 CMES, vol.119, no.2, pp.373-394, 2019

Optimization of Well Position and Sampling Frequency for Groundwater Monitoring and Inverse Identification of Contamination Source Conditions Using Bayes' Theorem

Shuangsheng Zhang[1, 5], Hanhu Liu[1], Jing Qiang[2, *], Hongze Gao[3, *], Diego Galar[4] and Jing Lin[4]

Abstract: Coupling Bayes' Theorem with a two-dimensional (2D) groundwater solute advection-diffusion transport equation allows an inverse model to be established to identify a set of contamination source parameters including source intensity (M), release location (X_0, Y_0) and release time (T_0), based on monitoring well data. To address the issues of insufficient monitoring wells or weak correlation between monitoring data and model parameters, a monitoring well design optimization approach was developed based on the Bayesian formula and information entropy. To demonstrate how the model works, an exemplar problem with an instantaneous release of a contaminant in a confined groundwater aquifer was employed. The information entropy of the model parameters posterior distribution was used as a criterion to evaluate the monitoring data quantity index. The optimal monitoring well position and monitoring frequency were solved by the two-step Monte Carlo method and differential evolution algorithm given a known well monitoring locations and monitoring events. Based on the optimized monitoring well position and sampling frequency, the contamination source was identified by an improved Metropolis algorithm using the Latin hypercube sampling approach. The case study results show that the following parameters were obtained: 1) the optimal monitoring well position (D) is at (445, 200); and 2) the optimal monitoring frequency (Δt) is 7, providing that the monitoring events is set as 5 times. Employing the optimized monitoring well position and frequency, the mean errors of inverse modeling results in source parameters (M, X_0, Y_0, T_0) were 9.20%, 0.25%, 0.0061%, and 0.33%, respectively. The optimized monitoring well position and sampling frequency can effectively safeguard the inverse modeling results in identifying the contamination source parameters.

[1] School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, 221116, China.

[2] School of Mathematics, China University of Mining and Technology, Xuzhou, 221116, China.

[3] GHD Services, Inc, 455 Philip Street, Waterloo, Ontario, Canada.

[4] Division of Operation and Maintenance Engineering, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, 97187 Luleå, Sweden.

[5] Xuzhou Urban Water Resources Management Office, Xuzhou, 221018, China.

[*] Corresponding Authors: Jing Qiang. Email: jingqiangsd@hotmail.com;
Hongze Gao. Email: hongze.gao@ghd.com.

It was also learnt that the improved Metropolis-Hastings algorithm (a Markov chain Monte Carlo method) can make the inverse modeling result independent of the initial sampling points and achieves an overall optimization, which significantly improved the accuracy and numerical stability of the inverse modeling results.

Keywords: Contamination source identification, monitoring well optimization, Bayes' Theorem, information entropy, differential evolution algorithm, Metropolis Hastings algorithm, Latin hypercube sampling.

1 Introduction

Sudden releases of contaminants to groundwater occur from time to time, which pose a serious threat to human health and can damage the environment. However, the source release locations, times, source strength and intensity are often unknown partially due to the detections of contaminants in groundwater come long after the release occurrences. Therefore, a better evaluation of source conditions is of great and practical significance to the remediation of groundwater contamination.

Using a groundwater flow and solute transport model to back-estimate the potential source conditions is here referred as contamination source inverse identification. Essentially, it is to use groundwater monitoring data to back-calculate the model input parameters (or to define the source conditions). A number of researchers employed the Bayes' Theorem solving inverse problems, such as, Sohn et al. [Sohn, Small and Pantazidou (2000)]; Zeng et al. [Zeng, Shi, Zhang et al. (2012)]; Chen et al. [Chen, Izady, Abdalla et al. (2018)]. Other methods solving inverse problems include geostatistics method [Snodgrass and Kitanidis (1997); Lin, Le, O'Malley et al. (2017)]; differential evolution algorithm [Ruzek and Kvasnicka (2001); Ramli, Bouchekara and Alghamdi (2018); Zhao (2007)]; genetic algorithm [Giacobbo, Marseguerra and Zio (2002); Mahinthakumar and Sayeed (2005); Bahrami, Ardejani and Baafi (2016); Bozorg-Haddad, Athari, Fallah-Mehdipour et al. (2018)]; the simulated annealing algorithm [Dougherty and Marryott (1991); Marryott, Dougherty and Stollar (1993)], and so on. Those above can be divided into deterministic methods and nondeterministic methods. Among the nondeterministic methods, the Bayesian statistical method uses the monitoring data to adjust the model input parameter, and then combines the sets of parameters prior probability density function with the sample likelihood function, so that forming a statistical method that is very flexible, intuitive, robust and easy to understand. This is why the Bayesian statistical method or Bayes' Theorem has become extensively employed in engineering applications.

Using the Bayesian statistical approach to reversely identify the model parameters (or the source conditions in this study), it is often necessary to solve the posterior estimation value or posterior distribution of the parameters. When the parameter dimension is not particularly high, a numerical integral method or normal approximation method can be used to solve the posterior estimation value or posterior distribution of the parameters [Tanner (1996)]. However, with the increase of the parameter dimension, the computational complexity of the numerical integral method will increase exponentially, which results in the difficulty of the solution process. So, the Monte Carlo Method

[Roberts and Casella (2004)] with independent samplings as an approximate solution is often used to solve the question. The Markov chain Monte Carlo method, as an efficient and rapid sampling method, has been widely used such as by Metropolis et al. [Metropolis, Rosenbluth, Rosenbluth et al. (1953)]; Hastings [Hastings (1970)]; Haario et al. [Haario, Saksman and Tamminen (2001)]; Xu et al. [Xu, Jiang, Yan et al. (2018)].

In real engineering applications, it is impractical to obtain a desired number of monitoring wells and samples due to the limitations of funds and/or restrictions of site conditions. Therefore, in reality, there is always insufficient data available to use, which could potentially result in ill-posed characteristic of inverse problem [Carrera and Neuman (1986)]. Therefore, to address the above limitations, the monitoring well locations together with the sampling frequency need to be optimized to potential satisfy the inverse parameters evaluation requirements. To do so, a few steps should follow. First, the data for monitoring plan need to be quantified by setting up an objective function [Zhang (2017)], of which the most commonly selected objective function is signal-to-noise ratio (SNR) [Czanner, Sarma, Eden et al. (2008)], i.e., the relative entropy from the Bayesian formula [Huan and Marzouk (2013); Zhang (2017); Lindley (1956)]. The SNR and relative entropy are applicable to nonlinearity models with an assumption that the parameter distribution is non-Gaussian [Huan and Marzouk (2013); Lindley (1956)]. Groundwater flow and transport models are mostly nonlinear so that the SNR approach and Bayesian formula using relative entropy apply. However, the SNR approach only factor in the effect of monitoring errors on the monitoring data, while the relative entropy does not include the influence of the prior distribution of parameters on the posterior distribution. To overcome the limitations of both the SNR approach and the relative entropy approach, the information entropy by Shannon [Shannon (1948)] was employed since it measures of information uncertainty. The greater of the information uncertainty, the greater of the information entropy will be. The information entropy in the posterior distribution of model parameters is used as a measurable value of the monitoring data. The smaller of the information entropy, the smaller of the uncertainty of the model parameters will be, and the better of the parameter inverse estimation results will expect.

This paper provides an example problem with an instantaneous release of a contaminant in a confined groundwater aquifer. The information entropy of the posterior distribution of the model parameters based on the Bayesian formula was used as the measurement value of the monitoring data. And the optimization of the monitoring well position and sampling frequency was solved by the Monte Carlo method and a differential evolution algorithm under the condition of single well monitoring and determining monitoring times. Relying on the optimized monitoring well program, the contamination source is identified by an improved Metropolis algorithm based on Latin hypercube sampling. This paper could provide reference for the optimization design of groundwater contamination monitoring wells and inverse problem of contamination source identification.

2 Study methods

2.1 Parameter inversion based on bayesian formula

The Bayesian formula [Berger (1995)] 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,

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

Assuming that

- The number of the unknown parameters in the model are m, namely $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)$;
- 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:

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

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

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

Assuming that the number of monitoring values in the model are n, namely $\boldsymbol{d} = (d_1, d_2, \cdots, d_n)$. Where d_i indicates the i[th] monitoring value. $C_i(x, y, t|\boldsymbol{\alpha})$ indicates the corresponding i[th] predictive value. Then $\varepsilon_i = d_i - C_i(x, y, t|\boldsymbol{\alpha})$ indicates the measurement error, $i = 1, 2, \cdots, n$. Assuming that ε_i obeys normal distribution with mean $\mu = 0$; and standard deviation $\sigma = 0.05$; and each ε_i is mutually independent, so the conditional probability density function $p(\boldsymbol{d}|\boldsymbol{\alpha})$ can be expressed as follows:

$$p(\boldsymbol{d}|\boldsymbol{\alpha}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\sum_{i=1}^{n} \frac{\left[d_i - C_i(x, y, t|\boldsymbol{\alpha}) \right]^2}{2\sigma^2} \right\} \tag{4}$$

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

$$p(\boldsymbol{\alpha}|\boldsymbol{d}) = \frac{\prod_{i=1}^{m} p(\alpha_i)}{(2\pi\sigma^2)^{n/2}\, p(\boldsymbol{d})} \exp\left\{-\sum_{i=1}^{n} \frac{\left[d_i - C_i(x,y,t|\boldsymbol{\alpha})\right]^2}{2\sigma^2}\right\} \tag{5}$$

Due to $\frac{\prod_{i=1}^{m} p(\alpha_i)}{(2\pi\sigma^2)^{n/2}\, p(\boldsymbol{d})}$ is a fixed value, and independent of parameters $\boldsymbol{\alpha}$, is expressed as λ. Then, the Eq. (5) can be written as:

$$p(\boldsymbol{\alpha}|\boldsymbol{d}) = \lambda \exp\left\{-\sum_{i=1}^{n} \frac{\left[d_i - C_i(x,y,t|\boldsymbol{\alpha})\right]^2}{2\sigma^2}\right\} \tag{6}$$

Eq. (6) can be viewed as a function about parameters $\boldsymbol{\alpha}$ under the condition that the measured value is fixed. Since it is difficult to draw the explicit expression of Eq. (6) by a numerical integral method, the Markov Chain Monte Carlo method is employed to solve the equation.

The core of Markov Chain Monte Carlo method is Monte Carlo simulation method and Markov chain sampling method. When the sample points are sufficient as "probability events", the probability can be approximately represented as frequency, which is the essence of the Monte Carlo simulation method. Employing the Markov chain sampling method can ensure that the Markov chain select more data points in the area with high probability, which can save the workload of the Monte Carlo simulation. The Metropolis algorithm is a classic Markov Chain Monte Carlo method, and therefore widely used. The Metropolis algorithm is utilized to find the solution in this article. The details of the Metropolis algorithm can be found in literature [Metropolis, Rosenbluth, Rosenbluth et al. (1953); Hastings (1970)].

2.2 Monitoring well design optimization

The optimization of the monitoring plan mainly includes monitoring well quantity, position and monitoring frequency. As an example, a single well is used to illustrate the general approach:

Assuming that the monitoring well position is $\boldsymbol{D}$, from which the monitoring data is still recorded as $\boldsymbol{d}$. Then the Bayesian formula can be rewritten as:

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

Since the prior distribution $p(\boldsymbol{\alpha}|\boldsymbol{D})$ of parameter $\boldsymbol{\alpha}$ suggests a preliminary set of unknown parameters (that is not affected by the monitoring well position, namely $p(\boldsymbol{\alpha}|\boldsymbol{D}) = p(\boldsymbol{\alpha})$), the Eq. (7) becomes the following:

$$p(\boldsymbol{\alpha}|\boldsymbol{d},\boldsymbol{D}) = \frac{p(\boldsymbol{\alpha})\, p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D})}{\int p(\boldsymbol{\alpha})\, p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D}) d\boldsymbol{\alpha}} \tag{8}$$

Normalizing the integral constant $\int p(\boldsymbol{\alpha})\, p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D}) d\boldsymbol{\alpha}$ results in the probability of

monitoring data $\boldsymbol{d}$ at position D expressed as $p(\boldsymbol{d}|\boldsymbol{D})$, in the following format:

$$p(\boldsymbol{d}|\boldsymbol{D}) = \int p(\boldsymbol{\alpha})\, p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D}) d\boldsymbol{\alpha} \tag{9}$$

Assuming that the probability density is a function of a random variable X [expressed as $f(x)$], the information entropy of X in the interval $[a,b]$ can be defined as follows [Shannon (1948)]:

$$H(X) = -\int_a^b f(x) \ln f(x) dx \tag{10}$$

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

$$H(\boldsymbol{D}, \boldsymbol{d}) = -\int p(\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{D}) \ln p(\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{D}) d\boldsymbol{\alpha} \tag{11}$$

The left side of Eq. (11) contains monitoring data $\boldsymbol{d}$, which can be considered as a random variable, with a probability density function $p(\boldsymbol{d}|\boldsymbol{D})$. In order to obtain a function only containing variable $\boldsymbol{D}$, both sides of Eq. (11) are multiplied by $p(\boldsymbol{d}|\boldsymbol{D})$. Then the expectation $E\left(H(\boldsymbol{D}, \boldsymbol{d})\right)$ of the information entropy $H(\boldsymbol{D}, \boldsymbol{d})$ can be written as:

$$\begin{aligned} E\left(H(\boldsymbol{D}, \boldsymbol{d})\right) &= -\int [\int p(\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{D}) \ln p(\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{D}) d\boldsymbol{\alpha}] p(\boldsymbol{d}|\boldsymbol{D}) d\boldsymbol{d} \\ &= -\iint p(\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{D})\, p(\boldsymbol{d}|\boldsymbol{D}) \ln p(\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{D}) d\boldsymbol{\alpha} d\boldsymbol{d} \end{aligned} \tag{12}$$

$E\left(H(\boldsymbol{D}, \boldsymbol{d})\right)$ is associated only with the monitoring well position $\boldsymbol{D}$, and is a continuous function on $\boldsymbol{D}$. Therefore $E\left(H(\boldsymbol{D}, \boldsymbol{d})\right)$ can be expressed as $E(\boldsymbol{D})$. The optimal monitoring well position $\boldsymbol{D}^*$ can be found by obtaining the minimum value of $E(\boldsymbol{D})$. According to the concept of information entropy, we can use the monitoring value $\boldsymbol{d}^*$ from monitoring well at the position $\boldsymbol{D}^*$ to back-calculate these unknown parameters $\boldsymbol{\alpha}$ under the condition that the information entropy of the posterior distribution of $\boldsymbol{\alpha}$ is minimal as an objective.

It is difficult to derive the explicit expression of Eq. (12). So, the Monte Carlo method [Huan and Marzouk (2013)] is used to find an approximately solution.

Using Eq. (8), Eq. (12) can be rewritten as follows:

$$\begin{aligned} & E(\boldsymbol{D}) \\ &= -\iint p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D})\, p(\boldsymbol{\alpha}) \ln[p(\boldsymbol{\alpha})\, p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D}) / p(\boldsymbol{d}|\boldsymbol{D})] d\boldsymbol{\alpha} d\boldsymbol{d} \\ &= -\iint p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D})\, p(\boldsymbol{\alpha}) [\ln p(\boldsymbol{\alpha}) + \ln p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D}) - \ln p(\boldsymbol{d}|\boldsymbol{D})] d\boldsymbol{\alpha} d\boldsymbol{d} \\ &= -\iint p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D})\, p(\boldsymbol{\alpha}) \ln p(\boldsymbol{\alpha}) d\boldsymbol{\alpha} d\boldsymbol{d} - \iint p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D})\, p(\boldsymbol{\alpha}) [\ln p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D}) - \ln p(\boldsymbol{d}|\boldsymbol{D})] d\boldsymbol{\alpha} d\boldsymbol{d} \end{aligned} \tag{13}$$

Since the prior distribution $p(\boldsymbol{\alpha})$ is given by Eqs. (2) and (3), the information entropy $p(\boldsymbol{\alpha})$ in Eq. (13) can be written as $-\iint p(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{D})\, p(\boldsymbol{\alpha}) \ln p(\boldsymbol{\alpha}) d\boldsymbol{\alpha} d\boldsymbol{d} = -\ln p(\boldsymbol{\alpha})$. This suggests that the greater of information entropy is, the greater of the uncertainty of $\boldsymbol{\alpha}$.

When $p(\alpha)$ remains unchanged, $-\ln p(\boldsymbol{\alpha})$ stays the same. In order to get the minimum value of $E(\boldsymbol{D})$, we only calculate the minimum value of $-\iint p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D})\,p(\boldsymbol{\alpha})[\ln p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D})-\ln p(\boldsymbol{d}|\boldsymbol{D})]d\boldsymbol{\alpha}d\boldsymbol{d}$.

If we name $U(\boldsymbol{D})$ as follows:

$$U(\boldsymbol{D})=-\iint p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D})\,p(\boldsymbol{\alpha})[\ln p(\boldsymbol{d}|\boldsymbol{\alpha},\boldsymbol{D})-\ln p(\boldsymbol{d}|\boldsymbol{D})]d\boldsymbol{\alpha}d\boldsymbol{d} \tag{14}$$

the monitoring value $\boldsymbol{d}$ in the Bayesian formula could reduce the uncertainty of parameters $\boldsymbol{\alpha}$, based on Lindley (1956).

Eq. (14) can be solved by Monte Carlo method as follows:

$$U(\boldsymbol{D})\approx-\frac{1}{N}\sum_{i=1}^{N}[\ln p(\boldsymbol{d}^i|\boldsymbol{\alpha}^i,\boldsymbol{D})-\ln p(\boldsymbol{d}^i|\boldsymbol{D})] \tag{15}$$

Firstly, we can randomly draw N samples from the prior distribution $p(\boldsymbol{\alpha})$ of unknown parameter $\boldsymbol{\alpha}$, namely $\boldsymbol{\alpha}^i$ $(i=1,2,\cdots N)$. Secondly, for each $i\in N$, we can get a sample $\boldsymbol{d}^i$ from the conditional probability density function $p(\boldsymbol{d}|\boldsymbol{\alpha}^i,\boldsymbol{D})$ according to Eq. (4), and a total of N. Finally, each group of $\boldsymbol{\alpha}^i$ and $\boldsymbol{d}^i$ is brought into Eq. (4), and we can obtain the $p(\boldsymbol{d}^i|\boldsymbol{\alpha}^i,\boldsymbol{D})$ in Eq. (15). From Eq. (9), $p(\boldsymbol{d}^i|\boldsymbol{D})$ in Eq. (15) can be rewritten as $p(\boldsymbol{d}^i|\boldsymbol{D})=\int p(\boldsymbol{\alpha})\,p(\boldsymbol{d}^i|\boldsymbol{\alpha},\boldsymbol{D})d\boldsymbol{\alpha}$, which will be solved by the Monte Carlo method as follows:

$$p(\boldsymbol{d}^i|\boldsymbol{D})\approx\frac{1}{N}\sum_{j=1}^{N}p(\boldsymbol{d}^i|\boldsymbol{\alpha}^j,\boldsymbol{D}) \tag{16}$$

2.3 Improved metropolis algorithm based on latin hypercube sampling

The Metropolis algorithm is a classic MCMC method, which is widely used. However, it is prone to local optimization, or difficult to converge by using Metropolis algorithm to invert the model parameters. Moreover, sampling efficiency is low. This study proposes an improved Metropolis algorithm based on Latin hypercube sampling.

2.3.1 Metropolis algorithm

The Metropolis algorithm was first proposed by Metropolis et al. [Metropolis, Rosenbluth, Rosenbluth et al. (1953)], and further revised by Hastings [Hastings (1970)]. Specific steps are as follows:

(1) The initial samples $\boldsymbol{Z}^i$ $(i=1)$ are generated randomly according to the prior distribution of parameters $\boldsymbol{\alpha}$.

(2) Make the uniform distribution $U(\boldsymbol{Z}^i-step,\boldsymbol{Z}^i+step)$ as the proposal distribution, and satisfy the symmetric random walk. Here $step$ represents the size of a random walk. Then a candidate sample $\boldsymbol{Z}^*$ is generated from $U(\boldsymbol{Z}^i-step,\boldsymbol{Z}^i+step)$, while u is extracted from the uniform distribution $U(0,1)$ randomly.

(3) In case $u < \min\left\{1, \frac{p(\boldsymbol{Z}^* \mid d)}{p(\boldsymbol{Z}^i \mid d)}\right\}$, let $\boldsymbol{Z}^{i+1} = \boldsymbol{Z}^*$. Otherwise, $\boldsymbol{Z}^{i+1} = \boldsymbol{Z}^i$. $p(\boldsymbol{Z}^* \mid d)$. Here $p(\boldsymbol{Z}^i \mid d)$ is calculated using Eq. (6).

(4) Repeat Steps (2) and (3) till the given iterative times are reached.

2.3.2 Improved metropolis algorithm based on latin hypercube sampling method

In order to prevent the local optimization of the inversion results, or to generate problems that are difficult to converge, this study uses the Latin hypercube sampling method [Dai (2011)] to optimize the sampling process to ensure the randomness and uniformity of the initial points of the sample. Latin hypercube sampling is a multi-dimensional hierarchical random sampling method with good dispersion uniformity and representativeness. When we want to extract q sets of samples from the prior distribution ranges $[A_i, B_i](i = 1, 2, \cdots, m)$ of m dimensional model parameters α by Latin hypercube sampling method**,** the specific steps are as follows:

(1) Dividing m prior distribution ranges $[A_i, B_i](i = 1, 2, \cdots, m)$ of m dimensional model parameters into q ranges, which can be recorded as $\left[A_{ij}, B_{ij}\right](i = 1, 2, \cdots, m; j = 1, 2, \cdots, q)$. Thus, $m \times q$ ranges are produced.

(2) Extracting α_{ij} from the range $[A_{ij}, B_{ij}]$ randomly, and total $m \times q$ numbers are produced. The following matrix is formed:

$$\Psi_{mq} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1q} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mq} \end{bmatrix}$$

(3) Arranging the row vector $[\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{iq}](i = 1, 2, \cdots, m)$ in the matrix Ψ_{mq} randomly into a new row vector $[\beta_{i1}, \beta_{i2}, \cdots, \beta_{iq}](i = 1, 2, \cdots, m)$. Thus, Ψ_{mq} will be transformed into a new matrix, which is:

$$\Phi_{mq} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mq} \end{bmatrix}$$

(4) Each column vector of matrix Φ_{mq} is a set of samples; and q sets of samples are combined together.

The specific steps of the improved Metropolis algorithm based on Latin hypercube sampling method are as follows:

(1) A number of *q* sets of initial samples are randomly obtained from the value ranges of model parameters by the Latin hypercube sampling method.

(2) Taking the q sets of samples as initial points in Step (1), *q* Markov Chains are

generated by the Metropolis algorithm.

(3) The averages of the calculated results of q Markov Chains are taken as the final results.

The improvement of this method lies in reducing the influence of the initial point values on the result by the Metropolis algorithm as much as possible, which accords with the idea of the Monte Carlo method.

2.3.3 Convergence judgment of the improved multi-chain metropolis algorithm

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

$$\hat{R}_i = \sqrt{\frac{g-1}{g} + \frac{q+1}{q} \cdot \frac{B_i}{W_i}} \tag{17}$$

where,

- $\hat{R}_i$ $(i = 1, 2, \cdots, m)$ is the i^{th} parameter judgment indicator;
- 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.

3 Example application

3.1 Example overview

In a homogeneous and confined aquifer, the (2D) groundwater solute advection-diffusion equation [Zheng and Gordon (2009)] can be expressed as:

$$\frac{\partial}{\partial x}(D_x \frac{\partial C}{\partial x}) + \frac{\partial}{\partial y}(D_y \frac{\partial C}{\partial y}) - \frac{\partial}{\partial x}(uC) = \frac{\partial C}{\partial t} \tag{18}$$

where,

- C is the concentration of a contaminant at the monitoring position (x, y) for the time t,(mg/L);
- t is the elapsed time from the release time (day);
- D_x and D_y are diffusion coefficients in the vertical and horizontal directions (m^2/day), respectively;

- u is an average groundwater flow velocity (m/day).

The analytical solution to Eq. (18) for the instantaneous release of a contaminant at a position can be expressed as:

$$C(x,y,t)=\frac{M}{4\pi kh\sqrt{D_x D_y}\,t}\exp\left[-\frac{(x-ut)^2}{4D_x t}-\frac{y^2}{4D_y t}\right] \tag{19}$$

where,

- M is the source release intensity (g);
- h is the thickness of the aquifer (m);
- x and y are the vertical and horizontal distances from the contamination source (m), respectively;
- k is the effective porosity of the aquifer.

Providing that the aquifer is homogeneous with a coordinate system in 2D as shown on Fig. 1; the release location of a contaminant is in the region of S; groundwater flow direction follows the x-axis; and the monitoring well position is $\boldsymbol{D}$; the regional hydrogeological parameters are presented in Tab. 1.

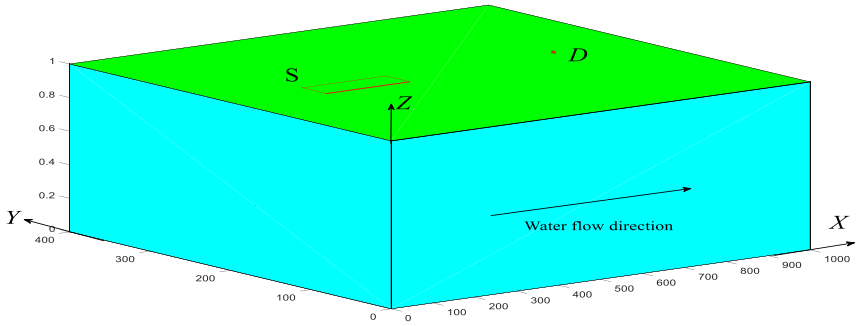


Figure 1: Schematic diagram of the example problem

Table1: Known hydrological parameters in the studied area

Parameter	D_x (m^2/d)	D_y (m^2/d)	u (m/d)	k	h (m)
Value	1.5	0.3	5	0.3	1

Table 2: Value range of parameter $\boldsymbol{\alpha}$ to be solved in the studied area

Parameter	Contamination source intensity M (g)	Contamination source position X_0 (m)	Contamination source position Y_0 (m)	Release time T_0 (days)
Value	[800,1600]	[200,400]	[185,215]	[40,60]

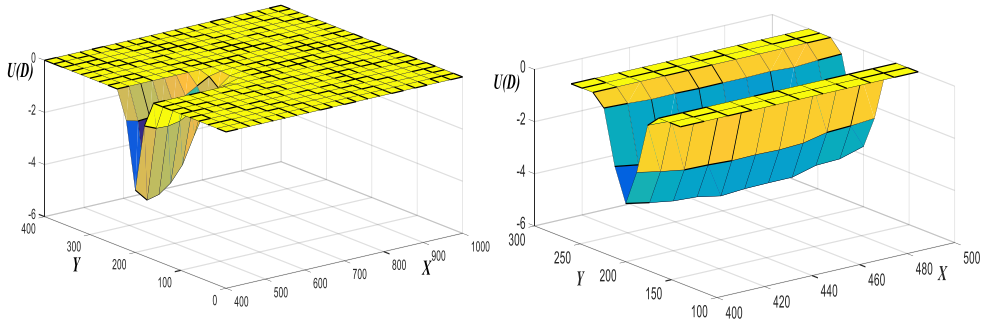
3.2 Monitoring well position optimization

A certain moment without pollution discharged in S is taken as the initial time, at this time $t=0$. The prior distribution ranges of the parameters to be solved in the studied area are shown in Tab. 2. The monitoring duration starts from $t=70$ and ends at $t=90$. A monitoring time interval is 4 days. Providing the solute migration velocity is 5 m/day, the possible range of optimal monitoring well position $\boldsymbol{D}^*$ is within this range: $\Omega=\{400\le X_1\le 1000,\ 0\le Y_1\le 400\}$.

The optimization problem of monitoring well position can be generalized as follows:

$$E(\boldsymbol{D}^*)=\min_{D\in\Omega}E(\boldsymbol{D})=-\ln p(\boldsymbol{\alpha})+\min_{D\in\Omega}U(\boldsymbol{D}) \tag{20}$$

$E(\boldsymbol{D})$ as continuous function of $\boldsymbol{D}$, so Eq. (20) can be solved by a differential evolution algorithm [Zhao (2007)]. In order to reduce the computational complexity, the two-step Monte Carlo method is used to reduce the range of position parameter $\boldsymbol{D}$ in the first step. In the first step, Ω is divided into 400 small rectangles; and $U(\boldsymbol{D})$ in the rectangular grid nodes can be calculated applying Eqs. (15) and (16) by Monte Carlo method (the number of samples, *N*, is 5,000). Then filtering out the region with $U(\boldsymbol{D})<0$, a narrowed range of optimal monitoring well position $\boldsymbol{D}^*$ is within $\Omega'=\{400\le X_1\le 670, 140\le Y_1\le 260\}$ as shown on Fig. 2(a). In the second step, a smaller range of optimal monitoring well position $\boldsymbol{D}^*$ is expressed as $\Omega''=\{400\le X_1\le 500, 196\le Y_1\le 204\}$ (Fig. 2(b)) following the computation process in the first step, but with a number of samples of 10,000 by the Monte Carlo method. Finally, the minimum value of the objective function $U(\boldsymbol{D})$ can be solved by the Monte Carlo method and differential evolution algorithm in Ω''. As a result, the minimum value $U(\boldsymbol{D}^*)=-4.5369$, and the optimal monitoring well position $\boldsymbol{D}^*=(445,200)$.



(a) The first step (b) The second step

Figure 2: Three dimensional stereograms of the two-step Monte Carlo method to solve $U(\boldsymbol{D})$

In order to verify the effectiveness of the monitoring well design optimization using the Bayesian formula, the information entropy $E(\boldsymbol{D})$ and the relative root mean square error of inversion results $RT(D)$ are used as an evaluation to compare the results of using 8

monitoring well positions randomly selected in $\Omega' = \{400 \le X_1 \le 670, 140 \le Y_1 \le 260\}$ with this settings for $\boldsymbol{D}^*$

(1) Information Entropy $E(\boldsymbol{D})$: The coordinates of the 8 monitoring wells are shown in Tab. 3. From Eqs. (13), (15), (16), $U(\boldsymbol{D})$ and $E(\boldsymbol{D})$ of the above 8 monitoring wells, $\boldsymbol{D}^*$ can be solved in Tab. 3. As illustrated in Tab. 3, the information entropy at $\boldsymbol{D}^*$ = (445,200) is the smallest.

(2) Relative Root Mean Square Error of Inversion Results $RT(D)$: $RT(\boldsymbol{D}_i)$ is the relative root mean square error between the posterior mean estimate M_{D_i} and real parameter R, which demonstrates the comprehensive influence of D_i on the inversion results. 20 sets of parameters are randomly and evenly obtained from the prior distribution of parameter $\boldsymbol{\alpha}$ as the real values (Tab. 4). Corresponding to the 8 monitoring positions in Tab. 3, 160 sets of monitoring values with measurement error $\boldsymbol{\varepsilon}$ can be calculated using Eq. (18). Then the parameter $\boldsymbol{\alpha}$ can be inversely estimated using the generated monitoring values (with a length of the Markov chain of 25,000). In order to ensure the accuracy of inversion results, only the last 2,000 samples after a stabilization trend are used to calculate the posterior mean value M_{D_i}. Then applying the M_{D_i} and R in Tab. 4 to Eq. (20) as follows:

$$RT(\boldsymbol{D}_i) = \sqrt{\left[\sum_{j=1}^{20}\sum_{k=1}^{4}\left(\frac{M_{D_i}(j,k) - R(j,k)}{R(j,k)}\right)^2\right] / 80} \tag{21}$$

Where,

- j is the j[th] set of parameter $\boldsymbol{\alpha}$;
- k is the k[th] component of parameter $\boldsymbol{\alpha}$.

According to Eq. (21), $RT(D_i)$, the inversion results of the 8 monitoring wells can be solved, as shown in Tab. 3.

Table 3: $\boldsymbol{D}_i$, $E(\boldsymbol{D}_i)$ and $RT(\boldsymbol{D}_i)$ of 8 monitoring wells

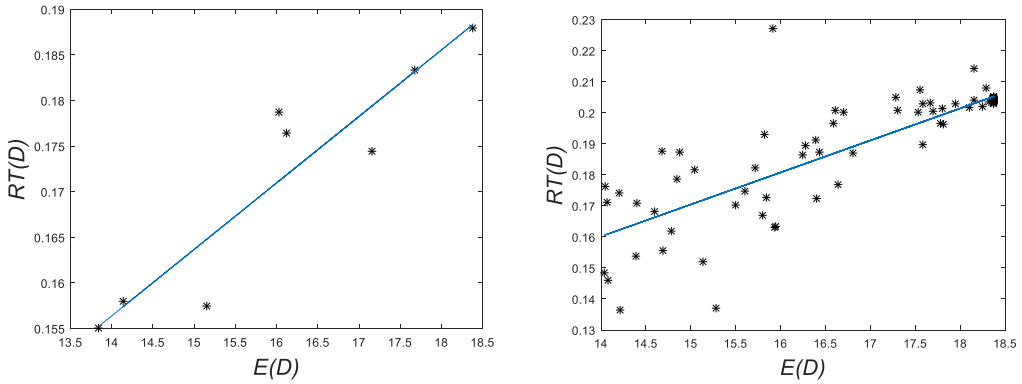
i	1	2	3	4	5	6	7	8
$\boldsymbol{D}_i$	(445,200)	(657,248)	(470,184)	(590,191)	(550,200)	(420,190)	(460,195)	(406,180)
$U(\boldsymbol{D}_i)$	-4.5369	$-4.92*10^{-3}$	-2.3546	-0.7079	-2.2596	-3.2233	-4.2440	-1.2195
$E(\boldsymbol{D}_i)$	13.8430	18.3749	16.0253	17.6720	16.1203	15.1556	14.1359	17.1604
$RT(\boldsymbol{D}_i)$	0.1550	0.1880	0.1787	0.1833	0.1764	0.1575	0.1580	0.1744

Table 4: 20 sets of real parameters obtained from the prior distribution

Number	Contamination source intensity M (g)	Contamination source position X_0 (m)	Contamination source position Y_0 (m)	Contamination release time T_0 (days)
1	1212.8	306.5	200.1	51.8
2	1252.7	275.5	205.9	47.5
3	1404.8	287.8	186.5	43.7
4	1098.9	268.0	190.7	59.4
5	967.7	327.1	196.1	55.1
6	1320.6	254.1	190.0	45.7
7	1434.7	363.0	187.8	52.6
8	1555.4	298.9	202.2	53.0
9	1334.7	358.1	211.9	54.1
10	949.5	208.0	192.8	40.8
11	1084.4	246.7	193.5	42.8
12	1366.7	227.7	197.1	46.5
13	876.5	350.5	191.0	49.9
14	1022.1	367.5	203.9	57.7
15	1598.9	264.5	194.1	44.4
16	1132.3	307.9	200.0	48.1
17	1470.9	344.8	201.9	57.3
18	1458.0	323.5	214.2	50.6
19	916.9	251.5	204.7	47.2
20	1154.5	374.5	188.6	49.1

$RT(\boldsymbol{D}_i)$ and $E(\boldsymbol{D}_i)$ in Tab. 3 are plotted on Fig. 3(a). It can be seen that they approximately fit linearly based on the scattered plot on Fig. 3(a). A good positive linear relationship $RT(\boldsymbol{D})=0.0073E(\boldsymbol{D})+0.0543$ can be found with a correlation coefficient of 0.93. Due to the influence of calculation errors unavoidably generated by the Monte Carlo method and the Markov chain Monte Carlo method when $E(\boldsymbol{D})$ and M_{D_i} were solved, the relationship between $RT(\boldsymbol{D}_i)$ and $E(\boldsymbol{D}_i)$ by the 8 monitoring wells can be not sufficiently established as a linear fit. Therefore, as an example with a set of 100 monitoring wells were selected randomly within the prior distribution of parameter $\boldsymbol{\alpha}$ to calculate the $E(\boldsymbol{D})$ and $RT(\boldsymbol{D})$. A good positive linear relationship of $RT(\boldsymbol{D})=0.01032E(\boldsymbol{D})+0.0156$ was established with a correlation coefficient of 0.85, which suggests that a smaller $E(\boldsymbol{D})$ and a better inversion result for parameter $\boldsymbol{\alpha}$. So,

the information entropy $E(\boldsymbol{D})$ could be a good indicator for a monitoring well design optimization.



(a) 8 monitoring wells (b) 100 monitoring wells

Figure 3: The fitting diagram of the relationship between $E(\boldsymbol{D})$ and $RT(D)$

3.3 Monitoring frequency optimization

Assuming that monitoring well position has been determined, with $\boldsymbol{D}^* = (445, 200)$,which is the best estimated monitoring position. The monitoring count is set as 5 times, and the first monitoring time is set as $t = 70$. The monitoring interval (Δt) is defined as a time interval between 2 monitoring time with only an integer value. The information entropy of posterior distribution of parameter $\boldsymbol{\alpha}$ under a monitoring interval can be expressed as $E(\boldsymbol{D}^*, \Delta t)$, and therefore $E(\boldsymbol{D}^*, \Delta t) = -\ln p(\boldsymbol{\alpha}) + U(\boldsymbol{D}^*, \Delta t)$. The different monitoring intervals are expressed as $RT(\boldsymbol{D}^*, \Delta t)$. The calculation methods of $E(\boldsymbol{D}^*, \Delta t)$ and $RT(\boldsymbol{D}^*, \Delta t)$ are the same as that of $E(\boldsymbol{D})$ and $RT(D)$ in Section 3.2. The calculation results are shown in Tab. 5. It can be seen from Tab. 5 that both $U(\boldsymbol{D}^*, \Delta t_i)$ and $RT(\boldsymbol{D}^*, \Delta t)$ reach minimum values when $\Delta t = 7$. The optimal monitoring frequency is obtained to be 7 at the optimized monitoring well position $\boldsymbol{D}^* = (445, 200)$.

Table 5: Δt_i 、 $U(\boldsymbol{D}^*, \Delta t_i)$ 、 $E(\boldsymbol{D}^*, \Delta t_i)$ and $RT(\boldsymbol{D}^*, \Delta t_i)$ of 12 monitoring cycles

Δt_i	1	2	3	4	5	6
$U(\boldsymbol{D}^*, \Delta t_i)$	-2.1623	-3.0180	-3.8329	-4.5369	-5.1250	-5.4204
$E(\boldsymbol{D}^*, \Delta t_i)$	16.2176	15.3619	14.5470	13.8430	13.2549	12.9595
$RT(\boldsymbol{D}^*, \Delta t_i)$	0.1829	0.1726	0.1641	0.1550	0.1570	0.1552
Δt_i	7	8	9	10	11	12
$U(\boldsymbol{D}^*, \Delta t_i)$	-5.4888	-5.3106	-4.9502	-4.5897	-4.2665	-3.9655
$E(\boldsymbol{D}^*, \Delta t_i)$	12.8911	13.0639	13.4297	13.7902	14.1134	14.4144
$RT(\boldsymbol{D}^*, \Delta t_i)$	0.1548	0.1561	0.1627	0.1690	0.1617	0.1651

The scatter plot for $RT(\boldsymbol{D}^*, \Delta t_i)$ and $E(\boldsymbol{D}^*, \Delta t_i)$ in Tab. 5 is presented on Fig. 4. A good linear relationship of $RT(\boldsymbol{D}^*, \Delta t)=0.0076E(\boldsymbol{D}^*, \Delta t)+0.0563$ is found with a correlation coefficient of 0.91. This demonstrates further that the information entropy of the posterior distribution of parameters is an important parameter for the optimal design of monitoring wells.

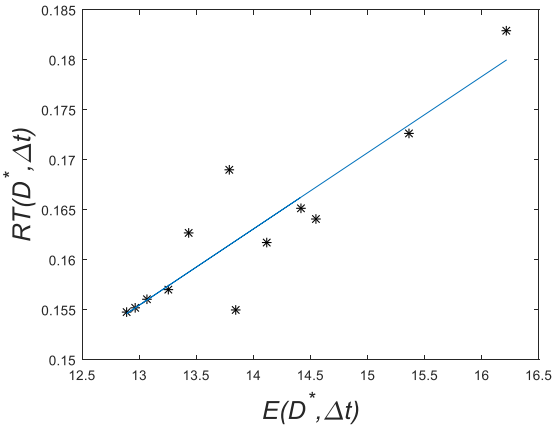


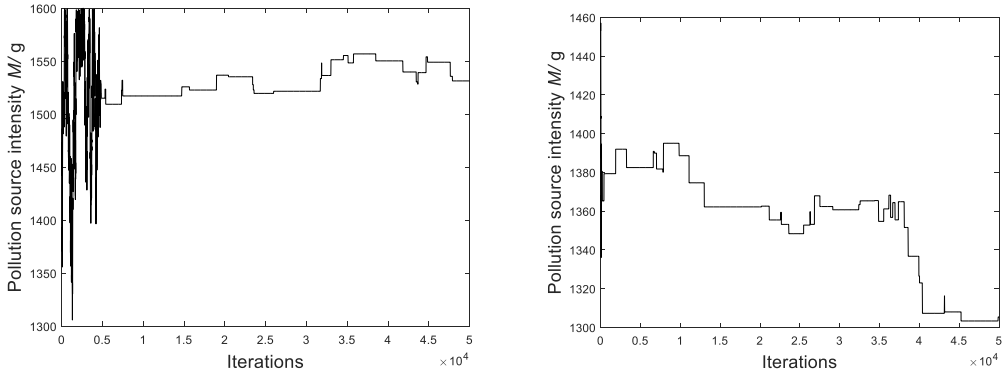
Figure 4: The fitting diagram of the relationship between $E(\boldsymbol{D}^*, \Delta t)$ and $RT(\boldsymbol{D}^*, \Delta t)$

3.4 Contamination source identification based on the optimized monitoring plan

From Sections 3.2 and 3.3, we know that the optimal monitoring well plan for the example in this paper is that monitoring well position $\boldsymbol{D}^* = (445, 200)$, and monitoring frequency is 7.

3.4.1 Solution results by the metropolis algorithm

In this section, the model parameters are solved by adopting the classic Metropolis algorithm to build the Markov chains. However, the iterative curves of the 4 parameters M, X_0, Y_0, T_0 to be estimated vary widely because of the different initial sample points by the Markov chains. Taking M as an example, different initial sample, such as an initial sample Set 1 of $(M, X_0, Y_0, T_0) = (1500.1, 263.6, 193.2, 53.5)$ and a sample Set 2 of $(M, X_0, Y_0, T_0) = (1449.4, 368.7, 212.5, 58.9)$ are randomly selected, the iterative curves are produced as shown on Fig. 5(a) and Fig. 5(b).



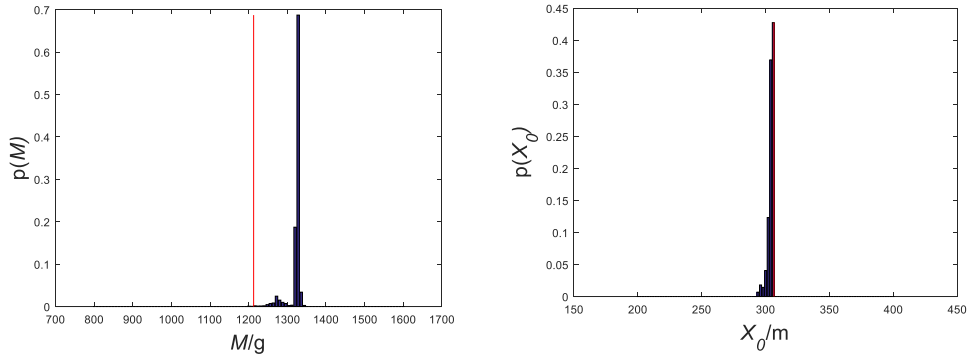
(a) Initial sample Set 1 (b)Initial sample Set 2

Figure 5: Iterative curves of model parameter M based on the Metropolis algorithm with different initial samples

It can be seen from Fig. 5(a) that when the initial sample point is M_1=1500.1 g and the Metropolis algorithm is iterated to 5000 times, the parameter M value gradually becomes stable, and eventually converges to $M \approx 1520g$. However, the value is quite different from the true value (M=1212.8 g), with an error of 25.3%. Fig. 5(b) shows that when the initial sample point is M_1=1449.4 g and the iteration reache 50,000 times, the parameter M value still does not converge. Therefore, when the classical Metropolis algorithm is used to solve the contamination source intensity, the solution result is greatly affected by the initial sample point. This suggests that it is prone to local optimum or difficult to converge. What's more, it is speculated that similar problems may occur to another 3 parameters X_0, Y_0, T_0 .

3.4.2 Solution results by the improved multi-chain metropolis algorithm

Giving that the solution results of the Metropolis algorithm are affected by the initial sample points, the improved multi-chain Metropolis algorithm is used to solve the same problem. The posterior probability histograms of the contamination source parameters are drawn by the improved multi-chain Metropolis algorithm, as shown on Fig. 6 (where the red line shows the true value).



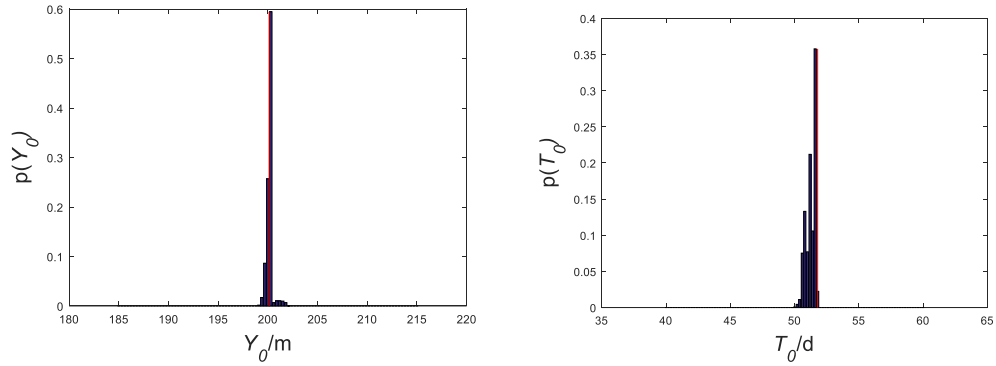


Figure 6: Posterior probability histograms of model parameters based on improved Metropolis algorithm

In the inversion process, each Markov chain has a length of 50,000 for a total of 40 chains. When the evolutionary generations of the parallel Markov chains reach 40,000, the convergence judgment index of 4 parameters are as $\hat{R}_i \approx 1.0 < 1.2\ (i = 1,2,3,4)$. At this time, the Markov chains of all parameters have converged. Then the previous unstable 40,000 Markov chains results are excluded, so that only the last 10,000 stable results are used to perform the posterior statistical analysis. The results are shown in Tab. 6. From Tab. 6, it can be seen that the Metropolis algorithm based on the Latin hypercube sampling method can achieve overall optimization in inversion results and can effectively improve the accuracy of inversion results.

Table 6: Posterior statistical results of model parameters based on improved Metropolis algorithm

Project		Mean	Mean error %	Median value	The median error %
Contamination source intensity M /g		1324.28	9.20	1324.61	9.22
Contamination source coordinates/m	X_0 /m	305.70	0.25	305.83	0.21
	Y_0 /m	200.14	0.0061	200.14	0.0066
Leakage time from first monitoring time T_0 /d		51.62	0.33	51.65	0.27

3.4.3 Influence of prior distribution range on the parameter inversion results

Taking the parameter M as an example, 10 sets of parameters prior distribution ranges (as shown in Tab. 7) are selected to verify the influence of the prior distribution range on the parameter inversion results. All of the observations are observed through the optimal observation scheme. $U(\boldsymbol{D}^*|\Phi)$, $E(\boldsymbol{D}^*|\Phi)$ and $RT(\boldsymbol{D}^*|\Phi)$ with different prior distribution ranges are calculated in the process as in Section 3.2. The values are shown in Tab. 8, and the linear fit between $RT(\boldsymbol{D}^*|\Phi)$ and $E(\boldsymbol{D}^*|\Phi)$ is shown on Fig. 7.

Table 7: 10 sets of parameters prior distribution ranges

Number	Contamination source intensity M (g)	Contamination source coordinates		Leakage time from first monitoring time T_0 (days)
		X_0 (m)	Y_0 (m)	
Φ1	[800,1600]	[200,400]	[185,215]	[40,60]
Φ2	[900,1600]	[200,400]	[185,215]	[40,60]
Φ3	[1000,1600]	[200,400]	[185,215]	[40,60]
Φ4	[1100,1600]	[200,400]	[185,215]	[40,60]
Φ5	[1200,1600]	[200,400]	[185,215]	[40,60]
Φ6	[1300,1600]	[200,400]	[185,215]	[40,60]
Φ7	[1400,1600]	[200,400]	[185,215]	[40,60]
Φ8	[1500,1600]	[200,400]	[185,215]	[40,60]
Φ9	[700,1600]	[200,400]	[185,215]	[40,60]
Φ10	[600,1600]	[200,400]	[185,215]	[40,60]

Table 8: $U(\boldsymbol{D}^*|\Phi), E(\boldsymbol{D}^*|\Phi)$ and $RT(\boldsymbol{D}^*|\Phi)$ with different parameters prior distribution ranges

Number	$\boldsymbol{D}^*$	$U(\boldsymbol{D}^*\|\Phi)$	$-\ln p(\boldsymbol{\alpha})$	$E(\boldsymbol{D}^*\|\Phi)$	$RT(\boldsymbol{D}^*\|\Phi)$
Φ1	(445,200)	-5.4888	18.3799	12.9151	0.1548
Φ2	(445,200)	-5.5332	18.2463	12.7131	0.1591
Φ3	(445,200)	-5.5596	18.0922	12.5326	0.1431
Φ4	(445,200)	-5.6548	17.9099	12.2551	0.1257
Φ5	(445,200)	-5.6425	17.6867	12.0442	0.1190
Φ6	(445,200)	-5.7471	17.3990	11.6519	0.1074
Φ7	(445,200)	-5.7029	16.9936	11.2907	0.1040
Φ8	(445,200)	-5.7811	16.3004	10.5193	0.0994
Φ9	(445,200)	-5.3915	18.4976	13.1061	0.1645
Φ10	(445,200)	-5.3664	18.6030	13.2366	0.1683

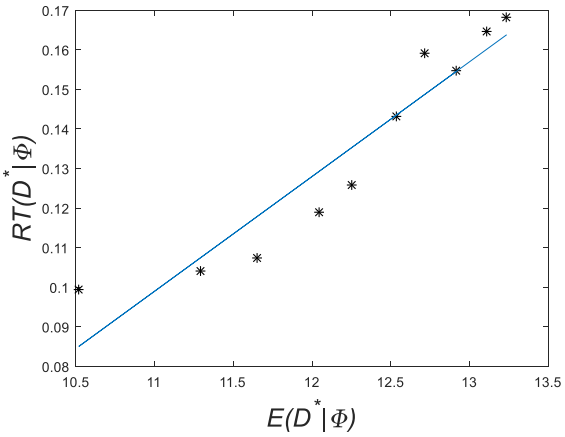


Figure 7: The fitting diagram of the relationship between $RT(\boldsymbol{D}^*|\Phi)$ and $E(\boldsymbol{D}^*|\Phi)$

A good linear relationship $RT(\boldsymbol{D}^*|\Phi)=0.0290E(\boldsymbol{D}^*|\Phi)-0.2198$ can be found with a correlation coefficient of 0.95 (Fig. 7). So, the smaller of $E(\boldsymbol{D}^*|\Phi)$ is, the better the inversion results will be. Based on Eqs. (13) and (14) $\boldsymbol{\alpha}$ obeys uniform distribution; and $E(\boldsymbol{D}^*|\Phi)$ increases with the increase of the prior distribution ranges of $\boldsymbol{\alpha}$; which makes the inversion error increase. Therefore, in order to reduce the inversion error, it is crucial to keep the prior distribution within a smaller range.

In Tab. 6, the parameters X_0, Y_0, T_0 have better inversion results since the mean errors are all less than 0.5%. However, the inversion mean error of M is as high as 9.20%. It is largely because that the prior distribution range of M is larger than that of X_0, Y_0 and T_0 according to the above theory. Therefore, in order to improve the accuracy of the inversion results, the prior distribution ranges of parameters should be minimized through good field investigation or experiment.

4 Conclusion

(a) Using the Bayesian formula to solve the inverse problem of contamination source identification could effectively avoid the loss of "truth values" comparing with that in the deterministic model solution process, which could improve the reliability in identifying the contamination source information (i.e., the contamination source intensity, position and release time).

(b) Both the relative root mean square error of inversion results and the information entropy of posterior distribution of parameters show a good positive linear relationship, suggesting that the information entropy can be seen as a potent indicator for inversion results. The smaller of the information entropy is, the better of the parameter inversion result will be. Therefore, the optimize design method using the Bayesian formula and information entropy can be an effective method to design a groundwater contamination monitoring well plan.

(c) The computational efficiency of the optimal design of groundwater monitoring wells

could be significantly improved by coupling the two-step Monte Carlo method with the differential evolution algorithm.

(d) Utilizing the improved Metropolis algorithm based on Latin hypercube sampling to solve the inverse problem of groundwater contamination source conditions could achieve the overall optimal results with a significantly improved accuracy. And its reliability and stability are better than those of the classic Metropolis algorithm.

Acknowledgment: This work was supported by Major Science and Technology Program for Water Pollution Control and Treatment (No. 2015ZX07406005), Also thanks to the National Natural Science Foundation of China (No. 41430643 and No. 51774270) and the National Key Research & Development Plan (No. 2016YFC0501109).

References

Bahrami, S.; Ardejani, F. D.; Baafi, E. (2016): Application of artificial neural network coupled with genetic algorithm and simulated annealing to solve groundwater inflow problem to an advancing open pit mine. *Journal of Hydrology*, vol. 526, pp. 471-484.

Berger, J. O. (1995): Recent development and applications of Bayesian analysis. *Proceedings 50th ISI, Book I*.

Carrera, J.; Neuman, S. P. (1986): Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resources Research*, 1986, vol. 22, no.2, pp. 211-227.

Chen, M. J.; Izady A.; Abdalla, O. A.; Amerjeed, M. (2018): A surrogate-based sensitivity quantification and Bayesian inversion of a regional groundwater flow model. *Journal of Hydrology*, vol. 557, pp. 826-837.

Czanner, G.; Sarma, S. V.; Eden, U. T.; Brown, E. N. (2008): A signal-to-noise ratio estimator for generalized linear model systems. *Lecture Notes in Engineering & Computer Science*, vol. 2171, no. 1.

Dai, Y. B. (2011): *Uncertainty Analysis of Vehicle Accident Reconstruction Results Based on Latin Hypercube Sampling (Ph.D. Thesis)*. Changsha University of Science & Technology, China.

Dougherty, D. E.; Marryott, R. A. (1991): Optimal groundwater management: 1. Simulated annealing. *Water Resources Research*, vol. 27, no. 10, pp. 2493-2508.

Gelman, A.; Rubin, D. B. (1992): Inference from iterative simulation using multiple sequences. *Statistical Science*, no. 7, pp. 457-472.

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. *Annals of Nuclear Energy*, vol. 29, no. 8, pp. 967-981.

Haario, H.; Saksman, E.; Tamminen, J. (2001): An adaptive Metropolis algorithm. *Bernoulli*, vol. 7, no. 2, pp. 223-242.

Bozorg-Haddad, O.; Athari, E.; Fallah-Mehdipour, E.; Loaiciga, H. A. (2018): Real-time water allocation policies calculated with bankruptcy games and genetic programing. *Water Science and Techonology: Water Supply*, vol. 18, no. 2, pp. 430-449.

Hastings, W. K. (1970): Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, vol. 57, no. 1, pp. 97-109.

Huan, X.; Marzouk, Y. M. (2013): Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, vol. 232, no. 1, pp. 288-317.

Lin, Y. Z.; Le, E. B.; O'Malley, D.; Vesselinov, V. V.; Tan, B. T. (2017): Large-scale inverse model analyses employing fast randomized data reduction. *Water Resources Research*, vol. 53, no. 8, pp. 6784-6801.

Lindley, D. V. (1956): On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, vol. 27, no. 4, pp. 986-1005.

Mahinthakumar, G. K.; Sayeed, M. (2005): Hybrid genetic algorithm-local search methods for solving groundwater source identification inverse problems. *Journal of Water Resources Planning and Management*, vol. 131, no. 1, pp. 45-57.

Marryott, R. A.; Dougherty, D. E.; Stollar, R. L. (1993): Optimal groundwarter management: 2. Application of simulated annealing to a field-scale contamination site. *Water Resources Research*, vol. 29, no. 4, pp. 847-860.

Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller A. H; Teller E. (1953): Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092.

Ramli, M. A. M.; Bouchekara, H. R. E. H.; Alghamdi, A. S. (2018): Optimal sizing of PV/wind/diesel hybrid microgrid system using multi-objective self-adaptive differential evolution algorithm. *Renewable Energy*, vol. 121, pp. 400-411.

Roberts, C. P.; Casella, G. (2004): *Monte Carlo Statistical Methods (Second Edition)*. Springer-Verlag, USA.

Ruzek, B.; Kvasnicka, M. (2001): Differential evolution algorithm in the earthquake hypocenter location. *Pure and Applied Geophysics*, vol. 158, no. 4, pp. 667-693.

Sebastiani, P.; Wynn, H. P. (2000): Maximum entropy sampling and optimal Bayesian experimental design. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 62, pp. 145-157.

Shannon, C. E. (1948): A mathematical theory of communication. *Bell System Technical Journal*, vol. 27, no. 4, pp. 379-423.

Snodgrass, M. F.; Kitanidis, P. K. (1997): A geostatistical approach to contaminant source identification. *Water Resources Research*, vol. 33, no. 4, pp. 537-546.

Sohn, M.D.; Small, M. J.; Pantazidou, M. (2000): Reducing uncertainty in site characterization using bayes monte carlo methods. *Journal of Environmental Engineering*, vol. 126, no. 10, pp. 893-902.

Tanner, M. A. (1996): *Tools for Statistical Inference: Methods for the Expectation of Posterior Distribution and Likelihood Functions*. Springer-Verlag, USA.

Xu, W. T.; Jiang, C.; Yan, L.; Li, L. Q.; Liu, S. N. (2018): An adaptive metropolis-hastings optimization algorithm of Bayesian estimation in non-stationary flood frequency analysis. *Water Resources Management*, vol. 32, no. 4, pp. 1343-1366.

Zeng, L. Z.; Shi, L. S.; Zhang, D. X.; Wu, L. S. (2012): A sparse grid based Bayesian method for contaminant source identification. *Advances in Water Resources*, vol. 37, pp. 1-9.

Zhang, J. J. (2017): *Bayesian Monitoring Design and Parameter Inversion for Groundwater Contaminant Source Identification (Ph.D. Thesis).* Zhejiang University, China.

Zhang, J. J.; Zeng, L. S.; Chen C.; Chen, D. J.; Wu, L. S. (2015): Efficient Bayesian experimental design for contaminant source identification. *Water Resources Research*, vol. 51, no. 1, pp. 576-598.

Zhang, J. J.; Li, W. X.; Zeng, L. Z.; Wu, L. S. (2016): An adaptive Gaussian process-based method for efficient Bayesian experimental design in groundwater contaminant source identification problems. *Water Resources Research*, vol. 52, no. 8, pp. 5791-5984.

Zhao, G. (2007): *Differential Evolution Algorithm with Greedy Strategy and Its Applications (Ph.D. Thesis)*. Harbin Institute of Technology, China.

Zheng, C. M.; Gordon, D. (2009): *Applied Contaminant Transport Modeling (Second Edition)*. Higher Education Press, China.