Journal of Environmental Management 140 (2014) 83–92

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Using CV-GLUE procedure in analysis of wetland model predictive uncertainty

Chun-Wei Huang [a], Yu-Pin Lin [a,*], Li-Chi Chiang [b], Yung-Chieh Wang [a]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Department of Civil and Disaster Prevention Engineering, National United University, Miaoli 36003, Taiwan, ROC

A R T I C L E I N F O

Article history:
Received 4 September 2013
Received in revised form
21 February 2014
Accepted 17 March 2014
Available online 12 April 2014

Keywords:
Predictive uncertainty
Uncertainty analysis
GLUE
Characteristic CV
Constructed wetland
Paddy

A B S T R A C T

This study develops a procedure that is related to Generalized Likelihood Uncertainty Estimation (GLUE), called the CV-GLUE procedure, for assessing the predictive uncertainty that is associated with different model structures with varying degrees of complexity. The proposed procedure comprises model calibration, validation, and predictive uncertainty estimation in terms of a characteristic coefficient of variation (characteristic CV). The procedure first performed two-stage Monte-Carlo simulations to ensure predictive accuracy by obtaining behavior parameter sets, and then the estimation of CV-values of the model outcomes, which represent the predictive uncertainties for a model structure of interest with its associated behavior parameter sets. Three commonly used wetland models (the first-order K–C^* model, the plug flow with dispersion model, and the Wetland Water Quality Model; WWQM) were compared based on data that were collected from a free water surface constructed wetland with paddy cultivation in Taipei, Taiwan. The results show that the first-order K–C^* model, which is simpler than the other two models, has greater predictive uncertainty. This finding shows that predictive uncertainty does not necessarily increase with the complexity of the model structure because in this case, the more simplistic representation (first-order K–C^* model) of reality results in a higher uncertainty in the prediction made by the model. The CV-GLUE procedure is suggested to be a useful tool not only for designing constructed wetlands but also for other aspects of environmental management.

Software availability

Name: CV-GLUE uncertainty estimation tool for wetland water quality modeling
Programming language: Visual Basic 2010
Availability: From the authors upon request for research and practical purposes.

1. Introduction

The estimation of the predictive uncertainty of ecosystem models is important and challenging to both the scientific community and decision-makers (Zak and Beven, 1999). Uncertainties in model predictions arise mostly from data, model parameters, and model structure (Rankinen et al., 2006; Refsgaard et al., 2006; Lindenschmidt et al., 2007; Freni et al., 2008, 2009; Parasuraman and Elshorbagy, 2008; David, 2009). These uncertainties contribute to total predictive uncertainty and exhibit various interrelationships, which cannot be easily represented in additive terms. The inadequate establishment or selection of a model structure (conceptual model) often leads to a significant error and uncertainty because a model not only simplifies the real world, but also determines the complexity of relevant parameter sets and input data. Strong simplifications in model representations introduce major uncertainties into environmental assessments (Verburg et al., 2013). Therefore, the analysis of uncertainty due to a model structure should be emphasized during the evaluation of a model's predictive accuracy when used in environmental management and design (Van der Sluijs, 2006).

Many studies have studied uncertainty that is related to model structure. Those studies either explicitly identify the uncertainty that is associated with a model structure (Håkanson, 2000; Lindenschmidt et al., 2007) or elucidate the implicit relationship between predictive uncertainty and model complexity (Snowling and Kramer, 2001; Lindenschmidt, 2006; Parasuraman and

* Corresponding author. Tel./fax: +886 2 3366 3467.
E-mail addresses: yplin@ntu.edu.tw, yplin@staff.pccu.edu.tw (Y.-P. Lin).

http://dx.doi.org/10.1016/j.jenvman.2014.03.005
0301-4797/

Elshorbagy, 2008). Håkanson (2000) discussed inherent uncertainties in two system dynamics-based lake eutrophication model structures, using the characteristic coefficient of variation (characteristic CV) to quantify the uncertainties. The characteristic CV is a simple standard measure of uncertainty that can be used to represent the uncertainty within local data or the values of local parameters (Håkanson, 2000) and it has benefits in comparing uncertainties in different units. Snowling and Kramer (2001) developed a procedure for evaluating modeling uncertainty that took into account model complexity, sensitivity, and predictive error. They hypothesized that as a model became more complex in with more parameters and variables, the predictive error decreased and the overall sensitivity increased. In their study, the overall sensitivity was specified as the area of an envelope bounded by $\pm$ one standard deviation of the output mean. Lindenschmidt (2006) confirmed this hypothesis by analyzing various eutrophication models and found that increasing the number of parameters increased the overall model sensitivity. In their studies, overall model sensitivity referred to the sum of the sensitivities of the outputs to the values of all parameters. Generally, explicit structural uncertainty analysis provides information on the structural uncertainty of a specific model (Håkanson, 2000), while implicit model structure uncertainty analysis yields information that helps in the selection of an appropriate model structure from simple to complicated (Snowling and Kramer, 2001; Lindenschmidt, 2006; Parasuraman and Elshorbagy, 2008). However, the cited studies made few references to structural uncertainty due to the simplification of reality.

Monte-Carlo simulation has been widely used to evaluate the uncertainty of water quality models (Håkanson, 2000; Rankinen et al., 2006; Lindenschmidt et al., 2007; Freni et al., 2008, 2009; Dean et al., 2009; Wang et al., 2010, 2012) and hydrological models (Beven and Binley, 1992; Krysanova et al., 2006; McMichael and Hope, 2007; Blasone et al., 2008; Hassan et al., 2008; Chu et al., 2010). In these studies, the Generated Likelihood Uncertainty Estimation (GLUE) procedure (Beven and Binley, 1992) has been commonly used to quantify the parameter uncertainty based on the results of model calibration. The GLUE procedure is not only comprehensive in estimating the likelihood of all possible outcomes for a specific distribution of inputs but also practical in determining behavioral parameter sets of a model. Based on the results of the GLUE procedure, a further analysis can be performed under a certain predictive accuracy without "non-behavior" (unacceptable) parameter sets (Rankinen et al., 2006; Blasone et al., 2008). However, most relevant studies have tended to focus on parameter uncertainty as a proxy for the predictive uncertainty of a specific model structure.

The aim of this study is to develop a GLUE-based procedure for quantifying the predictive uncertainty of existing wetland water quality models with various model complexities. The relationship between predictive uncertainty and complexity of model structures under the same limitations and assumptions was evaluated to suggest effective models. Three wetland water quality models were used to demonstrate the effectiveness of the GLUE-based procedure. They are the first-order $K–C^*$ model ($K–C^*$ model), the plug flow model with dispersion (PFD model), and the system dynamics-based wetlands water quality model (WWQM model). The $K–C^*$ model is often used for designing constructed wetlands (Kadlec, 2000). This model is based on the assumption that the concentration profiles of nutrients decrease exponentially with distance, instead of real nutrient interactions in wetlands. The $K–C^*$ model is practical due to its simple model structure with very few measurable driving variables. However, the inadequacy of the $K–C^*$ model lies in the deterministic, constant rate of areal removal and a lack of internal chemical and physical process information in the model structures (Kadlec, 2000). More complex model structures than that of the $K–C^*$ model have been designed to capture the chemical and physical processes of wetlands in greater detail, such as the plug flow with dispersion model (PFD model) (Kadlec and Knight, 1996; Kadlec, 2000; Person and Wittgren, 2003) and the system dynamics model (Wynn and Liehr, 2001; Dardona, 2004; Marsili-Libelli and Checchi, 2005; Chavan and Dennett, 2008; Wang et al., 2012). To determine the effect of model complexity on the modeling of wetland water quality, the proposed CV-GLUE procedure was applied to analyze the predictive uncertainty of the aforementioned models using water quality data obtained by sampling a constructed wetland with paddy cultivation. This work shows that the proposed procedure provides comprehensive information on the predictive uncertainty of different model structures to assist in the evaluation of models used in the design and management of constructed wetlands.

2. Material and methods

2.1. CV-GLUE procedure

The CV-GLUE procedure is developed to analyze predictive uncertainty. It is based on the GLUE procedure and uses two-stage Monte-Carlo simulations. The interface for implementing the CV-GLUE procedure is developed in Visual Basic 2010. This procedure comprises six steps. First, the structure of the model is defined (step 1); then, the model is calibrated and validated using genetic algorithms (GA) (Goldberg, 1989) to optimize the parameter set (step 2). The value of each optimal parameter is used as a central value in setting up an approximate range of parameters that is used in the next step. The purpose of step 2 is to help modelers obtain the behavior parameter sets more efficiently when local data do not suffice to identify the ranges of related parameters (Wang et al., 2010, 2012). The results are compared with values in the literature to confirm that those values fall within those ranges. More details of the GA procedure in step 2 are described in Appendix S1.

The first-stage Monte-Carlo simulation consists of two steps (step 3 and step 4). In step 3, the approximate range is established as zero (-100%) to double the value ($+100\%$) of each optimal value of each parameter. The values of many parameter sets are estimated using Nash and Sutcliffe (1970) efficiency (NSE) as the likelihood measure. In step 4, the behavior parameter sets are determined after an arbitrary predictive accuracy (NSE > 0.6) is specified as the acceptability threshold in both calibration and validation processes. Based on these behavior parameter sets, various acceptability thresholds from 0.6 to 0.8 are established for further analyses in the following step. Therefore, the effect of the subjective choice of the acceptability threshold on predictive accuracy can be assessed. The likelihood function is defined as follows:

$$L(\theta_i|Y) = 1 - \sigma_i^2/\sigma_{\mathrm{obs}}^2, \tag{1}$$

where σ_i is the error variance associated with the ith realization of a selected model; σ_{obs} is the observed variance for the period of interest, and θ_i is the selected behavior parameter sets. In this study, the values of a total of 1,000,000 parameter sets were estimated from a uniform distribution, to select at least 10,000 behavior parameter sets.

Based on the behavior parameter sets that were obtained in the first-stage Monte-Carlo simulation, the second-stage Monte-Carlo simulation was performed to estimate the predictive uncertainty of model structure of interest (step 5). The predictive uncertainty of a considered structure $Y = f(X_1, X_2, \ldots, X_m)$ depends on the uncertainties of the individual state variables X_i, $i = 1$, m (Tsai and