European Water 57: 293-297, 2017.
© 2017 *E.W. Publications*

Surrogate modelling for simplification of a complex urban drainage model

M. Mahmoodian[1,2*], J.P. Carbajal[3], V. Bellos[4,6], U. Leopold[1], G. Schutz[5] and F. Clemens[2]
[1] *Luxembourg Institute of Science and Technology, Belval, Luxembourg*
[2] *Technical University of Delft, Delft, The Netherlands*
[3] *Swiss Federal Institute of Aquatic Science and Technology, Eawag, Switzerland*
[4] *CH2M, Swindon, UK*
[5] *RTC4Water, Belval, Luxembourg*
[6] *National Technical University of Athens, Athens, Greece*
[*] *e-mail: mahmood.mahmoodian@list.lu*

Abstract: Urban drainage modelling typically requires development of highly detailed and complex models due to the nature of the underlying drainage processes. This makes activities such as model calibration, uncertainty quantification analysis and usage in real-time control (RTC) challenging and computationally expensive. The focus of this paper is to develop a surrogate model to simplify and accelerate a complex model, and make it available for RTC in future studies. Hence, only the output of the model which is relevant for RTC is considered. Surrogate models may lead to larger uncertainties in the model predictions but can significantly decrease simulation runtime. Therefore, quantification of this uncertainty is addressed here as well. We use the detailed InfoWorks ICM software as the simulator for surrogate modelling. The case study area is within the Haute-Sûre catchment in Luxembourg. First, we ran the InfoWorks ICM model to produce a dataset of inputs and outputs of the simulator. Second, a surrogate model is developed based on this data. The surrogate model is able to give an estimation of the wastewater volume in the storage tank which can be used to control the combined sewer overflow (CSO) volume. The preliminary results show that the introduced surrogate model provides a reliable method to decrease model complexity and runtime significantly. It also allows for a simple quantification of the uncertainty induced by model simplification for different simulations. However, further investigation is required to find the optimum model simplification in this regard.

Key words: Surrogate modelling, urban drainage, combined sewer overflow (CSO), uncertainty

1. INTRODUCTION

Most of the urban drainage systems which have been constructed during the 19th and 20th centuries are classified as Combined Sewer Systems (CSS) (Burian 1999), which means that the sewer network combines the runoff produced after a rainfall event with wastewater produced from households and industry. One significant disadvantage for this kind of drainage systems, is that during rain events with high precipitation, parts of the wastewater may overflow and spill into the receiving water body before passing through the wastewater treatment plant (WWTP). This event, which is known as Combined Sewer Overflow (CSO), can cause serious environmental impacts on the receiving waters and its ecosystems (see e.g. Toffol 2006).

One way of tackling this problem is to take advantage of a dynamic management approach such as model based Real-Time Control (RTC) to increase the performance of the urban drainage system. In a model-based RTC, the simulator is run several times, and frequently, to produce predictions of the outcomes of many reasonable actions, therefore a computationally expensive simulator may hinder the application of RTC. Hence, we need fast simulators, even if this entails larger uncertainties in the predictions. There are two main strategies to achieve a fast model (Asher et al. 2015): a) to develop a simple, conceptual model tailored to RTC (e.g. Mahmoodian et al. 2016); b) to simplify/reduce the already existing computationally expensive models to construct the so-called surrogate models (e.g. Carbajal et al. 2016; van Daal-Rombouts et al. 2016).

In this paper, we combine these two approaches in a hybrid way. The strategy for developing a surrogate model or emulator consists of the following steps: a) identification of the variables to be emulated; b) development of a simplified model in which every component contributing to the variables identified in step (a), is replaced by a function; and c) definition of these functions, which can be ad hoc or based on training data obtained with the detailed simulator.

2. MATERIALS AND METHODS

In this study, our main focus is the CSO locations in the urban drainage system. The components such as WWTP or the receiving water body are not considered here. We apply the strategy described previously, to construct an emulator for the dynamics of the volume in a storage tank and the overflow of a CSO weir. These two are the variables of interest as required by the step a) of the strategy. The detailed simulator used herein is a model of an urban catchment located in the area of Nocher-Route-Dahl, Luxembourg which includes these structures (tank with weir). Step b) requires the development of a simplified model. For the case of the storage tank, an intuitive model is given by the mass balance equation:

$$\frac{dV}{dt} = D(t, d_c) + R(t, \alpha, \tau) - P(t, p_c) - C(t, V_{max}, \alpha, \tau) \tag{1}$$

where V is the storage tank volume, and is driven by an inflow and an outflow. The inflow is composed of the dry weather flow (*D)* and the inflow generated by rainfall (*R)*. The outflow is composed of the outflow generated by a pump (or a controllable valve) installed in the storage tank (*P)* and the CSO volume which overflows through the weir (*C)*. Next, we give the explicit expression of each component.

The wastewater inflow depends on several properties of the catchment (e.g. population), and is characterised by a daily pattern. Since this pattern is well identified, it can be described by:

$$D(t, d_c) = d_c d(t) \tag{2}$$

where $d(t)$ is the daily pattern unit waveform of wastewater flow and d_c is a scaling constant (equal to 0. 66L/s in the specific case study).

The R component is the inflow to the tank due to rainfall. This function aims at implementing a short-cut for the transformations that the upstream network, to which the tank is connected, apply on the runoff flowing into the pipes. Two major transformations are the delay introduced by the upstream network physical properties (e.g. lengths, slopes, etc.) and the scaling of the rain-to-runoff process. Herein we learn this function from training data provided by the detailed simulator playing the role of a virtual but reference reality implemented using the InfoWorks ICM® software. We use this simulator to obtain the inflow to the tank when the rain events have a constant intensity and a predefined duration. The training data consists of 40 different constant rainfall intensities (from 2.5 to 100 mm/h with a 2.5 mm/h step) with a 4-hour duration. The reasons why we use these synthetic rainfall scenarios are: 1) The available observed rainfall data are not representing the worst case scenarios (high intensities and long durations); and 2) In this way, it is possible to analyse the relationship between rainfall characteristics and the system response in a simpler approach. It is noticed that the function is independent from duration of the rainfall. In this conditions, the inflow to the storage tank volume depends only on the rainfall intensity r and a lag τ in a linear way, and not in the duration of the rainfall. Therefore, R is defined as follows:

$$R(t, \alpha, \tau) = \alpha r(t - \tau) \tag{3}$$

where the value of α obtained from the training data is 0.294 (it is noted that r is in mm/h, t and τ is in min and R is in m^3), whereas the lag as τ=30min. The lag for the case study is defined based on cross-correlation analysis of the rainfall time series (input) and tank volume and CSO time series (output).

The P component is the pump flow, which depletes water from the tank at a constant discharge determined by the manufacturer. Therefore, P takes the value 0 (pump is off) or p_c (pump is on). In this study, p_c has a value of 6 L/s.

The C component is the CSO volume which overflows through a weir when the storage tank volume reaches the maximum capacity V_{max}. The equation of the flow over the weir should be in the following form:

$$C(t, C_D, L, A, V_{max}) = \begin{cases} \frac{C_D g^{\frac{1}{2}} L}{A^{3/2}} (V(t) - V_{max})^{\frac{3}{2}} & if V \geq V_{max} \\ 0 & otherwise \end{cases} \tag{4}$$

where C_D is the discharge coefficient of the weir, g is the gravitational acceleration, L is the length of the weir and A is the area of the tank.

However, due to the integration scheme weir overflow is significantly overestimated. Therefore, C is calculated from the training data, similar to R:

$$C(t, V_{max}, a, \tau) = \begin{cases} ar(t-\tau) & if V \geq V_{max} \\ 0 & otherwise \end{cases} \tag{5}$$

3. RESULTS

As it was mentioned earlier, the purpose of developing an emulator in this study was to be able to apply it for prediction of output with real (unseen) inputs. Hence, in this section we are going to use the emulator for prediction of storage tank and CSO volumes using a real rainfall time series recorded by a rain gauge located in the catchment during October of 2007 until December of 2009 and compare them with the corresponding results derived by the InfoWorks. Figure 1 shows the comparison between the emulator and the simulator for three indicative events, taken from the entire time series, for both of storage tank and CSO volume.

As it can be observed in Figure 1a, the emulator, at the current stage of development, is able to capture the ascending part of storage tank volume in a considerably high accuracy. This part is related to tank filling stage. More importantly, the peaks are emulated correctly as well. The latter is more important in our future model-based RTC application, because if we can predict accurately the time in which the tanks in the network would fill up during or after a potential rainfall event, we can use this information to manipulate the actuators of the network (e.g. pumps or control valves) in an optimised way to avoid or mitigate possible CSO events.

According to Figure 1b, the emulator is able to capture the large CSO events (first row) relatively better than the smaller events. Although, the time period of CSO events, is predicted accurately in all three cases.

At the current stage of development, the main problem with the emulator is that it depletes the storage tank faster than the simulator. Since the CSO component is estimated relatively well and its volume is much less than the storage tank volume, this problem must be related to the simplification of P component of the emulator which is the pumping outflow. This can be a result of the fact that the simulator does not use the pump's maximum capacity (6 L/s) during all the time steps in which the pump is on. More investigation is going to be done regarding this issue to improve the emulator.

For quantification of uncertainty induced by model simplification, the Normalised Root Mean Squared Error (NRMSE) is calculated between the results derived by the emulator and the simulator for both of storage tank and CSO volumes in respect to time. NRMSE is derived via dividing RMSE by the range of the variable (storage tank or CSO volume). This can be considered as a simple structural uncertainty quantification. The normalised RMSE, for the storage tank volume is found equal to 0.072, whereas for the CSO volume 0.002, correspondingly. This indicates that the emulator gives better results regarding CSO volume prediction and, in fact, this aspect is more interesting for us regarding future RTC system design. Although, a better comparison would be done by comparing the simulator and emulator results with real measured data, if available. In addition to the RMSE, an effort is made to define the distribution of the residuals between the emulator and the simulator. Figure 2 shows these distributions.

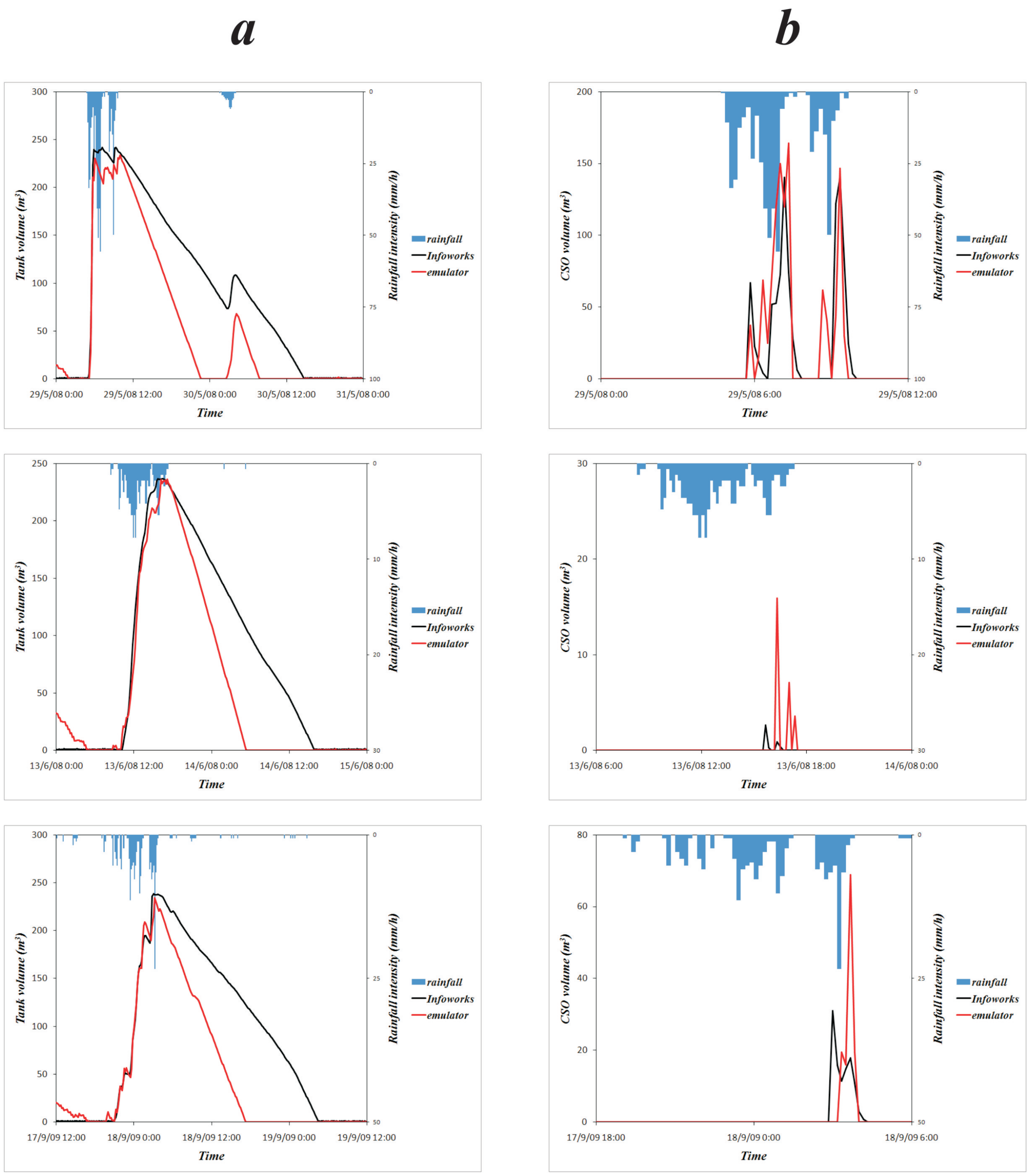

Figure 1. Comparison between emulator and simulator: a) for the storage tank volume; b) for the CSO volume.

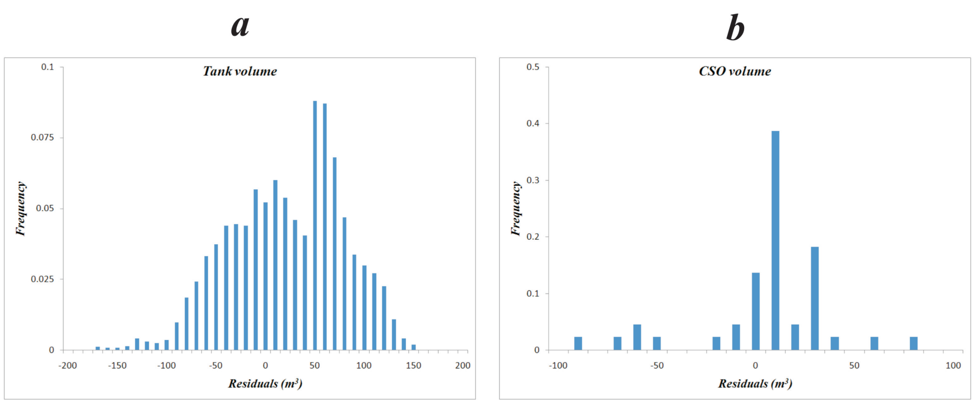

Figure 2. Distribution of the residuals between emulator and simulator: a) for the storage tank volume; b) for the CSO volume

4. CONCLUSIONS

Herein, we developed an emulator for the emulation of a storage tank volume and CSO flow in the downstream part of an urban catchment at Nocher-Route-Dahl in Luxembourg. The component functions of the emulator were learnt using simulations from a detailed model implemented in InfoWorks. The emulator provides satisfactory results in terms of speeding up the simulations. As an example, the speed up it provides on a 2-years-long time series of observed values is approximately 1300 (i.e. the emulator is 1300 times faster than the simulator). This would be an interesting aspect regarding application of the emulator for RTC purpose, in which we are required to speed up the optimisation process by employing faster models.

In contrast with some previous studies, in which the input was the inflow to the storage tank or WWTP (e.g. Mahmoodian et al. 2016; Vanrolleghem et al. 2005), the surrogate herein uses rainfall measurements (or forecasts) as inputs. This will provide an RTC system with longer reaction time to compute the optimal control actions and the consequent behaviour of the system actuators.

The simplistic emulation approach which is introduced in this study, is on its early stage of development. The approach would get more complex for more detailed case studies with several inputs and outputs of interest in various locations within the sewer network. In future steps, we will focus on producing more precise results and reducing the uncertainty induced by model simplification. Finally, we will develop the surrogate model to consider wastewater quality modelling in addition to its quantity modelling to be applied in RTC practices.

ACKNOWLEDGMENTS

This research was done as part of the Marie Curie ITN – Quantifying Uncertainty in Integrated Catchment Studies (QUICS) project. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607000. JPC received funding from the EmuMore Project URL: kakila.bitbucket.io/emumore.

REFERENCES

Asher, M. J., Croke, B. F. W., Jakeman, A. J., Peeters, L. J. M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resources Research, 51(8): 5957–5973.

Burian, S. J., Nix S. J., Durrans R. S., Pitt R .E., Fan C., Field R., 1999. Historical Development of Wet Weather Flow Management. Journal of Water Resources Planning and Management,125: 3-13.

Carbajal, J. P., Leitão, J. P., Albert, C., 2016. Appraisal of data-driven and mechanistic emulators of nonlinear hydrodynamic urban drainage simulators. Environmental Modelling and Software, 92: 17-27.

Van Daal-Rombouts, P., Sun S., Langeveld J., Bertrand-Krajewski J., Clemens F., 2016. Design and performance evaluation of a simplified dynamic model for combined sewer overflows in pumped sewer systems. Journal of Hydrology, 538: 609–624.

Innovyze, 2017. InfoWorks ICM. Available at: http://www.innovyze.com/products/infoworks_icm/.

Mahmoodian, M., Delmont, O.,Schutz, G., 2017. Pollution-based model predictive control of combined sewer networks, considering uncertainty propagation. International Journal of Sustainable Development and Planning, 12(1): 98–111.

Toffol, S., 2006. Sewer system performance assessment – an indicators based methodology. Universität Innsbruck.

Vanrolleghem, P. A., Benedetti, L., Meirlaen, J., 2005. Modelling and real-time control of the integrated urban wastewater system. Environmental Modelling and Software, 20(4): 427-442.