Performance-evaluation of urban drainage models

Pedersen, A. N.; Borup, M.; Brink-Kjær, A.; Mikkelsen, P. S.

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Pedersen, A. N., Borup, M., Brink-Kjær, A., & Mikkelsen, P. S. (2021). *Performance-evaluation of urban drainage models*. Paper presented at 15th International Conference on Urban Drainage.

Performance-evaluation of urban drainage models

A. N. Pedersen[1,2*], M. Borup[3], A. Brink-Kjær[1] & P. S. Mikkelsen[2]

[1]VCS Denmark, Vandvaerksvej 7, 5000 Odense C, Denmark
[2]DTU Environment, Technical University of Denmark, Bygningstorvet, Bygning 115, 2800 Kgs. Lyngby, Denmark
[3]Kruger, Gladsaxevej 363, 2860 Søborg, Denmark
**Corresponding author email: anp@vandcenter.dk*

Highlights

- Evaluation of signatures from model and sensors provide insight into the model performance.
- Combination of signatures can indicate the model performance for a specific objective.
- Some signatures are harder to simulate than others and need to be further investigated.

Introduction

By introducing digital twins (Pedersen et al., 2021) in the work with the urban drainage system there is an expectancy that the simulation models included in the digital twin perform sufficiently all the time. But the water industry currently has no good and generally agreed way to validate and quantify this. The simulation models have been used in the community for decades and results from hydrodynamic urban drainage models form the basis of many decisions in redesign and upgrading of the current urban drainage systems. They are expected to be able to simulate the current situation when they are well calibrated. However, what is often happening in the utilities is only heuristic calibration of parameters in the model in cases where sufficient new observations are available. With the introduction of digital twins, the simulation models need to perform well all the time with many objectives in mind, or at least give an insight in the model performance for the current situation in order to make good, informed decisions. We therefore need to understand the model performance across many operational conditions and across multiple locations by learning from the locations where monitoring occurs. We need to get an overview of under which criteria the model performs acceptably and which not. By extracting so called signatures (Gupta et al., 2008) that summarize certain characteristics of a timeseries, we are able to identify and diagnose the model performance for specific objectives, instead of applying average distance measures such as RMSE on entire time series.

For this paper we focus on the objective of quantifying combined sewer overflow and use signatures relevant to this objective when evaluating the model performance in a simple way.

Methodology

Observations and model results from six internal and external overflow structures from a suburban area in Bellinge, Odense, Denmark are analysed. The model applied is a Mike Urban model with MOUSE engine and with rain input from rain gauges in the area. The data and model is open and available for use by anyone interested (Pedersen, in review).

Signatures are calculated for rain-induced events (Pedersen et al., in process), and focus is here on three signatures that can be derived by water level sensors: peak level, duration of overflow, and AUC (area under curve) from above crest level – which can be considered a surrogate of the overflow volume. We chose these specific signatures since they all describe independent important characteristics when assessing overflows.

Events with high rain input uncertainty, as calculated based on the spatial statistics between nearby rain gauges (Pedersen et al., in process), are not considered in the model evaluation as the model is not expected to be able to replicate these events.

Signatures for modelled and observed values, respectively, are plotted against each other for multiple events, and an ordinary least squares linear regression with intercept 0 is applied on the resulting scatter plot.

$$\hat{y}_i = \beta_i x \qquad \text{(Eq. 1)}$$

, where $\hat{y}_i$ are the modelled and *x* the measured values. The slope parameter, β_i, is evaluated, and the 95% confidence intervals are calculated based on t-tests. If the confidence interval covers a slope of 1 the model performs very well. A general performance analysis for the objective "overflow" is made for all signatures *i* of relevance with an average of the absolute slope difference.

$$Performance = \frac{\sum_{i=0}^{n} |1-\beta_i|}{n} \qquad \text{(Eq. 2)}$$

, where *n* is the number of signatures of relevance.

Results and discussion

The modelled and measured results for the three defined signatures are plotted directly against each other event by event in Figure 1. For the peak level we are only interested in the events where overflow was observed (the blue area on Figure 1, left), even though the model simulates many false positive overflows (indicated with FP in Figure 1) as well.

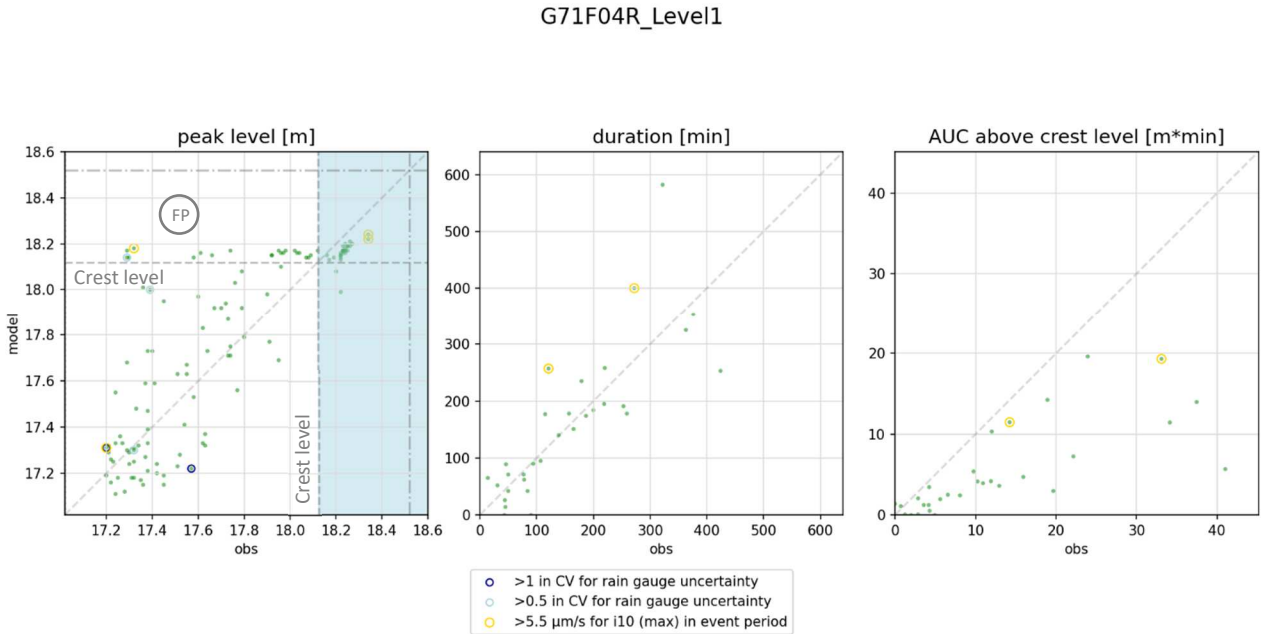

Figure 1. Signature graphs with relevance to overflow for the location G71F04R Level1. Peak level (left), where the blue area indicates that an overflow is observed (level > 18.12). Duration of the overflow (middle) and AUC (area under curve) from the crest level (right). The spatial variability of rainfall is indicated with yellow and blue circles (CV = Coefficient of Variation).

The results from the statistics are shown in Table 1. As the optimal performance has a slope of 1, the results where the 95% confidence band covers 1 is highlighted with green. The general performance is calculated by Eq. 2 in the table and shows that G71F04R and G80F66Y performs the best, whereas G71F06R, G71F68Y and G80F11B is not performing that well. Less considerations should be given to the statistics for G80F11B due to the few events recorded for this sensor. What can be seen for G71F04R and G80F66Y is that the peak level and the duration of overflow generally performs well, but the AUC calculated from the crest level is much higher for the observations. This could indicate that the weir formulae in the model allow too much water to flow over the crest for a given water level. However, this statement needs to be investigated further.

Table 1. Results statistics from the 6 overflow structures in the Bellinge case area. Length of the period with data, number of observed events, number replicable of events where the rain input uncertainty is considered acceptable for a realistic comparison, regression slope (eq. 1) (mean, and 95% confidence level in brackets) for each of the three signatures (peak level, duration, AUC), and general model performance (eq. 2). Green colour: estimated regression slope within confidence band *Only two years considered (2018-08-01 – 2020-10-01) due to changes in throttle pipes between G71F04R (G71F090) and G71F06R in August 2018.

				Regression slope β_i (mean, 95% confidence band)			
	Period (years)	No. Observed events	No. Replicable events	Peak level [m]	Duration overflow [min]	AUC above crest level [m*min]	General performance
G71F04R_Level1	2*	31	29	0.94 [0.93:0.96]	0.95 [0.81:1.09]	0.36 [0.27:0.44]	0.75
G71F05R_LevelInlet	10	740	488	0.91 [0.90:0.92]	0.34 [0.32:0.36]	0.31 [0.28:0.33]	0.52
G71F06R_LevelInlet	2*	125	87	1.08 [1.07:1.10]	1.95 [1.72:2.17]	2.24 [1.93:2.55]	0.24
G71F68Y_LevelPS	10	60	27	0.89 [0.82:0.96]	0.18 [0.10:0.27]	0.03 [-0.02:0.09]	0.37
G80F11B_Level1	1	4	3	0.79 [0.61:0.98]	0.19 [0.14:0.23]	0.01 [0.01:0.02]	0.33
G80F66Y_Level1	1	15	13	1.06 [0.95:1.18]	1.03 [0.68:1.37]	0.51 [0.18:0.83]	0.81

As an analysis of overflow usually is based on some extreme precipitation event, these are of course interesting to simulate. But since these events typically have high spatial variability, having rain input from rain gauges only result in highly uncertain input. Applying confidence bands as an indication of where the model perform well, can have some limitations if the modelled events have large standard errors. The analysis may be improved by incorporating the coefficient of determination to indicate this.

Conclusions and future work

With the provided method we are able to indicate where the model fits well in general, and in detail for the specific signatures. The observed events can be replicated to different degrees for the different locations. The rain input from rain gauges can be very uncertain and the events with high uncertainty are disregarded in the analysis. By combining signatures, we are able to get a general model performance for the objective in interest. From the results it can generally be seen that the signature AUC does not perform as well as the other signatures, and this underline the need for more specific diagnostic tools to evaluate model performance for different objectives.

Future work will aim to improve the evaluation with both different uncertainty classes, as well as ways to differentiate the weight of the different events or signatures. One analysis can seldom stand alone, and therefore we need to provide different analysis of the signatures to provide strength of the method. Different objectives will be addressed with relevant signatures in the coming months.

The data and from these six overflows are openly available to everyone (Pedersen et al, in review), and we hope that anyone with interest will help investigating if the models can provide the necessary information.

References

Gupta H. V., Wagener T. & Liu Y. (2008) Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes, 22(18), 3802–3813. doi:10.1002/hyp.6989

Pedersen A. N., Borup M., Brink-Kjær A., Christiansen L. E. & Mikkelsen P. S. (2021) Living and Prototyping Digital Twins for Urban Water Systems: Towards Multi-Purpose Value-Creation Using Models and Sensors. Water, 13. doi:10.3390/ w13050592

Pedersen A. N., Pedersen J. W., Borup M., Brink-Kjær A., Christiansen L. E. & Mikkelsen P. S. (in process) Using multi-event hydrologic and hydraulic signatures from sensors to diagnose uncertainty in integrated urban drainage models.

Pedersen A. N., Pedersen J. W., Vigueras-Rodriguez A., Brink-Kjær A., Borup M. & Mikkelsen P. S. (in review) The Bellinge data set: Open data and models for community-wide urban drainage systems research. Earth Syst. Sci. Data Discuss. doi:10.5194/essd-2021-8