Physics and Chemistry of the Earth 42–44 (2012) 31–41

Contents lists available at SciVerse ScienceDirect

Physics and Chemistry of the Earth

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

Uncertainty estimation of a complex water quality model: The influence of Box–Cox transformation on Bayesian approaches and comparison with a non-Bayesian method

Gabriele Freni [a,*], Giorgio Mannina [b]

[a] Facoltà di Ingegneria ed Architettura, University of Enna "Kore", Cittadella Universitaria, 94100 Enna, Italy
[b] Dipartimento di Ingegneria Civile, Ambientale ed Aerospaziale, Università di Palermo, Viale delle Scienze, 90146 Palermo, Italy

ARTICLE INFO

Article history:
Available online 5 September 2011

Keywords:
Bayesian inference
Water-quality modelling
Generalised Likelihood Uncertainty Estimation (GLUE)
Integrated urban drainage systems
Receiving water body
Wastewater treatment plant

ABSTRACT

In urban drainage modelling, uncertainty analysis is of undoubted necessity. However, uncertainty analysis in urban water-quality modelling is still in its infancy and only few studies have been carried out. Therefore, several methodological aspects still need to be experienced and clarified especially regarding water quality modelling. The use of the Bayesian approach for uncertainty analysis has been stimulated by its rigorous theoretical framework and by the possibility of evaluating the impact of new knowledge on the modelling predictions. Nevertheless, the Bayesian approach relies on some restrictive hypotheses that are not present in less formal methods like the Generalised Likelihood Uncertainty Estimation (GLUE). One crucial point in the application of Bayesian method is the formulation of a likelihood function that is conditioned by the hypotheses made regarding model residuals. Statistical transformations, such as the use of Box–Cox equation, are generally used to ensure the homoscedasticity of residuals. However, this practice may affect the reliability of the analysis leading to a wrong uncertainty estimation. The present paper aims to explore the influence of the Box–Cox equation for environmental water quality models. To this end, five cases were considered one of which was the "real" residuals distributions (i.e. drawn from available data). The analysis was applied to the Nocella experimental catchment (Italy) which is an agricultural and semi-urbanised basin where two sewer systems, two wastewater treatment plants and a river reach were monitored during both dry and wet weather periods. The results show that the uncertainty estimation is greatly affected by residual transformation and a wrong assumption may also affect the evaluation of model uncertainty. The use of less formal methods always provide an overestimation of modelling uncertainty with respect to Bayesian method but such effect is reduced if a wrong assumption is made regarding the residuals distribution. If residuals are not normally distributed, the uncertainty is over-estimated if Box–Cox transformation is not applied or non-calibrated parameter is used.

1. Introduction

Up to day, although several models have been developed, integrated urban water quality modelling still presents numerous difficulties (Mannina and Viviani, 2010; Freni et al., 2009a; Mannina, 2005; Rauch et al., 2002). The field data needed for model calibration/validation are generally limited and monitoring campaigns are usually characterised by measurements carried out at the watershed outlet, thus being representative of the combined effects of both the accumulation and transport of pollutants throughout the overall system (Ashley et al., 2004; Kanso et al., 2006; Freni et al., 2009a). The unbalance between data availability and model complexity reduces the reliability of model results, transferring uncertainty to the model outputs (among others, Freni et al., 2011). Data and parameters uncertainties are often lumped in all the cases where no data are available for specifically assessing uncertainty connected to measures.

Uncertainty analysis became a very useful tool for evaluating model reliability and the growing interest of researchers on this topic is demonstrated by the increasing literature production of recent years. However, in the field of urban drainage modelling, uncertainty analysis is still in its infancy and only a few studies (among others: Lindblom et al., 2007; Freni et al., 2008, 2010; Schellart et al., 2008; Willems, 2008; Deletic et al., 2009a,b; Dotto et al., 2009; Kleidorfer et al., 2009) have been carried out compared with other research field such in the case of hydrology. Indeed, the assessment of uncertainties in urban drainage models is not wide spread in practice and is usually an academic exercise (Deletic

* Corresponding author.
E-mail addresses: gabriele.freni@unikore.it (G. Freni), mannina@idra.unipa.it (G. Mannina).

1474-7065/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pce.2011.08.024

et al., 2011). This is mainly because the techniques required for this analysis are so numerous, highly complex, poorly understood, and some are still highly underdeveloped. Clear and comprehensive comparisons of these techniques when applied to typical drainage models would therefore be desirable.

Uncertainty of a model can be represented in two ways: by giving a range (or a band) of values or by providing a posterior distribution of parameter values. The former entails to likely enclose the true value of a specific simulated variable: lower uncertainty is connected with stricter uncertainty bands. Conversely, the latter consists in providing a posterior distribution of parameter values in which the "true" value should be included. Larger bands and more uniform distributions are usually caused by too complex models coupled insufficient and/or poor quality data. Using the concept of uncertainty, the "best" model is the one able to correctly simulate a specific variable minimising the width of uncertainty bands.

None of the uncertainty analysis methods presented in literature are universally accepted in urban drainage modelling (Freni et al., 2009b). These methods range from classical Bayesian techniques (among others: McCarthy et al., 2008; Haydon and Deletic, 2009), to the pseudo-Bayesian ones (among others: Freni et al., 2008; Thorndahl et al., 2008).

All Bayesian techniques start from prior knowledge regarding the modelling error (that is assumed to reflect a user-selected probability distribution) and parameter distributions thus updating them according to available data by means of Bayes' theorem application. The most used pseudo-Bayesian technique is the Generalised Likelihood Uncertainty Estimation (GLUE) proposed by Beven and Binley (1992). GLUE, like other pseudo-Bayesian approaches, neglects any prior knowledge because of the complexity of the model and of the physical system. Bayesian methods are thus conditioned by prior assumptions, that have to be verified and may provide unreliable results in terms of final model uncertainty, and by transformations applied to the modelling outputs and to data in order to make such hypotheses applicable (Freni et al., 2009b). The use of transformation functions, such as the Box–Cox transformation, is used to reduce the heteroscedasticity of modelling errors making the hypothesis of normally distributed errors and allowing the use of a normal Bayesian likelihood function.

The application of such transformation may reduce the advantages provided by Bayesian methods in the cases where the residuals distribution is heavily skewed.

Bearing in mind the considerations discussed above, with the aim to explore the applicability of new uncertainty methods derived by other research fields, the present paper is aimed to explore the impact of the application of the Box–Cox transformation on Bayesian uncertainty analysis for environmental water quality models. To accomplish such a goal, Box–Cox parameter setting was initially calibrated on the basis of the evaluation of residuals homoscedasticity. Further, some non-calibrated values of Box–Cox parameter were considered as well in order to simulate different grades of residual skewness. A comparison among the results of the Bayesian and GLUE analyses were carried out. As a common hypothesis for all the study, the developed uncertainty analysis is based on the principle that input data uncertainty, model structural uncertainty and measured uncertainty can be lumped into model parameter uncertainty (Lindblom et al., 2007).

The model has been applied to the Nocella experimental catchment (Italy) where both quantity and quality data were available.

2. Materials and methods

2.1. The adopted model

In the present study, a bespoke urban integrated model developed during previous studies was applied (Mannina, 2005). The structure of the adopted model will be briefly described, further details of the model can be found in Mannina (2005) and Mannina and Viviani (2010). The model is able to assess the main phenomena that take place throughout the three components of the integrated system, i.e. sewer system (SS), wastewater treatment plant (WWTP) and receiving water body (RWB). The integrated model is composed mainly of three sub-models for the simulation of the components; each sub-model is divided into a quantity and quality module for the simulations of the hydrographs and pollutographs.

The model equations and parameters are presented in Tables 1 and 2. More specifically, the equations have been grouped considering the three main sub-models, i.e. SS, WWTP and RWB.

The SS sub-model starts to evaluate the net rainfall, from the measured hyetograph, by a loss function (taking into account surface storage and soil infiltration). From the net rainfall, the model simulates the net rainfall–runoff transformation process and the flow propagation with a cascade of one linear reservoir and a linear channel, representing the catchment, and a linear reservoir, representing the sewer network (Eqs. (1) and (2)).

To simulate the build-up on the catchment surfaces an exponential function (Eq. (3)) was adopted (Alley and Smith, 1981). The solid wash-off caused by overland flow during a storm event was simulated with the formulation in Eq. (4) proposed by Jewell and Adrian (1978). The solids deposits in the sewers during dry weather have been evaluated by adopting an exponential law (Eq. (5)). Regarding the sewer sediments erosion as well as transport in order to have a realistic and correct approach a particular care has been taken about sediments transformation in sewers, considering their cohesive-like behaviour linked to organic substances and to the physical–chemical changes during the sewer transport (Crabtree, 1989; Ristenpart, 1995). In particular, the transport equation proposed by Parchure and Mehta (1985) (Eq. (6)) couplet to the bed sediment structures hypothesised by Skipworth et al. (1999) (Eq. (7)) to simulate the sediment erosion rate was considered. The pollutographs at the outlet of the sewer system have been evaluated by hypothesising the complex catchment sewer network as a reservoir (Eqs. (8) and (9)) and by considering the transport capacity of the flow (Eq. (10)). Finally, the WWTP inflow has been computed taking into account the presence of CSO device: its behaviour has been simulated by the Eq. (11), where CSO efficiency has been taken into account by introducing the dilution coefficients, rd1 and rd2. The SWT inflow and outflows quali–quantitative characteristics have been considered by the Eqs. (12) and (13).

Regarding the WWTP sub-model substrate and microorganisms concentration in the activated sludge tank have been calculated with mass balances based on Monod's theory (Eqs. (14) and (15). The sedimentation tank performance has been simulated using the solid flux theory according to the methodology proposed by Takács et al. (1991). In particular, the solids concentration profile is obtained by dividing the settler into 50 horizontal layers of constant thickness. Within each layer the concentration is assumed to be constant and the dynamic update is performed by imposing a mass balance for each layer. Three sets of equations are adopted for the layers depending on their depth: the first set (Eq. (16)) is used for the upper region of the tank (the clarification region), the second one (Eq. (17)) for the lower layer (the thickening zone) and the third for the feed layer (Eq. (18)). The settling velocity function (Eq. (19)) proposed by Takács et al. (1991) is employed. Regarding the RWB sub-model the exemplified form of the De Saint Venant equation (kinematic wave) for the quantity module (Eq. (20)) and the dispersion advection equation for the quality module were adopted Eqs. (21) and (22). The dispersion coefficient E has been evaluated by using the expression suggested by Elder (Eq. (23), where only the vertical velocity gradient is considered important in stream flow (Brown and Barnwell, 1987).