BAYESIAN CALIBRATION OF A RWQM1-BASED RIVER WATER QUALITY MODEL

CRISTINA MARTIN (1,2), MARC B. NEUMANN (1), PETER A. VANROLLEGHEM (1), EDUARDO AYESA (2)
(1): modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 Avenue de la Médecine, Québec, QC, G1V 0A6, Canada.
(2): Environmental Engineering Department, CEIT and Tecnun (University of Navarra), Manuel de Lardizábal 15, 20018, San Sebastián, Spain.

Uncertainty-based modelling was used to describe the biochemical activity in the Tajo River (Spain). The biochemical model is a reduced version of the River Water Quality Model no. 1 (RWQM1). The transport model includes a *Particulate Transport Factor* (*PTF*), which disconnects the hydraulic and solids' retention times along the river and enables the description of biomass accumulation in some river zones. The water quality model is implemented in the WEST simulation platform while the calibration approach, based on the SCEM-UA algorithm, is coded in Matlab. A successful model calibration has been performed using longitudinal profiles of ammonium, nitrates and dissolved oxygen for winter and summer periods. The main feature of the approach is to map the uncertainty to only three, transport-related parameters: the *PTF*, the volumetric oxygen mass transfer coefficient in winter (K_La February), and its equivalent in summer (K_La August). It is found that the uncertainty about the optimal *PTF* value is considerably lower than that of the K_La February or K_La August parameters. The uncertainty propagation shows that the model is able to properly describe the oxygen and nitrogen in-river profiles with very low uncertainty (narrow confidence intervals). The quality of the results encourages further use of the model for the assessment of water quality management scenarios in the Tajo basin.

INTRODUCTION

River water quality modelling has been subject of research for a long time. Since the development of the Streeter Phelps model (Streeter and Phelps [1]) many improved models and new software packages have been proposed (Cox [2]). The RWQM1 (Reichert *et al.* [3]) represents a big step forward in river water quality modelling. It includes the different biomass species as explicit state variables and guarantees a closed mass balance by defining the model components in terms of the mass fractions of elemental compounds (C, H, O, N and P). However, the complexity of the model structure has not encouraged a widespread use of the model and few model applications have been published (Benedetti *et al.* [4]; Borchardt and Reichert [5]; De Keyser *et al.* [6]; De Schepper *et al.* [7]; Martin *et al.* [8]; Reichert [9]; Vanrolleghem *et al.* [10]).

The Bayesian approach is being proposed for model calibration in a wide range of contexts (Omlin and Reichert [11]; Beven and Freer [12]; Feyen *et al.* [13]). Besides the optimum parameter values, it provides an assessment of the associated uncertainty.

Additionally, the reliability of the model responses can be easily assessed by forward propagation of the parameter uncertainty (Beck [14]). This is especially interesting when the model needs to be used for the exploration of other possible scenarios.

This paper proposes the use of RWQM1 to describe the biochemical activity of the Tajo River (Spain). A Bayesian Monte Carlo approach is used to calibrate the model and the uncertainty of the estimated values is assessed. The research work has delivered a complete software structure which is ready for use in other case studies.

MATERIALS AND METHODS

River Water Quality Model

The Tajo River is represented by a series of 36 Continuously Stirred Tank Reactors (CSTR) which ensures a hydraulic residence time distribution (HRTD) close to a plug flow reactor. The temporal change of the model components (water quality parameters) is calculated as a function of the transport and the biochemical reaction components (Eq. 1).

$$\frac{\partial X_i^j}{\partial t} = \frac{\partial X_i^j}{\partial t}\Big|_{transport} + \frac{\partial X_i^j}{\partial t}\Big|_{biochemical} \quad (1)$$

where X_i^j represents the concentration of biochemical model component *i* in tank *j*.

The biochemical model is a reduced version of RWQM1 (Reichert *et al.* [3]). It includes 14 components and 13 biochemical transformations to describe the organic material biodegradation, nitrification and denitrification processes, as well as biomass population dynamics and phosphate precipitation. The model assumes no significant changes of pH during the process. The model components are defined by means of their elemental mass fractions, in terms of C, H, N, O and P. This model building methodology ensures closed mass balances throughout the system. The model stoichiometric matrix, $[E_{p,i}]$, evaluates the amount of component X_i that appears or disappears for each transformation process, *p*. On the other hand, the kinetic vector evaluates the rate at which each transformation takes place (Eq. 2).

$$\frac{\partial X_i^j}{\partial t}\Big|_{biochemical} = \left[E_{p,i}\right]^T \cdot \left\{\rho_p\right\} \quad (2)$$

The transport model distinguishes between soluble and particulate components by means of a Particulate Transport Factor *PTF* (Martín *et al.* [8]). This number (between 0 and 1) defines the fraction of particulates in the bulk liquid. Eq. (3) describes the transport equations for soluble and particulate components:

$$\frac{\partial X_{iSol}^j}{\partial t}\Big|_{transport} = \frac{Q}{V_j}\left(X_{iSol}^{j-1} - X_{iSol}^j\right)$$

$$\frac{\partial X_{iPar}^{j}}{\partial t}\Big|_{transport} = \frac{Q}{V_j} \cdot PTF \cdot \left(X_{iPart}^{j-1} - X_{iPar}^{j} \right) \quad (3)$$

The *PTF* parameter leads to a disconnection between the *HRT* (Hydraulic Retention Time) and the *SRT* (Sludge Retention Time) of the river since a fraction (1-*PTF*) of particulate components is retained in the tank at each time step. This parameter allows for the accumulation of particulate components in some zones of the river. This effect might represent many different phenomena: the solids deposition, the inert material wash out, the biomass growth on the river bed, etc. Obviously, the *PTF* parameter should be adjusted depending on the case study.

Integrated Monte Carlo Methodology

The Bayesian approach treats the model parameters as random variables (Omlin and Reichert [11]) where the probability of each value (θ) represents a degree of belief. Before model calibration, some prior probability distribution, p(θ), defines the modeller's knowledge about possible parameter values. The calibration process updates the prior beliefs based on the evidence of experimental data (in the following y), providing the joint posterior distribution of the parameters, p(θ|*y*).

Assuming i.i.d. residuals (*e*), and using uniform prior distributions, Gamma integrals can be used to integrate out the influence of the residual error (see Box and Tiao [15]) and the posterior density leads to Eq. (4):

$$p(\theta|e) \propto \left[\sum_{i=1}^{m} e_i^2 \right]^{-\frac{m}{2}} \quad (4)$$

where *m* is the number of experimental data. For more information about the Bayesian inference scheme, the reader is referred to Box and Tiao [15] and to Thiemann *et al.* [16].

The *Integrated Monte Carlo Methodology* (*IMCM*), recently proposed by Martin and Ayesa [17], is a complete software package intended to automatically solve calibration problems related to biochemical models. It includes four modules: Markov Chain Monte Carlo (MCMC), Moving Feasible Ranges (MFR), Statistical Analysis of the Joint Posterior Distribution (SAD) and Uncertainty Propagation Analysis (UPA).

The MCMC module is based on the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm, initially proposed by Vrugt *et al.* [18]. SCEM-UA is a widely-used Markov Chain Monte Carlo algorithm that uses a multi-chain approach based on the Metropolis Sampler (Metropolis *et al.* [19]) and Shuffled Complex Evolution algorithm from Duan *et al.* [20]. The main drawback of the algorithm is that the Markov Chain does not evolve freely in the entire parameter space but only within the bounds defined by the prior parameter distributions (uniform distributions). These bounds are defined by the modeller and define the region in which the joint posterior distribution is going to be sampled. If these ranges are well suited, presumably a single SCEM-UA run will completely define the posterior distribution, but otherwise, a new SCEM-UA trial has to be launched using some

new updated ranges. The MFR algorithm automates this process by launching the SCEM-UA algorithm several times until the joint posterior distribution is properly addressed.

Once the joint posterior distribution is obtained, the SAD module estimates a long list of parameter statistics (mean values, standard deviations, coefficients of variation, etc.) enabling the assessment of the parameters uncertainty. Finally, the UPA module performs the uncertainty propagation analysis from model parameters to model responses. The module assesses the uncertainty around the optimum model response (using the optimum parameter values) by the Confidence Intervals associated to Parameter Uncertainty (CIPU) and the Confidence Intervals associated to Total Uncertainty (CITU). The CIPU bounds demarcate the zone in which the model responses are most probably going to be found (with respect to parameter uncertainty) while the CITU predict the zone in which a new set of experimental data would be expected (by including the effect of the residual errors).

Software Set Up and Calibration Strategy

The simulation platform WEST (Vanhooren *et al.* [21], www.mikebyDHI.com) is used to implement the proposed water quality model. The Fix Volume Tank default WEST model is modified to account for the parameter *PTF*, and the biochemical model transformations are also incorporated. The *IMCM* is implemented in Matlab R2007b, using nested functions written as independent m-files. The WEST and Matlab packages are connected by using the TornadoMEX (WEST provided) API function which calls *ad hoc* scripts of the case study.

The SCEM-UA algorithm in the MCMC module uses 10,000 model simulations (Figu
algo
MCI
poste

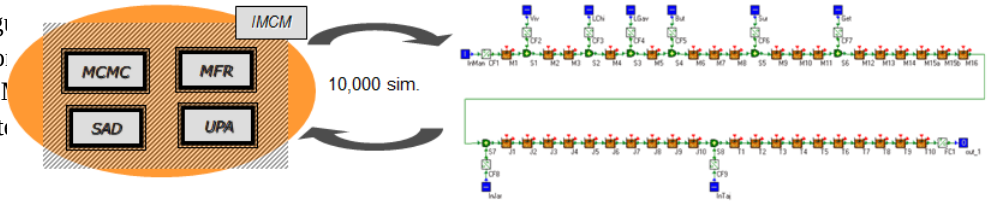

Figure 1. Schematic representation of *IMCM* and WEST simulation platforms.

In the case of the Tajo basin, the main goal lies in the description of the longitudinal nitrogen and oxygen profiles in summer and winter scenarios. During model calibration only three parameters were estimated: the *Parameter Transport Factor* (*PTF*), the volumetric oxygen mass transfer coefficient in winter (K_La February), and its equivalent in summer (K_La August). The parameter *PTF* is supposed to be the same in each river stretch. On the other hand, the transfer of oxygen to the liquid phase (K_La) varies seasonally as it depends on the water level, water temperature, etc. The objective is to explore to which extent the model can describe the experimental data and to assess the uncertainty involved in that estimation.

RESULTS AND DISCUSSION

Steady-state simulations of NH_4, NO_3 and DO concentration profiles along the river branches in February and August of 2003 are used for the calibration of *PTF*, K_La February

and K_La August parameters. Figure 2 shows the calibration result. The optimum parameter values (*PTF* = 0.17, K_La February = 13.7 day^{-1} and K_La August = 12.2 day^{-1}) are estimated and the posterior distributions are well defined. It can also be concluded that the uncertainty around the *PTF* optimum value is considerably lower (Coefficient of Variation, CV=0.076) than that of the K_La August (CV=0.479) or K_La February (CV=0.136).

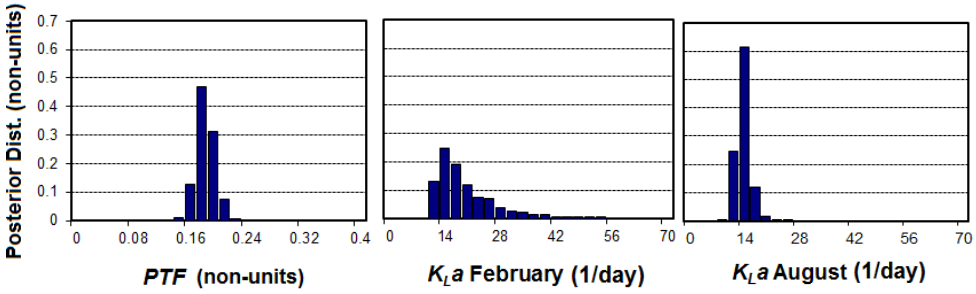

Figure 2. Marginal posterior distributions of the calibrated parameters: PTF, K_La February and K_La August

The parameter *PTF* defines the transport of particulate components along the river. Once the transport is defined, the spatial profile of the different microorganism species is dynamically estimated by the model, since the biomass maintains a constant equilibrium with the available substrate provided by the wastewater treatment plant (WWTP) discharges. In this manner, for high *PTF* values long river distances are needed to achieve a given biodegradation, while low *PTF* values are associated with highly concentrated biochemical activity. The final value of the *PTF* parameter reflects that the biodegradation rates observed in the river (by means of the NH_4, NO_3 and DO experimental profiles) matches those described by the model. Obviously, this parameter turns out to be highly identifiable. The K_La February and K_La August parameters determine the in-river dissolved oxygen concentration. The K_La February is less identifiable than the K_La August because the dissolved oxygen concentrations in winter are close to the saturation limit and therefore, high values of the K_La February return similar DO in-river profiles.

Figure 3 shows the uncertainty propagation results. The optimum parameter values were reasonably able to reproduce the experimental profiles of NH_4, NO_3 and DO (dashed line). It is also easy to observe that WWTPs' discharges lead to a high NH_4 concentration in the Manzanares River, which is subsequently nitrified along the Jarama and Tajo stretches. Comparing February to August profiles, it can be seen that higher summer temperatures prompt a faster biodegradation of NH_4 leading to a faster increase in NO_3. Additionally, the in-river DO concentration declines in August compared to February both in the discharge zone as well as the Jarama stretch due to the increase of nitrification activity and lower solubility.

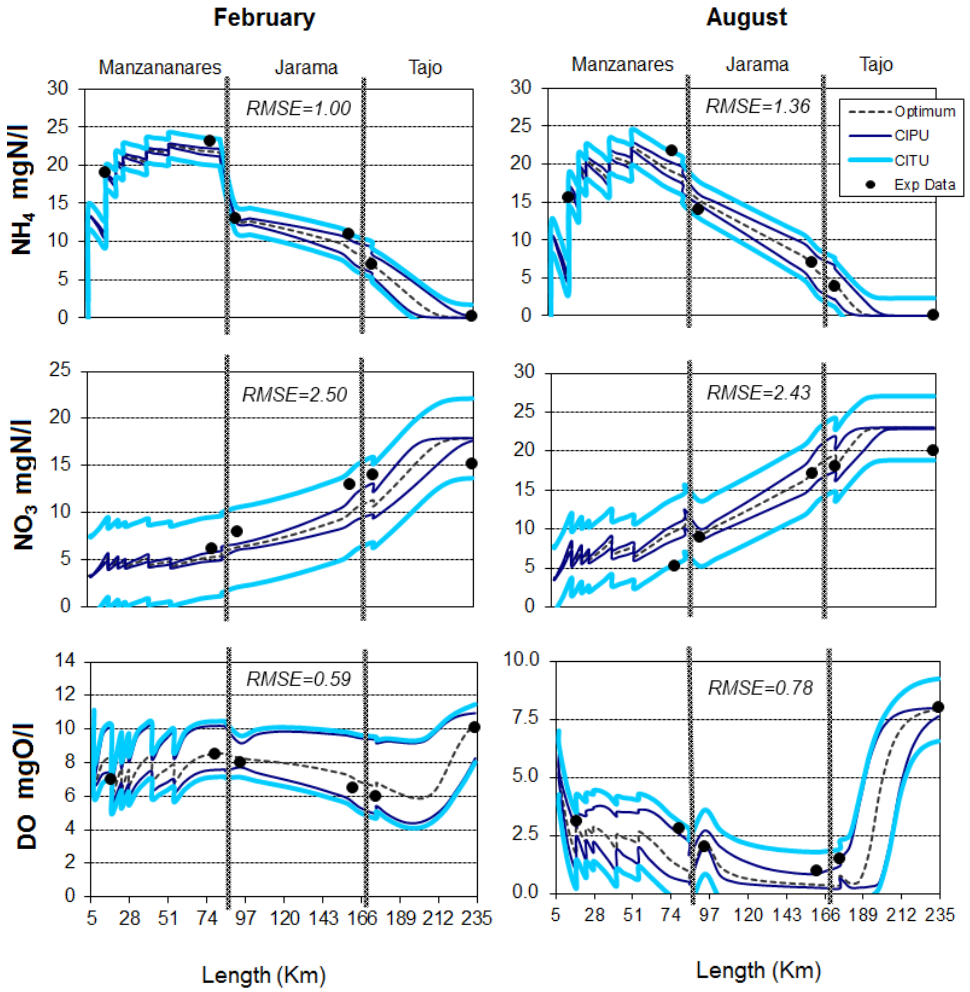

Figure 3. Uncertainty propagation on NH_4 (mgN/l), NO_3 (mgN/l) and DO (mgO/l) model simulation responses. CIPU (dark lines) associated with parameter uncertainty and CITU (light lines) representing the total uncertainty in terms of parameters and residual errors.

Figure 3 also depicts the Confidence Intervals: CIPU (dark lines) associated with parameter uncertainty and CITU (light lines) representing the total uncertainty in terms of parameters and residual errors. On the one hand, the CITU bounds bracket most of the experimental data, indicating appropriate model behaviour. On the other hand, the CIPU ranges are most of the times quite narrow indicating that there is little uncertainty around the optimum simulation response. In this respect, it is evident that the CIPU become wider when assessing the uncertainty of the model to describe the oxygen concentration in February: winter temperatures make oxygen concentrations approach the saturation limit, which leads to considerable uncertainty in the definition of the $K_L a$ February optimum value, and therefore higher uncertainty in the oxygen simulation responses. The CIPU

bounds can also serve to detect possible inconsistencies in experimental data, as they are depicting the zone where the model responses are expected to be. For example, assuming that the NH_4 discharged by WWTPs is completely nitrified along the studied branch, the CIPU indicate that NO_3 in-river concentrations should reach higher concentrations at the end of the Tajo branch both in February and August.

The reasonably low uncertainty expected in the model responses (see CIPU bounds) encourages further utilization of the model for exploration purposes. For example, as most of the WWTPs in the studied area were designed for C removal only, the investigation of a new scenario where they are upgraded for C and N removal would be very interesting.

CONCLUSIONS

The biochemical activity in the Tajo Case Study has been described using a RWQM1-based model. The *Integrated Monte Carlo Methodology* (*IMCM*), based on Bayesian inference, has served to calibrate the model using ammonium, nitrate and oxygen data. The study allowed estimating the uncertainty in the parameters and in the model responses.

- Only three transport-related parameters have been calibrated: the *Particulate Transport Factor* (*PTF*), the volumetric oxygen mass transfer in winter (K_La February), and the volumetric oxygen mass transfer in summer (K_La August). The *PTF* parameter allows including the effect of the biomass accumulation in some zones of the river, and therefore, describing the high biochemical activity observed in the Jarama and Tajo branches.
- The *Integrated Monte Carlo Methodology* (*IMCM*) has properly yielded the joint posterior distribution of the parameters. In this sense, the optimum *PTF* value was more certainly defined than the ones of K_La August or K_La February parameters. *IMCM* has facilitated the uncertainty evaluation not only with respect to the parameter values but also in terms of the model responses. The model is now able to describe the expected variability of the data around the main process behaviour.

The suitability of the results encourages further use of the model for the exploration of new scenarios, such as the one in which the main WWTPs of the zone were upgraded to fulfil N removal.

ACKNOWLEDGEMENTS

This work was made possible through the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Peter Vanrolleghem holds the Canada Research Chair in Water Quality Modeling.

REFERENCES

[1] Streeter W.H. and Phelps E.B., "*A Study of the Pollution and Natural Purification of the Ohio River*", Public Health Bull.146, U.S. Public Health Service, Washington D.C, USA, (1925).

[2] Cox B.A., "A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers", *Sci. Total Env.*, Vol. 314, No. 316, (2003), pp 335-377.

[3] Reichert P., Borchardt D., Henze M., Rauch W., Shanahan P., Somlyódy L. and Vanrolleghem P.A., "*River Water Quality Model No.1*". Scientific and Technical Report No. 12, IWA Publishing, London, UK (2001).

[4] Benedetti L., De Keyser W., Nopens I. and Vanrolleghem P.A., "Probabilistic modelling and evaluation of wastewater treatment plant upgrades in a water quality based evaluation context". *J. Hydroinformatics*, Vol. 12, (2010), pp 380-395.

[5] Borchardt D. and Reichert P., "River Water Quality Model no. 1 (RWQM1): Case Study I. Compartimentalisation approach applied to oxygen balances in the River Lahn (Germany)". *Wat. Sci. Tech.*, Vol. 43, No. 5, (2001), pp 41-49.

[6] De Keyser W., Gevaert V., Verdonck F., Nopens I., De Baets B., Vanrolleghem P.A., Mikkelsen P.S. and Benedetti L., "Combining multimedia models with integrated urban water system models for micropollutants". *Wat. Sci. Tech.*, Vol. 62, No. 7, (2010), pp 1614-1622.

[7] De Schepper V.C.J., Holvoet K.M.A., Benedetti L., Seuntjens P. and Vanrolleghem P.A. "Extension of the River Water Quality Model N°1 with the fate of pesticides". *J. Hydroinformatics*, Vol. 14, (2012), pp 48-64.

[8] Martín C., Cardona C.M., San Martin D., Salterain A. and Ayesa E., "Dynamic simulation of water quality in rivers based on the IWA RWQM1. Application of the new simulator Calhidra 2.0 to the Tajo River", *Wat. Sci. Tech.*, Vol. 54, No. 11-12, (2006), pp 75-83.

[9] Reichert P., "River Water Quality Model no. 1 (RWQM1): Case Study II. Oxygen and nitrogen conversion processes in the River Glatt (Switzerland)", *Wat. Sci. Tech*., Vol. 43, No. 5, (2001), pp 51-60.

[10] Vanrolleghem P.A., Benedetti L. and Meirlaen J. "Modelling and real-time control of the integrated urban wastewater system". *Environ. Modelling & Software* Vol. 20, (2005), pp 42-442.

[11] Omlin M. and Reichert P., "A comparison of techniques for the estimation of model prediction uncertainty", *Ecol. Model.*, Vol. 115, No. 1, (1999), pp 45-59.

[12] Beven K. and Freer J., "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology", *J. Hydrol.*, Vol. 249, No. 1-4, (2001), pp 11-29.

[13] Feyen L., Vrugt J.A., Nuallain B.O., van der Knijff J. and De Roo A., "Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model". *J. Hydrol.*, Vol. 332, No. 3-4, (2007), pp 276-289.

[14] Beck M.B., "Water quality modeling: A review of the analysis of uncertainty". *Water Resour. Res.*, Vol. 23, No. 8, (1987), pp 1393-1442.

[15] Box G.E.P. and Tiao G.C., "*Bayesian Inference in Statistical Analysis*". Addison-Wesley-Longman, Harlow, UK, (1973).

[16] Thiemann M., Trosset M., Gupta H. and Sorooshian S., "Bayesian recursive parameter estimation for hydrologic models", *Water Resour. Res.*, Vol. 37, No. 10, (2001), pp 2521-2535.

[17] Martin C. and Ayesa E., "An Integrated Monte Carlo Methodology for the calibration of water quality models", *Ecol. Model.*, Vol. 221, (2010), pp 2656-2667.

[18] Vrugt J.A., Gupta H.V., Bouten W. and Sorooshian S., "A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrology model parameters", *Water Resour. Res.*, Vol. 39, No. 8, (2003), pp 1201.

[19] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E., "Equation of state calculations by fast computing machines", *J. Chem. Phys.*, Vol. 21, No. 6, (1953), pp 1087-1092.

[20] Duan Q., Gupta V.K. and Sorooshian S., "Effective and efficient global optimization for conceptual rainfall-runoff models", *Water Resour. Res.*, Vol. 28, No. 4, (1992), pp 1015-1031.

[21] Vanhooren H., Meirlaen J., Amerlink Y., Claeys F., Vangheluwe H. and Vanrolleghem P.A., "WEST: Modelling biological wastewater treatment", *J. Hydroinform.*, Vol. 5, No. 1, (2003), pp 27-50.