ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230620000748
Original scientific paper

Traceability of River Water Pollution Based on MFO and M-H Algorithms

Dongyan JIA*, Jinling SONG, Lisha DONG, Yan KANG, Xiaoning ZENG

Abstract: The work proposed a novel model to accurately trace the pollution sources of water pollution incidents based on moth-flame optimization and Metropolis-Hastings sampling algorithms. The model first utilized moth-flame optimization to estimate the parameters of the pollutant migration-diffusion model by minimizing the error between monitored and predicted concentration. It then traced the optimal pollution source location, discharge volume, and time using the M-H sampling algorithm. Simulation experiments demonstrated the model achieved significantly lower errors in tracing pollution source information compared to a previous method, with relative errors within 1.33%. The new model provides an accurate and efficient approach to tracing water pollution incidents and overcomes the limitations of previous methods. It exhibits substantial potential in identifying pollution sources within real-world aquatic environments as well as facilitating prompt responses to mitigate environmental and health impacts.

Keywords: M-H sampling; migration-diffusion model; moth-flame optimization algorithm; parameter calibration; traceability; water pollution

1 INTRODUCTION

Water pollution incidents devastate the aquatic ecological environment and lead to significant losses and even social upheaval; thus, emergency treatment for water pollution [1-7] has become imperative. The promulgation of the *Action Plan for the Prevention and Control of Water Pollution* (the *Water Ten Articles*) has put forward higher requirements for the rapid handling of such incidents. Determining the pollution-source location and the discharge amount and time of pollutant immediately is the key to prompt the response to the water pollution incident. Quickly identifying pollution-source information is significant in minimizing the impact of water pollution incidents.

Research on water pollution incidents focuses on spatio temporal dynamic changes in water quality [8, 9], pollution causes [10], pollutant sources [11-13], and the traceability of water pollution sources [14-19]. Accidental pollutants will convene and spread in the river as water pollution incidents happen. Therefore, it will change the water quality of the downstream river section from the location of the pollution source. Revealing the origins of water pollution incidents involves discerning the source and magnitude of pollutants in watersheds based on monitored water-quality data.This problem is also known as water pollution traceability [20].

Current research on water pollution traceability is generally divided into two categories: the mechanism-based pollution-source traceability method and the non-mechanism-based pollution-source traceability method. The former depends on the diffusion model of river pollutants and uses the direct method [21], optimization method [22-24], and probabilistic statistical method [25, 26]. The method can solve issues such as pollution- -source location and pollutant discharge, but it is usually limited by the scope of monitoring. The standard genetic algorithm and mathematical analysis are used to construct the inversion model of pollution-source information [22]. The optimal value of pollution source information is searched using an improved genetic algorithm after roughly estimating the emission location and emission time of the pollution source [23]. The homology algorithm and genetic algorithm are combined for the inversion of pollution source information [24]. All these three methods use the parameters of the migration-diffusion model that are not rate-determined, which leads to a certain bias in the final traceability results. The backward-localization probability-density function is used to identify the discharge location of pollution sources in surface water [25]. However, the method is easily affected by the irregularity of the river, the diffusion coefficient of pollutants, and the velocity of water flow. The transverse water-flow velocity and transverse-diffusion coefficient are the sensitive parameters of the pollutant-migration diffusion model, showing certain uncertainties [26]. A pollution-source information-identification framework considering parameter uncertainty is proposed combining Bayesian reasoning and Markov chain Monte Carlo method. However, the computational-time complexity of the model increases exponentially as the number of unknown parameters increases in the posterior probability density function for identifying pollution sources. The non-mechanism-based pollution-source traceability method no longer considers the diffusion model of pollutants but seeks pollution sources from its links with related objects. The cross-correlation analysis method is used to draw a chart of cross-correlation water quality indices [27]. Association rules (Apriori) are used to track pollutants in common industries in the research area to assess the complex relationship between pollutant point sources and water quality monitoring data. On this basis, a comprehensive short- and long-term memory network (LSTM) is used to construct an identification model of industrial point sources of pollutants. The knowledge map construction technology is used to integrate the knowledge content and structure in water pollution [28]. A knowledge map of water-pollution traceability is constructed for traceability research on water pollution incidents in inland rivers. The accuracy of the pollution-source traceability results of these two methods depends entirely on the level of detail and accuracy of the relationship network between pollutant entities. Moreover, the workload is large at the early stage and realization poses a greater challenge. Although current research on tracing water pollution incidents has its characteristics, the limitations of each method have resulted in unsatisfactory traceability effects. Therefore, further research is needed to improve traceability methods for water pollution.

The mechanism-based pollution-source tracking method relies on the pollutant migration diffusion model. It infers the source of pollution according to measured pollutant values at each monitoring point. The traceability of pollution source information is solved as the "inverse problem" of the water pollution migration-diffusion model. Therefore, the accuracy of predicting pollution-source information depends on the precision of the pollutant migration-diffusion model and the "inverse problem" modeling method for water pollution. The accuracy of the pollutant migration and diffusion model depends on the accuracy of the parameters. Only if the model parameters are consistent with the hydrological environment at the time can the accuracy of the calculation results of the pollutant migration and diffusion model be guaranteed. The calibration of model parameters aims to enhance the accuracy of these parameters. The parameter calibration method is currently the most commonly used optimization algorithm. Compared with other algorithms, the moth flame optimization algorithm has strong parallel optimization ability, excellent global performance, and a propensity to avoid local extremes. Based on this algorithm, the parameter calibration can be improved. Besides, the "inverse problem" modeling process of the water pollution migration and diffusion model is easily affected by uncertain factors such as water quality monitoring data, model parameters, and water pollution event information [29]. Bayesian estimation can deal with the uncertainty well. Therefore, the modeling method based on Bayesian estimation can improve the inversion accuracy of water-pollution-source parameters. A water pollution traceability model is proposed for water pollution incidents of point sources emitted instantaneously. This model is based on a moth-flame optimization algorithm and M-H (Metropolis-Hastings) sampling method to trace the source information of water pollution more precisely. Fig. 1 shows the technical roadmap of the SI-RSWP model.

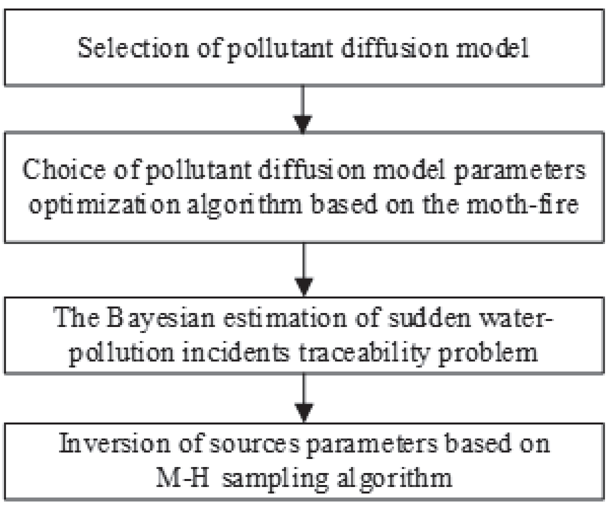


Figure 1 Technical roadmap of the SI-RSWP model

The work mainly includes the following three contents according to the technical roadmap of the SI-RSWP model.

(1) A parameter calibration method was proposed for the pollutant migration-diffusion model parameters based on moth-flame optimization algorithm. The parameters of the pollutant migration-diffusion model were refined through optimization to closely align with the actual hydrological conditions of the river. The accuracy of the model was improved.

(2) The traceability of river water pollution was transformed into a Bayesian estimation problem to construct the likelihood function. The pollutant discharge, pollution-source location, and sewage discharge time were sampled and tested according to the M-H algorithm in the Markov Monte Carlo method. The optimal estimated values of each parameter were finally determined.

(3) An experiment was designed to track the point-source pollution source discharged instantaneously on shore. The proposed inversion model and the algorithm setting model parameter calibration were validated by comparing experimental results with actual values. The traceability effects of related methods were also analyzed.

2 PARAMETER CALIBRATION OF THE POLLUTANT MIGRATION-DIFFUSION MODEL BASED ON THE MOTH-FLAME OPTIMIZATION ALGORITHM

The performance of the pollutant migration-diffusion model heavily depends on its parameters, such as the flow rate and the river diffusion coefficient. The hydrological data of the rivers in practice change with the season and weather, but they cannot be measured as frequently as the water-quality monitoring. If the used model parameters do not match the hydrological environment at the time, the inversion result will deviate greatly from the real result. Therefore, the parameter calibration of the model reduces the actual survey work and enhances the generalization and precision of the pollutant migration-diffusion model. The rate is generally determined by an optimization algorithm. The moth-flame optimization algorithm was used to optimize the parameters of the pollutant migration-diffusion model in the experiment.

2.1 Moth-Flame Optimization Algorithm

The moth-flame optimization algorithm (MFO) is a new type of population-based intelligent optimization algorithm [30]. It simulates the flight-positioning mechanism of moths that fly in a spiral around the flame and eventually converge on the flame. The moth is a candidate solution to the optimization problem in the MFO algorithm. The variable is the spatial position of the moth, and the moth's spatial position and corresponding fitness values are stored in Eqs. (1) and (2).

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,d} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,d} \\ \vdots & \vdots & \cdots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,d} \end{bmatrix} \tag{1}$$

$$OM = \left[OM_1 \quad OM_2 \quad \cdots \quad OM_n \right]^T \tag{2}$$

where n represents the population size; d represents the dimension of the optimization problem.

Since each moth updates its position based on only one flame, Eq. (3) shows that the flame-position matrix is akin to the moth matrix; the flame fitness value is stored in the matrix in Eq. (4).

$$F = \begin{bmatrix} F_{1,1} & F_{1,2} & \cdots & F_{1,d} \\ F_{2,1} & F_{2,2} & \cdots & F_{2,d} \\ \vdots & \vdots & \cdots & \vdots \\ F_{n,1} & F_{n,2} & \cdots & F_{n,d} \end{bmatrix} \tag{3}$$

$$OF = \begin{bmatrix} OF_1 & OF_2 & \cdots & OF_n \end{bmatrix}^T \tag{4}$$

The moth is the actual search subject in the MFO algorithm, and it moves in the search space. The flame is the optimal position obtained by the moth until the current iteration. The individual moth iteratively updates its position around the flame until the optimal solution is found. The mathematical description is divided into a flame-catching behavior and a flame abandonment behavior.

(1) Flame-catching behavior. Moths with phototaxis M_i will move towards the nearest bright light (flame) F_j in nature. The logarithmic spiral in Eq. (5) is the movement trajectory of the moth catching flame.

$$M_i = S\left(M_i, F_j\right) \tag{5}$$

where M_i is the position of the i^{th} moth; F_j is the position of the j^{th} flame; S is the spiral function. Eq. (6) shows the expression.

$$S\left(M_i, F_j\right) = D_i \cdot e^{bt} \cdot \cos\left(2\pi \cdot t\right) + F_j \tag{6}$$

where $D_i = |F_i - M_j|$ represents the distance between the i^{th} moth M_i and the j^{th} flame F_j; b is the spiral shape constant; t is a random number between [−1, 1]. $t = -1$ means the closest to the flame; $t = 1$ means the farthest from the flame.

(2) Flame abandonment behavior. The flames are sorted according to the fitness value after the flame list is updated, and then the moths update their position relative to the corresponding flame. That is, the first moth updates its position relative to the optimal flame, and the last moths use the worst flame to update their position. This can avoid the moths moving towards only one flame and causing the algorithm to fall into local optimality.

Eq. (7) is used to adaptively reduce the number of flames during the iteration process. It balances global exploration and local exploitation and ensures the convergence speed of the algorithm.

$$flame_{nu} = \left\{ N - t \cdot \frac{N-1}{T} \right\} \tag{7}$$

where t represents the current number of iterations; N represents the maximum number of flames; T represents the maximum number of iterations; { } represents rounding.

The fitness value is calculated in the MFO algorithm according to the objective function of the optimization problem. Eq. (8) shows that the objective function is generally expressed as a function related to the parameters to be optimized.

$$f\left(X\right) = f\left(X_1, X_2, \cdots X_d\right) \tag{8}$$

where (X_1, X_2, ..., X_d) represents the parameters to be optimized.

The steps of the MFO algorithm are as follows:

(1) Initialize the algorithm parameters: the number of moths (n), the dimension (d), and the maximum number of iterations (T). Randomly initialize the population.

(2) Calculate the fitness value of the moths in the population according to Eq. (8), and sort them in ascending order by the fitness value.

(3) Use Eq. (7) to update the number of flames.

(4) Use Eq. (6) to update the position of the moths.

(5) If the given number of iterations is desired, the algorithm ends and the optimal solution is obtained; otherwise, return to step (2).

2.2 Pollutant Migration-Diffusion Model

The diffusion of pollutants in rivers is a complex 3D model problem. However, the gradient of pollutant concentration is negligible in vertical and horizontal directions if the river's length exceeds its depth and width. The complex 3D diffusion problem can be simplified to a 1D migration-diffusion problem in an unbounded space of an instantaneous point source in this case. The river in the experiment has an elongated shape, so a 1D migration-diffusion model is used as the pollutant migration-diffusion model. Eq. (9) shows the 1D migration-diffusion model.

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = D\frac{\partial^2 C}{\partial x^2} + KC \tag{9}$$

where u represents the flow rate; D represents the diffusion coefficient of the river; K represents the attenuation coefficient of pollutants; C_i, in the boundary conditions, represents the concentration of (x_i, y_i) at each entrance and exit of the river.

The analytical solution of the 1D migration-diffusion model was used to predict the pollutant concentration at each point in the experiment. Eq. (10) shows the analytical solution of the 1D migration-diffusion model for the pollution emitted instantaneously in point sources.

$$c\left(x,t\right) = \frac{M}{\sqrt{4\pi Dt}} \exp\left[-\frac{\left(x-ut\right)^2}{4Dt} \right] \tag{10}$$

where $c(x, t)$ represents the pollutant concentration of the river at the time t and the section x; M represents the total pollutant mass of the pollution source; D represents the diffusion coefficient of the river; u represents the water flow speed.

2.3 Parameter-Calibration Algorithm of the Pollutant Migration-Diffusion Model Based on MFO

The fitness value is calculated based on the objective function, according to the principle of the MFO algorithm. The MFO's objective function is determined according to the pollutant prediction requirements while the MFO is used to rate the parameters of the pollutant migration-diffusion model. Then the MFO algorithm is used to optimize the value of the model parameters.

The model parameters are optimized to minimize the difference between predicted and observed monitoring data by finding an optimal set of parameters from a given space. Therefore, Eq. (11) shows that the sum of the two error squares can be used to construct the objective function.

$$\min f\left(y, y^{*}, \theta\right)=\min \sum_{i=1}^{n} \sum_{j=1}^{m}\left(y_{ij}-y_{ij}^{*}\right)^{2} \tag{11}$$

where $f(y, y^*, \theta)$ is the objective function; y is the theoretically calculated value of the pollutant concentration by Eq. (10); y^* is the concentration observation value; θ is the parameters to be determined; n is the number of the parameters to be determined; m is the number of the observed values.

A model parameters optimization algorithm is proposed based on MFO pollutant migration-diffusion model parameters (MPO-MFO) and the principle of the moth-flame optimization algorithm and the pollutant migration-diffusion. The algorithm steps are as follows.

(1) Determine the parameters to be determined by the rate according to the 1D migration-diffusion model.

(2) Obtain the pollutant concentration data of the monitoring section and process the data preliminarily. Set the upper and lower limit values of the model parameters based on the measured data and prior knowledge, and then initialize the parameters to be determined.

(3) Use Eq. (11) as the objective function of the moth-flame optimization algorithm. The 1D migration-diffusion model parameters are optimized within the range of values to obtain the optimal values of each parameter.

3 TRACEABILITY OF RIVER WATER POLLUTION INCIDENTS

The traceability of pollution-source information in river water pollution incidents is the "inverse problem" of the water-pollution migration-diffusion model. The "inverse problem" is not qualitative for the uncertainty of the river system [31]. Bayesian estimation deals with uncertainties in the "inverse problem" by expressing all uncertainties as probability distributions based on probability theory. Therefore, the experiment transforms the problem of traceability of river water pollution into a Bayesian estimation problem. The likelihood function is constructed by the prior pollution-source information and the pollutant-monitored data based on the pollutant migration-diffusion model with determined parameter rates. Then the posterior probability density function of the unknown pollutants is derived, and the relevant pollution-source information is obtained by sampling method.

3.1 Posterior Probability Density Function of the Pollution-Source Parameters

The upper and lower limits of the parameters can be obtained through the prior information in a river water pollution event, although the parameters' distribution type is unknown. Therefore, the uniform distribution on the upper and lower limit interval can be used to represent the distribution of unknown parameters. The work studied pollutant emissions, pollution-source location, and pollutant discharge time. It is defined as $X(M, x_0, t_0)$. Eqs. (12) to (14) show the probability density functions corresponding to the prior distribution of the three unknown parameters, according to the prior knowledge.

$$p(M)=\begin{cases}0.0526, & 13<M<32 \\ 0, & \text{other}\end{cases} \tag{12}$$

$$p(x_0)=\begin{cases}0.2857, & 0<x_0<3.5 \\ 0, & \text{other}\end{cases} \tag{13}$$

$$p(t_0)=\begin{cases}1, & 0<t_0<1 \\ 0, & \text{other}\end{cases} \tag{14}$$

An observation error is superimposed on the concentration measurement data to meet the actual situation. It is assumed that the observation error obeys a normal distribution with a standard deviation of σ, and each measurement point is independent of the other. Eq. (15) shows the likelihood function of the pollutant traceability problem.

$$P(Y|X)= \prod_{i=1}^{N} P(Y|X_i) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\sum_{i=1}^{N} \frac{[Y_i - C_i(M,x,t\,|\,X)]^2}{2\sigma^2}\right] \tag{15}$$

where Y_i represents pollutant concentration data measured at the i^{th} measurement point; C_i represents the theoretical value of the forward simulation by utilizing the pollutant migration-diffusion model for the i^{th} measurement point; N represents the number of observed pollutant concentration data.

Eq. (16) converts the traceability problem into a posterior probability density function of river-water-pollution traceability by the Bayesian formula to solve pollution-source unknown parameter X.

$$P(X\,|\,Y)=\frac{P(Y\,|\,X)*P(X)}{P(Y)}= \propto \frac{P(X)}{\left(\sqrt{2\pi}\sigma\right)^N} \exp\left[-\sum_{i=1}^{N} \frac{[Y_i - C_i(M,x,t\,|\,X)]^2}{2\sigma^2}\right] \tag{16}$$

Eq. (17) represents the posterior probability density function, with the number of observed values ($N = 20$) and the hydrological water-quality parameters substituted.

$$P(M,x_0,t_0|Y)= \frac{2.732*10^{-4}}{\left(\sqrt{2\pi}\sigma\right)^{20}} \exp\left[-\sum_{i=1}^{20} \frac{[Y_i - C_i(M,x,t\,|\,X)]^2}{2\sigma^2}\right] \tag{17}$$

3.2 Pollution Source-Parameter Inversion Algorithm Based on the M-H

The pollution-source parameters are sampled by a sampling algorithm based on the density function of traceability parameter ($X(M, x_0, t_0)$) for river water pollution events. The density-function distribution of each parameter is obtained, and the optimal estimate of the pollution-source parameters is determined according to the distribution. The M-H algorithm in the Markov Monte Carlo method was selected to sample the pollution-source parameters in the work. Its core idea was to construct a Markov chain, with target distribution $\pi(x)$ as the invariant distribution. Fig. 2 shows the complete calculation process of the inversion algorithm for pollution source parameters (IA-PSP).

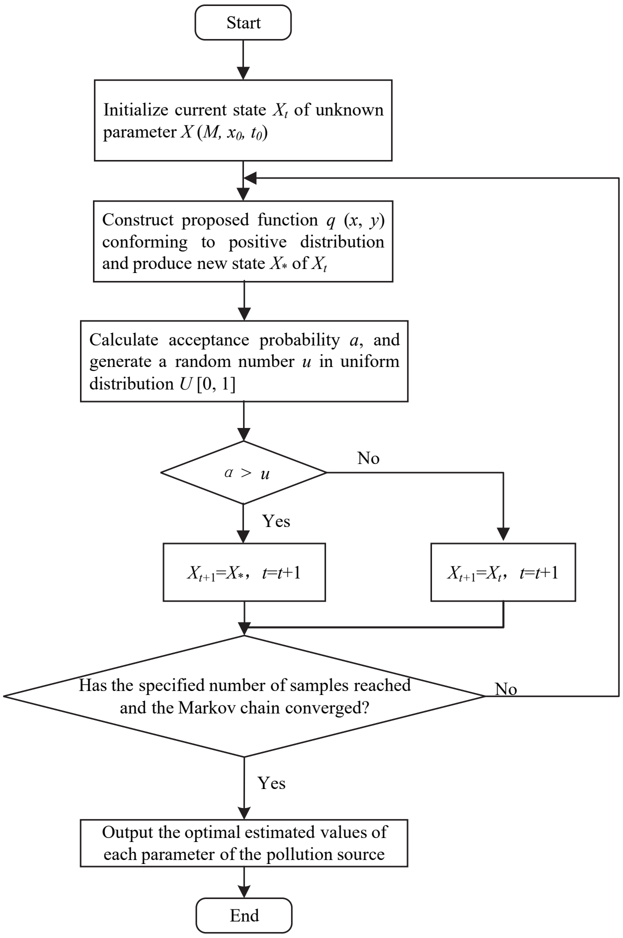


Figure 2 IA-PSP flow

The IA-PSP consists of seven steps, and the details of each step are as follows.

(1) Initialize the algorithm parameters. Assume that the current state of the unknown parameter X is $X_t = x$, with next state $X_{(t+1)}$ and symmetrical proposal function $q(x, y)$.

(2) Use the proposal function to generate new state X_t from current state X^* by walking randomly.

(3) Calculate the acceptance probability of X^*: $\alpha\left(X_t, X_*\right) = \min\left\{1, \frac{\pi\left(X_*\right) q\left(X_*, X_t\right)}{\pi\left(X_t\right) q\left(X_t, X_*\right)}\right\}$, and target distribution area $\pi(x)$ is the posterior probability density function of Eq. (16).

(4) Extract a random number u from the uniform distribution U[0, 1] and compare it with the acceptance probability α. If $\alpha > u$, then $X_{(t+1)} = X^*$; otherwise, $X_{(t+1)} = X_t$.

(5) Repeat steps (2)-(4) until the Markov chain converges or reaches the maximum number of cycles to obtain a sample set of the pollution-source parameters.

(6) The distribution interval with the maximum samples is counted based on the sample-set distribution of each pollution-source parameter. This is the maximum probability interval of each parameter. The maximum probability interval of the pollutant quality is denoted as $MPR\_M = [M_{\min}, M_{\max}]$; the pollutant discharge location is denoted as $MPR\_x_0 = [x_{0\min}, x_{0\max}]$; the pollutant discharge time is denoted as $MPR\_t_0 = [t_{0\min}, t_{0\max}]$.

(7) Select the intermediate value of the maximum probability interval of each parameter. The pollutant-quality BE_M, the pollutant discharge-location BE_x_0, and the pollutant discharge-time BE_t_0 are used as the optimal estimates of each parameter. Return to Step (1) .

4 WATER POLLUTION TRACEABILITY EXPERIMENT

A river section without tributaries was chosen for the simulation experiments to verify the accuracy of the SI-RSWP when the section allows for the 1D constant flow calculations. The algorithm effect was verified by comparing the inversion results with the true values.

An instantaneous point pollution source was constructed at $x = 0$ in the experimental simulation process. The monitoring data of two sections, located 2 and 2.5 KMs away from the pollution source at different times, were used to determine the parameters rate and develop a pollution-traceability model. Each monitoring section was sampled 10 times, with an interval of 6 minutes, and 20 monitoring data sets were obtained.

4.1 Experimental Data

The known hydrological parameters include river diffusion coefficient D and river average flow rate u in the experiment. The parameters to be reversed include pollutant mass M, pollutant discharge-time t_0, and pollutant discharge-location x_0. Pollutant discharge time represents the time difference between when the pollution occurred and when abnormal pollutant concentration was first observed; the pollutant discharge location represents the distance between the pollution source and the first observation point. Tab. 1 shows the values of each parameter.

Table 1 Hydrological parameter values

Parameters	River-diffusion coefficient D / km^2/h	Average river-flow rate u / km/h	Pollutant quality M / kg	Pollutant discharge-time t_0 / h	Pollutant discharge-location x_0 / km
Types	Known	Known	To be reversed	To be reversed	To be reversed
Values	2	5	17.88	0.5	2

4.2 Parameter Calibration and Result Analysis of the Pollutant Migration-Diffusion Model

An upper and lower limit is set on the parameter value in the experiment to reduce model-parameter rate-determination blindness based on the previous cases' initial values. Tab. 2 shows the specific value.

Table 2 Initial value setting of the model parameters

Parameters	Value lower limit	Value upper limit
River-diffusion coefficient D / km^2/h	0.01	8
River flow rate u / km/h	0.01	15

MPO-MFO, the parameter rate setting algorithm of the pollutant migration-diffusion model based on MFO, is used in the experiment to set the model parameter rate. The relevant parameters of different moth-flame optimization algorithms are replaced by running the experimental program multiple times. Final algorithm parameters are as follows after comparative experiments. Moths' number of n = 30, dimension d = 2, and iterations' maximum number T = 100. The algorithm performed 100 repeated operations under the same experimental conditions to verify the reliability of the calculation results. Fig. 3 shows the rate-setting results of the two parameters to be determined.

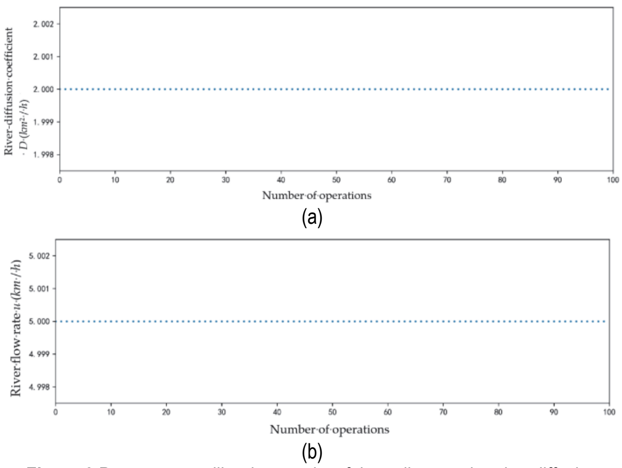


Figure 3 Parameters calibration results of the pollutant migration-diffusion model based on the MFO

Fig. 3 shows that the experimental results are relatively stable in 100 experiments, and the parameter calibration results are very close to the true value.

Table 3 Comparison of calibration results and true values

Parameters	True values	Mean of rate-setting results	Absolute error
River-diffusion coefficient D / km^2/h	2	2.00000047	0.00000047
River flow rate u / km/h	5	4.99999959	0.00000041

Tab. 3 shows the mean and error analysis of the results of these 100 experiments. The model parameter results calculated by the MPO-MFO algorithm are close to the true values. The relative errors of the river diffusion coefficient and water flow velocity compared to their respective mean values are 0.0000235 and 0.0000082%, respectively, which are small and negligible. The introduction of the moth-flame optimization algorithm can enhance the accuracy of parameter calibrations for the pollutant migration-diffusion model based on the MFO.

4.3 Pollution-Source Parameter Inversion and the Result Analysis

About 20000 samples were repeatedly sampled in the traceability process on the IA-PSP pollution source parameter inversion algorithm. Figs. 4 and 5 display the iteration curve and probability distribution histogram of each pollution-source parameter, respectively.

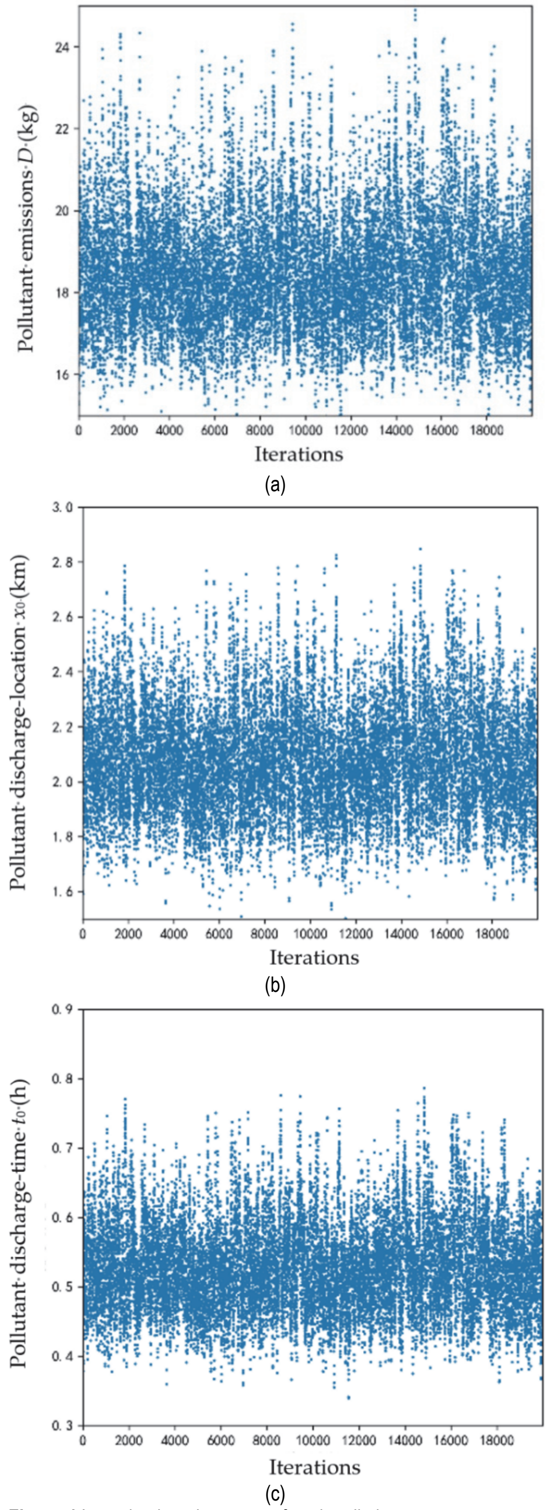
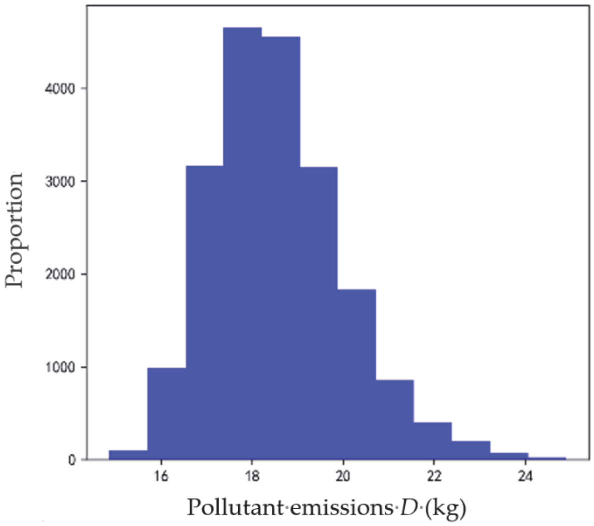
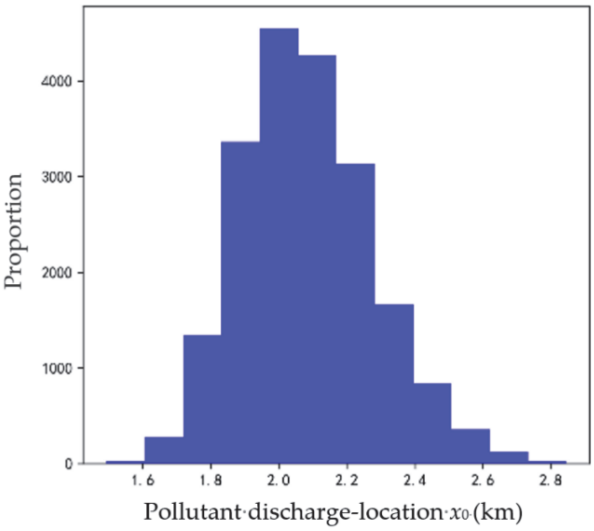


Figure 4 Inversion iteration curve of each pollution-source parameter

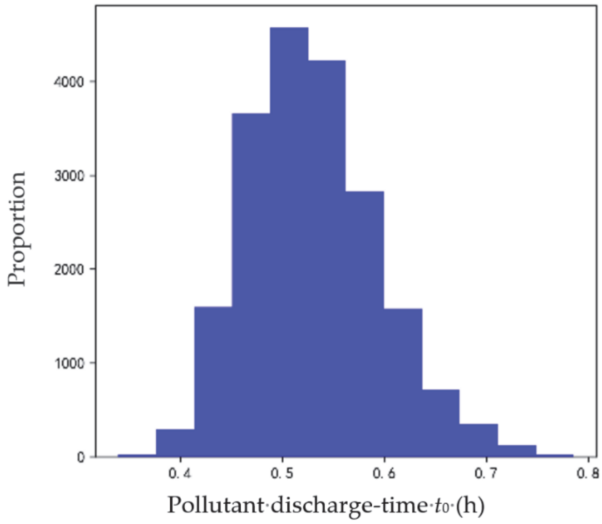
The posterior probability-density distribution of each parameter is relatively concentrated in the results obtained by the IA- PSP pollution-source parameter inversion algorithm (Fig. 4).



(a)



(b)



(c)

Figure5 Histogram of the probability distribution of each pollution-source parameter

Fig. 5a shows that the maximum probability interval of pollutant discharge-quality MPR_M = [17.37493, 18.21118885], the distribution interval with the largest sample size. Internal's intermediate value BE_M = 17.79306 is taken as the optimal estimate of pollutant discharge quality. Fig. 5b indicates that of pollutant discharge-location MPR_x_0 = [1.944384, 2.05718757], with BE_x_0 = 2.0007858 as the optimal estimate. Fig. 5c presents that of the pollutant discharge-time MPR_t_0 = [0.48802858, 0.52525398], with BE_t_0 = 0.5066413 as the optimal estimate.

The pollution-source parameters from the IA-PSP algorithm are the traceability results of the SI-RSWP model for this pollution event, according to its technical route. Therefore, the water pollution incident is likely to occur 0.5066413 hours before the first monitoring at 2.0007858 km from the initial observation point, with a pollutant discharge quality of approximately 17.79306 kg. Tab. 4 shows the comparison between the traceability results and the incident's true values.

Table 4 Comparison of the traceability results and the true values

Parameters	True values	Traceability results	Absolute errors	Relative errors	Average absolute error
Pollutant emissions D / kg	17.88	17.79306	0.08694	0.486%	0.0314557
Pollutant discharge-time t_0 / h	0.5	0.5066413	0.0066413	1.3283%	
Pollutant discharge-location x_0 / km	2	2.0007858	0.0007858	0.0393%	

The absolute error and the relative error between the parameter values and the true values are significantly smaller, and they are obtained by the SI- RSWP traceability model (Tab. 4). The inversion result of the pollutant discharge location is very accurate; the relative error is only 0.0393%. It is highest for the pollutant discharge-time, but still acceptable as the difference from the true value is less than 5 minutes in the traceability result. Besides, the average absolute error obtained by the SI-RSWP traceability model is 0.0314557. Compared with the true pollution-source parameters, the traceability results of the SI-RSWP model are more accurate. The moth-flame optimization algorithm is introduced to calibrate the hydrological parameters of the pollutant migration and diffusion model before the SI-RSWP model inverts water-pollution-source parameters. The consistency of the model parameters and the hydrological environment at that time is guaranteed, thereby enhancing its generalization characteristics. On this basis, the posterior probability density function of unknown pollutant parameters is constructed combined with the a priori information of the pollution source. The M-H sampling algorithm is used to obtain the optimal inversion values of each parameter of pollution sources.

To sum up, it is proved that the SI- RSWP traceability model is accurate and practical for water-pollution incidents.

4.4 Model Comparison

The accuracy of the SI-RSWP model was evaluated by comparing it with the water-pollution traceability method.

The method was based on the BAS algorithm and proposed by Wang et al. [19]. Tab. 5 shows the comparison results.

Table 5 Comparison of the traceability results (absolute error) of the two traceability methods

Parameters	Absolute errors	
	SI-RSWP	Traceability method based on the BAS algorithm
Pollutant emissions D / kg	0.08694	0.452
Pollutant discharge-time t_0 / h	0.0066413	0.1167
Pollutant discharge-location x_0 / km	0.0007858	0.02543

The absolute error in the SI-RSWP model is significantly smaller than in the BAS algorithm-based method, indicating that the SI-RSWP model is more accurate. This is because the SI- RSWP traceability model incorporates a moth-flame optimization algorithm to accurately determine the parameters of the pollutant migration-diffusion model and improve its precision. However, the BAS algorithm fails to rate the parameters because it is greatly affected by the complexity of the pollutant migration-diffusion model. Additionally, the SI-RSWP traceability model utilizes the M-H sampling algorithm to solve pollution-source parameter inversion uncertainty through the probability distribution. The accuracy of the BAS algorithm is significantly impacted by the reduced step size and causes less-precise traceability results.

Compared with the existing research results, the SI-RSWP traceability model calibrates the parameters of the pollutant migration and diffusion model, which improves the accuracy of the model. Besides, the inversion model of pollution source information constructed based on Bayesian estimation can better deal with the uncertainty of the inverse problem. Therefore, the SI-RSWP traceability model exhibits high accuracy and demonstrates reduced susceptibility to uncertainty factors. However, the SI-RSWP traceability model cannot be applied to all types of water pollution. It can perform well for instantaneous single-point pollution sources; the traceability effect is not ideal for other types of water pollution. This is also the direction we will continue to explore in the future.

5 CONCLUSIONS

The work presented a traceability model for river water pollution incidents based on moth-flame optimization and M-H sampling algorithms. The moth-flame algorithm estimated pollution diffusion parameters, while the Metropolis-Hastings sampling traced the optimal source location, discharge volume, and time. Simulation experiments demonstrated the model's ability to rapidly and precisely identify pollution sources, overcoming the limitations of previous techniques. These results highlight the potential of the proposed model to enhance responses to water pollution events by identifying responsible polluters. Future work should evaluate the model on real-world monitoring data and a 2D pollutant migration-diffusion model. Overall, this paper makes valuable contributions toward tracing and controlling water pollution. In summary, a valuable approach is proposed for tracing sudden river pollution incidents, which is demonstrated through simulation experiments.

Acknowledgment

The work was supported by the S&T Program of Hebei (Grant No. 21370103D and 21373301D), the 2023 Scientific Research Project of Higher Education Institutions in Hebei Province (Grant No. ZC2023123), the Open Fund Project of Hebei Innovation Center for Smart Perception and Applied Technology of Agricultural Data (Grant No. ADIC2023Y004), OpenProject of Hebei Key Laboratory of Software Engineering (Grant No. KF2307), andCreative Research Groups Foundation of Hebei Normal University of Science and Technology.

6 REFERENCES

[1] Zhao, W. X. (2022). *Two-dimensional water quality numerical stimulation of the river*. Chongqing: Chongqing University.

[2] Yan, B., Liu, Y., Gao, Z., & Liu, D. (2022). Simulation of sudden water pollution accidents in Hunhe River Basin Upstream of Dahuofang Reservoir. *Water*, *14*(6), 925. https://doi.org/10.3390/w14060925

[3] Setiyono, J. & Natalis, A. (2021). Ecocides as a serious human rights violation: A study on the case of river pollution by the palm oil industry in Indonesia. *International Journal of Sustainable Development and Planning*, *16*(8), 1465-1471. https://doi.org/10.18280/ijsdp.160807

[4] Oceng, R., Andarani, P., & Zaman, B. (2023). Quantifying Plastic Waste and Microplastic Contamination in African Aquatic Systems: An Imperative for Sustainable Waste Management. *Acadlore Transactions on Geosciences*, *2*(2), 94-112. https://doi.org/10.56578/atg020204

[5] Sari, M. M., Inoue, T., Harryes, R. K., Suryawan, I. W. K., Yokota, K., Notodarmojo, S., & Priyambada, I. B. (2022). Potential of recycle marine debris in pluit emplacement, Jakarta to achieve sustainable reduction of marine waste generation. *International Journal of Sustainable Development and Planning*, *17*(1), 119-125. https://doi.org/10.18280/ijsdp.170111

[6] Obradovic, V. & Vulevic, A. (2023). Water Resources Protection and Water Management Framework in Western Balkan Countries in Drina River Basin. *Acadlore Transactions on Geosciences*, *2*(1), 24-32. https://doi.org/10.56578/atg020103

[7] Wan, L., Yuan, Y., Sun, H., & Qu, J. (2023). Simulation Study on Drop and Explosion Performance of Marine Steam-Water Separator. *International Journal of Simulation Modelling*, *22*(3), 497-507. https://doi.org/10.2507/IJSIMM22-3-CO12

[8] Kumar, A., Taxak, A. K., Mishra, S., & Pandey, R. (2021). Long term trend analysis and suitability of water quality of River Ganga at Himalayan hills of Uttarakhand, India. *Environmental Technology & Innovation*, *22*, 101405. https://doi.org/10.1016/j.eti.2021.101405

[9] Dhawde, R., Surve, N., Macaden, R., Wennberg, A. C., Seifert-Dähnn, I., Ghadge, A., & Birdi, T. (2018). Physicochemical and bacteriological analysis of water quality in drought prone areas of Pune and Satara districts of Maharashtra, India. *Environments*, *5*(5), 61-80. https://doi.org/10.3390/environments5050061

[10] Nazari-Sharabian, M., Ahmad, S., & Karakouzian, M. (2018). Climate change and eutrophication: a short review. *Engineering, Technology and Applied Science Research*, *8*(6), 3668-3672.

[11] Zhang, L. L., Liu, W. P., Zheng, B., Xu, A., & Sun, B. D. (2020). Features of roof rainwater runoff pollution in a northern coastal city under the effects of multiple factors. *International Journal of Sustainable Development and Planning*, *15*(4), 431-438. https://doi.org/10.18280/ijsdp.150403

[12] Kumar, A., Matta, G., & Bhatnagar, S. (2021). A coherent approach of water quality indices and multivariate statistical models to estimate the water quality and pollution source apportionment of river ganga system in Himalayan region, Uttarakhand, India. *Environmental Science and Pollution Research*, *28*, 42837-42852. https://doi.org/10.1007/s11356-021-13711-1

[13] García, A. C., Alban, C. A. P., Benalcázar, J. R. T., Rodríguez, A. C., Lorente-Leyva, L. L., & Aleaga, A. M. L. (2021). Control of pollutant emissions from a boiler through the percentage of oxygen. *Journal Européen des Systèmes Automatisés*, *54*(3), 469-474. https://doi.org/10.18280/jesa.540310

[14] Kou, X., Ding, J., Li, Y., Li, Q., Mao, L., Xu, C., & Zhuang, S. (2021). Tracing nitrate sources in the groundwater of an intensive agricultural region. *Agricultural Water Management*, *250*, 106826. https://doi.org/10.1016/j.agwat.2021.106826

[15] Jakóbczyk-Karpierz, S., Ślósarczyk, K., & Sitek, S. (2020). Tracing multiple sources of groundwater pollution in a complex carbonate aquifer (TarnowskieGóry, southern Poland) using hydrogeochemical tracers, TCE, PCE, SF6 and CFCs. *Applied Geochemistry*, *118*, 104623. https://doi.org/10.1016/j.apgeochem.2020.104623

[16] Shen, J., Deng, S., & Wu, J. (2021). Identifying Pollution Sources in Surface Water Using a Fluorescence Fingerprint Technique in an Analytical Chemistry Laboratory Experiment for Advanced Undergraduates. *Journal of Chemical Education*, *99*(2), 932-940. https://doi.org/10.1021/acs.jchemed.1c00534

[17] Hou, G., Yan, H., & Yu, Z. (2023). Application of AI Identification Method and Technology to Boron Isotope Geochemical Process and Provenance Tracing of Water Pollution in River Basins. *Sustainability*, *15*(7), 5942. https://doi.org/10.3390/su15075942

[18] Sun, C., Li, C. Q., Bai, B., Yang, G., & Wang, Q. (2020). Bayesian inference for source determination of sudden water pollution events. *China Rural Water and Hydropower*, *2020*(8), 71-75. https://doi.org/10.3969/j.issn.1007-2284.2020.08.014

[19] Wang, Z. H., Gong, L., Kang, C. T., Wang, H., & Yang, Y. Q. (2020). Tracing source of sudden water pollution in rivers and canals based on BAS algorithm. *Water Resources Protection*, *36*(5), 87-92. https://doi.org/10.3880/j.issn.1004-6933.2020.05.013

[20] Wei, X. S., Chen, Y. H., Chang, M., Gao, H. J., & Zhang, L. Y. (2022). Research progress on the monitoring and traceability technology of water pollution in river basins. *Environmental Monitoring in China*, *38*(5), 27-37. https://doi.org/10.3880/10.19316/j.issn.1002-6002.2022.05.04

[21] Hamdi, A. (2012). Inverse source problem in a 2D linear evolution transport equation: detection of pollution source. *Inverse Problems in Science and Engineering*, *20*(3), 401-421. https://doi.org/10.1080/17415977.2011.637207

[22] Zhang, S. P. & Xin, X. K. (2017). Pollutant source identification model for water pollution incidents in small straight rivers based on genetic algorithm. *Applied Water Science*, *7*, 1955-1963.

[23] Jing, L., Kong, J., Wang, Q., & Yao, Y. (2018). An improved contaminant source identification method for sudden water pollution accident in coaster estuaries. *Journal of Coastal Research*, *85*, 946-950. https://doi.org/10.2112/SI85-190.1

[24] Li, J., Lu, W., & Fan, Y. (2021). Groundwater pollution sources identification based on hybrid homotopy-genetic algorithm and simulation optimization.*Environmental Engineering Science*, *38*(8). https://doi.org/10.1089/ees.2020.0117

[25] Cheng, W. P. & Jia, Y. (2010). Identification of contaminant point source in surface waters based on backward location probability density function method. *Advances in Water Resources*, *33*(4), 397-410. https://doi.org/10.1016/j.advwatres.2010.01.004

[26] Guozhen, W., Zhang, C., Li, Y., Haixing, L., & Zhou, H. (2016). Source identification of sudden contamination based on the parameter uncertainty analysis. *Journal of Hydroinformatics*, *18*(6), 919-927.

[27] Sun, L. B. (2019). Exploring the application of artificial intelligence in the construction of smart libraries.*Jiangsu Science & Technology Information*, *36*(13), 20-22.

[28] Mao, W. S. (2020). The traceability method of water pollution incidents based on the knowledge-map relation inference. *Lanzhou: Lanzhou Jiaotong University*, *2021*(1). https://doi.org/10.27205/d.cnki.gltec.2020.000534

[29] Zhu, Z., Li, Y., Sun, Y.,Liu, Z., & Zhang, C. (2023). A Decision Support Framework for Pollution Source Detection via Coupled Forward-Inverse Optimization and Multi-Information Fusion. *Water Resources Research*, *59*(6), 1-18. https://doi.org/10.1029/2022WR032866

[30] Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-based systems*, *89*, 228-249. https://doi.org/10.1016/j.knosys.2015.07.006

[31] Zhou, R., Sun, Y., Chen, W., Zhang, K., Shao, S., & Zhang, M. (2023). A multi-objective decision model for water pollution load allocation under uncertainty. *Water*, *15*(2), 309. https://doi.org/10.3390/w15020309

Contact information:

Dongyan JIA, Lecture
(Corresponding author)
School of Mathematics and Information Technology,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
Hebei Innovation Center for Smart Perception and Applied Technology of Agricultural Data,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
E-mail: jdy_1983@163.com

Jinling SONG, Associate Professor
School of Mathematics and Information Technology,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
Hebei Innovation Center for Smart Perception and Applied Technology of Agricultural Data,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
E-mail: songjinling99@126.com

Lisha DONG, Lecture
School of Mathematics and Information Technology,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
E-mail: dlsjy0114@126.com

Yan KANG, Lecture
School of Mathematics and Information Technology,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
Hebei Innovation Center for Smart Perception and Applied Technology of Agricultural Data,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
E-mail: kangyan_1983@163.com

Xiaoning ZENG, Associate Professor
Institution School of Mathematics and Information Technology,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
E-mail: qhdzxn@163.com