Identification of source information for sudden hazardous chemical leakage accidents in surface water on the basis of particle swarm optimisation, differential evolution and Metropolis–Hastings sampling

Haidong Yang\textsuperscript{a,b}, Jinjin Li\textsuperscript{a}, Biyu Liu\textsuperscript{a,b,*}, Luying Chen\textsuperscript{a}

(a. School of Economics and Management, Fuzhou University, Fuzhou 350116, China; b. Department of Business Administration, Technology and Social Sciences, Luleå University of Technology, Luleå, 97187, Sweden)

Abstract: A quick and accurate identification of source information on sudden hazardous chemical leakage accident is crucial for early accident warning and emergency response. This study firstly regards source identification problem of sudden hazardous chemical leakage accidents as an inverse problem and constructs a source identification model based on the Bayesian framework. Secondly, a new identification method is designed on the basis of particle swarm optimisation (PSO), differential evolution (DE) and the Metropolis–Hastings (M-H) sampling method. Lastly, the designed method, i.e. PSO-DE-MH, is verified by an outdoor experiment analyses in a section of the South–North Water Transfer Project. Results show that the number of iterations, the average absolute error, the average relative error and the average standard deviations of the identification results obtained by PSO-DE-MH are less than those of PSO-DE and DE-MH. Moreover, the relative error and the sampling relative error of the identification results under five different measurement errors (MEs) ($\sigma = 0.01, 0.05, 0.1, 0.15, 0.2$) are less than 9.5% and 0.2%, respectively. The designed method is effective even when the standard deviation of the ME increases to 0.2. Therefore, the designed method can effectively and accurately obtain the source information of sudden hazardous chemical leakage accidents. This study provides a new idea and method to solve the difficult problems of emergency management.

Keywords: Emergency identification; Bayesian inference; Particle swarm optimisation; Differential evolutionary; Sudden hazardous chemical leakage accident

1 Introduction

With the implementation of the ‘One Belt, One Road’ strategy and the acceleration of urban construction, the production, storage, transportation and utilisation of hazardous chemicals in China have also increased (Zhao et al., 2018; Sun et al., 2019). However, the production, storage, transportation and utilisation of hazardous chemicals easily trigger the sudden hazardous chemical leakage accidents (SHCLAs) in surface water (Zhang et al., 2018; Yoo and Choi, 2019). According to the China Ecological Environment Bulletin in 2019, an average of 84 sudden environmental

* Corresponding author at: School of Economics and Management, Fuzhou University, Fuzhou, China. E-mail: jasperseu@fzu.edu.cn, Tel: +86 (0591) 22866403, Fax: +86 (0591) 22866403.
pollution accidents every four working days was directly dispatched or disposed by the Ministry of Ecological Environment, including many SHCLAs, such as the leakage of sodium cyanide in Yiyang County, Jiangxi Province and the leakage of cadmium in the Beijiang River, Guangdong Province. Many heavy SHCLAs have occurred in China since 2010. For example, in July 2010, an SHCLA was caused by chemical leakage in the Songhuajiang River, Jilin Province; in June 2011, a phenol leakage accident occurred in Xinanjiang, Zhejiang Province; in January 2012, a cadmium leakage accident occurred in the Longjiang River, Guangxi Province; in January 2013, an aniline leakage accident happened in the ZhuoZhang River, Shanxi Province; in May 2014, a waste lye leakage accident occurred in Jingjiang, Jiangsu Province; in November 2015, water pollution in the Jialing River was caused by tailing leakage in Longnan City, Gansu Province (Wang et al., 2015); in March 2016, a chemical leakage to irrigation sources in Chizhou, Anhui Province and a diesel oil leakage accident in Hanzhong, Shanxi Province occurred; in May 2017, a thallium leakage accident occurred in the Jialing River, Sichuan Province; in August 2018, a waste lye leakage accident happened in Nanyang City, Henan Province; in May 2019, water pollution outbreak occurred in the Beiluo River, Shanxi Province. These accidents have a series of uncertainties, such as leakage time, leakage position, leakage intensity, damage degree and influence range, which usually damage the water ecological environment and threaten human health, even affecting social stability (Yu et al., 2015; Andre and Badia, 2015; Wang et al., 2019; Lei et al., 2019). Many hazardous chemicals, such as organic matter, heavy metals and bacteria, are difficult to detect with the naked eye before causing harm (Wu et al., 2020). Once an SHCLA occurs, emergency decision-makers only grasp the law of hazardous chemical's migration and diffusion in water at the first time and then issue an accurate warning level and formulate appropriate emergency response measures (Yang et al., 2018; Lei et al., 2019). Amongst them, the historical reconstruction process of SHCLA is the key to reflect the migration and diffusion law of hazardous chemicals in water and track the responsible persons (Liu and Wang, 2020). Therefore, the ability to identify the leakage source information of SHCLA quickly and accurately plays an important role in the process of formulating emergency disposal and rescue measures.

At present, many scholars study the source information identification problem of sudden water pollution accidents (SWPAs) from the perspective of an inverse problem (Yang et al., 2016; Ashyralyev and Emharab, 2018). According to the mathematical characteristics of the existing emergency identification methods, which can be divided into four types, namely, the analytical, regularisation, deterministic and stochastic methods (Yang et al., 2016; Wang et al., 2016; Liu and Wang, 2020).

The analytical method is a method to obtain identification results by using the speed of
hazardous chemical diffusion and the concentration of the source of hazardous chemical leakage (Alapati and Kabala, 2000). For example, Skaggs and Kabala (1995) studied the reconstructed problem of a single point source release process at some observation point by using the quasireversibility method and found that this reconstructed problem can be changed into a well-posed problem. Although this method has high computational efficiency, its accuracy is relatively poor (Wu et al., 2020).

The regularisation method is a classical inverse method that obtains a numerical or analytic solution by converting ill-posed problems into a well-posed problem (Jana and Nair, 2020; Ma et al., 2017). For example, Skaggs and Kabala (1994) explored the release process of a pollutant base on a one-dimensional homogeneous steady flow by using the Tikhonov regularisation method. Wei et al. (2010) designed a new method based on optimal perturbation regularisation to identify the fractional diffusion equation of multipoint SWPAs. Hamdi and Adel (2007) studied the two-dimensional source identification problem of SWPAs by using the classical least square regularisation method. This method is used to solve the emergency identification problem, but it sacrifices the precision of the partial solution.

The deterministic method based on optimisation is a method to obtain the identification results of pollution sources by minimising the differences between simulated and observed values (Guan et al., 2006; Ayvaz and Tamer, 2016; Huang et al., 2018). For example, Parolin et al. (2015) studied the source identification problem of SWPAs using the Luus–Jaakola algorithm, the particle collision algorithm, ant colony optimisation and the golden section method, and the effectiveness of these methods were verified by a case study on the Macae estuary on the southeast coast of Brazil. Zhang and Xin (2017) studied the source identification problem of single- and multipoint source SWPAs on the basis of the genetic algorithm (GA). Jing et al. (2018) studied the source identification problem of SWPAs that occur in rivers by using an improved GA. Sun et al. (2019) studied the emergency identification problem of SWPAs that occur in rivers on the basis of the variable decoupling method and GA. Pan et al. (2020) studied the emergency identification problem of SWPAs that occur in groundwater, including the intensity and the permeability coefficient, by using a simulation–optimisation method. The source identification problem of SWPAs is an ill-posed problem; however, if a small error in observation value or identification model exists, the identification result can be obtained with a larger error by using the deterministic method (Hazart et al., 2014).

Multivariate nonlinear regression and the associated maximum likelihood method are two stochastic methods firstly used to identify the source information of SWPAs, whereas statistical induction, minimum relative entropy and the probability method are the three most commonly
used stochastic methods at present (Jing et al., 2020; Wu et al., 2020). Statistical induction can
analyse the uncertainty of emergency identification problems, but it needs substantial
measurement data. Minimum relative entropy is a method that can analyse the uncertainty of the
source identification problem of SWPAs and change the problem into a new problem, which can
be solved by a priori distribution. The inverse probability density method based on adjoint
equation is one of the most popular stochastic methods based on probability theory. For example,
Skiba and Parra-Guevara (2017) considered various factors related with oil spill, and proposed a
bioremediation strategy for an oil-polluted marine ecosystem based on adjoint equation approach.
Hu et al. (2019) used this method to identify the source identification problem of SWPAs and
analysed computational divergence during the process of emergency identification. Ghane et al.
(2016) and Jing et al. (2020) studied the inverse problem of the intensity of SHCLAs that occur in
rivers by using the inverse probability density method based on an adjoint equation and verified
the effectiveness of the method. Emergency identification problems are relatively difficult to
identify under high-dimensional unsteady flow, and the stability of the method requires further
study. The Bayesian inference method is another stochastic method based on probability theory,
and it transforms emergency identification problems into the posterior estimation of unknown
parameters on the basis of Bayesian inference and Markov chain Monte Carlo sampling technique
(Bayesian-MCMC) (Han et al., 2014; Yang et al., 2016; Yu et al., 2020; Zhang et al., 2020). For
example, Jiang et al. (2017) studied the source identification problem of SWPAs that occur in
rivers on the basis of Bayesian theory and the tracer experiment. Guo et al. (2016) studied the
source identification problem of SWPAs that occur at the South-to-North Water Diversion project
by using Bayesian-MCMC. The identification results obtained by Bayesian-MCMC have strong
randomness and avoid the decision risk caused by ‘optimal’ parameter distortion; however, its
calculation increase exponentially with the increase in parameters.

Considering the urgency of dealing with SWPA, some scholars have designed coupling methods
based on the optimisation and stochastic methods (Chen et al., 2007; Wang et al., 2016; Yang et al.,
2016; Cao and Yun, 2017; Wu et al., 2020). Two types of coupling methods exist: one is based on
different optimisation methods, whereas the other is based on the stochastic method and one
optimisation method. For example, Cao and Yun (2017) studied single-point source identification
problem and multipoint source identification problem by coupling particle swarm optimisation
(PSO) and differential evolution (DE) and verified the accuracy of this designed method through
several examples. Wang et al. (2018) analyzed the advantage and disadvantage of fuzzy adaptive
Kalman filter and weighted recursive least squares algorithm, developed an inverse analysis
method for the real-time monitoring of pollutant diffusion. Yang et al. (2016) studied the
multipoint source identification of SWPA by coupling DE and the Metropolis–Hastings (M-H) sampling method and verified the effectiveness and accuracy of the designed method by conducting outdoor experiments. Wu et al. (2020) explored the identification problem of SWPAs in rivers or lakes and designed a new identification method based on an adaptive Metropolis sampling method and DE and verified the accuracy of this designed method by conducting a case study. Although the first type of coupling method has better searching ability than the deterministic method, it cannot deal with uncertainties during the process of source identification; the other type of coupling method has the same ability to deal with uncertainties as the stochastic method, but its searching ability needs to be strengthened. As typical SWPAs, SHCLAs have general characteristics similar to sudden accidents but also have the characteristics of sprawl, transformation and coupling (Chen et al., 2007; Wang et al., 2016). Therefore, in consideration of resource limitations and the urgency of emergency response, designing a faster and more accurate emergency identification method is necessary.

Based on the above analysis and the characteristics of different methods, this study designs a new identification method on the basis of the M-H sampling method, PSO and DE. In this method, a set of possible solutions is firstly found from the solution space by PSO, and then the most likely solution set, which is close to the true value, is found from a set of possible solutions by DE; the identification results are finally obtained by M-H sampling from the most likely solution set. The new approach not only uses the search ability of PSO and the propagation thought of DE to improve the accuracy and search speed in the posterior space but also uses the sample point update strategy to reduce the uncertainty in the process of emergency identification problem. To verify the effectiveness and accuracy of the new method, this study takes outdoor experiment as a case study and comparatively analyses the identification results by using the new method, PSO-DE and DE-MH. The calculation results show that the new method effectively improves calculation efficiency, thus meeting the needs of emergency management.

This paper is organised as follows: Section 2 describes the source identification problem of SHCLAs and constructs the emergency identification model. Section 3 designs a new coupling method on the basis of the searching ability of DE and PSO under the Bayesian inference framework. Section 4 presents the process and results of a case study. Section 5 provides the conclusions.

2 Problem description and model formulation

The unknown parameters of hazardous chemical leakage sources (e.g. positions, intensity and time) can be identified through observation and prediction values (Jun et al., 2007), where prediction values can be obtained on the basis of the laws of the pollutants’ migration and diffusion and the available pollutants’ concentration distribution information. Therefore, in this
section, the leakage source identification problem of SHCLA is first described, and then the laws of hazardous chemical migration and diffusion in water is discussed, an emergency identification model of SHCLA is finally established.

### 2.1 Description of source identification problem

With the rapid development of economy and society, the transportation and utilisation of hazardous chemicals are increasing, thus greatly increasing the occurrence probability of SHCLAs. We take an SHCLA that occurs in a river or a canal as an example; if we consider a section with a length of \(L\) and width of \(B\), then the geometric shape of the section does not change much. \(S\) represents the leakage source information of hazardous chemicals (e.g. intensity, position and leakage time), as shown in Figure 1. The section has \(m\) observation points, and \(q_j(t)\) is used to represent the concentration distribution function of water quality index obtained from the \(j\)th observation point at time \(t\).

![Fig. 1. Schematic of the emergency identification of SHCLA](image)

Figure 1 shows that once the observation values from the \(j\)th (\(j=1, 2, \ldots, m\)) observation point suddenly appears abnormal, that is, once SHCLA occurs, emergency decision-makers urgently need to identify leakage source information \(S\) and then formulate emergency response measures to ensure the safety of water quality downstream. In other words, the source identification problem of SHCLA is how to obtain the source information \(S\) by the observation values \(q_j(t)\) quickly and accurately.

### 2.2 Emergency identification model

Two-dimensional convection–diffusion equations are widely used in water quality models and are commonly used to simulate the migration and diffusion of hazardous chemicals in water after SHCLA occurs.
where $C(x, y, t)$ is the concentration of hazardous chemicals in position $(x, y)$ at time $t$ (g/L); $u_x$, $u_y$ represent the longitudinal and transverse velocity of water flow (m/s), respectively; $D_x$, $D_y$ are the longitudinal dispersion and transverse diffusion coefficients (m$^2$/s), respectively; $K(t)$ is the degradation function of hazardous chemicals.

However, the diffusion of hazardous chemicals in the water body is restricted by boundary and reflection due to the constraint of shore and bottom. Assume that the initial conditions are $C=C_{|t=0}$ and $C=C_{|t=\infty}$. If the bilateral reflection on both sides of the water body is considered and the leakage position of SHCLA is taken as the origin point, then the solution of Equation (1) is shown as Equation (2).

$$C(x, y, t) = C_1(x, y, t) + C_2(x, y, t) + C_3(x, y, t) + C_4,$$  \hspace{1cm} (2)

$$C_1(x, y, t) = \frac{M}{4\pi h \sqrt{D_x D_y t}} \exp[-kt - \frac{(x-u, t)^2}{4D_x t} - \frac{(y-u, t)^2}{4D_y t}],$$  \hspace{1cm} (3)

$$C_2(x, y, t) = \sum_{n=1}^{\infty} \frac{M}{4\pi h \sqrt{D_x D_y t}} \exp[-kt - \frac{(x-u, t)^2}{4D_x t} - \frac{(2nb + y-u, t)^2}{4D_y t}],$$  \hspace{1cm} (4)

$$C_3(x, y, t) = \sum_{n=1}^{\infty} \frac{M}{4\pi h \sqrt{D_x D_y t}} \exp[-kt - \frac{(x-u, t)^2}{4D_x t} - \frac{(2nb - 2nB - y-u, t)^2}{4D_y t}],$$  \hspace{1cm} (5)

where $M$ is the leakage intensity of hazardous chemicals, g; $b$ is the distance between the leakage point and the shore, m; $B$ and $h$ are respectively the width and average depth of the water body, m; $n$ is the number of reflections across the boundary; $x$ and $y$ are respectively the longitudinal distance and transverse distance from the predicted point to the leakage position, m; $C_1(x, y, t)$ is the concentration increment generated by the leakage source of hazardous chemicals at point $(x, y)$; $C_2(x, y, t)$ is the concentration increment generated by reflection near the shore at the point $(x, y)$; $C_3(x, y, t)$ is the concentration increment generated by reflection from the far shore at the point $(x, y)$; $C_4$ is the background concentration of water.

SHCLAs have many leakage modes, and one of which is instantaneous nonshore leakage mode (e.g., the leakage accident occurs when a vehicle carrying hazardous chemicals is driving on the bridge due to traffic accidents, machine malfunction or other reasons). If the background
concentration of the water body is ignored, then the concentration of hazardous chemicals at the
downstream point \((x, y)\) can be expressed as Equation (6).

\[
C(x, y, t) = \frac{M}{4\pi h} \exp\left(-Kt - \frac{(x-u)^2}{4Dt}\right) \left[\exp\left(-\frac{(y-v)^2}{4Dt}\right)\right] (\text{6})
\]

If the hazardous chemical leakage mode is an instantaneous shore leakage, then the
correlation of the hazardous chemical at the downstream point \((x, y)\) can be expressed as
Equation (7).

\[
C(x, y, t) = \frac{M}{2\pi h} \exp\left(-Kt - \frac{(x-u)^2}{4Dt} - \frac{(y-v)^2}{4Dt}\right) (\text{7})
\]

If an SHCLA has \(q(q \geq 1)\) leakage sources and \(k(k \geq 1)\) observation points, \(d=\{d_1, d_2, \ldots, d_k\}\),
\(g=\{g_1, g_2, \ldots, g_k\}\) and \(h=\{h_1, h_2, \ldots, h_k\}\) respectively denote the set of observation value, prediction
and true values of the observation points, then there exist measurement errors (MEs) and
prediction errors (PEs). Given the characteristics of symmetry, unimodality, boundedness and
offsetting, the ME and PE are usually assumed to obey the normal Gauss distribution (Guo et al.,
2009; Kastner et al., 2013). Therefore, the probability density of ME and PE at the \(i\)th observation
point are respectively expressed as Equation (8).

\[
\begin{align*}
  f(d_i | h_i, S) &\propto \frac{1}{(2\pi \sigma_{d,i}^2)^{1/2}} \exp\left[-\frac{(d_i - h_i)^2}{2\sigma_{d,i}^2}\right] \\
  f(h_i | g_i, S) &\propto \frac{1}{(2\pi \sigma_{h,i}^2)^{1/2}} \exp\left[-\frac{(h_i - g_i)^2}{2\sigma_{h,i}^2}\right]
\end{align*}
\] (8)

where \(\sigma_{d,i}\) and \(\sigma_{g,i}\) are respectively the standard deviations of the ME and PE at the \(i\)th
observation point.

Given that \(k(k \geq 1)\) observation points are independent of each other, then the likelihood
function can be described as Equation (9) (Carrera and Neuman, 1986a; Aster et al., 2005).

\[
L(d | S) = \prod_{i=1}^{k} f(d_i | h_i, S) = \prod_{i=1}^{k} \prod_{g_i} f(d_i | h_i, S) f(h_i | g_i, S) dg_i
\]

\[
\propto \frac{1}{(2\pi)^{1/2} \prod_{i=1}^{k} (\sigma_{d,i}^2 + \sigma_{g,i}^2)^{1/2}} \exp\left[-\sum_{i=1}^{k} \frac{(C_i(x, y, t) - C_i(x, y, t | S))^2}{2(\sigma_{d,i}^2 + \sigma_{g,i}^2)}\right]
\] (9)

where \(C_i(x, y, t) = g_i, C_i(x, y, t | S) = d_i\), in which \(d\) is the observation vector with length \(k\).

The maximum likelihood estimator (MLE) is a method that can be used to estimate parameters
by using given observation values. Its principle is to find a parameter vector, which maximises the
probability of observation values (Zeunert and Günter, 2020). To improve the evaluation and
convergence of parameters, the MLE is usually determined by minimising the negative
log-likelihood function (MNLL) (Carrera and Neuman, 1986b). Therefore, the emergency
identification model of SHCLA should be expressed as Equation (10).

\[ MNLL = \arg \min_{S} [-2 \ln L(d \mid S)] \approx \min_{\lambda} \sum_{i=1}^{k} \left[ \ln(\sigma_{f,i}^2 + \sigma_{i}^2) + \frac{(C_i(x,y,t) - C_{\text{obs}}^i)}{\sigma_{f,i}^2 + \sigma_{i}^2} \right] \] (10)

If we substitute Equation (6) or Equation (7) into Equation (10), then the source information \( S \)
can be identified or estimated by using the M-H sampling method. However, some limitations,
such as large amount of computations and low sampling efficiency, exist when the M-H sampling
method is used to solve the leakage source identification problem of SHCLA. To obtain a more
accurate source information of SHCLA, the M-H sampling method needs to adjust the Markov
chain to reach the convergence domain through a vast number of calculations, and this
requirement reduces the identification efficiency and then affects the emergency effect of SHCLA.
A vast number of calculations outside the convergence domain can be avoided if we can improve
the efficiency of M-H sampling. Given that DE and PSO have strong global and local searching
ability respectively, a new identification method that combines PSO, DE and M-H sampling is
designed to identify the source of SHCLA in this study.

3 Emergency identification method

This new identification method is designed by combining PSO, DE and M-H sampling on the
basis of Bayesian inference. To explain the proposed identification method, we firstly discuss the
procedure of PSO, DE and M-H sampling, then explore the searching mechanism and operation
process of the new method, and finally discuss the index to diagnose the convergence of the new
method.

3.1 Particle swarm optimisation

PSO, which was proposed in 1995, is an evolutionary algorithm based on the behaviour of birds
(Eberhart and Kennedy, 1995a). The basic idea of PSO is to find the optimal solution through the
cooperation and information sharing amongst individuals in the population (Eberhart and Kennedy,
1995b). In contrast to other evolution algorithms, PSO is easy to implement with few evolution
parameters (Gupta et al., 2016; Jain et al., 2018). In PSO, it firstly initialises a group of random
particles, where each particle has its position and velocity (Tang et al., 2007; Hou and Jiang, 2017).
Then, the optimal solution is found by updating the iteration. In each iteration, the updated
velocity of each particle is determined mainly by \( p_{\text{best}} \) and \( g_{\text{best}} \), where \( p_{\text{best}} \) denotes the best
solution found so far by a particle in the position, and \( g_{\text{best}} \) indicates the best value found so far by
all the particles in the population. Figure 2 shows the procedure of PSO.
The convergence speed of PSO is fast and easy to realise and thus has shown its unique advantages in solving and applying various problems. However, it easily stagnates, and the convergence accuracy is low in the process of convergence.

**3.2 Differential evolution**

DE is a population-based evolutionary algorithm that searches solutions randomly over a continuous space through real vector coding (Storn and Price, 1997). The core idea of DE is to obtain the mutation operator by the difference of the multiple pairs of vectors selected arbitrarily in the population (Ali and Törn, 2004). Figure 3 illustrates the procedure of DE.

DE has fast convergence rate and strong robustness and thus can be used to solve optimisation problems over continuous spaces (Bergey and Ragsdale 2005; Rönkkönen et al. 2005; Qin et al. 2009).
Given the advantage of greedy optimisation, DE has improved optimisation performance, but it greatly reduces the ability of the population to resist the local extremum attraction. As a result, DE is more suitable for accelerating convergence in the later state of the algorithm.

### 3.3 Metropolis–Hastings sampling method

M-H sampling is a method to estimate unknown parameters by constructing a Markov chain (Raje and Krishnan 2012). To make the method ergodic in solution space, however, new samples are usually generated on the basis of a proposal distribution (Brooks and Roberts, 1998). But the proposal distribution should satisfy the following three conditions: (a) its probability density function is a constant, (b) it can be calculated, and (c) it can generate random numbers. Figure 4 illustrates the procedure of M-H sampling method.

**Fig. 3.** Procedure of DE
However, the sampling efficiency of M-H sampling is not high because the acceptance rate of the constructed Markov chain during the transfer process may be small; thus, the Markov chain easily stands during the sampling process and refuses to jump large numbers.

3.4 PSO-DE-MH

Numerous uncertainties exist in the source identification problem of SHCLA, leading to the poor efficiency or low accuracy of the identification results obtained by the existing emergency
identification methods. Therefore, to improve the efficiency and accuracy of identification results, a new method is proposed in this section by combining PSO, DE with M-H on the basis of different searching mechanisms.

For M-H, the calculating efficiency and accuracy of the identification results largely depend on the convergence rate of the Markov chain. However, the convergence rate of the Markov chain is slow because M-H needs to use less unknown parameters for prior information to obtain the exact target region, especially for complex or high-dimensional models.

PSO and DE exactly improve the convergence rate of the Markov chain constructed by M-H. For example, the new testing parameters are examined by updating the position and velocity of particles based on the movement rule of PSO, or they are determined by mutation and cross-operation based on the propagation idea of DE. To consider the uncertainties of source identification problem fully and ensure that the parameter space is searched efficiently, we firstly search the possible solutions that are close to true values on the basis of PSO, and then the possible solutions closer to the true values are obtained by an improved mutation formula based on the propagation idea of DE. Therefore, the new testing parameters $X_{pbest}^{(i)}$ are generated in accordance with Equation (11).

$$X_{pbest}^{(i)} = X_{i,pbest}(r_i) + E \times (X_{i,pbest}(r_j) - X_{i,pbest}(r_k)) + \varepsilon,$$  \hspace{1cm} (11)

where $X_{i,pbest}$ is the $i$th generation population consisting of the individual optimal $Y_{best}$ of all particles generated by PSO; $r_1$, $r_2$, $r_3$ are mutually different integers in the $i$th generation population; $E$ is a scaling factor of the difference vector and is a positive real number; $\varepsilon$ is the given disturbance that reflects the uncertainties existing in the identification problem.

In Equation (11), scaling factor $E$ and disturbance $\varepsilon$ are the tuning parameters. Scaling factor $E$ has a great influence on identification solutions. A small $E$ value converges prematurely and falls into local optimum, whereas a large $E$ value easily slows down the search speed and skips global optimal solutions (Vitaliy, 2006). In addition, if the value of $\varepsilon$ is too large, then it reduces the accuracy of identification solutions (Roberts and Rosenthal, 2004). An interaction exists between $E$ and $\varepsilon$. When we choose a large value of $E$ and ignore $\varepsilon$, the search process becomes random; when the chosen $\varepsilon$ is too large, it cannot reflect the influence of $E$ on the sampling method. Therefore, neither $E$ nor $\varepsilon$ should be set too large or too small. Regarding the proposed problem in this study, we choose $E$ as a constant value in $[0, 2]$ and set $\varepsilon$ to $1\%$–$20\%$ of the range of prior distribution (Storn and Price, 1997; Roberts and Rosenthal, 2004). Moreover, the size of $NP$ affects the searching efficiency. Generally, to maintain the balance between variety of convergence rate and population, the size of $NP$ should be set to $[10, 50]$, where $NP \in [10, 35]$ is suitable for low-dimensional problems and $NP \in [35, 50]$ for high-dimensional problems (Mohamed and Sabry,
2012). Figure 5 shows the process of generating new testing parameters with three dimensions by PSO and DE. \( X^{(i)}(j) \) in Figure 5 represents the individuals in the \( j \)th population with the \( i \)th iteration, \( j \in [1, NP] \).

Figure 5: Process of generating new testing parameters \( X_{**} \) with three dimensions by PSO and DE

PSO-DE-MH is an integrated method based on PSO, DE and M-H. In other words, we firstly find the approximate location of the optimal solution by PSO, and then the optimal solution is found by using the propagation idea of DE and the sampling idea of M-H. To coordinate the global and local search capability, the new method adopts a nonlinear change strategy, including inertial weight, learning factor and scaling factor. The detailed procedure of the new method to solve this considered problem is as follows:

(i) Initialising the particle swarm

In accordance with the number of variables \( S \), the population size \( (NP) \), the maximum number of particle swarm movement \( \text{iter}_{\text{max}} \) and the maximum sampling times \( I_{\text{max}} \) are determined, initialising velocity \( V_{i}^{\text{iter}} \) \((i=1,2,\ldots, I_{\text{max}}; \text{iter}=1, 2, \ldots, \text{iter}_{\text{max}}) \) and position \( Y_{i}^{\text{iter}} \) \((i=1,2,\ldots, I_{\text{max}}; \text{iter}=1, 2, \ldots, \text{iter}_{\text{max}}) \).

(ii) Calculating the inertial weight

Inertial weight \( w \) is a key parameter that affects the performance of the algorithm. To make the particle swarm have good global and local search ability at the initial and later stages of iteration,
respectively, a dynamic inertial weight is adopted, as shown in Equation (12).

\[ w_{iter} = w_{min} - (w_{max} - w_{min})(\frac{iter - iter_{max}}{iter_{max}})^3, \text{iter} \in [1, iter_{max}], \]  

(12)

where \( \text{iter} \) is the number of particle swarm movement, \( w_{max} \) is the maximum value of inertial weight, \( w_{min} \) is the minimum value of inertial weight. Generally, \( w_{max} = 0.9 \), and \( w_{min} = 0.4 \).

(iii) Calculating the time-varying learning factor

The learning factors \( c_1 \) and \( c_2 \) affect the individual optimal and global optimum of each particle, respectively. Venter and Sobiesczanski-Sobieski (2004) donated \( c_1 \) as self-confidence and \( c_2 \) as swarm-confidence. The smaller learning factor limits the movement of the particle, whereas the larger learning factor causes the particle to be divergent. A time-varying learning factor is adopted in the designed method, as shown in Equation (13).

\[
\begin{align*}
    c_1 &= c_{11} - c_{12} \times (\frac{2 \cdot \text{iter} - \text{iter}_{max}}{\text{iter}_{max}})^3, \\
    c_2 &= c_{21} - c_{22} \times (\frac{2 \cdot \text{iter} - \text{iter}_{max}}{\text{iter}_{max}})^3,
\end{align*}
\]  

(13)

where \( c_{11}, c_{12}, c_{21} \) and \( c_{22} \) are constants. Generally, \( c_{11} = c_{12} = 1.5 \), and \( c_{21} = c_{22} = 0.5 \).

(iv) Calculating and updating the fitness, position and movement speed of the particle swarm

\[
\begin{align*}
    Y_{i,iter+1} &= w \cdot Y_{i,iter} + c_1 \times \text{rand}_i(Y_{i,pbest} - Y_{i,iter}) + c_2 \times \text{rand}_i(Y_{i,gbest} - Y_{i,iter}), \\
    Y_{i,iter+1} &= Y_{i,iter} + Y_{i,iter+1},
\end{align*}
\]  

(14)

where \( Y_{i,iter} \) is the individual optimum of the particle swarm after \( \text{iter} \) movement, and \( Y_{i,iter} \) is the optimal value of \( Y_{i,pbest} \).

(v) Repeating steps (ii)–(iv) until \( \text{iter} = \text{iter}_{max} \) or \( Y_{i,iter} \) satisfies the minimum limit

(vi) Combining the individual optimal \( Y_{i,iter} \) (iter = 1, 2, …, iter\(_{max}\)) to form the \( i \)th iteration population \( X^{(i)} \) and calculating the concentration of hazardous chemical and conditional probability density of \( X^{(i)} \)

(vii) Selecting randomly three different particles from a new particle swarm and then performing mutation in accordance with Equation (15) to obtain a new test parameter \( X^{(*)} \)

\[
X^{(*)} = X^{(i)}(r) + E \cdot (X^{(i)}(r) - X^{(i)}(r)) + \varepsilon,
\]  

(15)

where \( E \) is a scaling factor; \( r_1, r_2 \) and \( r_3 \) are respective mutually different integers in the \( i \)th generation population.

(viii) Obtaining the accept probability \( A(X^{(i)}(r_1), X^{(*)}) \) at which the Markov chain moves from \( p(X^{(i)}(r_1)) \) to \( p(X^{(*)}) \) by Equation (16)
\[ A(X^{(i)}(r_i), X^{(i)}) = \min\{1, \frac{p(X^{(i)})}{p(X^{(i)}(r_i))}\} \]  

1. Generating a random number \( R (R \in [0,1]) \) that follows a uniform distribution.
2. If \( R < A(X^{(i)}(r_i), X^{(i)}) \), then the testing parameters are accepted, and \( X^{(i)} = X^{(i)}(r_i) \);
3. otherwise, \( X^{(i)} = X^{(i)}(r_i) \).
4. (x) Repeating steps (i)–(ix) until the predetermined iterations are completed.
5. Figure 6 shows the detailed operation process of PSO-DE-MH.
Combining the individual optimal $Y_{i,\text{pbest}}^{\text{iter}}$ and global optimal $Y_{i,\text{gbest}}^{\text{iter}}$ to form the $i$th Iteration

Calculating the Pollutant concentration corresponding to the new parameters

Calculating the posterior probability density function of the new Parameters

Generating a new testing parameters $X^{(*)}$

Calculating the fitness of each particle

Calculating the velocity parameter and updating particle position

Calculating individual optimal $Y_{i,\text{pbest}}^{\text{iter}}$ and global optimal $Y_{i,\text{gbest}}^{\text{iter}}$ according to fitness

Determining Inertia Weight $w$ and learning factor $c_1$ and $c_2$

Calculating the fitness of each particle

Calculating the posterior probability density function of the $i$th Iteration

Is the terminal constraints satisfied?

Calculating the concentration of hazardous chemical corresponding to the $i$th Iteration

Fig. 6. Operation process of the PSO-DE-MH method
3.5 Convergence diagnostic method

Scale Reduction Score (SRS) is a method to judge the convergence of an algorithm by calculating a quantitative diagnostic indicator (Gelman and Rubin, 1992). The proceeds of the method as follows:

(i) Initializing the number of sampling sequences $k$ and is the number of iterations within each sequence $n$

(ii) Calculating the average variances of all the sampling sequences $H$ by Equation (17)

$$H = \frac{n}{k-1} \sum_{i=1}^{k} \left( \frac{1}{n} \sum_{j=1}^{n} x_{ij} - \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij} \right)^2,$$  

(17)

where $x_{ij}$ is the $j$th sample value of the $i$th sampling sequence.

(iii) Calculating the average value of the $k$ within-sequence variances $W$ by Equation (18)

$$W = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{1}{n-1} \sum_{j=1}^{n} \left( x_{ij} - \frac{1}{n} \sum_{j=1}^{n} x_{ij} \right)^2 \right)$$  

(18)

(iv) Calculating an index called scale reduction score (SRS) by using Equation (19)

$$\sqrt{SRS} = \sqrt{\frac{n-1}{n} + \frac{k+1}{nk} H}.$$  

(19)

(v) If $\sqrt{SRS} \approx 1$, then the generated samples converge to the posterior distribution of the unknown parameters which need to be identified. Otherwise, the generated samples do not converge to the posterior distribution.

4 Results and discussion

4.1 Outdoor experiment analyses

As we all know, China is a country with extremely uneven water resources. In order to balance water resources, China has been building the South–North Water Transfer Project (SNWTP) since 2002, which is divided into East Route Project (ERP), Middle Route Project (MRP) and West Route Project (WRP), as shown in Figure 7(a). The starting point of MRP is the Danjiangkou Reservoir in Hubei Province, its water supply area comprises the Henan province, Hebei province, Beijing city and Tianjin city (Yang, et al., 2020). The total distance of MRP is 1276 km, which including many hydraulic structures, such as over 1300 bridges, 88 diversion sluice gates, 53 exist sluice gates, 26 aqueducts, and so on. Moreover, MRP is constructed in the form of three-dimensional cross arrangement, along which there are 1640 cross buildings, such as river channel crossing, left bank drainage, canal crossing, railway crossing and highway crossing. It is worth noting that, according to the survey of china south-to-North water diversion group, more than 600 pollution sources along the MRP, which can easily prone to the occurrence of SHCLA.
To verify the proposed method, we choose 3km canal section as the experimental object, which along the direction of DongyangGe to the bridge of Baiyun in Baoding city, Hebei province, and take DongyangGe sluice gate as observation point D, as shown as Figure 7(b). During the experiment, we use sucrose as tracer, the system and pollutant characteristics shown in Table 1. The first time the studied section was discovered to be polluted is set as 10:00 am, and the measurement is conducted every 10 min thereafter until 2:00 pm. The concentration distribution of hazardous chemicals at observation point D, as shown in Figure 8.

Table 1 Parametrization of the experimental canal section.

| Parameter          | Description                          | Value     |
|--------------------|--------------------------------------|-----------|
| $h$                | Canal depth                           | 3.5m      |
| $W$                | Canal width                           | 40m       |
| $D_x$              | Dispersion coefficient in x-direction | 1225 m²/min |
| $D_y$              | Dispersion coefficient in y-direction | 36 m²/min  |
| $u_x$              | Flow velocity in x-direction          | 21.6m/min |
| $u_y$              | Flow velocity in y-direction          | 0.12m/min |
| $K$                | Decay rate                            | 1/14400   |
| $m$                | Intensity                             | $10^6$g/ m³ |
| $x_0$              | x-coordinate of source location       | 5000m     |
| $y_0$              | y-coordinate of source location       | 21m       |
| $t_0$              | Leakage time                          | 120min    |
Figure 8 shows that the sequence value of pollutant concentration at observation point D are approximately normally distributed, and the highest value is approximately 3.1777 g/m³, occurring at approximately 11:50 am.

4.2 Analysis of emergency identification results

In accordance with the characteristics of SHCLA and the prior information of the leakage source, all unknown parameters are uniformly distributed and independent of each other. Then, the posterior probability density function of the unknown parameters can be expressed as Equation (20).

$$\sigma_i(m, x_i, y_i, t_i) = \begin{cases} \lambda \exp\left[\sum_{i=1}^{n} \left(-\frac{(C(m, x_i, y_i, t_i) - d_i)^2}{2(\sigma_{d,i}^2 + \sigma_{g,i}^2)}\right)\right] & 0.4 \times 10^6 < m < 1.6 \times 10^6, \\ 0 & 0 < x_i < 10^4, 0 < y_i < 40, 0 < t_i < 200 \end{cases} \quad (20)$$

where $d_i$ is the observation value, $\lambda$ is the proportionality constant, $\sigma_{d,i}$ and $\sigma_{g,i}$ are respectively the standard deviations of ME and PE.

To simplify the calculations, only ME is considered in this section and is assumed to follow a Gaussian distribution with $\sigma_{d,i} = 0.01$. Moreover, we set the parameters $E$, $NP$ and $\varepsilon$ as stated in the rule in Section 3.4. After repeated computation, we set $E = 0.1$, $NP = 10$, $\varepsilon = 0.01$, $w_{\text{max}} = 0.9$, $w_{\text{min}} = 0.4$, $c_{11} = c_{12} = 1.5$, $c_{21} = c_{22} = 0.5$, $iter_{\text{max}} = 3000$ and $l_{\text{max}} = 10,000$. In accordance with the operation process of PSO-DE-MH, the Frequency distribution and the iterative curve of pollution sources, such as the intensity $m$, position $(x_0, y_0)$ and leakage time $t_0$, are illustrated in Figures 9-11, respectively. The comparison of identification results and truth values is shown in Table 2. At the same time, according to Equations (18)–(20), we calculate the scale reduction scores of the unknown parameters with $k = 4$ and $n = 10000$. Calculating result shows $\sqrt{\text{SRS}} = 1.033$, that is, the proposed method (i.e. PSO-DE-MH) has convergence.
Fig. 9. the Frequency distribution of Intensity (a) and Leakage time (b) based on PSO-DE-MH

Fig. 10. the Frequency distribution of x-coordinate of source location (a) and y-coordinate of source location (b) based on PSO-DE-MH

Fig. 11. Iterative curve of unknown parameters based on PSO-DE-MH
Table 2 Comparison between identification results based on PSO-DE-MH and the true values

| Parameters                          | Identification result | Position \((x_0, y_0)\)  | Leakage time \(t_0\) |
|-------------------------------------|-----------------------|--------------------------|----------------------|
| Intensity \(m\)                     | \(0.994 \times 10^6\ g/m^3\) | \((4992.8m, 20.7381m)\) | 119.6473min          |
| Absolute error \(g/m^3\)           | \(6 \times 10^3\)     | \((7.2m, 0.2619m)\)      | 0.3527min            |
| Relative error (%)                  | 0.6%                  | (0.144%, 1.25%)          | 0.29%                |
| Sampling standard deviation         | 0.0244                | (0.0611, 0.0546)         | 1.3551               |

As shown in Figure 9(a), the generated samples of source intensity \(m\) are almost fall in \([0.995 \times 10^6, 1.015 \times 10^6]\). Among them, 75.52% of the generated samples are fall in \([0.995 \times 10^6, 1.015 \times 10^6]\). Figure 9(b) shows that the generated samples of leakage time \(t_0\) are almost fall in \([120, 125]\). Among them, 61.21% of the generated samples are fall in \([120, 123]\). According to Figure 10, the generated samples of Leakage position \((x_0, y_0)\) are almost respectively fall in \([4990, 5130]\) and \([21.05, 22.45]\). Among them, over 60 percent of generated samples are respectively fall in the \([4990, 5080]\) and \([21.05, 21.85]\). According to Figure 11, the identification results approach the truth values of the parameters after approximately 300 iterations, and the constructed Markov chain has the characteristics of stable and piecewise smooth. As shown in Table 2, if we adopt PSO-DE-MH to solve the identification problem of hazard chemical leakage source, then the relative error and the sampling standard deviation of unknown parameters are less than 1.3% and 1.4, respectively.

Therefore, the identification results with the proposed method are close to the true values, and the generated samples also converge to the posterior distribution.

4.3 Comparative analysis with PSO-DE and DE-MH

To verify the effect of PSO-DE-MH, DE-MH and PSO-DE are used in this study to solve this emergency identification problem of SHCLA simultaneously. The iterative curve of the unknown parameters is shown in Figures 12 and 13, respectively. We perform the following operations for the above three methods: (i) The identification results were analysed after eliminating the unstable results, as shown in Table 3. (ii) The calculation errors of the unknown parameters were extracted at intervals, as shown in Figure 14.

![Fig. 12. Iterative curve of unknown parameters based on DE-MH](image-url)
Fig. 13. Iterative curve of unknown parameters based on PSO-DE

Table 3 Comparison between the true values and identification results based on PSO-DE and DE-MH

| Parameters                   | Intensity $m$ (g/m³) | Position $(x_0, y_0)$ (m) | Leakage time $t_0$ (min) |
|------------------------------|----------------------|---------------------------|-------------------------|
|                              | PSO-DE   | DE-MH      | PSO-DE      | DE-MH     | PSO-DE | DE-MH       |
| Identification result        | 928900   | 1010300    | (4907.5,    | (5066.2,   | 113.77 | 123         |
|                              |          |            | 20.39       | 21.75)     |         |             |
| Absolute error               | 71100    | 10300      | (92.5, 0.61)| (66.2, 0.75)| 6.23   | 3           |
| Relative error (%)           | 7.11     | 1.03       | (1.85, 2.89)| (1.32, 3.59)| 5.19   | 2.5         |
| Sampling standard deviation  | 0.08     | 0.02       | (0.58, 0.81)| (0.03, 0.48)| 1.76   | 1.45        |

Fig. 14. Relative error analysis of inversion results based on PSO-DE-MH (a), DE-MH (b) and PSO-DE (c)
Table 4 Calculation error comparison amongst PSO-DE-MH, DE-MH and PSO-DE

| Error          | Intensity m | Position (x_0,y_0) | Leakage time t_0 |
|----------------|-------------|--------------------|------------------|
|                | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Maximum (%)    | 27.17 | 27.17 | 27.16 | (37.7, 3.58) | (4.44, 7.31) | (37.7, 3.58) | 6.39 | 8.22 | 6.39 |
| Minimum (%)    | 0.04  | 0.35  | 5.93   | (0.01, 0.81)    | (0.55, 0.05)  | (0.74, 2.64)  | 0.03  | 1.05  | 4.00  |
| Mean (%)       | 1.35  | 1.58  | 8.94   | (1.28, 1.30)    | (1.44, 3.51)  | (5.41, 2.97)  | 1.04  | 2.72  | 5.27  |

Note: ①PSO-DE-MH; ②DE-MH; ③PSO-DE

Figures 11–14 and Tables 2–4 show that compared with DE-MH and PSO-DE, PSO-DE-MH has the following advantages:

(i) The identification results quickly stabilise after the start of iteration. The identification results are stable after 300 iterations, which are 200 and 1700 iterations less than that of PSO-DE and DE-MH, respectively.

(ii) The maximum probability of sampling is close to the true values. The leakage intensity m, positions (x_0, y_0) and time t_0 of hazardous chemicals are most likely at 0.9 × 10^6–1.1×10^6 g, 4.5–5.5 km, 20.5–21.5 m and 118–122 min.

(iii) The identification results have high computational accuracy. The relative error, sample standard deviation and mean error of the unknown parameters are less than 1.3%, 1.4 and 1.5%, respectively. The leakage intensity m is less than 1%, 0.03 and 1.4%; the leakage position x_0 are less than 0.2%, 0.07 and 1.3%; the leakage position y_0 are less than 1.3%, 0.06 and 1.4%; and those of the leakage time are less than 0.3%, 1.4 and 1.1%, respectively.

(iv) The sampling obtained by PSO-DE-MH are closest to the true values, followed by DE-MH and PSO-DE. The average relative error, average standard deviation and average mean error of the leakage source information which include intensity, position and leakage time by the proposed method are 0.57%, 0.3738 and 1.24%, respectively, which are less than the results obtained by DE-MH (2.11%, 0.4955, 5.65%) and PSO-DE (3.2%, 0.8059, 2.31%).

4.4 Analysis of ME impact on identification results

For SHCLA, the effectiveness of the identification method largely depends on the constructed likelihood function, which is directly related to the distribution of ME. Therefore, the effects of ME on the identification results must be discussed (Lu et al., 2019). The identification results are obtained with limited sampling; thus, the sampling relative error (SRE) can be used to measure the effects of ME on the identification method (Cao et al., 2015) and can be expressed as Equation (21).
where $\sigma$ is the standard deviation of ME, $V$ is the true value, and $n_1$ is the number of sampling.

To analyse the effects of ME on the identification results, we choose five different standard deviations of ME (0.01, 0.05, 0.1, 0.15 and 0.2) to discuss the effects on the identification results. The mean and standard deviation of the leakage intensity $m$, leakage positions $(x_0, y_0)$ and leakage time $t_0$ under the five different standard deviations of ME are calculated respectively by PSO-DE-MH, as shown in Table 5. The relative error and SRE with five different standard deviations of ME are also calculated on the basis of the proposed method, as shown as Figure 15 and Figure 16.

**Table 5** Statistics of pollution sources with five different standard deviations of ME

| Statistics $\sigma$ | Intensity $m$ $(10^6 \text{ g/m}^3)$ | Position $(x_0, y_0)$ (m) | Leakage time $t_0$ (min) | Intensity | Position $(x_0, y_0)$ | Leakage time $t_0$ |
|---------------------|---------------------------------|---------------------------|--------------------------|-----------|----------------------|------------------|
| 0.01                | 0.994                           | (4992.8, 20.7381)         | 119.6473                 | 0.0244    | (0.0611, 0.0546)     | 1.3551           |
| 0.05                | 1.012                           | (5043.5, 20.5816)         | 121.2372                 | 0.0278    | (0.0899, 0.1344)     | 3.3088           |
| 0.10                | 1.0218                          | (5063.1, 20.4840)         | 122.1162                 | 0.0350    | (0.1395, 0.2255)     | 5.8676           |
| 0.15                | 1.0749                          | (5084.1, 21.6775)         | 123.0903                 | 0.0394    | (0.1964, 0.5917)     | 8.3764           |
| 0.2                 | 1.0926                          | (5098.5, 21.9888)         | 124.5197                 | 0.0499    | (0.2956, 0.4181)     | 13.1481          |

**Fig.15.** Variation diagram of the relative errors with different of MEs
As shown in Table 5, Figure 15 and Figure 16, the standard deviation, relative error and SRE of the identification results with five different standard deviations of ME are less than 14, 10% and 2%, respectively. Figure 15 shows that the relative errors of unknown parameters increase with the standard deviations of ME. When the standard deviation of ME is greater than 0.1, the effects on the leakage intensity \( m \) and leakage position \( x_0 \) are maximum and minimum, respectively. Moreover, the relative errors of identification results are greater than 1%. The reason is that when \( \sigma \geq 0.1 \), ME has a significant effect on the information of leakage source through the accept probability of the proposed method. If ME is larger, then the accept probability is close to 1, which leads to the samples being far away from the true values. When \( \sigma < 0.1 \), the effects on the leakage positions \( y_0 \) and \( x_0 \) are maximum and minimum, respectively. Moreover, the relative errors of the identification results are less than 2.5%. The reason is that when \( \sigma < 0.1 \), ME mainly affects the accuracy of the identification results by the proposal distribution interval. If the range of proposal distribution of unknown parameters is small, then the identification results are affected by ME. Figure 16 shows that SRE also increases with the increase of the standard deviation of ME, the increasing speed of SRE is in turn leakage intensity, leakage position and leakage time. Because ME is an important factor affecting the sampling effectiveness of the proposed method by the accept probability.

Although ME is an important factor that affects the accuracy of identification results, PSO-DE-MH remains effective even when the standard deviations of ME increase to 0.2.

**5 Conclusion**

For SHCLA in the surface water, emergency rescue depends on the accuracy and effectiveness of
the identification of leakage sources. Therefore, it is the premise and foundation to build and
design a scientific and reasonable emergency identification model and method to deal with this
kind of accidents effectively. In this study, the emergency identification model is built on the basis
of laws of pollutants and diffusion and Bayesian inference, and a new method called PSO-DE-MH
is designed by integrating the advantages of PSO, DE and MH-MCMC. The proposed method is
tested with an outdoor experiment in a section of SNWTP. The results show that the proposed
method can effectively improve the calculation efficiency and identify the accuracy of source
information for SHCLA. These results can be used for identifying the source information and
establishing emergency response measures. The specific conclusions are as follows:

(i) The emergency identification model is proposed from the perspective of the probability statistic
and provides a new method with strong search ability that avoids the decision risk caused by
‘optimal’ parameter distortion. However, there are some shortcoming in the constructed model,
such as the failure to accurately capture the characteristics of water quality changes and the failure
to fully describe the mechanism of pollutant migration and transformation in water bodies. In any
case, these shortcomings do not prevent the application of this model to the identification of
various source information. The results show that the proposed method can replace the original
model with the identification model based on Bayesian inference in the study of the source
identification of SHCLA. The reliability of this method is greatly affected by sample size,
structural parameters and other factors.

(ii) In contrast to the M-H sampling method under the Bayesian framework, PSO and DE are used
in this study to ensure that the samples can be quickly sampled around the true values. Hence,
PSO-DH-MH does not need to have high approximation accuracy in the entire research area and
only needs to have high accuracy in important areas, especially near the true values. This approach
not only reduces the calculation cost but also improves the accuracy of inversion.

(iii) New samples generated by the proposed method can improve the accuracy of the
identification model. During this process, the ranges of variables to be identified are controlled by
the proposal distribution, and the search space of the identification problem solution is gradually
reduced by the search ability of PSO and the propagation thought of DE, thus decreasing the
selecting cost of new samples and improving convergence.

(iv) One great advantage of PSO-DE-MH is its low computational cost. This study shows that the
identification results obtained by the proposed method quickly stabilise after the start of iteration.
The more accurate the prior information of the leakage source is, the higher the effectiveness of
the sampling by PSO-DE-MH is.

(v) Another great advantage of PSO-DE-MH is high accuracy. The uncertainty of information of
leakage sources increases with the increase in ME. This study shows that the average relative error 
and the average mean error of the leakage source information by PSO-DE-MH are 0.57% and 
1.24% when the standard deviation ME is 1%. Even if the standard deviation of ME is increased 
to 20%, PSO-DE-MH can still accurately identify the leakage source’s information.

In conclusion, compared with the existing models and methods, the proposed identification model 
and method have better noise immunity and thus can be used to solve the leakage source 
identification problem of SHCLAs. Considering the increasing complexity of water environments 
and the diversity of observation data, how to construct a more practical emergency identification 
model based on complex hydrodynamic water quality models and diversity data must be the focus 
of further research.

Ethics approval and consent to participate
Not applicable

Consent for Publication
Not applicable

Availability of data and material
All data generated or analysed during this study are included in this article.

Competing Interests
The authors declare that they have no competing interests.

Funding
This study is supported by the National Social Science Fund of China with Research Fund No. 
17BGL179.

Authors Contributions
All authors contributed to the study conception and design. Material preparation was performed by 
Haidong Yang and Biyu Liu. Data collection and analysis were performed by Jinjin Li and Luying 
Chen. Software design and test were performed by Haidong Yang and Jinjin Li. The first draft of 
the manuscript was written by Haidong Yang and all authors commented on previous versions of 
the manuscript. All authors read and approved the final manuscript.
Acknowledgements

The authors are very thankful for constructive comments and suggestions from anonymous reviewers and editors.

Reference

Aster RC, Borchers B, Thurber C (2005) Parameter Estimation and Inverse Problems (International Geophysics). Academic Press. ISBN 0120656043. URL http://www.ees.nmt.edu/outside/courses/GEOP529_book.html

Alapati S, Kabala ZJ (2000) Recovering the release history of a groundwater contaminant using a non-linear least-squares method. *Hydrological Processes*, 14(6), 1003 – 1016. http://doi.org/10.1002/(SICI)1099-1085(20000430)14:6<1003::AID-HYP981>3.0.CO;2-W

Ali MM, Törn A (2004) Population set-based global optimization algorithms: some modifications and numerical studies. *Computers & Operations Research*, 31(10), 1703-1725. http://doi.org/10.1016/S0305-0548(03)00116-3

Andrle M, El Badia A (2015) On an inverse source problem for the heat equation application to a pollution detection problem. ii. *Inverse Problems in Science & Engineering*, 23(3), 389-412. http://doi.org/10.1515/iip.2002.10.6.585

Ashyralyev A, Emharab F (2018) Source identification problems for hyperbolic differential and difference equations. *Journal of Inverse and Ill-Posed Problems*, 27(3), 301-315. http://doi.org/10.1515/jipip-2018-0020

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

Bergey PK, Ragsdale C (2005) Modified differential evolution: a greedy random strategy for genetic recombination. *Omega*, 33(3), 255–265. http://doi.org/10.1016/j.omega.2004.04.009

Brooks SP, Roberts GO (1998) Assessing convergence of Markov chain Monte Carlo algorithms. *Statistic & Computing*, 8(4), 319–335. http://doi.org/10.1103/PhysRevE.72.016715

Cao HG, Yun WG (2017) Research of the abrupt waters pollution source based on optimization algorithm of PSO-DE. *China Environmental Science*, 37(10), 3807-3812.

Cao R, José A, Vilar, JM, et al (2015) Sampling Error Estimation in Stratified Surveys. *Open Journal of Statistics*, 3(3), 200-212. http://doi.org/10.4236/ojs.2013.33023

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady state conditions: 1. maximum likelihood method incorporating prior information. *Water Resources Research*, 22(2), 199-210. http://doi.org/10.1029/WR022i002p00199

Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms. *Water Resources Research*, 22(2), 211-227. http://doi.org/10.1029/WR022i002p00211

Chen LP, Jiang JC, Yin L (2007) Numerical simulation water pollution diffusions for sudden hazardous chemical. *Journal of Hydrodynamics (Ser. A)*, 22(6), 761-765. http://doi.org/10.1007/s10483-007-0101-x

Datta B, Chakrabarty D, Dhar A (2011) Identification of unknown groundwater pollution sources using classical optimization with linked simulation. *Journal of Hydro-environment Research*, 5(1), 25–36. http://doi.org/10.1016/j.jher.2010.08.004

Eberhart RC, Kennedy J (1995a) A new optimizer using particle swarm theory. *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, IEEE, 39-43.

Eberhart RC, Kennedy J (1995b) Particle swarm optimization. *Proceedings of the IEEE conference on Neural Network, IEEE*, 1942-1948.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. http://doi.org/10.2307/2246093
Ghane A, Mazaheri M, Jamal MVS (2016) Location and release time identification of pollution point source in river networks based on the Backward Probability Method. Journal of Environmental Management, 180, 164-171. http://doi.org/10.1016/j.jenvman.2016.05.015

Guan J, Aral M, Maslia M, Grayman W (2006) Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. Journal of Water Resources Planning & Management, 132(4), 252-262. http://doi.org/10.1061/(ASCE)0733-9496(2006)132:4(252)

Guo S, Yang R, Weng W (2009) Source inversion of toxic gas dispersion in urban areas based on the MCMC method. Journal of Tsinghua University (natural science edition), 49(5), 629-634.

Guo ZW, Zhang C, Li, Y, et al (2016) Source identification of sudden contamination based on the parameter uncertainty analysis. Journal of Hydro Informatics, 18(6), 919-927. http://doi.org/10.2166/hydro.2016.002

Gupta S, Singh RK, Sharan SN (2016) An approach to implement PSO to optimize outage probability of coded cooperative communication with multiple relays. Alexandria Engineering Journal, 55(3), 2805-2810. http://doi.org/10.1016/j.aej.2016.07.018

Hamdi, Adel (2007) Identification of point sources in two-dimensional advection-diffusion-reaction equation: application to pollution sources in a river stationary case. Inverse Problems in Science & Engineering, 15(8), 855-870. http://doi.org/10.1080/17415970601162198

Han JC, Huang GH, Zhang H, Li Z, Li YP (2014) Bayesian uncertainty analysis in hydrological modeling associated with watershed subdivision level: a case study of slurry model applied to the xiangxi river watershed, China. Stochastic Environmental Research and Risk Assessment, 28(4), 973-989. http://doi.org/10.1007/s00477-013-0792-0

Hazart A, Giovannelli JF, Dubost S, Chatellier L (2014) Inverse transport problem of estimating point-like source using a Bayesian parametric method with MCMC. Signal Processing, 96, 346-361. http://doi.org/10.1016/j.sigpro.2013.08.013

Hou J, Jiang B (2017) Predictions and estimations under a group of linear models with random coefficients. Communication in Statistics-Simulation and Computation, 47(1-2), 510-525. http://doi.org/10.1080/03610918.2017.1283704

Hu Y, Zhang WJ, Ren HT, et al (2019) Adjoint method for source term inversion in reverse process of pollutant convection diffusion. Advances in Science and Technology of Water Resources, 39(2), 7-11.

Huang LX, Wang LC, Zhang YY, Xing LT, Hao QC, Xiao Y, Zhu HH (2018) Identification of groundwater pollution sources by a SCE-UA algorithm-based simulation/optimization model. Water, 10(2), 193. http://doi.org/10.3390/w10020193.

Jain NK, Nangia U, Jain J (2018) A Review of Particle Swarm Optimization. Journal of the Institution of Engineers, 99(4), 1-5. http://doi.org/10.1007/s40031-018-0323-y

Jiang JP, Dong FJ, Liu RT (2017) Applicability of Bayesian inference approach for pollution source identification of river chemical spills: a tracer experiment based analysis of algorithmic parameters, impacts and comparison with Frequentist approaches. China Environmental Science, 37(10), 3813-3825.

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

Jing P, Yang Z, Zhou W, Huai W, Lu X (2020) Inverse estimation of finite-duration source release mass in river pollution accidents based on adjoint equation method. Environmental science and Pollution Research, 27(13), 14679-14689. http://doi.org/10.1007/s11356-020-07841-1

Jana A, Nair MT (2020) A truncated spectral regularization method for a source identification problem. Journal of Analysis, 28, 279-293. http://doi.org/10.1007/s41478-018-0080-y

Jun K, Kang J, Lee K (2007) Simultaneous estimation of model parameters and diffuse pollution sources for river...
water quality modeling. Water Science and Technology, 56(1), 155–162. http://doi.org/10.2166/wst.2007.447
Kastner CA, Braumann A, Man PLW, Mosbach S, Brownbridge GPE, Akroyd J, et al (2013) Bayesian parameter
estimation for a jet-milling model using metropolis-hastings and wang-landau sampling. Chemical Engineering
Science, 89, 244-257. http://doi.org/10.1016/j.ces.2012.11.027
Lei Y, Zhou X, Xie, L (2019) Emergency monitoring and disposal decision support system for sudden pollution
accidents based on multimedia information system. Multimedia Tools and Applications, 78(8), 11047-11071.
http://doi.org/10.1007/s11042-018-6665-2
Liu X D, Wang, J (2020) Advances in methods for identifying surface water pollution sources. Advances in Water
Science, 31(2), 302-311.
Lu ZR, Zhou J, Wang L (2019) On choice and effect of weight matrix for response sensitivity-based damage
identification with measurement and model errors. Mechanical Systems & Signal Processing, 114, 1-24.
Ma D, Tan W, Zhang Z, Hu J (2017) Parameter identification for continuous point emission source based on
tikhonov regularization method coupled with particle swarm optimization algorithm. Journal of Hazardous
Materials, 325, 239-250. http://doi.org/10.1016/j.jhazmat.2016.11.071
Mohamed AW, Sabry HZ (2012) Bayesian parameter uncertainty modeling in a macroscale hydrologic model and its
effect of weight matrix for response sensitivity based linked simulation-optimization method. Chinese Journal of Environmental Science, 40(4), 1698-1705.
Parolin RDS, Silva N, Antônio J, Rodrigues, PPGW, Llanes Santiago O (2015) Estimation of a contaminant source
in an estuary with an inverse problem approach. Applied Mathematics & Computation, 260, 331-341.
Qin AK, Huang VL, Suganthan, PN (2009) Differential evolution algorithm with strategy adaptation for global
numerical optimization. IEEE Transactions on Evolutionary Computation, 13(2), 398-417.
Raje D, Krishnan R (2012) Bayesian parameter uncertainty modeling in a macroscale hydrologic model and its
impact on Indian river basin hydrology under climate change. Water Resources Research, 48(8), 2838-2844.
Roberts GO, Rosenthal JS (2004) General state space Markov Chains and MCMC algorithms. Probability
Surveys, 1(1), 20-71. http://doi.org/10.1214/15495780410000024
Rönkkönen J, Kukkonen S, Price KV (2005) Real-parameter optimization with differential evolution. IEEE
Congress on Evolutionary Computation, 1, 506 - 513. http://doi.org/10.1109/CEC.2005.1554725
Singh RM, Datta B (2006) Identification of groundwater pollution sources using GA-based linked simulation
optimization model. Journal of Hydrologic Engineering, 11(2), 1216-1227.
Skaggs TH, Kabala ZJ (1995) Recovering the history of a groundwater contaminant plume: method of
quasi-reversibility. Water Resources Research, 31(11), 2669-2673. http://doi.org/10.1029/95wr02383
Skaggs TH, Kabala ZJ (1994) Recovering the release history of a groundwater contaminant. Water Resources
Research, 30(1), 71-79. http://doi.org/10.1029/93WR02656
Skiba YN, Parra-Guevara D (2017) Application of adjoint approach to oil spill problems. Environmental Modeling
& Assessment, 22(4), 379-395. http://doi.org/10.1007/s10666-016-9540-4
Storn R, Price K(1997) Differential evolution--a simple and efficient heuristic for global optimization over
continuous spaces. Journal of Global Optimization, 11(4), 341–359. http://doi.org/10.1023/A:1008202812138
Sun J, Zhang WS, Jing L, Han X, Wang Q, Kong J (2019) Estuarine pollutant tracing method based on

- 31 -
evolutionary algorithm and water quality model. Science, Technology and Engineering, 19(14), 384-391.
2. Tang Y, Guan X (2007) Parameter estimation for time-delay chaotic system by particle swarm optimization. Chaos Solitons & Fractals, 40(3), 1391-1398. http://doi.org/10.1016/j.chaos.2007.09.055
3. Venter G, & Sobieszczanski-Sobieski J (2004) Multidisciplinary optimization of a transport aircraft wing using particle swarm optimization. Structural and Multidisciplinary Optimization, 26(1), 121-131. http://doi.org/10.1007/s00158-003-0318-3.
4. Vitaliy F (2006) Differential evolution in search of solutions. Springer, New York. http://doi.org/10.1007/978-0-387-36896-2.
5. Wang J, Shih Y, Wang PY, Yu YH, Huang CP (2019) Hazardous wastes treatment technologies. Water Environment Research, 91(5), 1177-1198. http://doi.org/10.1002/wer.1213.
6. Wang J, Zhao S, Li X, Cai F, Zhang L. (2016). Comparison of three methods for assessing damage of water resource in sudden water pollution accident. Environmental Science & Technology, 39(03), 200-204.
7. Wang Y, Zhang W, Engel BA, Peng H, Theller L, Shi Y, et al (2015) A fast mobile early warning system for water quality emergency risk in ungauged river basins. Environmental Modelling & Software, 73, 76-89. http://doi.org/10.1016/j.envsoft.2015.08.003
8. Wei H, Chen W, Sun H, Li X (2010) A coupled method for inverse source problem of spatial fractional anomalous diffusion equations. Inverse Problems in Science & Engineering, 18(7), 945-956. http://doi.org/10.1080/17415977.2010.492515
9. Wu W, Ren J, Zhou X, Wang J, Guo M (2020) Identification of source information for sudden water pollution incidents in rivers and lakes based on variable-fidelity surrogate-dream optimization. Environmental Modelling & Software, 104811. http://doi.org/10.1016/j.envsoft.2020.104811
10. Yang HD, Liu BY, Huang JH (2018) Forecast model parameters calibration method for sudden water pollution accidents based on improved Bayesian-Markov chain Monte Carlo. Control and Decision, 33(4), 679-686. http://doi.org/10.13195/j.kzyjc.2017.0213
11. Yang HD, Shao DG, Liu BY, et al (2016) Multi-point source identification of sudden water pollution accidents in surface waters based on differential evolution and Metropolis-Hastings-Markov Chain Monte Carlo. Stochastic Environmental Research and Risk Assessment, 30(2), 507-522. http://doi.org/10.1007/s00477-015-1191-5
12. Yang YL, Lei XH, Long Y, Tian Y, Zhang YH, Yao Y, Hou XS, Shi MS., Wang PW, Zhang CL, Wang H, Quan YL, Lei XH, Long Y, Tian Y, Zhang YH, Yao Y, Hou XS, Shi MS., Wang PW, Zhang CL, Wang H, Quan Y (2020) A novel comprehensive risk assessment method for sudden water accidents in the middle route of the north-south water transfer project (china). Science of the Total Environment, 698, 134167-.
13. Wang XD, Zhang DQ, Chen LY (2018) Real-time monitoring of pollutant diffusion states and source using fuzzy adaptive kalman filter. Water, Air, & Soil Pollution, 229: 238. https://doi.org/10.1007/s11270-018-3885-z
14. Yoo B, Choi SD (2019) Emergency evacuation plan for hazardous chemicals leakage accidents using GIS-based risk analysis techniques in South Korea. International Journal of Environmental Research and Public Health, 16(11), 1948. http://doi.org/10.3390/ijerph16111948
15. Yu HL, Wu YZ, Cheung SY (2020) A data assimilation approach for groundwater parameter estimation under Bayesian maximum entropy framework. Stochastic Environmental Research and Risk Assessment, 34(2), 709-721. http://doi.org/10.1007/s00477-020-01795-z
16. Yu LA, Li L, Wu JQ, et al (2015) Emergency policy exploration for network public opinion crisis in water pollution accident by hazardous chemicals leakage based on systematic dynamics. Systems Engineering-Theory & Practice, 35(10), 2687-2697. http://doi.org/CNJLSUN-XTLL.0.2015-10-029
17. Zeuner S, Günter M (2020) Influence of the spatial and temporal monitoring design on the identification of an instantaneous pollutant release in a river. Advances in Water Resources, 146, 103788.
Zhang SP, Xin XK (2017) Pollutant source identification model for water pollution incidents in small straight rivers based on genetic algorithm. *Applied Water Science*, 7(4), 1955-63. https://doi.org/10.1007/s13201-015-0374-z

Zhang GX, Yang RH, Meng XF (2018) Construction of the emergency decision system for hazardous chemicals leakage based on "scenario-response" model. *Environmental Protection of Chemical Industry*, 38(5), 609-614.

Zhao L, Qian Y, Hu QM, Jiang R, Li M, Wang X (2018) An analysis of hazardous chemical accidents in China between 2006 and 2017. *Sustainability*, 10(8), 1-18. https://doi.org/10.3390/su10082935

Zhang S, Qiang J, Liu H, Li Y (2020) Optimization design of groundwater pollution monitoring scheme and inverse identification of pollution source parameters using bayes' theorem. *Water, Air, & Soil Pollution*, 231(1), 1-17. https://doi.org/10.1007/s11270-019-4369-5