

Identification of Water Quality Model Parameter Based on Finite Difference and Monte Carlo

Dongguo Shao¹, Haidong Yang^{1*}, Biyu Liu²

¹State Key Laboratory of Water Resources & Hydropower Engineering Science,
Wuhan University, LuoJia Hill, Wuhan, China

²Management Science & Engineering Department, Economics and Management School,
Jiulonghu Campus, Southeast University, Nanjing, China
Email: *yangcatfix@126.com

Received September 22, 2013; revised October 27, 2013; accepted November 21, 2013

Copyright © 2013 Dongguo Shao *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Identification results of water quality model parameter directly affect the accuracy of water quality numerical simulation. To overcome the difficulty of parameter identification caused by the measurement's uncertainty, a new method which is the coupling of Finite Difference Method and Markov Chain Monte Carlo is developed to identify the parameters of water quality model in this paper. Taking a certain long distance open channel as an example, the effects to the results of parameters identification with different noise are discussed under steady and un-steady non-uniform flow scenarios. And also this proposed method is compared with finite difference method and Nelder Mead Simplex. The results show that it can give better results by the new method. It has good noise resistance and provides a new way to identify water quality model parameters.

Keywords: Parameters Identification; Water Quality Model; Long Distance Water Transfer Project; Finite Difference Method; Monte Carlo

1. Introduction

In order to allocate water resources rationally, many Long Distance Water Transfer Projects (LDWTPs) have been built or are being constructed in China. But there are numerous controls and cross buildings along the process in LDWTP, so it is highly possible that sudden water pollution incident happens [1]. Once these incidents occur, it is necessary to reveal the rule of pollutants transport and diffusion based on water quality parameters quickly, and then put forward emergency disposal counter measures, otherwise it will cause inestimable consequences [2]. Therefore, it is crucial to identify water quality model's parameters. Many approaches have been proposed to identify the parameters of water quality model, such as theoretical formula method, empirical formula method and tracer test method, etc. [3]. The tracer test method belongs to the category of inverse problems which include methods such as moments, fitting, optimization and the uncertainty analysis [4]. With the development of computer technology, identification

method based on Optimization has been widely used, such as Simplex method [5], Particle Swarm Optimization [6] *et al.* But there are many strict limit conditions and the non-identifiability will increase with the increase of parameters' number when using these methods. In addition, LDWTP is complex and giant system, in which there are full of uncertain factors, and it don't exist analytical expressions of pollutant concentration. Therefore, it is necessary to find a better identification method to research the rule of pollutants transport and diffusion in LDWTP.

Therefore, a new method which named FDM-MCMC is proposed based on Finite Difference Method (FDM) and Markov Chain Monte Carlo (MCMC) in this paper to identify the water quality model parameters. And in order to verify this proposed algorithm's accuracy, efficiency and anti-noise capability, the results of parameters identification with different scenarios are analyzed by numerical simulation. Finally this proposed method is compared with Finite Difference Method and Nelder Mead Simplex (FDM-NMS).

*Corresponding author.

2. Materials and Methods

2.1. Description of the Problem

LDWTP is a series system consists of many canal pools which are divided by gates, as shown in **Figure 1** [7]. If pollutants in canals are attenuated by the first order kinetics, the law of pollutants' transportation and diffusion is described by:

$$\frac{\partial(AC)}{\partial t} + \frac{\partial(QC)}{\partial x} = \frac{\partial}{\partial x} \left(AE_x \frac{\partial C}{\partial x} \right) + F(x, t) \quad (1)$$

where A is cross section area, C is the concentration of pollutant, Q is cross section flow, t is the time, x is the longitudinal coordinate, E_x is the longitudinal dispersion coefficient, $F(x, t)$ is the source term. $F(x, t) = S_1 + W_c - KAC$, S_1 is internal source (sink), K is reaction rate, W_c is the exogenous input terms.

According to **Figure 1**, LDWTP is a complicated non-linear system and the concentration of pollutant can be expressed as:

$$C(x, t) = C(x, t|\theta) + e \quad (2)$$

where θ is a set of parameters set, which is difficult to be measured. And the concentration of pollutants is on the section of canal pool (see **Figure 2**).

According to **Figure 2**, Equation (1) can be transformed into (3) [5]:

$$\begin{aligned} & \frac{(AC)_i^{k+1} - (AC)_i^k}{\Delta t} + \frac{(QC)_{i-1}^{k+1} - (QC)_{i-1}^k}{\Delta x_{i-1}} \\ &= \left(\overline{S_1 + W_c} \right)_i^{k+1} - \left(AKC \right)_i^{k+1} + \left[\left(AE_x \right)_i^{k+1} \frac{C_{i+1}^{k+1} - C_i^{k+1}}{(\Delta x_i + \Delta x_{i-1})/2} \right. \\ & \left. - \left(AE_x \right)_{i-1}^{k+1} \frac{C_i^{k+1} - C_{i-1}^{k+1}}{(\Delta x_{i-1} + \Delta x_{i-2})/2} \right] \times \frac{1}{\Delta x_{i-1}} \end{aligned} \quad (3)$$

where C_i^{k+1} is the pollutant concentration of the i th canal pool at $k + 1$ moment, so Equation (3) can be expressed as:

$$a_i C_{i-1} + b_i C_i + c_i C_{i+1} = d_i \quad (i = 1, 2, \dots, N) \quad (4)$$

where a_i, b_i, c_i are coefficients; d_i is a constant. Equation

(4) is a linear implicit difference equation which is made up of N equations and it can be solved by combining with the upstream and downstream boundary conditions.

Therefore, the identification problem of water quality model parameters in LDWTP can be solved by a limited concentration measurement data. However, since the uncertainty of the model structure and the observation data, the FDM-MCMC method is developed to solve this problem.

2.2. Methods

Uncertainty identification method based on the Bayesian theorem can avoid the decision risk caused by the distortion of the "optimal" parameters in a certain extent [8]. So according to Bayesian theorem, it can be stated as follows:

$$p(\theta | y) = p(\theta) p(y | \theta) / p(y) \propto p(\theta) p(y | \theta) \quad (5)$$

where θ is the unknown parameter; y is the observed data; $p(y|\theta)$ is the likelihood function; $p(\theta|y)$ is the parameter's posterior probability density function; $p(\theta)$ is the parameter's joint priori probability density function.

Assuming ε_i is the measurement error, e_i is the prediction error, and these errors are independent and obey Laplace distribution. So the problem of model parameters can be transformed to solve the parameters' posterior probability density function:

$$\begin{aligned} p(\theta | y) &= L(y | \theta) \cdot p(\theta) \propto \frac{p(\theta)}{\prod_{i=1}^M (2\sigma_{f,i} + 2\sigma_{d,i})} \\ &\times \exp \left[-\sum_{i=1}^M \frac{|C_i(x, t) - C_i(x, t | \theta)|}{(\sigma_{f,i} + \sigma_{d,i})} \right] \end{aligned} \quad (6)$$

where M is the number of observation data, $\sigma_{d,i}, \sigma_{f,i}$ are the standard deviation of ε_i and e_i , respectively.

Therefore, the estimated values of the unknown parameters can be obtained by Equation (6) and MCMC method which is based on sampling random method. Since Metropolis-Hastings is a sampling method in a generation-rejection sample forms, a new identification method based on Bayesian-Markov Chain Monte Carlo to

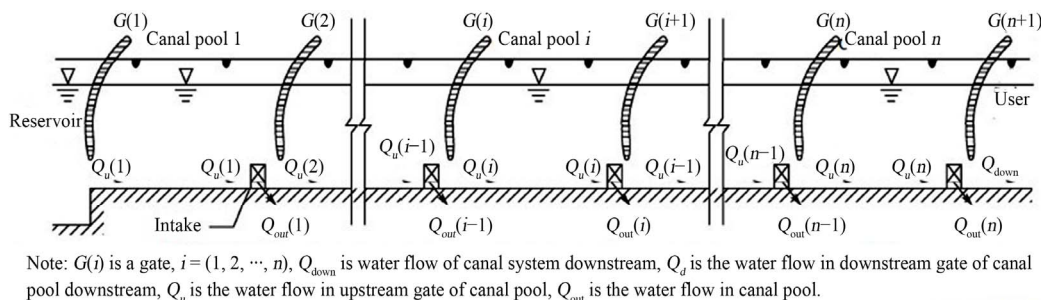
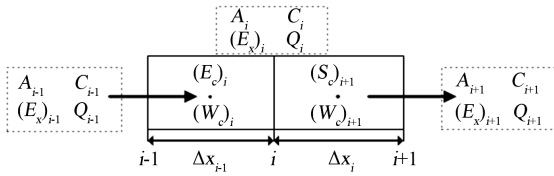


Figure 1. Sketch of the controlled canal system.



Note: $(S)_p, (W)_i$ are the internal source and sink and exogenous input terms of the i th canal pool, respectively; $Q_p, A_p, (E_x)_i$ are respectively water flow, cross section area and the longitudinal dispersion coefficient of the i th canal pool.

Figure 2. The finite difference format to the wind of water quality model equation.

identify the parameters in this paper [9]. The detailed solving steps are as follows:

- 1) The study area is divided into N canal pools by spatial discretization and each canal pool section has only a little change over time and space;
- 2) Determining sample space and $p(\theta^{(i)})$ of the unknown parameters;
- 3) Generating initial values $\theta^{(i)}(1), \theta^{(i)}(2), \dots, \theta^{(i)}(S)$;
- 4) Obtaining the conditional probability density by setting the Proposal distribution $q(\theta^{(i)}(S), \theta^{*(i)}(S))$, generating $\theta^{*(i)}(S)$, and calculating the $\theta^{(i)}(S)$ and $\theta^{*(i)}(S)$ corresponding to the pollutant density;
- 5) Finding the likelihood function which can reflect the relationship between the model parameters and measurement data, and then calculating the posterior probability density function;
- 6) Getting the accept probability $A(\theta^{(i)}, \theta^{*(i)})$ at which Markov Chain moves from $\theta^{(i)}$ to $\theta^{*(i)}$ as following [9]:

$$A(\theta^{(i)}, \theta^{*(i)}) = \min \left\{ 1, \frac{p(\theta^{*(i)}) p(\theta^{(i)} | \theta^{*(i)})}{p(\theta^{(i)}) p(\theta^{*(i)} | \theta^{(i)})} \right\} \tag{7}$$

$$= \min \left\{ 1, \frac{p(\theta^{*(i)})}{p(\theta^{(i)})} \right\}$$

7) Generating a random number which belongs to 0 - 1 and obeys uniform distribution. If $R < A(\theta^{(i)}, \theta^{*(i)})$, then setting $\theta^{(i+1)} = \theta^{*(i)}$, otherwise, $\theta^{(i+1)} = \theta^{(i)}$;

8) Repeating steps from 1) to 7) until it reaches a pre-determined iterations.

3. Results

For LDWTP, the “twin” experimental is an effective means to identify parameters of water quality model [10]. However, the longitudinal dispersion coefficient is becoming more and more important in the sudden water pollution accident [11]. Therefore, an open channel with 3km length is taken as an example in this paper. Assuming the inflow water concentration in the upstream ($x = 0$) is 1.0 mg/L and free outflow in downstream. The flow

field distribution is $u = 0.5 + 0.001x$ and the channel can be dispersed into $N = 6$ canal pools according to the channel’s geometry features whose true values of longitudinal dispersion coefficient $(E_x)_i (i = 1, \dots, N)$ are 50, 70, 90, 110, 130 and 140 m^2/s .

3.1. Steady Non-Uniform Flow

Assuming flow Q is equivalent to 10 m^3/s , and then the identification results are obtained as error level $\sigma = 0.1$ by using FDM-MCMC and FDM-NMS method, respectively, as show in **Table 1**.

From **Table 1**, while $\sigma = 0.1$ the average relative error obtained by FDM-MCMC and FDM-NMS method are respectively 4.99% and 8.25%.

3.2. Unsteady Non-Uniform Flow

If the flow of this open channel is a function of time, it is $Q(t) = 10 + 0.001t$, the identification results of obtained by FDM-MCMC and FDM-NMS under different noise are shown in **Table 2**. Here we take relative standard deviation (RSD) as the accuracy of the identification values:

$$RSD = \lambda/\mu \times 100\% \tag{8}$$

where λ is standard Deviation; μ is the mean of identification value. So when $\sigma = 0.1$, RSD obtained by the two methods are shown in **Table 3**.

From **Table 2**, the average relative errors are respectively 3.41% and 8.26% by FDM-NMS and are respectively 2.93% and 5.14% by FDM-MCMC. From **Table 3**, the average RSD is respectively 27.30% and 52.96% by FDM-NMS.

4. Discussion

Therefore, comparing with FDM-NMS, the FDM-MCMC has the following advantages:

- 1) Wider applicability

In the two scenarios, When $\sigma = 0.1$, the average relative errors are less than 6% by the FDM-MCMC. So the FDM-MCMC has strong applicability. It is not only ap-

Table 1. Identification results by FDM-NMS method.

	N	1	2	3	4	5	6
	Value (m^2/s)	50.786	48.738	89.635	128.299	119.320	149.217
A	Relative error (%)	1.57	16.09	0.41	16.64	8.22	6.58
	Value (m^2/s)	49.880	69.734	100.774	108.782	123.737	155.974
B	Relative error (%)	0.24	0.38	11.97	1.11	4.82	11.41

Note: “value” denotes “identification value”; “A” denotes FDM-NMS method; “B” denotes FDM-MCMC method.

Table 2. Identification results by FDM-NMS method.

N	$\sigma = 0.05$				$\sigma = 0.1$			
	Value (m ² /s)		Relative error (%)		Value (m ² /s)		Relative error (%)	
	A	B	A	B	A	B	A	B
1	49.710	50.109	0.22	0.58	49.710	49.815	0.58	0.37
2	69.934	69.649	0.50	0.09	72.479	69.654	3.54	0.49
3	93.245	93.557	3.95	3.58	95.515	99.364	6.13	10.40
4	112.977	110.355	0.32	2.70	110.229	105.666	0.21	3.94
5	132.976	122.762	5.57	2.29	137.794	127.255	5.99	2.11
6	124.288	149.846	7.03	11.2	186.376	158.931	33.13	13.52

Note: "A" and "B" have the same meaning as in Table 1.

Table 3. RSD of identification value by the two methods %.

N	1	2	3	4	5	6
FDM-NMS	8.61	18.62	14.16	20.55	9.99	62.19
FDM-MCMC	17.81	21.52	23.39	35.16	33.32	30.77

pliable to the constant flow, but also applicable to the unsteady flow.

2) Higher accuracy

The average errors by FDM-MCMC are less than by FDM-NMS method, which are respectably 3.26% and 3.12%. So the precision of identification results by FDM-MCMC are higher than by FDM-NMS in the same circumstances.

3) Stronger anti-noise ability

The average relative errors obtained by FDM-NMS and FDM-MCMC method are both less than the corresponding measurement errors. But the average relative standard deviation obtained by FDM-MCMC is less 25.96% than by FDM-NMS.

In summary, the FDM-MCMC has wide applicability, high identification accuracy and anti-noise ability. It can be better to identify the parameters of water quality model in LDWTP.

5. Conclusion

With the improvement of water quality model function, the identification difficulty is becoming more and more. Therefore, the Finite Difference Method is adopted to solve water quality model in LDWTP, then based on the Bayesian inference the unknown parameters' posterior probability density function is identified, and further the corresponding statistics are obtained by sampling with MCMC to identify water quality model parameters in LDWTP. According to the numerical results, it improves significantly not only on the global convergence and

convergence rate, but also the identification accuracy is relatively higher by the FDM-MCMC in solving water quality model parameter identification problem with higher nonlinear degree. Because the FDM-MCMC has the characteristic of strong targeted to improve, it is applicable to other nonlinear identification problem. In addition, it will cause large amount of calculation for the large required sample number. So as to reduce the computational complexity greatly, combing with other inversion methods and making it possible to apply in real-time identification mode would be the future research direction.

6. Acknowledgements

The research work was supported by the national science and technology major special project (No.2012ZX07205005) and the public welfare industry research project of the ministry of water resources of the People's Republic of China (No.201101063).

REFERENCES

- [1] L. Jijian, W. Xu, L. Chanyu and M. Chao, "Emergency Regulation for Sudden Water Pollution Accidents of Open Channel in Long Distance Water Transfer Project," *Journal of Tianjin University (Science and Technology)*, Vol. 46, No. 1, 2013, pp. 44-50.
- [2] Y. Xiaohua, Y. Zhifeng and L. Jianqiang, "A New Method for Parameter Identification in Water Environment Model," *Advances in Water Science*, Vol. 14, No. 5, 2003, pp. 554-557.
- [3] L. Xiaodong, Y. Qi and X. Hongqin, "Advance in Inverse Problem of Environmental Hydraulics," *Advances in Water Science*, Vol. 20, No. 6, 2009, pp. 885-893.
- [4] J. Ferrer, M. A. Pérez-Martín, S. Jiménez, T. Estrela and J. Andreu, "GIS-Based Models for Water Quantity and Quality Assessment in the Júcar River Basin, Spain, including Climate Change Effects," *Science of the Total Environment*, No. 440, 2012, pp. 43-59.
- [5] X. Hongqin, Z. Chen, L. Xiaodong and G. Li, "Finite Difference Method-Simplex Method for Determination of Longitudinal Dispersion Coefficient in Natural River," *Journal of PLA University of Science and Technology (Natural Science Edition)*, Vol. 13, No. 2, 2012, pp. 214-218.
- [6] A. Afshar, N. Shojaei, and M. Sagharjooghifarahani, "Multiobjective Calibration of Reservoir Water Quality Modeling Using Multiobjective Particle Swarm Optimization (MOPSO)," *Water Resources Management*, Vol. 27, No. 7, 2013, pp. 1931-1947.
<http://dx.doi.org/10.1007/s11269-013-0263-x>
- [7] L. Mengkai, W. Chang De and F. Xxiaobo, "Analysis on the Hydraulic Response of Long Distance Canal Control System during Ice Period," *Transactions of the CSAE*, Vol. 27, No. 2, 2011, pp. 20-27.
- [8] Z. Song, L. Guohua, M. Xinwei, et al., "Identification of Parameters for Standard $k-\epsilon$ Turbulence Model Based on Bayesian Inference," *Journal of Sichuan University: En-*

- gineering Sciences*, Vol. 42, No. 4, 2010, pp. 78-82.
- [9] D. McCarthy, A. Deletic, V. Mitchell and C. Diaper, "Sensitivity Analysis of an Urban Stormwater Microorganism Model," *Water Science and Technology*, Vol. 62, No. 6, 2010, pp. 1393-1400.
<http://dx.doi.org/10.2166/wst.2010.349>
- [10] M. A. Friedrichs, "A Data Assimilative Marine Ecosystem Model of the Central Equatorial Pacific: Numerical twin experiments," *Journal of Marine Research*, Vol. 59, No. 6, 2001, pp. 859-894.
<http://dx.doi.org/10.1357/00222400160497544>
- [11] K. K. Khatua and K. C. Patra, "Flow Distribution in Meandering Compound Channel," *ISH Journal of Hydraulic Engineering*, Vol. 15, No. 3, 2009, pp. 11-26.
<http://dx.doi.org/10.1080/09715010.2009.10514956>