Water Research 221 (2022) 118782

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

A learning-based approach towards the data-driven predictive control of combined wastewater networks – An experimental study

Krisztian Mark Balla [a,b,*], Jan Dimon Bendtsen [a], Christian Schou [d], Carsten Skovmose Kallesøe [a,b], Carlos Ocampo-Martinez [c]

[a] Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, Aalborg 9220, Denmark
[b] Controls Group, Technology Innovation, Grundfos Holding A/S, Poul Due Jensens Vej 7, Bjerringbro 8850, Denmark
[c] Automatic Control Deptartment, Universitat Politècnica de Catalunya, Llorens i Artigas 4-6, Planta 2, Barcelona 08028, Spain
[d] Digital Water, Water Utility, Grundfos Holding A/S, Poul Due Jensens Vej 7, Bjerringbro 8850, Denmark

ARTICLE INFO

Keywords:
Smart water system
Data-driven modelling
Control
Uncertainty
Sewer network

ABSTRACT

Smart control in water systems aims to reduce the cost of infrastructure expansion by better utilizing the available capacity through real-time control. The recent availability of sensors and advanced data processing is expected to transform the view of water system operators, increasing the need for deploying a new generation of data-driven control solutions. To that end, this paper proposes a data-driven control framework for combined wastewater and stormwater networks. We propose to learn the effect of wet- and dry-weather flows through the variation of water levels by deploying a number of level sensors in the network. To tackle the challenges associated with combining hydraulic and hydrologic modelling, we adopt a Gaussian process-based predictive control tool to capture the dynamic effect of rain and wastewater inflows, while applying domain knowledge to preserve the balance of water volumes. To show the practical feasibility of the approach, we test the control performance on a laboratory setup, inspired by the topology of a real-world wastewater network. We compare our method to a rule-based controller currently used by the water utility operating the proposed network. Overall, the controller learns the wastewater load and the temporal dynamics of the network, and therefore significantly outperforms the baseline controller, especially during high-intensity rain periods. Finally, we discuss the benefits and drawbacks of the approach for practical real-time control implementations.

1. Introduction

The primary function of sewers is to convey wastewater (and stormwater in case of combined networks) towards treatment facilities before releasing it to the environment. Population growth, urbanization, and changing precipitation patterns due to climate change cause major impacts on these networks with increased wastewater and rain loads (Eggimann et al., 2017; Yuan et al., 2019). These loads often result in harmful overflows and degraded treatment performance, threatening the ecological health of the water bodies and damaging the infrastructure (Schütze et al., 2002). Advanced strategies for sewer control are designed on historical weather observations, raising the question of how to operate these infrastructures in the wake of ongoing urbanization and climate change.

1.1. Motivation

To handle the increased load on old infrastructure (without substantial investment), a possible solution is to use advanced control methods, relying on real-time data and system-wide optimization techniques (Yuan et al., 2019). The increased collection and utilization of data enabled the real-time management of urban water systems, forming a basis for advanced decision-making tools (Kitchin, 2014). In the context of sewer networks, these tools aim to prepare the system for high-intensity storm events to optimally utilize the maximum available storage capacity. From a control-theoretic perspective, proactive control, e.g., Model Predictive Control (MPC), has high relevance in sewers, however, in practice reactive control is the most commonly implemented approach (Lund et al., 2018). Decision making by using weather forecasts is a widely used method by researchers in the water community (Campisano et al., 2013).

* Corresponding author at: Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, Aalborg 9220, Denmark.
E-mail address: kballa@grundfos.com (K.M. Balla).

https://doi.org/10.1016/j.watres.2022.118782
Received 13 March 2022; Received in revised form 15 May 2022; Accepted 17 June 2022
Available online 20 June 2022

A significant issue with traditional MPC is the need for a well-maintained system model. At small utilities, such models are often economically out of reach, and therefore neither decision-support nor advanced control techniques are used by the practitioners (Lund et al., 2018). Easy commissioning, therefore, has a great impact in practice, yet it is an unresolved issue when it comes to controlling wastewater networks.

Overflows in sewers often occur due to bottlenecks induced by the slow filling times of storage elements and the significant delays of the sewage transport (Ocampo-Martinez, 2010). The uncertainty associated with the weather forecasts is also an issue, often deteriorating the prediction capabilities of MPC. Consequently, handling the rain and wastewater load via control is a challenging task, not only due to the forecast uncertainty but also due to the uncertainty of the modelling.

To justify the need for autonomous and easy-commissionable control strategies, we introduce first the existing methods. Then, we detail our contributions and specify the control and modelling methods used throughout the paper. Finally, the proposed approach is evaluated on an experimental setup, using real rain and wastewater flow data from real-world utilities.

1.2. State of the art

Instrumentation forms the basis of system-wide planning and automation in urban water systems (Eggimann et al., 2017; Yuan et al., 2019). The data-driven transformation of water system management has resulted in the deployment of a high number of sensors, enabling online monitoring and data processing at many water utilities (Campisano et al., 2013). The most widely used instrumentation in sewers is flow and level sensors, often placed in tanks and manholes (Banik et al., 2017). Flow sensors are typically used for calibrating hydraulic models for planning and decision support (Mignot et al., 2012; Yuan et al., 2019), as well as for modelling the hydrologic processes, e.g., rain running off the catchments (Li et al., 2019). Placement of the actuators and the sensors is a non-trivial task in distributed-parameter systems such as sewer networks. For instance, Leitão et al. (2018) discusses the identification of flow control devices to enable in-sewer storage. In addition to physical sensors, software sensors have also been developed for flow estimation, utilizing mainly weather radar data, pump operation, and the water level variation through level sensors (Ahm et al., 2016; Chen et al., 2014; Kallesøe and Knudsen, 2016; Kisi et al., 2013; Rjeily et al., 2017).

Real-time control in sewer networks converts the latest sensor measurements to operational decisions by the use of controllable assets, e.g., pumps, gates, and valves (Ocampo-Martinez, 2010). The foundation of all predictive decision-making techniques is the underlying dynamic model of the system (Lund et al., 2018). The most intuitive approach to obtain such models is to consider the physics behind the process and apply first-principle hydraulic and hydrologic laws (Todini, 2007), while maintaining the intuition behind the modelling (Balla et al., 2022). However, such models often rely on a high level of detail involving many parameters, and therefore keeping them up-to-date is expensive and laborious (Schütze et al., 2002). Besides, one of the most commonly applied first-principle modelling techniques relies on sets of partial differential equations (Xu et al., 2012; 2011), often requiring prohibitively many sensors for proper calibration. Other physically-based techniques attempt to conceptualize parts of the network, e.g., by considering segments of the system as virtual volumes (Joseph-Duran et al., 2015; Mollerup et al., 2016), and to simplify the model based on skeletonization of the network (Thrysøe et al., 2019; Zhang et al., 2021).

As a result of the increased data availability, data-driven modelling and control techniques have gained popularity within the urban water systems community (Eggimann et al., 2017). Data-driven models (often termed as black-box) are described by their input-output characteristics, where inputs typically comprise the rain forecasts, while the outputs are the corresponding flows (Kitchin, 2014). Neural networks have been applied in modelling the system hydraulics (Dawson and Wilby, 2001; Mounce et al., 2014; Vidyarthi et al., 2020) and the hydrology as well (Chang et al., 2001; Duncan et al., 2012; Rjeily et al., 2017). One of the strengths of neural networks in water systems is their generally high performance of learning complex and nonlinear input-output relations. On the other hand, although generating solutions with neural networks is efficient, they lack the physical interpretability of parameters and depend heavily on data quality.

MPC is a well-suited approach for the optimal mitigation of sewer volumes and regulating the flows with the use of rainfall forecasts (Beeneken et al., 2013; Lund et al., 2018). Characterization of the forecast uncertainties has been reported by considering a multiple scenario approach in both sewers and water resource management (Balla et al., 2020; Tian et al., 2017). In Löwe et al. (2014, 2016) and Vezzaro and Grum (2014), the incorporation of stochastic grey-box models for rainfall-runoff has been considered to reduce combined overflows. Additionally, characterization of the forecast uncertainties by learning the underlying dynamics of the flows with Gaussian processes have been reported in Wang et al. (2016b) for water distribution systems, and in Troutman et al. (2017) for flow prediction in combined sewers.

Reinforcement learning has shown promising results in both combined (Ochoa et al., 2019) and storm water networks (Mullapudi et al., 2020), while iterative learning control has been used to learn the return periods of rain events (Cui et al., 2015). Nevertheless, relatively few studies report on learning-based control in sewer systems. Learning-based control is therefore a research area promising a potential alternative or supplement to the real-time control of wastewater networks.

1.3. Contribution

This paper aims to enable fully automated decision-making in combined sewer systems. The key innovation behind the proposed method relates to its ability to learn and make decisions in real time based on level sensor feeds and weather forecasts. Specifically, the contributions are the following:

- A novel data-driven control approach based on the combination of hydraulic modelling and Gaussian processes,
- An economically and practically feasible predictive controller using solely in-sewer water level observations,
- Uncertainty assessment regarding the system states via propagating the uncertainty through the predictions,
- Experimental overflow control validation.

The proposed solution has two clear benefits. First, in contrast to black-box modelling, the basic hydraulic laws are combined with data-driven techniques. The structure of the model preserves intuition by incorporating the physically measurable levels familiar to practitioners working in the water sector. By utilizing only the easy-accessible physical description of the network, we make our method robust towards forecast uncertainty as well as data deficiency. Additionally, the purchase and maintenance costs related to flow sensors are often expensive in comparison to level sensors (Zhang et al., 2021). In this work, the level-to-flow conversion is bypassed by establishing direct relations between the rain and water levels as well as by relating in-sewer water level measurements to the level variations in the storage tanks at the pumping stations. As opposed to hydraulic modelling, we focus on obtaining a prediction model to evaluate the water levels without the need of the network dimensions and other particular parameters.

2. Problem statement

The overall concept of the method is shown in Fig. 1, where in-sewer level sensors are deployed at critical locations in manholes and basins. The network topology is defined by a tree graph (Thrysøe et al., 2019),

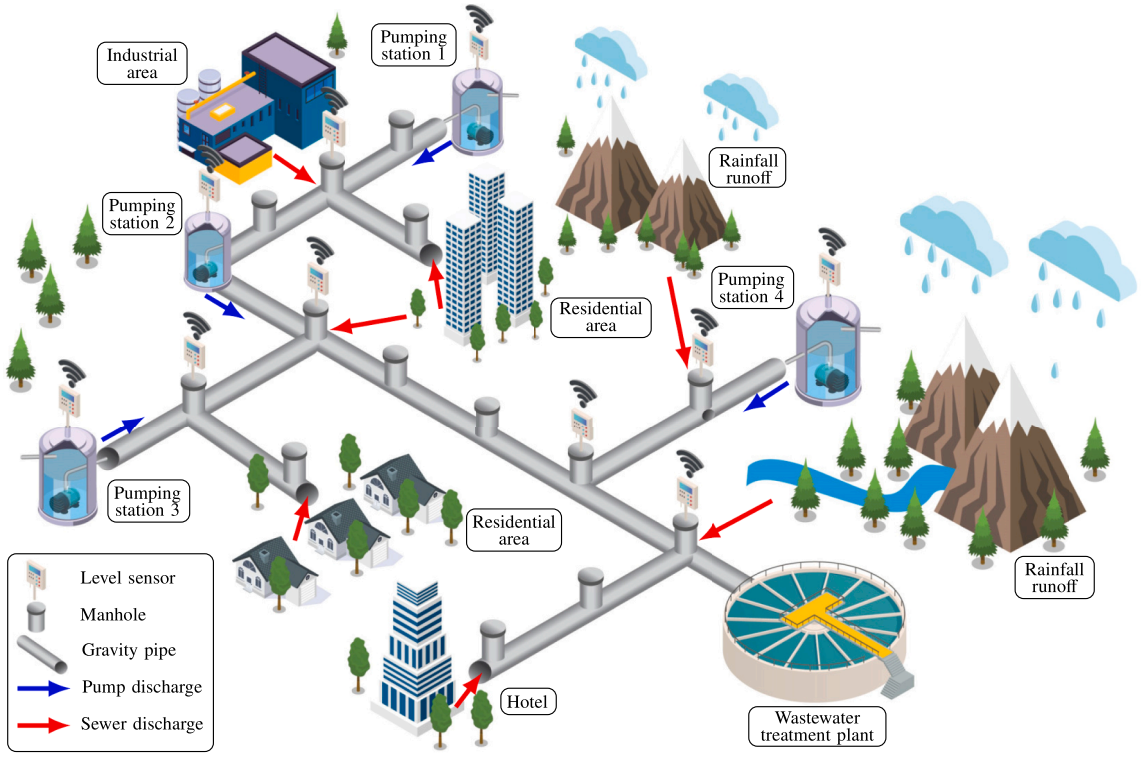

Fig. 1. An illustration of a pumped wastewater network, where water level sensors are deployed in critical points.

where pumping stations are connected via gravity sewers.

Note that the topology is simplified based on the high-level piping layout (Thrysøe et al., 2019), while the infiltration of rain and wastewater is concentrated on network nodes (manholes) being affected by the discharge. The discharged waste- and storm-water are collected and pumped from station to station until the root (treatment plant) is reached. Specifically, we consider the full scale of the network, however, only the main sewer lines between the pumping stations are modelled.

The configuration of our proposed control method is shown in Fig. 2. The models behind the controller are the physical model (Section 3.1) and the data-driven model (Section 3.2). The former incorporates the physical knowledge about the dimension of basins, while the latter describes the effect of rain, wastewater and the uncertainties in forms of residuals by using sensor (h), estimation (Q) and rain forecast (d) data. Opposed to classical methods that handle the inflows (or disturbances) by building individual forecasting blocks, we consider the translation of rain to level variation incorporated in the controller. The Gaussian Process-based MPC controller block (GP-MPC) (Section 3.4.3) stands for the optimization algorithm behind the MPC problem, using a relevant cost function (Section 3.4.2) and the operational and physical constraints (Section 3.4.1). The decision support block (Section 3.4.4) is an information panel providing performance measures of the closed-loop control performance to, e.g., network operators in case the algorithm is used as an offline decision-support tool. As shown, the controller provides flow setpoints to the pumping stations, where the pumps operating in parallel move the water volumes at the rate of the optimal flow (Q_{optimal}). In this study, we focus on variable speed wastewater pumping, hence the hybrid behavior induced by the traditional fixed-speed pump operation is not within the scope of the proposed predictive control algorithm. Since only water level sensors are deployed in the wastewater network, the loop is closed with an observer or pumpflow estimator, allowing for using soft sensing techniques or estimating the

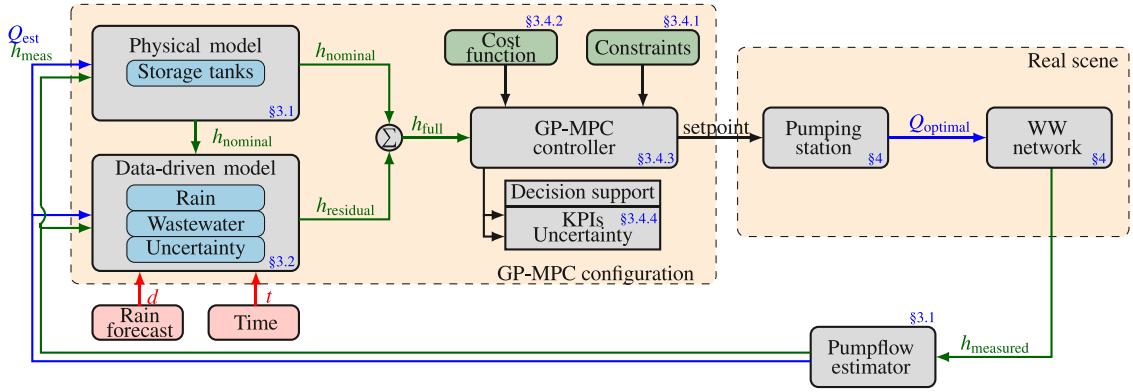

Fig. 2. Closed-loop topology of the GP-MPC controller. Signals denoted with blue are flow variables, green signals are water levels and red signals denote the rain forecast and time. The pipe network (plant) is represented by the *WW (Wastewater) network* block and each block is labeled by the number of section where detailed description of the functionality is given.

pump flows in the proposed output-feedback scheme. In the following, we present the control scheme by describing each building block.

3. Methods

3.1. Physical modelling

The nominal model structure is described by the physical laws of wastewater transport. The information we use are the the topological layout of the network, the size of storage tanks, and the estimated pump flow. Hydraulic storage elements are described with simple mass-balances. Specifically, the level change induced by pump operation is given by

$$\boldsymbol{h}_t(t+1) = \boldsymbol{A}_t\boldsymbol{h}_t(t) + \boldsymbol{B}_t\boldsymbol{Q}(t), \tag{1}$$

where $\boldsymbol{h}_t \in \mathbb{R}^{N_t}$ is the vector of water levels in storage tanks at discrete time t, with N_t being the number of tanks and $\boldsymbol{Q} \in \mathbb{R}^{N_Q}$ is the vector of pump flows representing the sum of flows for each pump at the N_Q pumping stations. The parameter matrices $\boldsymbol{A}_t \in \mathbb{R}^{N_t \times N_t}$ and $\boldsymbol{B}_t \in \mathbb{R}^{N_t \times N_Q}$ are defined by the physical size of the storage elements, i.e., the diameter and the discretization time step or sampling time. The mass balance is described by Eq. (1) with the exception that the effect of inflows, i.e., rain runoff and domestic wastewater are in general unknown, hence not considered as part of the nominal storage dynamics.

The discharged flow of each pump at a pumping station can be accurately approximated with a polynomial expression of each pump sitting in a basin. The pump flow Q is related to the relative pressure, the power and the speed of the pump, described by the following expressions:

$$Q = sa_0\frac{1}{\omega} + sa_1\frac{\Delta p}{\omega} + sa_2\frac{P_p}{\omega^2} + sa_3\omega, \tag{2}$$

where Q is the flow to be estimated in m^3/s, ω is the rotational speed in rad/s, Δh is the level difference between the wastewater basin and the outlet point, and p is the relative pressure to atmospheric pressure, obtained by measuring the inlet pressure and the level in the wet well. Note that p is in mWc, i.e., meter water column. Constants a_i are pump parameters describing the pump curve of the specific pump, assumed to be known in this work. Furthermore, sP_p is the sum of the power consumption of P_p of individual pumps, s denoting the number of running pumps. Several implementations of flow estimation in wastewater pumping stations exist, demonstrating high accuracy in practice (Kallesøe and Knudsen, 2016).

The governing dynamics of the discharged flow propagation in pipes is assumed to be unknown in the nominal model. Therefore, the nominal part of the mass-balance for the entire network is given by the combination of the vector of tank levels $\boldsymbol{h}_t$ and the vector of water levels $\boldsymbol{h}_p \in \mathbb{R}^{N_p}$ in manholes, where system parameters related to pipe dynamics are zero. The full nominal model is given in the standard linear state-space form

$$\boldsymbol{h}(t+1) = \boldsymbol{f}(\boldsymbol{h}(t), \boldsymbol{Q}(t)) = \boldsymbol{A}\boldsymbol{h}(t) + \boldsymbol{B}\boldsymbol{Q}(t), \tag{3}$$

where f represents the known part or nominal dynamics of the wastewater network and $\boldsymbol{h} \in \mathbb{R}^{N_h}$ is the vector of combined water levels where $N_h = N_t + N_p$ corresponds to the number of water level sensors deployed in the entire network. Note that the structure of the state-space model is created based on the network topology, i.e., using the piping layout. In the case of several pumping stations connected by transport pipes, building the mass-balance model can be easily automated by stacking the vectors of suitable dimensions of water levels in Eq. (1).

3.2. Data-driven modelling

The exogenous effect of dry and wet-weather flows are governed by unknown dynamics that are excluded from the nominal model in Eq. (3). These exogenous flows induce variations in the levels in basins and the manholes. Consequently, the flow inside the combined sewer conduits is characterized by the sum of dry-weather discharge (domestic wastewater) and wet-weather discharge (delayed rainfall-runoff), i.e., $q(t) = q_r(t) + q_{ww}(t)$, where q_r and q_{ww} are physically present due to rainfall-runoff and domestic wastewater production, respectively, while q is the combined sewer flow. We consider both combined and storm-water networks, wherein the latter case the network is not influenced by dry-weather flow. However, we do not take into account the groundwater infiltration explicitly as we rather consider it implicitly in the dry-weather flows.

Given water level sensor data h, pump flow estimate Q and rainfall forecast d, the problem is formed by the need to learn the model parts which can complement the nominal dynamics described in Eq. (3). Note that our method does not require to generate the q_r rain-impacted flows and the q_{ww} diurnal flows, as we propose to learn the relation between rain and level in the sewers. With the learned model, we aim to predict the evolution of water levels, i.e., the system states. For this reason, we assume that the entire network dynamics are composed of a nominal and an additive, unknown part. The former represents the known hydraulics of the sewer network, while the latter represents the rain and wastewater flow infiltrating into the system, the pipes transporting the water volumes, moreover the forecast and model uncertainty. The combined network model is given by

$$\boldsymbol{h}(t+1) = \boldsymbol{f}(\boldsymbol{h}(t), \boldsymbol{Q}(t)) + \boldsymbol{B}_p\boldsymbol{g}(\boldsymbol{h}(t), \boldsymbol{Q}(t), \boldsymbol{d}(t), t) + \boldsymbol{w}(t), \tag{4}$$

where $\boldsymbol{g}$ is a nonlinear vector function governing the unknown dynamics, $\boldsymbol{d} \in \mathbb{R}^{N_d}$ is the vector of rainfall forecasts at N_d different locations and $\boldsymbol{w}$ is the process noise $\boldsymbol{w} \sim \mathcal{N}(0, \Sigma_w)\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_w)$, following Gaussian white noise distribution. Besides, $\boldsymbol{B}_p$ is a matrix mapping the nonlinear dynamics $\boldsymbol{g}$ to the full state vector $\boldsymbol{h}$. Simply stated: if there is a storage tank where the level variation $\boldsymbol{h}$ is not affected by uncertain inflows, $\boldsymbol{B}_p$ maps the lower dimensional outputs of the function $\boldsymbol{g}$ to the full state vector $\boldsymbol{h}$ by simply contributing zero to the nominal dynamics.

To generate input data for learning the unknown function $\boldsymbol{g}$, we use the level sensor measurements $\boldsymbol{h}$, the flow estimates $\boldsymbol{Q}$, and the forecast of rainfall $\boldsymbol{d}$.Note that the weather forecast $\boldsymbol{d}$ indicates how the levels are varying inside the sewers due to the infiltration of rain q_r. Since the dry-weather flow q_{ww} typically follows a diurnal pattern, the cyclical behavior correlates to time, i.e., it is likely that the level patterns are similar at the same time of the day at a different week. For this reason, we provide the time of the day t as an input to the unknown function $\boldsymbol{g}$. In this way, we provide information about the time periodicity of the dry-weather level pattern variations. For ease of notation, let us define the input data as a vector

$$\boldsymbol{z} = [\boldsymbol{h}^\intercal, \boldsymbol{Q}^\intercal, \boldsymbol{d}^\intercal, t]^\intercal. \tag{5}$$

To create residuals, we use the data provided by the level sensors and subtract the nominal dynamics, i.e.,

$$\boldsymbol{y}(t) = \boldsymbol{g}\Big(\boldsymbol{z}(t)\Big) + \boldsymbol{w}(t) = \boldsymbol{B}_p^\dagger\Big(\boldsymbol{h}(t+1) - \boldsymbol{f}\Big(\boldsymbol{h}(t), \boldsymbol{Q}(t)\Big)\Big), \tag{6}$$

where $\boldsymbol{y} \in \mathbb{R}^{N_y}$ is the vector of residuals of size N_y, corresponding to the number of water levels influenced by either dry- or wet-weather inflows. Besides, the mapping matrix $\boldsymbol{B}_p^\dagger$ is inverted with the Moore-Penrose pseudo-inverse. By using the level sensors distributed along the network, we aim to capture the dry- and wet-weather sewer level dynamics in the residuals.

The input-output set is constructed with data under nominal operation, i.e.,

$$\mathscr{D} = \{(\boldsymbol{z}(i), \boldsymbol{y}(i)) \mid i = 1, ..., M\}, \tag{7}$$

where M is the number of collected data points. Note that the nominal

operation via the on/off control of the pumps might create spikes in the residual signal $\boldsymbol{y}$ and therefore removing the outliers is recommended as part of data pre-processing.

A powerful way to represent the input-output mapping of $\boldsymbol{g}$ by taking into account the forecast uncertainties is to model the relation as a Gaussian Process. Rather than claiming that the input-output relation above belongs to a specific mathematical model structure, a Gaussian Process is a nonparametric, probabilistic model, based on data. Instead of parameterizing the unknown dynamics governing the residuals, we characterize their distribution. Hence, the residual model representing one entry of $\boldsymbol{y}$ is given by

$$y \sim GP\big(m(z), \boldsymbol{\Sigma}_{GP} + \boldsymbol{I}\sigma_n^2\big), \tag{8}$$

where the distribution of the Gaussian process is fully characterized by its mean function $m(z)$ and covariance $\boldsymbol{\Sigma}_{GP}$. We consider the mean $m(z)$ as a constant, equivalent to a model bias. The noise variance is denoted by σ_n^2 and $\boldsymbol{I}$ is the identity matrix of suitable dimension. The mean and covariance are defined by

$$m(\boldsymbol{z}(i)) = \mathbb{E}\{g(\boldsymbol{z}(i))\}, \tag{9a}$$

$$\Sigma_{GP}(i,j) = \mathrm{cov}(g(\boldsymbol{z}(i)), g(\boldsymbol{z}(j))) \approx k(\boldsymbol{z}(i), \boldsymbol{z}(j)), \tag{9b}$$

where the mean $m(z)$ and the covariance matrix $\boldsymbol{\Sigma}_{GP}$ are obtained by evaluating the mean and covariance functions given all measured data pairs in $\mathscr{D}$. The expected value operator is denoted by $\mathbb{E}\{\cdot\}$. The covariance function or kernel k establishes a measure of similarity between the function values of g. Specifically, our model makes use of the kernel to approximate the covariance of the residual signals. In this setting, we assume that the sewer dynamics exhibit smooth and continuous behavior (based on the slow sewer dynamics), and therefore a squared exponential kernel is used to approximate the covariance function in Eq. (9) (Rasmussen and Williams, 2018). The squared exponential kernel is given by

$$k(\boldsymbol{z}(i), \boldsymbol{z}(j)) = {\sigma_f}^2 \exp\left(-\frac{1}{2}(\boldsymbol{z}(i) - \boldsymbol{z}(j))^{\intercal} \boldsymbol{S}\boldsymbol{\Lambda}^{-1}\boldsymbol{S}^{\intercal}(\boldsymbol{z}(i) - \boldsymbol{z}(j))\right), \tag{10}$$

where the kernel is characterized by its hyper-parameters ${\sigma_f}^2$ and $\Lambda^{-1} = \mathrm{diag}({\sigma_{L,1}}^{-2}, \ldots, {\sigma_{L,N_z}}^{-2})$ denoting the signal variance and the length scale matrix, respectively. Note that we use automatic relevance determination, meaning that we use different length scale parameters for different dimensions of the input vector z (Rasmussen and Williams, 2018). Hence, the relative importance of contribution for each input is assessed.

Using all input dimensions in Eq. (5) for characterizing each residual is computationally demanding, considering that each level sensor and pumping data is used to evaluate the kernel function. Consequently, the mapping matrix $\boldsymbol{S}$ maps only the physically meaningful entries in our data set $\mathscr{D}$, therefore the number of hyper-parameters (σ_L) used for the parameterization of the level residuals is reduced.

It can be shown that the posterior distribution over all possible realization of the unknown dynamics $\boldsymbol{g}$ is given by Bayes' Rule

$$\mathbb{P}\{\boldsymbol{g} \,|\, z, \boldsymbol{y}\} = \frac{\mathbb{P}\{\boldsymbol{g}\}\mathbb{P}\{\boldsymbol{y} \,|\, z, \boldsymbol{g}\}}{\mathbb{P}\{\boldsymbol{y} \,|\, z\}}. \tag{11}$$

Given our problem formulation, the posterior distribution simplifies to (Rasmussen and Williams, 2018)

$$\mathbb{P}\{\boldsymbol{g} \,|\, z, \boldsymbol{y}\} \sim GP(m(z), \boldsymbol{\Sigma}_{GP} + \boldsymbol{I}\sigma 2_n) \tag{12}$$

The hyper-parameters of the above problem are learned by maximizing the marginal likelihood $\mathbb{P}\{\boldsymbol{y} \,|\, z\}$, typically done via numerical approximations, as the analytical evaluation of the above problem is intractable (Chalupka et al., 2013).

Once the hyper-parameters are identified, the Gaussian process model is used to predict the level residual y^* at a test point z^*, using the relation $y^* = g(z^*)$. The problem of predicting the residual corresponds to finding the probability distribution of $\mathbb{P}\{y^* \,|\, \mathscr{D}, z^*\}$, given the training data $\mathscr{D}$, a testing input z^* and the hyper-parameters. By using the kernel to approximate the covariance between the training and testing points, the mean and variance of the Gaussian process are reformulated, i.e.,

$$\mu_{GP}(z^*) = m(z^*) + \boldsymbol{K}_{z^*z}\big(\boldsymbol{K}_{zz} + \boldsymbol{I}{\sigma_n}^2\big)^{-1}(\boldsymbol{y} - m(\boldsymbol{z})), \tag{13a}$$

$$\boldsymbol{\Sigma}_{GP}(z^*) = \boldsymbol{K}_{z^*z^*} - \boldsymbol{K}_{z^*z}\big(\boldsymbol{K}_{zz} + \boldsymbol{I}{\sigma_n}^2\big)^{-1}\boldsymbol{K}_{zz^*}, \tag{13b}$$

where $\boldsymbol{K}_{zz^*} = \boldsymbol{k}(z, z^*)$ and $\boldsymbol{K}_{z^*z} = \boldsymbol{K}^{\intercal}_{zz^*}$ are the covariances between the training and testing points, furthermore $\boldsymbol{K}_{z^*z^*}$ is the autocovariance of the testing point. Note that the above derivation of the GP-based residual model provides a framework to predict the physically meaningful level residuals $\boldsymbol{y}$ based on the similarity of the measured (training) and newly experienced (target) data points regarding the pump flows $\boldsymbol{Q}$, water levels $\boldsymbol{h}$ and the weather forecasts d.

3.3. Probabilistic prediction model

To plan the state of the sewer network over multiple prediction steps ahead, we aim to predict the effect of wet- and dry-weather discharges on the water levels. Predicting multiple steps ahead with the GPs representing $\boldsymbol{g}$ and nominal model $\boldsymbol{f}$ means that the mean and the variance of the previously predicted states are used to predict the next states. Hence, we feed back stochastic variables as inputs. In general, the resulting water level distribution is non-Gaussian, as we propagate the stochastic states through the nonlinear kernel stated in Eq. (10). The resulting distribution is approximated, such that the water levels $\boldsymbol{h}$ and the GP dynamics are approximated at each prediction step t as jointly Gaussian, i.e.,

$$\begin{pmatrix}\boldsymbol{h}(t)\\ GP(t)\end{pmatrix} \sim \mathscr{N}\Big(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t)\Big) = \left(\begin{bmatrix}\boldsymbol{\mu}_h(t)\\ \boldsymbol{\mu}_{GP}(t)\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_h(t) & \boldsymbol{\Sigma}_{h,GP}(t)\\ \boldsymbol{\Sigma}_{GP,h}(t) & \boldsymbol{\Sigma}_{GP}(t)\end{bmatrix}\right), \tag{14}$$

where $\boldsymbol{\Sigma}_{h,GP} = (\boldsymbol{\Sigma}_{GP,h})^{\intercal}$ are the cross-covariances between the physical and the residual water levels, $\boldsymbol{\mu}_h(t)$ is the vector of mean water levels and $\boldsymbol{\Sigma}_h(t)$ is the covariance matrix of the water levels at time t. The covariance of the water levels provides an extra measure of how uncertain we are about our prediction. The pump flows $\boldsymbol{Q}$ are treated as deterministic variables.

To find the transition probability of the full water level states (including the nominal and residual contributions), we apply the first-order Taylor expansion of the approximated joint Gaussian distribution shown in Eq. (14) around the mean $\boldsymbol{\mu}_h(t)$ of the water levels at time step t (Hewing et al., 2020). Note that since we are concerned with the probabilistic description of the water level evolution in our prediction, the characterization of the states is done by considering the expected value $\boldsymbol{\mu}_h$ and the governing uncertainty around it, i.e., the variance of the water levels $\boldsymbol{\Sigma}_h$. Hence, the mean and variance dynamics of the water levels result in

$$\boldsymbol{\mu}_h(t+1) = f(\boldsymbol{\mu}_h(t), \boldsymbol{Q}(t)) + \boldsymbol{\mu}_{GP}(t), \tag{15a}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{h}}(t+1) = \big[\nabla_h \boldsymbol{f}(\boldsymbol{\mu}_h(t), \boldsymbol{Q}(t)), \boldsymbol{B}_p\big]\boldsymbol{\Sigma}(t)\big[\nabla_h \boldsymbol{f}(\boldsymbol{\mu}_h(t), \boldsymbol{Q}(t)), \boldsymbol{B}_p\big]^{\intercal}, \tag{15b}$$

where $\boldsymbol{\Sigma}$ is the joint covariance matrix in Eq. (14) and ∇_h denotes the first-order partial derivative with respect to the water levels. Note that the expected value in terms of the mean water levels is given by the sum of the mass-balance based nominal dynamics $\boldsymbol{f}$ and the contribution of the residual dynamics $\boldsymbol{\mu}_{GP}$. By inserting the model into the nominal dynamics $\boldsymbol{f}$, the final form of the mean-variance dynamics describing the water level evolution becomes

$$\boldsymbol{\mu}_h(t+1) = \boldsymbol{A}\boldsymbol{\mu}_h(t) + \boldsymbol{B}\boldsymbol{Q}(t) + \boldsymbol{B}_p\boldsymbol{\mu}_{GP}(t), \tag{16a}$$

$$\boldsymbol{\Sigma}_h(t+1) = \boldsymbol{A}\boldsymbol{\Sigma}_h(t)\boldsymbol{A}^\intercal + \boldsymbol{B}_p\boldsymbol{\Sigma}_{GP,h}(t)\boldsymbol{A}^\intercal + \boldsymbol{A}\boldsymbol{\Sigma}_{h,GP}(t)\boldsymbol{B}_p{}^\intercal + \boldsymbol{B}_p\boldsymbol{\Sigma}_{GP}(t)\boldsymbol{B}_p{}^\intercal, \tag{16b}$$

where the co-variance update and the cross co-variance between the Gaussian process and the water levels are given by

$$\boldsymbol{\mu}_{GP}(t) = \boldsymbol{\mu}_{GP}(\tilde{\boldsymbol{z}}(t)) \tag{17a}$$

$$\boldsymbol{\Sigma}_{h,GP}(t) = \boldsymbol{\Sigma}_h(\nabla_h \boldsymbol{\mu}_{GP}(\tilde{\boldsymbol{z}}(t)))^\intercal \tag{17b}$$

$$\boldsymbol{\Sigma}_{GP}(t) = \boldsymbol{\Sigma}_{GP}(\tilde{\boldsymbol{z}}(t)) + \nabla_h \boldsymbol{\mu}_{GP}(\tilde{\boldsymbol{z}}(t))\boldsymbol{\Sigma}_h(t)(\boldsymbol{\mu}_{GP}(\tilde{\boldsymbol{z}}(t)))^\intercal, \tag{17c}$$

where the input vector is given by $\tilde{\boldsymbol{z}} = [\boldsymbol{\mu}_h{}^\intercal, \boldsymbol{Q}^\intercal, \boldsymbol{d}^\intercal, t]^\intercal$.

3.4. Predictive control

Regarding the control of large-scale water systems, the popularity of MPC is to a great extent due to the fact that physical and operational constraints are handled in the optimization problem. According to the predictions with the model and the rain forecasts, the MPC algorithm optimizes the manipulated variables (flows or levels) over a given prediction horizon H_p of chosen length. Inputs are computed and evaluated to obtain a future response from the water system. Then, the sequence of future responses is evaluated based on our cost function and we optimize until our numerical solution yields suitable inputs according to a system management policy, which can be identified as optimal. The optimal inputs are then sent to the actuators (pumps or gates) and the entire process is repeated in a receding horizon fashion. In our work, constraints are formed on the physical flow limits of pumps and the physical dimensions of the network, e.g., the capacity of storage tanks and manholes. The disturbances are considered as the wet- and dry-weather flows affecting the sewer network in terms of wastewater flow and rain runoff, among which the latter is of highly stochastic nature. The forecast of these exogenous signals is typically done in terms of nowcasting and forecasting. Nowcasts are obtained by rainfall radars, providing sufficient spatial and temporal reliability up to two hours, while forecasts span over a longer time horizon.

3.4.1. Constraints

Both physical and operational constraints are formulated for the optimization problem associated with the GP-MPC strategy. We consider the sum of each pump unit at the pumping station, hence the constraint on the manipulated flows is given by

$$\boldsymbol{H}_Q\boldsymbol{Q}(t) \leq \boldsymbol{b}_Q, \tag{18}$$

where $\boldsymbol{b}_Q = [\boldsymbol{Q}_{\max}{}^\intercal, \boldsymbol{Q}_{\min}{}^\intercal]^\intercal \in \mathbb{R}^{2N_Q}$ is the vector of upper and lower flow bounds at each pumping station, i.e., the maximum and minimum sum of flow that a station can provide. Furthermore, the matrix $\boldsymbol{H}_Q = [\boldsymbol{I}_{N_Q}, -\boldsymbol{I}_{N_Q}]^\intercal$ maps the vector of pump flow variables $\boldsymbol{Q}$ to the suitable dimensions of $\boldsymbol{b}_Q$.

Constraints on the system states pose limitations on the maximum and minimum water levels. Often the bounds correspond to the capacity of a manhole or a basin. From the physical point of view, it is evident that a combined wastewater network is best prepared for a high-intensity rain event if basins are emptied beforehand. Keeping the water levels as low as possible is particularly important before a storm event, as water volumes might need to be used to the maximum capacity of the piping network. Considering the uncertain nature of rain forecasts and the dynamic nature of wastewater flow patterns, the goal of the controller is to reject the wet- and dry-weather inflows. In this study, we adapt some ideas from predictive control in water distribution networks (Grosso et al., 2014; Wang et al., 2016a), where we introduce an operational constraint. This operational criterion keeps the levels in storage tanks within a specific safety range instead of forcing them to a reference. The functionality of this constraint is to allow the controller to operate the level freely by penalizing only level values which violate the safety bounds. The safety bounds and the operating capacity are illustrated in Fig. 3.

While the minimum and maximum level values of the physical capacity constraints are evident, the determination of the safety bounds is crucial to achieving a proper performance of the closed-loop control strategy. We argue that the safety bounds are placed best at the lower region of tanks, as the system remains emptied and prepared in case of an unexpected storm event. (Furthermore to limit odor problems due to retention.) While finding the optimal placement of the safety region is out of scope here, it is reserved for future simulation studies.

Introducing the nonlinear kernel and propagating the uncertainties with the Gaussian processes result in system states (water levels) being probabilistic, following a Gaussian distribution. Hence the state constraints need to be treated stochastic. In this study, we formulate probabilistic constraints in terms of chance constraints (Wang et al., 2016a), i.e.,

$$\mathbb{P}\{\boldsymbol{H}\boldsymbol{h}(t) \leq \boldsymbol{b}\} \geq \boldsymbol{\alpha}, \tag{19a}$$

$$\mathbb{P}\{\boldsymbol{H}_s\boldsymbol{h}(t) \leq \boldsymbol{b}_s\} \geq \boldsymbol{\alpha}_s, \tag{19b}$$

where Eq. (19a) describes the constraint on the physical capacity of storage elements while Eq. (19b) describes the constraint on the safety region. The operator $\mathbb{P}\{\}$ is the probability that the inequality is satisfied with $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_s$ being the confidence levels. Furthermore, the mapping matrices $\boldsymbol{H}_h = [\boldsymbol{I}_{N_h}, -\boldsymbol{I}_{N_h}]^\intercal$ and $\boldsymbol{H}_s = [\boldsymbol{I}_{N_{h_t}}, -\boldsymbol{I}_{N_{h_t}}]^\intercal$ map the vector of water levels $\boldsymbol{h}$ to suitable size of $\boldsymbol{b} = [\boldsymbol{h}_{t,\max}{}^\intercal, \boldsymbol{h}_{t,\min}{}^\intercal]^\intercal$ and $\boldsymbol{b}_s = [\boldsymbol{h}_{s,\max}{}^\intercal, \boldsymbol{h}_{s,\min}{}^\intercal]^\intercal$ water level bounds, respectively. Note that the lower bounds corresponding to $\boldsymbol{b}$ are defined by the minimum volume in the wet wells to avoid the dry-run of the pumps.

Under our assumptions that $\boldsymbol{h}$ is jointly Gaussian with the residuals $\boldsymbol{y}$, the above probabilistic expressions can be reformulated as convex, deterministic constraints (Hewing et al., 2020; Wang et al., 2016a). The constraints are given by

$$\boldsymbol{H}\boldsymbol{\mu}_h(t) \leq \boldsymbol{b} + \boldsymbol{H}_\epsilon\boldsymbol{\epsilon}(t) - \boldsymbol{c} \odot \boldsymbol{H}\mathrm{diag}(\boldsymbol{\Sigma}_h(t))^{\frac{1}{2}}, \tag{20a}$$

$$\boldsymbol{H}_s\boldsymbol{\mu}_h(t) \leq \boldsymbol{b}_s + \boldsymbol{H}_\xi\boldsymbol{\xi}(t) - \boldsymbol{c}_s \odot \boldsymbol{H}_s\mathrm{diag}(\boldsymbol{\Sigma}_h(t))^{\frac{1}{2}}, \tag{20b}$$

where the actual water level values are replaced by their expected or mean values $\boldsymbol{\mu}_h$. Furthermore, we introduce a term called the vector of critical values $\boldsymbol{c} = \phi(\boldsymbol{\alpha})^{-1}$, where $\phi(\cdot)$ is the vector of inverse cumulative distribution function (or quantile) of the standard Gaussian distribution evaluated at $\boldsymbol{\alpha}$. These quantiles can be precomputed and used as constant values. The operator $\odot$ denotes element-by-element multiplication and the slack terms $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_{\max}{}^\intercal, \boldsymbol{\epsilon}_{\min}{}^\intercal]^\intercal$ and $\boldsymbol{\xi} = [\boldsymbol{\xi}_{\max}{}^\intercal, \boldsymbol{\xi}_{\min}{}^\intercal]^\intercal$ denote vectors of relaxation variables standing for safety violation and overflow, respectively. The mapping matrices $\boldsymbol{H} \in \mathbb{R}^{2N_h \times N_h}$ and $\boldsymbol{H}_s \in \mathbb{R}^{2N_t \times N_t}$ map the mean water level $\boldsymbol{\mu}_h$ and variance $\boldsymbol{\Sigma}_h$ to the suitable dimensions of the maximum and minimum water level bounds $\boldsymbol{b}$ and $\boldsymbol{b}_s$. Note that the additional term in Eq. (20) corresponds to the tightening of the original bounds, conditioned on the evolution of the water level variances along the prediction horizon. As expected, the longer we predict into the future, the higher the variances grow due to the model and forecast uncertainties. To avoid recursive infeasibility, the slack variables ξ and ϵ

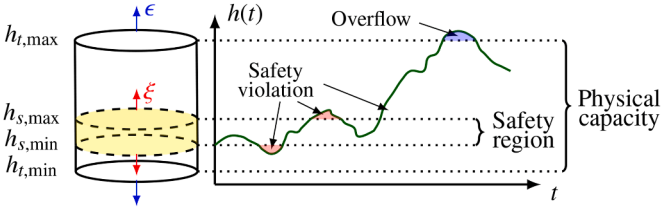

Fig. 3. Safety and capacity constraints, where blue and red arrows are constraint relaxations for overflow (ϵ) and safety violation (ξ).

are utilized to soften the constraints.

3.4.2. Cost function

The cost function is the key component in the design of the GP-MPC. In general, the formulation of the control problem relates to the manipulation of water volumes to avoid undesirable overflows and water surges outside the main sewer lines. From the control point of view, we focus on the rejection of the stochastic meteorological (rain-runoff) and human (wastewater flow) loads, aiming to avoid the physical constraint violations resulting in overflows or water surges. Although here we propose a specific objective function, there is a flexibility of either removing or adding control objectives simply by adding new control goals. For example, the control strategy may vary according to the infrastructure design, e.g., the inclusion of treatment plant objectives may be crucial to add in combined networks with high wastewater load. In this work, we focus on the following operational and management criteria (listed in decreasing order of priority)

I. Minimise overflow in storage elements
II. Minimise safety volume violation
III. Minimise the water level in storage elements
IV. Minimise the control action of pumps

The predefined objectives are aggregated in a multi-objective cost function to fulfill all control criteria. As the evolution of the water levels is described by an approximated joint Gaussian probability distribution, the cost function is formulated on stochastic variables. The overall cost of the control problem is formed as expected values, given by

$$\mathscr{J}(t) = \mathbb{E}\left\{ \underbrace{W_1 \| \boldsymbol{\epsilon}(t)\|^2_{\lambda_1}}_{\text{I.}} + \underbrace{W_2 \| \boldsymbol{\xi}(t)\|^2_{\lambda_2}}_{\text{II.}} + \underbrace{W_3 \| \boldsymbol{h}(t)\|^2_{\lambda_3}}_{\text{III.}} + \underbrace{W_4 \| \Delta\boldsymbol{Q}(t)\|^2_{\lambda_4}}_{\text{IV.}} \right\} \tag{21}$$

where the different control objectives are prioritized through the W weighting constants. Furthermore, these weights also normalize each objective such that water levels and flows become comparable in magnitude. *Cost* I. represents the overflow penalty, where the use of slack variable ϵ represents the water level exceeding the physical bounds of the basins. The amount of overflow shared between pumping stations is prioritized with the diagonal λ_1 matrix, where λ_1 is diagonal and $\boldsymbol{0} \leq \lambda \leq \boldsymbol{I}$, similarly to all λ matrices. Moreover, the weight constant W_1 is significantly higher than any other weights, as using the overflow variables is undesirable. *Cost* II. corresponds to the safety slack, while *Cost* III. penalizes the level in storage tanks and manholes. By adjusting λ_3, the filling sensitivity of storage tanks or manholes can be adjusted, meaning that storage nodes prone to overflows are filled slower and emptied faster than less sensitive storage elements. Note that *Cost* IV on minimizing the pumpflows is formulated on the variation of the signal $\Delta Q(t) = Q(t) - Q(t-1)$, accounting for integral action enabling smooth system response.

The slack variables representing overflow ϵ and the safety violation ξ are decision variables, similarly to the change of flow ΔQ for pumps. The decision variables are considered deterministic, therefore the only stochastic term in Eq. (21) is *Cost* III. Taking the expected value of the quadratic term results in the following expression (Hewing et al., 2020):

$$\mathbb{E}\left\{W_3 \| \boldsymbol{h}(t)\|^2_{\lambda_3}\right\} = W_3\left[\|\boldsymbol{\mu}_h(t)\|_{\lambda_3}{}^2 + \mathrm{tr}\{\lambda_3 \boldsymbol{\Sigma}_h(t)\}\right], \tag{22}$$

where tr{} is the trace operator and the expected value results in the mean $\boldsymbol{\mu}_h$ and the covariance $\boldsymbol{\Sigma}_h$ of the water level values.

3.4.3. Optimization problem

Bringing together the approximations of the water levels and the Gaussian processes, furthermore the expected values of both the constraints and cost function, we introduce the tractable form of the optimization problem behind the GP-MPC algorithm (indicated in Fig. 2). The problem is given by

$$\underset{\substack{\Delta\boldsymbol{Q}(0),\dots,\Delta\boldsymbol{Q}(H_p-1)\\ \boldsymbol{\epsilon}(0),\dots,\boldsymbol{\epsilon}(H_p-1)\\ \boldsymbol{\xi}(0),\dots,\boldsymbol{\xi}(H_p-1)}}{\text{Minimize}} \sum_{t=i}^{i+Hp-1} W_1 \| \boldsymbol{\epsilon}(i)\|^2_{\lambda_1} + W_2 \| \boldsymbol{\xi}(i)\|^2_{\lambda_2} + W_3\left[\|\boldsymbol{\mu}_h(i)\|^2_{\lambda_3} + \mathrm{tr}\{\lambda_3 \boldsymbol{\Sigma}_h(i)\}\right] + W_4 \| \Delta\boldsymbol{Q}(i)\|^2_{\lambda_4}, \tag{23a}$$

subject to

$$\boldsymbol{\mu}_h(i+1) = f(\boldsymbol{\mu}_h(i), \boldsymbol{Q}(i)) + \boldsymbol{\mu}_{GP}(i), \tag{23b}$$

$$\boldsymbol{\Sigma}_h(i+1) = \left[\nabla_h \boldsymbol{f}(\boldsymbol{\mu}_h(i), \boldsymbol{Q}(i)), \boldsymbol{B}_p\right] \boldsymbol{\Sigma}(i) \left[\nabla_h \boldsymbol{f}(\boldsymbol{\mu}_h(i), \boldsymbol{Q}(i)), \boldsymbol{B}_p\right]^{\mathsf{T}}, \tag{23c}$$

$$\Delta\boldsymbol{Q}(i) = \boldsymbol{Q}(i) - \boldsymbol{Q}(i-1), \tag{23d}$$

$$\boldsymbol{H}_Q \boldsymbol{Q}(i) \leq \boldsymbol{b}_Q, \tag{23e}$$

$$\boldsymbol{H}\boldsymbol{\mu}_h(i) \leq \boldsymbol{b} + \boldsymbol{H}_\epsilon \boldsymbol{\epsilon}(i) - \boldsymbol{c} \odot \boldsymbol{H}\mathrm{diag}(\boldsymbol{\Sigma}_h(i))^{\frac{1}{2}}, \tag{23f}$$

$$\boldsymbol{H}_s \boldsymbol{\mu}_h(i) \leq \boldsymbol{b}_s + \boldsymbol{H}_\xi \boldsymbol{\xi}(i) - \boldsymbol{c}_s \odot \boldsymbol{H}_s \mathrm{diag}(\boldsymbol{\Sigma}_h(i))^{\frac{1}{2}}, \tag{23g}$$

$$\boldsymbol{\epsilon}(i) \geq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\xi}(i) \geq \boldsymbol{0}, \tag{23h}$$

$$\boldsymbol{\mu}_{GP}(i), \boldsymbol{\Sigma}_{GP}(i) \text{according to Eq. (13)} \tag{23i}$$

$$\boldsymbol{\Sigma}(i) \text{according to Eq. (14)} \tag{23j}$$

$$\boldsymbol{\mu}_h(0) = \boldsymbol{h}(i), \quad \boldsymbol{\Sigma}_h(0) = 0, \tag{23k}$$

where the minimization is solved at time t for every $i = 0, \dots, H_p - 1$ along the prediction horizon H_p in a receding horizon fashion. Note that the optimization problem is subject to the dynamic network equations in Eqs. (23b) and (23c), forming equality constraints. Furthermore, the rain forecast d is included in these constraints, as the mean and variance $\boldsymbol{\mu}_{GP}$, $\boldsymbol{\Sigma}$ are constructed based on the z training data vector. After solving the optimization problem in Eq. () at state $\boldsymbol{h}(0)$, the resulting decision variables form an optimal control sequence of the change in pumpflows $\boldsymbol{u} = [\Delta\boldsymbol{Q}^{\mathsf{T}}(0), \Delta\boldsymbol{Q}^{\mathsf{T}}(1), \dots, \Delta\boldsymbol{Q}^{\mathsf{T}}(H_p - 1)]$, where only the first row of $\boldsymbol{u}$ is used. Note that the vector of slacks $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$ are also decision variables obtained via the optimization.

3.4.4. Key performance indicators

The proposed approach is aimed for the online, automatic control of combined or separated wastewater networks. The approach, however, serves as a toolchain not only for closed-loop control but as a decision support for water practitioners (information panel in Fig. 2). We aim to support decision making by providing Key Performance Indicators (KPIs) for predicting overflows, assess the uncertainty of the predicted water levels and to provide information about the safety region in storage tanks. The KPIs are given by

$$\mathrm{KPI}_\xi = \frac{1}{H_p} \sum_{k=1}^{N_Q} \sum_{i=0}^{H_p-1} \xi_k(i), \tag{24a}$$

$$\mathrm{KPI}_\epsilon = \frac{1}{H_p} \sum_{k=1}^{N_Q} \sum_{i=0}^{H_p-1} \epsilon_{\max,k}(i), \tag{24b}$$

$$\mathrm{KPI}_{\Delta Q} = \frac{1}{H_p} \sum_{i=0}^{H_p-1} \Delta\boldsymbol{Q}^{\mathsf{T}}(i) \Delta\boldsymbol{Q}(i), \tag{24c}$$

$$\mathrm{KPI}_{\Sigma} = \frac{1}{H_p}\sum_{i=0}^{H_p-1}\mathrm{tr}\{\boldsymbol{\Sigma}_h(i)\}, \tag{24d}$$

where the performance indicator in Eq. (24a) is related to the safety bound violation, the KPI in Eq. (24b) for overflows, Eq. (24c) assesses the smooth performance of the pumping and the KPI in Eq. (24d) is related to the amount of uncertainy along the prediction horizon H_p, respectively. The KPI indicating the level of potential overflow is only assessed for the slack variable $\epsilon_{\max}$, corresponding to the level violation for the upper capacity limit of basins and manholes. Note that all KPIs are averaged along the prediction horizon and considered for the entire wastewater network. Ideally, the KPIs accounting for overflows and safety violation should be zero, meaning that the pumps counteract the wet and dry-weather flow disturbances and they respect the safety requirements. In practice, the stochastic disturbances are complicated to forecast and the uncertainty in the model and in the forecast are always present.

Additionally, our KPIs are also intended to indicate whether it is necessary to activate a fallback scenario in case extreme weather conditions are experienced for which the infrastructure was not designed. In this case, we cannot assure that our controller learns to react safely.

3.4.5. Implementation

The control algorithm and the interfacing software to the experimental setup are available on an open-source web repository (https://github.com/csocsidior/LB-GP-based_WWnetwork_control). The data collected during the experiments have also been attached to the web repository to allow practitioners and researchers to evaluate our implementation. Additionally, a simulator environment replicating the topology of our laboratory equipment is also provided. The algorithm has been implemented on a Windows OS desktop computer with a 3.6 (GHz), Intel Xeon machine with 64 GB RAM, and the software has been written in `Matlab`. The real-time control algorithm has been interfaced with `Simulink`, and the data was obtained and locally managed at each unit of the experimental setup with a `CODESYS` soft-PLC in real time (3S-Smart Software Solutions GmbH). The optimization problem related to the GP-MPC controller has been solved via direct multiple shooting in the symbolic framework `CasADI` (Andersson et al., 2019) with a pimer-dual interior point solver `IPOPT` (Wächter and Biegler, 2006). For finding the hyperparameters of the Gaussian processes, we used the `fitrgp` toolbox in `Matlab`.

4. Case study

To show the practical feasibility of the learning-based framework, the controller is deployed on a laboratory setup, emulating a combined wastewater network. This laboratory setup enables us to prototype our control solution serving as proof-of-concept without the risk of compromising the operation of real-world infrastructure. (A detailed description of the test setup can be found in Val Ledesma et al. (2021).) Besides, the experimental tests conducted in this paper are inspired by a real wastewater network topology located in Gram, Denmark, proposing a realistic control problem. The configuration is shown in Fig. 4.

The replicate of the network segment is a 1 : 80 scale of the real infrastructure. Therefore, the resolution of the time scale and the wet- and dry-weather flows are scaled down accordingly. Specifically, the diurnal pattern of wastewater is scaled to 19 min, corresponding to one day in real life. While the data acquisition is done at every 0.5 s, the control time step is 10 s, equivalent to sending a control signal every 12

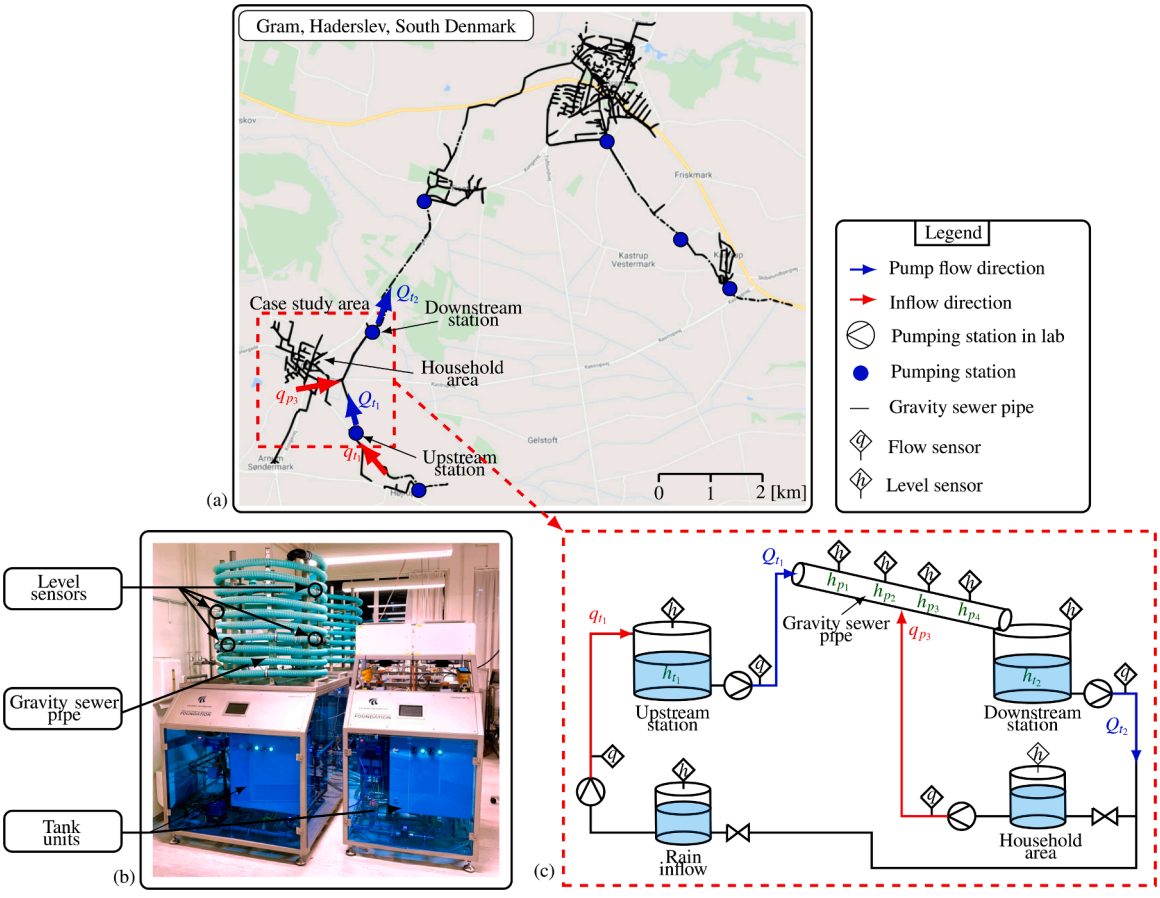

Fig. 4. Case study area of a combined wastewater network in Gram, Denmark (a) and the equivalent representation of the considered network segment by the Smart Water Infrastructures Laboratory, where (b) is showing the experimental setup and (c) the schematics of the topology.

min in real life. Besides, the data used in the experiment are real wastewater and rain precipitation[1]. It is important to note that, due to the limitations of the experimental setup, the proper characterization of the runoff dynamics is out of scope in this study. Therefore, both the rain and wastewater discharges have been emulated based on the historical flow measurements. However, the forecast of the rain is provided from outside of the optimization problem statement, and therefore our method and structure are still valid and remain the same for real-world implementation where the synchronization of the rain and wastewater flows is characterized properly.

The experimental setup consists of an upstream and downstream pumping station connected via a sewer pipe, where the water volumes are transported with pumps. Lateral inflow from household areas enters the system by discharging wastewater at the middle point of the pipeline. As indicated in Fig. 4, the laboratory setup is equipped with level sensors distributed along the open-channel pipes and storage tanks together with flow sensors at the pumping stations. Although flow estimation is given by Eq. (2), we simply use the available sensors on the setup.

Following the methods in Section 3, the nominal model of the network is assessed. Specifically, we have $N_Q = 2$ pumping stations, $N_p = 1$ level sensor in pipes and $N_t = 2$ at the stations. The dry- and wet-weather flows enter the system at $N_d = 2$ points, where rain infiltrates the system at pumping station t_1. The training data array for identifying the data-driven part of the model in Eq. (5) is constructed as $z(t) = [\boldsymbol{h}^\intercal(t), \boldsymbol{Q}^\intercal(t), d(t), t]^\intercal$, where the level and flow signals at time t are given by

$$\boldsymbol{h}(t) = \left[h_{t_1}(t), h_{t_2}(t), h_{p_3}(t)\right]^\intercal, \tag{25}$$

$$\boldsymbol{Q}(t) = \left[Q_{t_1}(t), Q_{t_2}(t)\right]^\intercal. \tag{26}$$

Out of the four available level sensors in manholes, we use h_{p_3} placed after the connection of the lateral inflow pipe. We argue that the sensor measurement located at this point captures sufficient information to model how the pump Q_{t_1}, and disturbance flows q_{p_3} enter the channel. Then, the nominal parameters of the wastewater network are given by

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_{2\times 2} & \boldsymbol{0}_{2\times 1} \\ \boldsymbol{0}_{1\times 2} & 0 \end{bmatrix}, B = \begin{bmatrix} \frac{T_s}{\tau_{t_1}} & 0 \\ 0 & \frac{T_s}{\tau_{t_2}} \\ \boldsymbol{0}_{1\times 2} & \end{bmatrix}, \tag{27}$$

where T_s denotes the sampling time of the controller, while τ_1 and τ_2 are the storage tank parameters representing the geometry and size of the tanks. It is important to note that the experiments are carried out such that the water recirculates in the system, meaning that the flows and volumes need to be balanced. For this reason, the controlled pumps cannot turn off to zero flows, as expected in a real-world implementation. Instead, the operating range of the pumped flows is lifted to a value where the network can run for long experiments without emptying the *Rain inflow* and *Household area* auxiliary tanks.

5. Results and discussion

5.1. Residual model training

Given the physical model, the residuals $\boldsymbol{y} \in \mathbb{R}^3$ can be constructed based on the water level measurements $\boldsymbol{h} \in \mathbb{R}^3$. As stated in Section 3.2, beyond the sensor availability, knowledge of the physical system plays a significant role in the training efficiency of the model. To find the hyperparameters for each GP, the dimension of the training data set is reduced according to Eq. (10) by using the slicing matrices. These matrices define which dimensions of the original training set z influence the given residual based on the topological layout of the system. Matrices $\boldsymbol{S}$ are mapped to a lower dimension than N_z, where the hyper-parameters σ_L are picked that are physically relevant to parameterize the given residual. In our specific case study, to train the GP on residual y_1 (corresponding to the upstream tank t_1), the predictors $Q_{t1}, (4)$, the rain forecasts $d, (6)$, and the time $t, (7)$ are used, hence only three hyper-parameters are employed instead of using the full size $N_z = 7$ inputs and the corresponding hyper-parameters. This is well-aligned with our physical insights, as we can observe from the visual inspection of the water level variation in the upstream tank that both the dry- and wet-weather flows and the corresponding pumps influence the signal. The illustration of the feature selection is shown in Fig. 5. The collected measurement data for training is obtained under the nominal operation of the network. We consider the nominal operation of pumping stations when pumps operate with threshold-based control rules, most commonly applied by wastewater utilities (Lund et al., 2018).

To test the modelling capabilities of the Gaussian process model fitted to the residuals, the collected data have been divided into a training and validation sets. The GP models have been trained on 80% of the collected data set, corresponding to 60 days of on/off operation. The rest of the data (15 days) have been used for validating the results. Fig. 6 shows the three residuals constructed from the measurement data $\boldsymbol{h}$ obtained via the level sensors. It is seen from these results that the predictions with the GP model match the level residual observations within the validation period in the two tanks and the pipes. Furthermore, except for some outlier points, the confidence interval characterized by the variance of the GP process covers the distribution of the data points well. The variations in the data are primarily due to the noise and the measuring precision of the sensors. As seen in residual y_1, removing the effect of the nominal dynamics from the original level signal results in the daily diurnal level variation patterns induced by the dry-weather discharges, and level peaks due to the wet-weather rain precipitation. It is worth noting that the performance of our level predictions using rain forecasts is underpinned by the fact whether we observed rain episodes similar to the current forecast before. Besides, note that our experimental test setup has physical limitations of how different rain flow profiles we can create. This might partly explain why our final model exhibited such suitable performance in predicting the combined level variations (e.g., residual y_1 in Fig. 6).

The data describing residual y_2 are related to the level variations induced by the discharged pump flow coming from the upstream station t_1 and the lateral inflow coming from the household area. Note that the level variations due to the pumping activity of such pumps sitting at the downstream station are not visible on the signal in Fig. 6 as we removed the pump flow time series scaled to the level variations in the tank. It is seen from the signal that the diurnal lateral flows (q_{p_3}) coming from upstream induce the level variations in the downstream tank. Lastly, residual y_3 describes exactly the level variations in the pipe without any modifications, as the dynamics of flow propagation in pipes are not characterized by any physically-based nominal dynamics in our study. Incorporating physical knowledge (e.g., travel time, level attenuation) into the pipe residual model is of course possible in specific cases, but

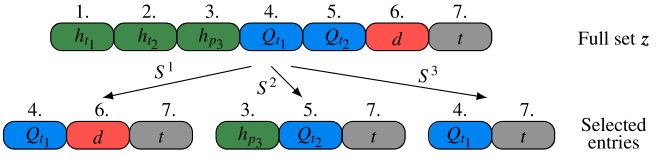

Fig. 5. Feature selection with the slicing matrices $\boldsymbol{S}$.

[1] The rain data has been obtained through the Danish Meteorological Institute's Open Data application interface (https://confluence.govcloud.dk/display/FDAPI). The wastewater data has been obtained from the utility *Fredericia Spildevand og Energi A/S* in Denmark.

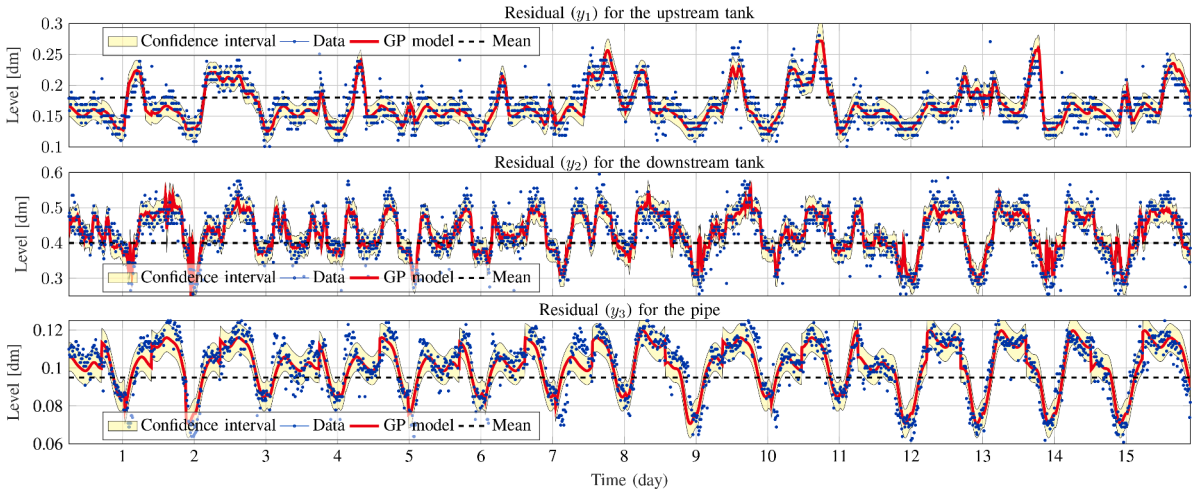

Fig. 6. Validation of the GP model with the residuals regarding the level variations in the two tanks and in the sewer pipe, respectively.

ignored in this initial evaluation; investigation of this will be reserved for future studies. As seen, the variations mainly occur due to the dry-weather lateral inflow from the household area (which we desire to capture through this signal), while the jumps observed in both the predictions and the data are due to the pumping cycles coming from the upstream station.

So far, we verified our assumptions on the input selection based on our physical insights. However, it is crucial to make sure that our model captures the correlation between each input dimension of the training set z used for the residual predictions. Since the GP models are used to solve an optimization problem through multiple-step predictions, we need to make sure that the decision variables are properly captured in the model. Hence, the following measure is introduced to measure the relevance of each input on the corresponding residuals:

$$\bar{r}_i = \frac{\exp\left(-\sigma_{L,i}\right)}{\sum_{j=1}^{N_L} \sigma_{L,j}}, \tag{28}$$

where $\bar{r}_i$ is the normalized relevance of the i^{th} predictor and N_L is the number of length-scale hyperparameters used for the given output residual. The relevant data inputs receive positive values between one and zero, while a value close to zero indicates irrelevant input data.

The comparison of input relevance corresponding to each GP model is shown in Fig. 7. As seen in residual y_1, the time input t (used to describe the diurnal variation of wet-weather flows) is dominant compared to the rain forecasts d and to the pumping activity Q_{t_1}. This fact is in line with our expectations as the majority of the residual data incorporates information about the diurnal wastewater activity, while the rain peaks appear less often in the time series. It is also seen that the pump flow data are quite irrelevant when we predict with the model. This verifies our method since the effect of the pump dynamics is part of the nominal model, hence it should not affect the residual.

The relevance bars of residual y_2 show that the level variation in the sewer pipe discharging to the downstream tank (h_{p_3}) has a high relevance, verifying our initial assumptions, as the only discharge source is the flow gravitated down from the upstream tank. Note, however, that our model shows some correlation between the nominal pump flows Q_{t_2} and the time input t. A possible explanation for this fact might be that in case of high loads, both pumping stations turn on approximately at the same time, meaning that Q_{t_2} and h_{p_3} inhabit similar characteristics. Moreover, we select the time input to model each residual, in case there are some additional periodic components in the signal not described by the level sensor in the gravity pipe. Lastly, the water level variation in the sewer pipe is induced by the pumps upstream Q_{t_1} and by the lateral inflow q_{p_3}, which we model inherently by providing time t as an input. It is worth noting that we do not distinguish between weekdays and weekends. This means that the predicted diurnal patterns represent an average model, which considers the similarity between any days in our training set.

5.2. Closed-loop control experiment

The experimental evaluation of the learning-based predictive controller has been carried out with an $H_p = 20$ steps horizon, which is equivalent to a four-hour ahead prediction in real life. It should be noted that the computational complexity of solving the optimization problem in Eq. (22) is highly dependent on the GP model used for learning the dry-weather flows and the unmodelled dynamics. From the implementation point of view, propagating the uncertainty depends on the number of data points that we use in our optimization problem, as μ_{GP} and Σ_{GP} are conditioned on the observed data and therefore evaluating Eq. (13) has a cost growing with the number of points. To overcome this issue, we select a subset of $M = 80$ data points from the $\mathscr{D}$ training set with a criteria that these points need to be close to the previously

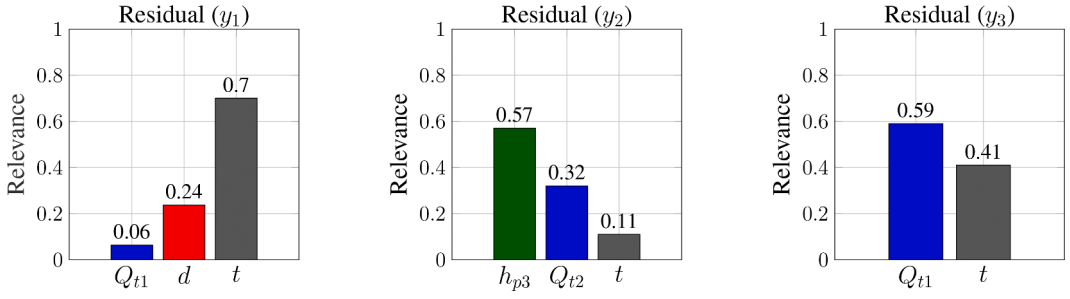

Fig. 7. Relevance of the regressors showing the effect of the input data on the residuals.

predicted state trajectories. Hence, we assume that the previous solution trajectory will lie close to the current one, which is fair considering that wastewater networks inhabit slowly-varying dynamics. Although several sparse GP approximations exists (Hewing et al., 2020), here we implement the most simple version and reserve more advanced sparse approximations for future studies. Furthermore, we add new level, pump flow and forecast points at every second control step to our data dictionary $\mathscr{D}$, i.e., we continuously learn new state-action-forecast pairs. Note that the controller is launched after the model is pre-trained on the 60 days of training data previously obtained from the nominal operation, hence the point selection already has a wide feature-space to select from.

The closed-loop control results obtained from our experimental setup aim to show the benefits of distributing the water level sensors in combination with using the residual-based physical and GP-based data-driven techniques to learn the dynamics of a network-scale control problem. To assess the performance of the GP-MPC, the method is compared with a standard baseline controller, meaning that we emulate the same scenarios and run the two different controllers under the same physical and control properties. We aim to emphasize how an easy-commissionable controller can challenge the simple baseline controllers, especially when the sewer system is facing capacity problems under heavy hydrologic load (Lund et al., 2018). In our implementation, both controllers act globally and compute the flow reference signals to the local PI controllers governing the pumps. To stretch both controllers to their capacity limits, a period equivalent to 18 days in real life with heavy rain periods has been chosen, forcing the network to overflow due to its insufficient storage capacity. The results of the experiment are shown in Fig. 8. The figure compares both control scenarios by showing the forecasts and the discharged inflows entering the system (a-b), the water level in each tank (c-d), the volume of actual overflow escaping from the tanks (e-f), and finally the control decisions at the two pumping stations made by the learning-based GP-MPC (g-h) and by the on/off baseline controller (i-j).

Overflows are triggered several times while running the baseline controller due to the lack of collaboration between the upstream and downstream pumping stations. Opposed to on/off operation, it is clear that the GP-MPC at the upstream tank shifts the timing of the pumping under heavy rain events. Note that the controlled flow at the upstream station Q_{t_1} rarely reaches its upper flow limit and often reduces the outflow, thereby saving the capacity downstream. By delaying the flows from the upstream station, the GP-MPC controller allowed the downstream tank to drain and to spend less time overflowing. This shift in time and the flow reduction is observed between Day $1-2$, and between Day $14-17$. Looking at the control actions between the latter period (Day 14 and 15), the system is exposed to an extreme event, where a high-intensity and long-duration rain event is about to be forecasted. During this episode, the control actions at the upstream pumping station start to oscillate when the controller realizes that the safety bounds need to be violated and the upstream tank need to use the slack variables for overflows to reduce the overall accumulated spilled volumes.

The KPI corresponding to the control actions is verifying this behavior, shown in Fig. 9. For various practical reasons, this action is undesirable. However, the upstream pumping station indeed overflows at Day 16, while the water levels at the downstream basin hover just below the upper physical level limits. A possible explanation for this behavior might be the type of rain event forecasted at Day 14. In Fig. 9 in the last row, the uncertainty predicted by the controller is assessed in comparison to the different forecasts. Short uncertainty peak can be explained by the low performance of the point selection, meaning that

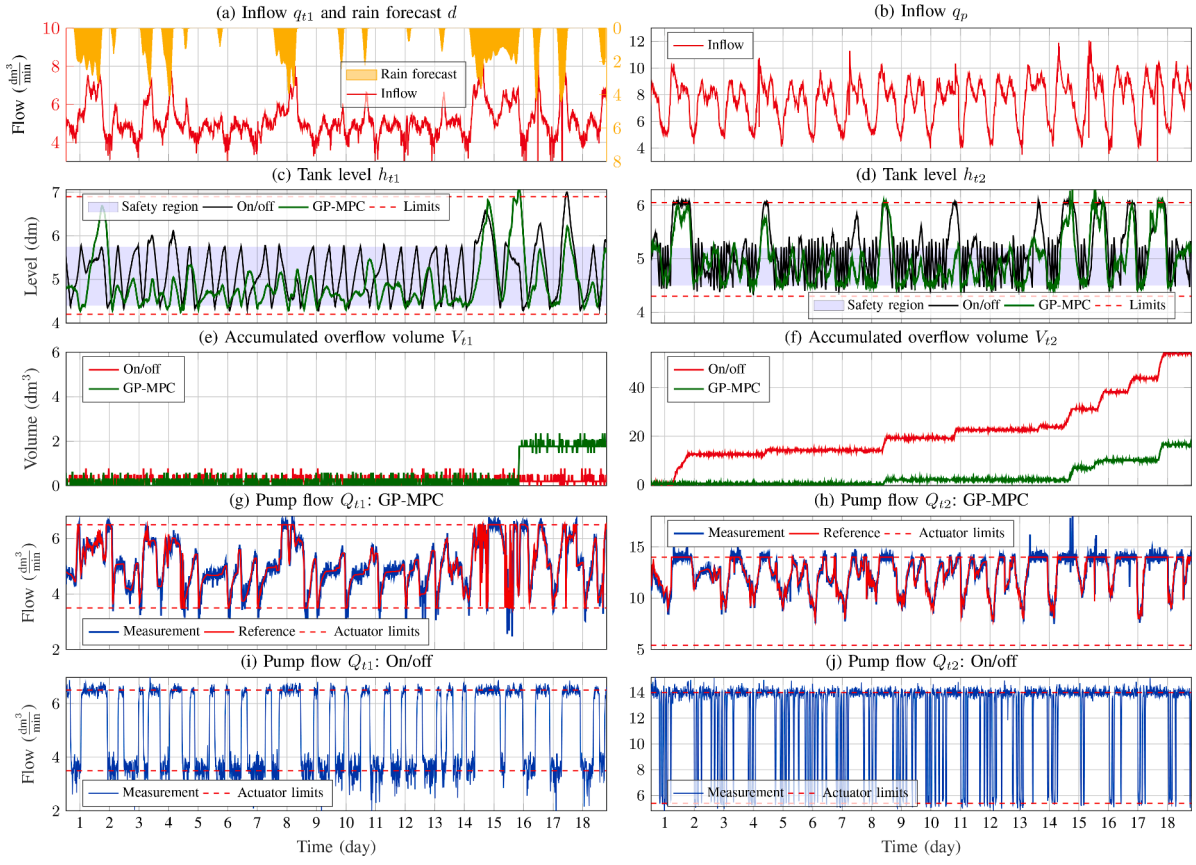

Fig. 8. Performance comparison between the GP-MPC and On/off controller, operating the experimental setup representing the case study area during an 18-days period with heavy rain-events. Inflows, water levels in the basins, accumulated overflow and the control actions are shown regarding the upstream pumping station (left column) and the downstream pumping station (right column).

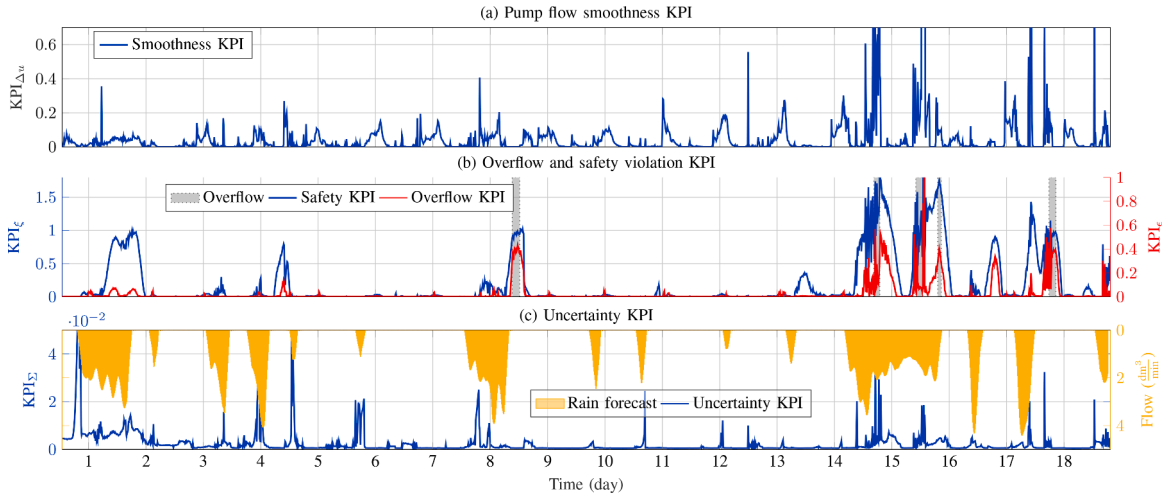

Fig. 9. Key performance indicators provided by the user in run time, showing at each control step the smoothness of the control action, overflow, safety violation and uncertainty averaged over the prediction horizon.

the points selected from the feature space $\mathscr{D}$ are not suitably representing the currently forecasted scenario. This is visible at most times instants under rain forecasts. However, the uncertainty remains high during the two longest rain events between Day $1-2$ and $14\ -\ 16$, respectively. This fact indicates that even though the point selection with $M = 80$ points allows solving the optimization problem under two seconds on average, the prediction quality and smoothness of the control action are degraded significantly. Moreover, our experimental tests confirm that the performance of the GP-MPC is quite sensitive to the formulation and tuning of the objective function. As seen from the KPIs between Day $14-15$, the uncertainty remains high during the rain events, indicating that the points we use for the predictions do not describe the forecasted scenario in a proper way. At the same time, the rain has a long duration and its intensity triggers overflow in the predictions. However, as the uncertainty grows due to the bad description of the data, the controller attempts to minimize the variance to the cost of not reporting overflows.

Note that between Day $9-14$ there is a relatively dry period, where the controller at both stations makes the outflows of the pumps mimic the daily diurnal flow variations induced by the wastewater flowing into the system. This indicates that the Gaussian process part of the model predicts an average wastewater inflow with an uncertainty bound that fits the actual inflows quite well. Thereby the pumps exhibit smooth control actions resulting in smooth variations inside the safety region defined in the tanks (marked with the blue area in Fig. 8 (c-d)).

The results illustrated here show a number of benefits and challenges to using the GP-MPC scheme to learn and predict the dry- and wet-weather flows from the level variations occurring in combined wastewater networks. Arguably, the major benefit of learning from the level data is the ability to launch the controller without developing control models relying on the level to flow conversion. However, as the experimental tests have shown, the adoption of the method is challenged by several practical issues. Since the effect of the inflows is handled by the Gaussian processes, the contribution of the data-driven decision-making cannot be easily explained and explicit guarantees cannot be given. However, using sparse approximations of the available training data sets is anticipated to increase the quality of predicting the residuals. To further improve the robustness of the controller, instead of choosing M exact points for the covariance matrices, it is anticipated that approximating the original training data matrix with an M dimensional sparse matrix based on the point selection will improve the uncertainty propagation.

6. Conclusions

This paper introduced a Gaussian process-based predictive control algorithm for the real-time control of wastewater networks. While flow modelling with Gaussian processes has been successfully used in water systems before, to our knowledge this is the first instance where the methods have been applied and verified experimentally in real-time control without the use of any flow sensors. The methods proposed here and our experimental tests showed promising results in using the domain knowledge combined with the data-driven model to make automated decisions on a network scale. The proposed control architecture has the potential to serve as either an online or an offline decision-support tool to control actuators in wastewater networks, predict overflows and assess the uncertainty of the decisions. To that end, the formulations and real-time results provided by this paper should serve as a basis to support data-driven predictive control as a feasible solution in wastewater networks.

Further research on how the flow and level propagation in the transport pipes could be incorporated into the nominal dynamics promises a potential for improvement in the robustness of the proposed study. This is a matter of future work.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Krisztian Mark Balla reports financial support was provided by Innovation Fund Denmark. Krisztian Mark Balla reports financial support was provided by Grundfos Holding AS. Carsten Skovmose Kallesoe reports financial support was provided by Grundfos Holding AS. Christian Schou reports financial support was provided by Grundfos Holding AS. Carlos Ocampo-Martinez reports financial support was provided by Advanced Learning-Based supervision for Efficiency and Safety in smart infrasTructures L-BEST Project. Carsten Skovmose Kallesoe reports a relationship with Grundfos Holding AS that includes: employment and equity or stocks. Christian Schou reports a relationship with Grundfos Holding AS that includes: employment and equity or stocks. Krisztian Mark Balla reports a relationship with Grundfos Holding AS that includes: employment.

Acknowledgments

The authors would like to acknowledge the Poul Due Jensen Foundation for providing the Smart Water Laboratory for testing, the

wastewater utility in Fredericia, Denmark for providing the historical data for our experiments, and the Provas Utlity in Gram, Denmark for inspiration of the network structure. This work was funded by Innovation Fund Denmark and Grundfos Holding A/S (Ref. 9065-00018A) as part of a Danish Industrial Ph.D. project. The work of C. Ocampo-Martinez has been supported by the project PID2020-115905RB-C21 (L-BEST) funded by MCIN/ AEI /10.13039/501100011033.

References

Ahm, M., Thorndahl, S., Nielsen, J.E., Rasmussen, M.R., 2016. Estimation of combined sewer overflow discharge: A software sensor approach based on local water level measurements. Water Sci. Technol. 74 (11), 2683–2696.
Andersson, J.A., Gillis, J., Horn, G., Rawlings, J.B., Diehl, M., 2019. CasADi: a software framework for nonlinear optimization and optimal control. Math. Program. Comput. 11 (1), 1–36.
Balla, K.M., Schou, C., Bendtsen, J.D., Ocampo-Martínez, C., Kallesøe, C.S., 2022. A nonlinear predictive control approach for urban drainage networks using data-driven models and moving horizon estimation. IEEE Trans. Control Syst. Technol. (Early Access) 1–16.
Balla, K.M., Schou, C., Dimon Bendtsen, J., Kallesoe, C.S., 2020. Multi-scenario model predictive control of combined sewer overflows in urban drainage networks. 2020 IEEE Conference on Control Technology and Applications (CCTA). IEEE, Montréal, pp. 1042–1047.
Banik, B.K., Alfonso, L., Di Cristo, C., Leopardi, A., Mynett, A., 2017. Evaluation of different formulations to optimally locate sensors in sewer systems. J. Water Resour. Plann. Manag. 143 (7), 04017026.
Beeneken, T., Erbe, V., Messmer, A., Reder, C., Rohlfing, R., Scheer, M., Schuetze, M., Schumacher, B., Weilandt, M., Weyand, M., 2013. Real time control (RTC) of urban drainage systems - A discussion of the additional efforts compared to conventionally operated systems. Urban Water J. 10 (5), 293–299.
Campisano, A., Cabot Ple, J., Muschalla, D., Pleau, M., Vanrolleghem, P.A., 2013. Potential and limitations of modern equipment for real time control of urban wastewater systems. Urban Water J. 10 (5), 300–311.
Chalupka, K., Williams, C.K., Murray, I., 2013. A framework for evaluating approximation methods for Gaussian process regression. J. Mach. Learn. Res. 14 (1), 330–350.
Chang, F.J., Liang, J.M., Chen, Y.C., 2001. Flood forecasting using radial basis function neural networks. IEEE Trans. Syst., Man Cybern. Part C 31 (4), 530–535.
Chen, J., Ganigué, R., Liu, Y., Yuan, Z., 2014. Real-time multistep prediction of sewer flow for online chemical dosing control. J. Environ. Eng. 140 (11), 04014037.1–04014037.9.
Cui, Y., Jin, M., Li, D., Xi, Y., Cen, L., 2015. Iterative learning predictive control for urban drainage systems. Chinese Control Conference, CCC, pp. 4107–4112.
Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Prog. Phys. Geogr. 25 (1), 80–108.
Duncan, A.P., Chen, A.S., Keedwell, E.C., Djordjević, S., Savić, D.A., 2012. Urban flood prediction in real-time from weather radar and rainfall data using artificial neural networks. Weather Radar and Hydrology: IAHS Red Book Proceedings. Exeter, pp. 568–573.
Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy De Vitry, M., Beutler, P., Maurer, M., 2017. The potential of knowing more: A review of data-driven urban water management. Environ. Sci. Technol. 51 (5), 2538–2553.
Grosso, J.M., Ocampo-Martinez, C., Puig, V., Joseph, B., 2014. Chance-constrained model predictive control for drinking water networks. J. Process Control 24 (5), 504–516.
Hewing, L., Kabzan, J., Zeilinger, M.N., 2020. Cautious model predictive control using Gaussian process regression. IEEE Trans. Control Syst. Technol. 28 (6), 2736–2743.
Joseph-Duran, B., Ocampo-Martinez, C., Cembrano, G., 2015. Output-feedback control of combined sewer networks through receding horizon control with moving horizon estimation. Water Resour. Res. 24 (5), 504–516.
Kallesøe, C.S., Knudsen, T., 2016. Self calibrating flow estimation in waste water pumping stations. Proceedings of the 2016 European Control Conference (ECC2016). IEEE, Aalborg, pp. 55–60.
Kisi, O., Shiri, J., Tombul, M., 2013. Modeling rainfall-runoff process using soft computing techniques. Comput. Geosci. 51, 108–117.
Kitchin, R., 2014. The real-time city? Big data and smart urbanism. GeoJournal 79 (1), 1–14.
Leitão, J.P., Carbajal, J.P., Rieckermann, J., Simões, N.E., Sá Marques, A., de Sousa, L.M., 2018. Identifying the best locations to install flow control devices in sewer networks to enable in-sewer storage. J. Hydrol. 556, 371–383.
Li, J., Sharma, K., Liu, Y., Jiang, G., Yuan, Z., 2019. Real-time prediction of rain-impacted sewage flow for on-line control of chemical dosing in sewers. Water Res. 149, 311–321.
Löwe, R., Thorndahl, S., Mikkelsen, P.S., Rasmussen, M.R., Madsen, H., 2014. Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar. J. Hydrol. 512, 397–407.
Löwe, R., Vezzaro, L., Mikkelsen, P.S., Grum, M., Madsen, H., 2016. Probabilistic runoff volume forecasting in risk-based optimization for RTC of urban drainage systems. Environ. Model. Softw. 80, 143–158.
Lund, N.S.V., Falk, A.K.V., Borup, M., Madsen, H., Steen Mikkelsen, P., 2018. Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management. Crit. Rev. Environ. Sci. Technol. 48 (3), 279–339.
Mignot, E., Bonakdari, H., Knothe, P., Lipeme Kouyi, G., Bessette, A., Rivière, N., Bertrand-Krajewski, J.L., 2012. Experiments and 3D simulations of flow structures in junctions and their influence on location of flowmeters. Water Science and Technology 66 (6), 1325–1332.
Mollerup, A.L., Mikkelsen, P.S., Sin, G., 2016. A methodological approach to the design of optimising control strategies for sewer systems. Water Sci. Technol. 83, 103–115.
Mounce, S.R., Shepherd, W., Sailor, G., Shucksmith, J., Saul, A.J., 2014. Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. Water Sci. Technol. 69 (6), 1326–1333.
Mullapudi, A., Lewis, M.J., Gruden, C.L., Kerkez, B., 2020. Deep reinforcement learning for the real time control of stormwater systems. Adv Water. Resour. 140, 103600.
Ocampo-Martinez, C., 2010. Model Predictive Control of Wastewater Systems, 1st. Springer, Barcelona.
Ochoa, D., Riano-Briceno, G., Quijano, N., Ocampo-Martinez, C., 2019. Control of urban drainage systems: Optimal flow control and deep learning in action. Proceedings of the American Control Conference, pp. 4826–4831.
Rasmussen, C.E., Williams, C.K.I., 2018. Gaussian Processes for Machine Learning. MIT press Cambridge, MA.
Rjeily, Y.A., Abbas, O., Sadek, M., Shahrour, I., Chehade, F.H., 2017. Flood forecasting within urban drainage systems using NARX neural network. Water Sci. Technol. 76 (9), 2401–2412.
Schütze, M.R., Butler, D., Beck, M.B., 2002. Modelling, Simulation and Control of Urban Wastewater Systems, 2nd. Springer.
3S-Smart Software Solutions GmbH. Codesys. https://www.codesys.com.
Thrysøe, C., Arnbjerg-Nielsen, K., Borup, M., 2019. Identifying fit-for-purpose lumped surrogate models for large urban drainage systems using GLUE. J. Hydrol. 568, 517–533.
Tian, X., Negenborn, R.R., van Overloop, P.J., Maestre, J.M., Sadowska, A., van de Giesen, N., 2017. Efficient multi-scenario model predictive control for water resources management with ensemble streamflow forecasts. Adv. Water Resour. 109, 58–68.
Todini, E., 2007. Hydrological catchment modelling: Past, present and future. Hydrol. Earth Syst. Sci. 11 (1), 468–482.
Troutman, S.C., Schambach, N., Love, N.G., Kerkez, B., 2017. An automated toolchain for the data-driven and dynamical modeling of combined sewer systems. Water Res. 126 (1), 88–100.
Val Ledesma, J., Wisniewski, R., Kallesøe, C.S., 2021. Smart water infrastructures laboratory: Reconfigurable test-beds for research in water infrastructures management. Water (Switzerland) 13 (13), 1875.
Vezzaro, L., Grum, M., 2014. A generalised Dynamic Overflow Risk Assessment (DORA) for Real Time Control of urban drainage systems. J. Hydrol. 515, 292–313.
Vidyarthi, V.K., Jain, A., Chourasiya, S., 2020. Modeling rainfall-runoff process using artificial neural network with emphasis on parameter sensitivity. Model. Earth Syst. Environ. 6, 2177–2188.
Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. 106 (1), 25–57.
Wang, Y., Ocampo-Martinez, C., Puig, V., 2016. Stochastic model predictive control based on Gaussian processes applied to drinking water networks. IET Control Theory Appl. 10 (8), 947–955.
Wang, Y., Ocampo-Martinez, C., Puig, V., Quevedo, J., 2016. Gaussian-process-based demand forecasting for predictive control of drinking water networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 69–80.
Xu, M., Negenborn, R.R., van Overloop, P.J., van de Giesen, N.C., 2012. De Saint-Venant equations-based model assessment in model predictive control of open channel flow. Adv. Water Resour. 49 (December), 37–45.
Xu, M., van Overloop, P.J., van de Giesen, N.C., 2011. On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow. Adv. Water Resour. 34 (2), 282–290.
Yuan, Z., Olsson, G., Cardell-Oliver, R., van Schagen, K., Marchi, A., Deletic, A., Urich, C., Rauch, W., Liu, Y., Jiang, G., 2019. Sweating the assets The role of instrumentation, control and automation in urban water systems. Water Res. 155 (1), 381–402.
Zhang, Q., Zheng, F., Jia, Y., Savic, D., Kapelan, Z., 2021. Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems. Water Res. 188 (1), 116544.