Advances in Water Resources 88 (2016) 109–123

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Multiple-response Bayesian calibration of watershed water quality models with significant input and model structure errors

Feng Han[a,b], Yi Zheng[a,b,*]

[a] College of Engineering, Peking University, Beijing 100871, PR China
[b] Institute of Water Sciences, Peking University, Beijing 100871, PR China

ARTICLE INFO

Article history:
Received 18 March 2015
Revised 15 October 2015
Accepted 3 December 2015
Available online 11 December 2015

Keywords:
Water quality modeling
Non-point source pollution
Bayesian inference
Markov chain Monte Carlo
Multiple-response calibration
Uncertainty analysis

ABSTRACT

While watershed water quality (WWQ) models have been widely used to support water quality management, their profound modeling uncertainty remains an unaddressed issue. Data assimilation via Bayesian calibration is a promising solution to the uncertainty, but has been rarely practiced for WWQ modeling. This study applied multiple-response Bayesian calibration (MRBC) to SWAT, a classic WWQ model, using the nitrate pollution in the Newport Bay Watershed (southern California, USA) as the study case. How typical input and model structure errors would impact modeling uncertainty, parameter identification and management decision-making was systematically investigated through both synthetic and real-situation modeling cases. The main study findings include: (1) with an efficient sampling scheme, MRBC is applicable to WWQ modeling in characterizing its parametric and predictive uncertainties; (2) incorporating hydrology responses, which are less susceptible to input and model structure errors than water quality responses, can improve the Bayesian calibration results and benefit potential modeling-based management decisions; and (3) the value of MRBC to modeling-based decision-making essentially depends on pollution severity, management objective and decision maker's risk tolerance.

1. Introduction

Watershed water quality (WWQ) models, such as Soil and Water Assessment Tool (SWAT) [1,2], Watershed Analysis Risk Management Framework (WARMF) [3–5], and Hydrological Simulation Program - Fortran (HSPF) [6], can provide spatially and temporally distributed simulations of hydrology and water quality variables. WWQ modeling helps enhance our understanding of watershed processes of pollutants, and build explicit linkages between pollution causes and water quality effects. The models have been widely used in addressing water quality management issues [7–11], such as Total Maximum Daily Loads (TMDLs) planning [5]. However, their applications suffered significant modeling uncertainty resulting from inaccurate forcing inputs, model structural inadequacy, uncertain model parameters and observational errors (e.g., errors in water quality measurements) [12–16]. During the past decade, modeling uncertainty has been extensively discussed with regard to hydrology [17], but has received much less attention for water quality [16,18].

Bayesian inference using Markov chain Monte Carlo (MCMC) sampling is an advanced approach for model calibration and uncertainty analysis, which requires an explicit statistical model of residuals (i.e., error model) for rigorous likelihood evaluation. It provides posterior parameter distributions and can be used to assess both parametric and predictive uncertainties. In the field of hydrological modeling, various Bayesian inference approaches have been proposed and applied [19–26], but their applications to WWQ modeling have been limited [27–29]. WWQ modeling integrates pollution simulation with hydrological simulation. However, in existing WWQ models, watershed processes of soil erosion, chemical reactions and pollutant transport are often accounted for by simple equations, which reflects the current knowledge gaps [12–14]. Thus, WWQ modeling generally involves much higher model structure errors than pure hydrological modeling. On the other hand, in WWQ modeling, model inputs of both point sources (e.g., wastewater discharge) and non-point sources (e.g., fertilizer application and atmospheric deposition) loadings are often highly inaccurate due to the lack of loading data at desired temporal and spatial resolutions. For example, it is impractical to continuously monitor chemical concentrations in effluents, and in many cases only yearly or monthly loading estimates are available for daily-step simulations. Another real-world example is that bookkeeping of fertilizer uses is usually poor, and therefore it is impossible to know exactly when and where the historical fertilizer application occurred. The amount and timing of the application are often estimated based on sales data and/or plant growth cycle.

* Corresponding author at: Room 1001, Wangkezhen Building, Peking University, Beijing 100871, China. Tel./fax: +86 10 62768255.
E-mail address: yizheng@pku.edu.cn (Y. Zheng).

http://dx.doi.org/10.1016/j.advwatres.2015.12.007
0309-1708/

Therefore, WWQ modeling also involves much higher input errors than pure hydrological modeling.

Water quality observations (e.g., instream nitrate concentrations) are critical to the calibration of WWQ models, but are often scarce and poorly measured. In contrast, high-frequency (e.g., daily) flow observations at gaging stations are more available and reliable, representing a general supplement to water quality observations. However, using multiple types of observational data to constrain water quality simulation has been a highly ad hoc practice. A recent study [30] conducted a Bayesian calibration for sediment modeling using both flow and sediment observations. It implemented several sequential calibration (i.e., calibrate the model for the flow response first, and then for the sediment response) strategies in parallel, and then fused the results through an ensemble approach. This sequential calibration approach is sound and useful, but a more straightforward alternative would be multiple-response Bayesian calibration. A few studies have shown that including multiple hydrological responses (e.g., flow, soil moisture, etc.) in the Bayesian calibration can improve identifiability of model parameters and adequateness of uncertainty assessment [31,32]. As discussed before, water quality simulation in general involves greater and more complicated modeling uncertainty than hydrological simulation. Thus, the successful experience of multiple-response Bayesian calibration in hydrological modeling may not be fully transferable to WWQ modeling, and further studies are highly desired.

In this study, we investigated the impact of input and model structure errors on model calibration, prediction and management decisions. We considered a particular modeling situation, in which input and model structure errors are very significant and would severely bias the simulation results. This situation is very common in watershed-scale modeling of poorly monitored water quality parameters (e.g., nutrients, metals, pesticides, etc.), but has been rarely investigated in a multiple-response Bayesian calibration context. SWAT and Differential Evolution Adaptive Metropolis (DREAM$_{(ZS)}$), a state-of-the-art MCMC algorithm developed in the field of hydrological modeling [25,33,34], were employed as the WWQ model and MCMC algorithm, respectively. The nitrate pollution in Newport Bay watershed (Southern California, USA) was the study case. A series of numerical experiments were designed and implemented. Overall, the study demonstrated the critical role of input and model structure errors in assessing modeling uncertainty, identifying posterior parameter distributions and making management decisions, and demonstrated the feasibility and importance of performing multiple-response calibration for WWQ models in a management context. It has also been concluded that interpretation of the modeling uncertainty would depend on water quality management concerns.

2. Multiple-response Bayesian calibration

The relationship between a model output (i.e., model response) and its corresponding observation can be expressed as

$$\mathbf{Z} = \mathbf{Y}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{Z}$ and $\mathbf{Y}(\boldsymbol{\theta})$ are the observed and simulated values of the concerned response; $\boldsymbol{\theta}$ is a vector of uncertain model parameters; and $\boldsymbol{\varepsilon}$ is a lumped residual error term. Bayes' rule can be adopted, as

$$p(\boldsymbol{\theta}|\mathbf{Z}) \propto p(\boldsymbol{\theta})L(\boldsymbol{\theta}, \mathbf{Z}) \tag{2}$$

where $p(\boldsymbol{\theta}|\mathbf{Z})$ and $p(\boldsymbol{\theta})$ represent the posterior and prior distributions of $\boldsymbol{\theta}$, respectively; and $L(\boldsymbol{\theta}, \mathbf{Z})$ represents the likelihood function mathematically determined by the error model of $\boldsymbol{\varepsilon}$. In the context of multiple-response calibration, different error models are required for different responses, and therefore multiple likelihood functions are needed [31]. If the residual errors of multiple responses are assumed to be independent, the combined likelihood function (denoted as $L_{multiple}$) is the product of individual likelihood functions [21,23,31]:

$$L_{multiple} = \prod_{i=1} L^i\left(\boldsymbol{\theta}, \mathbf{Z}^i\right) \tag{3}$$

where i indicates the ith response, and $\boldsymbol{\theta}$ is a common set of random model parameters for the multiple responses.

To derive the individual and combined likelihood functions, it is critical to determine the error model of each response. In flow modeling, normal error models have often been assumed for $\boldsymbol{\varepsilon}$ (e.g., [20,24,34]), but have been criticized for being unrealistic [35]. Recent studies proposed more realistic but complex error models [19,22,26,36–38] for flow modeling. For example, in an autocorrelated, heteroscedastic, and non-Gaussian error model [22], the heteroscedasticity is reflected by a linear equation, $\sigma_t = \sigma_0 + \sigma_1 y_t$, where y_t is the simulated flow at time t, σ_t is the estimated standard deviation of the residual error ε_t, and σ_0 and σ_1 are two unknown hyper-parameters. Auto-correlation of residuals is depicted by a first order autoregressive model (i.e., AR(1)), and another hyper-parameter, the lag-1 autoregressive parameter ϕ, is then included. A skew exponential power (SEP) distribution was employed to deal with the skewness and heavy tail issue. The SEP distribution contains two tunable parameters, the skewness parameter ξ and the kurtosis parameter β. More details about the SEP distribution can be found by Schoups and Vrugt [22]. Hence, there are five hyper-parameters (σ_0, σ_1, ϕ, ξ, β), hereafter denoted as a vector $\boldsymbol{\varphi}$, to be inferred in the Bayesian calibration. This error model leads to the following log-likelihood function [22,37]:

$$\log L\left(\boldsymbol{\theta}, \boldsymbol{\varphi}, \mathbf{Z}\right) = n\log\frac{2\sigma_\xi \omega_\beta}{\xi + \xi^{-1}} - \sum_{t=1}^{n} \log\sigma_t - c_\beta \sum_{t=1}^{n} |a_{\xi,t}|^{2/(1+\beta)} \tag{4}$$

where n is the number of observations; $a_{\xi,t} = \xi^{-sign(\mu_\xi+\sigma_\xi a_t)}(\mu_\xi + \sigma_\xi a_t)$, with a_t being an independent and identically distributed random error with zero mean and unit standard deviation; and μ_ξ, σ_ξ, ω_β and c_β are all functions of the skewness parameter ξ and the kurtosis parameter β (see [22] for details). According to Evin et al. [37], applying the AR(1) model to standardized residuals (i.e., $\eta_t = \frac{\varepsilon_t}{\sigma_t}$) instead of raw heteroscedastic residuals can lead to more stable predictive distributions. We followed Evin et al.'s approach and a_t is calculated as

$$a_t = \frac{\eta_t - \phi\eta_{t-1}}{\sqrt{1-\phi^2}} \tag{5}$$

A few studies have specifically discussed error models for water quality responses in the context of Bayesian inference. By Wellen et al. [30], normality and independence of error were simply assumed for sediment modeling. As suggested by Schoups and Vrugt [22], in Bayesian calibration, one could gradually increase complexity of the error model, from a simple Gaussian, homoscedastic and non-autocorrelated model to Eq. (4), until posterior checks confirmed that the residual errors are consistent with the error model assumptions.

3. Data and methods

3.1. Newport Bay Watershed

The Newport Bay Watershed (NBW) is located in Orange County, southern California (Fig. 1). It is a highly urbanized watershed with an area of about 400 km^2. As of 2001, around 70% of NBW was residential, commercial and industrial areas, and agricultural and orchard areas accounted for no more than 8%. It has a typical Mediterranean climate featured by short, mild winters, and dry summers. The annual rainfall depth is about 330 mm, occurring mostly between November and April. About 95% of the freshwater flow volume into the upper