# Identification of Irregular Non-Point Contaminant Sources Using Ensemble Smoother with Multiple Data Assimilation and Closed Cubic B-Spline Curve Approximation

Wenjun Zhang[a,b,d], Teng Xu[a,b,*], J. Jaime Gómez-Hernández[f], Zi Chen[e], Chunhui Lu[a,c], Guodong Zhang[a,b]

[a]The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, China

[b]College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, China

[c]Yangtze Institute for Conservation and Development, Hohai University, Nanjing, China

[d] National Marine Environmental Monitoring Center, 116023 Dalian, China

[e] Nanjing Center, China Geological Survey, Nanjing, China

[f] Institute of Water and Environmental Engineering, Universitat Politècnica de València, Valencia, Spain

* **Corresponding author**.

*E-mail address:* teng.xu@hhu.edu.cn (Teng Xu)

**Abstract:**

Reconstructing the spatial and temporal characteristics of non-point sources is essential for early warning and precise restoration of groundwater pollution. Currently, the identification of non-point contaminant sources is primarily focused on those with simple regular spatial architecture, such as rectangular or circular shapes, while the characterization of irregular non-point contaminant sources has been overlooked, which is attributed to their complex and diverse spatial structures that cannot be readily captured by predefined geometric parameters such as centroid coordinates and side lengths. To address this challenge, we propose a novel approach, an ensemble smoother with multiple data assimilation combined with a closed cubic B-spline curve approximation (ES-MDA-Bs), for the identification of irregular non-point sources. The effectiveness of ES-MDA-Bs in identifying irregular non-point sources in synthetic confined aquifers is assessed and the influence of varying numbers and placements of knot points on its performance for this purpose explored. Then, a comparative analysis between ES-MDA-Bs and the ensemble smoother with multiple data assimilation combined with a rotating ellipse approximation (ES-MDA-RE) ---which has previously demonstrated competency in identifying ellipse-like sources---is performed. The results show that ES-MDA-Bs, with sufficient knot points to approximate an irregular shape, can effectively identify both the irregular spatial structure of the source and its release parameters, i.e., the initial release time, release duration, and mass loading. ES-MDA-Bs clearly outperforms ES-MDA-RE in identifying irregular non-point sources. The accuracy and uncertainty of ES-MDA-Bs are improved as the number of knot points increases.

**Keyword**: Irregular contaminant source identification; Ensemble smoother with multiple data assimilation; Groundwater contamination

## 1. Introduction

Groundwater contamination threatens surrounding ecosystems and may disrupt economic activities. A fundamental approach to mitigate this threat involves tracing the contaminant back to its source and try to accurately reproduce the contaminant plume evolution within an aquifer. Source tracking involves reconstructing its spatial and temporal release characteristics, which are essential for accurate risk assessments and the development of effective remediation strategies (Srivastava and Singh, 2015).

Based on their characteristics, contaminant sources can be classified as either point or non-point. Point sources refer to localized and discrete origins of pollution, often resulting from accidents at specific sites, such as chemical spills or leaks from landfill sites. In contrast, non-point sources are diffuse and continuous, originating from activities that cover large areas, such as fertilizer leaching from agricultural irrigation or effluents from chemical plants. Compared to point sources, non-point sources exert broader geographical and longer-lasting effects on the aquifer, and their spatial distribution often exhibits irregular shapes. Thus, accurately characterizing the irregular spatial structure of a non-point source is a critical factor for effective pollution control.

In the past, due to technical and computational limitations, researchers working on the estimation of source parameters have concentrated in either single point sources (e.g., Hwang and Koerner, 1983; Jha and Datta, 2015; Xu and Gómez-Hernández, 2016, 2018; Chen et al., 2018; Hou et al., 2021; An et al., 2022; Bai and Tahmasebi, 2022; Chang et al., 2022; Ge et al., 2023; Chang et al., 2024) or multiple discrete point sources (e.g., Gorelick et al., 1983; Mahar and Datta, 1997; Hwang et al., 2020; Jamshidi et al., 2020; Anshuman and Eldho, 2022; Kontos et al., 2022; Anshuman and Eldho, 2023; Chen et al., 2023a; Li et al., 2023a; Singh and Mahor, 2023; Wu et al., 2025).

Recently, non-point source identification studies have gradually gained attention. However, the limited studies on non-point source reconstruction have predominantly focused on sources with regular shapes. The existing research on non-point sources can be categorized into two groups based on how the sources are treated during the identification process. The first group involves only identifying the release information parameters of sources, with the spatial structure and location assumed to be known. For instance, in Butera and Tanda (2003) and Pan et al. (2023a), the release history was estimated from a square-shaped source that is uniform in space but changes over time. Both Zhang et al. (2024a) and Zheng et al. (2024) provided accurate estimations of the release strength from a linear contaminant source. Other studies found contaminant release intensities from multiple discrete potential sources with known locations and irregular shapes (e.g., Gzyl et al., 2014; Pan et al., 2021, 2022, 2023b, 2025; Li et al., 2023b; Luo et al., 2023). The second group focuses on simultaneously identifying both the release information and the spatial structure of regular-shaped sources. For instance, Mahinthakumar and Sayeed (2005) and Jin et al. (2009) determined the location and shape of a prismatic source by identifying the coordinates of two diagonally-opposed vertices, followed by estimating its release concentration. Mirghani et al. (2009) and Mirghani et al. (2012) reconstructed a cubic areal source by identifying its centroid and side length, and reproduced its release information by calculating the initial concentration. Ayvaz (2016) was the first to try to figure out irregularly shaped sources. He used a new simulation-optimization method to find both the release information and the locations of discrete elements that, when put together, made up the shape of the source. However, when dealing with large-scale non-point sources, this approach may introduce significant errors and computational burdens. Due to the diverse spatial characteristics of non-point sources, inherent uncertainties are substantial, rendering traditional point source identification methods inadequate for addressing non-point source issues. Notably, to the best of our knowledge, current scientific literature lacks effective methodologies capable of systematically characterizing the spatial structure of irregular non-point sources. Hence, a logical and streamlined method is suggested for accurately describing the spatial distribution of irregular sources of any size and shape, as long as some background information about where the contamination came from is known.

In our earlier work (Xu et al., 2022), we successfully applied the rotation ellipse approach, which involved approximating a non-point source with an ellipse with five parameters (center point coordinates, semi-major and semi-minor axes, and rotation angle). In this work, we propose a major improvement by shifting to the use of spline curves, which are characterized by a number of knot points, to identify irregular shapes.

The ensemble smoother with multiple data assimilation is combined with a closed cubic B-spline curve approximation (ES-MDA-Bs) to deal with irregular non-point source identification problems. It was decided to use the ensemble smoother with multiple data assimilations (ES-MDA) to solve the inverse problem because it has been shown to be good at finding sources in both synthetic (e.g., Xu et al., 2021, 2022) and experimental cases (Chen et al., 2023b). To see how well ES-MDA-Bs can approximate the source's spatial distribution, we try out different configurations by changing the number and placement of knot points. We also perform a comparison between ES-MDA-Bs and the ensemble smoother with multiple data assimilation combined with a rotating ellipse approximation (ES-MDA-RE) to see how well different approximation methods work for finding irregular non-point sources. Note that our study aims to investigate the capability of ES-MDA-Bs to estimate the location, spatial structure, and release information of an irregular non-point source.

The structure of this paper is organized as follows: In Section 2, we talk about the algorithms behind both ES-MDA-Bs and ES-MDA-RE in detail. In Section 3, we set up three scenarios to test how well ES-MDA-Bs and ES-MDA-RE work in a confined aquifer. Finally, in Section 4, we compare and analyze the results, and in Section 5, we talk about what needs to be done in the future. Finally, in Section 6, we wrap up the paper with a full summary.

## 2. Methodology

We propose the ES-MDA-Bs algorithm to identify the parameters of a non-point source with an irregular spatial structure. To demonstrate the effectiveness of ES-MDA-Bs, we use the ES-MDA-RE algorithm for comparison. The computational procedures of both algorithms are described in the following sections.

### 2.1 Ensemble smoother with multiple data assimilation combined with a closed cubic B-spline curve approximation (ES-MDA-Bs)

ES-MDA-Bs is a data assimilation method designed to identify complex source with a spatial irregular shape. It integrates two key components: ES-MDA for parameter identification and a closed B-spline curve to delineate the source shape through a number of knot points. This approach assimilates concentration observations to determine the coordinates of the knot points on the spline curve and the source release parameters, enabling the reconstruction of both the spatial structure and release history of the source.

#### 2.1.1 Ensemble smoother with multiple data assimilations (ES-MDA)

ES-MDA is an ensemble-based data assimilation approach proposed by Emerick and Reynolds (2013), that incorporates an iterative scheme to handle the non-linear state equations in hydrological modeling. It is an evolution of the ensemble smoother (Evensen and Van Leeuwen, 2000), which, in turn, evolves from the ensemble Kalman filter (Li et al., 2012). This approach assimilates observations of state variables from all time steps into the ensemble of model simulations at once. The computational process can be typically divided into two key steps: forecast and update.

During the forecast step, a forward model, represented by equation $\Psi(\bullet)$, forecasts the state variables $C_j^f$ (in our case, solute concentrations) for all time steps at the $j^{th}$ assimilation iteration, using both the initial state variables $C_0$ and the model parameters (in our case, the parameters defining the source location and release history) determined during the previous iteration $P_{j-1}^a$, as model inputs. Note that forecasting from $C_0$ implies that each forward model simulation is run from the same initial state at time 0.

$$C_j^f = \Psi(C_0, P_{j-1}^a) \tag{1}$$

During the update step, observation data from all time steps are assimilated to refine the model parameters. The assimilation process adjusts the model parameters using the discrepancy between forecasted states at observation locations $C_j^{f,o}$ and observation data in the same wells. The objective is to obtain an updated parameter vector $P_j^a$ at the $j^{th}$ iteration based on the parameter state from the previous iteration $P_{j-1}^a$ aimed to reducing the discrepancies between forecasts and observations,

$$P_j^a = P_{j-1}^a + K_j(C^o + \sqrt{a_j}\varepsilon_j - C_j^{f,o}) \tag{2}$$

with

$$K_j = D_{PC,j}(D_{CC,j} + a_j R)^{-1} \tag{3}$$

where $C^o + \sqrt{a_j}\varepsilon_j$ is the observed data already affected by an observation error $\varepsilon_j$, which is damped by factor $a_j$. According to Emerick and Reynolds (2013), the damping factors should be chosen as to satisfy $\sum_{j=1}^{Na}\frac{1}{a_j} = 1$ and in decreasing order following an arithmetic or a geometric regression (specifically, $a_j$ is equal to the total number of assimilation iterations $Na$ for all iterations). $K_j$ is the Kalman gain and depends on the error covariance $R$, the cross-covariance $D_{PC,j}$ between parameters and states at the observation locations at all time steps and the auto-covariance $D_{CC,j}$ between states at observation locations at all time steps at the $j^{th}$ iteration.

In this study, we evaluate the performances of ES-MDA for two different implementations: The ellipse-base method and the spline-base one. The source parameters to be identified $P$ will be the initial release time $T^0$ [T], the release duration $\Delta T$ [T], the constant release mass-loading $M$ [MT$^{-1}$], and the geometric parameters $Z$ ($Z = Z_s$ or $Z_e$, Z varies depending on the source shape approximation method, as will be explained in the following subsections).

$$P = [Z, T^0, \Delta T, M]^T \tag{4}$$

ES-MDA serves as the core inversion algorithm in this study. The workflow is summarized in Figure 1, where the source parameters $P$ vary according to the source shape approximation method employed.

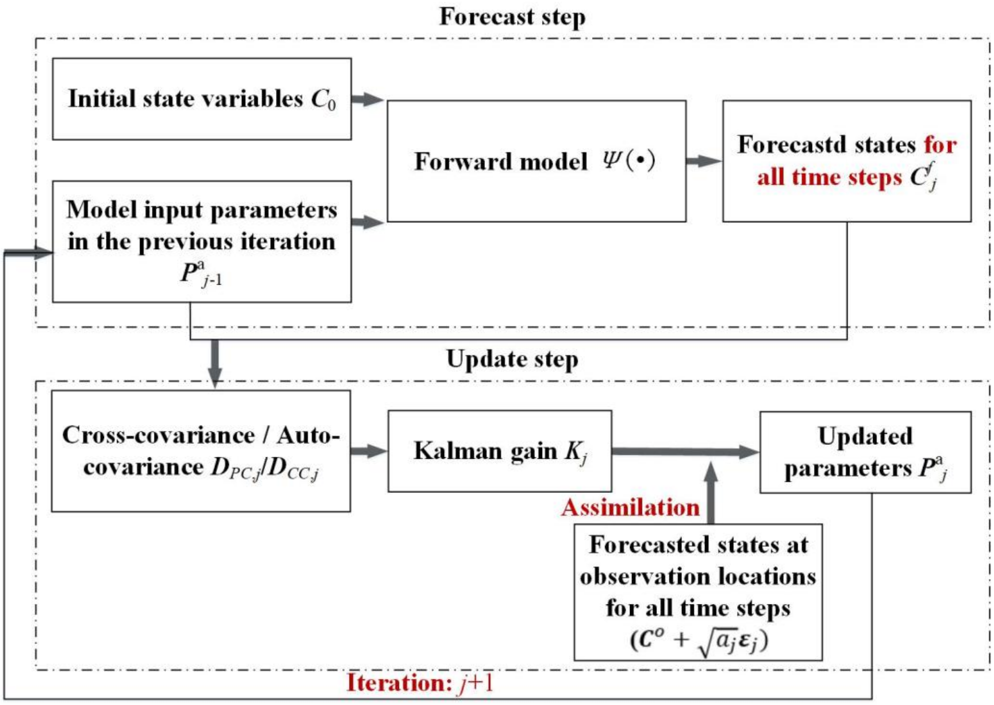
Forecast step
Initial state variables $C_0$
Forward model $\Psi(\bullet)$
Forecastd states for all time steps $C^f_j$
Model input parameters in the previous iteration $P^a_{j-1}$
Update step
Cross-covariance / Auto-covariance $D_{PC_j}/D_{CC_j}$
Kalman gain $K_j$
Updated parameters $P^a_j$
Assimilation
Forecasted states at observation locations for all time steps $(C^o + \sqrt{a_j}\varepsilon_j)$
Iteration: $j$+1

**Figure 1**: Workflow of ES-MDA.

### 2.1.2 Closed cubic B-spline curve approximation

The B-spline curve, initially proposed by Isaac Jacob (Schoenberg, 1946), has gained significant popularity in computer-aided design and numerical analysis due to its versatility (e.g., De Boor and De Boor, 1978; Piegl and Tiller, 1996; Park and Lee, 2007; Farin, 2014). The curve, characterized by degree $k$, is expressed as the sum of basis functions $B_{ia,k}(t)$ multiplied by control points $A_{ia}$. The curve within data point segment $[D_i, D_{i+1})$ can be written as

$$Z_{S_i}(t) = \sum_{ia=i}^{i+k} B_{ia,k}(t)\, A_{ia} \tag{5}$$

where $ia$ is the minimum number of control points required to ensure the smoothness of the curve; $i$ is the number of knot points, $i = 0, \dots, n-1$. The basis function $B_{ia,k}(t)$ are given by

$$B_{ia,0}(t) = \begin{cases} 1 & t \in [D_i, D_{i+1}) \\ 0 & t \in otherwise \end{cases} \quad , k = 0 \tag{6}$$

and

$$B_{ia,k}(t) = \frac{1}{k!}\sum_{j=0}^{i+k-ia}(-1)^j \binom{k+1}{j}(t+i+k-ia-j)^k, k > 0 \tag{7}$$

where $\binom{k+1}{j}$ is the binomial coefficient representing the number of combinations of $j$ elements that can be chosen from a set of $(k+1)$ distinct elements. The cubic B-spline curve $(k=3)$ offers a good balance between smoothness and computational efficiency, making it well-suited for a wide range of applications.

By substituting $k = 3$ into Eq. (5), the cubic B-spline is given by

$$Z_{S_i}(t) = \frac{1}{6}(1-t)^3 A_i + \frac{1}{6}(3t^3 - 6t^2 + 4)A_{i+1} + \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1)A_{i+2} + \frac{1}{6}t^3 A_{i+3} \tag{8}$$

where the coordinates of $A_i$, $A_{i+1}$, $A_{i+2}$ and $A_{i+3}$ are computed to ensure that the curve passes through the knot points $D_i$ and $D_{i+1}$, and that there is first and second derivative continuity at these points. For readers unfamiliar with splines and the distinction between control points and knot points: To elaborate, control points are used to construct the curve and establish the relationship between the knot points and the resulting B-spline curve (see Eq. (5)). Conversely, knot points represent specific locations that the curve must pass through. The relationship between control points and knot points can be formally described in Eq. (9). For further details on fitting knot points to a curve, please refer to the following references (e.g., de Boor, 1978; Lyche and Mørken, 1999; Juhász and Hoffmann, 2004).

$$D_i = \frac{1}{6}(A_i + 4A_{i+1} + A_{i+2}) \tag{9}$$

Algorithm 1: ES-MDA-Bs

```
Generate initial ensembles of source parameters ($P = [Z_s, T^0, \Delta T, M]^T$).
for $j = 1: N_a$
    Forecast concentrations in the observation wells ($C_j^{f,o}$).
    Update source parameters ($P$) based on the misfits between forecasts and
    observations ($C^o + \sqrt{a_j}\varepsilon_j - C_j^{f,o}$).
end for
```

Consequently, the core steps of the ES-MDA-Bs algorithm are summarized in Algorithm 1. Note that the spatial shape of the source location ($Z_s$) is defined by a set of $n$ knot points to be identified

$$Z_s = [\{D_i\}, i = 0, \dots, n-1]^T \tag{10}$$

## 2.2 Ensemble smoother with multiple data assimilation combined with a rotating ellipse approximation (ES-MDA-RE)

In our previous work (Xu et al., 2022), we proposed ES-MDA-RE to identify irregular shapes that could be approximated by an ellipse. The assimilation process is similar to ES-MDA-Bs, with the primary difference that ES-MDA-RE uses a different parameterization of the contaminant source shape. While in ES-MDA-Bs the unknowns are the coordinates of the $n$ knot points, in ES-MDA-RE the unknowns ($Z_e$) are the parameters defining the ellipse: the center point coordinates $(Xs, Ys)$ [L, L], the lengths of the semi-major and semi-minor axes $Ra$ [L] and $Rb$ [L], and the clockwise rotation angle with respect to the $x$-axis $B$ [°].

$$Z_e = [Xs, Ys, Ra, Rb, B]^T \tag{11}$$

Algorithm 2: ES-MDA-RE

```
Generate initial ensembles of source parameters ($P = [Z_e, T^0, \Delta T, M]^T$).
for $j = 1{:}N_a$
    Forecast concentrations in the observation wells ($C_j^{f,o}$).
    Update source parameters ($P$) based on the mists between forecasts and observations ($C^o + \sqrt{a_j}\varepsilon_j - C_j^{f,o}$).
end for
```

## 3. Application

We construct a two-dimensional synthetic confined aquifer, discretized into a grid of 80×80×1 cells, with each cell size measuring 10 [L]×10 [L]×80 [L]. (No specific units will be used; any set of consistent units will result in the results presented.) In this study, the release of a non-reactive contaminant is modeled under a steady-state groundwater flow condition, and solute movement is driven solely by advection and dispersion. Consequently, the governing equations for the forward model include the steady-state groundwater flow based on Darcy's law and the principle of continuity,

$$Q = -\nabla \cdot (K\nabla H) \tag{12}$$

where $Q$ represents sources and sinks per unit volume [T$^{-1}$]; $\nabla \cdot$ is the divergence operator; $\nabla$ is the gradient operator; $K$ is the hydraulic conductivity [LT$^{-1}$]; and $H$ is the hydraulic head [L]; and the transport equation,

$$\frac{\partial(\theta_e \mathrm{C})}{\partial t} = \nabla \cdot [\theta_e(\alpha_m + \beta v) \cdot \nabla C] - \nabla \cdot (\theta_e v C) - q_c C_s \tag{13}$$

where $C$ is the contaminant source concentration [ML$^{-3}$]; $t$ is the simulation time [T]; $\alpha_m$ is the molecular diffusion coefficient [L$^2$T$^{-1}$]; $\beta$ is the dispersivity tensor [L]; $q_c$ is the volumetric flow rate per unit cross-section [LT$^{-1}$]; and $C_s$ is the concentration of the sources or sinks [ML$^{-3}$]; $\theta_e$ [-] is the effective porosity and $v$ [LT$^{-1}$] is the velocity given by $v = (-K\nabla H)/\theta_e$.

To simulate groundwater flow and solute transport processes, the forward model requires specific input parameters to solve for the state variables. Figure 2 presents the natural logarithm of the hydraulic conductivity field, which exhibits a heterogeneous Gaussian distribution generated using the GCOSIM3D program, a multivariate multi-Gaussian sequential simulation code (Gómez-Hernández and Journel, 1993). The Gaussian distribution is defined by the parameters listed in Table 1. The boundary conditions for the aquifer are as follows: the east and west boundaries are set to constant heads of 80 [L] and 300 [L], respectively, while the north and south boundaries are impermeable. Porosity is homogeneous and equal to 0.3 [-]. For the solute transport process, the initial concentration across the domain is set to zero, and the dispersion coefficient $\beta$ is assumed to be constant in both space and time. The total transport simulation time is 10950 [T], divided into 100 stress periods of 109.5 [T] each. The non-reactive contaminant source releases at a constant mass-loading rate from the 10th time step (approximately 985.5 [T]) to the 30th time step (around 3285.0 [T]), resulting in a release duration of 21-time steps (approximately 2299.5 [T]), as detailed in

Table 2. The observational errors are assumed to follow a normal distribution with zero mean and a variance of 0.01. Groundwater flow and solute transport processes are simulated using the MODFLOW and MT3D programs, respectively.

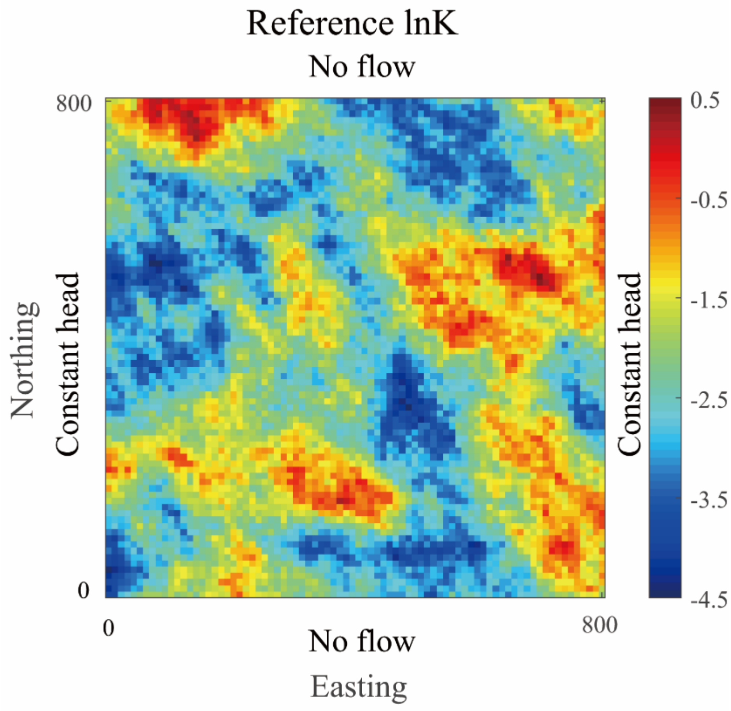

**Figure 2**: Reference log-conductivity field. An indication of the type of boundary conditions used for the solution of the flow equation is also shown.

**Table 1.** Parameters used to generate the lnK field

| | Mean | Std.dev. | Variogram | $\lambda_{max}$ | $\lambda_{min}$ | Angle |
|---|---|---|---|---|---|---|
| lnK | -2 | 1 | Spherical | 300 | 200 | 135 |

**Table 2.** Source release parameters

| Parameters | Values |
|---|---|
| Initial release time [$T^0$] | 985.5 |
| Release duration time [$\Delta T$] | 2299.5 |
| Mass loading rate [$M$] | 100 |

As mentioned, this study focuses on identifying non-point source with complex irregular spatial distribution. The reference shape for the source is presented in Figure 3 (a). To demonstrate the performance of ES-MDA-Bs in identifying irregular non-point sources, three scenarios are developed: in scenario S1 and S2, the spatial structure of the source is approximated using 10 and 5 knot points, respectively, which form a continuous path based on the initial position vector $D_i$ (Eq.10) and are identified by ES-MDA-Bs; and in scenario S3, the source is assimilated to an ellipse and its 5 defining parameters identified using ES-MDA-RE. Scenarios S1 and S2 are designed to evaluate the impact of the number of knot points on the approximation of the irregular source spatial structure using ES-MDA-Bs. Scenario S3 serves to compare the performance of the ellipse-based method and the spline-based one, eventually highlighting the advantage of using ES-MDA-Bs over ES-MDA-RE for irregular source identification. In all scenarios, the number of assimilation steps is also tested. The basic parameters defining the three scenarios are listed in Table 3.

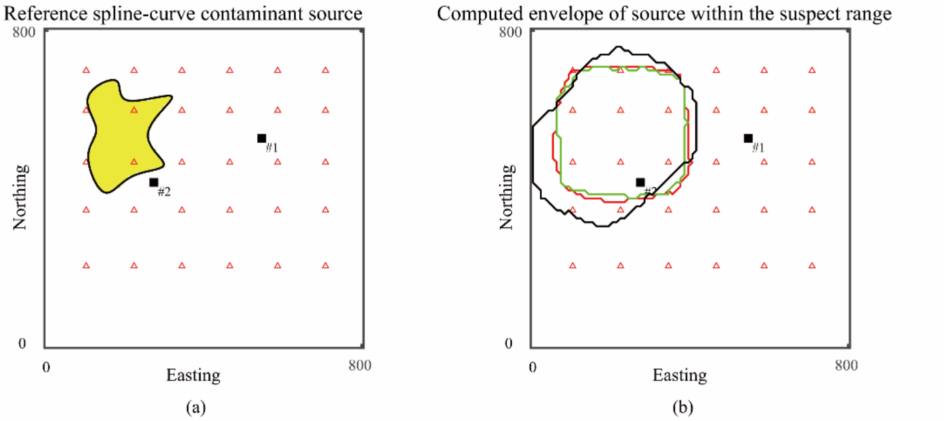

**Figure 3**: Reference contamination source (a) and the computed envelope of generated shapes within the suspect range for knot points and elliptical control parameters, used to initialize the parameters prior to the start of the assimilation process (b). The red triangles are the observation locations. The black squares are the validation locations (not used during the assimilation process). The suspects are in (b) correspond to scenarios S1 (black), S2 (green) and S3 (red).

For each scenario, we generate an ensemble of 500 realizations, consisting of 23-tuples for S1, 13-tuples for S2, and 8-tuples for S3. Within each tuple, the values are randomly drawn from a uniform distribution and used as initial inputs for the inversion process. The parameter ranges are broad to account for significant uncertainty and are chosen from uniform distributions as indicated next: $x$-coordinates of knot points $Xs \in U[80, 380]$, $y$-coordinates of knot points $Ys \in U[400, 700]$, initial release time $T^0 \in U[0, 3175.5]$, release duration $\Delta T \in U[1204.5, 6679.5]$ and mass-loading rate $M \in U[90, 150]$, $x$-coordinate of ellipse center $Xsc \in U[160, 200]$, $y$-coordinate of ellipse center $Ys \in U[480, 580]$, semi-major axis $Ra \in U[110, 210]$, semi-minor axis $Rb \in U[50, 120]$ and clockwise rotation angle $B \in U[0, 90]$. The prior suspected parameter ranges are summarized in Table 3. The envelops of all 500 realizations of initial guesses of the contaminant source are shown in Figure 3 (b). It is important to recall that the update step is non-convex, meaning that the updated parameter values can be outside the initial uncertainty intervals if the dynamics of the system indicate that the source was initially clearly off the reference.

**Table 3.** Definition of scenarios

| Scenario | | S1 | | | | S2 | | | | S3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of assimilation iterations [$la$] | | 0 | 1 | 4 | 7 | 0 | 1 | 4 | 7 | 0 | 1 | 4 | 7 |
| Approximation | | Spline curve | | | | | | | | Rotating ellipse | | | |
| Number of knots [$lk$] | | 10 | | | | 5 | | | | / | | | |
| Suspect Range | $x$-coordinate of knot [$Xs$] | 80-380 | | | | | | | | / | | | |
| | $y$-coordinate of knot [$Ys$] | 400-700 | | | | | | | | / | | | |
| | $x$-coordinate of center point of ellipse [$Xsc$] | / | | | | | | | | 160-200 | | | |
| | $y$-coordinate of center point of ellipse [$Ysc$] | / | | | | | | | | 480-580 | | | |
| | Semi-major axis of ellipse [$Ra$] | / | | | | | | | | 110-210 | | | |
| | Semi-minor axis of ellipse [$Rb$] | / | | | | | | | | 50-120 | | | |
| | Clockwise rotation angle [$B$] | / | | | | | | | | 0-90 | | | |
| | Initial release time [$T^0$] | 0-3175.5 | | | | | | | | | | | |
| | Release duration time [$\Delta T$] | 1204.5-6679.5 | | | | | | | | | | | |
| | Mass-loading rate [$M$] | 90-150 | | | | | | | | | | | |

## 4. Results

Figure 4 shows the probability of the source location in all three scenarios before any updates and after 0th, 1st, 4th, and 7th assimilation iterations. The statistic $Pr_i$ is employed to quantify the probability of the source occurring in any given unit $i$ within the aquifer, calculated as the ensemble mean of an indicator $Ir_{j,i}$.

$$Pr_i = \frac{1}{N}\sum_{j=1}^{N} Ir_{j,i} \tag{14}$$

where $N$ denotes the number of realizations included in the ensemble; the indicator $Ir_{j,i}$ denotes the presence of the source in cell $i$ for the $j^{th}$ realization, where a value of 1 indicates the source is present in cell $i$, and 0 indicates its absence. The analysis of Figure 4 reveals a significant level of uncertainty in the initial source distribution for all scenarios, which progressively diminishes with each assimilation iteration, reaching a minimum in the final step. Ultimately, S1 and S2 show clear irregular spatial structures of the potential source, while S3 does not. These findings demonstrate that the spline curve is more effective in approximating the spatial distributions of irregular non-point sources than the rotating ellipse, though its accuracy may be limited by the number of knot points.

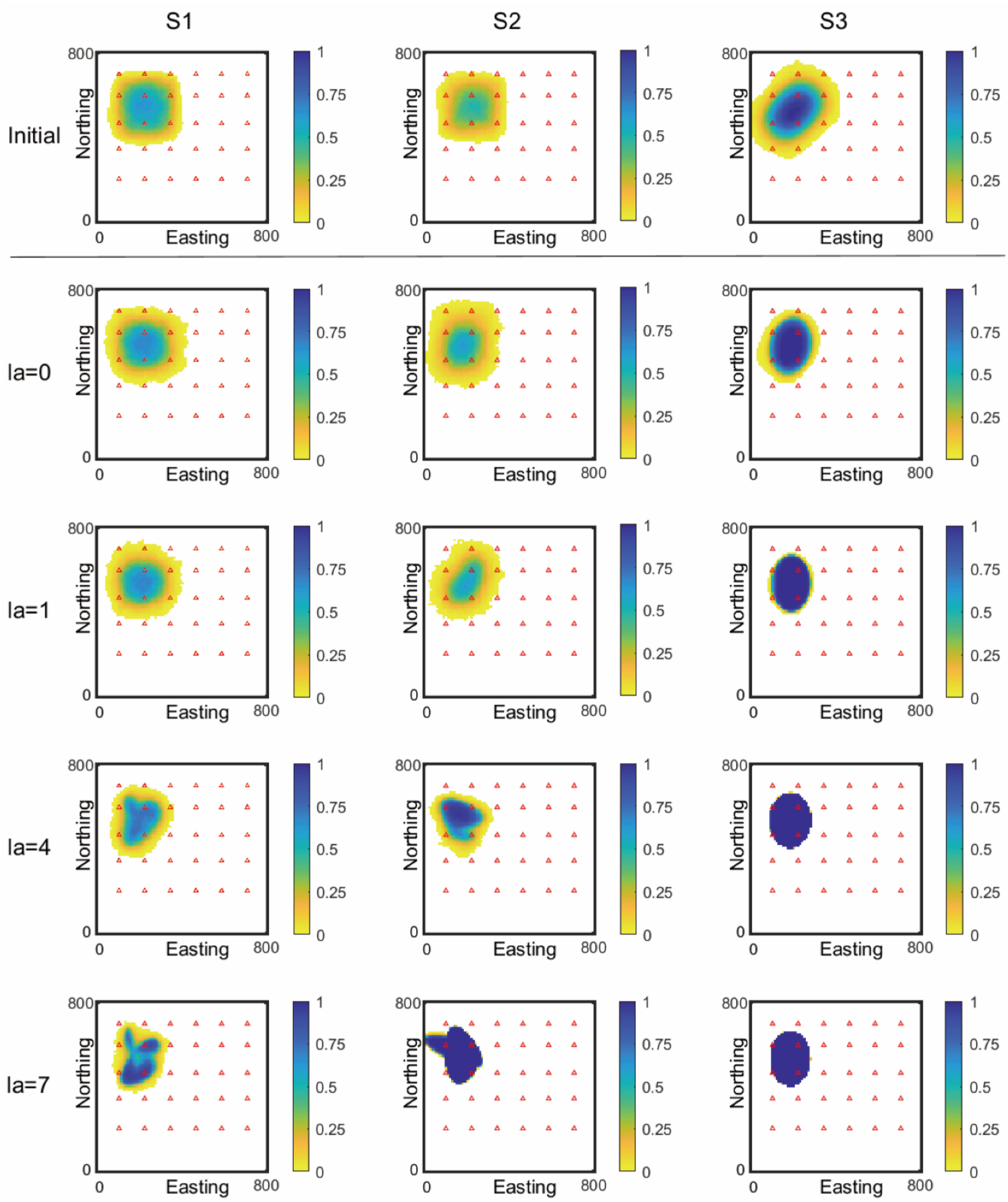

**Figure 4**: Probability of source location as computed from the source positional parameters updated at the 0th, 1st, 4th, and 7th assimilation iterations in scenarios S1-S3. The left column corresponds to ES-MDA-Bs with 10 knot points, the center column corresponds to ES-MDA-Bs with 5 knot points, the right column corresponds to ES-MDA-RE.

Figures 5 presents a statistic assessing the accuracy of location and shape updates of the potential source through assimilating concentration observations at the 0th, 1st, 4th, and 7th iterations across all three scenarios. This statistical measure, the Total Ensemble Absolute Bias ($TEAB$), quantifies accuracy across the domain by integrating the absolute deviations of the estimated source indicator $Ir_{j,i}$, from the reference indicator $Ir_{ref,i}$, for each spatial unit

$i$. Specifically, $TEAB$ for any unit $i$, $TEAB_i$, is computed as:

$$TEAB_i = \sum_{j=1}^{N} |Ir_{j,i} - Ir_{ref,i}| \tag{15}$$

where $Ir_{ref,i}$ denotes the presence of reference source in cell $i$. Note that an integral value of 0 indicates a fully accurate reproduction of source location and spatial structure, while a value of 500 means that the source does not appear in any realization. The analysis of Figure 5 shows a gradual convergence of the potential source structure toward the reference spatial distribution with increasing assimilation iterations in S1. After the 7th iteration, minor deviations appear at high-curvature edges in the updated source locations. These deviations cause the updated positions to shift inward, leading to a slight underestimation of the spatial distribution extent of the source. In contrast, S2 shows minimal reduction in deviation between estimated and true indicator values near the source boundary over successive iterations, with a noticable northwest misfit emerging on areas of marked curvature, resulting in an inaccurate restructuring of the source spatial distribution. These findings suggest that when using ES-MDA-Bs with limited-configured knot points for source localization, significant misjudgments occur in high-curvature areas, despite its capacity to estimate irregular spatial structures. Incorporating additional knot points enables ES-MDA-Bs to perform optimally, though it tends to slightly underestimate the spatial extent. Unlike S1 and S2, S3 captures source locations in elliptical areas, with restricted resolution of the irregularities.

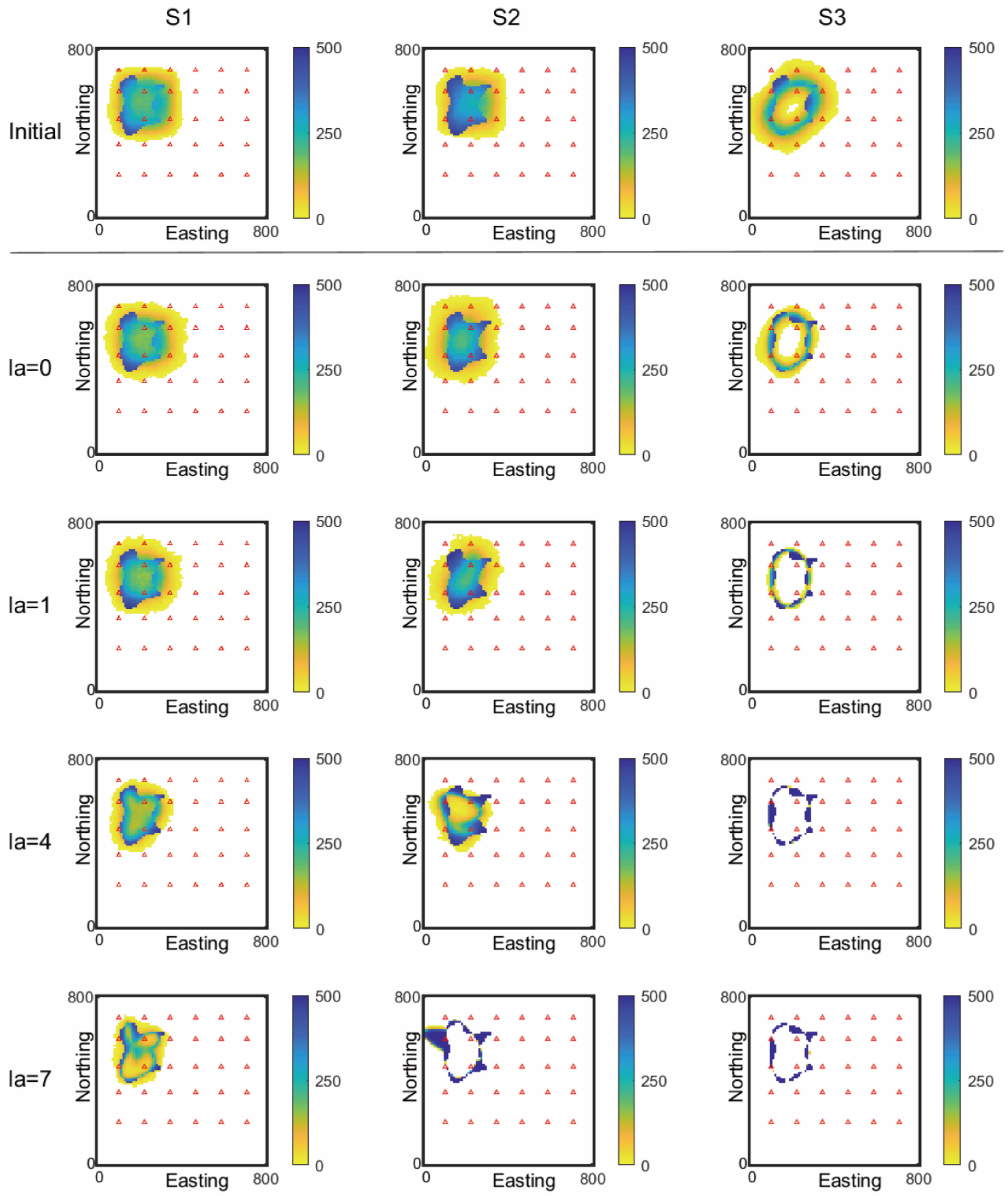

**Figure 5**: $TEAB$ computed with the initial and updated ensembles of positional parameters after the 0th, 1st, 4th, and 7th data assimilation iterations in scenarios S1-S3.

Figure 6 presents the ensemble variance of source indicator values for all three scenarios, providing a measure of variability in the indicator within the ensemble at each potential source location $i$, that can be written as:

$$EnsVar_i = \sigma_{Ir_i}^2 \qquad (16)$$

where $\sigma_{Ir_i}$ denotes the ensemble standard deviation of source indicator at location $i$. Figure 6 shows a significant reduction in initial spatial variability in all three scenarios as the number of iterations increases, reaching a minimum after the 7th iteration. In the final step, S1 displays the largest variance along the edges of the source, which is reasonable. Conversely, S2 and S3 show a minimal spatial variability but lacked accuracy, suggesting a severe underestimation of uncertainty. These findings demonstrate that, when using ES-MDA-Bs to estimate the spatial structure of an irregular source, increasing the number of identified knot points can mitigate uncertainty underestimation and improve spatial accuracy.

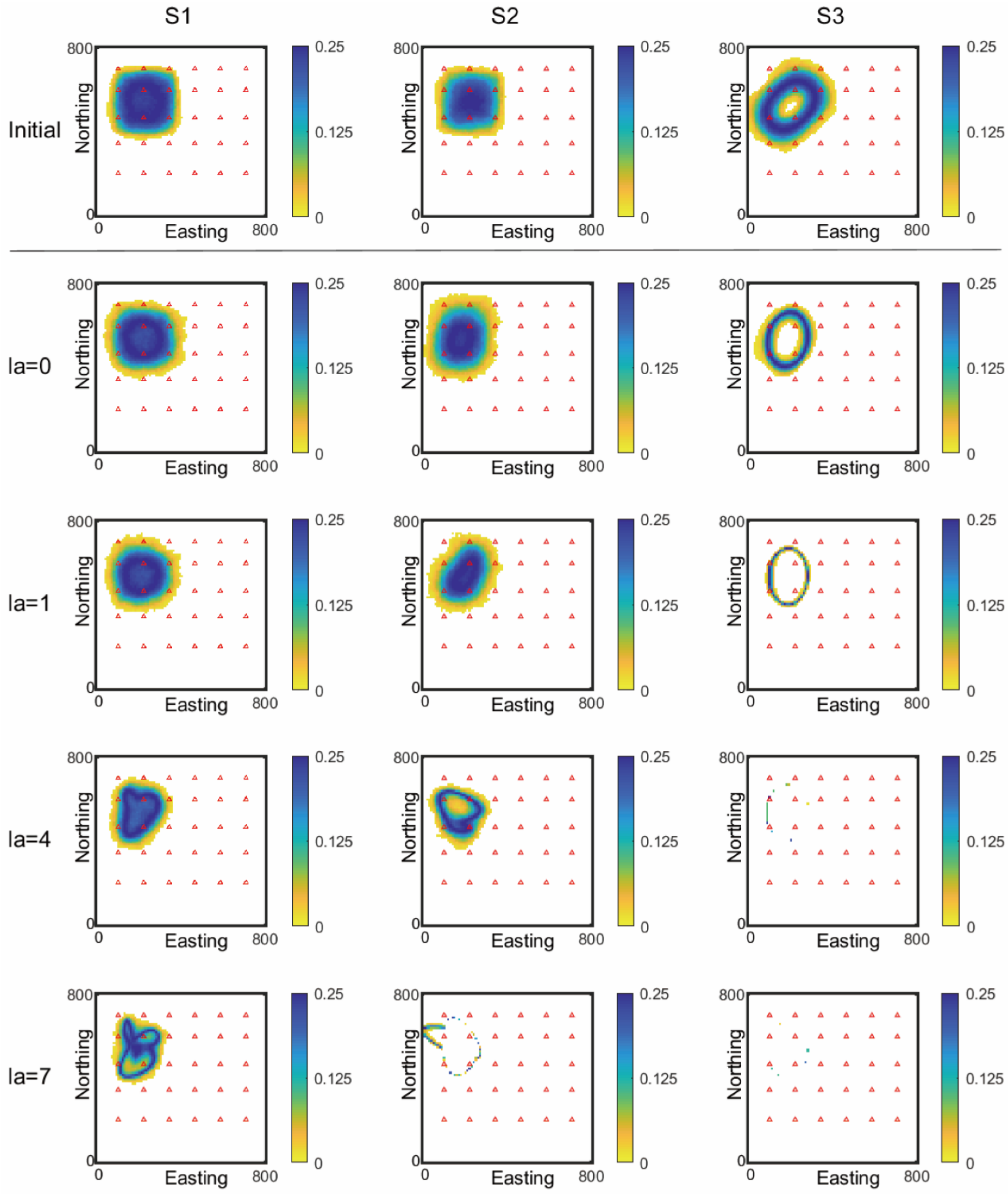

**Figure 6**: Ensemble variance computed with the initial and updated ensembles of positional parameters after the 0th, 1st, 4th, and 7th data assimilation iterations in scenarios S1-S3.

To further clarify the influence of the number of knot points on the reconstruction accuracy and ensemble variability in identifying source location and spatial structure, we evaluated the sensitivity of domain-integrated $TEAB$ and ensemble variance to the number of spline knots. Figure 7 displays the domain-integrated $TEAB$ and ensemble variance computed using ES-MDA-Bs with 5, 8, 10, and 12 knots after the 7th iteration. As the number of knots increases, uncertainty initially rises from a minimum level, while reconstruction accuracy improves and stabilizes around 10 knots but deteriorates sharply beyond this point. This trend is attributed to excessive spline flexibility, which increases sensitivity to local oscillations caused by ensemble uncertainty in knot placement during data assimilation, potentially inducing spatial distortion. Although the spatial extents of source structures reconstructed with 8 and 10 knots are similar, ensemble variance grows with knot count, reflecting an enhanced capacity of the algorithm to explore a wider solution space. These results indicate that insufficient knot configurations restrict the ability of ES-MDA-Bs to capture structural variability, leading to underestimation of uncertainty. However, increasing knot density to mitigate this limitation introduces a critical trade-off between uncertainty representation and reconstruction fidelity. Therefore, in this study, an ES-MDA-Bs configuration with 10 knots achieves reliable performance in estimating both the location and spatial structure of irregular sources. To develop general guidelines for optimal knot configuration, future work will investigate adaptive strategies for determining both knot number and placement.

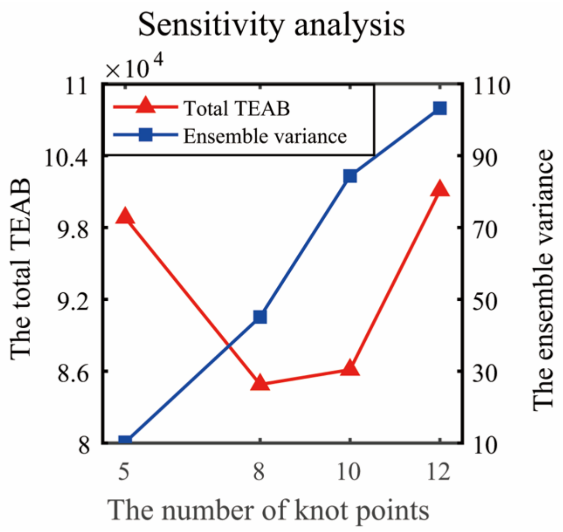

**Figure 7**: Sensitivity analysis of the integrated $TEAB$ and ensemble variance with respect to the number of spline knots, based on the updated ensembles of positional parameters obtained after the 7th data assimilation iteration using ES-MDA-Bs.

As mentioned above, when limited-configured knot points are used for approximation, a clear misfit with underestimated uncertainty appears due to filter inbreeding. This issue stems from ES-MDA-Bs, as an ensemble-based data assimilation algorithm, suffering from spurious correlations in ensemble covariance when the ensemble size is smaller than the number of assimilated observations. To address it, we employed localization technique to the ensemble-based algorithm, with detailed explanations provided in our previous work (Zhang et al., 2024b). Here, we evaluate the capability of ES-MDA-Bs combined with localization (LES-MDA-Bs) in S2 to improve the accuracy of spatial structure identification. Figure 8 presents three statistics, probability, $TEAB$ and ensemble variance of the location and shape of the potential source in S2 using LES-MDA-Bs before and after 0th, 1st, 4th, and 7th assimilation iterations. As the assimilation iterations increase, the potential source gradually aligns with the reference spatial extent, with obvious deviations and uncertainty underestimation removed. This result indicates that while LES-MDA-Bs with limited-configured knot points can capture the general extent of an irregular source, though it struggles to reconstruct the detailed spatial structure.

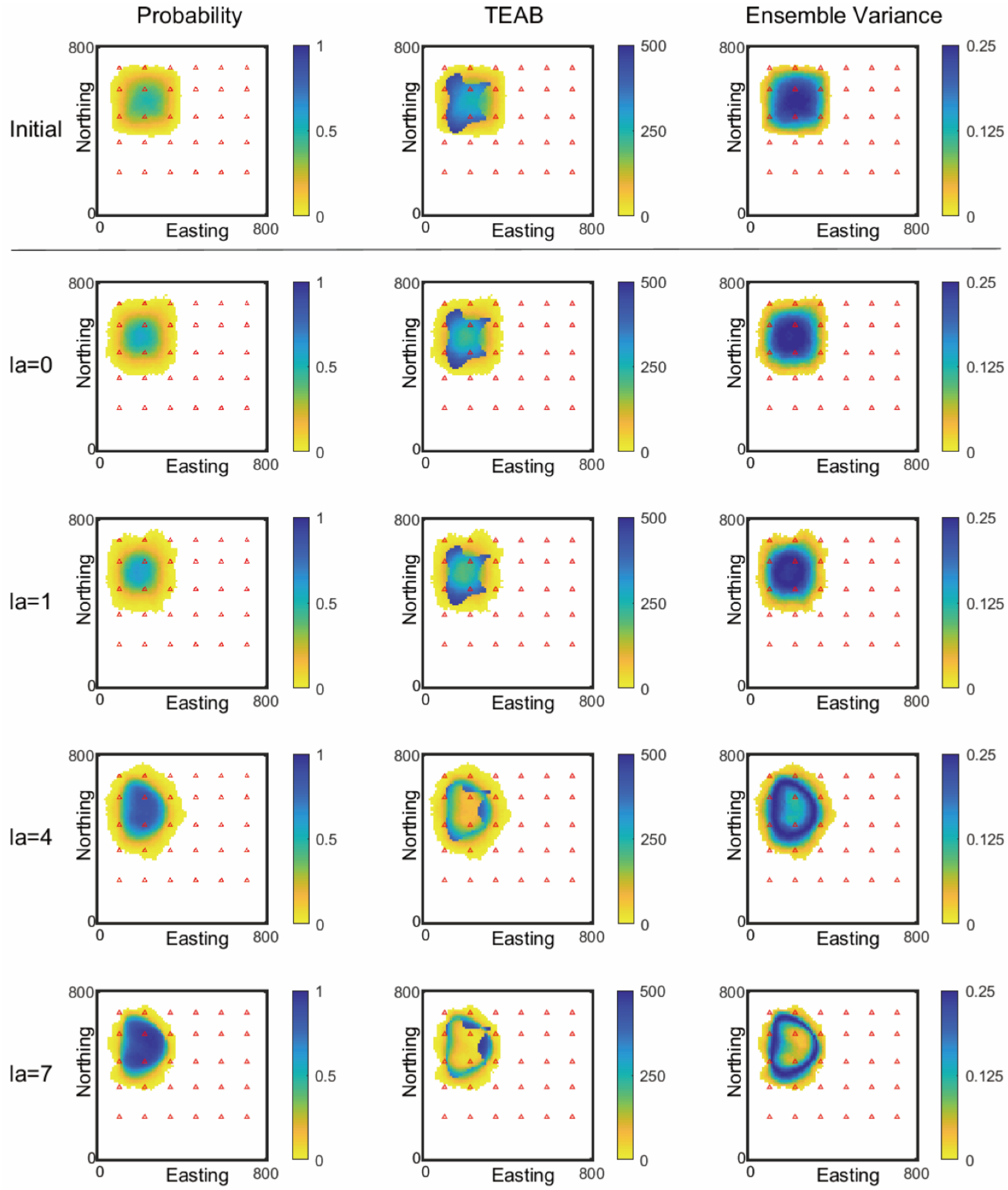

**Figure 8**: Probability, $TEAB$ and ensemble variance computed with the initial and updated ensembles of positional parameters after the 0th, 1st, 4th, and 7th data assimilation iterations in scenario S2 using LES-MDA-Bs.

Figures 9 and 10 show three statistical measures, the average absolute bias ($AAB$), the ensemble spread ($ESp$), and boxplot, of the release information parameters ($T^0$, $\Delta T$, $M$), in all three scenarios before and after 0th, 1st, 4th, and 7th updating iterations. Specifically, the $AAB$, which quantifies the accuracy of the release parameters, can be written as:

$$AAB = \frac{1}{N}\sum_{j=1}^{N}|Pr_j - Pr_{ref}| \qquad (17)$$

where $Pr_j$ denotes the source release information value for the $j^{th}$ realization; $Pr_{ref}$ is the corresponding reference value. And the $ESp$, measuring the variability within the ensemble for release parameters, can be written as:

$$ESp = \sqrt{\sigma_{Pr}^2} \tag{18}$$

where $\sigma_{Pr}^2$ denotes the ensemble variance of the release information value. When the ratio of $ESp/AAB$ equals to 1, indicates a satisfactory performance of the smoother (Xu et al., 2013, 2022). Figure 9 shows a progressive decline in $AAB$, $ESp$ and the ratio of $ESp/AAB$ with each assimilation, reaching their minimum values after the 7th iteration. Specifically, for scenario S1, the *AAB* of the mass-loading rate exhibits a slight increase after the 4th iteration but subsequently decreases to a lower level by the final update. This temporary fluctuation may result from localized ensemble perturbations introduced during the assimilation process, which can occasionally lead to variations in intermediate estimates. However, the overall trend remains stable, as indicated by the *ESp/AAB* ratio of the mass-loading rate, which remains close to 1 throughout the iterations. This suggests that the algorithm consistently performs well in estimating the mass-loading rate, despite minor fluctuations during individual updates. Additionally, it can be observed that the *ESp/AAB* ratio in S1 stabilizes closer to 1 after the final update compared to that in S3, indicating that ES-MDA-Bs demonstrates greater stability in identifying source information. However, a slight bias is observed in the final release mass-loading rate of S1 compared to the reference value. To further clarify this bias, the boxplots have been shown in Figure 10. This figure illustrates how ensemble medians of the release parameters progressively align with reference values, with uncertainty decreasing to a minimum in the final step. However, due to the underestimation of source spatial extent as mentioned, the final mass-loading rate in S1 remains overestimated, consistent with the mass conservation principle.

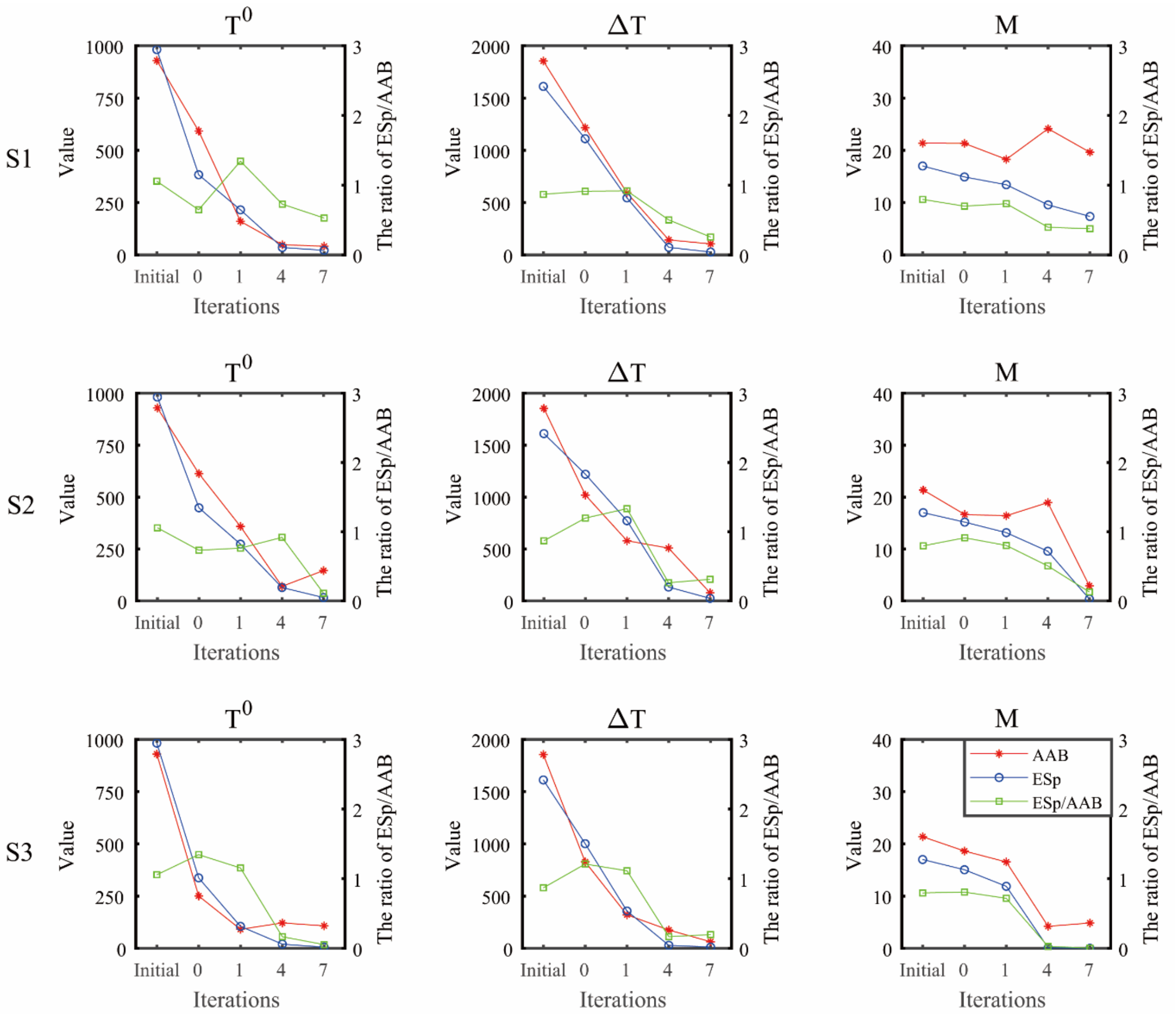

**Figure 9**: $AAB$, $ESp$ and the ratio of $ESp/AAB$ computed with the initial and updated ensembles of release information parameters of contaminant source, including $T^0$, $\Delta T$ and $M$, after the 0th, 1st, 4th, and 7th data assimilation iterations in scenarios S1-S3. The red line corresponds to $AAB$, the blue line corresponds to $ESp$ and the green line corresponds to $ESp/AAB$.

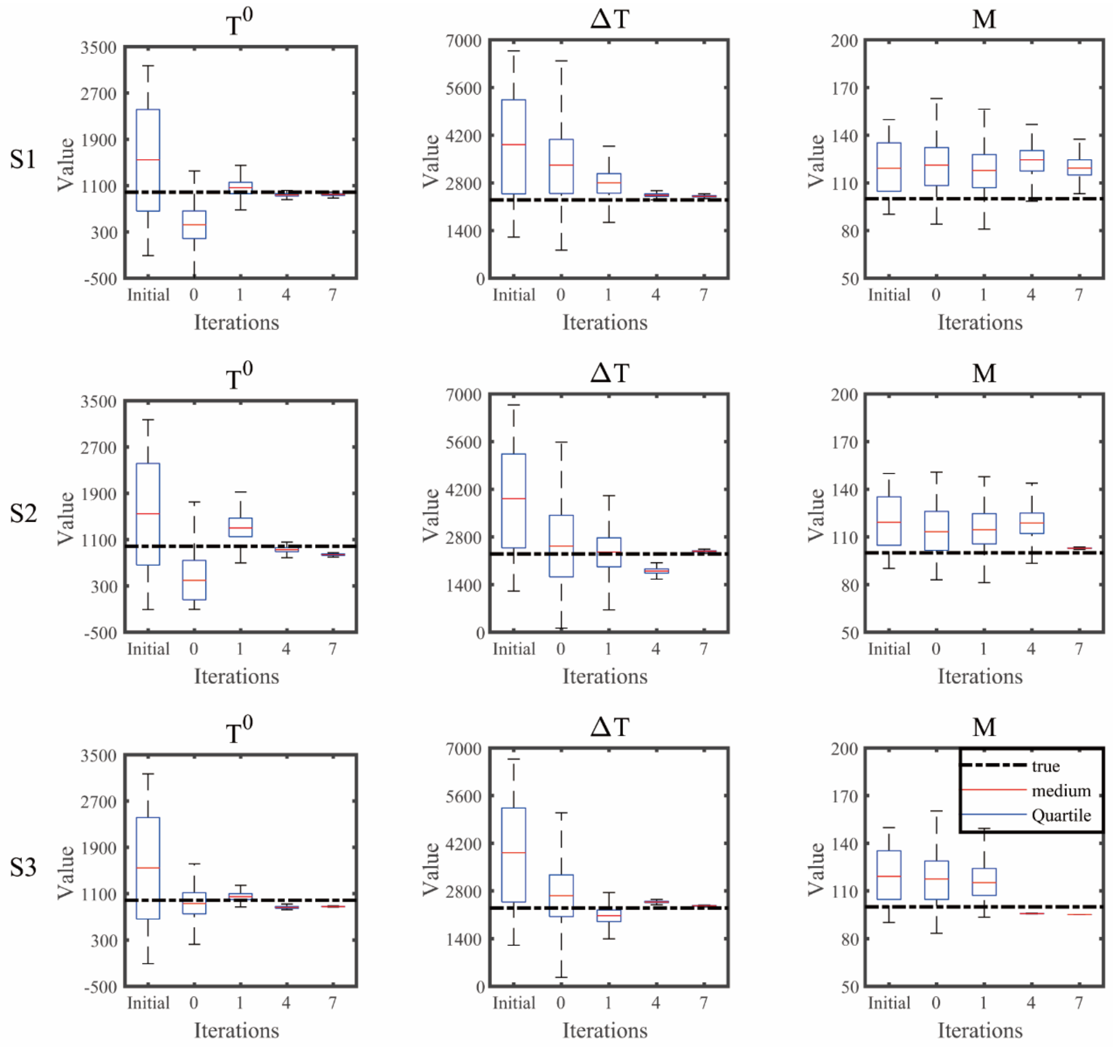

**Figure 10**: Boxplots computed with the initial and updated ensembles of release information parameters of contaminant source, including $T^0$, $\Delta T$ and $M$, after the 0th, 1st, 4th, and 7th data assimilation iterations in scenarios S1-S3. The dashed horizontal black line corresponds to the reference value.

To verify the accuracy of irregular source identification, two validation wells (#1, #2) are selected to predict the concentration evolution over time for the three scenarios, as shown in Figures 11 and 12. The initial concentrations exhibits significant uncertainties within the 5% to 95%, which decrease as assimilation iterations are performed. In S2 and S3, the confidence intervals nearly overlap but deviate from the true evolution. Particularly in well #2, the significant deviations indicate an underestimation of concentration uncertainty. In contrast, the reference evolutions for S1 consistently fall within the confidence intervals, demonstrating superior performance of ES-MDA-Bs with adequately equipped knot points in predicting concentration.

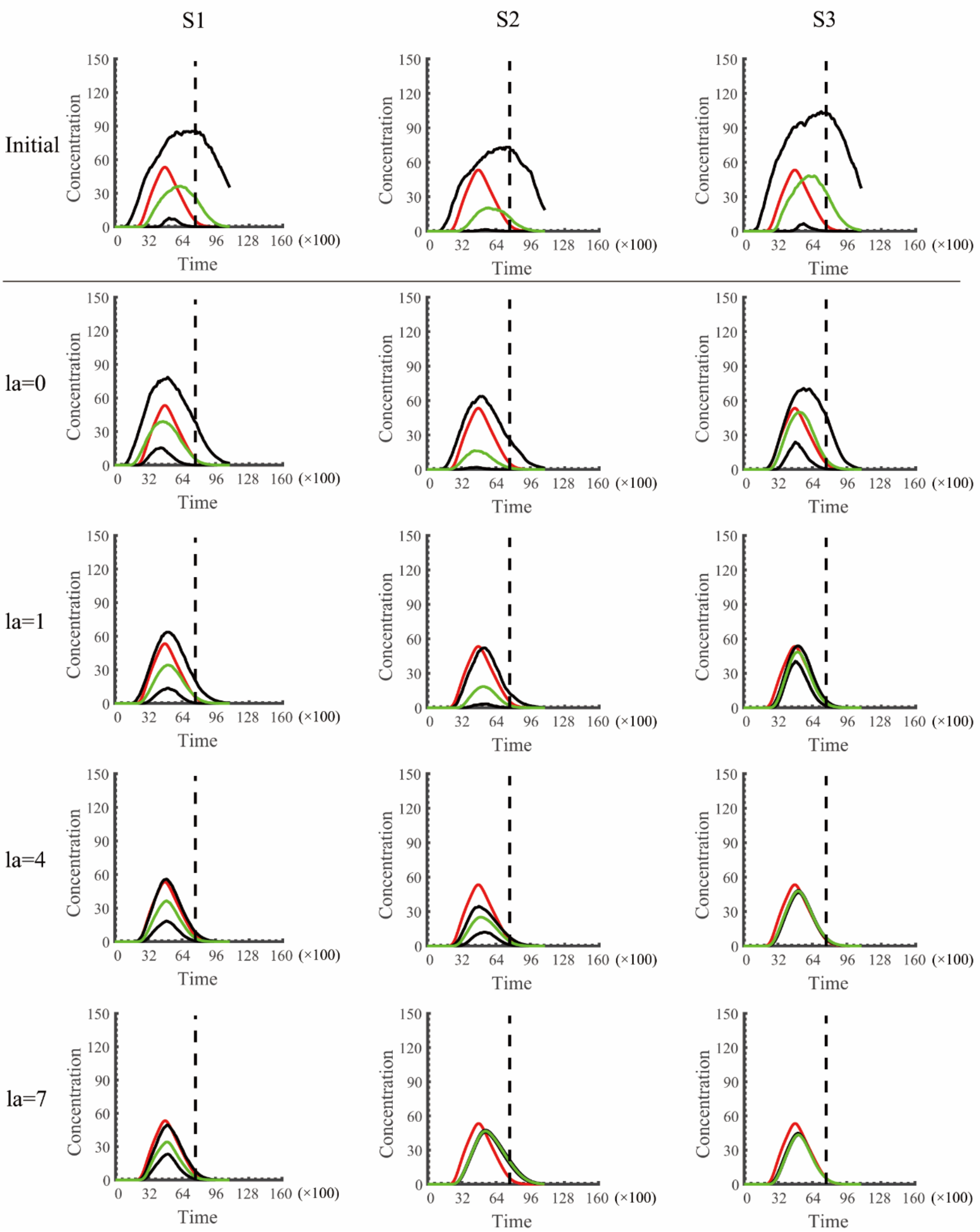

**Figure 11**: Time evolution of solute concentrations at the verification well #1 computed with the initial and updated ensembles of source parameters after the 0th, 1st, 4th, and 7th data assimilation in scenarios S1-S3. The left column corresponds to S1, the center column corresponds to S2 and the right column corresponds to S3. The red line corresponds to the reference field. The black lines correspond to the 5 and 95 percentiles of all realizations, and the green line corresponds to the median. The vertical dashed lines mark the end of the assimilation period.

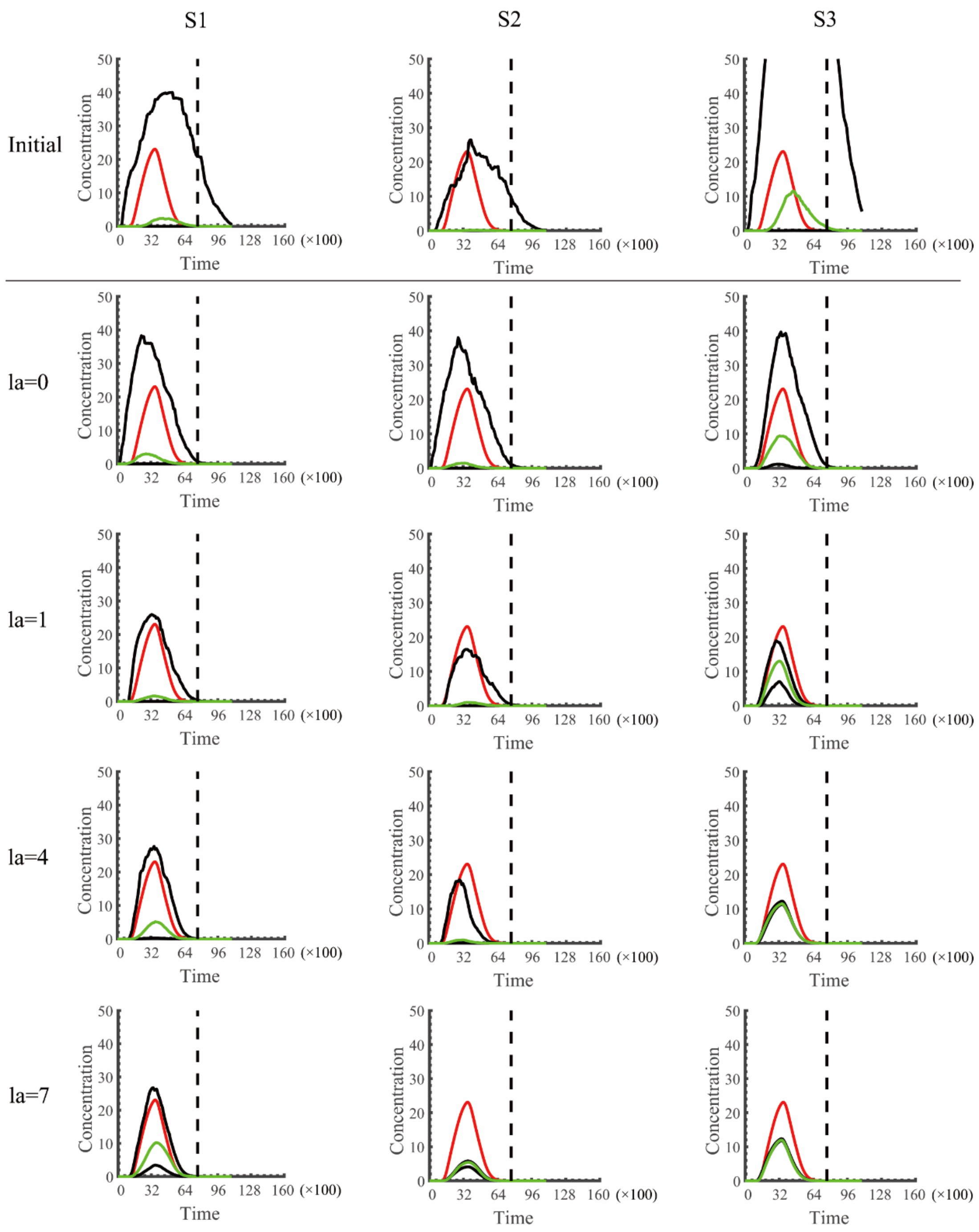

**Figure 12**: Time evolution of solute concentrations at the verification well #2 computed with the initial and updated ensembles of source parameters after the $0^{th}$, $1^{st}$, $4^{th}$, and $7^{th}$ data assimilation in scenarios S1-S3. The left column corresponds to S1, the center column corresponds to S2 and the right column corresponds to S3. The red line corresponds to the reference field. The black lines correspond to the 5 and 95 percentiles of all realizations, and the green line corresponds to the median. The vertical dashed lines mark the end of the assimilation period.

Figure 13 presents a comparative analysis between the reference and the ensemble mean of the contaminant plume for the three scenarios at the $10^{th}$, $30^{th}$, and $50^{th}$ time steps. The concentration plume is predicted using the updated source parameters from the 7th iteration. The results show that S2 significantly over-predicts the concentration, while the S1 and S3 fit well with the reference. However, S3 fails to capture the source irregularities derived from the underlying source parameters, as previously discussed. This discrepancy may be attributed to the fact that the predicted plume shape is influenced by multiple factors, such as uncertainties in source location, spatial structure, and release history, as well as the accuracy of these input parameters. These findings further confirm the accuracy of ES-MDA-Bs with appropriately configured knot points to reproduce the contaminant plume.

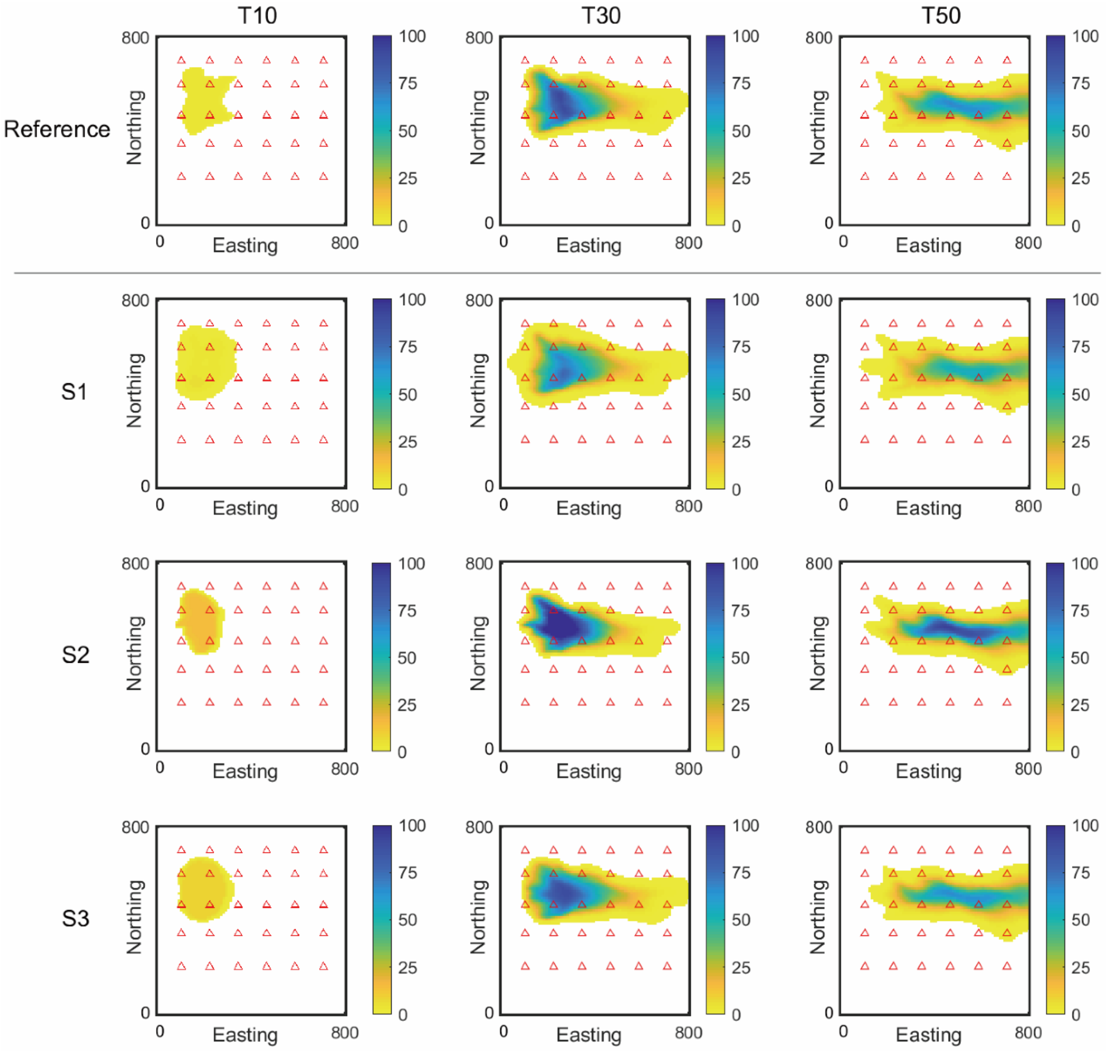

**Figure 13**: Contaminant plume at the 10th, 30th, and 50th simulation time steps, computed with the reference and the updated parameters after the 7th assimilation iteration in scenarios S1-S3. From top to bottom: plume in reference field; ensemble mean of plumes calculated by updates in S1; ensemble mean of plumes calculated by updates in S2; ensemble mean of

plumes calculated by updates in S3.

To further evaluate the robustness and generalizability of the ES-MDA-Bs framework configured with 10 knot points for identifying irregular non-point sources, a new synthetic confined aquifer is designed, characterized by distinct hydrogeological properties and contaminant source configurations compared to scenarios S1-S3 presented in Section 3. Figure 14(a) illustrates the natural logarithm of a heterogeneous hydraulic conductivity field, generated as a multi-Gaussian random field using the parameters specified in Table 1. In the groundwater flow simulation, constant head boundaries are assigned to the eastern and western edges of the aquifer, set at 90 [L] and 280 [L], respectively. For solute transport, the longitudinal dispersivity is held constant in both space and time, while the ratio of horizontal transverse to longitudinal dispersivity is reduced from 0.5 to 0.4, reflecting a potential reduction in lateral dispersion due to altered velocity gradients induced by the modified hydraulic conductivity distribution. As noted, a uniquely shaped non-point source, shown in Figure 14(b), is introduced in this setup. The plausible range for each basic source parameter is detailed in Table 3.

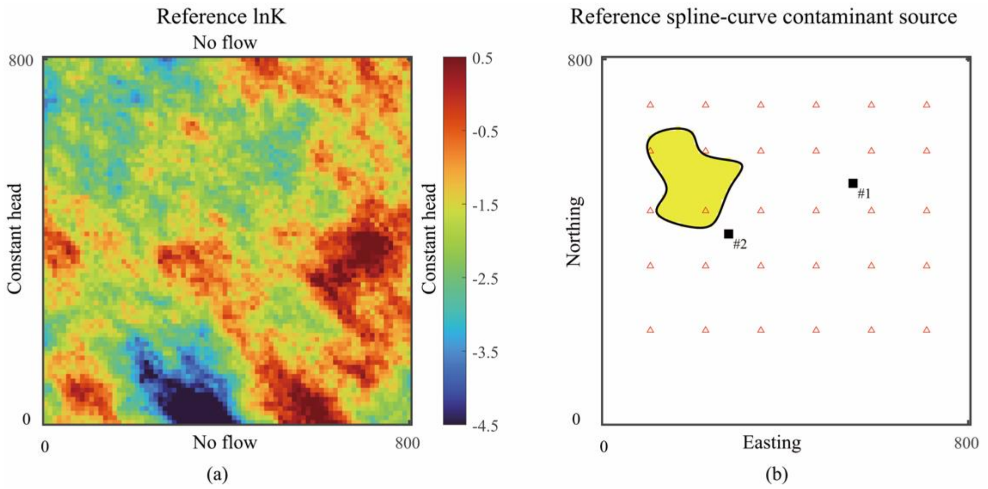

**Figure 14**: Reference log-conductivity field (a) and reference contamination source (b) in the new synthetic aquifer. An indication of the type of boundary conditions used for the solution of the flow equation is also shown in (a). The red triangles are the observation locations and the black squares are the validation locations shown in (b).

Figure 15 presents three statistical metrics—probability, $TEAB$, and ensemble variance of the potential source—computed using the updated ensembles of positional parameters following the 7th data assimilation iteration. A distinctly irregular spatial configuration is evident in the updated spatial extent, characterized by pronounced ensemble variability along the spline boundaries and localized positional discrepancies predominantly situated along high-curvature regions in the southern boundary—both features are consistent with expected behavior. These results further confirm that, as previously observed, ES-MDA-Bs with 10 knot points effectively effectively mitigates uncertainty underestimation while maintaining a high degree of accuracy in spatial reconstruction, exhibiting only minor deviations in proximity to the source margins.

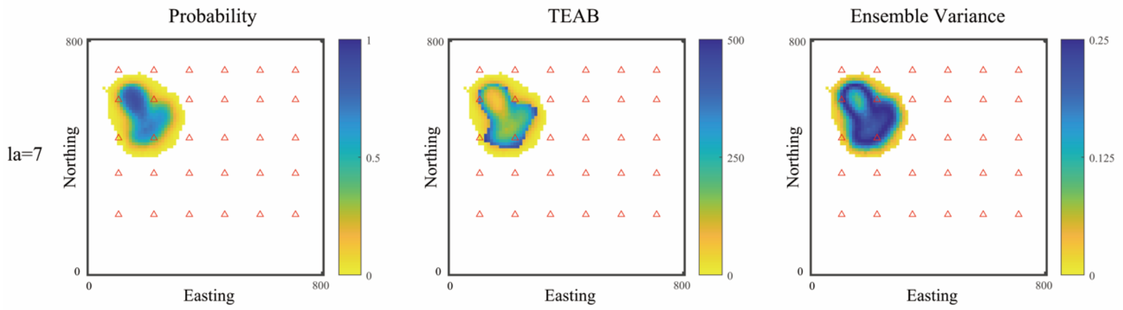

**Figure 15**: Probability, TEAB, and ensemble variance of the potential source, calculated using the updated ensembles of positional parameters following the 7th data assimilation iteration in the new synthetic aquifer.

Figure 16 illustrates a comparative analysis between the reference contaminant plume and the ensemble mean plume predicted using the updated source parameters from the 7th iteration, conducted in the new synthetic aquifer at the 10th, 30th, and 50th time steps. The ensemble prediction exhibits strong agreement with the reference plume in terms of temporal evolution, spatial extent, and concentration levels. These findings further validate the robustness of the ES-MDA-Bs method, configured with 10 knot points, in accurately identifying randomly shaped, irregular non-point sources—including their location, structural characteristics, and release history—and in reliably predicting contaminant plume distribution across diverse synthetic scenarios.

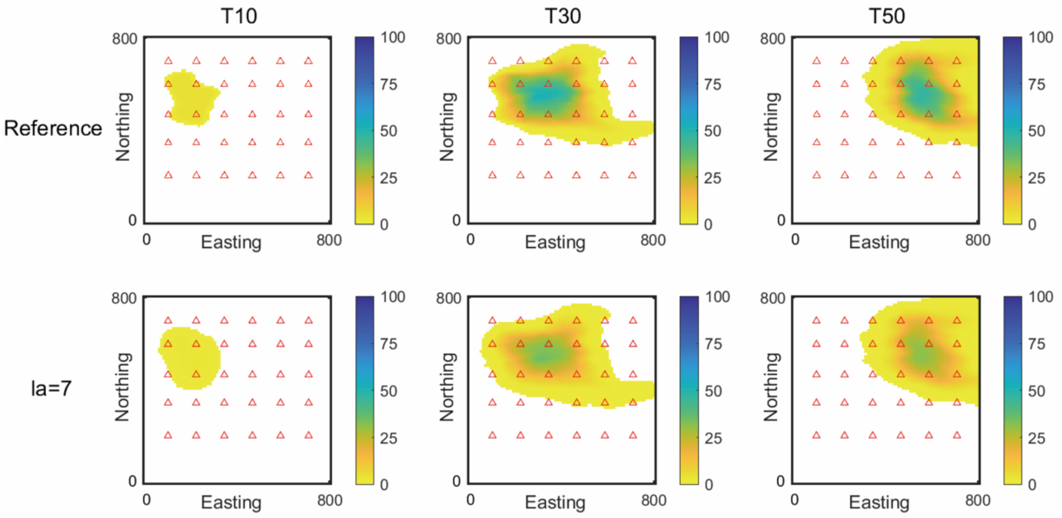

**Figure 16**: Contaminant plume at the 10th, 30th, and 50th simulation time steps, computed with the reference and the updated parameters after the 7th assimilation iteration in the new synthetic aquifer. From top to bottom: plume in reference field; ensemble mean of plumes calculated by updates in the new synthetic aquifer.

However, transitioning the method from synthetic cases to real-world applications represents a significant challenge. In the next phase, we will adhere to the established research trajectory commonly adopted in groundwater point source identification studies (e.g., Xu and Gómez-Hernández, 2016; Chen et al., 2023b), progressing systematically from theoretical validation to sandbox experiments, and ultimately to large-scale field implementations.

## 5. Discussion

The aforementioned results demonstrate that the proposed ES-MDA-Bs effectively reproduces the complex spatial structure and the release information of a non-point source. However, this study represents an initial step in exploring irregular non-point sources. To support real-world applications, several key aspects still need further consideration:

(1) Performance evaluation and algorithm enhancement for heterogeneous release: In this study, ES-MDA-Bs is applied to identify an irregular non-point source with homogeneous releases. However, in real-world settings, source releases typically exhibit spatial heterogeneity and temporal fluctuations. Therefore, our future research will focus on evaluating and enhancing the ES-MDA-Bs algorithm, such as by integrating localization

techniques, to accurately identify the locations, spatial distributions, and varying release characteristics of complex irregular non-point sources.

(2) Computational efficiency improvement: Data assimilation algorithms face computational challenges due to the time-consuming nature of multiple runs on forward model. Implementing a suitable surrogate model can alleviate this burden while maintaining accuracy in reproducing state variable fields. In future work, we will develop a new algorithm that integrates ES-MDA-Bs with an enhanced surrogate model to efficiently and accurately reconstruct irregular non-point sources.

(3) Impact of observation noise level: This study adopts a fixed observation noise level to isolate and evaluate the performance of ES-MDA-Bs in identifying irregular non-point sources. However, given the intrinsic characteristics of ES-MDA algorithms, an underestimated noise level may result in excessively large Kalman gains, leading to overfitting of spurious residuals and potential ensemble collapse during sequential updates. Therefore, future work will investigate the robustness of ES-MDA-Bs under varying noise levels and develop adaptive strategies to mitigate errors induced by noise misspecification.

(4) Optimization of observation well placement: Practical constraints—such as geological heterogeneity and economic limitations—often restrict the number and spatial configuration of observation wells in field applications. Future research could focus on overcoming these challenges by developing a multi-objective optimization framework for well placement, integrated with inverse modeling techniques (e.g., ES-MDA-Bs), to efficiently address complex non-point source identification problems under cost constraints.

## 6. Summary

In this paper, we evaluate the performance of ES-MDA-Bs in identifying a complex irregular non-point source. Our findings demonstrate that ES-MDA-Bs effectively determines the location, spatial structure and release information of the source, offering improved accuracy over ES-MDA-RE. With the updated source parameters, the contaminant concentration plume is successfully reproduced through forward prediction.

We further demonstrate that well-configured knot points significantly enhance the accuracy of source reconstruction and effectively mitigates uncertainty underestimation. Conversely, limited-configured knot points induce filter inbreeding, which can be mitigated by employing the localization technique. Besides, we accurately estimate the initial release time and release duration, though a slight overestimation of release mass-loading is observed. Moreover, ES-MDA-Bs outperforms ES-MDA-RE in approximating the spatial distribution of irregular sources, which is restricted to identify ellipse-like sources.

Building upon the research conducted by Xu et al. (2022), this paper extends the study by proposing a new algorithm, ES-MDA-Bs, to accurately reproduce the location, spatial structure, and release information of a complex irregular non-point source. This research has made progress toward real-world applications. For future work, we intend to explore which surrogate model combined with ES-MDA-Bs can improve the computational efficiency of identifying complex irregular non-point sources. In addition, further testing and refining the algorithm's performance to accommodate changes in source release patterns will be a priority.

## Acknowledgments

Financial support to carry out this work was received from the National Natural Science Foundation of China (grant no. 42377046). J. J. Gómez-Hernández acknowledges the financial support funded by MCIN/AEI/10.13039/50110001103 through grant PID2019-109131RB-I00. C. Lu acknowledges the National Natural Science Foundation of China (grant no. 51879088).

## References

An, Y., Y. Zhang, and X. Yan (2022), An integrated bayesian and machine learning approach application to identification of groundwater contamination source parameters, Water, 14 (15), 2447. doi: 10.3390/w14152447.

Anshuman, A., and T. Eldho (2022), Entity aware sequence to sequence learning using lstms for estimation of groundwater contamination release history and transport parameters, Journal of Hydrology, 608, 127,662. doi: 10.1016/j.jhydrol.2022.127662.

Anshuman, A., and T. Eldho (2023), A parallel work ow framework using encoder-decoder lstms for uncertainty quantification in contaminant source identification in groundwater, Journal of Hydrology, 619, 129,296. doi: 10.1016/j.jhydrol.2023.129296.

Ayvaz, M. T. (2016), A hybrid simulation-optimization approach for solving the areal groundwater pollution source identification problems, Journal of Hydrology, 538, 161-176. doi: 10.1016/j.jhydrol.2016.04.008.

Bai, T., and P. Tahmasebi (2022), Characterization of groundwater contamination: A transformer-based deep learning model, Advances in Water Resources, 164, 104,217. doi: 10.1016/j.advwatres.2022.104217.

Butera, I., and M. G. Tanda (2003), A geostatistical approach to recover the release history of groundwater pollutants, Water Resources Research, 39 (12). doi: 10.1029/2003WR002314.

Chang, Z., W. Lu, and Z. Wang (2022), Study on source identification and source-sink relationship of lnapls pollution in groundwater by the adaptive cyclic improved iterative process and monte carlo stochastic simulation, Journal of Hydrology, 612, 128,109. doi: 10.1016/j.jhydrol.2022.128109.

Chang, Z., Guo, Z., Chen, K., et al. (2024). A comparison of inversion methods for surrogate-based groundwater contamination source identification with varying degrees of model complexity. Water Resources Research, 60(4). doi: 10.1029/2023WR036051.

Chen, Z., J. J. Gómez-Hernández, T. Xu, and A. Zanini (2018), Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble kalman filter, Journal of hydrology, 564, 1074-1084. doi: 10.1016/j.jhydrol.2018.07.073.

Chen, Z., L. Zong, J. J. Gómez-Hernández, T. Xu, Y. Jiang, Q. Zhou, H. Yang, Z. Jia, and S. Mei (2023a), Contaminant source and aquifer characterization: An application of es-mda demonstrating the assimilation of geophysical data, Advances in Water Resources, p. 104555. doi: 10.1016/j.advwatres.2023.104555.

Chen, Z., T. Xu, J. J. Gómez-Hernández, A. Zanini, and Q. Zhou (2023b), Reconstructing the release history of a contaminant source with different precision via the ensemble smoother with multiple data assimilation, Journal of Contaminant Hydrology, 252, 104,115. doi: 10.1016/j.jconhyd.2022.104115.

de Boor, C. (1978), A Practical Guide to Splines, Springer-Verlag.

De Boor, C., and C. De Boor (1978), A practical guide to splines, vol. 27, springerverlag New York.

Emerick, A. A., and A. C. Reynolds (2013), Ensemble smoother with multiple data assimilation, Computers & Geosciences, 55, 3-15. doi: 10.1016/j.cageo.2012.03.011.

Evensen, G., and P. J. Van Leeuwen (2000), An ensemble kalman smoother for non-linear dynamics, Monthly Weather Review, 128 (6), 1852-1867. doi: 10.1175/1520-0493(2000)1282.0.CO;2.

Farin, G. (2014), Curves and surfaces for computer-aided geometric design: a practical guide, Elsevier.

Ge, Y., W. Lu, and Z. Pan (2023), Groundwater contamination source identification based on sobol sequences-based sparrow search algorithm with a bilstm surrogate model, Environmental Science and Pollution Research, 30 (18), 53,191-53,203. doi: 10.1007/s11356-023-25890-0.

Gómez-Hernández, J. J., and A. G. Journel (1993), Joint sequential simulation of multigaussian fields, in Geostatistics Tróia'92: Volume 1, pp. 85-94, Springer. doi: 10.1007/978-94-011-1739-5_8.

Gorelick, S. M., B. Evans, and I. Remson (1983), Identifying sources of groundwater pollution: An optimization approach, Water Resources Research, 19 (3), 779-790. doi: 10.1029/WR019i003p00779.

Gzyl, G., A. Zanini, R. Fraczek, and K. Kura (2014), Contaminant source and release history identification in groundwater: a multi-step approach, Journal of contaminant hydrology, 157, 59-72. doi: 10.1016/j.jconhyd.2013.11.006.

Hou, Z., W. Lao, Y. Wang, and W. Lu (2021), Homotopy-based hyper-heuristic searching approach for reciprocal feedback inversion of groundwater contamination source and aquifer parameters, Applied Soft Computing, 104, 107,191. doi: 10.1016/j.asoc.2021.107191.

Hwang, H.-T., S.-W. Jeen, D. Kaown, S.-S. Lee, E. A. Sudicky, D. T. Steinmoeller, and K.-K. Lee (2020), Backward probability model for identifying multiple contaminant source zones under transient variably saturated flow conditions, Water Resources Research, 56 (4), e2019WR025,400. doi: 10.1029/2019WR025400

Hwang, J. C., and R. M. Koerner (1983), Groundwater pollution source identification from limited monitoring well data: Part 1|theory and feasibility, Journal of hazardous materials, 8 (2), 105-119. doi: 10.1016/0304-3894(83)80050-8.

Jamshidi, A., J. M. V. Samani, H. M. V. Samani, A. Zanini, M. G. Tanda, and M. Mazaheri (2020), Solving inverse problems of unknown contaminant source in groundwater-river integrated systems using a surrogate transport model based optimization, Water, 12 (9), 2415. doi: 10.3390/w12092415.

Jha, M., and B. Datta (2015), Application of dedicated monitoring{network design for unknown pollutant-source identification based on dynamic time warping, Journal of Water Resources Planning and Management, 141 (11), 04015,022. doi: 10.1061/(ASCE)WR.1943-5452.0000513.

Jin, X., G. Mahinthakumar, E. M. Zechman, and R. S. Ranjithan (2009), A genetic algorithm-based procedure for 3d source identification at the borden emplacement site, Journal of Hydroinformatics, 11 (1), 51-64. doi: 10.2166/hydro.2009.002.

Juhász, I., and M. Hoffmann (2004), Constrained shape modification of cubic b-spline curves by means of knots, Computer-Aided Design, 36 (5), 437-445. doi: 10.1016/S0010-4485(03)00116-7.

Kontos, Y. N., T. Kassandros, K. Perifanos, M. Karampasis, K. L. Katsifarakis, and K. Karatzas (2022), Machine learning for groundwater pollution source identification and monitoring network optimization, Neural Computing and Applications, 34 (22), 19,515-19,545. doi: 10.1007/s00521-022-07507-8.

Li, J., Z. Wu, H. He, and W. Lu (2023a), Identifying groundwater contamination sources based on the hybrid grey wolf gradient algorithm and deep belief neural network, Stochastic Environmental Research and Risk Assessment, 37 (5), 1697-1715. doi:

10.1007/s00477-022-02360-6.

Li, L., H. Zhou, H.-J. Hendricks Franssen, and J. J. Gómez-Hernández (2012), Modeling transient groundwater flow by coupling ensemble kalman filtering and upscaling, Water Resources Research, 48 (1). doi: 10.1029/2010WR010214.

Li, Y., W. Lu, Z. Pan, Z. Wang, and G. Dong (2023b), Simultaneous identification of groundwater contaminant source and hydraulic parameters based on multilayer perceptron and flying foxes optimization, Environmental Science and Pollution Research, pp. 1-15. doi: 10.1007/s11356-023-27574-1.

Luo, C., W. Lu, Z. Pan, Y. Bai, and G. Dong (2023), Simultaneous identification of groundwater pollution source and important hydrogeological parameters considering the noise uncertainty of observational data, Environmental Science and Pollution Research, pp. 1-16. doi: 10.1007/s11356-023-28091-x.

Lyche, T., and K. Mørken (1999), The sensitivity of a spline function to perturbations of the knots, BIT Numerical Mathematics, 39, 305-322. doi: 10.1023/A:1022346030560.

Mahar, P. S., and B. Datta (1997), Optimal monitoring network and ground-water pollution source identification, Journal of water resources planning and management, 123 (4), 199-207. doi: 10.1061/(ASCE)0733-9496(1997)123:4(199).

Mahinthakumar, G., and M. Sayeed (2005), Hybrid genetic algorithm-local search methods for solving groundwater source identification inverse problems, Journal of water resources planning and management, 131 (1), 45-57. doi: 10.1061/(ASCE)0733-9496(2005)131:1(45).

Mirghani, B. Y., K. G. Mahinthakumar, M. E. Tryby, R. S. Ranjithan, and E. M. Zechman (2009), A parallel evolutionary strategy based simulation-optimization approach for solving groundwater source identification problems, Advances in Water Resources, 32 (9), 1373-1385. doi: 10.1016/j.advwatres.2009.06.001.

Mirghani, B. Y., E. M. Zechman, R. S. Ranjithan, and G. Mahinthakumar (2012), Enhanced simulation-optimization approach using surrogate modeling for solving inverse problems, Environmental Forensics, 13 (4), 348-363. doi: 10.1080/15275922.2012.702333.

Pan, Z., W. Lu, Z. Chang, et al. (2021), Simultaneous identification of groundwater pollution source spatial-temporal characteristics and hydraulic parameters based on deep regularization neural network-hybrid heuristic algorithm, Journal of Hydrology, 600, 126, 586. doi: 10.1016/J.JHYDROL.2021.126586.

Pan, Z., W. Lu, and Y. Bai (2022), Groundwater contamination source estimation based on a refined particle filter associated with a deep residual neural network surrogate, Hydrogeology Journal, 30 (3), 881-897. doi: 10.1007/s10040-022-02454-z.

Pan, Z., W. Lu, H. Wang, and Y. Bai (2023a), Groundwater contaminant source identification based on an ensemble learning search framework associated with an auto xgboost surrogate, Environmental Modelling & Software, 159, 105,588. doi: 10.1016/j.envsoft.2022.105588.

Pan, Z., W. Lu, and Y. Bai (2023b), Groundwater contaminated source estimation based on adaptive correction iterative ensemble smoother with an auto lightgbm surrogate, Journal of Hydrology, 620, 129,502. doi: 10.1016/j.jhydrol.2023.129502.

Pan, Z., Guo, Z., Chen, K., et al. (2025). A deep adaptive bidirectional generative adversarial neural network (Bi-GAN) for groundwater contamination source estimation. Journal of Hydrology, 132753. doi: 10.1016/j.jhydrol.2025.132753.

Park, H., and J.-H. Lee (2007), B-spline curve fitting based on adaptive curve refinement using dominant points, Computer-Aided Design, 39 (6), 439-451. doi: 10.1016/j.cad.2006.12.006.

Piegl, L., and W. Tiller (1996), The NURBS book, Springer Science & Business Media.

Schoenberg, I. J. (1946), Contributions to the problem of approximation of equidistant data by analytic functions. part b. on the problem of osculatory interpolation. a second class of analytic approximation formulae, Quarterly of Applied Mathematics, 4 (2), 112-141.

Singh, P., Mahor, V., Lakshmaiya, N., Shanker, K., Kaliappan, S., & Muthukannan, M., et al. (2024). Prediction of groundwater contamination in an open landfill area using a novel hybrid clustering-based ai model. Environment Protection Engineering, 50(1). doi: 10.37190/epe240106.

Srivastava, D., and R. M. Singh (2015), Groundwater system modeling for simultaneous identification of pollution sources and parameters with uncertainty characterization, Water resources management, 29, 4607-4627. doi: 10.1007/s11269-015-1078-8.

Wu, Y., Li, M., Xie, H., et al. (2025). Characterizing multi-source heavy metal contaminated sites at the Hun River basin: An integrated deep learning and data assimilation approach. Journal of Hydrology, 648, 132349. doi: 10.1016/j.jhydrol.2024.132349.

Xu, T., and J. J. Gómez-Hernández (2016), Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble kalman filtering, Water Resources Research, 52 (8), 6587-6595. doi: 10.1002/2016WR019111.

Xu, T., and J. J. Gómez-Hernández (2018), Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble kalman filter, Advances in Water Resources, 112, 106-123. doi: 10.1016/j.advwatres.2017.12.011.

Xu, T., J. J. Gómez-Hernández, H. Zhou, and L. Li (2013), The power of transient piezometric head data in inverse modeling: An application of the localized normal-score enkf with covariance in ation in a heterogenous bimodal hydraulic conductivity field, Advances in Water Resources, 54, 100-118. doi: 10.1016/j.advwatres.2013.01.006.

Xu, T., J. J. Gómez-Hernández, Z. Chen, and C. Lu (2021), 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, 595,125,681. doi: 10.1016/j.jhydrol.2020.125681.

Xu, T., W. Zhang, J. J. Gómez-Hernández, Y. Xie, J. Yang, Z. Chen, and C. Lu (2022), Non-point contaminant source identification in an aquifer using the ensemble smoother with multiple data assimilation, Journal of Hydrology, 606, 127,405. doi: 10.1016/j.jhydrol.2021.127405.

Zhang, X., Jiang, S., Zheng, N., et al. (2024a). Integration of DDPM and ILUES for simultaneous identification of contaminant source parameters and non‐Gaussian channelized hydraulic conductivity field. Water Resources Research, 60(9), e2023WR036893. doi: 10.1029/2023WR036893.

Zhang. W., T. Xu, Z. Chen, J. J. Gómez-Hernández, C. Lu, J. Yang, Y. Ye, and M. Jing (2024b), Simultaneous identification of a non-point contaminant source with gaussian spatially distributed release and heterogeneous hydraulic conductivity in an aquifer using the les-mda method, Journal of Hydrology, 630, 130,745.

doi:10.1016/j.jhydrol.2020.125681.

Zheng, N., Li, Z., Xia, X., et al. (2024). Estimating line contaminant sources in non-Gaussian groundwater conductivity fields using deep learning-based framework. Journal of Hydrology, 630, 130727. doi: 10.1016/j.jhydrol.2024.130727.