Inverse problem of an instantaneous pollution source in a wide and shallow river and the analysis on inversion accuracy

L. X. Han[1,2], Y. Zhu[2], W. L. Jin[2] & F. X. Zhang[2]
[1]*Key Laboratory of Shallow Lakes of Ministry of Education, Hohai University, China*
[2]*College of Environment, Hohai University, China*

Abstract

According to the hydrodynamic characteristics of a wide and shallow river and the transfer characteristics of soluble pollutants, an inverse problem of instantaneous source in a wide and shallow river is established to confirming the pollution source intensity and location. A mathematical solving method is set out by using a differential evolution algorithm. A typical case is set out and the solving method is verified by a two-dimensional theoretical solution. Based on the scene analysis method, further analysis is made to analyze the influence of monitoring sites, monitoring frequency and monitoring accuracy on inversion accuracy.
Keywords: wide and shallow river, instantaneous pollutant source, source intensity, inverse problem, inversion accuracy.

1   Introduction

In recent years, sudden pollution accidents of surface water have occurred frequently for some accidents (such as shipping chemicals leakage), due to their indeterminacy, the occurrence and intensity of the sources or some other basic parameters which probably cannot be determined earlier resulting in warning and treatment for the accidents becoming more difficult. So, identifying the pollution source information scientifically and accurately is the key foundation work for early warning and emergency treatment in sudden water pollution accidents.

When a water pollution accident occurs, the pollution source information can be worked backwards from the emergency monitoring data of water quality,

which, according to the mathematical language, uses some information of the pollutants' temporal and spatial distribution to reverse the source term in the mathematical model transport and transformation of pollutants. In the subject of environmental hydraulics, it can be divided into the inverse problem of source term identification (Liu *et al.* [1]). This research is being given more and more attention by researchers. For example, Akcelika *et al.* [2] used the variational finite element method to solve the source inversion problem for convective-diffusive transport; Wang and Zabaras [3] put forward the Markov random field model for source identification in flows through porous media; Cheng and Jia [4] used the location probability density function method to identify the contaminant point source in surface waters; Han and Zhu [5] structured and solved a source inversion problem with constraint conditions in a river network area; Chen *et al.* [6] used the correlation coefficients optimization method to reverse the point instantaneous source in one-dimensional river flow; Zhu *et al.* [7] did the research of pollution source identification and model parameter inversion by the Bayesian approach; Wang and Qiu [8] researched the uniqueness, stability and inversion algorithm of the source term identification problem in a basin. In recent years, some modern optimization algorithm such as the genetic algorithm was widely used in solving the source term inverse problem, Min *et al.* [9] transformed the inverse problem of source term identification of a convection-diffusion equation to an optimization problem and solved the problem by using the genetic algorithm; Han [10] used sewage treatment cost minimum as the objective function and solved the inverse problem of pollutant emissions with constraint conditions by using the genetic algorithm; Jae *et al.* [11] used the genetic algorithm to establish a river water quality management model for optimizing regional wastewater treatment. The differential evolution algorithm as a relatively new optimization algorithm is also beginning to be used for the source term identification of surface water pollution.

Generally speaking, previous researches mainly focus on control and identification of the pollution source in the case of knowing the pollution discharge position and lack of the research and discussion about the influence of emergency monitoring scheme, monitoring data accuracy of water quality to the accuracy of source term inversion.

Based on the transport and transformation characteristics of a sudden instantaneous source in a wide and shallow river and according to the downstream monitoring data of water quality, this paper structures an inverse problem of instantaneous source term identification and inverses the occurrence and intensity of the source by a differential evolution algorithm. The influence of the monitoring plan (including monitoring sites layout, monitoring frequency) and monitoring error level is also discussed in the paper, which can provide an important technical reference to the formulation and optimization of an accident emergency monitoring plan.

2 Two-dimensional water quality transportation and diffusion model of instantaneous source in a wide and shallow river

Assuming that the waters of the accident is a river which can be called wide and shallow because its width–depth ratio is big enough, the background concentration of pollutants 0, at the time of $t=0$, on the two-dimensional coordinates (x_m, y_m), the pollutants instantaneous pull into the water which weighs M, so the change law of pollutants concentration increment C changing with time and space can be described by the two-dimensional convection – diffusion equation:

$$\begin{cases} \dfrac{\partial C}{\partial t} = \dfrac{\partial}{\partial x}(D_x \dfrac{\partial C}{\partial x}) + \dfrac{\partial}{\partial y}(D_y \dfrac{\partial C}{\partial y}) - \dfrac{\partial(u_x C)}{\partial x} - \dfrac{\partial(u_y C)}{\partial y} - KC + M\delta(x - x_m, y - y_m, t) \\ \text{Initial Condition:} C(x,y,t)\big|_{t=0} = 0 \\ \text{Boundary Condition:} C(x,y,t)\big|_{x=0} = C_0, \dfrac{\partial C(x,y,t)}{\partial x}\bigg|_{x=L} = 0 \end{cases} \quad (1)$$

where: u_x, u_y is vertical and horizontal velocity component; D_x, D_y is vertical and horizontal diffusion coefficient of pollutants; K is the degradation coefficient of pollutants; $M\delta(x - x_m, y - y_m, t)$ is the source term, where M is instantaneous source emissions, x_m, y_m are the coordinates of the point source in the two-dimensional space; $\delta(\bullet)$ is the Dirac Function.

Eqn (1) constitutes a two-dimensional water quality simulation mathematical model of instantaneous source in a wide and shallow river, which is called a forward problem, its theoretical solution is:

$$C(x,y,t) = \frac{M}{4\pi ht\sqrt{D_x D_y}} \exp\left[-\frac{(x - x_m - u_x t)^2}{4D_x t} - \frac{(y - y_m - u_y t)^2}{4D_y t} \right] \exp(-Kt) \quad (2)$$

3 The construction of source term inverse problem and the mathematical solution method

3.1 Construction of the mathematical inverse problem of instantaneous source in a wide and shallow river

In sudden water pollution accidents, it is often difficult to monitor the occurrence and intensity of the pollution source in time, which means M and x_m, y_m in eqn (1) are unknown. So we cannot use the water quality model to forecast the water quality impact of an accident at an earlier time. In order to identify the source term, we can arrange several monitoring sites downstream of the accident, so that

we can get a group of monitoring data of pollutants concentration, which is described in mathematical language as:

$$C(x,y,t) = C_{obs}(x_i, y_i, t_j) \quad (i = K, j = 1, 2, \cdots N_i) \tag{3}$$

where: $C_{obs}(x_i, y_i, t_j)$ is the monitoring data of monitoring point "i" and monitoring moment "j"; K is the total number of monitoring sites; N_i is the total number of monitoring times on the monitoring point "i".

Through the monitoring pollutant concentrations to determine the source intensity M and source location (x_m, y_m), we can constitute an inverse problem of instantaneous source identification in a wide and shallow river. Assume that S is an undetermined vector function constituted by M and (x_m, y_m), denote the solution of definite problem eqn (1) which corresponds to S as $C(S, x, y, t)$, then the above inverse problem can be converted to an optimization problem as follows:

$$\min \sum_{i=1}^{K} \sum_{j=1}^{N_i} \left[C_{obs}(x_i, y_i, t_j) - C_{obs}(S, x_i, y_i, t_j) \right]^2 \tag{4}$$

That is: using the square sum of the difference between the monitoring value and calculated value at a different monitoring point and monitoring moment as the objective function, calculate the optimal solution which can meet the conditions.

3.2 Solution method of the inverse problem of source term identification: differential evolution algorithm

A differential evolution algorithm uses the evolution law of "natural selection, survival of the fittest", belonging to the evolutionary algorithm. It can solve a discontinuous coefficients identification problem, including the partial differential equation inverse problem of source term identification, parameter identification, and initial conditions identification, etc. Using a differential evolutionary algorithm to solve a partial differential equation inverse problem basically does not depend on the knowledge of search space or other auxiliary information, and the individuals are evaluated only by a specific fitness function, which drives the individuals to evolve, and there are no particularly stringent requirements to the fitness function itself.

3.3 The example of a source term inverse problem and solving method verification

As shown in fig. 1, we can assume that the study area in the wide and shallow river is 3000m long and 150m wide, the average depth h=1.0m, the average flow velocity u=0.65m/s, the dispersion coefficient D_x=4.0m^2/s, D_y=0.5 m^2/s. Assume that a pollution accident occurred at the location of x_m = 500m, y_m = 50m at t=0 moment, the pollutants were instantly pulled into the water, its weight M=2000kg, degradation coefficient K=0.0001/d.

Monitoring scheme: we can assume that the monitoring time is 1200~3700 seconds after the accident occurred, the monitoring location is A1(2000, 40), A2(2000, 80), A3(2000, 120), the monitoring frequency is every 500 seconds and all monitoring points monitor 6 times.
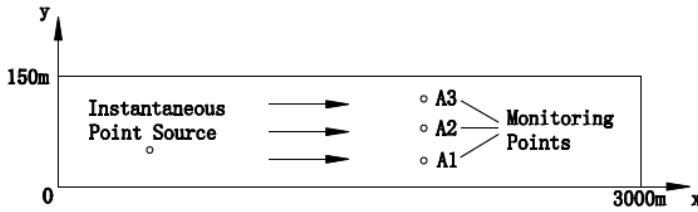
Figure 1: Schematic diagram of a calculating example of inversion in a wide and shallow river.

Monitoring error: use the analytical solution of direct problem eqn (2) to calculate the pollutant concentration at monitoring point "i" and monitoring moment "j" called $C'_{obs}(x_i, y_i, t_j)$ (i=1, 2,…, j, j=1, 2,…, N_i, K is the total number of monitoring points, N_i is the monitoring frequency at monitoring point "i") as the theoretical value of monitoring data. Considering the monitoring error, a disturbance should be appended to the theoretical value; the monitoring data which appended the monitoring error is:

$$C_{obs}(x_i, y_i, t_j) = C'_{obs}(x_i, y_i, t_j) + e \cdot \omega_{i,j} \cdot C'_{obs}(x_i, y_i, t_j) \tag{5}$$

where: $C_{obs}'(x, y, t)$ is the theoretical value of monitoring data which is obtained from the analytical solution of the direct problem; $C_{obs}(x, y, t)$ is the monitoring data which is appended to the monitoring error; e is the monitoring error level, e=0.05; $\omega_{i,j}$ is the random number which conforms to the standard normal distribution.

Parameter selection: the population size is 20, crossover probability is 0.1, crossover factor is 0.5, and evolutionary generations are 500. Use the differential evolution algorithm program to calculate, the inversion result is shown in table 1.

Table 1: Inversion result of source intensity and location.

	M/kg	x_m/m	y_m/m
use the monitoring data without error	2000.000	500.000	50.000
use monitoring data which error	2016.823	497.008	52.485
the theoretical value	2000	500	50

The inversion results show that when using the monitoring data without error which directly gets the analytical solution of the direct problem, the inversion results are the same as the theoretical value; when using the monitoring data with some error, the inversion results have some small error, but can still reflect the realities of the pollution sources. So, the solving method can inverse the source intensity and location accurately.

4 The influence of the monitoring scheme and monitoring accuracy to the inversion accuracy

The accurate inversion of pollution source after the water pollution has happened depends on the accurate description of water quality at the monitoring points. So the monitoring scheme (the monitoring point's number and the monitoring frequency) and the monitoring accuracy have important influences on the inversion accuracy.

4.1 The influence of the monitoring point's number to the inversion accuracy

In order to analyze the influence of the monitoring point's number to the inversion accuracy independently, we can use the scenario analysis method. On the premise of setting the same starting and ending time (1200s–3700s) and monitoring frequency (every 500s), respectively assume three kinds of working conditions: setting only one monitoring point (A_2 (2000, 80)), setting two monitoring points (A_1 (2000, 40) and A_2 (2000, 80)) and setting three monitoring points (A_1 (2000, 40), A_2 (2000, 80), A_3 (2000, 120)), use the numerical experiments to quantitatively analyze the influence of the monitoring point's number to the inversion accuracy. For researching the problem easily, we may not consider the monitoring error and describe the inversion accuracy of source intensity by the relative error E_M and source location by the absolute error E_x, E_y.

The inversion accuracy calculated by different monitoring point's number is shown in table 2.

Table 2:   Inversion result of different numbers of monitoring sites

monitoring point	M/kg	E_M	x_m/m	E_x	y_m/m	E_y
only A_2	5058.872	1.529	497.770	2.230	87.061	37.061
A_1 and A_2	2000.079	0.000	499.997	0.003	50.001	0.001
A_1, A_2, A_3	2000.000	0.000	500.000	0.000	50.000	0.000

The inversion results in table 2 show that there is a big inversion error when setting only one monitoring point, and cannot correctly reflect the reality of the pollution sources; but when setting two or more than two monitoring points, the inversion results are very close to the reality of pollution sources.

4.2 The influence of the monitoring frequency to the inversion accuracy

On the premise of setting the same monitoring points (A_1 and A_2) and the same starting and ending time (1200s–3700s), respectively assume that the monitoring frequency is monitoring 2, 3, 4, 5, 6 times and without monitoring error, use these five working conditions to quantitatively analyze the influence of the monitoring frequency to the inversion accuracy. The results are shown in table 3.

The inversion results show that, in the case of setting two monitoring points, when monitoring less than 4 times, the inversion results have a large error and

cannot correctly reflect the reality of the pollution sources; when monitoring 4 times, the inversion results are more accurate. In general, under the condition of fixed monitoring points, the higher the monitoring frequency, the higher the inversion accuracy. So when making a monitoring scheme, the monitoring frequency should be increased as far as possible to ensure the reliability of inversion results.

Table 3:   Inversion result of different monitoring frequencies.

Monitoring times	Monitoring time interval/s	M/kg	E_M	x_m/m	E_x	y_m/m	E_y
2	2400	13020.245	5.510	562.950	62.950	49.029	0.071
3	1200	4784.767	1.392	245.867	244.133	49.241	0.759
4	800	2026.784	0.013	498.889	1.111	49.988	0.012
5	600	1998.766	0.001	499.947	0.053	49.976	0.024
6	400	2000.000	0.000	500.000	0.000	50.000	0.000

4.3 The influence of the monitoring error level to the inversion accuracy

When implementing the emergency monitoring scheme for a pollution accident, the water quality monitoring data is bound to have a certain error or disturbance. So after the theoretical value of monitoring data is generated by the analytical solution of direct problem, a monitoring error should be added and the new data can be used as the emergency monitoring data. That is, first use the standard normal distribution program to generate a set of random numbers called ω_q, and then give different values to the error level e, and substitute it into eqn (5) to structure the emergency monitoring data.

Through calculation, the trend of relative error of source intensity E_M with the change of e is shown in fig. 2.
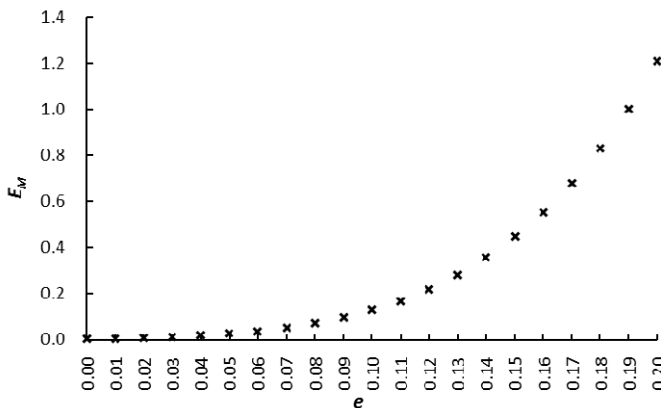

Figure 2:   Trend of E_M with the change of e.

In addition, in order to quantitatively analyze the relationship between the inversion accuracy of source location and the monitoring accuracy, we can calculate the absolute distance between the inversion source position and the real source location called L, shown in the formula below.

$$L = \sqrt{\left(x_{true} - x_{estimated}\right)^2 + \left(y_{true} - y_{estimated}\right)^2} \tag{6}$$

Through calculation, the trend of L with the change of e is shown in fig. 3.
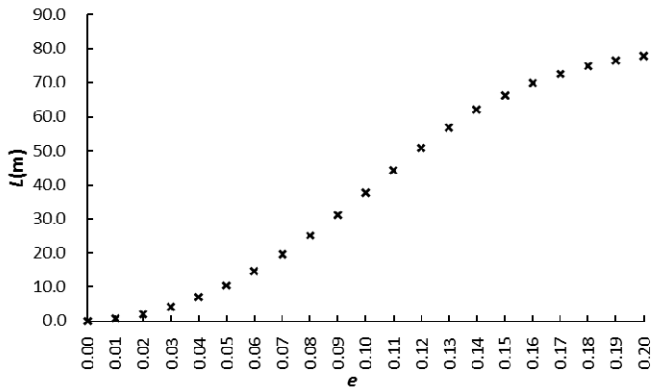

Figure 3:   Trend of L with the change of e.

Figs 2 and 3 show that E_M increases with e increasing and the growth rate was accelerating; L is also, but its growth rate has certain fluctuations. It may be due to randomness of the random number generated by the computer. It can be seen that when using the differential evolution algorithm to inverse the pollution source, the lower the monitoring date accuracy, the lower the inversion accuracy and the higher the monitoring date accuracy, the higher the inversion accuracy. So we should reduce the monitoring error as far as possible, so as to improve the inversion accuracy and guarantee the reliability of the inversion result.

5   Conclusions

For a wide and shallow river, based on the two-dimensional water quality model, this paper constructs an inverse problem of the pollution source's intensity and location, gives the solution mathematical method and verifies it. The numerical experimental study shows that the inversion accuracy depends on the influencing factors such as the number of monitoring points, monitoring frequency and the water quality monitoring accuracy. The mathematical inversion of pollution source based on a certain monitoring scheme and monitoring accuracy can provide technical support for determination of the instantaneous source's intensity and location. In this paper, the study method we used is basing the

direct problem of theoretical analytical solutions for the water quality model, but also applies to the inverse problem constructed by the water quality model solved by the numerical method.

Acknowledgement

References

[1] Liu, X.D., Yao, Q., Xue, H.Q., *et al.*, Advance in inverse problems of environmental hydraulics. *Advances in Water Science*, **20(6)**, pp. 885-890, 2009 (in Chinese).

[2] Akcelika, V., Biros, G., Ghattas, O., *et al.*, A variational finite element method for source inversion for convective-diffusive transport. *Finite Elements in Analysis and Design*, **39**, pp. 683-705, 2003.

[3] Wang, J. & Zabaras, N., Using Bayesian statistics in the estimation of heat source in radiation. *International Journal of Heat and Mass Transfer*, **48**, pp. 15-29, 2005.

[4] Cheng, W.P. & Jia, Y., Identification of contaminant point source in surface waters based on backward location probability density function method. *Advances in Water Resources*, **33(4)**, pp. 397-410, 2010.

[5] Han, L.X., & Zhu, D.S., The pollutant source control method for network water environment management. *Journal of Hydraulic Engineering,* **(10)**, pp. 28-31, 2001 (in Chinese).

[6] Chen, Y.H., Wang, P., Jiang, J.P., *et al.*, Contaminant point source identification of rivers chemical spills based on correlation coefficients optimization method. *China Environmental Science*, **31(11)**, pp. 1802-1807, 2011 (in Chinese).

[7] Zhu, S., Liu, G.H., Wang, L.Z., *et al.*, A bayesian approach for the identification of pollution source in water quality model coupled with hydrodynamics. *Journal of Sichuan University*, **41(5)**, pp. 30-35, 2009 (in Chinese).

[8] Wang, Z.W. & Qiu, S.F., Stability and numerical simulation of pollution point source identification in a watershed. *Chinese Journal of Hydrodynamics*, **23(4)**, pp. 364-371, 2008 (in Chinese).

[9] Min, T., Zhou, X.D., Zhang, S.M., *et al*, Genetic algorithm to an inverse problem of source term identification for convection-diffusion equation. *Chinese Journal of Hydrodynamics*, **19(4)**, pp. 520-524, 2004 (in Chinese).

[10] Han, L.X., Inverse problem on amount of pollutant into natural channel. *Advances in Water Science*, **12(1)**, pp. 39-44, 2001 (in Chinese).

[11] Jae, H.C., KI, S.S. & Sung, R.H., A river water quality management model for optimizing regional wastewater treatment using a genetic algorithm. *Journal of Environmental Management*, **73(3)**, pp. 229-242, 2004.