# A Toolchain for the Data-driven Decision Support in Waste Water Networks - A Level-based Approach

Balla, Krisztian Mark; Lemée, Henrik; Schou, Christian; Kallesøe, Carsten

IWA World Water Congress & Exhibition
11-15 SEPTEMBER 2022 | COPENHAGEN, DENMARK | www.worldwatercongress.org

# A Toolchain for the Data-driven Decision Support in Waste Water Networks – A Level-based Approach

**Krisztian M. Balla*, Henrik Lemée**, Christian Schou***, and Carsten S. Kallesøe******

* Grundfos, Poul Due Jensens vej 7, 8850 Bjerringbro, Denmark and Aalborg University, Fredrik Bajers vej 7, 9220 Aalborg, Denmark (kballa@grundfos.com, kmb@es.aau.dk)
** Ishøj Forsyning, Baldersbækvej 6, 2635, Ishøj (hemle@ishoj.dk)
*** Grundfos, Poul Due Jensens vej 7, 8850 Bjerringbro, Denmark, (cschou@grundfos.com)
**** Grundfos, Poul Due Jensens vej 7, 8850 Bjerringbro, Denmark and Aalborg University, Fredrik Bajers vej 7, 9220 Aalborg, Denmark (ckallesøe@grundfos.com, csk@es.aau.dk)

**Abstract:** This paper aims to enable automated decision-making in combined wastewater and stormwater networks. The proposed concept is based on the deployment of in-sewer water level sensors distributed at critical locations in basins and manholes. With the use of level sensors and weather forecast feeds, we aim to learn how rain infiltrates into the network and build decision support to optimally manage the operation of storage elements, i.e., basins. For that, a data-driven probabilistic framework based on Gaussian Processes is developed. The presented framework enables practitioners to build system knowledge (actuator state, tank dimensions) into the design while leaving the uncertain parts (rain runoff) to be handled by the Gaussian Processes. The paper highlights the practical feasibility of the toolchain through a pilot project with Ishøj Spildevand in Denmark, where the prediction capabilities are tested for five months period, providing a proof of the proposed concept.

**Keywords**: Decision support; Smart water systems; Data-driven modelling; Climate adaptation

Up to today, most sewer systems operate without any form of global supervision or optimization. Utilities, however, are constantly challenged by the increased amount of wastewater and the more frequent high-intensity precipitation due to growing urbanization and climate change. Nonetheless, thanks to the ongoing digital developments in the water sector, wastewater operators have adopted advanced data acquisition and data processing for system monitoring, raising the question of how to build decision-making tools in the wake of the digital transformation in the urban water sector (Eggimann et al. 2017).

## Motivation & background

One way to handle intensive load on sewer systems (without infrastructure expansion) is to use system-wide optimization based on real-time data to avoid or, at least, attenuate water surges. From the control perspective, proactive methods, such as predictive control has high relevance in preparing the sewers for high-intensity rain events. However, reactive techniques (based on simple feedback rules) are the most widely implemented methods in practice (Eggimann et al. 2017).

An issue with predictive control is often the need for a well-calibrated high-fidelity or physical network model. Such models are available at some mid- or large-size water utilities, but often economically out of reach for smaller operators. For that reason, neither decision support nor control tools are used by practitioners, which clearly shows that plug & play solutions have a high impact in practice.

## Methods

The proposed model behind the toolchain consists of a nominal part (a hydraulic description of basins and the flow provided by controllable assets, e.g., pumps) and an unknown part (rain, wastewater, forecasts uncertainty, and all hidden dynamics, described by Gaussian processes). Opposed to other traditional toolchain approaches that handle the wet and dry-weather flows as separate forecast blocks, we consider the translation of rain intensity to level variation incorporated in the control design. Hence, the data-driven part of the model is only fitted to the difference of the measured water levels and the known nominal dynamics, i.e., to the residuals. For this, the available level measurements, information about the controllable assets, the high-level topology of the network, and the rain forecast feeds are used. Using a nonparametric and stochastic modelling tool, such as Gaussian Processes, enables us to use only the measured and forecasted data to conclude our control decision. A simplified diagram of the proposed toolchain is shown in Figure 1.

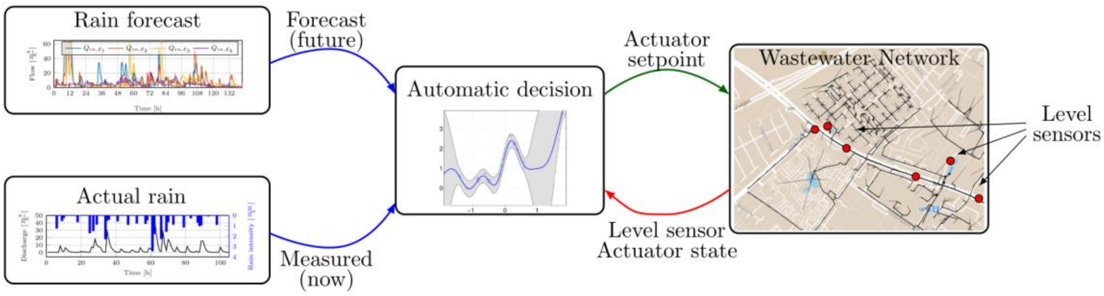


**Figure 1** Closed-loop automatic decision support strategy using water level sensors and feedback from actuators.

The decision (support) algorithm evaluates the forecasts, the actual rain, and the actual level measurements. Based on a similarity measure between the observations, the algorithm provides predictions of the water levels in the network, which can be used for decision on the network operation. Simply stated: the similarity between rain and wastewater patterns between historic data and current forecasts are evaluated, meaning that the current forecast will likely result in an observation similar to those historic rain events having the same length and intensity. Hence, we do not only conclude on mean value predictions but also provide our decision confidence to the utility. If similar events happened before then our confidence of the prediction is high, while if a new event is observed that is not learned yet, the confidence is low.

**Case study**

The experimental evaluation of the toolchain has been carried out in a pilot project under the collaboration of Grundfos Holding A/S, Aalborg University, and Ishøj Spildevand in Denmark. Five level sensors have been deployed in the network for a period spanning five months, providing 30 [sec] measurement resolution, while the rain data has been obtained from the Danish Meteorological Institute's (DMI) service at a 1 [min] resolution. The area and the sensor placements are shown in Figure 2.

The network is a stormwater system transporting the water from the city of Ishøj to the sea. There is a main transport line, along which there are stormwater basins with high volume capacity. The main focus of the utility is to carry out climate adaptation on their system, partly due to the following operational management issues:

*Problem 1.* Water volumes accumulate downstream, increasing the risk of flooding in case of high intensity rain events and high sea levels downstream

*Problem 2.* Without control, the capacity of *Basin 1.* is not utilized, therefore all volumes are bypassing and propagate downstream.

To solve these problems, better information about the behaviour of the system is needed. For this reason, we deployed level sensors in the two basins and three sensors between the basins to learn how rain infiltrates into the network, and most importantly: how the upstream level variations affect the levels downstream.

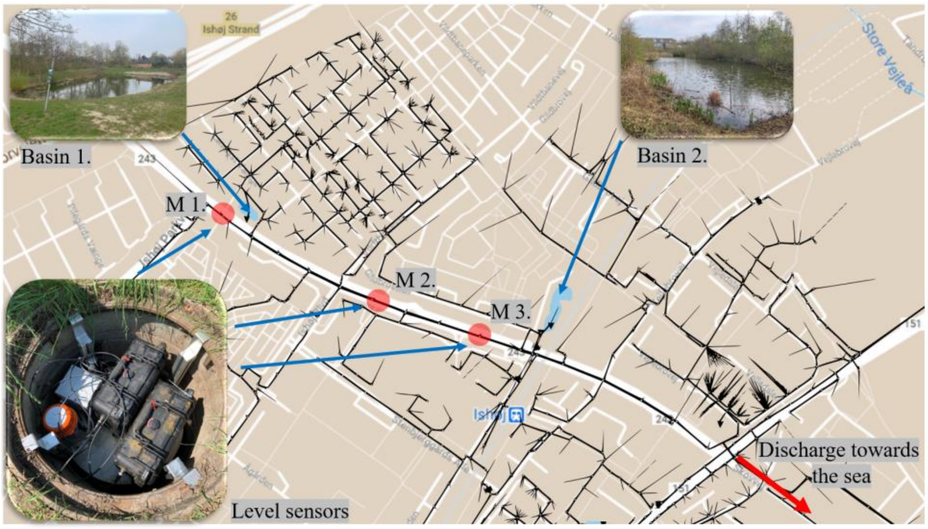


**Figure 2** Storm water network in Ishøj, where red dots denote the placement of water level sensors (M1-3.).

## Results

A visualization interface has been developed on top of the Grafana time series visualization package (Grafana, 2017) shown in Figure 3.

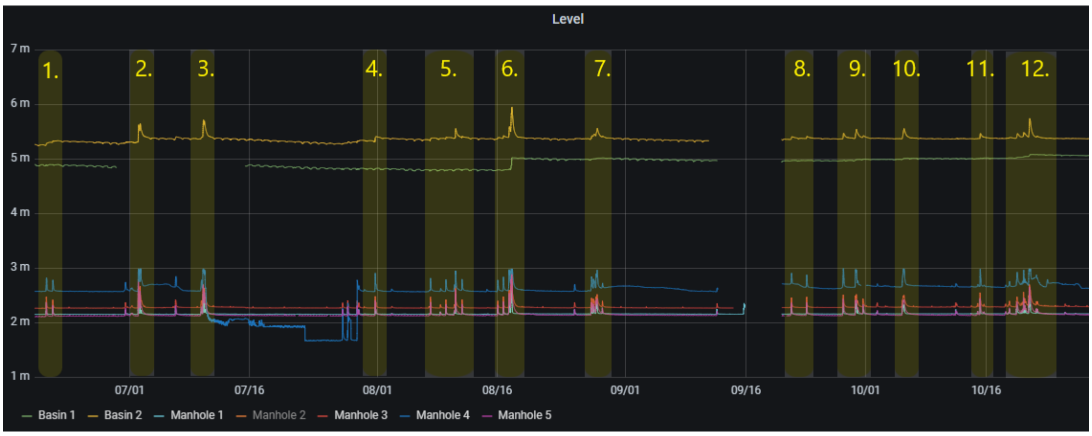


**Figure 3** Grafana visualization interface developed for the case study.

From the collected dataset, we chose 12 rain periods, which we used to evaluate the predicting capabilities of the proposed approach. An example of the training and validation results is shown in Figure 4. As shown, the water levels are trained and validated on approximately half of the collected data, respectively. An interesting event is encircled in blue where the predictions show high uncertainty before a long and high-intensity rain event. This is partly because such an event has not been encountered in the data we used for the training. Moreover, 1-hour predictions are shown in Figure 5. (Note that a more detailed analysis of the results is reserved for the full paper. )

## Conclusions

The results show that the current solution is capable of predicting reasonable levels and uncertainty measures with solely using water level and rain forecast feeds. A next step is to deploy controllable assets at Basin 1. and Basin 2. to build the control capabilities at the utility from our toolchain, serving as a backbone.

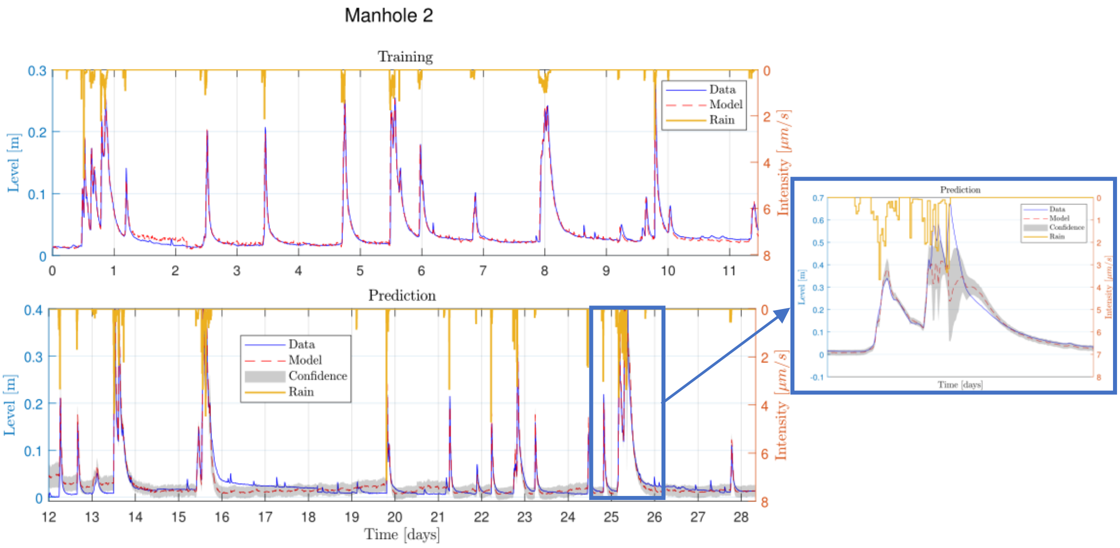


**Figure 4** Example of model training and predicted water level response with characterization of the uncertainty based on the 12 rain events over a five months test period from 16-June-2020 to 27 October 2020.

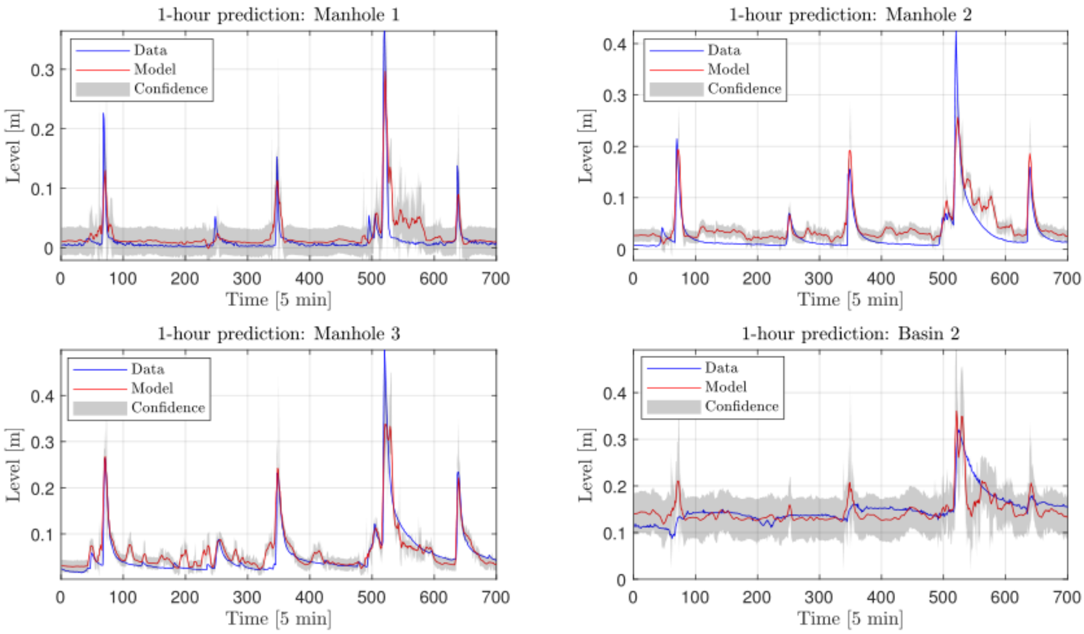


**Figure 5** 1-hour predictions verified over the validation set for the rain event between Day 13 and 14.

## REFERENCES

Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy DeVitry, M., Beutler, P., Maurer, M., 2017. The Potential of Knowing More: A Review of Data-Driven Urban Water Management. *Environmental Science and Technology* 51, 2538–2553

Grafana, 2017, URL: http://grafana.org/.