactive transport model in Case 3.

Parameters	Value
Model length (m)	500
Model width (m)	300
Grid spacing (m)	10×10
Total simulation time (days)	1200
Time steps	15
Porosity	0.3
Hydraulic gradient	0.002
Longitudinal dispersivity (m)	10
Horizontal transverse dispersivity (m)	3

The unknown parameters estimated in this case can be classified into three categories. The first category is the contaminant source characteristics which include the contaminant source strength (S_s) and contaminant source location (S_x, S_y). The second category is comprised of the four first-order rate constants (k_1, k_2, k_3, k_4) of the dechlorination reaction, known as the biochemical parameters. The third category consists of the KL terms which are drawn from the KL expansion of the log-conductivity field $Y = \ln(K)$.

To characterize the heterogeneity of the aquifer, the log-conductivity field $Y = \ln(K)$ is assumed to follow a Gaussian distribution, with mean $\mu_Y = 4$ and variance $\sigma_Y^2 = 1$. The identification of the hydraulic conductivity field can be realized by estimating the conductivity of each grid, so there are 1500 unknown parameters to be estimated. As the number of unknown parameters increases greatly, the computational cost can grow tremendously. To reduce the dimensionality and simplify the estimation of the hydraulic conductivity field, the KL expansion [60] is used to parameterize the log-conductivity field (Y), as used in Case 1. In this case, approximately 95% of the field variance for conductivity can be preserved by keeping the first 68 KL terms (i.e., $N_{KL} = 68$), which means that an approximative distribution of the conductivity field can be obtained with these 68 parameters. To obtain ample information from the measurements, eight observation wells are set in the model domain, as marked with red squares in Figure 11. The observation data, including the concentration of multiple contaminants (PCE, TCE, DCE and VC) and the hydraulic head of the specific grid, are collected from the eight observation wells at t = $[6, 8, 10, 12, 14] * 80$ (days). As a result, 168 observation data in total are collected from the observation wells with measurement errors $\varepsilon \sim N(0, 0.005^2)$.

In total, there are 75 uncertain parameters to be estimated in Case 3. The seven parameters which belong to the first and second parameter categories are uniformly distributed. Their corresponding ranges and reference true values are listed in Table 6. The prior distributions of the 68 KL terms follow the standard normal distribution. The joint estimation of the three types of model parameters is implemented with the ILUES algorithm.

Table 6. The prior distributions and reference true values of the model parameters estimated in Case 3. U stands for the uniform distribution.

Parameters	Prior Distribution	True Value
S_x	$U(60, 140)$	101.660
S_y	$U(100, 200)$	161.240
S_S	$U(200, 1000)$	498.830
$log_{10}(k_1)$	$U(-5, -2)$	−2.312
$log_{10}(k_2)$	$U(-5, -2)$	−2.517
$log_{10}(k_3)$	$U(-5, -2)$	−2.698
$log_{10}(k_4)$	$U(-5, -2)$	−3.043

Owing to the large quantity of unknown parameters, the ensemble size (N_e) of ILUES in this case is set to 1000. Based on the study in Case 2, instead of setting the factor α as a constant at all times, α is

randomly generated in the range of (0.1, 0.2) at each iteration of ILUES. The iteration number (N_{Iter}) of ILUES is set to 10. To demonstrate the advantage of ILUES in estimation accuracy, the performance of ensemble smoother with multiple data assimilation (ES-MDA) [52] is evaluated with the same numerical model for comparison. The setting of ensemble size (N_e) and iteration number (N_{Iter}) in ES-MDA is also the same as ILUES. As shown in Figure 12, even though the mean values of some estimated parameters (e.g., S_s, k_1, k_2, k_3) are close to the reference true values, the uncertainty in model parameters is still large in the last few iterations and some estimated parameters at the last iteration (e.g., S_x, S_y, k_4) deviate from the true values after the implementation of ES-MDA. In addition, a large number of samples in the last ensemble are still far from the true value, which means that ES-MDA is unable to achieve sufficient estimation accuracy. In contrast to the inversion results obtained from ES-MDA, the ILUES algorithm can provide more satisfactory results with higher estimation accuracy. As illustrated in Figure 13, the range of estimated values of model parameters gradually decreases as the number of iterations increases and it shrinks to a narrow range around the true value after ten iterations, which means that the uncertainty in model parameters is significantly reduced after ten iterations using ILUES. Furthermore, nearly all estimated values of model parameters approximate the true values after ten iterations, which demonstrates the superior accuracy of ILUES for model parameter estimation. In consequence, the joint estimation of contaminant source characteristics and biochemical parameters can be achieved with the ILUES algorithm.

To further verify the advantage of ILUES, the RMSE between the reference true values and the estimated values of selected model parameters is introduced to evaluate the estimation accuracy.

The RMSE of selected model parameters estimated by ILUES and ES-MDA are listed in Table 7. As clearly shown in Table 7, the RMSE values of estimated model parameters with the implementation of ILUES are all lower than those using ES-MDA. As lower RMSE values indicate higher estimation accuracy, a conclusion can be drawn that ILUES has an obvious advantage over ES-MDA on the joint estimation of contaminant source characteristics and biochemical parameters.
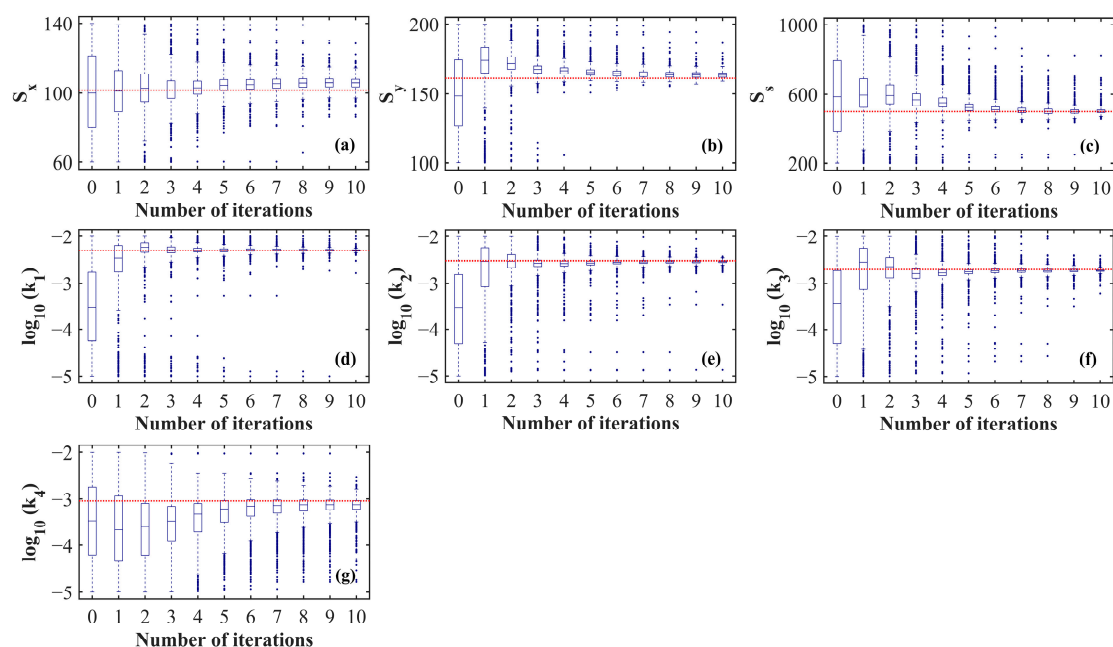
Figure 12. Box plots of model parameters estimated by ensemble smoother with multiple data assimilation (ES-MDA) in Case 3: (**a**,**b**) source location (S_x, S_y), (**c**) source strength (S_s), (**d**–**g**) the four rate constants (k_1, k_2, k_3, k_4). The red dashed lines denote the reference true values of corresponding model parameters.
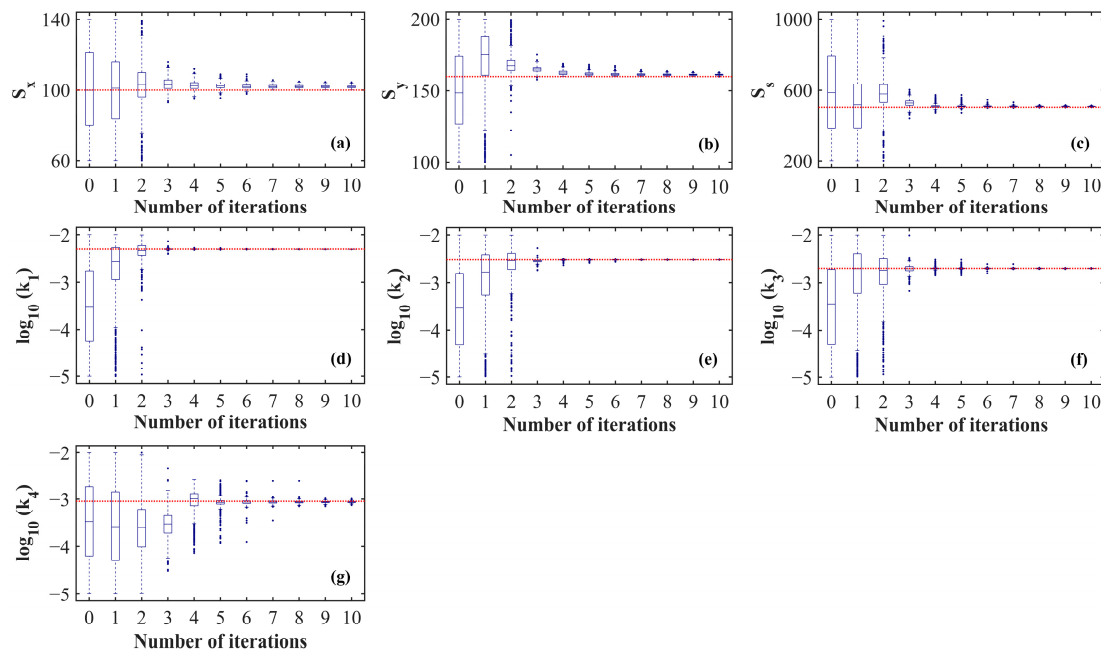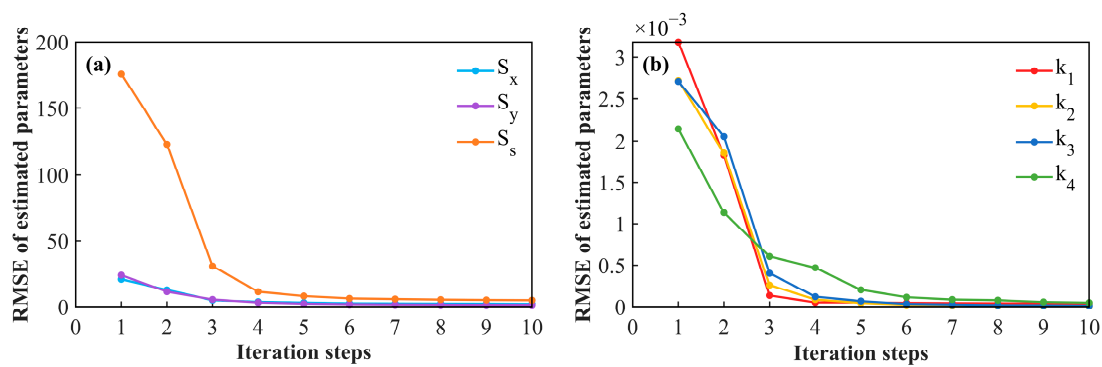
Figure 13. Box plots of model parameters estimated by ILUES in Case 3: (**a**,**b**) source location (S_x, S_y), (**c**) source strength (S_s), (**d–g**) the four rate constants (k_1, k_2, k_3, k_4). The red dashed lines denote the true values of corresponding model parameters.

**Table 7.** The RMSE of selected model parameters estimated by ILUES and ES-MDA.

	RMSE	
Parameters	**ILUES**	**ES-MDA**
S_x	2.02550	5.34737
S_y	1.37795	3.42515
S_s	4.89125	23.14605
k_1	0.00305	0.03113
k_2	0.00122	0.08848
k_3	0.00274	0.05292
k_4	0.02405	0.25088

The estimation accuracy of ILUES is also evaluated by the RMSE of model parameters considered in this case study. As illustrated in Figure 14, the RMSE values of all parameters decrease evidently with the increase of the iteration number and finally approximate 0 after 10 iterations. Thus, the ILUES algorithm can provide sufficient accuracy for the joint estimation of biochemical parameters and contaminant source characteristics of the reactive transport model.

Meanwhile, with the 68 KL terms inversed by ILUES, the hydraulic conductivity filed can be calculated from the KL expansion equation (Equation (17)). In this case, posterior samples in the ensemble of the 10th iteration are chosen to calculate the distribution of the log-conductivity field. Six posterior realizations, the mean estimate and the estimation variance of the log-conductivity field at the 10th iteration are presented in Figure 15. The reference log-conductivity field which is randomly generated at the beginning of the simulation is also presented in Figure 15. As illustrated in the figure, estimations of the log-conductivity field of the six posterior realizations are highly similar to the reference log-conductivity field and the estimation variance is reduced to a relatively low level after assimilation, which demonstrates the effectiveness of ILUES on hydraulic parameter estimation. The slight differences between the reference and the estimated log-conductivity field might be caused by the limited and sparse observation wells for each contaminant. To summarize, the ILUES algorithm can provide accurate joint estimations for the three types of model parameters mentioned above.

Figure 14. The RMSE of selected model parameters estimated in Case 3. (**a**) represents the RMSE of contaminant source characteristics at different iteration steps. (**b**) represents the RMSE of biochemical parameters at different iteration steps. Different coloured lines denote the RMSE of each parameter at different iteration steps.
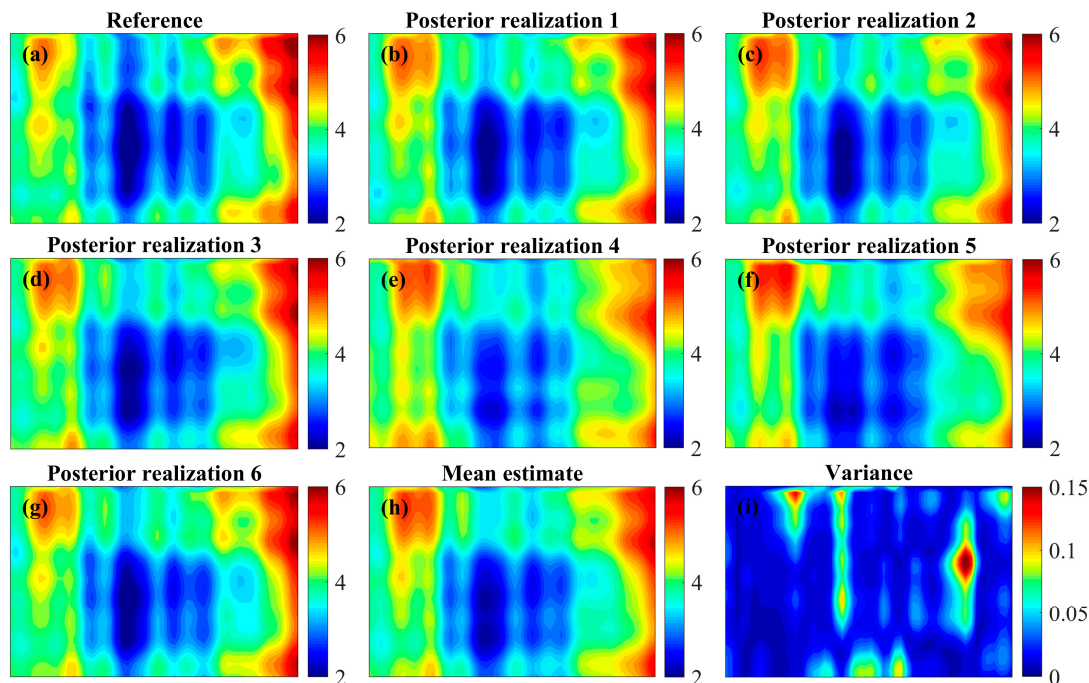
Figure 15. (**a**) is the reference log-conductivity field. (**b**–**g**) represent six posterior realizations of the log-conductivity field generated from the last ensemble. (**h**) represents the mean estimate of the log-conductivity field of the last ensemble. (**i**) represents the estimation variance of the log-conductivity field of the last ensemble.

4. Conclusions

In this study, an iterative local updating ensemble smoother (ILUES) method is employed to jointly estimate the hydraulic parameters, biochemical parameters and contaminant source characteristics of a reactive transport model. To better explore the parameter space, the original ILUES algorithm is modified by determining the local ensemble partly with a linear ranking selection scheme [56]. The reactive transport model is constructed to simulate the sequential anaerobic biodegradation process of tetrachloroethene (PCE). The applicability and accuracy of ILUES is evaluated by three numerical case studies. In all case studies, the uncertainty in model parameters is significantly reduced with the implementation of ILUES, which demonstrates the validity of ILUES for joint parameter estimation.

The inversion of unknown model parameters is of vital importance to predict the fate and transport of reactive contaminants in groundwater. The Markov Chain Monte Carlo (MCMC) is the

most widely used method to approximate the posterior distribution of unknown parameters. However, as MCMC methods typically require a large number of model evaluations, they are considered to be computationally intensive and may not be applicable to high-dimensional inverse modeling of reactive transport. The ensemble Kalman filter (EnKF) and ensemble smoother (ES), which are based on the theory of data assimilation, are more efficient alternatives to MCMC for parameter estimation in nonlinear systems. Nevertheless, for highly nonlinear problems, both the standard ES and standard EnKF methods cannot provide satisfactory inversion results with sufficient accuracy. Under these circumstances, the ILUES algorithm is developed to improve the accuracy and efficiency of parameter estimation for highly nonlinear reactive transport models.

It is noteworthy that the modified ILUES algorithm is compared with the ensemble smoother with multiple data assimilation (ES-MDA), which is an efficient iterative ensemble smoother method, in terms of the estimation accuracy of model parameters. Although the ES-MDA method can actually reduce the parameter uncertainty of the model, the ILUES algorithm provides better inversion results with higher accuracy, which demonstrates the advantage of ILUES.

The ensemble size N_e and factor α have obvious impacts on the estimation accuracy of model parameters. When α is either too small or too large, the inversion results are far from satisfactory. The estimation accuracy is usually better with $\alpha = 0.1$ and $\alpha = 0.2$ when the value of N_e is the same. Additionally, the estimation accuracy usually enhances with the increase of N_e, but the computational cost will increase in the meantime. Therefore, methods for determining the optimal balance between estimation accuracy and computational cost are worthy of further study.

In this paper, the reactive transport model is based on the sequential anaerobic biodegradation of PCE. The results can provide valuable references for solving inverse problems of other reactive contaminants (e.g., ionic reactions between inorganic contaminants). Although the ILUES algorithm is proved to be a valid method for solving inverse problems in this study, it may not be so useful in some practical contamination assessments. For example, when dealing with high-dimensional nonlinear problems and large-scale inverse modeling, the computational cost can be extremely high and even prohibitive. To address this issue, construction of the surrogate model which has a similar accuracy and a low computational cost is considered to be a hopeful solution.

The estimation accuracy of ILUES may also be hindered by the lack of information because only two types of measurements (contaminant concentration and hydraulic head) are assimilated in this study. In order to further improve the estimation accuracy of unknown parameters, our future works may focus on the simultaneous assimilation of multiple measurements (e.g., porosity, hydraulic conductivity and geophysical data).

Author Contributions: Conceptualization, S.J.; Formal analysis, X.X.; Funding acquisition, G.Z.; Methodology, X.X.; Project administration, S.J.; Resources, N.Z.; Software, R.Z.; Supervision, S.J.; Validation, R.Z.; Visualization, R.Z.; Writing—original draft, R.Z. All authors have read and agreed to the published version of the manuscript.

References

1. Moran, M.J.; Zogorski, J.S.; Squillace, P.J. Chlorinated Solvents in Groundwater of the United States. *Environ. Sci. Technol.* **2007**, *41*, 74–81. [CrossRef] [PubMed]
2. Conrad, M.E.; Brodie, E.L.; Radtke, C.W.; Bill, M.; Delwiche, M.E.; Lee, M.H.; Swift, D.L.; Colwell, F.S. Field Evidence for Co-Metabolism of Trichloroethene Stimulated by Addition of Electron Donor to Groundwater. *Environ. Sci. Technol.* **2010**, *44*, 4697–4704. [CrossRef]

3. Azubuike, C.C.; Chikere, C.B.; Okpokwasili, G.C. Bioremediation techniques-classification based on site of application: Principles, advantages, limitations and prospects. *World J. Microbiol. Biotechnol.* **2016**, *32*, 1–18. [CrossRef]

4. Tiehm, A.; Schmidt, K.R. Sequential anaerobic/aerobic biodegradation of chloroethenes-aspects of field application. *Curr. Opin. Biotechnol.* **2011**, *22*, 415–421. [CrossRef]

5. Bradley, P.M. History and Ecology of Chloroethene Biodegradation: A Review. *Bioremediat. J.* **2003**, *7*, 81–109. [CrossRef]

6. Vainberg, S.; Condee, C.W.; Steffan, R.J. Large-scale production of bacterial consortia for remediation of chlorinated solvent-contaminated groundwater. *J. Ind. Microbiol. Biotechnol.* **2009**, *36*, 1189–1197. [CrossRef]

7. Celia, M.A.; Kindred, J.S.; Herrera, I. Contaminant Transport and Biodegradation 1. A Numerical Model for Reactive Transport in Porous Media. *Water Resour. Res.* **1989**, *25*, 1141–1148. [CrossRef]

8. Schäfer, D.; Schäfer, W.; Kinzelbach, W. Simulation of reactive processes related to biodegradation in aquifers 1. Structure of the three-dimensional reactive transport model. *J. Contam. Hydrol.* **1998**, *31*, 167–186. [CrossRef]

9. Torlapati, J.; Clement, P. Using Parallel Genetic Algorithms for Estimating Model Parameters in Complex Reactive Transport Problems. *Processes* **2019**, *7*, 640. [CrossRef]

10. Hantush, M.M.; Mariño, M.A. Estimation of Spatially Variable Aquifer Hydraulic Properties Using Kalman Filtering. *J. Hydraul. Eng.* **1997**, *123*, 1027–1035. [CrossRef]

11. Michael, H.A.; Khan, M.R. Impacts of physical and chemical aquifer heterogeneity on basin-scale solute transport: Vulnerability of deep groundwater to arsenic contamination in Bangladesh. *Adv. Water Resour.* **2016**, *98*, 147–158. [CrossRef]

12. Ptak, T.; Liedl, R. *Modeling of Reactive Contaminant Transport in Hydraulically and Hydrogeochemically Heterogeneous Aquifers Using a Geostatistical Facies Approach BT—geoENV IV—Geostatistics for Environmental Applications*; Sanchez-Vila, X., Carrera, J., Gómez-Hernández, J.J., Eds.; Springer: Dordrecht, The Netherlands, 2004; pp. 247–258.

13. Wagner, B.J. Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling. *J. Hydrol.* **1992**, *135*, 275–303. [CrossRef]

14. Bailey, R.T.; Baù, D. Estimating spatially-variable first-order rate constants in groundwater reactive transport systems. *J. Contam. Hydrol.* **2011**, *122*, 104–121. [CrossRef] [PubMed]

15. Mo, S.; Zabaras, N.; Shi, X.; Wu, J. Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification. *Water Resour. Res.* **2019**, *55*, 3856–3881. [CrossRef]

16. Bailey, R.T.; Baù, D.A.; Gates, T.K. Estimating spatially-variable rate constants of denitrification in irrigated agricultural groundwater systems using an Ensemble Smoother. *J. Hydrol.* **2012**, *468–469*, 188–202. [CrossRef]

17. Dai, Z.; Samper, J. Inverse problem of multicomponent reactive chemical transport in porous media: Formulation and applications. *Water Resour. Res.* **2004**, *40*, 1–18. [CrossRef]

18. Carniato, L.; Schoups, G.; van de Giesen, N. Inference of reactive transport model parameters using a Bayesian multivariate approach. *Water Resour. Res.* **2014**, *50*, 6406–6427. [CrossRef]

19. Li, L.; Zhou, H.; Gómez-Hernández, J.J.; Hendricks Franssen, H.J. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter. *J. Hydrol.* **2012**, *428–429*, 152–169. [CrossRef]

20. 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. Contam. Hydrol.* **2018**, *216*, 50–57. [CrossRef]

21. Lan, T.; Shi, X.; Jiang, B.; Sun, Y.; Wu, J. Joint inversion of physical and geochemical parameters in groundwater models by sequential ensemble-based optimal design. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1919–1937. [CrossRef]

22. Srivastav, R.K.; Sudheer, K.P.; Chaubey, I. A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resour. Res.* **2007**, *43*, 1–12. [CrossRef]

23. Zhang, G.; Lu, D.; Ye, M.; Gunzburger, M.; Webster, C. An adaptive sparse-grid high-order stochastic collocation method for Bayesian inference in groundwater reactive transport modeling. *Water Resour. Res.* **2013**, *49*, 6871–6892. [CrossRef]

24. Chandra, R.; Azam, D.; Müller, R.D.; Salles, T.; Cripps, S. Bayeslands: A Bayesian inference approach for parameter uncertainty quantification in Badlands. *Comput. Geosci.* **2019**, *131*, 89–101. [CrossRef]

25. Emery, J.M.; Grigoriu, M.D.; Field, R.V. Bayesian methods for characterizing unknown parameters of material models. *Appl. Math. Model.* **2016**, *40*, 6395–6411. [CrossRef]

26. Sharma, S. Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. *Annu. Rev. Astron. Astrophys.* **2017**, *55*, 213–259. [CrossRef]

27. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [CrossRef]

28. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [CrossRef]

29. Cao, T.; Zeng, X.; Wu, J.; Wang, D.; Sun, Y.; Zhu, X.; Lin, J.; Long, Y. Integrating MT-DREAMzs and nested sampling algorithms to estimate marginal likelihood and comparison with several other methods. *J. Hydrol.* **2018**, *563*, 750–765. [CrossRef]

30. Shi, X.; Ye, M.; Curtis, G.P.; Miller, G.L.; Meyer, P.D.; Kohler, M.; Yabusaki, S.; Wu, J. Assessment of parametric uncertainty for groundwater reactive transport modeling. *Water Resour. Res.* **2014**, *50*, 4416–4439. [CrossRef]

31. Vrugt, J.A.; Gupta, H.V.; Bouten, W.; Sorooshian, S. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **2003**, *39*. [CrossRef]

32. Yan, X.; Dong, W.; An, Y.; Lu, W. A Bayesian-based integrated approach for identifying groundwater contamination sources. *J. Hydrol.* **2019**, *579*, 124160. [CrossRef]

33. Cui, T.; Marzouk, Y.M.; Willcox, K.E. Data-driven model reduction for the Bayesian solution of inverse problems. *Int. J. Numer. Methods Eng.* **2015**, *102*, 966–990. [CrossRef]

34. Guo, Q.; Dai, F.; Zhao, Z. Comparison of Two Bayesian-MCMC Inversion Methods for Laboratory Infiltration and Field Irrigation Experiments. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1108. [CrossRef] [PubMed]

35. Oyanagi, R.; Okamoto, A.; Tsuchiya, N. Multiple Kinetic Parameterization in a Reactive Transport Model Using the Exchange Monte Carlo Method. *Minerals* **2018**, *8*, 579. [CrossRef]

36. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]

37. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **1994**, *99*. [CrossRef]

38. Robert, S.; Künsch, H.R.; Group, F.; Robert, B.S.; Künsch, H.R. Localizing the Ensemble Kalman Particle Filter. *Tellus A Dyn. Meteorol. Oceanogr.* **2017**, *870*, 1–15. [CrossRef]

39. Houtekamer, P.L.; Mitchell, H.L.; Pellerin, G.; Buehner, M.; Charron, M.; Spacek, L.; Hansen, B. Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Mon. Weather Rev.* **2005**, *133*, 604–620. [CrossRef]

40. ELSheikh, A.H.; Pain, C.C.; Fang, F.; Gomes, J.L.M.A.; Navon, I.M. Parameter estimation of subsurface flow models using Iterative Regularized Ensemble Kalman Filter. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 877–897. [CrossRef]

41. Moradkhani, H.; Sorooshian, S.; Gupta, H.V.; Houser, P.R. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* **2005**, *28*, 135–147. [CrossRef]

42. Sun, L.; Seidou, O.; Nistor, I.; Liu, K. Review of the Kalman-type hydrological data assimilation. *Hydrol. Sci. J.* **2016**, *61*, 2348–2366. [CrossRef]

43. Wang, D.; Chen, Y.; Cai, X. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resour. Res.* **2009**, *45*, 1–13. [CrossRef]

44. Stroud, J.R.; Katzfuss, M.; Wikle, C.K. A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation. *Mon. Weather Rev.* **2018**, *146*, 373–386. [CrossRef]

45. Emerick, A.A.; Reynolds, A.C. Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **2013**, *55*, 3–15. [CrossRef]

46. van Leeuwen, P.J.; Evensen, G. Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. *Mon. Weather Rev.* **1996**, *124*, 2898–2913. [CrossRef]

47. Skjervheim, J.A.; Evensen, G. An Ensemble Smoother for assisted History Matching. In Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, TX, USA, 21–23 February 2011; Volume 2, pp. 1049–1063. [CrossRef]

48. Chen, Y.; Oliver, D.S. Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **2013**, *17*, 689–703. [CrossRef]

49. White, J.T. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environ. Model. Softw.* **2018**, *109*, 191–201. [CrossRef]

50. Ju, L.; Zhang, J.; Meng, L.; Wu, L.; Zeng, L. An adaptive Gaussian process-based iterative ensemble smoother for data assimilation. *Adv. Water Resour.* **2018**, *115*, 125–135. [CrossRef]

51. Chen, Y.; Oliver, D.S. Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother. *Math. Geosci.* **2012**, *44*, 1–26. [CrossRef]

52. Emerick, A.A.; Reynolds, A.C. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Comput. Geosci.* **2012**, *16*, 639–659. [CrossRef]

53. Li, L.; Davis, A. Data assimilation in groundwater modelling: Ensemble Kalman filter versus ensemble smoothers. *Hydrol. Process.* **2018**, *32*, 2020–2029. [CrossRef]

54. Ma, X.; Bi, L. A robust adaptive iterative ensemble smoother scheme for practical history matching applications. *Comput. Geosci.* **2019**, *23*, 415–442. [CrossRef]

55. Zhang, J.; Lin, G.; Li, W.; Wu, L.; Zeng, L. An Iterative Local Updating Ensemble Smoother for Estimation and Uncertainty Assessment of Hydrologic Model Parameters With Multimodal Distributions. *Water Resour. Res.* **2018**, *54*, 1716–1733. [CrossRef]

56. Yadav, S.L.; Sohal, A. Comparative Study of Different Selection Techniques in Genetic Algorithm. *Int. J. Eng. Sci. Math.* **2017**, *6*, 174–180.

57. Harbaugh, A.W.; Banta, E.R.; Hill, M.C.; McDonald, M.G. Modflow-2000, the US geological survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. *Open File Rep. U. S. Geol. Surv.* **2000**, *92*, 134.

58. Clement, T.P. *RT3D—A Modular Computer Code for Simulating Reactive Multi-Species Transport in 3-dimensional Groundwater Aquifers*; Pacific Northwest National Lab.: Richland, WA, USA, 1997; pp. 1–59. [CrossRef]

59. Clement, T.P.; Sun, Y.; Hooker, B.S.; Petersen, J.N. Modeling Multispecies Reactive Transport in Ground Water. *Groundw. Monit. Remediat.* **1998**, *18*, 79–92. [CrossRef]

60. Zhang, D.; Lu, Z. An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loève and polynomial expansions. *J. Comput. Phys.* **2004**, *194*, 773–794. [CrossRef]