

# Bayesian framework of parameter sensitivity, uncertainty, and identifiability analysis in complex water quality models

Haifeng Jia<sup>\*</sup>, Te Xu, Shidong Liang, Pei Zhao, Changqing Xu

School of Environment, Tsinghua University, Beijing, 100084, China



## ARTICLE INFO

### Article history:

Received 30 May 2017

Received in revised form

15 February 2018

Accepted 1 March 2018

### Keywords:

Water quality model

Parameter

High dimension

Sensitivity

Uncertainty

Identifiability

## ABSTRACT

An efficient Bayesian analytical framework was developed to address the challenges of uncertainty analysis and assess the parameter identification problems of complex water quality models with high-dimensional parameter space. The inclusion of a multi-chain Markov Chain Monte Carlo method and comprehensive global sensitive analysis (GSA) guarantees the results to be robust. A high-frequency synthetic data case study was conducted in the EFDC water quality module including 54 parameters. The comprehensive GSA identified 39 completely or partially sensitive parameters for reducing dimensionality, among which only nine were identifiable without significant bias. The fundamental causes of the parameter identification problem could be traced to the cognitive limitations of the real water quality assessment process instead of data scarcity. The framework is powerful for exploring these limitations, generating reminders for model users to use Bayesian estimates in future forecasts, and providing directions for model developers to perfect a model in future work.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

With rapid urbanization and economic development, water quality deterioration has become a global concern. Serious problems necessitate the prevention and control of water pollution, as well as aquatic ecosystem management. Water quality models are powerful mathematical tools for water quality assessment, pollution control, emergency preparedness and response, and aquatic environmental planning (Mirchi and Watkins, 2012; Melching et al., 2013; Xiao et al., 2015).

Water quality models are established based on an understanding of the relevant hydrodynamic, chemical, and biochemical pollutant migration and transformation processes in aquatic ecosystems, as well as the hypotheses on inaccessible behaviors (Beck, 1987; Melching et al., 1990; Walker et al., 2003; Lindenschmidt et al., 2007). These inaccessible but complicated behaviors are often parameterized based on different state variables (e.g., the Monod model uses two parameters to describe the microbial usage of BOD<sub>5</sub>). With deepening insight into such mechanisms and related processes, water quality models have become increasingly complex. On the one hand, non-monotonic and non-linear

relationships among state variables have replaced the initial monotonic and linear relationships, yielding extensive local optima. On the other hand, dimensions of the parameter space have increased dramatically, and redundant relationships have simultaneously arisen (Refsgaard et al., 2006; Freni et al., 2011), causing extensive equifinality (i.e., multiple “optimal” parameter vectors that yield similar goodness-of-fit).

Parameter identifiability is the possibility of learning the true values of underlying parameters with an infinite experimental dataset (Raue et al., 2009). Parameter identification for complex water quality models is inevitably challenging and parameter true values are often not learned because of the increased computation cycles and the aforementioned factors. Against such serious parameter identification problems (Omlin et al., 2001; Müller et al., 2002; Brun et al., 2002; Raue et al., 2009), an efficient and robust uncertainty analysis (UA) will aid both model users and developers to assess parameter identifiability, and eventually determine the cognitive limitations of the real behaviors of an aquatic ecosystem. In turn, targeted control of model imperfections can mitigate the adverse impacts of non-identifiability and strengthen the reliability of simulation results, thereby preventing decision-making errors (Wagener and Kollat, 2007; Ascough et al., 2008; Jiang et al., 2017).

The Bayesian method represents a modern branch of UA techniques developed based on Bayes' theorem (Eq. (1)). This method describes parameter uncertainty by deriving the posterior

<sup>\*</sup> Corresponding author.

E-mail address: [jhf@tsinghua.edu.cn](mailto:jhf@tsinghua.edu.cn) (H. Jia).

parameter distribution ( $\pi(\Theta|\mathbf{x})$ ) from a combination of prior parameter distribution ( $\pi(\Theta)$ ) and the likelihood function ( $p(\mathbf{x}|\Theta)$ ), in which empirical knowledge (such as past research experience, previous comparable experiments, and even intuition or belief) and sampling information are encoded, respectively.

$$\pi(\Theta|\mathbf{x}) \propto \pi(\Theta)p(\mathbf{x}|\Theta) \tag{1}$$

In general, the explicit functional form of the posterior distribution is unlikely to be derived analytically. Therefore, sampling is indispensable. The Markov Chain Monte Carlo (MCMC) method provides a series of efficient sampling algorithms to obtain the posterior parameter distribution (Tierney, 2015). Multi-chain MCMC methods, represented by the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016), are a substantially competitive branch. Multi-chain MCMC methods are based on a genetic algorithm integrated with the “population” concept. Each Markov chain uses a randomly sampled point from the prior parameter space as the initial population to initiate an evolution. Multiple Markov chains interact with one another to co-generate the transition kernel (Ter Braak, 2006; Ter Braak and Vrugt, 2008), different from single-chain MCMC methods. Hence, multi-chain MCMC methods have been used to improve sampling and searching capabilities, as well as to prevent premature convergence (i.e., being trapped in local optimum) in high-dimensional parameter spaces (Vrugt et al., 2009). These advantages resulted in the current popularity of multi-chain MCMC methods in hydrological research (Keating et al., 2010; He et al., 2011; Harrison et al., 2012; Joseph and Guillaume, 2013). However, relevant applications in complex water quality models still remain insufficient.

To further increase the efficiency of UA, dimensionality reduction is often initially required for complex water quality models. Sensitivity analysis (SA), which assesses the degree to which parameter uncertainty causes output variations, plays an important role in dimensionality reduction via parameter prioritization and fixing (Saltelli et al., 2008; Bilotta et al., 2012; Sun et al., 2012; Ganji et al., 2016). SA techniques are often categorized into local (LSA) and global sensitivity analysis (GSA) methods. LSA is a partial-derivative-based method to investigate the response of a small disturbance of each parameter around a specific location in parameter space on model output (Matott et al., 2009; Baroni and

Tarantola, 2014). A common method of conducting LSA is to utilize the one-factor-at-a-time (OAT) method (Yang, 2011). Although LSA is computationally economical and popular (Luo and Zhang, 2009; Jia et al., 2015; Abdul-Aziz and Al-Amin, 2016), it is not suitable for reducing the dimensionality of complex water quality models, because of its location dependence as well as the lack of knowledge on the suitable location, i.e., the parameter true value.

GSA investigates the effect of the variations over the entire prior parameter space on the model output (Saltelli et al., 2008; Zhan and Zhang, 2013; Pianosi et al., 2016). GSA does not rely on a pre-known suitable location; thus, it overcomes the limitations of LSA. Common GSA methods can be classified into four categories: (i) variance-based methods, such as the Sobol’s method; (ii) entropy-based methods, such as Kullback–Leibler (KL) entropy; (iii) derivative-based methods, such as the Morris screening method; and (iv) regression-based methods, such as standardized regression coefficients (SRC). Due to the ambiguous definition of “global sensitivity”, different GSA methods reveal different relationships between the parameters and model responses (Razavi and Gupta, 2015), which lead to varying results of GSA. Razavi and Gupta (2015, 2016) introduced important characteristics (i.e., local sensitivities and their global distribution, global distribution of model responses, and structural organization of the response’s surface) to interpret global sensitivity and indicated that existing GSA methods (such as the Sobol’s and Morris screening methods) only focus on one or a few of these characteristics while not considering the others. This indicates the strengths and weaknesses of a single GSA method. In addition to Razavi and Gupta (2015, 2016), several other researchers have also recommended comprehensive and complementary use of different GSA techniques for robust SA (Cloke et al., 2008; Pappenberger et al., 2008; Mishra et al., 2009; Neumann, 2012; Cosenza et al., 2013; Gamerith et al., 2013; Wainwright et al., 2013; Gan et al., 2014; Vanrolleghem et al., 2015; Sarrazin et al., 2016).

Peer-reviewed literature remains insufficient for proposing an efficient and robust UA parameter framework to assess the parameter identification problem (i.e., model imperfection) of complex water quality models. “Efficient” refers to the computational frugality of the UA process as well as the achievement of convergence with infinite iterations, which can be realized by reducing the dimensionality. “Robust” indicates that the framework

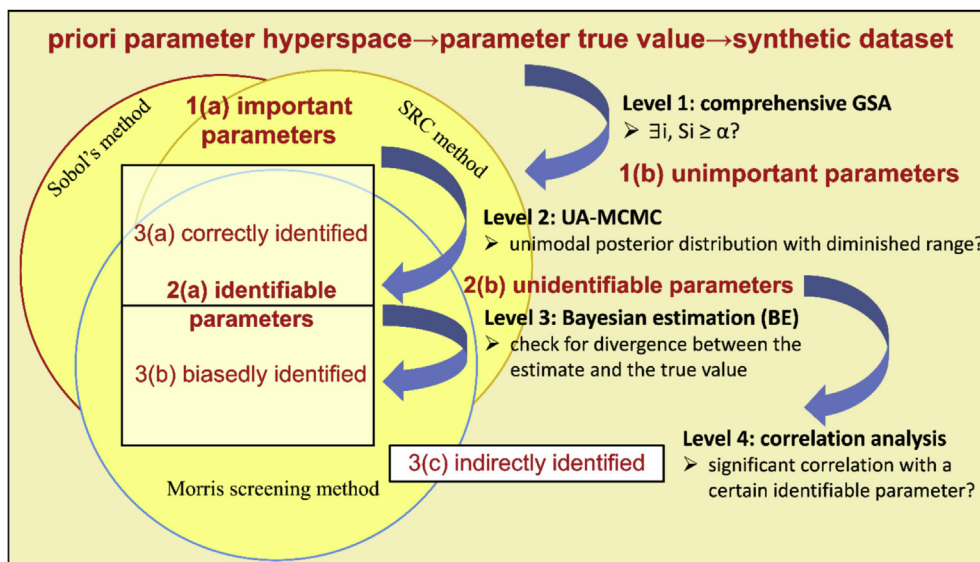


Fig. 1. Analytical framework of parameter sensitivity, uncertainty and identifiability in complex water quality models.

Download English Version:

<https://daneshyari.com/en/article/6962023>

Download Persian Version:

<https://daneshyari.com/article/6962023>

[Daneshyari.com](https://daneshyari.com)