

Source identification of sudden contamination based on the parameter uncertainty analysis

Wei Guozhen, Chi Zhang, Yu Li, Liu Haixing and Huicheng Zhou

ABSTRACT

It is important to identify the source information after a sudden water contamination incident occurs in a water supply system. The accuracy of the simulation model's parameters determines the accuracy of the source information. However, it is difficult to obtain the true value of these parameters by existing methods, so reduction of the errors caused by the uncertainty of these parameters is a crucial problem. A source identification framework which considers the uncertainty of the model's sensitive parameters and combines Bayesian inference and Markov Chain Monte Carlo (MCMC) algorithms simulation is established, and the South-to-North Water Diversion Project is taken as the case study in this paper. Compared with a framework which does not consider the uncertainty of the model's parameters, the proposed framework could solve the error caused by the wrong choice of model parameters and obtain more accurate results. In addition, the proposed framework based on traditional MCMC and that based on the Delayed Rejection and Adaptive Metropolis (DRAM-MCMC) are compared to prove that the DRAM-MCMC is more convergent and accurate. Lastly, the proposed framework based on DRAM-MCMC is proved to solve the problem with high practicality and generality in the studied long distance water diversion project.

Key words | Markov Chain Monte Carlo (MCMC), parameter uncertainty, source identification, sudden contamination incident

Wei Guozhen
Chi Zhang (corresponding author)
Yu Li
Liu Haixing
Huicheng Zhou
School of Hydraulic Engineering,
Dalian University of Technology,
Dalian 116024,
China
E-mail: czhang@dlut.edu.cn

INTRODUCTION

In recent years, sudden water contamination incidents in water supply systems or rivers, resulting from transport accidents, sewage pipe bursts, unregulated sewage treatment plant discharges, terrorism attack and extreme natural disasters (such as earthquake), have occurred more frequently and have caused environmental, economic and societal consequences, in particular in developing countries such as China (He *et al.* 2011; Tang *et al.* 2014). Rapid and accurate source identification of water contamination is crucial to reducing the potential impacts of such incidents and thus is an essential step to develop mitigation and adaptation strategies in water quality management.

The contamination source identification problem, which is normally regarded as an ill-posed and inverse problem, has attracted a great deal of attention. Many methods have

been developed in the literature, such as the classical regularization methods, simulation-optimization methods and Bayesian inference methods (Hamdi & Mahfoudhi 2013). The Tikhonov regularization method, one of the classical regularization methods, was first used to solve such problems, has been proved to be reliable and simple, and also can cope with the noise in the experimental data (Nguyen *et al.* 1999; Akçelik *et al.* 2003; Wang *et al.* 2013; Qi *et al.* 2016). With the development of computing power, optimization algorithms were used for source identification problems in combination with water quality simulation models, such as genetic algorithms (Khlaifi *et al.* 2009), heuristic harmony search algorithms (Ayvaz 2010) and parallel evolutionary strategies (Mirghani *et al.* 2009). Such algorithms are likely to find solutions that are more accurate

than those obtained from classical regularization methods. However, the regularization and optimization methods can only provide the point estimation but cannot take uncertainties in the inverse problem, which will increase the risk of obtaining the inaccurate identification and reduced reliability of finding the optimal solution because an increasing number of model parameters are uncertain. On the contrary, Bayesian approaches have a number of distinctive advantages and have been used in many areas (Wang *et al.* 2013; Zhang *et al.* 2015). This approach could provide a posterior probability distribution of the corresponding source parameters and quantify random errors in the data (Hassan *et al.* 2009). Takaishi (2013) combined the Bayesian inference and Markov Chain Monte Carlo (MCMC) method implemented by the importance sampling method into the generalized autoregressive conditional heteroscedasticity (GARCH) model, and they found that the methods could reduce the statistical error of the GARCH parameters. Wang & Harrison (2013) used the Bayesian and MCMC methods to identify the contaminant profile in Water Distribution Systems, they examined a statistical learning approach to build a regression model between the proposed parameters and likelihood for each pair of source and sensor nodes in the network and proved that the method was feasible and efficient. Shao *et al.* (2014) combined the Bayesian approaches and MCMC simulation to identify water quality model parameters, and the result shows that the method has high reliability and anti-noise capability. It can be imagined that the method combining the Bayesian approaches and MCMC is a good way to solve the contamination source identification problem.

In previous research, model parameters were regarded as deterministic, though they were obtained by the tracer tracking method, analogy method and empirical formula method generally, they had high uncertainties which resulted in the accuracy of the results directly (Van Griensven & Meixner 2007; Blasone *et al.* 2008; Xu *et al.* 2009; Zhang *et al.* 2013; Tian *et al.* 2014). Therefore, how uncertainties of model parameters are considered is crucially important. This paper aims to propose a framework, which replaces the deterministic value with the prior probability function of the model parameters obtained from the practical measure, analogy method and empirical formula method to make identification more accurate and faster.

Furthermore, the modified method of MCMC, i.e., the DRAM-MCMC, combining the Delayed Rejection, Adaptive Metropolis and MCMC (Haario *et al.* 2006), is proposed to improve the efficiency and accuracy of the source identification.

The remainder of this paper provides an overview of the framework of the source identification which considers the uncertainties of the parameters and gives the scheme design, followed by details of the case study. The sensitivity analysis results for the parameters and uncertainty analysis of different scenarios are then provided and the conclusions drawn.

METHODOLOGY

The contamination source identification is proposed to obtain the values of the contamination source terms by the mathematical methods according to the numerical model and the output of this model, which are obtained by the monitor or others methods. Previous papers gave values to the parameters of the numerical model, but neglected the uncertainty of the source terms. Thus, a new framework considering the uncertainty of the parameters is proposed, as shown in Figure 1 and presented as follows. Initially, a sensitivity analyzing method is used to analyze the sensitivity of the parameters and then the insensitive parameters are given constant values and directly applied into the water quality model. The ranges of the parameters of the model were assumed on the basis of practical measure, analogy method and empirical formula method. Following this, the prior probability function of the source terms and the sensitive parameters are put into the Bayesian inference and MCMC to identify the posterior probability function of the source terms.

Water quality model

In this paper, a long distance water diversion project is adopted as a case study, so the vertical diffusion of the water quality model is ignored, and a two-dimensional water quality model that describes the flow crosswise and lengthwise is used to simulate the effect of pollutant transport along the river. This model is mainly used for the

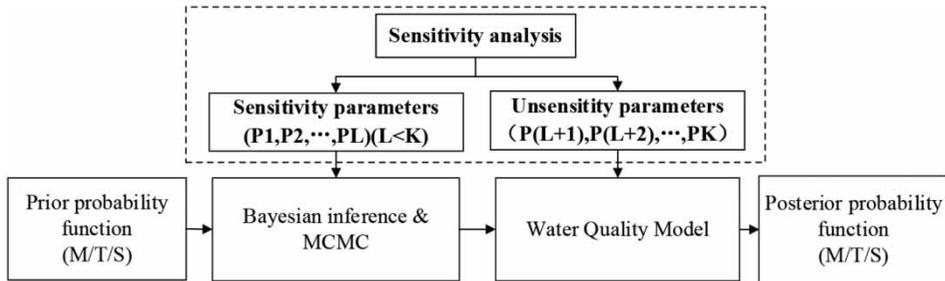


Figure 1 | The framework considering the model's insensitive parameters.

simulation of the transient point source discharge, and is described as

$$\frac{\partial C}{\partial t} + U_x \frac{\partial C}{\partial x} + U_y \frac{\partial C}{\partial y} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + KC \quad (1)$$

in which x represents the distance along the channel, y represents the distance along the cross-section, t is the time, C (g/l) is the pollutant concentration at the point (x, y, t) , U_x and U_y (m/s) denote the velocity of the river at the flow crosswise and lengthwise, respectively, D_x and D_y ($m \cdot s^{-2}$) are the vertical and horizontal mixed coefficients, respectively, and K represents the pollutant degradation coefficient, (d^{-1}).

Then the analytical expression of the pollutant concentration at the point (x, y, t) , i.e., $C(x, y, t)$ can be expressed as:

$$C(x, y, t) = \frac{M}{2U_x h \sqrt{D_x D_y t^2}} \exp(-Kt) \times \left\{ \exp \left[-\frac{(x - U_x t)^2}{4D_x t} - \frac{(y - U_y t)^2}{4D_y t} \right] \right\} \quad (2)$$

where h denotes the depth of river; M denotes the quality of the pollutants.

Bayesian inference and MCMC

Bayesian inference

The Bayesian inference is a useful approach where prior knowledge is taken into consideration naturally and allows to the user to obtain uncertainties about the estimated

parameter (Zio & Zoia 2008; Wang & Chen 2013; Zhao et al. 2014).

In this paper, the ranges of the model parameters were assumed on the basis of the practical measure, analogy method and empirical formula method. Due to a lack of data, uniform distributions within the parameter value ranges are used as prior distributions for these parameters following Haario et al. (2006) and Wang & Chen (2013). Various parameters are independent of each other. Therefore, according to the Bayesian inference, the total prior distribution is as follows:

$$P(\alpha) = \prod_{i=1}^n P(\alpha_i) = \begin{cases} \prod_{i=1}^n \frac{1}{\alpha_{i \max} - \alpha_{i \min}} & \alpha_i \in [\alpha_{i \min}, \alpha_{i \max}]; \\ 0 & \text{others;} \end{cases} \quad (3)$$

where $\alpha_{i \min}$ and $\alpha_{i \max}$ represent the upper and lower limitations of the i th parameter, respectively.

The noise of simulation error ε could be assumed to follow a distribution in accordance with the actual conditions. The distribution $P(\theta|\alpha)$ is the likelihood probability distribution. Then according to the Bayesian formula, the posterior probability distribution could be obtained:

$$P(\alpha|\theta) \propto P(\alpha) \times P(\theta|\alpha) = \prod_{i=1}^n P(\alpha_i) \times P(\theta|\alpha) \quad (4)$$

with the above equations, the posterior probability of parameters can be obtained. However, with the increasing number of unknown parameters, the calculation of numerical integration algorithm increases exponentially. Therefore, the MCMC method is applied to solve the problem.

MCMC sampling

The main methods to construct the Markov chain transition probability matrix include the Gibbs sampling algorithm, the Metropolis–Hastings algorithm and Metropolis algorithm (Cowles & Carlin 1996; Cowles & Rosenthal 1998; Zio & Zoia 2008; Haghighattalab *et al.* 2012). However, according to the previous studies (e.g., Haario *et al.* 2006; Mbalawata *et al.* 2015), the challenge of the standard methods is that it is very hard to find a good proposal distribution in complicated high-dimensional models. Haario *et al.* (2006) proposed a modified Metropolis algorithm Delayed Rejection and Adaptive Metropolis (DRAM-MCMC) and proved that the method combining with DRAM) was more efficient than MCMC. In the method, a higher stage candidate in DR is added to preserve the property and the reversibility of the Markov chain relative to the distribution of interest at each time step in the DRAM-MCMC method. And the advantage of DR could also save in terms of simulation time depend on exploiting the hierarchy between kernels. Moreover, AM, which has the correct ergodic properties, has been introduced into MCMC so that the likelihood probability distribution could be updated along the process using the full information cumulated. More details and theory can be seen in Haario *et al.* (2006).

In this paper, there are not enough data to construct the likelihood distribution. So the likelihood distribution in the traditional MCMC is assumed to follow a normal distribution as a result of lacking data (Wang & Chen 2013; Zhao *et al.* 2014), in DRAM-MCMC, it is assumed to follow a correlated Gaussian distribution according to the study by Haario *et al.* (2006). And the key problems are how to set the values of covariance and how many stages should be built. On the one hand, the covariance is made up of the initial covariance C_0 , the length of the initial non-adaptation N_0 , the target's dimension d and the standard optimal factor S_d . According to the study of Haario *et al.* (2006), the target's dimension d should be less than 15, so it is set to 6. The length N_0 of the initial non-adaptation period which is related to d is set to 1,000. The Gaussian proposal is started by the standard optimal factor $S_d = 2.4^2/d$ (Gelman *et al.* 1996). On the other hand, a second stage proposal is used in DR.

Scenario design

Three different scenarios to identify the source are designed for comparison. Scenario 1 and Scenario 2 are set to compare the advantage between the traditional framework which does not consider the uncertainty of the model parameters and the proposed framework in this paper. Scenario 2 and Scenario 3 are used to prove the superiority of the modified MCMC. The mode of the posteriori distribution of $\alpha = (\alpha_1, \alpha_2 \dots \alpha_n)$ is used as a representation of the characteristics of the contamination event. The number of the iteration was set to 50,000 to make these three schemes comparable in the paper. The detailed scenarios are shown in Table 1.

Scenario 1 (S1): the main purpose of this scenario is to illustrate that the uncertainty of parameters could easily lead to greater error in the results. It is difficult to select a single value in the range that is consistent with the truth value because of the uncertainties. In each identification, we chose a set of fixed parameters from the ranges of model parameters and use the traditional MCMC to estimate the source terms in this scenario without considering model parameter uncertainties.

Scenario 2 (S2) is based on the proposed framework and uses traditional MCMC, and Scenario 3 (S3) is based on the proposed framework and uses the modified MCMC, DRAM-MCMC.

According to the scenarios designed above, the accuracy and high efficiency of the proposed framework could be proved, because it is useful to solve the specified event. But we cannot prove that the proposed framework is suitable for other events, which differ in occurrence time, total load of the sewerage and position. Therefore, in order to prove this, 2,000 events with different levels of Load, Location and Time are designed to be identified with S3.

Table 1 | The design of three scenarios

| Scenario | S1 | S2 | S3 |
|-------------------------------------|------------------|------------------|-----------|
| Is parameter uncertainty considered | No | Yes | Yes |
| Sampling method | Traditional-MCMC | Traditional-MCMC | DRAM-MCMC |

CASE STUDY

In order to prove the accuracy and versatility of the scenario, the middle route of the South-to-North Water Diversion Project was chosen as a case study. The project transfers water from Danjiangkou reservoir in Hubei province and crosses the Yangtze River, Huaihe River, Haihe River, Yellow River basins and finally arrives at Beijing Tuancheng Lake. The total length of the project is 1,277 km, crossing nearly 150 cities and 151,000 hectares of land via open-channels, culverts and pipes. In addition, the span of the project is very large, and 936 different types of structures, including 44 railways and 571 bridges over the channel, have been built. Also, many factories have been built near the project especially in the Shijiazhuang province. Therefore, there is a high probability of sudden water pollution incident especially in Shijiazhuang province. Above all, a section in the Shijiazhuang province of 20 km length is taken as an example to identify the source term of the chemical contaminant.

According to the analysis of the possible contamination of the section and the field experiment and the analogy experiment, we could obtain the interval of the properties of the section, as listed in Table 2. Let us assume that sewage spilled into the section instantly. After a period, the downstream monitoring station detects a series of pollutant concentration data, as illustrated in Figure 2. The true fluid properties and hydraulic characteristics parameters are listed in Table 2, and the designed data of the spill, including occurrence time (Time), the total load of the sewage (Load), and the location of the event (Location) are all supposed to be known, and are 120 minutes, 1,000,000 g and 5,000 m, respectively.

Table 2 | Water quality model parameters

| Parameter | Interval | Default value |
|----------------------------|----------------|---------------|
| $D_x/(m^2 \cdot min^{-1})$ | [980, 1,470] | 1,225 |
| $D_y/(m^2 \cdot min^{-1})$ | [28.8, 43.2] | 36 |
| $U_x/(m \cdot min^{-1})$ | [17.28, 25.92] | 21.6 |
| $U_y/(m \cdot min^{-1})$ | [0.096, 0.144] | 0.12 |
| $K/(d^{-1})$ | [0.08, 0.12] | 0.1 |
| $h/(m)$ | [0.8, 1.2] | 1 |

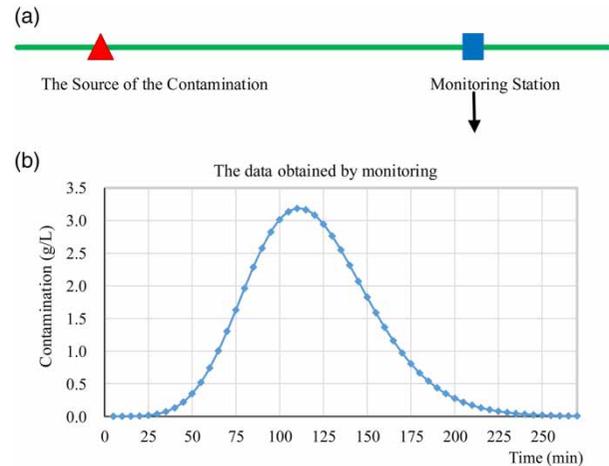


Figure 2 | The course of source identification.

RESULTS AND DISCUSSION

Sensitivity analysis

The sensitivity indices of six parameters are shown in Figure 3, where each panel represents Load, Location and Time, respectively. In each panel, the x-axis represents parameters, and y-axis represents the sensitivity indices, which measure individual parameter contributions to the variance of the metrics. In this paper, the parameters with sensitivity indices exceeding 10% are regarded as the sensitive parameter. As can be seen from Figure 3, for Load and Time, U_x is the only sensitive parameter, while for Location, there are two sensitive parameters, U_x and D_x . The sensitivity indices of other parameters, including U_y , D_y , h and K , are all below the 10% threshold, which indicates that they are insensitive parameters and their impact on the results is very small. This is because the source identification model is built for a long distance water transfer project where the horizontal distance is much longer than the longitudinal distance. During the period of translation, diffusion and decay have happened, so the influence of the horizontal will be much lower than that of the longitudinal so that it could even be ignored, and the longitudinal parameters U_x and D_x are important for the source identification model. In conclusion, the uncertainties of the sensitivity parameters U_x and D_x need to be considered, while the insensitivity parameters, U_y , D_y , h and K are regarded as constant.

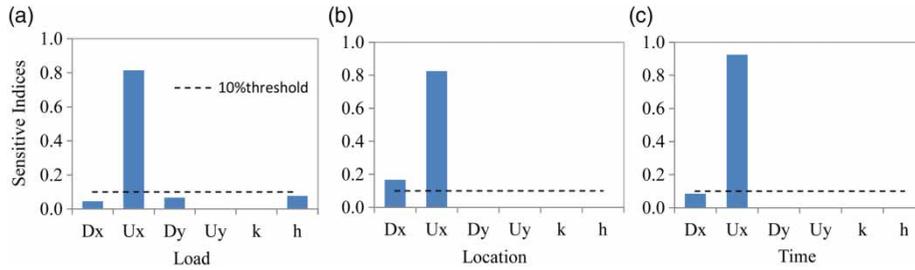


Figure 3 | Results of sensitivity analysis with regard to (a) load, (b) location and (c) time.

Uncertainty analysis

We put the sensitive parameter of the model into S2–S3, and every scenario runs 2,000 times. The maximum relative errors of all of the three source terms (Load, Location and Time) are chosen to evaluate the precision of the results. This is because if one of the three source terms does not satisfy the requirement of the precision, it will be misleading as a false alarm. Only when the maximum relative error satisfies the requirement, then all of the source terms which are identified can satisfy the accuracy requirement. The maximum relative errors of all of the three source terms and the statistical results, the frequency and cumulative frequency could be computed from 2,000 sets of results, as shown in Figure 4, in which x -axis represents the relative errors, the left y -axis represents the frequency and the right y -axis represents the cumulative frequency.

The assumption is that only the relative errors of all of the three source terms (Load, Location and Time) of less than 20% can be accepted. As can be seen from Figure 4(a), it is solved by S1 that only 27.2% of the 2,000 set results could be accepted. It is difficult to choose the correct values within these small ranges. More than 60% of the 2,000 set results are higher than 50% of the true results. This reveals that when the uncertainty of model parameters is not considered, it is very likely to fail to identify the contamination event.

Figure 4(b) shows that 63% of the results solved by S2 could be accepted, which implies that the scenario based on the traditional MCMC could reduce the uncertainty so as to make the results more accurate. The disadvantage of this scenario is that the relative error of more than 17% of the 2,000 set results exceed 100%. This means that the convergence of the scenario will be reduced if we consider the uncertainty of the sensitive parameters of the model,

which uses the range instead of the fixed value. In order to solve the problem, we propose the modified method DRAM-MCMC (Haario *et al.* 2006) in S3. And the results, as shown in the Figure 4(c), show that 96.55% of the 2,000 set results could be accepted, revealing that the DRAM-MCMC is more efficient and accurate than the traditional MCMC. Compared with the traditional MCMC, it could increase the convergence of the scenario because the DR keeps the Markovian property and reversibility by increasing the stage candidate and the AM improves ‘global’ adaptation based on the past history of the chain.

Lastly, in order to prove S3 is fit for the other events, 2,000 cases of different Load, Location and Time have been created, and the results are shown in Figure 5. It shows that 79.48% of the cases could provide acceptable results, having relative error of less than 20%, and the relative errors of 92.93% of the cases are under the level of 40%. Results therefore show that S3 is suitable for all kinds of events.

According to the analysis above, we could suggest that the commonality of the scenario based on DRAM-MCMC is better than that based on MCMC. In order to prove the availability of the scenario, we performed an analysis based on the data and obtained the mean and the standard deviation of relative error, as listed in Table 3. This shows that for the 80th quantile, the mean and the standard deviation of the relative errors are all less than 20%, relatively small, and the 90 quantile of the relative errors are all less than 31.18%.

CONCLUSIONS

According to previous research, we know that the uncertainty of parameters of the model, which is used to

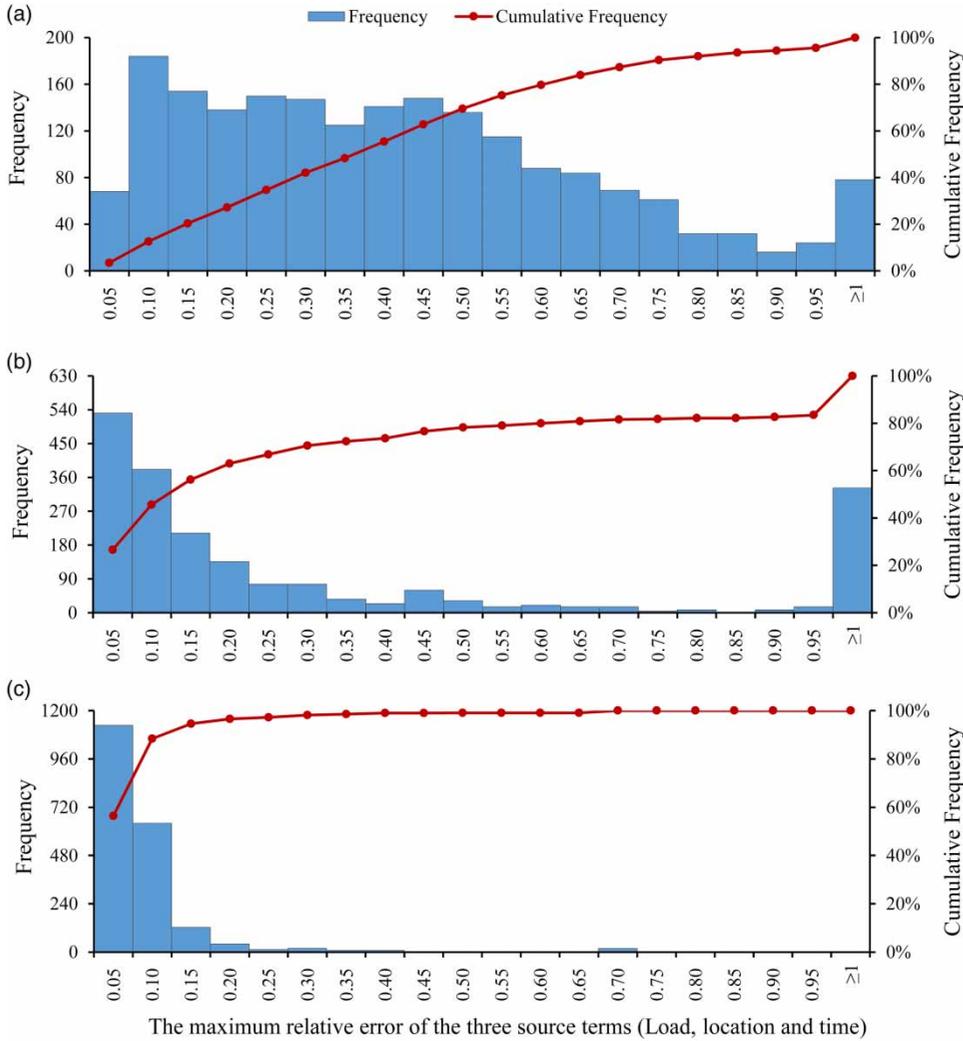


Figure 4 | The frequency and cumulative frequency of the maximum relative error of the three source terms, Load, Location and Time, under the three scenarios, illustrated for (a) S1, (b) S2 and (c) S3.

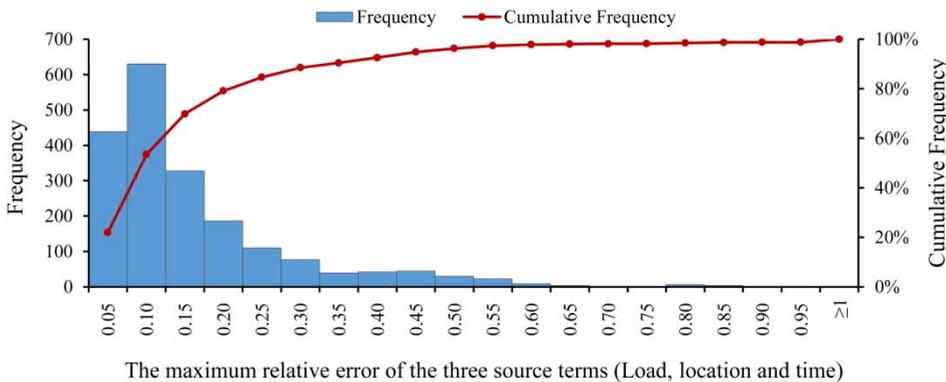


Figure 5 | The statistical results by the scenario based on DRAM-MCMC.

Table 3 | Result of the relative error

| | Distance (D) | Time (T) | Quality (Q) |
|--|-----------------|----------|----------------|
| Standard deviation of the relative error | 11.89% | 16.54% | 12.65% |
| 80th quantile of relative error | 17.80% | 19.08% | 14.20% |
| 90th quantile of relative error | 28.45% | 31.18% | 21.01% |
| Mean of relative error | 11.50% | 11.56% | 9.28% |

identify the source, leads to big errors in the result. So a framework considering the uncertainty of the model's insensitive parameters based on MCMC to improve accuracy of results is proposed. Two improvements have been used in this framework. One is that the sensitive parameters have been chosen by using the sensitivity analysis method to give the prior probability function instead of constant values. The other is that DRAM-MCMC is used to improve the accuracy of the result and the convergence of the identification.

This study provides a step-wise analysis of the source identification framework, which considers the uncertainty of the model's sensitive parameters. The main findings from this study can be summarized as follows: (1) from the sensitivity analysis of the framework, which does not consider the uncertainty of the model parameters, we could know that D_x and U_x are the sensitive parameters leading to very large uncertainties in the results; (2) through comparing the traditional framework which does not consider the uncertainty of the model's parameters and the proposed framework, we have found that the proposed framework which does consider the uncertainty obtains more accurate results than the traditional one; (3) a comparison of the proposed framework based on the traditional MCMC with the new DRAM-MCMC reveals that the proposed framework based on DRAM-MCMC has a better performance in improving the accuracy and convergence of the source terms; and, finally, (4) the proposed framework based on DRAM-MCMC is used to identify many different events, and the result shows that the 80th quantile, the mean and the standard deviation of the relative errors are all less than 20% which is very small, and the results prove that the proposed framework is effective for the case study of the South-to-North Water Diversion

Project, but it should be tested further on more case studies in the future.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 51320105010). Moreover, this study is partly funded by the national science and technology major project under grant 2014ZX03005001.

REFERENCES

- Akçelik, V., Biros, G., Ghattas, O., Long, K. R. & Waanders, B. V. B. 2003 A variational finite element method for source inversion for convective-diffusive transport. *Finite Elements in Analysis and Design* **39**, 683–705.
- Ayvaz, M. T. 2010 A linked simulation-optimization model for solving the unknown groundwater pollution source identification problems. *Journal of Contaminant Hydrology* **117**, 46–59.
- Blasone, R., Madsen, H. & Rosbjerg, D. 2008 Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* **353**, 18–32.
- Cowles, M. K. & Carlin, B. P. 1996 Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91**, 883–904.
- Cowles, M. K. & Rosenthal, J. S. 1998 A simulation approach to convergence rates for Markov Chain Monte Carlo algorithms. *Statistics and Computing* **8**, 115–124.
- Gelman, A., Roberts, G. O. & Gilks, W. R. 1996 Efficient Metropolis jumping rules. *Bayesian Statistics V*, 599–608.
- Haario, H., Laine, M., Mira, A. & Saksman, E. 2006 DRAM: efficient adaptive MCMC. *Statistics and Computing* **16**, 339–354.
- Haghighattalab, A., Minuchehr, A., Zolfaghari, A. & Khoshahval, F. 2012 Bayesian inference along Markov Chain Monte Carlo approach for PWR core loading pattern optimization. *Annals of Nuclear Energy* **50**, 150–157.
- Hamdi, A. & Mahfoudhi, I. 2013 Inverse source problem in a one-dimensional evolution linear transport equation with spatially varying coefficients: application to surface water pollution. *Inverse Problems in Science and Engineering* **21**, 1007–1031.
- Hassan, A. E., Bekhit, H. M. & Chapman, J. B. 2009 Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model. *Environmental Modelling & Software* **24**, 749–763.

- He, G., Zhang, L., Lu, Y. & Mol, A. P. J. 2011 Managing major chemical accidents in China: towards effective risk information. *Journal of Hazardous Materials* **187**, 171–181.
- Khlaifi, A., Ionescu, A. & Candau, Y. 2009 Pollution source identification using a coupled diffusion model with a genetic algorithm. *Mathematics and Computers in Simulation* **79**, 3500–3510.
- Mbalawata, I. S., Särkkä, S., Vihola, M. & Haario, H. 2015 Adaptive Metropolis algorithm using variational Bayesian adaptive Kalman filter. *Computational Statistics & Data Analysis* **83**, 101–115.
- Mirghani, B. Y., Mahinthakumar, K. G., Tryby, M. E., Ranjithan, R. S. & Zechman, E. M. 2009 A parallel evolutionary strategy based simulation–optimization approach for solving groundwater source identification problems. *Advances in Water Resources* **32**, 1373–1385.
- Nguyen, Y. T., Vu, T. D., Wong, H. K. & Yeow, Y. L. 1999 Solving the inverse problem of capillary viscometry by Tikhonov regularisation. *Journal of Non-Newtonian Fluid Mechanics* **87**, 103–116.
- Qi, W., Zhang, C., Fu, G. T. & Zhou, H. C. 2016 Quantifying dynamic sensitivity of optimization algorithm parameters to improve hydrological model calibration. *Journal of Hydrology* **553**, 213–223.
- Shao, D., Yang, H., Xiao, Y. & Liu, B. 2014 Water quality model parameter identification of an open channel in a long distance water transfer project based on finite difference, difference evolution and Monte Carlo. *Water Science & Technology* **69**, 587.
- Takaishi, T. 2013 Markov Chain Monte Carlo versus importance sampling in Bayesian inference of the GARCH Model. *Procedia Computer Science* **22**, 1056–1064.
- Tang, C., Yi, Y., Yang, Z. & Cheng, X. 2014 Water pollution risk simulation and prediction in the main canal of the South-to-North Water Transfer Project. *Journal of Hydrology* **519**, 2111–2120.
- Tian, Y., Booij, M. J. & Xu, Y. 2014 Uncertainty in high and low flows due to model structure and parameter errors. *Stochastic Environmental Research and Risk Assessment* **28**, 319–332.
- Van Griensven, A. & Meixner, T. 2007 A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment model. *Journal of Hydroinformatics* **4**, 277–291.
- Wang, G. & Chen, S. 2013 Evaluation of a soil greenhouse gas emission model based on Bayesian inference and MCMC: Model uncertainty. *Ecological Modelling* **253**, 97–106.
- Wang, H. & Harrison, K. W. 2013 Bayesian update method for contaminant source characterization in water distribution systems. *Journal of Water Resources Planning and Management* **139**, 13–22.
- Wang, J., Wei, T. & Zhou, Y. 2013 Tikhonov regularization method for a backward problem for the time-fractional diffusion equation. *Applied Mathematical Modelling* **37**, 8518–8532.
- Xu, Y., Tung, Y., Li, J. & Niu, S. 2009 Alternative risk measure for decision-making under uncertainty in water management. *Progress in Natural Science* **19**, 115–119.
- Zhang, C., Chu, J. G. & Fu, G. T. 2013 Sobols sensitivity analysis for a distributed hydrological model of Yichun River Basin, China. *Journal of Hydrology* **480**, 58–68.
- Zhang, J., Liu, P., Wang, H., Lei, X. & Zhou, Y. 2015 A Bayesian model averaging method for the derivation of reservoir operating rules. *Journal of Hydrology* **528**, 276–285.
- Zhao, T., Zhao, J., Lund, J. R. & Yang, D. 2014 Optimal hedging rules for reservoir flood operation from forecast uncertainties. *Journal of Water Resources Planning and Management* **140**, 04014041.
- Zio, E. & Zoia, A. 2008 Bayesian inference of BWR model parameters by Markov Chain Monte Carlo. *Annals of Nuclear Energy* **35**, 1929–1936.

First received 3 January 2016; accepted in revised form 11 May 2016. Available online 13 July 2016