Water Air Soil Pollut (2020) 231: 27
https://doi.org/10.1007/s11270-019-4369-5

Optimization Design of Groundwater Pollution Monitoring Scheme and Inverse Identification of Pollution Source Parameters Using Bayes' Theorem

Shuangsheng Zhang · Jing Qiang · Hanhu Liu · Yanyan Li

Received: 9 September 2019 / Accepted: 5 December 2019 / Published online: 11 January 2020

Abstract In the process of identifying groundwater pollution sources, in order to solve the problem that the monitoring data of monitoring wells was insufficient or the correlation between monitoring data and model parameters was weak, a monitoring well optimization method based on Bayesian formula and information entropy was proposed. Two-dimensional phreatic groundwater solute transport model was built and solved by using GMS software. To reduce the computational load of calling the numerical model repeatedly in the optimization design of the monitoring schemes and the identification process of the pollution sources, the Kriging method was used to establish the surrogate model of the numerical model. Under the condition of single well monitoring and determined monitoring frequency, with the target of optimization of monitoring position number D and monitoring time interval Δt, both the single-objective monitoring scheme with the minimum information entropy of the model parameter posterior distribution and the multi-objective monitoring scheme with the minimum information entropy and the shortest monitoring time were optimized respectively. According to the above-optimized monitoring schemes, the delayed rejection adaptive Metropolis algorithm was used to identify the pollution source parameters. The case study results showed that under the condition of pre-set single well monitoring with monitoring frequency of 10 times, the single-objective optimized monitoring scheme was $D = 37$ and $\Delta t = 20$ days. Under this monitoring scheme, the mean errors of inversion pollution source parameters $\alpha = (X_S, Y_S, T_1, T_2, Q_S)$ were 0.09%, 0.4%, 4.72%, 2.43%, and 9.29%, respectively. The multi-objective optimized monitoring scheme was $D = 37$ and $\Delta t =$ 2 days. Under this monitoring scheme, the mean errors of the inversion parameters $\alpha = (X_S, Y_S, T_1, T_2, Q_S)$ were 12.76%, 3.77%, 5.13%, 1.36%, and 7.68%, respectively. Compared with the monitoring scheme based on the single-objective optimization, although the inversion mean error of the five parameters based on the multi-objective optimized monitoring scheme increased by 2.75%, the monitoring time significantly reduced from 180 to 18 days.

Keywords Monitoring well optimization · Pollution source identification · Bayes' theorem · Information entropy · Kriging method · Delayed rejection adaptive Metropolis algorithm · Latin hypercube sampling

S. Zhang
College of Environmental Engineering, Xuzhou University of Technology, Xuzhou 221018, China

J. Qiang (✉)
School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China
e-mail: qiangjingcumt@163.com

H. Liu
School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

Y. Li
Jinan civil air defense architectural design research institute CO., LTD, Jinan 250014, China

1 Introduction

Identification of groundwater pollution sources is the process of reversing the location of the pollution source, the release intensity of the pollution source and the release time. And the process is carried out by establishing groundwater solute transport model and using the monitoring data in the monitoring well. Clearly, the essence of identification of groundwater pollution sources is to inverse and identify the solute transport model parameters by using the monitoring data. At present, the methods for solving the inverse problem mainly include Bayesian statistical method (Sohn et al. 2000; Chen et al. 2018), geostatistical method (Snodgrass and Kitanidis 1997), differential evolution algorithm (Ruzek and Kvasnicka 2001), genetic algorithm (Giacobbo et al. 2002), simulated annealing algorithm (Dougherty and Marryott 1991), and Kalman filter (Wang et al. 2018). Among them, the Bayesian statistical method aims to obtain parameter information from the monitoring data and combines the parameter prior probability density function with the sample likelihood function, so it is seen as a set of very flexible and intuitive methods for the inverse problem, and applied more and more extensively.

In the inversion of model parameters, it is often necessary to solve the posterior estimation value or posterior distribution of the parameters by Bayesian statistical methods. However, when the dimension of the model parameters is large, the numerical integration solution process is complicated and difficult. So the Monte Carlo method (MC) (Roberts and Casella 2004) is used for approximate solution. And the Markov chain Monte Carlo method (MCMC) (Metropolis et al. 1953; Hastings 1970; Tierney and Mira 1999; Mira 2002; Haario et al. 2001; Haario et al. 2006) is widely used as a classical sampling method. In recent years, some common methods of constructing Markov chains have been developed, such as Metropolis-Hastings algorithm (MH) (Metropolis et al. 1953; Hastings 1970), delay rejection algorithm (DR) (Tierney and Mira 1999; Mira 2002), adaptive Metropolis algorithm (AM) (Haario et al. 2001), and delay rejection adaptive Metropolis algorithm (DRAM) (Haario et al. 2006). The DRAM algorithm combines the DR algorithm and the AM algorithm, which not only ensures the local adaptation of the Markov chain but also guarantees the global adaptive adjustment of the chain. Wei et al. (2016) applied the DRAM algorithm to identify the source information after a sudden water contamination incident. Zhang (2017) used the DRAM algorithm to invert the parameters of the groundwater model and also pointed out its defects. For example, the DRAM algorithm was a single-chain MCMC algorithm, and suitable for the parameter posterior distribution to be a single-peak case. Therefore, an improved multi-chain delay rejection adaptive Metropolis algorithm based on Latin hypercube sampling (Gao 2008) was proposed.

On the other hand, the results of model parameters inversion are affected by monitoring scheme including monitoring well position, quantity, and monitoring frequency. However, the monitoring scheme is often limited by monitoring funds and other objective conditions, leading to ill-posedness (Carrera and Neuman 1986). In order to get an ideal model parameter inversion result, it is necessary to optimize the monitoring scheme. Firstly, an objective function needs to be defined to quantify the information amount of the monitoring scheme. Some objective functions have been developed so far, such as signal-to-noise ratio (SNR) (Gabriela et al. 2008) and relative entropy based on Bayesian formula (Huan and Marzouk 2013; Lindley 1956). However, the SNR only considers the interference effect of monitoring error on the monitoring data. And the relative entropy does not consider the influence of the prior distribution of parameters on the posterior distribution. Shannon (1948) pointed out that information entropy was a measure of information uncertainty. The greater the uncertainty, the larger the information entropy. This paper combined the Bayesian formula with information entropy (Shannon 1948; Zhang et al. 2019) to optimize the monitoring scheme.

In the process of the optimization design of monitoring scheme and the identification of pollution source, it is necessary to repeatedly call the groundwater solute transport model, making the calculation load very high. However, the application of the surrogate model can effectively reduce the calculation load. The commonly used methods of constructing surrogate model include polynomial regression (Knill et al. 1999) and Kriging (Kuhnt and Steinberg 2010; Luo et al. 2019). Kriging is an improved method of polynomial regression analysis, and the Kriging surrogate model can be established in MATLAB software by using the special DACE toolbox (Lophaven et al. 2002). So the Kriging method is used widely for constructing surrogate model.

In this paper, a two-dimensional solute transport simulation model for phreatic groundwater was established.

Under the condition of initial monitoring time and monitoring frequency, the Kriging method was used to establish a surrogate model of the solute transport simulation model. The optimized single-objective monitoring scheme MP1* with the minimum information entropy and the optimized multi-objective monitoring scheme MP2*with minimum information entropy and shortest monitoring time were calculated respectively. Then the improved multi-chain delay rejection adaptive Metropolis algorithm was used to identify the pollution source parameters based on the two optimized monitoring schemes. This paper will provide reference for the identification of groundwater pollution source and the optimization of monitoring schemes.

2 Study Methods

2.1 Bayesian Formula

The Bayesian formula is expressed as follows:

$$p(\boldsymbol{\alpha}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})}{p(\boldsymbol{d})} \propto p(\boldsymbol{d}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}) \tag{1}$$

Where,

- α is the unknown model parameter;
- d is the monitoring data;
- $p(\alpha|\, d)$ is the posterior probability density function of the model parameter;
- $p(\alpha)$ is the prior probability density function of the model parameter;
- $p(d|\, \alpha)$ is the conditional probability density function;
- $p(d) = \int p(d|\, \alpha)p(\alpha)\, d\alpha$ is the normalized integral constant, also called appearance probability of monitoring data d.

Assuming that

- The number of the unknown parameters in the model is m, namely $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)$;
- The environmental hydraulic model parameters are all distributed in a specific range;
- Each parameter obeys uniform distribution;
- $\alpha_1, \alpha_2, \cdots, \alpha_m$ are mutually independent.

So the prior probability density function of model parameter α_i can be defined as follows:

$$p(\alpha_i) = \begin{cases} \dfrac{1}{B_i - A_i}, & \alpha_i \in [A_i, B_i] \\ 0, & others \end{cases} \tag{2}$$

And the total prior distribution $p(\alpha)$can be expressed as follows:

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{m} p(\alpha_i) \tag{3}$$

- The monitoring data in the model is recorded as $d = (d_1, d_2, ..., d_n)$;
- $F(\alpha)$ indicates the calculated values of model under the condition of parameters α, and $\varepsilon = d - F(a)$ represents the error;
- $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ obeys normal distribution with the mean of 0;
- $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are mutually independent.

So the conditional probability density function can be expressed as follows:

$$p(d|\alpha) = \frac{1}{(2\pi)^{n/2}|C(\varepsilon)|^{1/2}} \exp\left\{-\frac{1}{2}(d - F(\alpha))^T C(\varepsilon)^{-1}(d - F(\alpha))\right\}, \tag{4}$$

where,

$$C(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix};$$

- $|\, C(\varepsilon)\,|$ is the determinant of matrix $C(\varepsilon)$;
- $C(\varepsilon)^{-1}$is the inverse matrix of matrix $C(\varepsilon)$;
- $\sigma_i > 0 (i = 1, 2, \cdots, n)$.

Combining the above functions (1), (2), (3), and (4), the posterior probability density function $p(\alpha|\, d)$ of α can be expressed as follows:

$$\begin{aligned} p(\alpha|d) &= \frac{\prod_{i=1}^{m} p(\alpha_i)}{p(d)(2\pi)^{n/2}|C(\varepsilon)|^{1/2}} \exp\left\{-\frac{1}{2}(d - F(\alpha))^T C(\varepsilon)^{-1}(d - F(\alpha))\right\} \\ &= \lambda \exp\left\{-\frac{1}{2}(d - F(\alpha))^T C(\varepsilon)^{-1}(d - F(\alpha))\right\} \end{aligned} \tag{5}$$

where $\lambda = \frac{\prod_{i=1}^{m} p(\alpha_i)}{p(\boldsymbol{d})(2\pi)^{n/2}|C(\varepsilon)|^{1/2}}$ is a fixed value, and independent of parameters α.

Equation (5) can be viewed as a function about parameters α under the condition that the measured value is fixed. Since it was difficult to draw the explicit expression of Eq. (5) by a numerical integral method, the Markov Chain Monte Carlo method was employed to solve the equation.

2.2 Optimization Design of Monitoring Scheme Based on Bayesian Formula and Information Entropy

The optimization design of the monitoring schemes mainly includes the optimization of the number, positions of the monitoring wells, and monitoring frequency. Under the condition of single well monitoring, both position D (D indicates the serial number of the monitoring wells) and monitoring time interval Δt of the monitoring wells will be optimized simultaneously.

Assume that

- The initial monitoring time is t1 (fixed value);
- Monitoring scheme MP = (D, Δt);
- Monitoring data is still recorded as d

The Bayesian formula can be rewritten as follows:

$$p(\boldsymbol{\alpha}|\boldsymbol{d}, D, \Delta t) = \frac{p(\boldsymbol{\alpha}|D, \Delta t)p(\boldsymbol{d}|\boldsymbol{\alpha}, D, \Delta t)}{\int p(\boldsymbol{\alpha}|D, \Delta t)p(\boldsymbol{d}|\boldsymbol{\alpha}, D, \Delta t)d\boldsymbol{\alpha}} \quad (6)$$

The prior distribution $p(\alpha|\, D, \Delta t)$ suggests a preliminary set of unknown parameters α, and is not affected by D and Δt. So $p(\alpha|\, D, \Delta t)$can be written as $p(\alpha)$. Equation (6) becomes the following:

$$p(\boldsymbol{\alpha}|\boldsymbol{d}, D, \Delta t) = \frac{p(\boldsymbol{\alpha})p(\boldsymbol{d}|\boldsymbol{\alpha}, D, \Delta t)}{\int p(\boldsymbol{\alpha})p(\boldsymbol{d}|\boldsymbol{\alpha}, D, \Delta t)d\boldsymbol{\alpha}} \quad (7)$$

where $\int p(\alpha)p(d|\, \alpha, D, \Delta t)d\alpha$ indicates the probability of monitoring value d obtained on the condition of position number of D and the monitoring time interval of Δt. So it can be denoted as $p(d|\, D, \Delta t)$ in the following format:

$$p(\boldsymbol{d}|D, \Delta t) = \int p(\boldsymbol{\alpha})p(\boldsymbol{d}|\boldsymbol{\alpha}, D, \Delta t)d\boldsymbol{\alpha} \quad (8)$$

Assuming that the probability density function of one-dimensional continuous random variable Θ is $f(\theta)$, the information entropy (Shannon 1948) of Θ in the interval $[a, b]$ can be defined as follows:

$$H(\Theta) = -\int_a^b f(\theta)\ln f(\theta)d\theta \quad (9)$$

So we can use the monitoring data d gotten at the position D to back-calculate the unknown parameters α, and then the posterior probability density function $p(\alpha|\, d, D, \Delta t)$ can be obtained. The information entropy of the posterior distribution α can be similarly expressed as follows:

$$H(D, \Delta t, \boldsymbol{d}) = -\int p(\boldsymbol{\alpha}|\boldsymbol{d}, D, \Delta t)\ln\ p(\boldsymbol{\alpha}|\boldsymbol{d}, D, \Delta t)d\boldsymbol{\alpha} \quad (10)$$

The left side of Eq. (10) contains monitoring data d, which could not be really obtained before the optimization design of the monitoring schemes. So d could be considered as a random variable, and the probability density function of d can be expressed as $p(d|\, D, \Delta t)$. In order to obtain a function only containing variable D and Δt, both sides of Eq. (10) are multiplied by $p(d|\, D, \Delta t)$, then integrated by d. And the expectation of information entropy $H(D, \Delta t, d)$ can be written as follows:

$$\begin{aligned} &E(H(D, \Delta t, d)) \\ &= -\int \left[\int p(\alpha|d, D, \Delta t)\ln p(\alpha|d, D, \Delta t)d\alpha\right] p(d|D, \Delta t)dd \\ &= -\iint p(\alpha|d, D, \Delta t)p(d|D, \Delta t)\ln p(\alpha|d, D, \Delta t)d\alpha dd \end{aligned} \quad (11)$$

where $E(H(D, \Delta t, d))$ is only affected by D and Δt, and is a continuous function on D and Δt. Therefore, $E(H(D, \Delta t, d))$ can be expressed as $E(D, \Delta t)$. And the optimal monitoring scheme MP* can be gotten by calculating the minimum value of $E(D, \Delta t)$. According to the concept of information entropy, we can use the monitoring value d^* from monitoring scheme MP* to back-calculate the unknown parameters α. At this time, the information entropy of the posterior distribution of α is the smallest, indicating that the uncertainty of α is also minimal, and the inversion effect is optimal.

The solving algorithm of Eq. (11) is very complicated, and it is difficult to obtain the expression. This paper will get the approximate result by using Monte Carlo method (Huan and Marzouk 2013; Zhang et al. 2019).

2.3 Improved Multi-chain Delay Rejection Adaptive Metropolis Algorithm Based on Latin Hypercube Sampling

2.3.1 Improved Multi-chain Delay Rejection Adaptive Metropolis Algorithm

Delayed rejection adaptive Metropolis algorithm (DRAM) was first proposed by Haario and others in 2006. The specific steps of the algorithm can be found in Haario et al. (2006). However, the single-chain DRAM algorithm easily causes the inversion result local convergence or no convergence (Zhang 2017). This paper proposes an improved multi-chain delay rejection adaptive Metropolis algorithm (multi-chain DRAM) based on Latin hypercube sampling. Latin hypercube sampling is a multi-dimensional hierarchical random sampling method with good dispersion uniformity and representation. The specific algorithm of Latin hypercube sampling is shown in the literature (Gao 2008).

Specific steps of the improved multi-chain DRAM algorithm based on Latin hypercube sampling are as follows:

- q sets of initial samples are randomly extracted from the prior ranges of model parameters by the Latin hypercube sampling method.
- Taking the q sets of samples as initial points in step (1), q parallel Markov Chains are generated by the DRAM algorithm.
- Convergence judgment. If the Markov chain satisfies the Gelman-Rubin convergence criterion (Gelman and Rubin 1992), the calculation terminates, otherwise the parallel sequence continues to evolve.
- The averages of the calculated results of q Markov Chains are taken as the final results.

2.3.2 Convergence Judgment of Improved Multi-chain DRAM Algorithm

In this study, the convergence of the last 50% sampling process by the multi-chain DRAM algorithm is guided by the Gelman-Rubin convergence diagnosis method (Gelman and Rubin 1992). The convergence indicator is as follows:

$$\hat{R}_i = \sqrt{\frac{g-1}{g} + \frac{q+1}{q} \cdot \frac{B_i}{W_i}}$$

where

- $\hat{R}_i\,(i = 1, 2, \cdots, m)$ is the judgment indicator of the i^{th} parameter;
- g is half the length of the Markov chain length in the multi-chain Metropolis algorithm;
- q is the number of Markov chains used for the judgment;
- B_i is the variance of the means of the last 50% samples in the q Markov chains of the i^{th} parameter;
- W_i is the average of the variance of the last 50% samples in the q Markov chains of the i^{th} parameter.

when $\hat{R}_i < 1.2$, the Markov chain converges; While when $\hat{R}_i \geq 1.2$, the Markov chain does not converge.

2.4 Kriging Surrogate Model and Sampling Method

In order to reduce the calculation load generated by repeatedly calling the groundwater solute transport numerical simulation model during the optimization design of the monitoring scheme and the parameter inversion process, the Kriging method (Lophaven et al. 2002) is used to construct the surrogate model of the numerical simulation model.

Both the establishment of the surrogate model and the selection of the initial samples of the improved DRAM algorithm need to adopt a certain sampling method to extract some samples. In the process of constructing the surrogate model, in order to ensure that the surrogate model can capture the trend of the object function, and the samples could be evenly distributed in the entire space of the prior distribution, the optimal Latin hypercube sampling method (Hickernell 1998) with centralization L_2 deviation (CL_2) as the optimizing index is used to extract samples. The improved DRAM algorithm is a multi-chain MCMC algorithm. The initial samples could be randomly extracted within the parameter prior distribution range by Latin hypercube sampling method (Kuhnt and Steinberg

2010), which could reduce the impact of randomly selecting samples on the inversion results.

3 Example Application

3.1 Model Establishment and Problem Overview

3.1.1 Model Establishment

Assuming that the study area was a rectangular area with 1000 m in length and 600 m in width, and the aquifer was a sandy aquifer with a thickness of 35 m (Table 1 for hydrogeological parameters), both the western boundary Γ_1 and the eastern boundary Γ_3 were the given head boundary. The eastern head was 25 m, and the western head was 30 m. Both the northern boundary Γ_2 and the southern boundary Γ_4 were the impermeable boundary.

A total of 58 monitoring wells were set up in the study area. The initial concentration of aquifer pollutant was zero. Pollutant x was found downstream of the study area on a certain day. The pollution source was initially determined in a certain upstream region S (a priori range). And the pollutant was continuously and constantly injected into the aquifer in the form of a water injection well (200 m^3/day) over a period of time. So the groundwater flow can be generalized as a two-dimensional homogeneous isotropic unsteady flow flowing from west to east. The schematic diagram of the study area is shown in Fig. 1.

The coordinate system was established with the southwest corner as the coordinate origin, and the numerical model of groundwater flow was established according to the hydrogeological conditions of the study area:

$$\begin{cases} \dfrac{\partial}{\partial x}\left[K(H-B)\dfrac{\partial H}{\partial x}\right]+\dfrac{\partial}{\partial y}\left[K(H-B)\dfrac{\partial H}{\partial y}\right]+w=\mu\dfrac{\partial H}{\partial t} & (x,y)\in\Omega,t\geq 0 \\ H(x,y,t)|_{t=0}=H_0(x,y) & (x,y)\in\Omega,t=0 \\ H(x,y,t)|_{\Gamma_1,\Gamma_3}=H_1(x,y,t) & (x,y)\in\Gamma_1,\Gamma_3,t\geq 0 \\ \left.\dfrac{\partial H}{\partial \tau}\right|_{\Gamma_2,\Gamma_4}=0 & (x,y)\in\Gamma_2,\Gamma_4,t\geq 0 \end{cases}$$

where K is the permeability coefficient, m/day: H is the water level, m: B is the aquifer floor elevation, m: w is the sources and sinks items: μ is the specific yield, dimensionless: Ω is the scope of the study area: $H_0(x, y)$ is the initial water level, m: H_1 is the known head on the first boundary, m: Γ_1, Γ_3 are the given head boundaries with Dirichlet boundary condition: Γ_2, Γ_4 are the impervious boundaries with Neumann boundary condition: τ is the outer normal direction for the Neumann boundary.

A numerical model of groundwater solute transport can be established based on the numerical model of groundwater flow. The boundaries of the simulated area can be generalized as follows: Γ_1 was the zero concentration boundary with Dirichlet boundary condition; Γ_3 was the convective diffusion flux boundary with Cauchy boundary condition; Γ_2 and Γ_4 were the zero diffusion flux boundaries with Neumann boundary condition. The solute transport model in the study area is as follows:

$$\begin{cases} n\dfrac{\partial c}{\partial t}=\dfrac{\partial}{\partial x}\left(nD_x\dfrac{\partial c}{\partial x}\right)+\dfrac{\partial}{\partial y}\left(nD_y\dfrac{\partial c}{\partial y}\right)-\dfrac{\partial}{\partial x}(v_x c)-\dfrac{\partial}{\partial y}(v_y c)+C_{inj}Q_{inj} & (x,y)\in\Omega,t\geq 0 \\ c(x,y,t)|_{t=0}=0 & (x,y)\in\Omega,t=0 \\ c(x,y,t)|_{\Gamma_1}=0 & (x,y)\in\Gamma_1,t>0 \\ D\left.\dfrac{\partial c}{\partial \tau}\right|_{\Gamma_2,\Gamma_4}=0 & (x,y)\in\Gamma_2,\Gamma_4,t>0 \\ \left.\left(-nD_x\dfrac{\partial c}{\partial x}+cv_x\right)\right|_{\Gamma_3}=f(x,y,t) & (x,y)\in\Gamma_3,t>0 \end{cases}$$

where D_x and D_y are the components of the hydrodynamic diffusion coefficient in the x and y directions, m^2/day: v_x and v_y are the percolation velocities of groundwater in the x and y directions respectively, m/day: n is the porosity of the aquifer medium, dimensionless: c is the mass concentration of the pollutant, mg/L: Q_{inj} is the amount of liquid injected into the aquifer, m^3/day: C_{inj} is the concentration of the pollutant entering the aquifer, mg/L: $f(x, y, t)$ indicates the solute mass passing through a unit area of flow sections in unit time only under the action of the hydrodynamic dispersion.

Table 1 Known hydrological parameters in the study area

Parameters	Longitudinal dispersion D_x/m	Transverse dispersion D_y/m	Permeability coefficient K/(m day^{-1})	Effective porosity$n$
Value	20	5	20	0.25
Parameters	Specific yield μ	Aquifer level H/m	Aquifer floor elevation B/m	
Value	0.15	35	0	

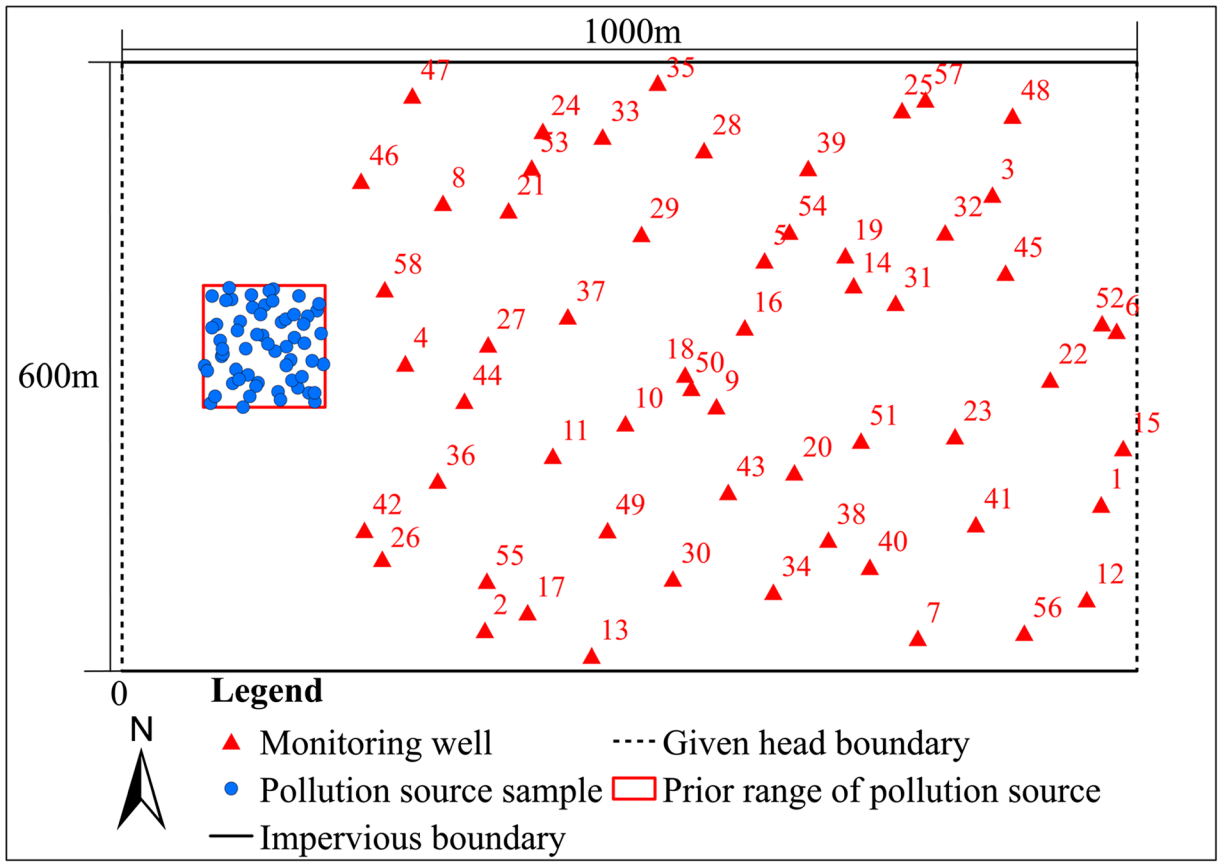

Fig. 1 Sketch of example model

The established groundwater flow and solute transport models were calculated by GMS (Groundwater Modeling System) software. In order to ensure that each grid center corresponded to a potential pollution source position, the study area was divided into 150 rows and 250 columns, and the basic cell side length is 4 m.

3.1.2 Problem Overview

For the potential pollution source parameter ranges, it was required to optimize the monitoring schemes using the existing 58 candidate monitoring wells. The optimized schemes contained single-objective and multi-objective monitoring schemes. The single-objective monitoring scheme was optimized with the minimum posterior distribution information entropy, and the multi-objective monitoring scheme was optimized with the minimum information entropy and the shortest monitoring time. Then the pollution source parameters, including the position of the pollution source, the start and stop time of emission pollutant, and the mass concentration of the pollutants, were identified based on the optimized schemes. That is, the unknown parameters of the pollution source $\alpha = (X_S, Y_S, T_1, T_2, Q_S)$ were solved, where (X_S, Y_S) is the position of the pollution source, m; T_1 and T_2 are the start and stop time of emission pollutant, d; Q_S is the mass concentration of the pollutant, mg/L.

3.1.3 Parameter Prior Range

The initial time was determined as a certain time when no pollution occurred. At this time, $t = 0$. Assuming that the prior distributions of the above five parameters $\alpha = (X_S, Y_S, T_1, T_2, Q_S)$ were evenly distributed, the prior ranges of the five parameters were as follows:

$80\ \text{m} \le X_S \le 200\ \text{m},\ 260\ \text{m} \le Y_S \le 380\ \text{m},\ 10\text{th day} \le T_1 \le 15\text{th day},\ 25\text{th day} \le T_2 \le 30\text{th day},\ 3,\ 000\ \text{mg/L} \le Q_S \le 3,\ 500\ \text{mg/L}.$

Table 2 Fifty sets of training input dataset obtained from the prior distribution

Serial number	X_s	Y_s	T_1	T_2	Q_s
1	128.6	358.2	10.7	27.9	3290.9
2	98.1	309.9	14.9	28.9	3236.3
3	192.0	355.0	10.5	25.7	3402.0
4	112.2	297.1	12.9	28.4	3146.5
5	133.9	284.1	13.8	25.5	3334.3
…	…	…	…	…	…
50	121.8	317.4	10.1	28.6	3284.8

3.2 Establishment of the Kriging Surrogate Model

Fifty sets of samples of α were evenly extracted from the prior distribution by using the optimal Latin hypercube sampling method. The samples were taken as the input dataset of the Kriging surrogate model (at this time,$CL_2(\Phi_{50,\ 5}) = 0.0054$). Fifty sets of samples are shown in Table 2.

Establishing the Kriging surrogate model for 58 candidate monitoring wells respectively.

The 50 sets of parameters in Table 2 were taken into the GMS software to obtain the daily pollutant mass concentrations of 58 candidate monitoring wells respectively within [450th day, 649th day]. The daily pollutant mass concentrations were seen as the output dataset of the Kriging surrogate models. Then the 50 sets of input and output dataset as the training samples were taken into MATLAB software. And the Kriging surrogate model of each monitoring well was trained by the DACE toolbox in the MATLAB software.

In order to test the accuracy of the Kriging surrogate models of 58 monitoring wells, 10 sets of parameters in the prior distributions of α in Table 3 were evenly extracted again by using the Latin hypercube sampling method, which were taken as the input values of the test samples. Then the 10 sets of parameters were taken into the GMS software to obtain the daily pollutant mass concentrations of 58 candidate monitoring wells respectively within [450th day, 649th day]. The daily pollutant mass concentrations were seen as the output values of the test samples, which were recorded as $y_{i,\ out}$, and $y_{i,\ out} = (y_{i,\ 1}, y_{i,\ 2}, \cdots, y_{i,\ 2000})$, where $i = 1, 2, \cdots, 58$ indicated the ith candidate monitoring well. Then the 10 sets of input values of the test samples were taken into the Kriging surrogate model to obtain output values, which were recorded as $\hat{y}_{i,out}$, and $\hat{y}_{i,out} = (\hat{y}_{i,1}, \hat{y}_{i,2}, \cdots, \hat{y}_{i,2000})$,where $i = 1, 2, \cdots, 58$ indicated the ith candidate monitoring well.

Taking the 58th monitoring well as an example, the output values of the test samples by numerical model were taken as the abscissa, and the output values of the Kriging surrogate model were taken as the ordinate. The comparison between the output values of the surrogate model and the test samples is plotted in Fig. 2. Figure 2 shows that the output values are concentrated in $y = x$, which indicates that the surrogate model can be a good substitute for the numerical model.

Then the coefficient of determination, the mean absolute error, and the root mean square error were used to further test and evaluate the accuracy of the surrogate models, as shown in Table 4.

(1). Coefficient of determination (R^2):

$$R_i^2 = 1 - \frac{\sum_{j=1}^{2000} \left(y_{i,j} - \hat{y}_{i,j}\right)^2}{\sum_{j=1}^{2000} \left(y_{i,j} - \overline{y}_i\right)^2}, i = 1, 2, \cdots, 58,$$

where $\overline{y}_i = \frac{\sum_{j=1}^{2000} y_{i,j}}{2000}$ represents the mean of the numerical model output values.

(2). Mean absolute error (MAE):

$$\mathrm{MAE}_i = \frac{\sum_{j=1}^{2000} \left|y_{i,j} - \hat{y}_{i,j}\right|}{2000}, i = 1, 2, \cdots, 58;$$

Table 3 Ten sets of testing input dataset obtained from the prior distribution

Serial number	X_s	Y_s	T_1	T_2	Q_s
1	132.8	331.5	14.4	29.4	3400.4
2	88.7	369.5	13.7	28.7	3309.5
3	178.8	341.8	11.7	29.9	3186.3
4	161.9	301.1	14.9	25.2	3295.5
5	148.4	364.7	10.8	25.8	3245.7
…	…	…	…	…	…
10	125.8	270.6	13.3	28.5	3125.4

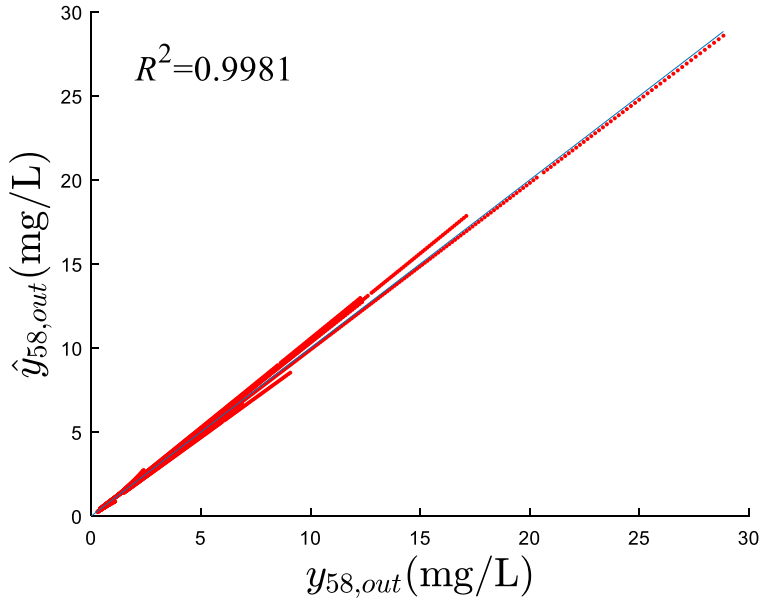

Fig. 2 Output comparison between the Kriging surrogate model and numerical model

(3). Root mean square error (RMSE):

$$\mathrm{RMSE}_i = \sqrt{\frac{\sum_{j=1}^{2000}\left(y_{i,j}-\hat{y}_{i,j}\right)^2}{2000-1}}, i = 1, 2, \cdots, 58.$$

It can be seen from the dataset in Table 4 that the Kriging surrogate model has higher prediction accuracy, indicating that the surrogate model can be a good substitute for the numerical model.

3.3 Optimization of Monitoring Schemes

3.3.1 Single-Objective Optimization Model Based on the Minimum Information Entropy

The value ranges of the parameters in the study area were described in Section 3.1.3. The first monitoring was set at time t1 = 450th day, and the monitoring was a total of 10 times. The interval between two adjacent monitoring was recorded as Δt, which is a positive integer, and 1 day ≤ Δt ≤ 20 days. That is, the pollution source identification task was completed before the 630th day. The purpose of this section was to select the optimal monitoring scheme MP1* = (D^*, Δt^*) from the 58 candidate monitoring well positions D and 20 monitoring intervals Δt. Therefore, the value ranges of the monitoring scheme (D, Δt) can be written as follows:

$$\Omega = \left\{ \begin{array}{l} 1 \le D \le 58,\ 1\ \text{day} \le \Delta t \le 20\ \text{days, and } D \text{ and } \Delta t \\ \text{are positive integers respectively} \end{array} \right\}.$$

It can be seen from Section 2.2 that the optimal monitoring scheme based on the minimum information entropy can be generalized to the minimum value of function (11), namely:

$$E\left(\mathrm{MP1}^*\right) = \min_{(D,\Delta t)\in\Omega} E(D, \Delta t) \quad (12)$$

The posterior probability density function in function (11) is given by Eq. (5), and the covariance matrix $C(\varepsilon)$ in Eq. (5) needs to be given. In the optimization design of monitoring schemes by using the Kriging surrogate model,

Table 4 R^2, MAE, and RMSE of the Kriging surrogate models for the 58 monitoring wells

Serial number	1	2	3	…	58	Mean
R^2	0.9798	0.9166	0.9947	…	0.9981	0.9804
MAE	0.0024	0.0007	0.0082	…	0.1606	0.1177
RMSE	0.0051	0.0014	0.0147	…	0.2178	0.1551

there must be error uncertainty in the surrogate models and measurement.

Assuming that

- The error $\varepsilon_i^{'}$ of the Kriging surrogate models for the i^{th} ($i = 1, 2, \cdots, 58$) monitoring well satisfies normal distribution $N\left(0, \left(\sigma_i^{'}\right)^2\right)$, the mean value $E\left(\varepsilon_i^{'}\right) = 0$, and mean square deviation $\sigma_i^{'} = \mathrm{RMSE}_i$;
- The measurement error $\varepsilon^{''}$ satisfies normal distribution $N(0, (\sigma^{''})^2)$, the mean value $E(\varepsilon^{''}) = 0$, and mean square deviation $\sigma^{''} = 0.01$;
- $\varepsilon_i^{'}$ and $\varepsilon^{''}$ are completely independent of each other.

The global error $\overline{\varepsilon}_i = \varepsilon_i^{'} + \varepsilon^{''}$ for the i^{th} ($i = 1, 2, \cdots, 58$) monitoring well satisfies normal distribution $N\left(0, \left(\sigma_i^{'}\right)^2 + \left(\sigma^{''}\right)^2\right)$. According to this, $C(\varepsilon)$ in Eqs. (4) and (5) can be determined.

The information entropy of all monitoring schemes can be obtained according to function (11). Because the values of D and Δt were all positive integers, the minimum value of objective function $E(D, \Delta t)$ was obtained, and $\min_{(D,\Delta t)\in\Omega} E(D, \Delta t) = 12.16$. So the optimal monitoring scheme MP1* = (37, 20). That is, the best monitoring well was no.37, and the best monitoring intervals was $\Delta t = 20$ days.

In order to verify the optimization design effect of the monitoring schemes based on Bayesian formula and information entropy, another 9 monitoring schemes were randomly selected from $\Omega = \{1 \le D \le 58,\ 1\ \mathrm{day} \le \Delta t \le 20$ days, and D and Δt were positive integers respectively$\}$. MP1* and the other new 9 monitoring schemes were represented by the symbol $(D_i, \Delta t_i)$ ($i = 1, 2, \cdots, 10$). Then the 10 monitoring schemes were evaluated by the information entropy $E(D, \Delta t)$ and the mean relative errors MRE$(D, \Delta t)$ of inversion results as indicators. However, it was unfair to evaluate the inversion results of the monitoring schemes by using a certain set of "the true values of parameters". So 20 sets of parameter values in the prior distributions of α were randomly and uniformly extracted by using Latin hypercube sampling method. The parameter values were recorded as "the true values of parameters" (Table 6), which was written as $\chi = [\chi(j,k)]_{20\times 5}$. Corresponding to the 10 monitoring schemes, 20 groups of " the true values of parameters" generated 200 groups of concentration monitoring values through Kriging surrogate models. Then the parameters α could be inversed by using the generated monitoring values and the improved DRAM algorithm (the number of parallel chain was 10). The length of each Markov chain is 34,000. When the length of the Markov chain is 30,000, the convergence judgment indexes of 5 parameters were $\hat{R}_i < 1.2\,(i = 1, 2, \cdots, 5)$. In order to ensure the accuracy of inversion results, only the last 4000 samples after a stabilization trend were used for posterior statistics, and the parameter posterior mean estimation $M_{(D_i,\Delta t_i)}$ of 20 sets of "the true values of parameters χ" can be calculated. $M_{(D_i,\Delta t_i)}$ was recorded as $M_{(D_i,\Delta t_i)} = \left[M_{(D_i,\Delta t_i)}(j,k)\right]_{20\times 5}$. Then the true values of parameters in Table 6 were brought into MRE(D_i, Δt_i). The expression was as follows:

$$\mathrm{MRE}(D_i, \Delta t_i) = \left[\sum_{j=1}^{20}\sum_{k=1}^{5}\left|\frac{M_{(D_i,\Delta t_i)}(j,k) - \chi(j,k)}{\chi(j,k)}\right|\right]/100 \quad (13)$$

MRE(D_i, Δt_i) of the 10 monitoring schemes were obtained by function (13). The results are shown in Table 5.

MRE(D_i, Δt_i) and $E(D_i, \Delta t_i)$ in Table 5 are fitted linearly, and they show a good positive linear relationship, which can be written as

MRE$(D, \Delta t) = 0.0072E(D, \Delta t) - 0.0345$ ($R^2 = 0.9101$), and shown in Fig. 3.

It can be seen from Fig. 3 and Table 5 that MRE(D, Δt) and $E(D, \Delta t)$ show a good positive linear relationship, which indicates $E(D, \Delta t)$ is an effective measure of the accuracy of the parameter inversion results. The smaller the $E(D, \Delta t)$, the higher the accuracy of the

Table 5 $E(D_i, \Delta t_i)$and MRE(D_i, Δt_i) of 10 monitoring schemes

i	$(D_i, \Delta t_i)$	$E(D_i, \Delta t_i)$	MRE(D_i, Δt_i)
1	(37, 20)	12.16	0.060
2	(23, 8)	17.13	0.086
3	(33, 10)	16.65	0.082
4	(50, 11)	12.88	0.053
5	(46, 15)	15.90	0.077
6	(3, 12)	17.23	0.089
7	(11, 6)	13.98	0.059
8	(38, 7)	17.61	0.093
9	(28, 12)	16.77	0.089
10	(41, 17)	16.41	0.088

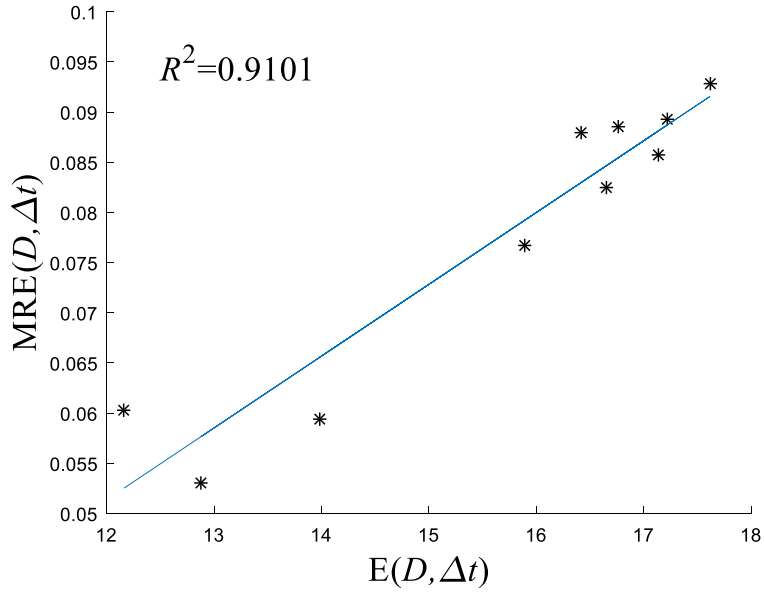

Fig. 3 The fitting diagram of the relationship between $E(D, \Delta t)$ and MRE$(D, \Delta t)$

parameter inversion. But when $E(D, \Delta t)$ gets the minimum, MRE$(D, \Delta t)$ is not the minimum. For example, $E(D_1, \Delta t_1)$=12.16 is less than $E(D_4, \Delta t_4)$=12.88, but MRE$(D_1, \Delta t_1)$=0.060 is greater than MRE$(D_4, \Delta t_4)$=0.053. The main reason is that although 20 sets of " the true values of parameters" in Table 6 are uniformly extracted from the prior distributions of α as far as possible by using Latin hypercube sampling method, the amount of "the true values of parameters" is relatively few, and it is impossible to make the values of parameters evenly distributed within the prior range in the true sense. The information entropy is solved by the Monte Carlo method (MC method). The Latin hypercube sampling method is used to extract 40,000 samples in the parameter prior ranges. So it is more reliable to take the minimum value of E(MP) as the index to select the optimal monitoring scheme. The minimum value of MRE(MP) cannot be used as the index to select the optimal monitoring scheme.

In summary, the smaller the information entropy $E(D, \Delta t)$, the smaller the uncertainty of the parameter posterior distribution, and the higher accuracy the inversion result. It fully verifies that the monitoring well optimization design method based on Bayesian formula and information entropy can be a good method for parameter inversion.

Table 6 Twenty sets of true values of parameters obtained from the prior distribution

Serial number	X_s	Y_s	T_1	T_2	Q_s
1	111.99	277.54	13.54	27.00	3485.45
2	88.55	300.65	14.53	29.14	3359.33
3	176.69	313.84	14.31	29.88	3228.53
4	195.77	306.27	10.48	27.14	3084.93
5	183.68	285.82	13.32	28.92	3313.37
…	…	…	…	…	…
20	121.53	368.70	12.27	28.27	3047.30

3.3.2 Multi-objective Optimization Model Based on Minimum Information Entropy and Minimum Monitoring Time

It not only needs to optimize the monitoring scheme with minimize information entropy but also requires the monitoring scheme to be the least time-consuming in order to find the pollution source as soon as possible. So the trade-off between information entropy and time-consuming of monitoring scheme should be considered. Assuming that the monitoring number of times was still set as 10, the multi-objective optimization model was established with the minimum information entropy and the shortest monitoring time. And the mathematical formula is as follows:

Objective 1: minimum information entropy

$$\min_{(D,\Delta t)\in\Omega} E(D, \Delta t) \tag{14}$$

Objective 2: shortest monitoring time

$$\min_{\Delta t\in[1,20],\text{positive inegers}} T = (10-1)\times\Delta t = 9\Delta t \qquad (15)$$

Generally speaking, the optimal solution of multi-objective optimization problem is not unique. The optimal solution set consists of the solutions whose reduction must be at the cost of increasing the value of other objective functions. The optimal solution set is called Pareto domain. The continuous multi-objective optimization problem can be solved by using the non-dominated sorting genetic algorithm with elite strategy (NSGA-II) (Deb et al. 2002). Since the values of D and Δt in the multi-objective optimization model are all positive integers, this paper used exhaustive method to get D^* with minimum information entropy under 20 kinds of Δt. The above 20 combinations of D^* and Δt are Pareto domains. The Pareto front is shown in Fig. 4.

If the time-consuming of monitoring scheme must be less than or equal to 20 days, the optimal monitoring scheme MP2* was calculated as follows: the best monitoring well was no.37, and the best monitoring intervals was $\Delta t = 2$ days, and the time-consuming $T = 18$ days. At this time, $E(\text{MP2}^*) = 13.59$.

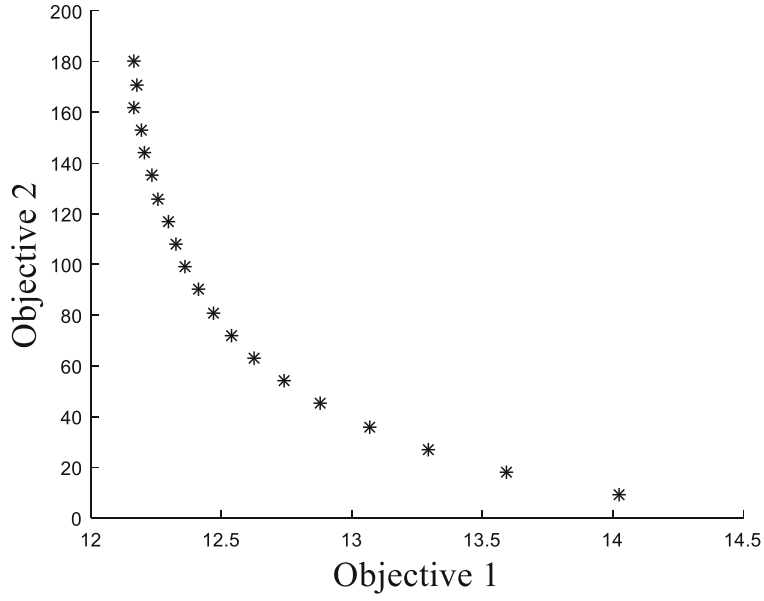

Fig. 4 Pareto front of the objective function

3.4 Identification of Pollution Source Based on Optimized Monitoring Schemes

Taking the true values of the first set of parameters in Table 6 ($X_S = 111.99$, $Y_S = 277.54$, $T_1 = 13.54$, $T_2 = 27.00$) as an example, the pollution source was identified by using the single-objective and multi-objective optimized monitoring schemes respectively.

3.4.1 Identification of Pollution Source Based on Single-Objective Optimized Scheme

Both the monitoring values obtained by single-objective optimized monitoring scheme MP1* and the improved DRAM algorithm (10 chains in total) were used to invert pollution source parameters. In the parameter inversion process, the length of each Markov chain was 34,000, among which the length of the non-adaptive Markov chain was 4000 and the length of the adaptive Markov chain was 30,000. The ergodic mean plots of model parameters based on MP1* are shown with the solid lines in Fig. 5. When the length of Markov chain was 30,000, the convergence judgment indexes of 5 parameters were $\hat{R}_i < 1.2\,(i = 1, 2, \cdots, 5)$, and all the Markov chains of all parameters converged. Then the previous unstable 30,000 Markov chains results were excluded. Only the last 4000 stable results were used to perform the posterior statistical analysis. The results are shown in Table 7.

3.4.2 Identification of Pollution Source Based on Multi-objective Optimized Scheme

Both the monitoring values obtained by multi-objective optimized monitoring scheme MP2* and the improved DRAM algorithm (10 chains in total) were used to invert pollution source parameters. In the inversion process, the Markov chain condition was set in the same way as in Section 3.4.1. The ergodic mean plots of model parameters based on MP2* are shown with the

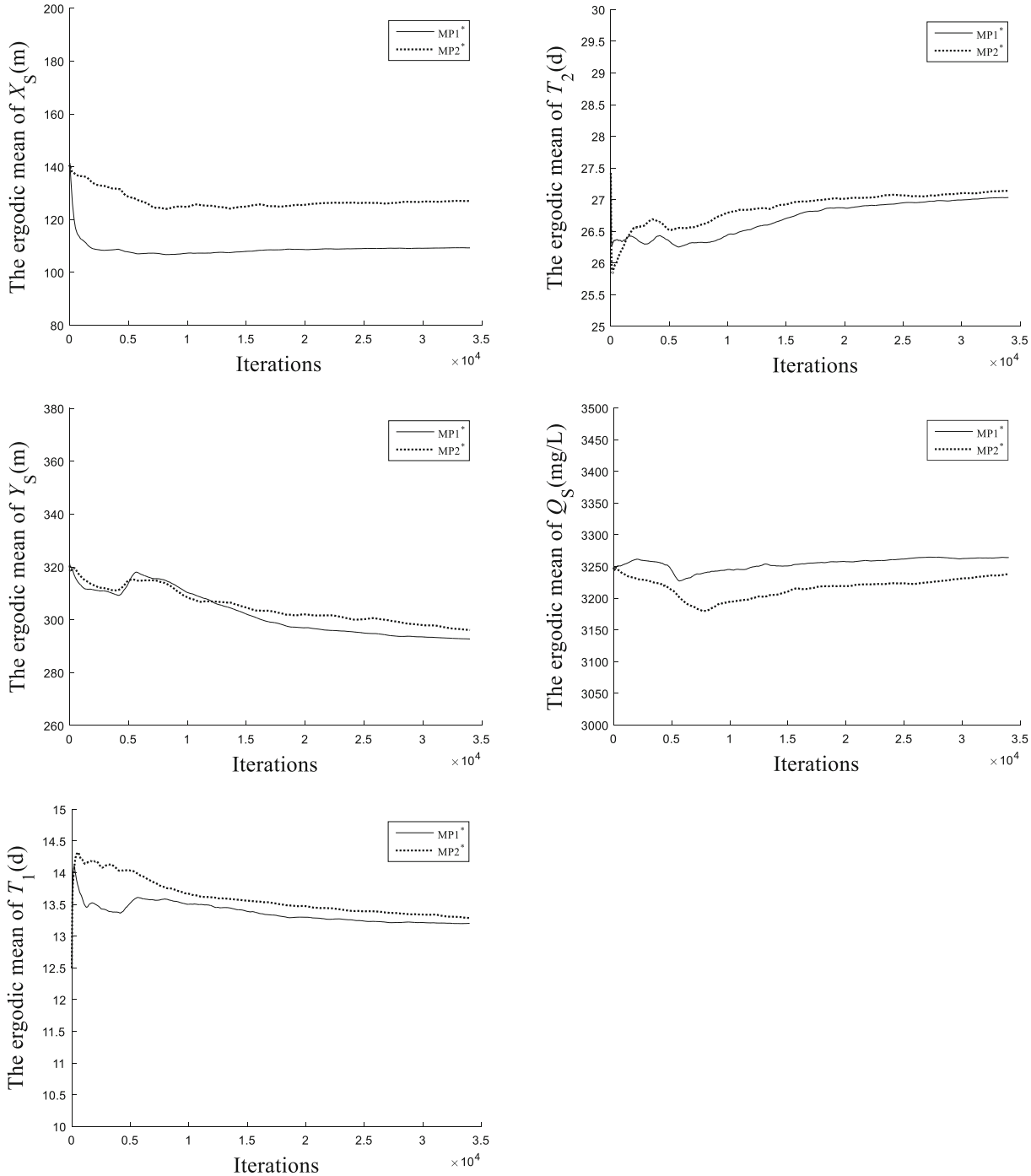

Fig. 5 Ergodic mean plots of model parameters based on MP1* and MP2*

Table 7 Posterior statistical results of model parameters based on MP1* and MP2*, and the convergence judgment indicators $\hat{R}_i$

True value of parameter α	Monitoring scheme MP1*					Monitoring scheme MP2*				
	Posterior mean $\overline{\alpha}$	Relative error of posterior mean %	Posterior median $\hat{\alpha}$	Relative error of posterior median %	$\hat{R}_i$	Posterior mean $\overline{\alpha}$	Relative error of posterior mean %	Posterior median $\hat{\alpha}$	Relative error of posterior median %	$\hat{R}_i$
$X_S = 111.99$	109.59	2.14	107.85	3.70	1.04	129.25	15.41	130.05	16.13	1.09
$Y_S = 277.54$	287.07	3.43	288.93	4.10	1.06	281.29	1.35	281.97	1.59	1.07
$T_1 = 13.54$	13.16	2.83	13.07	3.45	1.04	12.85	5.08	12.73	5.97	1.06
$T_2 = 27.00$	27.25	0.91	27.00	0.01	1.06	27.49	1.83	27.56	2.08	1.05
$Q_S = 3485.45$	3276.11	6.01	3284.46	5.77	1.15	3288.35	5.66	3274.68	6.05	1.11
Mean		3.06		3.41			5.87		6.36	

dotted line in Fig. 5, and the posterior statistical results are shown in Table 7.

It can be seen from Table 7 that the mean relative errors of the posterior mean of 5 parameters by MP1* and MP2* are 3.06% and 5.87% respectively. The precision of parameter inversion by MP1* is higher than that by MP2*, such as the inversion positions of the pollution source shown in Fig. 6. This is mainly due to $E(\text{MP1}^*) = 11.90 < E(\text{MP2}^*) = 13.35$. Moreover, when we observe pollutant concentrations at the no.37 monitoring well by forward simulation from the inversion pollution sources by MP1* and MP2*, we can see that the monitoring concentration residue between the pollution source by MP1* and the truth pollution source is less than that by MP2*. The comparison is shown in Fig. 7. It is further verified that the smaller the information entropy of the monitoring scheme, the smaller the uncertainty of the parameter posterior distribution, and the higher accuracy the inversion result.

Compared with the inversion results based on MP1*, the mean value of the relative errors of the posterior mean of 5 parameters increases by 2.81% by MP2*, but the monitoring time is shortened from 180 to 18 days. Therefore, the multi-objective optimized monitoring scheme is of more practical significance for the rapid identification of pollution source.

3.5 Sensitivity Analysis

As can be seen from Table 7, the relative errors of the posterior mean of parameters X_S, Y_S, T_1, and T_2 by MP1* are all small, but not Q_S. And the relative errors of the posterior mean of the parameter X_S and Q_S by MP2* are large. This is mainly due to the different sensitivities of

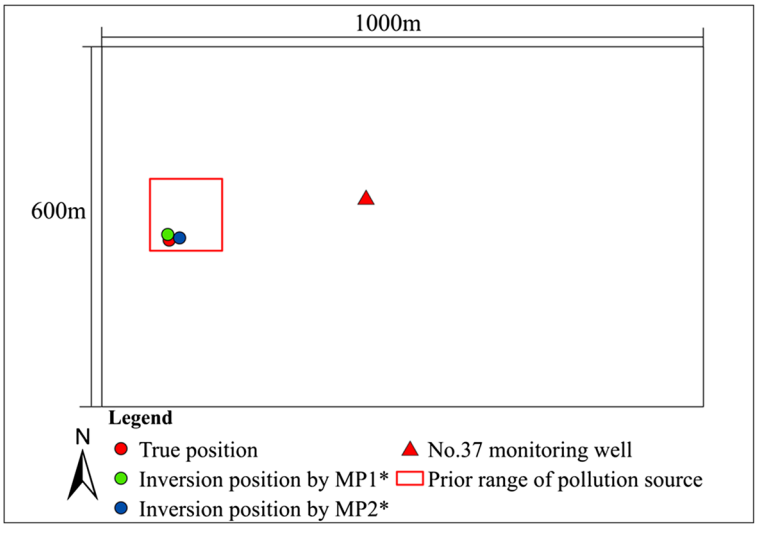

Fig. 6 Inversion positions of pollution source by MP1* and MP2*

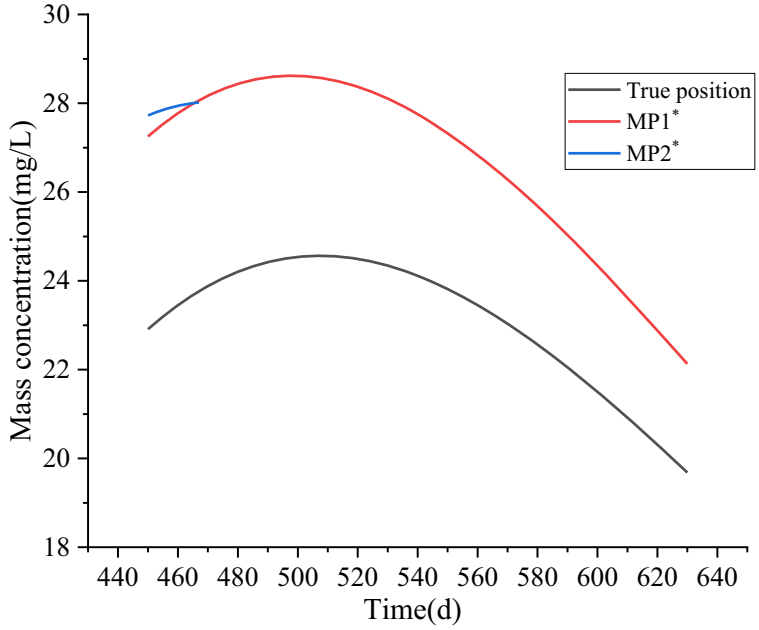

Fig. 7 Comparison of pollutant concentrations observed at the no.37 monitoring well from the inversion pollution sources by MP1* and MP2*

each parameter to the monitoring mass concentration values.

In order to avoid the defect that the local sensitivity analysis method does not consider the influence of the interaction between different parameters on the output results, the global sensitivity analysis method (Sobol' method) (Lenhart et al. 2002) and the Kriging surrogate model were used to obtain the first-order sensitivity coefficients of parameters α to the 10 monitoring datasets by MP1* and MP2* respectively. The results are shown in Table 8. According to the parameter sensitivity classification (Table 9) (Lenhart et al. 2002), the parameters X_S, Y_S, T_1, and T_2 by MP1* are medium sensitive parameters or sensitive parameters to the monitoring values, but Q_S is an insensitive parameter. Similarly, the parameters Y_S, T_1, and T_2 by MP2* are medium sensitive parameters or sensitive parameters, but X_S and Q_S are insensitive parameters.

4 Conclusion

The surrogate model of the numerical simulation model with high accuracy could be established by using the optimal Latin hypercube sampling and Kriging method. And the surrogate model could get the similar input-output relationship to the numerical simulation model with a small amount of calculation. So the surrogate model can significantly reduce the calculation load generated by repeatedly calling the groundwater solute transport numerical simulation model in the process of monitoring scheme optimization design and pollution source identification.

The mean relative errors of the parameter inversion results and the information entropy of the parameter posterior distribution show a good positive linear relationship, which indicates that information entropy is an effective measure of the accuracy of the inversion results. The smaller the information entropy, the higher the accuracy of the inversion results. The monitoring well optimization design method based on Bayesian formula and information entropy is an effective method to determine the monitoring scheme of groundwater pollution.

Compared with the single-objective optimized monitoring scheme, although the multi-objective optimized monitoring scheme can increase the error of the inversion results, it can significantly shorten the monitoring time. The multi-objective optimized monitoring scheme

Table 8 The first-order sensitivity coefficients of α

parameter	X_S	Y_S	T_1	T_2	Q_S
MP1*	0.080	0.490	0.187	0.194	0.014
MP2*	0.025	0.562	0.188	0.182	0.015

Table 9 Parameter sensitivity classification

classification	Highly sensitive parameter	Sensitive parameter	Medium sensitive parameter	Insensitive parameter
Ranges	$\lvert S_i \rvert \geq 1$	$0.2 \leq \lvert S_i \rvert < 1$	$0.05 \leq \lvert S_i \rvert < 0.2$	$0 \leq \lvert S_i \rvert < 0.05$

is of more practical significance for the rapid identification of pollution sources.

Multi-chain DRAM algorithm based on Latin hypercube sampling could avoid Markov chains falling into local optimum or the problem of difficulty in convergence. The algorithm could greatly enhance the accuracy of the parameter inversion results.

References

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resour Res, 22*(2), 211–227. https://doi.org/10.1029/wr022i002p00211.

Chen, M., Izady, A., Abdalla, O. A., & Amerjeed, M. (2018). A surrogate-based sensitivity quantication and Bayesian inversion of a regional groundwater fow model. *J Hydrol, 557*, 826–837. https://doi.org/10.1016/j.jhydrol.2017.12.071.

Deb, K., Pratap, A., Agarwal, S., & Meyarivant, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput, 6*(2), 182–197. https://doi.org/10.1109/4235.996017.

Dougherty, D. E., & Marryott, R. A. (1991). Optimal groundwater management: 1. Simulated annealing. *Water Resour Res, 27*(10), 2493–2508. https://doi.org/10.1029/91wr01468.

Gabriela, C., Sarma, S. V., Eden, U. T., & Brown, E. N. (2008). A signal-to-noise ratio estimator for generalized linear model systems. *Lect Notes Eng Computer Sci, 2171*(1), 1063–1069 http://www.iaeng.org/publication/WCE2008/WCE2008_pp1063-1069.pdf.

Gao. Y. (2008). Optimization methods based on Kriging surrogate model and their application in injection molding. Dalian: Dalian University of Technology.

Gelman, A. G., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat Sci, 7*, 457–472. https://doi.org/10.1214/ss/1177011136.

Giacobbo, F., Marseguerra, M., & Zio, E. (2002). Solving the inverse problem of parameter estimation by genetic algorithms: the case of a groundwater contaminant transport model. *Ann Nucl Energy, 29*(8), 967–981. https://doi.org/10.1016/S0306-4549(01)00084-6.

Haario, H., Saksman, E., & Tamminent, J. (2001). An adaptive Metropolis algorithm. *Bernoulli, 7*(2), 223–242. https://doi.org/10.2307/3318737.

Haario, H., Laine, M., & Mira, A. (2006). DRAM: efficient adaptive MCMC. *Stat Comput, 16*(4), 339–354. https://doi.org/10.1007/s11222-006-9438-0.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97–109. https://doi.org/10.2307/2334940.

Hickernell, F. A. (1998). A generalized discrepancy and quadrature error bound. *Math Comput, 67*(221), 299–322. https://doi.org/10.2307/2584985.

Huan, X., & Marzouk, Y. M. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems. *J Comput Phys, 232*(1), 288–317. https://doi.org/10.1016/j.jcp.2012.08.013.

Knill, D. L., Giunta, A. A., Baker, C. A., Grossman, B., Mason, W., Haftka, R., & Watson, L. T. (1999). Response surface models combining linear and Euler aerodynamics for supersonic transport design. *J Aircr, 36*(1), 75–86. https://doi.org/10.2514/2.2415.

Kuhnt, S., & Steinberg, D. M. (2010). Design and analysis of computer experiments. *AStA AdvStat Anal, 94*(4), 307–309. https://doi.org/10.1007/s10182-010-0143-0.

Lenhart, L., Eckhardt, K., Fohrer, N., & Frede, H. G. (2002). Comparison of two different approaches of sensitivity analysis. *Phys Chem Earth, 27*, 645–654. https://doi.org/10.1016/S1474-7065(02)00049-9.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Ann Math Stat, 27*(4), 986–1005. https://doi.org/10.1214/aoms/1177728069.

Lophaven, S. N., Nielsen, H. B., & Sondergaard, J. (2002). *Dace: a MATLAB Kriging toolbox*. Kongens Lyngby, Technical University of Denmark: Technical Report No. IMM-TR-2002-12.

Luo, J., Ji, Y., & Lu, W. (2019). Comparison of surrogate models based on different sampling methods for groundwater remediation. *J Water Resour Plan Manag, 145*(5), 04019015. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001062.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., & Teller, A. H. (1953). Equation of state calculations by fast computing machines. *J Chem Phys, 21*(6), 1087–1092. https://doi.org/10.1063/1.1699114.

Mira, A. (2002). Ordering and improving the performance of Monte Carlo Markov chains. *Stat Sci, 16*, 340–350. https://doi.org/10.1214/ss/1015346319.

Roberts, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (second edition). Springer https://doi.org/10.1007/978-1-4757-4145-2.

Ruzek, B., & Kvasnicka, M. (2001). Differential evolution algorithm in the earthquake hypocenter location. *Pure Appl Geophys, 158*, 667–693. https://doi.org/10.1007/PL00001199.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst Tech J, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

Snodgrass, M. F., & Kitanidis, P. K. (1997). A geostatistical approach to contaminant source identification. *Water*

Resour Res, 33(4), 537–546. https://doi.org/10.1029/96WR03753.

Sohn, M. D., Small, M. J., & Pantazidou, M. (2000). Reducing uncertainty in site characterization using Bayes Monte Carlo methods. *J Environ Eng-ASCE, 126*(10), 893–902. https://doi.org/10.1061/(ASCE)0733-9372(2000)126:10(893).

Tierney, L., & Mira, A. (1999). Some adaptive Monte Carlo methods for bayesian inference. *Stat Med, 18*, 2507–2515. https://doi.org/10.1002/(sici)1097-0258(19990915/30)18:17/18<2507::aid-sim272>3.

Wang, X., Zhang, D., & Chen, L. (2018). Real-time monitoring of pollutant diffusion states and source using fuzzy adaptive Kalman filter. *Water Air Soil Pollut, 229*(7), 238. https://doi.org/10.1007/s11270-018-3885-z.

Wei, G., Chi, Z., Yu, L., Liu, H., & Zhou, H. (2016). Source identification of sudden contamination based on the parameter uncertainty analysis. *J Hydroinf, 18*(6), 919–927. https://doi.org/10.2166/hydro.2016.002.

Zhang, J. (2017). Bayesian monitoring design and parameter inversion for groundwater contaminant source identification. Hangzhou: Zhejiang University- Thesis Doctorate.

Zhang, S., Liu, H., Qiang, J., Gao, H., Galar, D., & Lin, J. (2019). Optimization of well position and sampling frequency for groundwater monitoring and inverse identification of contamination source conditions using Bayes' Theorem. *Comput Model Eng Sci, 119*(2), 373–394. https://doi.org/10.32604/cmes.2019.03825.