

A geostatistical approach to groundwater pollution source identification considering first-order reaction

Yu-qiao Long^{a,b}, Ting-ting Cui^{a,b,*}, Wei Li^{a,b}, Yong-wei Gai^c

^aDepartment of Hydrology and Water Resources, Nanjing Hydraulic Research Institute, Nanjing 210029, China, email: ttcui@nhri.cn (T.-t. Cui)

^bState Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing 210098, China

^cWater Resources Service Center of Jiangsu Province, Nanjing 210029, China

Received 7 November 2017; Accepted 4 February 2018

ABSTRACT

Chemical reaction is a very important factor of the groundwater pollution source identification (PSI). However, the PSI method based on the geostatistics is always applied on the conservative pollutants. In this paper, the finite difference is employed to obtain the transfer function of complex transference of pollutant in groundwater, and a PSI method considering the first-order reaction is proposed. A numerical test is employed to analyse the result of the new method and the impact of reaction rate on the PSI problem. In the case, the new method could identify the release process perfectly. Accurate PSI result could be obtained under high concentration or low chemical reaction consumption of pollution. Though the PSI result is insensitive to the reaction rate when the reaction rate is between 10^{-4} and 10^{-3} , the more accurate reaction rate is still very important for the PSI problem. The method prompted in this paper has good agreement with the transport rule of pollutant, and could be very helpful for identifying groundwater pollution.

Keywords: Geostatistical; Groundwater; Pollution; Identification; First-order reaction

1. Introduction

Pollution source identification (PSI) refers to reconstructing the pollution source locations and releasing histories from observed concentration records [1]. As one of the first steps in environmental remediation project, PSI can be classified into three typical types [2]: namely finding the release history of a source, finding the location of a source and recovering the initial distribution of a contaminant plume. The PSI is helpful to making a cost-effective remediation strategy, partitioning the cleanup cost among liable parties [3].

The mathematical and simulation approaches of PSI have been extensively investigated in the past 30 years. Atmadja and Bagtzoglou [4] have subdivided the existing mathematical methods into four major groups, namely optimization, analytical and direct methods as well as probabilistic and

geostatistical approaches. Snodgrass and Kitanidis [5] used a probabilistic approach combining Bayesian theory and geostatistical techniques to estimate the pollution source function. The method is an improvement from some other methods in that the solutions are more general and make no blind assumptions about the nature and structure of the unknown source function. Limitation to this approach is that the location of the potential source must be known a priori [4]. Butera and Tanda [6] use the method to find the source function in a 2D problem. Michalak and Kitanidis [7] combine the method with the adjoint state method to identify the source function in a 3D problem. Butera et al. [8] extend the method to find both the source function and location. Though this method is extensively studied, most of these researches give their attention on the conservative pollutants, the PSI method considering the chemical reaction is seldom discussed.

The groundwater system could be taken as a chemical reaction system [9], the sorption/desorption and chemical reaction could be significant factor for PSI. Some

* Corresponding author.

biodegradation, radioactive decay, hydrolysis process could be taken as first-order reaction [10]. The first-order reaction kinetics equation is the common equation in contamination analysis [11]. In this paper, we give our effort on finding the source release history in a 1D homogeneous aquifer based on the geostatistical approach considering first-order reaction. Numerical experiment is carried out to test the method and evaluate the PSI result.

2. Theory

2.1. Transfer function considering the first-order reaction

The transport equation considering first-order reaction could be described as follows:

$$\frac{\partial nC}{\partial t} = \text{div}(nD \text{grad } C) - \text{div}(n \mathbf{u} C) - \lambda nC \quad (1)$$

where C is the contaminant concentration, [ML⁻³]; D is the dispersion tensor, [L²T⁻¹]; \mathbf{u} is the convection matrix, [LT⁻¹]; λ is the first-order reaction rate constant, [T⁻¹]; n is the porosity; t is the time, [T]; div is the divergence operator; grad is the gradient operator. The equation is a linear partial differential equation, because the equation is linear to all the unknown function and their derivatives [12].

When the contamination transport process is linear, the contamination concentration could be described by the transfer function as follows [8,13]:

$$C(\mathbf{x}, T) = \int_0^T s(t) f(\mathbf{x}, T-t) dt \quad (2)$$

where \mathbf{x} is the coordinate scalar; $s(t)$ is the actual release process; $f(\cdot)$ is the transfer function. The analytical transfer function could be obtained, if the contaminant is conservative [5,14]. Considering the reaction, some researchers amend the transfer function [15,16] or define the transfer function by transport experiment [17]. However, in many cases the characteristic of groundwater flow may not allow an analytical transfer function formulation, and the experiment could not define the transfer function because of the space time scale limits. In this paper, we use finite difference method to calculate the transfer function considering the first-order reaction.

Firstly, with a simple variable transformation, it could rewrite the Eq. (2) as:

$$C(\mathbf{x}, T) = \int_0^T s(T-t) f(\mathbf{x}, t) dt \quad (3)$$

If we assume a stepwise input function $s(T-t) = 1, t > 0, s(T-t) = 0, t \leq 0$, the breakthrough curve equation is

$$C_s(\mathbf{x}, T) = \int_0^T 1 \cdot f(\mathbf{x}, t) dt \quad (4)$$

Taking the time derivative of the Eq. (4), it results

$$\frac{\partial C_s(\mathbf{x}, t)}{\partial t} = f(\mathbf{x}, t) \quad (5)$$

A numerical model could be used to simulate the response of the aquifer to the stepwise input function. When the

breakthrough curve is obtained, the transfer function value at each time step could be calculated by the backward difference (Eq. (6)). The sensitivity matrix \mathbf{H} can be deduced by the transfer function value, and the pollution source release process could be determined by geostatistical approach.

$$\frac{dC_{t=i}}{dt} = \frac{C_{t=i} - C_{t=i-1}}{\Delta t} \quad (6)$$

2.2. Geostatistical model

The relation between the pollution source release process and the concentration observation could be generalized as follows [5]:

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{v} \quad (7)$$

where \mathbf{z} is an $m \times 1$ vector of observations. \mathbf{H} is a known sensitivity matrix assembled by transfer function. \mathbf{s} is an $n \times 1$ “state vector” obtained from the discretization of the unknown function that we wish to estimate. The measurement error is represented by the vector \mathbf{v} which is assumed to have zero mean and known covariance matrix \mathbf{R} . The expected value and covariance of \mathbf{s} could be expressed as Eqs. (8) and (9).

$$E[\mathbf{s}] = \mathbf{X}\boldsymbol{\beta} \quad (8)$$

$$\mathbf{Q}(\theta) = E[(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T] \quad (9)$$

where \mathbf{X} is a known $n \times p$ matrix and $\boldsymbol{\beta}$ are p unknown drift coefficients. $\mathbf{Q}(\theta)$ is a Gaussian function of unknown parameters θ .

The estimation procedure is divided into two parts. First the optimal structural parameters θ are found, and then the unknown function \mathbf{s} is estimated. The structural parameters θ are estimated by maximizing the probability of the measurements given θ :

$$p(\mathbf{z} | \theta) \propto |\boldsymbol{\Sigma}|^{-1/2} |\mathbf{X}^T \mathbf{H}^T \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{X}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}^T \boldsymbol{\Xi}^{-1} \mathbf{z}\right] \quad (10)$$

$$\boldsymbol{\Sigma} = \mathbf{H} \mathbf{Q} \mathbf{H}^T + \mathbf{R} \quad (11)$$

$$\boldsymbol{\Xi} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{X} (\mathbf{X}^T \mathbf{H}^T \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{H}^T \boldsymbol{\Sigma}^{-1} \quad (12)$$

$$\mathbf{H} = \Delta t \begin{bmatrix} f(\mathbf{x}_1, T-t_1) & f(\mathbf{x}_1, T-t_2) & \cdots & f(\mathbf{x}_1, T-t_n) \\ f(\mathbf{x}_2, T-t_1) & f(\mathbf{x}_2, T-t_2) & \cdots & f(\mathbf{x}_2, T-t_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_m, T-t_1) & f(\mathbf{x}_m, T-t_2) & \cdots & f(\mathbf{x}_m, T-t_n) \end{bmatrix} \quad (13)$$

Maximizing $p(\mathbf{z} | \theta)$ is equivalent to minimizing

$$L(\theta) = \frac{1}{2} \ln |\boldsymbol{\Sigma}| + \frac{1}{2} \ln |\mathbf{X}^T \mathbf{H}^T \boldsymbol{\Sigma}^{-1} \mathbf{H} \mathbf{X}| + \frac{1}{2} \mathbf{z}^T \boldsymbol{\Xi}^{-1} \mathbf{z} \quad (14)$$

The minimization can be achieved by taking derivatives of $L(\theta)$ with respect to θ and setting them to zero. Gauss–Newton iterations are used to find the minimization. When the iteration converges, $\mathbf{Q}(\theta)$ is known and solves the system

$$\begin{bmatrix} \Sigma & \mathbf{HX} \\ (\mathbf{HX})^T & 0 \end{bmatrix} \begin{bmatrix} \Lambda^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{HQ} \\ \mathbf{X}^T \end{bmatrix} \quad (15)$$

where Λ is a $m \times n$ matrix of coefficients and \mathbf{M} is $p \times n$ matrix of multipliers. The best estimates of the function \mathbf{s} and its covariance are

$$\hat{\mathbf{s}} = \Lambda \mathbf{z} \quad (16)$$

$$\mathbf{V} = -\mathbf{XM} + \mathbf{Q} - \mathbf{QH}^T \Lambda^T \quad (17)$$

The method does not enforce the nonnegativity of concentration. A transformation of the concentration is used to assure the nonnegativity of concentration. Define

$$\tilde{\mathbf{s}} = \alpha (\mathbf{s}^{1/\alpha} - 1) \quad (18)$$

The Eq. (1) in the transformed space becomes

$$\mathbf{z} = \mathbf{h} \left[\left(\frac{\tilde{\mathbf{s}} + \alpha}{\alpha} \right)^\alpha \right] + \mathbf{v} = \tilde{\mathbf{h}}(\tilde{\mathbf{s}}) + \mathbf{v} \quad (19)$$

Then the transfer function $\tilde{\mathbf{h}}(\tilde{\mathbf{s}})$ is not linear with respect to the transformed unknown $\tilde{\mathbf{s}}$. The best estimate of \mathbf{s} can be found by the quasi-linear procedure [5,18] and could be expressed as

$$\hat{\mathbf{s}} = \left(\frac{\tilde{\mathbf{s}} + \alpha}{\alpha} \right)^\alpha \quad (20)$$

3. Numerical case

The first-order reactive process is added to the 1D transport process case [3] to evaluate the method in this paper. We suppose the pollutant transport in a homogeneous aquifer, the actual mean velocity is a constant, the problem could be expressed as follows [11,19]:

$$\frac{\partial C}{\partial t} = D_L \frac{\partial^2 C}{\partial x^2} - u \frac{\partial C}{\partial t} - \lambda C, \quad 0 < x < +\infty, \quad t > 0 \quad (21)$$

$$C(x, t)|_{t=0} = 0, \quad 0 \leq x < +\infty \quad (22)$$

$$C(x, t)|_{x=0} = C_0, \quad t > 0 \quad (23)$$

$$C(x, t)|_{x \rightarrow +\infty} = 0, \quad t > 0 \quad (24)$$

where C is the pollutant concentration, $[\text{ML}^{-3}]$; D_L is the longitude dispersion coefficient ($D_L = 1$), $[\text{L}^2\text{T}^{-1}]$; u is the actual mean velocity ($u = 1$), $[\text{LT}^{-1}]$; λ is the first-order reaction rate constant ($\lambda = 10^{-4}$), $[\text{T}^{-1}]$; x is the transport distance ($x \in [0, 300]$), $[\text{L}]$; t is time, $[\text{T}]$.

Eq. (25) describes the true release history (Fig. 1). There are 20 observation points in the x -direction and the curve of observed concentration \mathbf{z} at $t = 300$ is shown in Fig. 2. The concentration at each observation point could be calculated by supposition method, and could be expressed as Eq. (26) [11]. The covariance of the measurement errors is expressed as $\mathbf{R} = \sigma^2_{\text{R}} \mathbf{I}$ ($\sigma^2_{\text{R}} = 1 \times 10^{-12}$). The \mathbf{Q} is expressed as Eq. (28).

$$s(t) = \exp \left[-\frac{(t-130)^2}{50} \right] + 0.3 \exp \left[-\frac{(t-150)^2}{200} \right] + 0.5 \exp \left[-\frac{(t-190)^2}{98} \right] \quad (25)$$

$$C(x, t) = \sum_{i=1}^n \frac{\Delta C_{i-1}}{2} \left\{ \exp \left[\frac{(u-w)x}{2D_L} \right] \operatorname{erfc} \left[\frac{x-w(T-t_{i-1})}{2\sqrt{D_L(T-t_{i-1})}} \right] + \exp \left[\frac{(u+w)x}{2D_L} \right] \operatorname{erfc} \left[\frac{x+w(T-t_{i-1})}{2\sqrt{D_L(T-t_{i-1})}} \right] \right\} \quad (26)$$

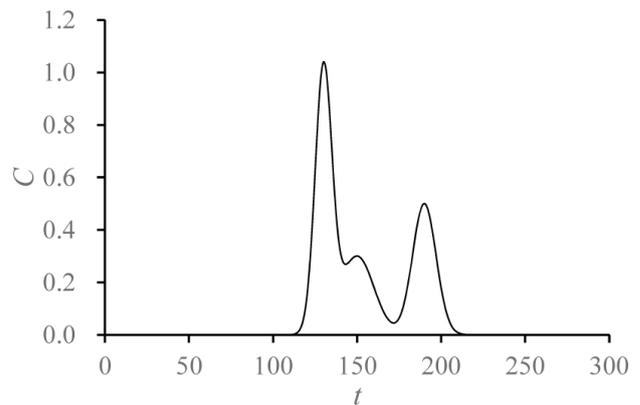


Fig. 1. Pollution source release curve.

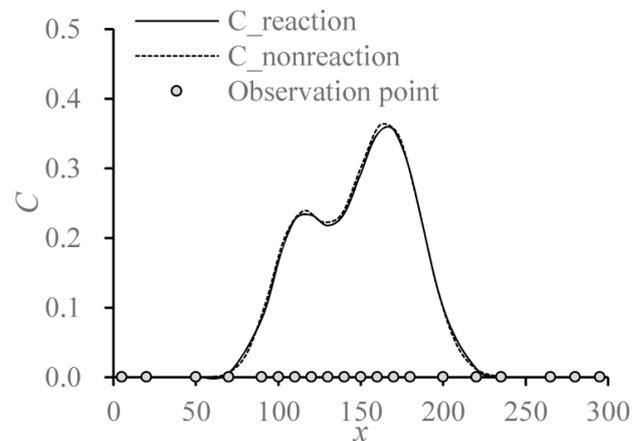


Fig. 2. Observation location and observed concentration at $t = 300$.

$$w = \sqrt{u^2 + 4\lambda D_L} \tag{27}$$

$$Q(t_i, t_j | \theta) = \sigma^2 \exp\left[-\frac{(t_i - t_j)^2}{l^2}\right] \tag{28}$$

4. Discussion

4.1. Identification result

Throughout the numerical case mentioned earlier, we obtain the breakthrough curves of each observation points (Fig. 3). The observation points OBW18–OBW20 are far from the pollution release source, and the breakthrough curves have not reached 1 at $t = 300$. The transfer function curves of these observation points are getting flat when the distance from release source increases (Fig. 4). The calculated release history has good agreement with the real release history (Fig. 5), their correlation r is 0.9986, so the method we proposed could identify the release history. The Euclidean distance de between the up and low bound of the 95% confidence

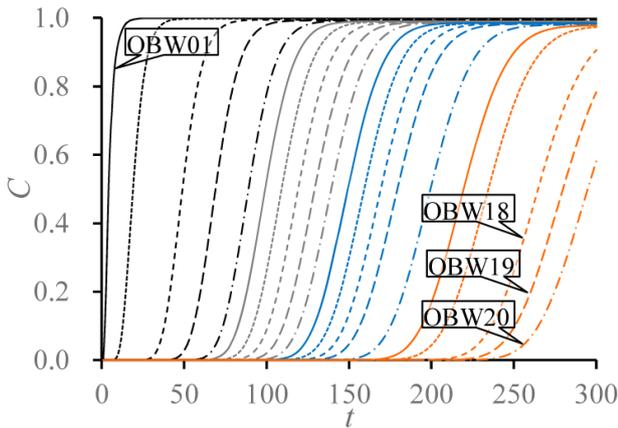


Fig. 3. The breakthrough curve at each observation point.

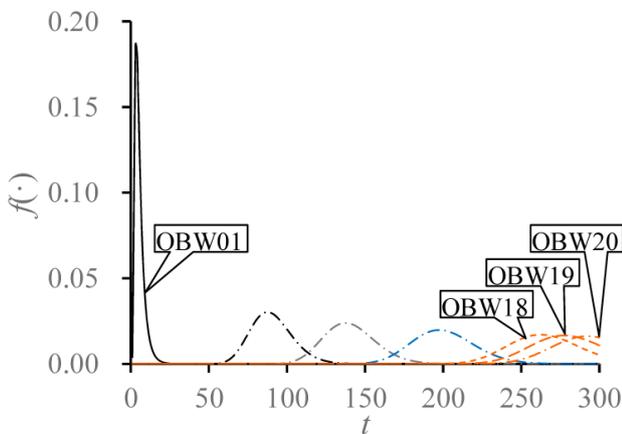


Fig. 4. The transfer function curve of each observation point.

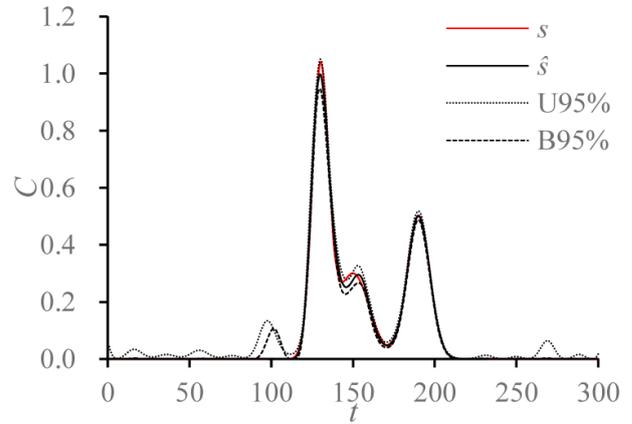


Fig. 5. The real and calculated release curves.

interval is used to evaluate the confidence interval. The de of the identified result is 5.48, it shows that the uncertainty of the calculated history is small. However, small uncertainty does not mean the calculated release history approximates the real history, because the model parameter might not be accurate.

4.2. The influence of first-order reaction rate constant

The reference [9] shows that the first-order reaction rate constant could be 10^{-2} to 10^{-4} commonly. To analyse the influence of reaction rate on PSI result, we set the constant changing between the scopes, and suppose that the reaction rate is known when we implement this PSI method [20,21]. Fig. 6 shows that the concentration curves at $t = 300$ are the similar. When the reaction rate is small, less pollutant is consumed, so concentration could be bigger than the concentration of big reaction rate.

The calculated release history has good agreement with the real release history during the reaction rate scope (Fig. 7). We analyse these results with relative deviation, because each PSI result is too close to the other. The real release

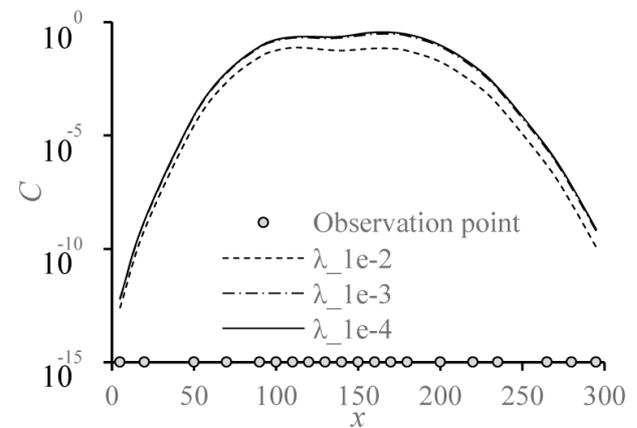


Fig. 6. The concentration curves of each λ at $t = 300$.

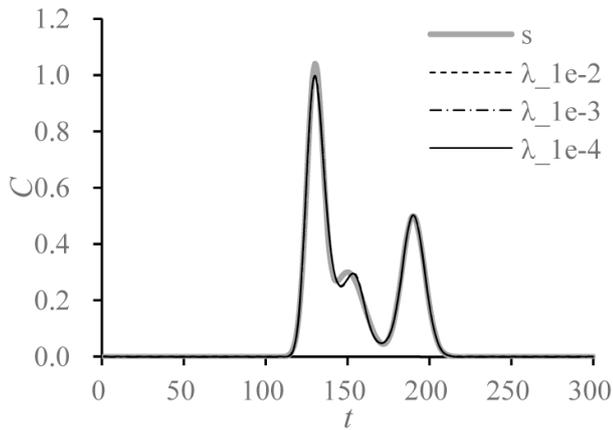


Fig. 7. The release curves of each λ when the λ are known.

concentration is 0 before and after the pollutant releasing, so we calculate the relative deviation of the relative deviation ($\Delta\lambda_1$ and $\Delta\lambda_2$) of the reaction rate 10^{-2} and 10^{-3} to 10^{-4} . Fig. 8 shows that $\Delta\lambda_1$ and $\Delta\lambda_2$ vibrate obviously before and after the pollutant releasing, and the $\Delta\lambda_1$ changes more conspicuously than $\Delta\lambda_2$. Because the concentration of reaction rate 10^{-4} (the denominator of the relative deviation) approaches 0 before and after the pollutant releasing, the small difference might lead to a big relative deviation. When the reaction rate equals 10^{-2} , more pollutant is consumed and the concentration is small, the PSI result is more easily influenced by calculate and observation error. So big release concentration and less pollutant consuming might benefit for identifying the release history.

4.3. The sensitivity of PSI result to the reaction rate

We suppose the real reaction rate is 10^{-4} , and it is unknown when the PSI method is implementing. By changing the reaction rate, the sensitivity of PSI result to the reaction rate is analysed. Fig. 9 shows when the reaction rate is equal or smaller than 10^{-3} , the calculated and real release history are almost the same, the correlation r is bigger

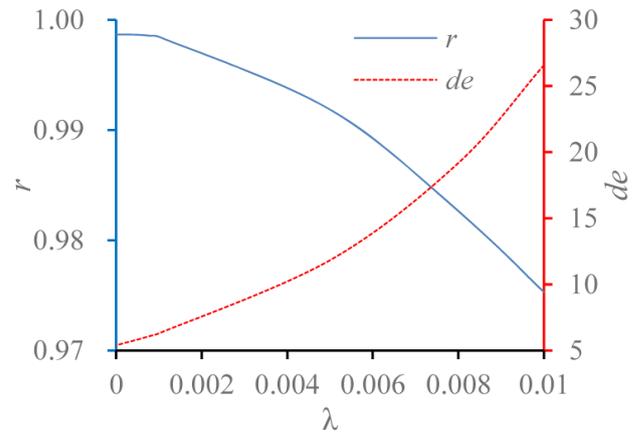


Fig. 9. The influence of λ on r and de .

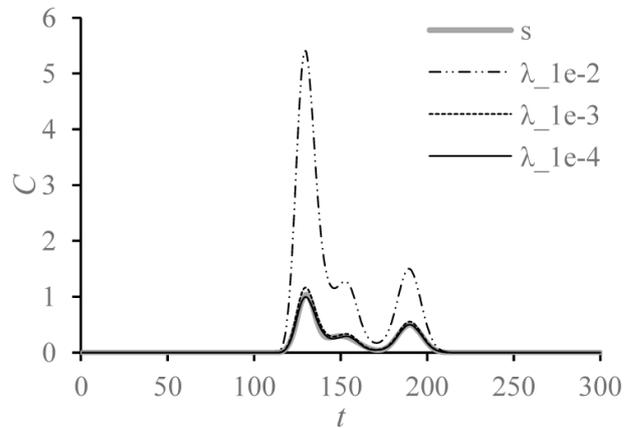


Fig. 10. The release curves of each λ when the λ are unknown.

than 0.9985, the Euclidean distance de is smaller than 6.3. When the reaction rate changes between 10^{-4} and 10^{-3} , r and de vary slightly [22]. It means the PSI result is not very sensitive to the reaction rate, a moderate result could be obtained [23]. When the reaction rate is bigger than 10^{-3} , the shape of release history curves is similar to the real history, but the peak is bigger than the real history (Fig. 10). If the estimated reaction rate is far from the real reaction rate, the calculated history cannot describe the real history. So, an accurate reaction rate is still an important key factor for identifying the release history with this method.

The reaction rate is big, more pollutant is consumed and the observation concentration reduces [24]. If the observation concentration at final time spot is the same, the concentration calculated with big reaction is higher than the concentration calculated with small reaction. It means the method is coincident with the law of transport with first-order reaction.

5. Conclusion

In this paper, the geostatistical approach considering the first-order reaction is used to identify the pollutant release history. The finite difference method is used to calculate the

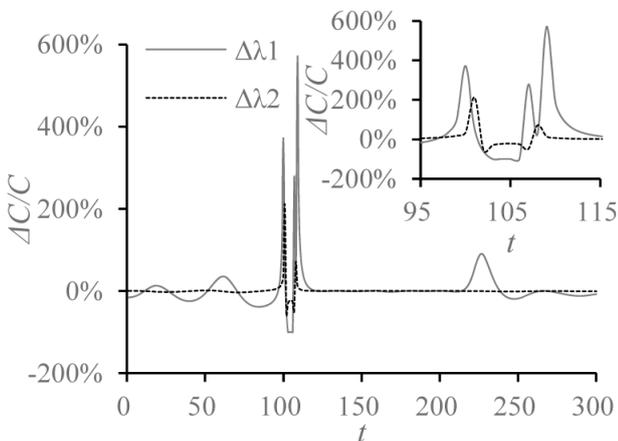


Fig. 8. The relative error of PSI results.

transfer function. The numerical experiment is carried out to test the method and evaluate the PSI result. It turns out that (1) the method proposed in this paper could identify the release history perfectly in the numerical case; (2) high concentration or small chemical reaction consumption of pollutant are helpful for finding accurate PSI result; (3) the PSI result is insensitive to the reaction rate when the reaction rate falls between 10^{-4} and 10^{-3} , but the accurate reaction rate is still very important for the PSI problem. The method prompted in this paper has good agreement with the transport law of pollutant with first-order reaction, and it might be helpful for groundwater pollution identification.

Acknowledgements

This study is financially supported by the Technical Demonstration Project of Ministry of Water Resources (SF-201706), the National Natural Science Foundation of China (grant nos. 51509157 and 51409161), the Natural Science Foundation of Jiangsu Province (grant no. BK20140080).

References

- [1] A.Y. Sun, S.L. Painter, G.W. Wittmeyer, A constrained robust least squares approach for contaminant source release history identification, *Water Resour. Res.*, 42 (2006) W04414.
- [2] E. Milnes, P. Perrochet, Simultaneous identification of a single pollution point-source location and contamination time under known flow field conditions, *Adv. Water. Res.*, 30 (2007) 2439–2446.
- [3] T.H. Skaggs, Z.J. Kabala, Recovering the release history of a groundwater contaminant, *Water Resour. Res.*, 30 (1994) 71–79.
- [4] J. Atmadja, A.C. Bagtzoglou, State of the art report on mathematical methods for groundwater pollution source identification, *Environ. Forensics*, 2 (2001) 205–214.
- [5] M.F. Snodgrass, P.K. Kitanidis, A geostatistical approach to contaminant source identification, *Water Resour. Res.*, 33 (1997) 537–546.
- [6] I. Butera, M.G. Tanda, A geostatistical approach to recover the release history of groundwater pollutants, *Water Resour. Res.*, 39 (2003) 1372.
- [7] A.M. Michalak, P.K. Kitanidis, Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resour. Res.*, 40 (2004) W08302.
- [8] I. Butera, M.G. Tanda, A. Zanini, Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach, *Stochastic Environ. Res. Risk Assess.*, 27 (2012) 1–12.
- [9] C.M. Zheng, G.D. Bennett, *Applied Contaminant Transport Modelling*, Wiley-Interscience, New York, 2002.
- [10] H. Wang, H. Zheng, F. Azuaje, Poisson-based self-organizing feature maps and hierarchical clustering for serial analysis of gene expression data, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 4 (2007) 163–175.
- [11] H. Wang, *Dynamics of Fluid Flow and Contaminant Transport in Porous Media*, Higher Education Press, Beijing, 2008.
- [12] S.Z. Lu, J.H. Shi, W.Y. Su, C.J. Yang, Y.Z. He, S.X. Yu, *Mathematics Dictionary*, Shanxi Education Press, Taiyuan, 2002.
- [13] A.J. William, H. Robert, *Soil Physics [M/OL]*, 6th ed., John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.
- [14] L. Ren, F. Yan, F. Zhang, A transfer function for the simulation and prediction of nitrate-nitrogen leached in soil, *J. Hydraul. Eng.*, 1 (2001) 21–27.
- [15] J. Utermann, E.J. Kladviko, W.A. Jury, Evaluating pesticide migration in tile-drained soils with a transfer function model, *J. Environ. Qual.*, 19 (1990) 707–714.
- [16] R.D. Zhang, The Transfer Function for Solute Transport, *Proceedings of Fourteenth Annual American Geophysical Union*, 1994, pp. 443–453.
- [17] L. Ren, M. Mao, Application of transfer function model to predict transport of Atrazine in soil, *J. Hydraul. Eng.*, 33 (2002) 39–45.
- [18] P.K. Kitanidis, Quasi-linear geostatistical theory for inverting, *Water Resour. Res.*, 31 (1995) 2411–2419.
- [19] S.C.A. Mana, M.M. Hanafiah, A.J.K. Chowdhury, Environmental characteristics of clay and clay-based minerals, *Geol. Ecol. Landsc.*, 1 (2017) 155–161.
- [20] G. Chelladurai, Influence of diets on growth and biochemical parameters of *Babylonia spirata*, *Geol. Ecol. Landsc.*, 1 (2017) 162–166.
- [21] N. Hashemi, Recognizing the potential of sustainable use of pasture resources in South Khorasan province with approach of carrying capacity, *Environ. Ecosyst. Sci.*, 1 (2017) 09–12.
- [22] A. Radan, M. Latifi, M. Moshtaghie, M. Ahmadi, M. Omid, Determining the sensitive conservative site in Kolah Ghazi National Park, Iran, to management wildlife by using GIS Software, *Environ. Ecosyst. Sci.*, 1 (2017) 13–15.
- [23] O. Adugna, D. Alemu, Evaluation of brush wood with stone check dam on gully rehabilitation, *J. CleanWAS*, 1 (2017) 10–13.
- [24] G. Farajollahi, M.R. Delavar, Assessing accident hotspots by using volunteered geographic information, *J. CleanWAS*, 1 (2017) 14–17.