CrossMark

A Hybrid Surrogate Modelling Strategy for Simplification of Detailed Urban Drainage Simulators

Mahmood Mahmoodian, et al. *[full author details at the end of the article]*

Abstract

Urban drainage modelling typically requires development of highly detailed simulators due to the nature of various underlying surface and drainage processes, which makes them computationally too expensive. Application of such simulators is still challenging in activities such as real-time control (RTC), uncertainty quantification analysis or model calibration in which numerous simulations are required. The focus of this paper is to present a rather simple hybrid surrogate modelling (or emulation) strategy to simplify and accelerate a detailed urban drainage simulator (UDS). The proposed surrogate modelling strategy includes: a) identification of the variables to be emulated; b) development of a simplified conceptual model in which every component contributing to the variables identified in step (a) is replaced by a function; c) definition of these functions, either based on knowledge about the mechanisms of the simulator, or based on the data produced by the simulator; and finally, d) validation of the results produced by the surrogate model in comparison with the original detailed simulator. Herein, a detailed InfoWorks ICM simulator was selected for surrogate modelling. The case study area was a small urban drainage network in Luxembourg. An emulator was developed to map the rainfall time series, as input, to a storage tank volume and combined sewer overflow (CSO) in the case study network. The results showed that the introduced strategy provides a reliable method to simplify the simulator and reduce its run time significantly. For the specific case study, the emulator was approximately 1300 times faster than the original detailed simulator. For quantification of the emulation error, an ensemble of 500 rainfall scenarios with 1 month duration was generated by application of a multivariate autoregressive model for conditional simulation of rainfall time series. The results produced by the emulator were compared to the ones produced by the simulator. Finally, as an indicator of the emulation error, distributions of Nash-Sutcliffe efficiency (NSE) between the emulator and simulator results for prediction of storage tank volume and CSO flow time series were presented.

Keywords Surrogate model · Model simplification · Emulator · Urban drainage · Combined sewer overflow (CSO)

1 Introduction

During a Combined Sewer Overflow (CSO) event, untreated wastewater spills into natural water bodies, which may cause serious negative impacts to the receiving waters and its

ecosystems (see e.g. De Toffol 2006). Most of the urban drainage systems which were built during the 19th and 20th centuries are combined systems and cause CSOs during intense or long rainfall events (Burian et al. 1999). Prevention and limitation of pollution of receiving waters due to CSO events was one of the main objectives of the Urban Wastewater Treatment Directive (EEC Council 1991). In 2000, the same issue was highlighted in the article 16 of the European Union's Water Framework Directive (WFD) (EEC Council 2000).

One way of reducing the frequency and volume of CSO events is to manage the urban drainage system in a dynamic way using Real-Time Control (RTC) (Schilling 1989) with application of model-based RTC (e.g. Fiorelli et al. 2013; Joseph-Duran et al. 2014a, b). In a model-based RTC, the simulator is run frequently to produce predictions of the outcomes of an extensive set of reasonable actions. Therefore, computationally expensive simulators limit the application of RTC making it unavoidable to replace them with alternative fast simulators.

Fast simulators can be achieved via three different general approaches: a) parallel computing and/or supercomputers (High Performance Computing, known as HPC); b) developing a simple and fast simulator according to the specific simulation requirements (e.g. Joseph-Duran et al. 2014b); or c) building a "surrogate model" based on the already existing detailed simulators. The latter approach is the focus of this study. In the literature, surrogate models are also known as emulators (O'Hagan 2006), meta-models (Blanning 1975), reduced models (Willcox and Peraire 2002), proxy models (Bieker et al. 2007), low fidelity models (Robinson et al. 2008), response surfaces (Regis and Shoemaker 2005) and so forth. So far, two comprehensive reviews about surrogate modelling approaches have been done in the field of water resources in general (Razavi et al. 2012) and groundwater modelling in specific (Asher et al. 2015). Based on these articles, four main categories of surrogate modelling approaches can be identified in the water sciences and engineering domain:

- *Data-driven approach*, in which the detailed or complex simulator is approximated through an empirical (or statistical) model which captures the input-output mapping of the original simulator. This category covers rather a broad range of methods. Some of common methods in this regard with their example application in the field of water engineering and management are: Artificial Neural Networks (ANN) (Sreekanth and Datta 2011) and Deep Learning (DL) (Li et al. 2016); Radial Basis Functions (RBF) (Christelis and Mantoglou 2016); Kriging (Zhao et al. 2016); and Gaussian Processes Emulators (GPE) (Carbajal et al. 2016). The main advantage of the data-driven methods is their generic and non-mechanistic nature. It means, one would only need to deal with the data generated by the simulator, rather than dealing with the mathematical descriptions behind the simulator. Besides, they result in considerably lower run time, in comparison with other surrogate modelling approaches. However, these methods are normally preferred when small number of parameters, varying in limited ranges, are involved in surrogate modelling process. Apart from popularity, data-driven methods have a main disadvantage which is their subjective (researcher dependent) structure. Besides, their applicability is normally limited to the ranges of the training dataset used.
- *Projection-based approach*, in which the dimensionality of the parameter space is reduced by projecting the governing equations of the simulator onto a basis of orthonormal vectors. For application in the field of water engineering and management, Balanced Truncation (BT) (e.g. Sahlan et al. 2013) and Proper Orthogonal Decomposition (POD) (e.g. Volkwein 2013; Xu et al. 2013) are among the most popular methods in this category. The main advantages of projection-based approaches are: their computational efficiency

once constructed, as well as producing an error bound after Model Order Reduction (MOR) in most of these techniques (Willcox and Megretski 2005). The major disadvantage is that they are highly mechanistic; meaning that one should initially define a clear mathematical description (e.g. a state-space model) for the given simulator which is subject to MOR. These approaches are rather difficult to be implemented in practice, especially if the commercial modelling software does not provide access to the source code or description of the implemented algorithms (which is normally the case, except for open-source software).

- *Hierarchical or multi-fidelity approach*, where the surrogate is developed, for instance, by ignoring some of the processes which are less relevant in a given case, or by reducing the numerical resolution of the model, (e.g. Meirlaen and Vanrolleghem 2002; Leitão et al. 2010). Here, the principal advantage is that sometimes these methods are able to maintain the detail and accuracy of the original detailed or complex simulator. The dominant disadvantage is that, these methods are also highly mechanistic and difficult to implement in practice. Besides, they are case-specific and it is more challenging to generalise and automate them to be applied for other given simulators of interest.

- *Hybrid approach,* in which different combinations of any of the above-mentioned approaches can be applied to develop the surrogate model. For instance, a data-driven approach can be mixed with a projection-based or multi-fidelity approach.

The purpose of surrogate modelling in this study is to reduce the computational cost of a detailed urban drainage simulator (UDS) and make it available for future applications such as RTC. Even though importance of surrogate modelling based on already existing, well-established, detailed UDSs (Bach et al. 2014) has been emphasised repeatedly (Meirlaen et al. 2002 and Schütze et al. 2004) still developing new simple and fast simulators for specific applications, such as RTC, is common (e.g. Joseph-Duran et al. 2014b; Mahmoodian et al. 2016). Nevertheless, in urban drainage modelling domain, there are few studies in which the potential of developing surrogate models based on existing detailed simulators have been shown.

Focusing on RTC application, (Meirlaen and Vanrolleghem 2002; Langeveld et al. 2013; van Daal-Rombouts et al. 2016) preferred the multi-fidelity approach. For example, (Langeveld et al. 2013) simplified parts of a detailed integrated UDS and successfully applied the surrogate model in RTC with focus on receiving water quality control. Few other researchers found the hybrid surrogate modelling approaches more practical for acceleration of computationally expensive UDSs. With focus on urban pluvial flood simulation, Bermúdez et al. (2018) developed a hybrid surrogate model, which applies ANN as the data-driven part, for acceleration of a 1D-2D detailed UDS. For the specific case study in this research, a simulation speed up factor of more than 10^4 with a low accuracy cost was achieved. Keupers et al. (2015) developed a hybrid surrogate model for a computationally demanding integrated river-sewer simulator in order to quantify the impact of CSO events on quality of the receiving water. In this study, the highly detailed quantity and quality modelling components of the integrated simulator were substituted by surrogate models which mostly had data-driven nature. A speed up factor of 1.10^4 was achieved for the specific case study.

Application of data-driven approaches in various aspects of urban water management domain has been growing rapidly during the past decade (Eggimann et al. 2017). Due to the advantages addressed above, data-driven surrogate modelling approaches are not exempt in

this regard (Fu et al. 2010; Gradano and Le Roux 2012; Nadiri et al. 2018). However, in most of data-driven approaches the input-output mapping is performed in a black (or grey) box manner, neglecting most of the mechanisms inside the simulator and solely focusing on the input-output data.

In the current study, we argue that, if it is possible to identify some of the modelling components directly from studying the mechanisms of the case study simulator, these components can be excluded from the data-driven analysis. Hence, in this article, we propose a novel hybrid surrogate modelling strategy, which is partly based on the ad-hoc information obtained from the detailed simulator under study and partly data-driven. The focus in this study is on wastewater quantity modelling. Based on the introduced hybrid surrogate modelling strategy, we developed an emulator for storage tank volume and CSO flow time series prediction based on upcoming rainfall time series in the case study catchment.

In the following sections of this document, first, a case study detailed UDS subject to surrogate modelling and a small urban drainage network are introduced; second, the surrogate modelling strategy is explained briefly together with step-by-step application for the specific case study in hand; third, the surrogate model is validated in comparison with the original UDS and the emulation error is quantified; and finally, a conclusion is made based on the achieved results and future potential studies are highlighted. Throughout the paper, the detailed or complex UDS will be addressed simply as "simulator" and accordingly the surrogate model will be also called the "emulator".

2 Case Study

2.1 Case Study Simulator

The case study simulator, InfoWorks ICM (Innovyze 2017), is an example of highly detailed commercial software which is commonly used for modelling urban drainage systems. Around two hundred different parameters and numerous processes are involved in this simulator which might make it computationally too expensive to be applied in applications such as model-based RTC. Figure 1 shows only the main elements of InfoWorks ICM and the involved modules. InfoWorks ICM was solely selected as a detailed case study simulator in this research. The advantages or disadvantages of this specific commercial software was not the focus of the research.

For the runoff modelling in this simulator it is possible to select among 15 types of runoff volume models and 13 types of runoff routing models (Wallingford procedure fixed percentage runoff model and Wallingford model were selected for the case study of this research respectively). Each of these models require their own specific parameters and inputs. The hydraulic model is based on the De Saint-Venant equations for conservation of mass and momentum (Innovyze 2017). The rainfall, which is the main input of the runoff sub-model, can be in forms of observed (recorded) or design rainfall. It should be noted that, the focus in this research is only on wastewater quantity modelling and wastewater quality modelling is neglected. In this study, it is assumed that the simulator represents the reality through "virtual reality" and the goal is to emulate it by focusing on inputs and outputs of interest. This assumption is the common practice in surrogate modelling (Kroll et al. 2017). It is assumed that a detailed simulator is in hand which is already calibrated with the observed measurements. However, this simulator is
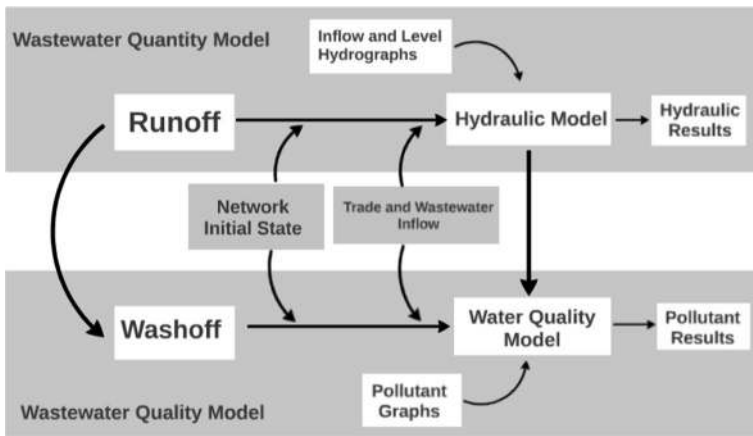
Fig. 1 Main components of the case study simulator (adapted from InfoWorks ICM documentation)

computationally expensive to be applied directly in applications such as model-based RTC or uncertainty propagation analysis. Hence, the focus is on developing a surrogate model based on this simulator to facilitate those applications.

2.2 Case Study Area

The case study area is the Nocher-Route-Dahl region, a small sub-catchment of an urban drainage network in the north of Grand Duchy of Luxembourg. The total area of this sub-catchment is equal to 54.125 ha with a contributing area (runoff surface) of 15.47 ha and a total population count of 1142. There are 220 pipes (with a total length of 10,724 m), 209 manholes and 3 CSO locations in this small case study area. Figure 2a which is drawn by InfoWorks ICM user interface, shows the modelled area. Here, the focus is on surrogate modelling for the CSO location 1 which has a retention tank together with a CSO structure. A similar procedure can be applied for other CSO locations in the catchment, since they have the same structure and similar components.

The structure of CSO location 1 in the case study is described next (see Fig. 2b). The inflow from the upstream sub-catchment flows into the main storage tank through a conduit which is connected to a rectangular weir structure for depleting the excess water in case of CSO events.
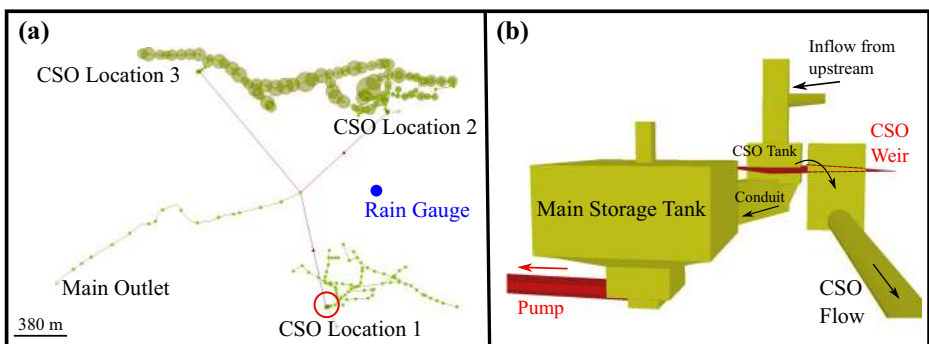

Fig. 2 **a** Nocher-Route-Dahl Case Study area; **b** Schematic view of the CSO location 1

The wastewater level in the main storage tank is controlled automatically by a fixed pump with maximum capacity of 6×10^{-3} m³/s. The pump operates based on user-defined switch on/off water levels inside the tank.

3 Method

The introduced hybrid surrogate modelling strategy in this study has four steps (see Fig. 3) and the description of this article follows the same steps in order to explain the strategy in detail. Steps A, B and C are described in the Section 3. Step D is included in Sections 4 and 5 of this article.

3.1 Identification of Variables of Interest to be Emulated

The first step to develop an emulator is to define the variables or inputs and outputs of interest based on desired application. Figure 4 presents our inputs and outputs of interest in the case study. The developed emulator should map the inputs to the outputs with an acceptable accuracy in comparison with the original simulator. The acceptable accuracy has to be defined based on the specific application of the emulator.

3.2 Development of a Simplified Conceptual Model

This step requires development of a simplified model. For case of the CSO location 1, the model can be given by the mass balance equation, as follows:

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

where V is the storage tank volume and is driven by inflow and outflow elements. 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 the pump (P) installed in the storage tank and the CSO flow which overflows through the weir (C). In the next step,
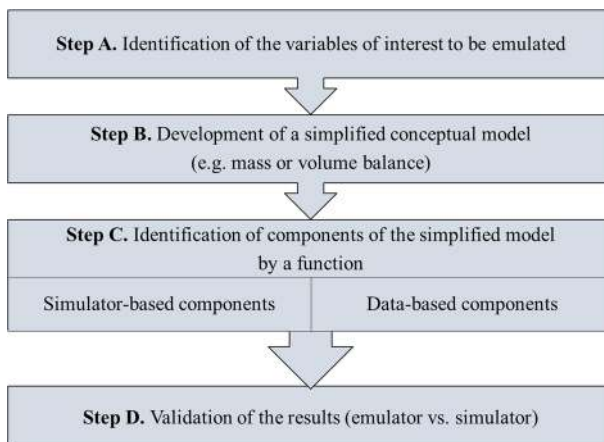

Step A. Identification of the variables of interest to be emulated

Step B. Development of a simplified conceptual model (e.g. mass or volume balance)

Step C. Identification of components of the simplified model by a function

Simulator-based components | Data-based components

Step D. Validation of the results (emulator vs. simulator)

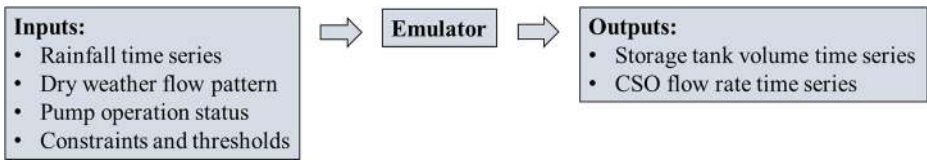**Fig. 3** Steps of the proposed hybrid surrogate modelling strategy

Fig. 4 Desired inputs and outputs to develop the case study emulator

the explicit expression of each component, including explanation of all the symbols in parentheses in Eq. (1), are introduced.

3.3 Identification of Simplified Model Components

In this step, components of the Eq. (1) should be identified either based on the knowledge from studying the mechanisms of the simulator at hand (simulator-based components) or based on the data generated by the simulator (data-based components). For the case study at hand, the flow components D, P and C of Eq. (1) are considered simulator-based components. While, R is a data-based component and it is identified (learned) based on synthetic data generated by the simulator.

3.3.1 Simulator-Based Components

The inflow generated by dry weather flow (D), which depends on demographic and hydraulic properties of the catchment, is characterized 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 of wastewater flow and d_c is a scaling constant (equal to 6.6×10^{-4} m³/s in the specific case study). This information can be extracted from running the simulator for dry weather flow situation (no rain).

Accordingly, the *P* component is the pump flow, which depletes water from the tank at an assumed constant discharge determined by the manufacturer. Therefore, *P* takes the value 0 (if the pump is off) or p_c (if the pump is on). p_c is the pump flow rate. In this study, p_c has a value of 6×10^{-3} m³/s. A similar approach can be considered for other types of system actuators such as orifices or controllable valves.

The CSO flow (C) runs over the weir if the storage tank volume reaches its maximum capacity (V_{max}). This component is given by the equation:

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

where C_D is the effective discharge coefficient of the weir, which can be obtained only by using values available from the design of the CSO structure (no learning involved). This component function can also be altered according to the CSO structure at hand (i.e. other types of weirs).

3.3.2 Data-Based Components

The inflow to the storage tank due to rainfall (R) implements a short-cut for all the transformations that the upstream network applies on the runoff flowing through the

sewer network. Two major transformations are the delay introduced by physical proper-
ties of the upstream network (e.g. lengths, slopes, etc.) and the scaling of the rainfall-
runoff process. These processes are simulated via detailed rainfall-runoff and routing
models in the original simulator and have the largest contribution to simulation compu-
tational cost.

To learn this function, the simulator was used to obtain the inflow to the storage tank
when the rainfall events have a constant intensity and a predefined duration. The training
data consists of 44 different constant rainfall intensities (from 2.6 to 100 mm/h) with a
4 h duration. This dataset was used since it was observed that: a) the R function is
independent from the rainfall event duration, i.e. tank filling behavior is always the same
for different rainfall durations; b) the inflow to the storage tank volume depends only on
the rainfall intensity r and a lag τ (Figs. 5 and 6).

Therefore, for the R component, the following model was proposed:

$$
\begin{aligned}
R(t,r) &= \alpha(r(t-\tau(r))) \\
\alpha(r) &= exp\big(a_\alpha + b_\alpha ln(r/r_{min}) + c_\alpha ln^2(r/r_{min}) + d_\alpha ln^3(r/r_{min})\big) \\
\tau(r) &= \begin{cases} a_\tau + b_\tau r + c_\tau r^{-d_\tau} & if\, r \geq r_{min} \\ 0 & otherwise \end{cases}
\end{aligned} \tag{4}
$$

where r_{min} is the minimum value of rainfall intensity in the training set. The structure of α is
given by a cubic polynomial fit on the logarithm of the training data, i.e. (rainfall intensity,
filling slopes) pairs.

The lag model could be defined as a constant delay using the traditional techniques
such as cross-correlation between the input (rainfall intensity) and output (tank volume)
signals. However, in this article, we recommend the application of time warping tech-
nique for the input rainfall intensity (Dürrenmatt et al. 2013). Time warping technique
helps accounting for the deformation of the signal in time as well as its delay. Figure 7
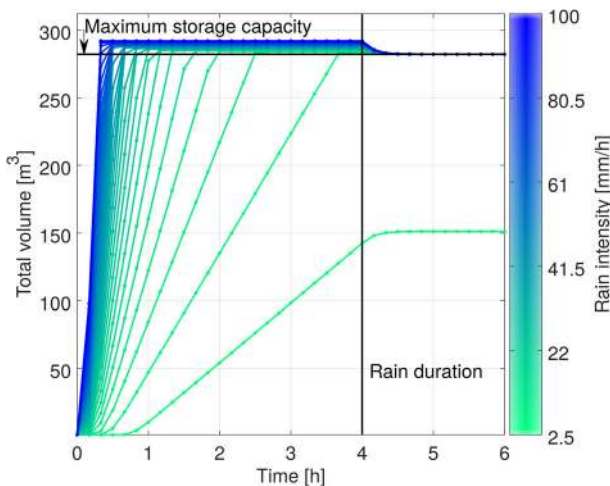shows time warping effect on a Gaussian rainfall signal, and compares it with a constant
lag (delay).

Fig. 5 Storage tank volume for various rainfall scenarios with different intensities and constant duration of 4 h
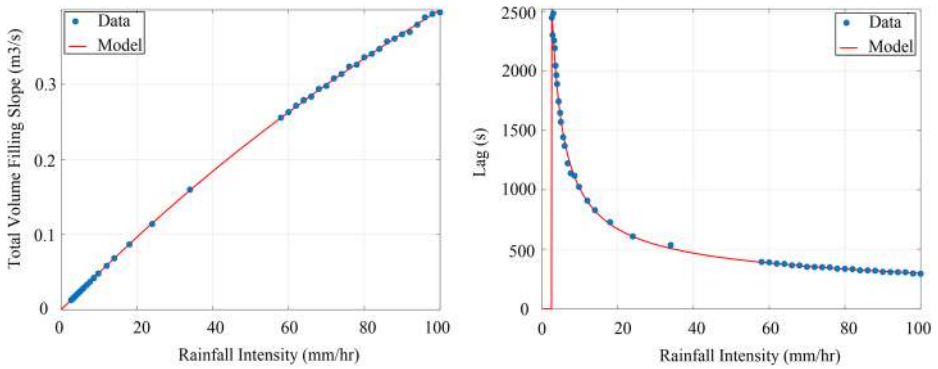(pump is off)

Fig. 6 Training data and model fitting results. Tank filling slope (left) and lag (right) as function of rainfall intensity. Circles show the training data, lines the fitted model

4 Validation

The last step of surrogate modelling strategy is to validate the results produced by the emulator in comparison with the ones generated by the original simulator. Hence, in this section, the emulator is applied to predict the storage tank volume and CSO flow rate time series using a real observed rainfall time series recorded by a rain gauge located in the catchment (Fig. 2a). The prediction results are compared to the corresponding results derived by the simulator.

Figure 8 depicts the comparison between the emulated and simulated tank volume time series for an entire year (2008). The emulator is able to capture the dynamics of storage tank volume with a considerably high Nash-Sutcliffe efficiency of 0.96 (NSE equal to one is the perfect match). The emulator is approximately 1300 times faster than the simulator in this specific case. It should be noted that, for this runtime comparison, only the hydrodynamic modelling by the simulator is taken into account (i.e. wastewater quality modelling is excluded).
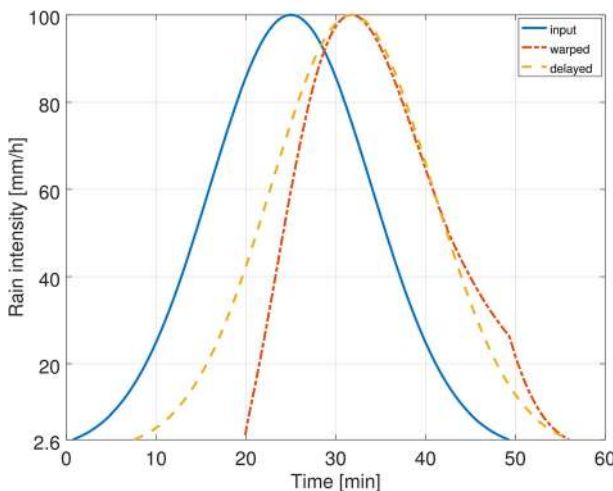

Fig. 7 Effect of the lag models on a Gaussian rainfall signal. The input signal is shown in continuous line, warped signal in dot-dashed line, and a signal with constant lag in dashed line
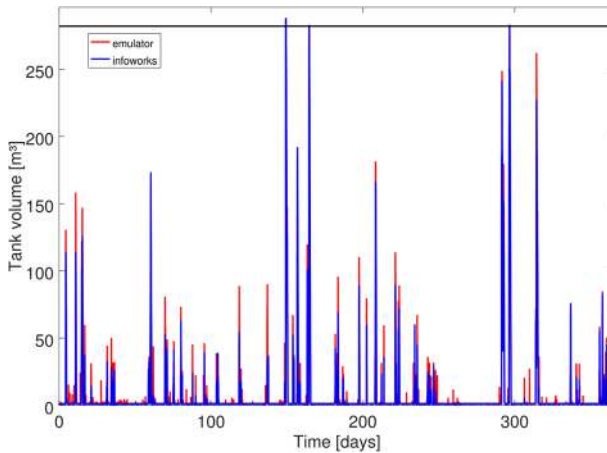
Fig. 8 Comparison between emulated and simulated tank volume time series (entire year 2008). Nash-Sutcliffe model efficiency 0.96. Root mean squared error 5.3 m³. Maximum absolute error (sign) 87 m³ (+)

The black horizontal line in Fig. 8 locates the maximum capacity of the storage tank volume (282 m³) which has been overpassed three times during the simulation period, indicating the occurrence of three CSO events. Figure 9 shows a detailed view of these events.

As it can be observed from Fig. 9, the quality of the emulator regarding CSO flow prediction is not as high as storage tank volume prediction. However, in Fig. 9 we are only focusing on three CSO events, which is not enough data to evaluate the accuracy of the emulator. Hence, in the next step, the emulator is validated using an ensemble of rainfall scenarios, which triggered more CSO events.
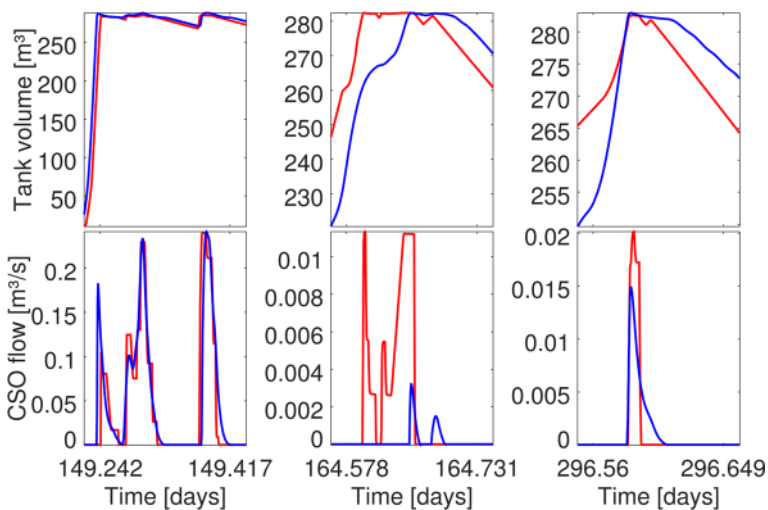

Fig. 9 Comparison between emulator (red) and simulator (blue) during CSO events: (top) storage tank volume (NSEs: 0.83, 0.33, 0.46); (bottom) CSO flow (NSEs: 0.70, −51, 0.51)

5 Emulation Error

In this section, a quantification of emulation error is performed in order to analyse the performance of the emulator with different rainfall scenarios. Based on the observed rainfall time series in the case study area, and application of a multivariate autoregressive time series model for conditional simulation of rainfall time series (Torres-Matallana et al. 2017), an ensemble of 500 rainfall scenarios of 1 year duration was generated. Since, normally, the most severe CSO events were observed during the month of August in the case study area, only the ensemble data of this month was considered for validation purpose in this section. The ensemble rainfall scenarios were used as input to the simulator, as well as the emulator. As a quantification of the emulation error, the Nash-Sutcliffe efficiency (NSE) values were calculated comparing the results produced by the simulator against the corresponding results of the emulator. The distribution of NSE for the ensemble runs is presented in the violin plots form in Fig. 10 in order to visualise the kernel probability density of the data at different values.

The results shown in Fig. 10 indicate the high accuracy of the emulator compared to the simulator. The predictions of CSO flows are not as precise as the ones for storage tank volume. The main reason for this is that, the emulator was developed only based on the storage tank filling data. In fact, the CSO flow is a side-product of the storage tank volume emulator, since it is calculated after surpassing the maximum capacity of the storage tank. This fact led to a delay of the CSO events by about 20 min forward (time resolution of simulations input and outputs was 10 min). The right panel of Fig. 10, shows the improvement on the NSE distribution obtained when the emulated CSO signals were shifted by this amount (20 min).
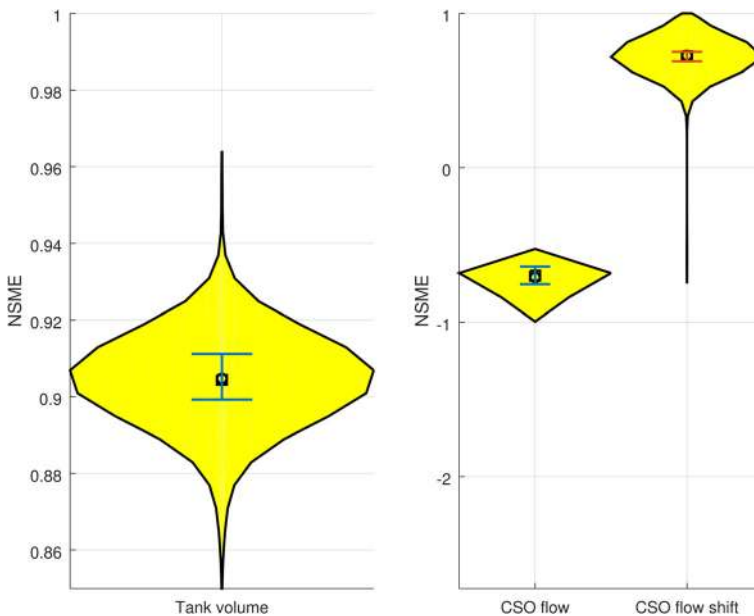

Fig. 10 Distribution of Nash-Sutcliffe efficiency (NSE) between emulator and simulator: (left) for the storage tank volume; (right) for the CSO flow without and with time shift correction

6 Discussion and Conclusions

The aim of the present research was to introduce a hybrid surrogate modelling strategy for acceleration of a computationally expensive UDS. A "hybrid" strategy was followed, since the component functions of the emulator were learned partly based on studying the mechanisms of the case study simulator at hand (simulator-based) and partly via synthetic input-output data generated by the simulator (data-based). Based on this strategy, an emulator was developed and validated for wastewater volume and CSO flow time series prediction for a small case study in Luxembourg. The novelty and added value of this research can be addressed in two main aspects. The first and the most important aspect is the simplicity of the introduced method and its hybrid nature. It means, most of the component functions of the emulator are quantified directly, and rather simply, using the knowledge obtained from studying the mechanisms of the simulator at hand. If one can quantify these components, with high certainty, directly from the simulator, there is no need to consider them as data-driven components. This is not the case in pure data-driven (black-box) surrogate modelling approaches. The second novelty of this research is regarding the lag or delay model for the R component of the emulator. In this research, time warping was applied instead of traditional cross-correlation technique. Time warping was useful to account for deformation of the emulator's output signal in time as well as its delay.

In compliance with the previous studies in application of hybrid surrogate modelling approaches which are partly data-driven (e.g. Bermúdez et al. 2018; Keupers et al. 2015), the introduced emulator in this research also provides satisfactory results in terms of speeding up the simulations with low accuracy cost (Fig. 10). It should be noted that the speed up factor depends on the case study at hand. As an example, for a 1-years-long time series simulation of observed values, the emulator herein provided a speed up factor of approximately 1300 (i.e. the emulator was 1300 times faster than the simulator). This speed up was achieved mainly because of: 1) making a shortcut for replacement of rainfall-runoff and routing models inside the original simulator, via R component of the emulator; and 2) by avoiding computation of unnecessary details (e.g. volumes and flows in all intermediate nodes and links of the network). This considerable speed up would be an outstanding aspect regarding applications such as RTC, uncertainty analysis or calibration in which numerous simulations are required.

In contrast with some previous research, in which the simulation input was the inflow to the storage tank or WWTP (e.g. Mahmoodian et al. 2016; Vanrolleghem et al. 2005), the emulator herein uses rainfall measurements (or forecasts) as inputs, and predicts the storage tank volume and CSO flow in advance. Hence, considering such an emulator in applications such as model-based RTC would provide a longer reaction time (e.g. to avoid potential upcoming CSO events).

Another advantage of the hybrid emulator introduced in this article can be highlighted in comparison with the previous works in which the rainfall event characteristics (e.g. volume, depth, duration, maximum intensity) were mapped directly to CSO events detection; either in form of binary detection of the CSO occurrence and duration (Schroeder et al. 2011; Thorndahl and Willems 2008) or in form of analog/digital detection of the CSO volume (Yu et al. 2013). Since, the introduced hybrid emulator in this article, was able to predict the storage tank volume as well as CSO flow time series and can be used for dry weather situation as well.

Finally, it should be emphasized that, the emulators or surrogate models are mainly tailored to specific cases and applications at hand. In surrogate modelling it is not intended to completely substitute a detailed simulator by an emulator. Besides, there is no universal and

Table 1 Summary of parameters values used to develop the emulator for CSO location 1

Symbol	Description	Value/range and unit
a	Effective inflow rain gain	2.94 e^{-1} m³/(s mm)
a_α, b_α, c_α, d_α	Rain gain model	4.364, 0.9880, 0.02118, −0.0090822
a_τ, b_τ, c_τ, d_τ	Lag model	228.13, −0.5687, 4904.5, −0.7947
C_D	Discharge coefficient of the weir	0.04612 $[(m^3)^{-1/2}/s]$
d_c	Dry weather flow scaling	6.6 e-4 m³/s
p_c	Pump rate	6.0 e^{-3} m³/s
r	Rainfall intensity	(2.6–100) mm/h
r_{min}	Minimum rainfall intensity [mm/h]	2.6 mm/h
V_{max}	Maximum tank volume at weir height (CSO threshold volume)	282 m³
τ	Rainfall inflow lag time	10 min

unique technique which can deal with all surrogate modelling challenges (Asher et al. 2015). Hence, in this study, we tried to introduce a simple and generic surrogate modelling strategy (see Fig. 3) which can be adapted according to the specific case studies or emulation purposes. For example, the hybrid emulator here was developed to predict the storage tank volume and CSO flow at a CSO location. Such an emulator can be useful for application in CSO management or model-based RTC. Development of the emulator would get more complex for more detailed case studies with several inputs and outputs of interest or by taking into account the spatial variability of rain within the urban drainage network. In such cases, one would require to estimate the data-based component functions (R) via other techniques such as non-linear regression, Artificial Neural Networks (ANNs) or Gaussian Process Emulators (GPEs).

The future steps of this research can be improvement of the emulator regarding aforementioned aspects as well as considering wastewater quality emulation to be applied in RTC practice in an integrated way. Another significant aspect to consider in future studies is uncertainty quantification and propagation for the emulator inputs and outputs.

6.1 Parameter Summary

To ease readability, Table 1 summarises the values of all parameters used for developing the case study emulator for CSO location 1.

Authors Contribution MM provided the case study and ran the simulations, contributed to data analysis and emulator design, and lead the writing and revision of this article. JPC and VB performed the data analysis and surrogate design, and contributed to the writing of the article. UL, GS, and FC supervised MM, and revised the drafts of the article.

Compliance with Ethical Standards

Conflict of Interest   None.

Resources   All the source code used to produce the emulation results reported herein can be downloaded from: https://bitbucket.org/KaKiLa/storagetank_emulator

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Asher MJ, Croke BFW, Jakeman AJ, Peeters LJM (2015) A review of surrogate models and their application to groundwater modeling, Water Resour Res 51:5957–5973. https://doi.org/10.1002/2015WR016967

Bach PM et al (2014) A critical review of integrated urban water modelling - urban drainage and beyond. Environ Model Softw 54:88–107. https://doi.org/10.1016/j.envsoft.2013.12.018. Elsevier Ltd

Bermúdez M et al (2018) Development and comparison of two fast surrogate models for urban pluvial flood simulations. Water Resour Manag 32(8):2801–2815. https://doi.org/10.1007/s11269-018-1959-8

Bieker HP, Slupphaug O, Johansen TA (2007) Real-time production optimization of oil and gas production systems: a technology survey. SPE Prod Oper 22(04):382–391

Blanning RV (1975) The construction and implementation of metamodels, simulation. Water Resources 24(6):177–184

Burian SJ, Nix SJ, Durrans SR, Pitt RE, Fan CY, Field R (1999) Historical development of wet-weather flow management. journal of water resources planning and management. American Society of Civil Engineers (ASCE), Reston, VA, 125(1):3–13.

Carbajal JP, Leitão JP, Albert C (2016) Appraisal of data-driven and mechanistic emulators of nonlinear hydrodynamic urban drainage simulators. Environ Model Softw 92:17–27. https://doi.org/10.1016/j.envsoft.2017.02.006

Christelis V, Mantoglou A (2016) Pumping optimization of coastal aquifers assisted by adaptive metamodelling methods and radial basis functions. Water Resour Manage 30:5845. https://doi.org/10.1007/s11269-016-1337-3

De Toffol S (2006) Sewer system performance assessment – an indicators based methodology. Universität Innsbruck

Dürrenmatt DJ, Del Giudice D, Rieckermann J (2013) Dynamic time warping improves sewer flow monitoring. Water Res 47(11):3803–3816. https://doi.org/10.1016/j.watres.2013.03.051

EEC Council (1991) Urban waste-water treatment directive. EEC Council Directive, (L), p 10. http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31991L0271

EEC Council (2000) EU water framework directive. Official Journal of the European Parliament L327(September 1996):1–82. https://doi.org/10.1039/ap9842100196

Eggimann S et al (2017) The potential of knowing more: a review of data-driven urban water management. Environ Sci Technol 51(5):2538–2553. https://doi.org/10.1021/acs.est.6b04267

Fiorelli D et al (2013) Optimised real time operation of a sewer network using a multi-goal objective function. Urban Water J 10(5):342–353. https://doi.org/10.1080/1573062x.2013.806560

Fu G, Makropoulos C, Butler D (2010) Simulation of urban wastewater systems using artificial neural networks. J Hydroinf. https://doi.org/10.2166/hydro.2009.151

Gradano JEA, Le Roux GAC (2012) Comparison of surrogate models for wastewater process synthesis. Computer Aided Chemical Engineering 30(June):1322–1326. https://doi.org/10.1016/B978-0-444-59520-1.50123-8

Innovyze (2017) InfoWorks ICM. Innovyze

Joseph-Duran B, Jung MN, Ocampo-Martinez C et al (2014a) Minimization of sewage network overflow. Water Resour Manage 28:41. https://doi.org/10.1007/s11269-013-0468-z

Joseph-Duran B, Ocampo-Martinez C, Cembrano G (2014b) Hybrid modeling and receding horizon control of sewer networks. Water Resour Res 51(October 2012):341–358. https://doi.org/10.1002/2013WR015119

Keupers I, Kroll S, Willems P (2015) Impact analysis of CSOs on the receiving river water quality using an integrated conceptual model. In: 10th international urban drainage modelling conference. Quebec, Canada, pp 205–218

Kroll S et al (2017) Semi-automated buildup and calibration of conceptual sewer models. Environ Model Softw. Elsevier Ltd 93:344–355. https://doi.org/10.1016/j.envsoft.2017.02.030

Langeveld JG et al (2013) Impact-based integrated real-time control for improvement of the Dommel River water quality. Urban Water J 10(5):312–329. https://doi.org/10.1080/1573062X.2013.820332

Leitão JP et al (2010) Real-time forecasting urban drainage models: full or simplified networks? Water Sci Technol 62(9):2106–2114. https://doi.org/10.2166/wst.2010.382

Li C, Bai Y, Zeng B (2016) Deep feature learning architectures for daily reservoir inflow forecasting. Water Resour Manage 30:5145. https://doi.org/. https://doi.org/10.1007/s11269-016-1474-8

Mahmoodian M, Delmont O, Schutz G (2016) Pollution-based model predictive control of combined sewer networks, considering uncertainty propagation. Int J Sustain Dev Plan. https://doi.org/10.2495/SDP-V0-N0-1-14

Mahmoodian M, Carbajal JP, Bellos V, Leopold U, Schutz G, Clemens F (2017) Surrogate modelling for simplification of a complex urban drainage model. European Water 57:293–297

Meirlaen J, Vanrolleghem PA (2002) Model reduction through boundary relocation to facilitate real-time control optimisation in the integrated urban wastewater system. Water Sci Technol 45(4–5):373–381 Article

Meirlaen J, Van Assel J, Vanrolleghem PA (2002) Real time control of the integrated urban wastewater system using simultaneously simulating surrogate models. Water Sci Technol 45(3):109–116

Nadiri AA et al (2018) Prediction of effluent quality parameters of a wastewater treatment plant using a supervised committee fuzzy logic model. J Clean Prod 180:539–549. https://doi.org/10.1016/j.jclepro.2018.01.139. Elsevier B.V.

O'Hagan A (2006) Bayesian analysis of computer code outputs: a tutorial. Reliab Eng Syst Saf 91(10):1290–1300

Razavi S, Tolson BA, Burn DH (2012) Review of surrogate modeling in water resources. Water Resour Res 48(7):W07401. https://doi.org/10.1029/2011WR011527

Regis RG, Shoemaker CA (2005) Constrained global optimization of expensive black box functions using radial basis functions. J Glob Optim 31(1):153–171

Robinson T et al (2008) Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. AIAA J 46(11):2814–2822

Sahlan S, Wahab NA, Darus IZM (2013) Results on frequency weighted model reduction techniques of activated sludge process. Proceedings - UKSim 15th International Conference on Computer Modelling and Simulation, UKSim 2013, pp 172–176. https://doi.org/10.1109/UKSim.2013.137

Schilling W (1989) Real time control of urban drainage systems-the state of the art. IAWPRC Task Group on Real-Time Control of Urban Drainage Systems, London

Schroeder K et al (2011) Evaluation of effectiveness of combined sewer overflow control measures by operational data. Water Sci Technol 63(2):325–330. https://doi.org/10.2166/wst.2011.058

Schütze M et al (2004) Real time control of urban wastewater systems—where do we stand today? J Hydrol 299(3–4):335–348. https://doi.org/10.1016/j.jhydrol.2004.08.010

Sreekanth J, Datta B (2011) Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management. Water Resour Manag 25:3201–3218. https://doi.org/10.1007/s11269-011-9852-8

Thorndahl S, Willems P (2008) Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first-order reliability method and parameterization of local rain series. Water Res 42(1–2):455–466. https://doi.org/10.1016/j.watres.2007.07.038

Torres-Matallana JA, Leopold U, Heuvelink GBM (2017) Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models. European Water 57:299–306

Van Daal-Rombouts P et al (2016) Design and performance evaluation of a simplified dynamic model for combined sewer overflows in pumped sewer systems. J Hydrol. Elsevier B.V. 538:609–624. https://doi.org/10.1016/j.jhydrol.2016.04.056

Vanrolleghem PA, Benedetti L, Meirlaen J (2005) Modelling and real-time control of the integrated urban wastewater system. Environ Model Softw 20:427–442. https://doi.org/10.1016/j.envsoft.2004.02.004

Volkwein S (2013) Proper orthogonal decomposition: theory and reduced-order modelling. University of Konstanz Department of Mathematics and Statistics

Willcox K, Megretski A (2005) Fourier series for accurate, stable, reduced-order models in large-scale linear applications. SIAM J Sci Comput 26(3):944–962

Willcox KE, Peraire J (2002) Balanced model reduction via the proper orthogonal decomposition. AIAA J 40(11):2323–2330. https://doi.org/10.2514/2.1570

Xu M, van Overloop PJ, van de Giesen NC (2013) Model reduction in model predictive control of combined water quantity and quality in open channels. Environ Model Softw 42:72–87. https://doi.org/10.1016/j.envsoft.2012.12.008

Yu Y et al (2013) Cluster analysis for characterization of rainfalls and CSO behaviours in an urban drainage area of Tokyo. Water Sci Technol 68(3):544–551. https://doi.org/10.2166/wst.2013.253

Zhao Y, Lu W, Xiao C (2016) A kriging surrogate model coupled in simulation-optimization approach for identifying release history of groundwater sources. J Contam Hydrol. Elsevier B.V. 185–186:51–60. https://doi.org/10.1016/j.jconhyd.2016.01.004

Affiliations

Mahmood Mahmoodian[1,2] · Juan Pablo Carbajal[3] · Vasilis Bellos[4,5] · Ulrich Leopold[1] · Georges Schutz[6] · Francois Clemens[2,7]

✉ Mahmood Mahmoodian
mahmood.mahmoodian@list.lu

[1] Luxembourg Institute of Science and Technology, Belvaux, Luxembourg

[2] Delft University of Technology, Delft, The Netherlands

[3] Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland

[4] CH2M, Swindon, UK

[5] National Technical University of Athens, Athens, Greece

[6] RTC4Water, Belval, Luxembourg

[7] Deltares, Delft, The Netherlands