Emulation of a Detailed Urban Drainage Simulator to Be Applied for Short-Term Predictions

Mahmoodian, Mahmood; Torres-Matallana, J. A.; Leopold, Ulrich; Schutz, Georges; Clemens, Francois

DOI
10.1007/978-3-319-99867-1_102

Publication date
2019

Document Version
Final published version

Published in
New Trends in Urban Drainage Modelling - UDM 2018

Emulation of a Detailed Urban Drainage Simulator to Be Applied for Short-Term Predictions

Mahmood Mahmoodian[1,2(✉)], J. A. Torres-Matallana[1,3], Ulrich Leopold[1], Georges Schutz[4], and Francois Clemens[2,5]

[1] ERIN Department, Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg
mahmood.mahmoodian@list.lu
[2] Sanitary Engineering Department, Technical University of Delft, Delft, The Netherlands
[3] Soil Geography and Landscape Department, Wageningen University, Wageningen, The Netherlands
[4] RTC4Water, Belval, Luxembourg
[5] Hydraulic Engineering Department, Deltares, Delft, The Netherlands

Abstract. The challenge of this study is to investigate on applicability of a data-driven Gaussian Process Emulator (GPE) technique to develop a surrogate model for a computationally expensive and detailed urban drainage simulator. The novelty is the consideration of (short) time series for the simulation inputs and outputs. Such simulation setup is interesting in applications such as Model Predictive Control (MPC) in which numerous, fast and frequent simulation results are required. Here, an emulator is developed to predict a storage tank's volume in a small case study in Luxembourg. Three main inputs are considered as the GPE's parameters: initial volume in the tank, the level in which the outlet pump of the tank must start to work, and the time series of expected rainfall in the upcoming 2 h. The output of interest is the total volume of the storage tank for the next 24 h. A dataset of 2000 input-output scenarios were produced using different possible combinations of the inputs and running the detailed simulator (InfoWorks® ICM). 80% of the dataset were applied to train the emulator and 20% to validate the results. Distributions of Nash-Sutcliffe efficiency and Volumetric Efficiency are presented as indicators for quantification of the emulation error. Based on the preliminary results, it can be concluded that the introduced technique is able to reduce the simulations runtime significantly while imposing some inevitable accuracy cost. More investigation is required to validate the more generic applicability of this technique for multiple outputs and interactions between different urban drainage components.

Keywords: Surrogate model · Gaussian process · Emulator · Urban drainage InfoWorks® ICM

G. Mannina (Ed.): UDM 2018, GREEN, pp. 592–596, 2019.
https://doi.org/10.1007/978-3-319-99867-1_102

1 Introduction

Acceleration of the model simulations while quantifying the uncertainty in the estimations, is still one of the challenges in the field of urban drainage modelling. Application of Gaussian Process Emulators (GPEs), as one of the promising surrogate modelling techniques, is gradually increasing to face this challenge. Among the most relevant and recent studies in this regard is a mechanism-based emulator, based on simulator's equations, which was introduced by Machac et al. (2016) for emulation of SWMM simulator. Later on, Carbajal et al. (2016) compared this mechanism-based GPE with a purely data-driven GPE. Based on the comparison results, it was asserted that data-driven GPE outperforms the mechanistic one in many applications. However, to the authors' knowledge, no data-driven GPE technique has been applied in the urban drainage modelling domain which can deal with inputs and outputs in time series format. The main reason for this is the high dimensionality problem in such case as a result of inclusion of numerous parameters. Although, in applications such as model-based real-time control (RTC), the number of simulation inputs and outputs are limited. Hence, the main motivation in our study is to investigate on a data-driven emulation technique which has the potential to be applicable for simulators with inputs and outputs in time series format and use it for short-term predictions specifically for RTC purpose. The current document briefly presents the methodology applied in this regard together with illustration of some achieved results.

2 Materials and Methods

2.1 Emulation Method

The GPE method used in this study is based on (Olson et al. 2015). In this approach the model output of interest (Y) can be formulated as a Gaussian process such that:

$$Y \sim N\left(\mu_{\beta}, \sum(\xi_y)\right) \tag{1}$$

where μ_β is a mean function which is considered linear in time, and ξ_y is a vector of covariance matrix parameters. The GPE inputs should be prepared in terms of various simulation parameter sets and the corresponding outputs can be in times series format. An R package named 'stilt' is used to develop the emulator in this study (Olson et al. 2015). Detailed information on the underlying mathematical framework of the method can be found in (Olson and Chang 2013).

2.2 Candidate Simulator, Case Study and Data

The candidate simulator subject to emulation in this study is InfoWorks® ICM which normally requires a detailed description of the structure and geometry of urban drainage networks together with numerous parameters and inputs. A small case study area from Haute Sûre catchment in Luxembourg, modelled by InfoWorks® ICM 8.5, is selected as the case study. The focus is to develop an emulator for short-term predictions of total storage volume in a combined sewer overflow (CSO) location in this case study. The introduced GPE method is able to map a set of parameters (inputs) to the desired outputs in times series format. It is observed that, in this case study, three inputs are playing the main role in changing the dynamics of the outputs of interest. These are including: the initial water volume in the storage tank (P1), the switch-on level for the fixed flow pump which controls the outflow of the tank (P2), and the expected rainfall event in the catchment. Since only short-term predictions are interested in this study, only short-term rainfall events are considered to test the applicability of the introduced GPE method. Hence, as an example, rainfall events with 2 h duration and time steps of 10 min are taken into account. The rainfall intensity (mm/h) in each time step is assumed as an individual parameter (i.e. 12 parameters: P3 to P14). The reason for consideration of rainfall time series as individual parameters is that the GPE technique applied in our research is only able to map individual discrete parameters (as input) to time series of the desired output. To our knowledge, taking into account continues time-series as input for GPE is very challenging due to high dimensionality problem in such case.

An ensemble of 100 rainfall time series were generated based on the observed rainfall time series in the case study area, consideration of different return periods and application of a multivariate autoregressive model and conditional simulation of rainfall time series (Torres-Matallana et al. 2017). Afterwards, 2000 parameter sets are configured by various possible combinations of 4 samples for P1, 5 samples for P2, and 100 samples for P3 to P14 (i.e. the 100 rainfall scenarios). Then, the parameter sets are used as input to run the InfoWorks® ICM simulator in order to build a dataset of 2000 input-output pairs to train and validate the emulator (80% for training, 20% for validation). Conditional Latin Hypercube Sampling (CLHS) technique, is implemented to sample the training and validation data from the ensemble dataset using the R package "clhs" (Roudier 2017).

3 Validation Results and Discussion

Figure 1 illustrates comparison of the results produced by the emulator and the original simulator (InfoWorks® ICM) regarding short-term predictions. Three random runs with different inputs are represented here as illustrative examples. Nash-Sutcliffe Efficiency (NSE) and Volumetric Efficiency (VE) are used as statistics for quantification of the emulation error. NSE or VE equal to 1 means a perfect match between the emulation and simulation time series. The results produced by the emulator and simulator are compared with each other using the validation dataset. Violin plots of NSE and VE are presented in Fig. 2 to visualise the kernel probability density of the data at different values.

As it can be observed from Fig. 2, the introduced GPE technique is capable of capturing the desired outputs in time series format with relatively high accuracy in comparison with the original detailed simulator. NSE and VE both are distributed towards 1 (Q_1, Q_2 and Q_3 quantiles of both distributions are located between 0.8 and 1 in all cases). However, there are few cases in which the deviation between emulator and simulator is significantly beyond the 95% confidence interval. Moreover, it was

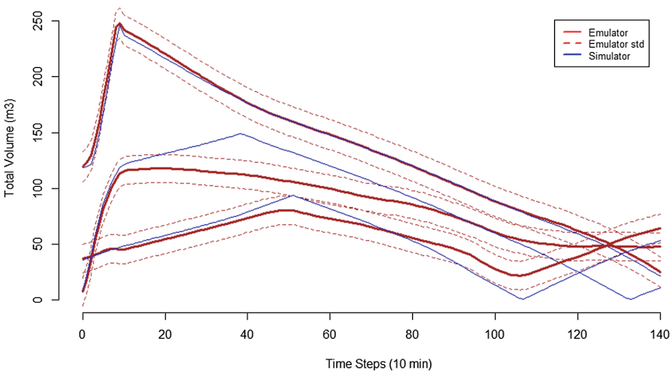

Fig. 1. Comparison of emulator vs. simulator results for three random sample scenarios from validation dataset.

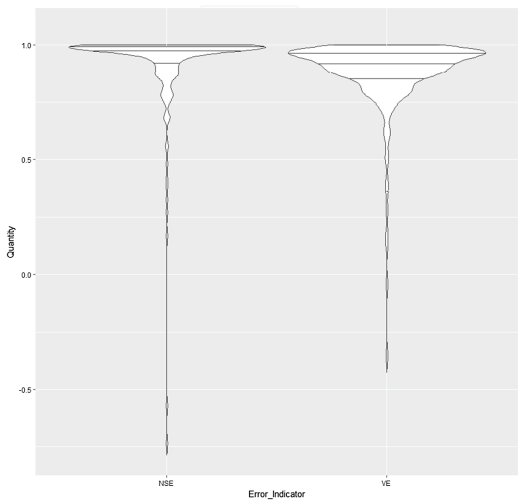

Fig. 2. Distribution of NSE (left) and VE (right) for validation dataset (horizontal lines indicate the 0.25, 0.5 and 0.75 quantiles).

observed that the emulator is approximately 300 times faster than the simulator in this small case study on the same computer. It should be noted that, for runtime comparison solely the hydrodynamic simulation's runtime by InfoWorks® is considered (excluding wastewater quality modelling). This runtime acceleration is mainly obtained by reducing the complexity and the numerical approach behind the detailed simulator and fitting a model solely based on the data achieved from snapshots of the simulator runs.

It should be mentioned that, the emulator fails to predict when the input parameters are beyond the training ranges. For instance, if the emulator is trained with rainfall intensities between 0 and 60 mm/h, it cannot be used to predict the output for an intensity of 70 mm/h. Besides, the emulator prediction results are worse (wider confidence interval) when there has not been sufficient training data in that range of inputs and outputs (data insufficiency). Hence, to achieve better emulation results, it is advised to implement a proper sampling design for the parameters to take into account a smoother coverage in the parameter space.

4 Conclusions

The main advantages of the introduced emulation technique are including: (1) its generic non-intrusive nature due to being solely based on data extracted from the simulator; (2) its applicability for inputs and outputs in (short) time series format; and (3) its capability for considerable simulation acceleration together with a low accuracy cost (if trained properly). However, this method can only be considered a proper surrogate modelling technique when a limited number of parameters are involved in changing the dynamics of the desired outputs. The future steps of this research will be investigation on the larger case study with various CSO locations and their interconnections; uncertainty quantification for rainfall input and its effect on emulation results; consideration of wastewater quality modelling; and finally application of such emulators in RTC of urban drainage systems.

Acknowledgements. 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.

References

Carbajal, J.P., Leitão, J.P., Albert, C.:. Appraisal of data-driven and mechanistic emulators of nonlinear hydrodynamic urban drainage simulators arXiv: 1609.08395v1 [stat.ME] 25 Sep 2016, 1–22 (2016)

Machac, D., Reichert, P., Albert, C.: Emulation of dynamic simulators with application to hydrology. J. Comput. Phys. **313**, 352–366 (2016). https://doi.org/10.1016/j.jcp.2016.02.046

Olson, R., Chang, W.: Mathematical framework for a separable Gaussian Process emulator (2013). www.geosc.psu.edu/~rtonkono

Olson, R., Chang, W., Keller, K., Haran, M.: R package " stilt." CRAN (2015)

Roudier, P.: R Package "clhs" - Conditioned Latin Hypercube Sampling. CRAN (2017). https://doi.org/10.1201/b12728-46

Torres-Matallana, J.A., Leopold, U., Heuvelink, G.B.M.: Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models. Eur. Water **57**, 299–306 (2017)