Real-time Data-driven Modelling and Predictive Control of Wastewater Networks

Balla, Krisztian Mark

Publication date:
2022

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

REAL-TIME DATA-DRIVEN MODELLING AND PREDICTIVE CONTROL OF WASTEWATER NETWORKS

BY
KRISZTIAN MARK BALLA

DISSERTATION SUBMITTED 2022

AALBORG UNIVERSITY
DENMARK

Real-time Data-driven Modelling and Predictive Control of Wastewater Networks

Ph.D. Dissertation

by

Krisztian Mark Balla

Dissertation submitted June, 2022

Dissertation submitted: June, 2022

PhD supervisors: Associate Professor Jan Dimon Bendtsen
Aalborg University

Professor Carsten Skovmose Kallesøe
Aalborg University and Grundfos Holding A/S

Lead Business Dev. Manager Christian Schou
Grundfos Holding A/S

PhD committee: Professor Roozbeh Izadi-Zamanabadi (chairman)
Aalborg University, Denmark

Professor Eric Kerrigan
Imperial College, United Kingdom

Professor Peter Steen Mikkelsen
Technical University of Denmark, Denmark

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-876-2

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

kballa@grundfos.com
cballa7@gmail.com
ORCID ID: 0000-0001-9338-7643

Printed in Denmark by Stibo Complete, 2022

To my parents and Cecilie, who only vaguely understand what I do for a living but are supportive nonetheless.

Abstract

Sewers convey wastewater and stormwater towards treatment facilities before releasing the water into the environment. The operation of wastewater networks is challenged by several factors, such as the rapid population growth, urbanization, and the increased load on the infrastructure due to changing weather conditions. Nonetheless, most sewers operate without any form of supervision or control. Consequently, urban flooding, sewer overflows, and suboptimal operation of the treatment plant are common in cities, and therefore untreated sewage discharge is a significant issue today.

One way to handle the increased load on the under-dimensioned sewer infrastructure (without substantial capital investment) is to better utilize the available network capacity via real-time modelling and control. Rather than expanding the capacity, assimilating data into the modelling and using real-time sensor and forecast feeds provide a solution with the capability of adapting to the changing conditions and environment. Predictive control is a widely-used control solution for the optimal management of water volumes in wastewater systems, however, often underpinned by the fact that a well-maintained model is required behind the controller. Easy commissioning has a great impact in practice, and therefore maintaining models with many details is often economically out of reach for smaller water utilities.

This thesis is concerned with the data-driven modelling and predictive control of combined wastewater and stormwater networks. One of the contributions of the thesis is to bridge the gap between hydraulic and hydrologic modelling, typically used individually as two disciplines but both necessary for predictive control in wastewater networks. To this end, we utilize grey-box techniques relying on the Saint-Venant partial differential equations to model the high-level piping layout of the network, combined with volume-based mass conservation dynamics. To tackle the challenges associated with the infiltrating disturbance flows, we propose to learn the dynamic effect of the wet- and dry-weather disturbances through the variations of easy-commissionable level sensors distributed through the network. We adopt a generic Gaussian process-based predictive control framework, where the data-driven Gaussian processes are obtained via training on the residuals generated between the real-time level measurements and the pre-identified physically-based, yet, grey-box models of the sewer hydraulics.

To show the practical feasibility of the approach, the closed-loop control has been tested in high-fidelity simulation environments and also on an experimental laboratory network, while the prediction capabilities have been validated through the course of a pilot project with a real-world water utility. The methods presented in this thesis have been verified experimentally without the use of any flow sensors deployed in the network. Compared to rule-based control methods (typically used at utilities), the proposed control methods have achieved a cumulative overflow decrease ranging from 10 to 28 percent under extreme load and over a one-month study period.

Resumé

For at undgå at spildevand og regnvand slippes ud i miljøet, er der etableret et kloaknetværk der transporterer vandet til renseanlæg, som renser vandet før det ledes ud i miljøet. Adskillige problemer, såsom hurtig befolkningstilvækst, urbanisering, samt øget belastning på kloaknettet på grund af skiftende vejrforhold, gør det vanskeligt at drive kloaknetværket. På trods af disse problemer er størstedelen af kloakkerne ikke uden opsyn og er ukontrollerede. Som følge heraf sker det utilsigtede kloakoverløb og ineffektiv spildevandsrensning på rensningsanlæggene, og udledning af urenset spildevand til følsomme recipienter et stort problem i dag.

En mulighed for at håndtere den øgede belastning af den underdimensionerede kloakinfrastruktur er at bruge realtidsmodellering og -styring for herved bedre at udnytte den tilgængelige netværkskapacitet. Assimilering af data i modellering og brug af realtidssensor- og prognose-input, i stedet for at tilføje kapacitet i kloakinfrastrukturen, giver en løsning, der tilpasser sig ændrede forhold og miljø. Prædiktive kontrol er en meget anvendt kontrolteknik til styring af vandmængder i spildevandssystemer. Dog kræver denne form for styring præcise og opdaterede modeller af kloakinfrastrukturen, hvilket hæmmer udbredelse af metoden. I praksis har nem idriftsættelse og vedligehold stor betydning for spildevandsselskaberne, og vedligeholdelse af modeller, til det niveau som er nødvendig til prædiktive kontrol, er derfor ofte ikke attraktiv for mindre selskaber.

Datadrevne modellering og prædiktive kontrol af spildevands- og regnvandsnetværk er fokus i denne afhandling. En af afhandlingens resultater er at bygge bro mellem hydraulisk og hydrologisk modellering, som normalt behandles separat, men begge er nødvendige for prædiktive kontrol af kloaknetværket. Til dette formål kombinerer vi volumenbaseret massebevarelsesdynamik med gråboksmodelleringsteknikker baseret på Saint-Venants ligningerne. Herved opnås en model, der beskriver netværkets rørtopologi, samt dets dynamiske forhold. Modellerne identificeres ud fra niveaumålinger placeret på strategiske steder i kloaknettet. Vi bruger et datadrevet Gaussisk-procesbaseret prædiktivt kontrolsystem, hvor de datadrevne Gaussiske processer læres fra residualler skabt mellem realtidsniveaumålinger og de præ-identificerede fysisk baserede gråbokse kloakmodeller.

Lukkesløjfe-styringen er blevet evalueret i high-fidelity modeller, samt på et

eksperimentelt laboratorienetværk for herved at demonstrere den praktiske anvendelighed, mens metodens evne til at forudsige volumendynamikken er blevet valideret gennem et pilotprojekt med et spildevandsselskab i den virkelige verden. Testen af de foreslåede kontrolmetoder har vist en kumulativ overløbsreduktion på mellem 10 og 28 procent under tung belastning sammenlignet med regelbaserede kontrolmetoder (normalt brugt på spildevandsselskaberne).

Összefoglaló

A csatornák a szennyvizet és a csapadékvizet szennyvíztisztító üzemek felé szállítják, mielőtt kiengednék a vizet a természetbe. A szennyvízhálózatok működését olyan tényezők nehezítik, mint a gyors népességnövekedés, a városiasodás, vagy az infrastruktúrára nehezedő nyomás, melynek legfőbb okozója az időjárási körülmények változása. Ennek ellenére a szennyvízcsatornák többsége bármiféle irányítás nélkül működik. Ennek következtében a városokban gyakran fordulnak elő árvizek és a csatornákból feljövő áradások. Továbbá, a szennyvíz tisztítóüzemek többnyire szuboptimálisan működnek, ezért a tisztítatlan szennyvíz természetbe való kiengedése napjaink súlyos problémájának számít.

Az alulméretezett csatornázási infrastruktúra kezelésének egyik módja a rendelkezésre álló hálózati kapacitás hatékonyabb kihasználása valós idejű modellezés és szabályozás révén. A kapacitás bővítése helyett a mérési adatok modellezésbe való beépítése, valamint a valós idejű szenzorok és időjárási előrejelzések használata olyan megoldást nyújt, amely képes adaptálni a változó körülményekhez és környezethez. A prediktív szabályozás egy széles körben alkalmazott szabályozási módszer a szennyvíz vízmennyiségének optimális kezelésére, azonban a szabályozó fenntartásához jól kalibrált modellek szükségesek. A gyakorlatban az irányítástechnika egyszerű üzembe helyezésének nagy jelentősége van, ezért a részletes modellek kalibrálása és fenntartása sok esetben gazdaságilag kivitelezhetetlen a kisebb vízművek számára.

Ezen szakdolgozat témája a kombinált szennyvíz- és csapadékvíz-hálózatok adatvezérelt modellezése és prediktív szabályozása. A dolgozat célul tűzi ki a hidraulikai és a hidrológiai modellezés közötti szakadék áthidalását. Ezeket jellemzően két külön tudományágként kezelik, azonban mind a kettő szükséges a szennyvízhálózatok prediktív szabályozásához. Ehhez a Saint-Venant parciális differenciálegyenleteken alapuló grey-box (szürke doboz) modellezési módszert alkalmazzuk a hálózat csővezeték-rendszerének modellezésére, kiegészítve térfogat alapú tömegmegőrzés-dinamikán alapuló modellekkel. A rendszert terhelő zavaró áramlásokat és a nedves és száraz időjárási zavarok (csapadék és szennyvíz) dinamikus hatásainak vizsgálatát a hálózatban elhelyezett, könnyen üzembe helyezhető szintérzékelők ingadozásain keresztül hajtjuk végre. Egy Gauss-folyamatokon alapuló prediktív szabályozási keretrendszert alkalmazunk, amelyben az adatvezérelt Gauss-folyamatokat a valós idejű vízállás

szintmérései és a csatorna hidraulikájának előre meghatározott, fizikai alapú szürke doboz modelljei közötti reziduumok révén kapjuk.

Jelen módszer gyakorlati kivitelezhetőségének bizonyítása érdekében a visszacsatolt (zárt) rendszerű irányítást valósághű szimulációs környezetben és kísérleti laboratóriumi hálózaton egyaránt teszteltük, továbbá az előrejelzési képességét a modellnek egy tényleges vízszolgáltatóval közösen végzett kísérleti projekt során hitelesítettük. A dolgozatban bemutatott módszereket kísérletileg igazoltuk, a hálózatban elhelyezett áramlásérzékelők használata nélkül. A klasszikus, szabályon alapuló irányítási stratégiákhoz képest (amelyet a vízművek a gyakorlatban jellemzően alkalmaznak), a jelen dolgozatban vizsgált szabályozási módszerekkel rendkívüli terhelés mellett 10 és 28 százalékkal alacsonyabb kumulatív áradási ráta volt mérhető az egy hónapig tartó vizsgálati időszak alatt.

Contents

Acknowledgments

I am grateful to Grundfos for the opportunity to conduct research for one of the pioneering companies in the field of hydraulics and optimization in water systems. Throughout this study, several people have joined my journey, to whom I must express gratitude for walking with me on the way and setting me on my current path.

I wish to express my gratitude to my advisors: Associate Professor Jan Dimon Bendtsen, Professor Carsten Skovmose Kallesøe, and Lead Business Dev. Manager Christian Schou, for the guidance they have provided throughout the past three years. I have been privileged to have so many advisors with so versatile expertise, each providing continued support and excellent guidance. I have grown both as a researcher and a person because of the enthusiasm and encouragement you consistently showed to me.

I would also like to thank my colleagues in the Controls department at Grundfos for their continued support and many insightful conversations. In particular, I am thankful for Hakon Børsting for providing all the possible help to conduct research in a large organization such as Grundfos; whether it was a travel to a conference, the purchase of an interesting book, or any sort of equipment needed for the success of this project.

I also want to recognize my peers at the Department of Electronic Systems at Aalborg University. Jorge Val Ledesma, Joakim Børlum Petersen, Mohammad Al Ahdab, Deividas Eringis, Henrik Glavind Clausen, Simon Thorsteinsson, Saruch Satishkumar Rathore, Chandramouli Santhanam, Yonghao Gui, and the entire staff at the Automation and Control section made the experience worthwhile. It was a wonderful experience to follow each of your research and get great ideas through fruitful discussions.

Furthere acknowledgment is given to the wastewater utilities Provas, Ishøj, and Fredericia for sharing the data with us and providing help and inspiration for our experimental tests.

Special recognition is given to Casper Houtved Knudsen and Adis Hodžić whom I supervised for a year during their M.Sc. project in very close collaboration with my study. Much of my work would not have been possible without your assistance, both in the laboratory and through your ideas and enthusiastic questions. (Often at 10. p.m. phone calls discussing math problems during the COVID-19 lockdown.)

While not officially noted as an advisor of this project, Carlos Ocampo-Martinez deserves an enormous amount of credit. Carlos has advised me for almost two years due to my delayed external stay as a consequence of the global pandemic. Luckily, in 2021 I had the privilege to visit the Institut de Robòtica i Informàtica Industrial at the Technical University of Catalonia in Barcelona. I would like to express my gratitude to Carlos for spending just as much time with me as his students and to make my stay prosperous. (Not to mention that my paper writing standards have increased significantly as a result of his eagle eyes.) Besides, I would like to acknowledge Shuang Zhang and Mohammad Sheikhsamad for making my visit not only fruitful but also fun.

Krisztian Mark Balla
Grundfos Holding A/S, Thursday 30$^{\text{th}}$ June, 2022

Part I

Summary

1 Background

In this chapter the introduction to waste- and stormwater networks is given. The chapter details the motivation, application background and the main objectives of the thesis.

1.1 Motivation

Throughout the past decades, governments, industries, and researchers have shown increasing interest to employ smart technologies, such as smart data acquisition, improved real-time monitoring, and decision-making in large-scale water systems [Li et al., 2020]. Due to the growing population, urbanization, and the changing weather conditions across the world, water has become a valuable asset while its scarcity is a significant issue for modern and developing societies. According to [UN DESA, 2019], the urban population of the world has grown from 1.019 billion in 1960 to 4.117 billion in 2017, and it is projected that it will reach 9.7 billion by 2050. As the demand for drinking water rises globally, so does the amount of discharged domestic wastewater. For instance, one of the most urbanized country in the world, China, has almost doubled its domestic wastewater discharge just over 14 years after the turn of the last century, as depicted in Figure 1.1[1].

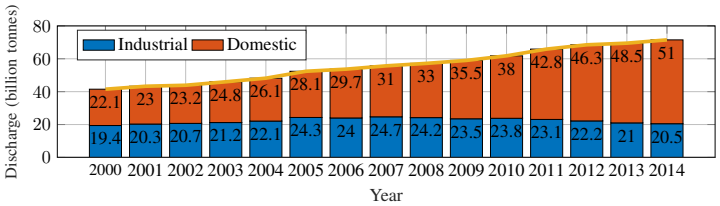


Figure 1.1: The evolution of the yearly domestic and industrial wastewater discharge over 14 years in the Republic of China. Data source: [Li et al., 2012].

One of the key drivers for the collection, transport, and treatment of discharged waste and stormwater is the emerging government regulations regarding the discharge of domestic and industrial wastewater to the environment. Even though regulations are tightening, the expansion and modernization of the sewer infrastructure do not follow the rapid urbanization growth [Henze and Arnbjerg-Nielsen, 2008]. For instance, many large cities still utilize century old piping infrastructure, meanwhile, urbanization and the excessive connections to the sewer net have grown to the extent that the load on the system surpasses its original design capacity [UNESCO World Water Assessment Programme, 2017].

[1]Despite the rising GDP and industrial activity, the industrial discharge is reported to be decreasing in China, indicating that the discharge is likely under-reported.

Of the US$271 billion in investments to meet the wastewater infrastructure needs of the United States, the expansion and correction of the current infrastructure account for 52% according to [US EPA, 2016]. As a consequence of the current situation, urban flooding and sewer overflows are common in large cities, and hence the discharge of untreated wastewater is a significant issue up until today, as depicted in Figure 1.2.

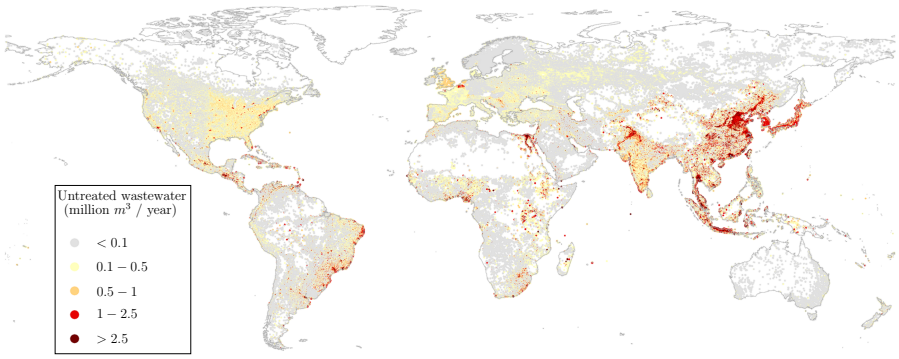


Figure 1.2: Untreated wastewater flows released to the environment based on a 5 arcmin resolution. Source: [Jones et al., 2021]

Besides the problem with the under dimensioned sewer infrastructure, another catalyst of the problem is the meteorological load through rain infiltrating into the systems [US EPA, 2021]. Figure 1.3 illustrates the change of land area affected by extreme single-day rain events over the past century.

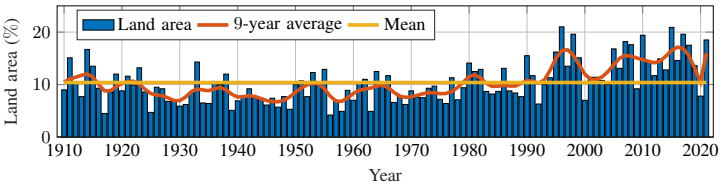


Figure 1.3: Percentage of land area in the United States where a large portion of annual precipitation has come from extreme single-day rain events. Data source: [NCEI, 2021]

According to [Kenward et al., 2016], climate models predict that the overall rain precipitation and the frequency of heavy or extreme rain events will continue to increase in certain parts of the world, which will increase the probability of sewer overflows.

Overflows in sewers often occur due to water surges and bottlenecks in the system. These bottlenecks are typically a consequence of the heavy infiltration, the out-dated design, and the lack of active control solutions to overcome the design issues. To handle the increased load on the existing infrastructure (without substantial re-design and expansion) and thereby utilize the in-sewer capacity in a better way, a solution might be to use advanced control methods, relying on real-time data and system-wide optimization.

1.2 Waste and Storm Water Networks

Open-channel hydraulic networks are systems where the water is transported through pipes with a free surface. Open-channel networks comprise several types of water systems in the water cycle, e.g., irrigation canals, rivers, lakes and sewer networks. In this thesis, our attention is mainly focused on sewer networks, however, the terms and characteristics discussed in the following can often resemble to processes in other open-channel applications.

The main purpose of sewer networks is to collect and transport the *sewage*[2] from the populated areas toward a treatment plant, where it is treated before the water is released back to the environment. The sewer network may also include stormwater collection, meaning that the rainwater runoff from roads, catchments, and roofs is collected together with the sewage. These networks together are called *combined sewer systems*, where the complexity of the operation is significantly increased compared to separated systems, in regards to the increased risks associated with incorrect network operation [Ocampo-Martinez, 2010]. In the thesis we deal with both types of systems, however, for generality, we refer to combined sewer networks in the rest of the study. An illustration of a simple combined sewer network scheme is depicted in Figure 1.4.

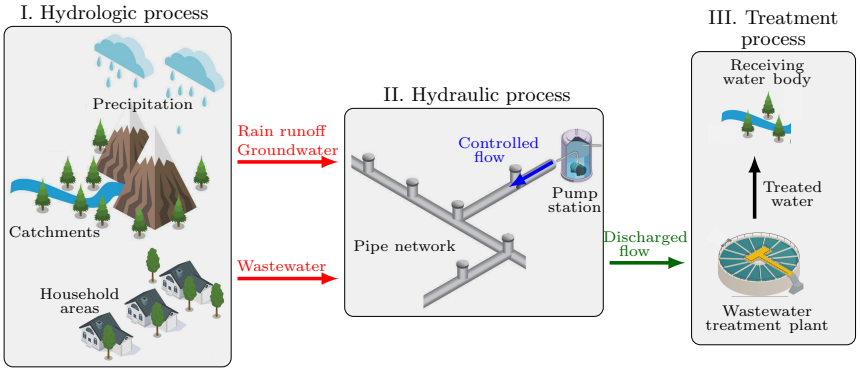


Figure 1.4: Wastewater collection and treatment process in combined sewer networks.

The illustration in Figure 1.4 points out the subprocesses and their inherent connection within the cycle of wastewater collection and treatment. Throughout the thesis, we adopt the terminology used in the field of control theory instead of the application-specific expressions. For instance, the exogenous flows (rain runoff, groundwater, wastewater) generated by the hydrologic processes provide a load on the hydraulic pipe network, which we call *disturbances*. These flows are then gravitated or mitigated utilizing controlled actuators to collect and eventually discharge to the root of the network, i.e., to the treatment plant. From a decision-making perspective, the outlet of the actuators is

[2]We define sewage as the liquid waste produced by human or industrial activity.

considered as *input* in the control problem. Moreover, the states we consider are restricted only to the hydraulic part of the collection process, typically characterized by the flow or level in hydraulic structures. The *output* of our control problem is the discharged flow provided to the treatment plant. Note that our consideration of the treatment processes is limited to the operational management goals, e.g., providing a smooth flow to the plant. Otherwise, the biological and chemical treatments are outside the scope of this study.

1.2.1 Hydrologic Process

Hydrologic processes typically comprise the rain running off catchments and roads. The infiltration of groundwater is a consequence of the slow precipitation and accumulation of the rain, hence groundwater is always handled by considering slow infiltration into the system. Moreover, the generation of sewage is due to human activity, however, we handle it together with the former two disturbances as they are the driving force of the hydraulic processes.

Both rainfall precipitation and urban catchments are spatially distributed to the extent that a precise description of the governing hydrologic processes is impossible. Generating disturbance forecasts governing the rain entering the sewers requires rainfall inputs, which require certain hydrologic instrumentation for the wastewater network. The most common way to measure rain intensity is through rain gauges and radars [Löwe et al., 2016]. Rain gauges are typically deployed in most sewer systems at several locations to capture local rain events. An illustration of a tipping bucket rain gauge is shown in Figure 1.5a. An alternative way to capture rain intensities is via radars using *nowcasting*[3], which can provide a lead time of up to 2 hours. Relating the forecasted rain intensities to the actual flow appearing in the sewer network is typically done through hydrologic modelling. Finding alternatives to hydrologic modelling is of major interest in the thesis.

1.2.2 Hydraulic Process

The hydraulic part of wastewater networks can be considered as a collection of subelements, where each part provides a specific function. The main part of sewers is the piping infrastructure, typically exhibiting a topology of an interconnected tree graph. Due to the extent and size of the piping network, the transport links are typically governed by significant time delays in terms of transport time between wastewater stations. Along the sewer lines, manholes are placed where disturbance flows enter the network. In between the main transport links, the wastewater stations are equipped with storage tanks and some actuators, in case there is active control available in the system. Otherwise, the sewage simply propagates by means of gravity towards the lowest elevation point. Most drainage and wastewater networks are driven by gravity, but actuation is needed in systems where the landscape is flat and therefore

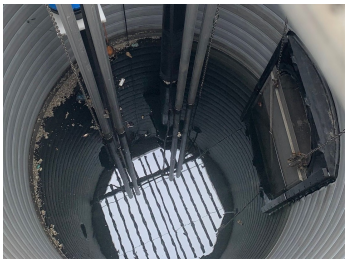
[3]Nowcasting is a form of weather forecast on a very short term period of up to two hours.

does not provide enough gravitation for the sewage. Examples for this are countries like Denmark and the Netherlands, where the sewage is transported with the help of pressurized pipelines, driven by the flow coming from pumping stations [van Heeringen et al., 2016], [van Nooijen and Kolechkina, 2018]. An illustration of a manhole, a pumping station and a wet pit with pumps is shown in Figure 1.5b, Figure 1.5c and Figure 1.5d, respectively.

(a) Tipping bucket rain gauge.

(b) Manhole (Ishøj, Denmark).

(c) Pumping station at the treatment plant(Biofos, Denmark).



(d) Pumps sitting in a wet pit (Ishøj, Denmark).

Figure 1.5: Instrumentation in wastewater networks.

Typical actuators in sewers are pumps, retention gates, weirs, and in some cases redirection gates [Ocampo-Martinez, 2010]. These assets aim to manipulate either the outflow or the downstream level in the system. Furthermore, wastewater networks are often equipped with retention tanks, where pumps sit in wet pits. These assets represent the controllable capacity in the system, storing or holding back water volumes. Besides, the most typical instrumentation of sewers includes flow and level sensors. Measuring flow in harsh environments such as sewers has a high maintenance cost and often provides inaccurate readings. Hence, velocity sensors are often utilized to enable more accurate flow computations.

1.2.3 Treatment Process

Treating the industrial wastewater and removing the pollutants is typically done by the treatment facility through physical, chemical, and biological processes. Figure 1.6a shows a treatment plant of a combined sewer network. The treatment process is of high importance, as it allows to release the treated

sewage to the receiving environment, e.g., lakes, rivers, and seas. An issue arises when the water provided to the treatment plant is combined. In situations when the rain load is high on the system, the treatment plant might get overwhelmed, furthermore the treating efficiency might degrade due to the dilution of the discharged flow [Schütze et al., 2002]. To overcome this issue, many networks are equipped with retention capabilities to smoothen the discharge flows. An illustration of a retention pond is shown in Figure 1.6b.

(a) Waste water treatment plant (Biofos, Denmark).

(b) Retention pond (Ishøj, Denmark).

Figure 1.6: Facilities in wastewater networks.

Designing control strategies to hold back volumes, avoid in-sewer overflows, and thereby avoid overloading the treatment plant is in focus in the thesis.

1.2.4 Real-time Sewer Optimization

The optimization problem solved behind any real-time sewer controller aims to improve the system performance in terms of avoiding overflows due to water surges, preventing discharges of combined overflows, utilizing the sewer storage capacity equally, and providing a smooth discharge to the treatment process. A typical closed-loop structure for sewer control is depicted in Figure 1.7.

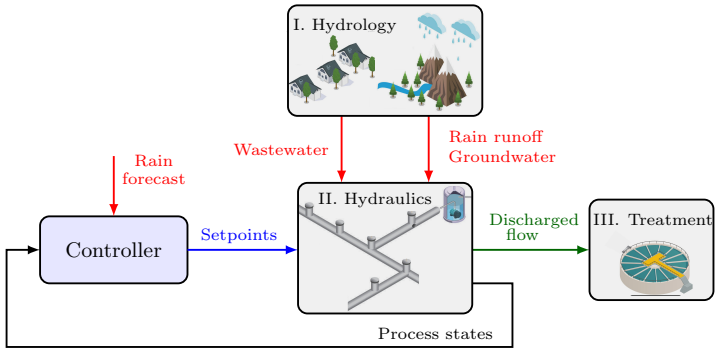


Figure 1.7: Automatic wastewater network optimization based on rain forecasts and state feedback. Red: disturbances, blue: input, green: output, black: states.

Having a controller that is easy to set into operation, as well as robust

in a real-world implementation is a significant criteria when it comes to the instrumentation and control of sewers. As shown in Figure 1.7, given the forecasts and state feedback, the ultimate goal of the optimization behind the controller is to provide the setpoints either to the personnel (decision-support) or directly to the controllable assets in the network (fully automated control).

This study has been initiated by Grundfos Holding A/S, whose vision is to provide digital solutions in terms of monitoring, decision-support, and eventually, automatic control solutions in wastewater networks. In industry, this vision is often termed as *Smart Water Management*. The thesis aims to enable a *Plug and Play* control paradigm that is easy to commission and can either support or fully automate the decision-making currently made by the utility operators. The specific objectives of the research are detailed in the following.

1.3 Research Objectives

The two primary objectives of the research is, on the one hand, to establish a real-time decision-making toolchain to minimize the effect of increased meteorological load on sewer networks. On the other hand, to reduce the cost of infrastructure expansion by better utilizing the available storage capacity [Balla et al., 2022d]. Specifically, our aim is to develop a real-time and data-driven modelling and control framework that is practically and economically feasible to implement for utility operators. To realize this vision, two hypotheses are formulated:

Hypothesis 1: *Structure preserving reduced models can represent the network dynamics such that the main sewer processes are captured, relying solely on in-sewer water level sensors, the topological network layout, and flow estimation data regarding pump operation from the pumping stations.*

A major difficulty with modelling large-scale rain and wastewater collection systems is to describe a large number of hydraulic elements without using many sensors and system parameters. Besides, pipe lengths, diameter, friction parameters, and bed slopes of the pipes are typically not available. Reduced and simplified models updated with a limited number of real-time sensor data should represent the key variables that describe the essential sewer processes for level and flow prediction in wastewater networks.

Hypothesis 2: *A physically-based yet data-driven model relying on the available system knowledge and the topological network layout can be used to predict the effect of hydraulic and hydrologic processes and their corresponding uncertainties in an integrated predictive controller.*

Model-based control approaches such as standard predictive control are well-suited techniques for the operational management of sewer networks. However, models behind the predictive controllers typically require well-maintained and properly calibrated network models, often economically out of reach for smaller

operators. Moreover, the prediction performance highly relies on the quality of disturbance forecasts and the dynamics capturing their effect on the system. Easy commissioning and the integrated handling of hydraulic and hydrologic processes thus have a high impact in practice.

The hypotheses lead to distinct research questions. The first two research questions are related to Hypothesis 1 while the latter two to Hypothesis 2, i.e,

Research Question 1: *Can we obtain easy-commissionable control models by only considering the main transport links, using a combination of simplified pipe models, graph theory, and the volume balance relations?*

Scalability is an issue in large-scale waste water networks, as the full-scale modelling is too complex in practice. Nevertheless, full-scale (high-fidelity) models often require prohibitively many sensors, meaning that their reliability is good for planning, however, less suitable for real-time control applications.

Research Question 2: *Can model parameters of reduced and conceptualized models be identified using only level sensor measurements, pump flow estimation, and information about the topological network structure?*

Keeping track of the volumes and flow transport in sewers is essential to optimize the network capacity. Level sensors are typically deployed in basins to monitor the level for rule-based pump control. However, we propose distributing additional level sensors inside the manholes along gravity sewer pipes. Learning the dynamics and the disturbances from the level variation at different locations in the system allows bypassing the level-to-flow calculations, often characterized by a high uncertainty and low accuracy in real-life applications.

Research Question 3: *Can standard predictive control be combined with learning models describing the interconnected process of hydraulic and hydrologic sewer dynamics using the rainfall intensities directly as input?*

Rain infiltrating into sewer networks and afterward the runoff flow propagating downstream in sewers is often seen as two individual modelling and control problems. Nevertheless, solving the sewer optimization problem is highly influenced by the non-linear and often season-dependent dynamics of the catchment runoff. The synthesis of the two problems has the potential to handle the direct effect of disturbances in the optimization behind the predictive controller.

Research Question 4: *Based on rain forecasts, can the in-sewer level measurements appropriately predict how the level variations in the storage elements evolve, e.g., in basins at pumping stations?*

Mass or volume balances in water networks are one of the most basic physical laws to describe storage transport and capacity. However, in the case of sewer networks, the volume balance of a storage tank is characterized not only by the actuators removing or adding volumes but by the infiltration of rain

and wastewater, making the actual volume balances quite uncertain towards the meteorological loads and forecasts. To overcome this issue, upstream level variations may be used to predict how the disturbances affect the volume balance in downstream tanks.

2 State of the Art

This chapter aims to provide an overview of the current state of the art and point out potential research gaps.

A *Plug and Play* solution for real-time control faces several challenges in wastewater networks, i.e., how well the network is instrumented, the quality of data, the modelling and physical attributes, and the complexity or computation demand of the control algorithm. In order to give a broad overview of these challenges, we summarize the state-of-the-art around the four main challenges illustrated in Figure 2.1.

(a) Modelling. **(b)** Instruments. **(c)** Data. **(d)** Control.

Figure 2.1: Four level of challenges in the real-time control of wastewater networks. Source: [Den Danske Vandklynge, 2022].

According to the review paper [Yuan et al., 2019], the above-mentioned groups of smart water system management do not frequently interact. Integrating the instrumentation and control of urban sewer systems offers a great potential to meet the growing complexity requirements of urban water management in cities of the future. In the following, we detail the four research areas respectively.

2.1 Instrumentation

In recent years, system-wide instrumentation of sewer networks has become a reality due to the rapid development of sensors, instruments, and the deployment of communication systems enabling real-time data acquisition. According to [Yuan et al., 2019], [Hill et al., 2014] and [Häck and Wiese, 2006], the instrumentation of large-scale water systems will enable to better utilize the controllable assets and defer the capital investment in expanding the existing infrastructure. [Häck and Wiese, 2006] also points out that the wide utilization

of software sensors[1] already available at the hands of utility operators will push further the development of setting advanced control algorithms into operation.

2.1.1 Physical Sensors

The instrumentation of sewer networks faces mainly practical challenges, e.g., the precision of the reading, and the optimal placement of sensors. For an overview of sewer sensors see [Yuan et al., 2019], and for a study on the potential and limitations of sensor deployments consult [Campisano et al., 2013].

Flow measurement in sewers is not trivial, as the flow in open channels is typically low, leading to large relative errors. As an example, [Aguilar et al., 2016] reported on an experimental test bench where the observed in-sewer flow discharge deviations were −56% to +62% under certain stormwater events, contrasting the nominal uncertainties provided by the sensor manufacturer of −10% to +10%.

Although the deployment of the sensors is a practical issue, there is some scientific work reported on choosing the optimal placement of sensors in sewer networks. For instance, [Mignot et al., 2012] reports on the optimal placement of flowmeters in local flow structures based on 3D simulations. Similarly, [Leitão et al., 2018] reports on an optimization method to support the decision of where to install flow control devices to prevent overflows in sewer networks.

2.1.2 Software Sensors

It is a general trend that researchers in the field of wastewater networks attempt to reconstruct the flow processes inside the sewer pipes to quantify the capacity utilization of the system. This is often done by using level, temperature, velocity, or conductivity sensors. The number of physical sensors is limited not only due to the large-scale nature of sewer networks but also because these sensors occasionally need maintenance [Bassø et al., 2020]. Maintenance in harsh environments such as sewers requires certified operators and preparation with equipment for several reasons, e.g., regulation and precautions before inspecting wastewater pipes. Flow sensors have direct contact with the fluid, meaning expensive and laborious work in regards of cleaning and maintenance. For a review on software sensors in the process industry, see [Kadlec et al., 2009].

Establishing a relation between level and flow is crucial to quantify volumes based on level measurements. For instance, [Ahm et al., 2016] proposed a software sensor to establish such relations in overflow structures, relying on computational fluid dynamics. Flow reconstruction based on the physical properties of sewer systems has been done in [Bassø et al., 2020], where velocity sensors have been used additionally to the level sensors to reduce the uncertainty of the estimated flows. For a similar idea, see also [Akgiray, 2004]. [Kallesøe and Knudsen, 2016] reports on a self-calibrating estimation of discharged gravity inflow to pumping stations, using the physical model of fixed-speed pumps and

[1] Equivalent to state observers in control theory.

level sensors in the wet pits. Similarly, [Chen et al., 2014] estimated the inflow to the pumping stations based on the on/off timing of the pumps.

In contrast with physically-based software sensor solutions, the development of data-driven techniques is also an emerging area. See for example [Zhang et al., 2018c], where conductivity sensors have been utilized to quantify rainfall-derived infiltration into sewers based on pollutographs. Long-short term memory neural networks have been used in [Palmitessa et al., 2021] to make a soft sensor for level tracking in overflow structures when observations are limited. See also [S.V. Lund et al., 2019] for runoff flow estimation via the assimilation of flow and level data.

Knowing the flow induced by the actively controlled assets, such as pumps, is also essential for real-time control applications. In [Kallesøe and Knudsen, 2016] and [Kallesøe and Eriksen, 2010], a robust approach for soft sensing pump flows has been developed based on the on/off operation of sewer pumps. In [Ahonen et al., 2010], the pump operational state has been estimated for supervisory purposes to maintain energy-efficient operation.

2.2 Data

With the increase of more sensors in water systems, it is often unclear which data is useful and how the data uncertainty affects predictions in real-time control. As [Hu et al., 2018] pointed out, more measurements do not necessarily improve the predictions, mainly because of the uncertainty introduced during the modelling. Similarly, [Kerkez et al., 2016] states that the performance of any real-time or data-driven control strategy is underpinned by the uncertainty related to rainfall forecasts or measurements, the modelling, and the sensor feeds. To overcome the issue with forecast uncertainties, [Jean et al., 2018] recommends data selection methods for the design of stormwater control in sewer networks. Additionally, [Molini et al., 2005] studies the impact of tipping-bucket rain gauge measurement errors and proposes a correction procedure for the underestimated rain intensity data sets.

Alternatives to the data selection and correction methods are data assimilation methods such as Kalman or Particle filtering, which is typically used in hydrologic modelling to account for data and modelling errors. For instance, [S.V. Lund et al., 2019] suggests using the Ensemble Kalman filter to update simplified hydrological models before forecasting, based on the most recent system observations to achieve precise runoff flows. Similar updating procedures are proposed in [Hutton et al., 2011] and [Refsgaard, 1997].

2.3 Modelling

Due to the large temporal and spatial scale of sewer networks, the combination of the three modelling disciplines (hydraulic, hydrologic, and quality-based treatment) is an extensive task, often termed integrated sewer modelling [Romero et al., 2021]. In the scientific literature, integrated modelling is an emerging area, whereas separated hydraulic, hydrologic, and treatment processes are well researched. Integrated modelling for combining hydrologic and hydraulic models is reported in [Luo et al., 2021], where a subdivision of three layers is proposed; a hydrologic modelling framework for the catchment runoffs, a hydraulic layer for suburbs, and a hydraulic layer for the city areas. For a review of quality-based integrated modelling see [Jia et al., 2021].

2.3.1 Hydraulics

Hydraulic modelling in sewers deals with the mechanical description of the fluid moving through the man-made hydrauliuc structures and conduits in the sewer network. This part of the sewer processes is of key importance due to the fact that it can be controlled by means of actuators. For a broad overview of hydraulic modelling in sewers, see [Chow, 2009].

Open-channel flow propagation in sewers is most accurately described via the Saint-Venant hyperbolic partial differential equations. These equations can describe the evolution of water level and flow in free-surface flow, however require precise information about the physical properties of the infrastructure, e.g., pipe geometry, length and slope. Besides, solving models based on the Saint-Venant equations for prediction in real-time is a computationally demanding and complex problem. In the scientific literature, the most common approach to tackle the computation and complexity burden of these equations is to apply approximations to them. For example in [Xu et al., 2012] and [Xu et al., 2011], the coupled partial differential equations are approximated by a linear state-space model and solved for control purposes. Similarly, [Schuurmans et al., 1995] and [Munier et al., 2008] assumed steady-state flow and identified operating points in open-channel canals to approximate the original equations with simple transfer functions. Nevertheless, the most simple approximation method is to assume kinematic waves in open channels. For an overview, see [Singh, 2001]. Another method is to assume diffusion waves to keep track of the water accumulation inside the pipes, i.e., the backwater. [Szymkiewicz and Gasiorowski, 2012] reports on the development of a simulation model based on the diffusion wave approximation and finite element methods, while [Duchesne et al., 2001] proposes a diffusion wave control model to support real-time decision making.

High-fidelity modelling in sewer networks is typically based on the spatial and time-discrete solution of the full Saint-Venant partial differential equations to simulate sewer flows and levels. Via using the available physical network attributes and measurement data for calibration, simulators such as the

`MIKE URBAN` collection system modeller [DHI, 2017], the `EPA SWMM` software tool [Niazi et al., 2017] and the `SWIFT` hydraulic package [Li et al., 2022] can be efficiently used to run offline simulations of large-scale wastewater systems. In recent years, several libraries have been developed to interact with these environments, including some interfacing layers with general-purpose programming languages such as the `MatSWMM` in MATLAB [Riaño-Briceño et al., 2016] and `PySWMM` in Python [McDonnell et al., 2020] to implement real-time control strategies in `EPA SWMM`. As an alternative to high-fidelity hydraulic simulators, [Bartos and Kerkez, 2021] reports on an interactive digital twin model for wastewater networks, where the framework not only includes the partial differential-based solvers but also provides a Kalman filtering scheme to efficiently update hydraulic states based on online data observations. Similarly, [Pedersen et al., 2021] reports on a digital twin modelling concept based on in-sewer level sensors to verify the performance of high-fidelity models and point out when and where simulation results are acceptable or differ from reality.

2.3.2 Hydrology

In the scientific literature, the hydrological input for real-time control studies is typically taken either by historical observations or generated by hydrologic models that map the rain to flow [Fletcher et al., 2013]. The input of rainfall forecasts are coming either from extrapolated radar rainfall estimates [Thorndahl et al., 2013], [Borup et al., 2016], [Thorndahl et al., 2017], or from numerical weather prediction models [Courdent et al., 2018]. The hydrologic model behind the input generation is often updated on historical rain gauge data, using simplified kinematic wave or unit hydrograph approximations of the runoff. For reference, see [Jean et al., 2018] and [Birkinshaw et al., 2021]. Machine learning techniques are also emerging in runoff modelling. For instance, [Liu et al., 2021] used statistical techniques, while [Dawson and Wilby, 2001] used artificial neural networks to relate rain to flow.

2.3.3 Conceptualizing

Conceptualizing the underlying sewer processes is often applied to construct models for control purposes. This results in loss of physical interpretation, however, allows the use of data for simpler identification purposes.

One of the most widely applied techniques is to collectively model the storage capacity of open-channel gravity pipes as interconnected storages. This method is often termed the virtual tank method and extensively utilized in [Ocampo-Martinez et al., 2013], [Ocampo-Martinez and Puig, 2009] and [Joseph-Duran et al., 2014b], for hybrid predictive control purposes. Besides, conceptual modelling is heavily used in the integrated (or quality-based) control of sewers. Due to the complexity of the biological and chemical processes in the transported sewage, conceptual pollutant transport and transformation models are typically coupled with the hydraulic model equations to consider the quality of

compounds. For reference, see the works in [Romero et al., 2021], [Sun et al., 2017] and [Guo et al., 2019]. Besides, [Petrucci and Tassin, 2015] reports on a conceptual flow-rate attenuation model in pipes, while [Wolfs et al., 2013] and [Wolfs et al., 2015] propose generic transfer functions for identification and automated recalibration. Furthermore, reduction by utilizing the sparsity of Saint-Venant based models is reported in [Xu et al., 2013], and the skeletonization of the network is proposed in [Thrysøe et al., 2019] via surrogate models.

2.3.4 Grey-box Methods

Large-scale sewer networks are complex systems, where the full physical description of the entire network is often too complicated to model accurately by first principles and white-box methods. Grey-box modelling has been applied widely in scientific research, where prior knowledge of the physical properties of a system can be encoded in a physically-based model or model structure. For an overview, consult [Rogers et al., 2017].

Grey-box modelling has been applied in sewer system modelling in [Thordarson et al., 2012] for probabilistic flow predictions, while the flow prediction problem is solved with stochastic differential equations assuming state-dependent diffusion in [Breinholt et al., 2011]. Additionally, [Bechmann et al., 2000] applied simple grey-box modelling for the first flush of wastewater to treatment plants. Other applications typically include large-scale systems, often described by spatially discretized partial differential equations. Applications fields include: household refrigerator systems in [Costanzo et al., 2013], district heating in [Nielsen and Madsen, 2006], pipe temperature transfer in [Kicsiny, 2017], and distillation process in [Bohlin, 1994].

White-box modelling, combined with Gaussian process regression is an active field of grey-box modelling research, where the residual between the nominal dynamics and the black box-box Gaussian process model is typically utilized to correct for errors. This approach retains the physical insight into the process, while also reducing the unmodelled errors. For a review on Gaussian processes, consult [Tulleken, 1993]. Gaussian process-based modelling has been done in water distribution networks for predicting water demands in [Wang et al., 2014] and [Wang et al., 2014]. Other fields of application include autonomous racing in [Hewing et al., 2020] and telescope positioning error correction in [Klenske et al., 2016].

2.3.5 Black-box methods

Due to the increased data availability and instrumentation, data-driven (or often termed black-box) techniques have gained popularity among the urban water systems community. For a general overview of machine learning techniques in wastewater networks consult [Granata and de Marinis, 2017]. Powerful tools, such as artificial neural networks are often applied to model wastewater flow

propagating inside the sewers [Sufi Karimi et al., 2019], the disturbances infiltrating into the network [El-Din and Smith, 2002], and generally to predict the risk and occurrence of overflows in critical points trained on historical rainfall data, as in [Zhang et al., 2018a], [Zhang et al., 2018b] and [Berkhahn et al., 2019]. Moreover, an emerging trend in black-box modelling of water systems is the generation of artificial data sets with the help of high-fidelity simulators, allowing to generate richer datasets, often not feasible to obtain through safety-critical real-life operations. Reports on this topic include [Zhang et al., 2018b], where a high-fidelity modelling environment has been used to update black-box models for predictions.

2.4 Control

Control of sewer networks is one of the most researched topics among the urban water society, mostly dominated by predictive optimal control. For a broad overview of real-time predictive control in sewers, see [Mollerup et al., 2016], [Ocampo-Martinez et al., 2013], [Schütze et al., 2004], [García et al., 2015], [Jacopin et al., 2001] and [van der Werf et al., 2021], while for case studies from real-world examples using high-fidelity simulations see [Gaborit et al., 2013] and [Cembrano et al., 2004].

2.4.1 Rule-based Control

Research in advanced predictive control is popular in the scientific literature, however reports on rule-based (or heuristic) control development are relatively few. In industrial applications, rule-based controllers are the most commonly implemented strategies for sewer networks. Although the performance of advanced global control is promising in real-world applications, the practical commissioning of such controllers is often unattractive to operators. See [Beeneken et al., 2013] for a review on the additional efforts of implementing real-time control solutions compared to rule-based controllers.

Event-driven local control algorithms of pressurized pumping stations are reported in a Dutch sewer network in [van Nooijen and Kolechkina, 2018], where a hierarchical control scheme is proposed to consider the behavior of the local event-driven pump stations. The study by [Fu et al., 2019] reports on a scenario-based approach, where hydraulic models are utilized to generate scenarios and thereby support decision-making to the utility operators. Similarly, [Klepiszewski and Schmitt, 2002] uses fuzzy logic in combined sewers with hydrodynamic and hydraulic simulators to set up a lookup table for admissible control rules. Tests over three storm events have shown that by adding new rules to the existing ones, the fuzzy-based controller does not provide operational advantages compared to the simple rule-based method. In [Poehler and Uwe Thamsen, 2017], a decentralized switching strategy for a network of five wet pit pumping stations is presented. An experimental implementation has

been proposed, which fits well into the existing instrumentation of a typical wet-pit pump control architecture. As opposed to on/off pump operation, in this decentralized framework the wet-pits communicate and exchange information about their water levels. Simulations testing with historical events have shown the improvement of the decentralized level-based decision-making mechanism.

2.4.2 Predictive Control

In the context of sewer network control, real-time control tools aim to prepare the sewer network for high-intensity rain or wastewater load to optimally utilize the maximum available storage capacity. From a control-theoretic perspective, predictive control has high relevance in sewers. In these frameworks, the future value of disturbances is typically given utilizing weather forecasts and an internal model is used to predict the states of the hydraulic elements in the network.

Several works in the scientific literature deal with solving the optimization of predictive control problems by using physically-based models. For instance, [Nielsen et al., 2020a], [Nielsen et al., 2020b], [Xu et al., 2011] and [Xu et al., 2012] use the simplified version of the well-known Saint-Venant partial differential equations to solve a linear convex optimization problem for the gravity-driven flow predictions in sewers. Nevertheless, [Schwanenberg et al., 2011] proposed a nonlinear predictive controller by using both an explicit and implicit time-stepping scheme of the full hyperbolic partial differential equations in large-scale river systems. To deal with the nonlinear and non-differentiable features of the Saint-Venant-based prediction model in a nonlinear predictive control problem, [Leirens et al., 2010] proposed a pattern search method to solve the optimization problem in sewer systems. Although solving the Saint-Venant equations to provide multiple-step predictions for the control problem is the most suitable way to incorporate nonlinear phenomena such as the wave attenuation, large time delays, and the backwater effect in sewers pipes, the computation demand is high and the modelling requires significantly high modelling expertise and data availability. The previously-mentioned features have been exploited in [Duchesne et al., 2004] and [Duchesne et al., 2003], where high-fidelity simulation models have been used together with a predictive controller to simulate surcharged flows and thereby exploit the capacity increase induced by the backwater in the sewer system pipes.

Hybrid predictive control has been proposed in [Puig et al., 2009], where the virtual tank approach has been adopted to conceptually model the entire sewer network as interconnected buffer tanks. The hybrid predictive control framework presented in [Joseph-Duran et al., 2014b] and [Joseph-Duran et al., 2014a] is capable of solving an optimization problem, where mainly redirection and retention gates are utilized as actuators. In these works, the combination of continuous and discrete type dynamics is considered, where the logical phenomena of overflow are triggered by redirection gates.

Nonetheless, most of the research articles reporting on predictive control

in sewer networks investigate the performance by considering historical disturbances, i.e., historical rain forecasts. Works on model-based optimization, where imperfect forecasts and the uncertainty governing both the modelling and predicting uncertainties are relatively few in sewer applications. Chance-constrained predictive control has been considered in [Svensen et al., 2021], where the uncertainty in weather forecasts has been used to propagate the uncertainty of the system states along the predictions. (For a review on stochastic model predictive control, see [Farina et al., 2016].) In [Courdent et al., 2015], the risk of overflows has been used in optimization based on a simplified sewer network model and the stored volume. [Grosso et al., 2014] reports on chance-constrained predictive control in drinking water systems by considering the uncertainty of the water demands. Besides, chance-constrained predictive control has been broadly used in other application fields; hydrogen-based microgrids in [Velarde et al., 2016], power grids in [Wang et al., 2019], and irrigation canals in [Nasir et al., 2019].

Another way to consider stochastic hydrological processes in optimization is to assume possible scenarios, estimate their probability and test the optimization under the different scenarios [Balla et al., 2020b]. A flow control problem has been studied in [Tian et al., 2017], where a multi-scenario approach has been implemented on a simulation model of a Dutch canal system [van Overloop et al., 2008]. In [Grosso et al., 2017], chance-constrained, tree-based, and multi-scenario stochastic MPC approaches have been compared and applied to drinking water networks. In [Tian et al., 2019], a multi-scenario model predictive control approach has been presented based on genetic algorithms for the regulation of water levels in a Dutch canal system. To overcome the computation issue due to the curse of dimensionality, tree-based predictive controllers have been proposed in [Raso et al., 2014] for level control in a collection reservoir, considering the hydrologic and forecasts uncertainties induced by sewage and rain infiltration. A similar approach is presented in [Velarde et al., 2019].

2.4.3 Learning Control

The available scientific literature on learning-based optimal control for large-scale water systems is sparse. Among the research community in control technology, learning-based or data-driven control techniques have emerged as a state-of-art technique for autonomous systems. For a review on Machine learning-based control in water resource systems, see [Labadie, 2014]. Reinforcement learning has shown some initial but promising results in canal systems to control gates for equal storage utilization in [Ochoa et al., 2019], while [Mullapudi et al., 2020] is the first reported scientific research in the practical application of deep reinforcement learning in stormwater networks. Deep learning with neural networks has also been reported in [El Ghazouli et al., 2022] and [Gudaparthi et al., 2020] for overflow prediction in sewers. Moreover, [Val Ledesma et al., 2021a] and [Val Ledesma et al., 2021b] report on real-time reinforcement learning for learning the unknown dynamics and water

demand in water distribution networks, supported with experimental results.

Iterative learning control has been widely used in applications where the control task is repetitive, hence the performance can be gradually improved by learning the results of the previous operations. Result in water distribution systems are available in [Jensen et al., 2018], while [Li et al., 2018] reports on the application of a dam-river channel irrigation system. Furthermore, [Cui et al., 2015] reports on learning rain events with similar return periods in combined sewer systems by utilizing iterative learning control. Additionally, [Grosso et al., 2013] reports on incorporating a learning and planning layer into a standard economic model predictive controller setup, where the demand forecasting for water distribution is done by neural networks. It also has self-tuning capabilities to tune the objectives in the control problem.

Learning the underlying dynamics to provide correction terms via the use of Gaussian processes has been an active field of research in the past two decades. For instance, [Wang et al., 2014] and [Wang et al., 2014] reports on the characterization of the uncertainty by learning the water demand in water distribution systems, while [Troutman et al., 2017] used the Gaussian processes to learn the infiltration flow dynamics in combined sewer systems. Predictive control with Gaussian processes has been widely used in other fields; autonomous racing is reported in [Kabzan et al., 2019] and [Hewing et al., 2020], telescope guiding to correct periodic errors in [Klenske et al., 2016], while trajectory tracking with robotic arms in [Carron et al., 2019].

2.5 Contributions

The application of predictive control in combined waste- and stormwater networks has received high academic attention. Nevertheless, apart from the large number of simulation studies with high-fidelity models, the real-time and industrial implementation of control methodologies are very few, making it difficult to evaluate the robustness and feasibility of the proposed methods in the state-of-the-art. Based on the state-of-the-art, the presented research in this thesis mainly focuses on the research gaps dealing with: the assessment of uncertainty governing the disturbances in forecasts, assimilating the data in both the hydraulic and hydrologic modelling, and the practical implementation of the control methods.

The main contributions are outlined via the associated scientific papers published throughout the project. To navigate between the contributions, a dependency chart is provided in Figure 2.2. All papers are provided in Part II.

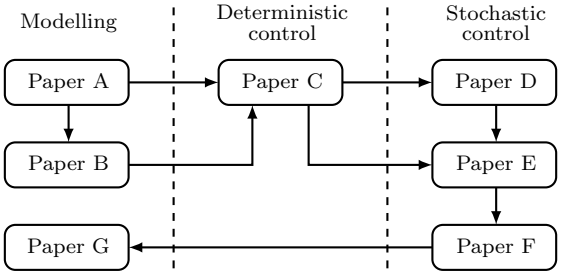


Figure 2.2: Dependency graph of the paper contributions grouped by research area.

Moreover, an overview is provided for each paper, summing up the topics of the content in Table 2.1.

	Modelling	Control	Simulation	Experiment	Real world
Paper A	✓		✓		✓
Paper B	✓			✓	
Paper C	✓	✓	✓		
Paper D		✓	✓		
Paper E	✓	✓		✓	
Paper F	✓	✓		✓	
Paper G	✓				✓

Table 2.1: Overview of the contributions by the topic of content.

Given the industrial application of the thesis, most of the contributions are dealing either with an experimental or real-world case study.

Paper A

K. M. Balla, C. S. Kallesøe, C. Schou, and J. D. Bendtsen. "Nonlinear grey-box identification with inflow decoupling in gravity sewers". *IFAC-PapersOnLine*, 53(2):1065–1070, 2020a. ISSN 24058963. doi:10.1016/j.ifacol.2020.12.1295.

Paper A is our first contribution investigating the grey-box modelling of gravity-driven sewage transport to identify the delays and the attenuation of the water flow. In this work, the downstream discharges are predicted coming from upstream pumping stations by decoupling the periodic wastewater and groundwater infiltration under dry-weather conditions. For this purpose, the physically-based Saint-Venant partial differential equations are reduced to a simplified form (Kinematic Wave approximation), where only four lumped parameters are used through the model fitting. Furthermore, simple Fourier series are used to capture the periodic effect of the disturbances.

To test our approach, two studies are presented. A numerical study on a simulation network with both wastewater and groundwater disturbance, and a real-life case study on data from the wastewater network in Gram, Denmark.

The model fitting indicates that there is an optimal number of sensors (states in the model) for which the approach is providing the highest accuracy.

Paper B

K. M. Balla, C. H. Knudsen, A. Hodzic, J. D. Bendtsen, and C. S. Kallesøe. "Nonlinear Grey-box Identification of Sewer Networks with the Backwater Effect: An Experimental Study". In *2021 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1202–1207, San Diego, 2021. IEEE. doi:10.1109/CCTA48906.2021.9658864.

Paper B extends Paper A by introducing an improved Kinematic Wave model from Paper A and a new model based on the Diffusion Wave approximation of the Saint-Venant equations. To this end, two system identification problems are presented using level sensors distributed along the sewer lines. Our motivation behind the more complex approximation with the Diffusion Wave model is to investigate whether the estimated model via level sensor data is capable of predicting a typical phenomenon in sewer networks: the backwater effect[2].

To compare the two approaches, data is gathered from an experimental setup representing a scaled version of a typical wastewater network topology.

The experimental comparison shows the feasibility of both approaches, furthermore the capability of the Diffusion Wave model to predict the accumulation of disturbances infiltrating in sewer pipes. We also highlight the drawbacks of the increased complexity coming with the extended Diffusion Wave model.

[2]Modelling of the backwater effect has the potential to utilize the volume capacity of pipes as extra storage under heavy loads when tanks cannot efficiently mitigate the overflows.

Paper C

K. M. Balla, C. Schou, J. D. Bendtsen, C. Ocampo-Martínez, and C. S. Kallesøe. "A Nonlinear Predictive Control Approach for Urban Drainage Networks Using Data-Driven Models and Moving Horizon Estimation". *IEEE Transactions on Control Systems Technology (Early Access)*, pages 1–16, 2022d. doi:10.1109/TCST.2021.3137712.

Paper C is building on the results obtained in Paper A and Paper B. In this article, we use the previous models in a predictive control framework. As pointed out in Paper A, the modelling of level variations in manholes through the coupled partial differential equations model requires as many level sensors, as many system states are used in our model. This inherently means expensive installation and maintenance costs. For that reason, we formulate joint parameter and state estimations via a moving horizon approach to obtain the unmeasured states and network parameters in real time.

To verify our results, we use a numerical study using a high-fidelity simulation model of a branch network with four interconnected pumping stations. To provide disturbances to our control problem, we use the catchment runoff dynamics of the simulator providing real rain and wastewater data.

The results obtained through the simulations show the dominance of our method compared to currently applied rule-based controllers, resulting in a 28% accumulative overflow decrease over the considered period. In this paper, we also report on the scalability and robustness of uncertain rain and wastewater flow infiltrations. The approach scales almost linearly with the number of pumping stations and the uncertain forecasts do not compromise the closed-loop performance significantly. This robustness is partially due to the physical guarantees that we can provide in both the moving horizon system identification and control, relying on the models developed in Paper A and Paper B.

Paper D

K. M. Balla, C. Schou, J. Dimon Bendtsen, and C. S. Kallesøe. "Multi-scenario model predictive control of combined sewer overflows in urban drainage networks". In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1042–1047, Montréal, 2020b. IEEE. ISBN 9781728171401. doi:10.1109/CCTA41146.2020.9206362.

Paper D is the first attempt to characterize the uncertainty due to the disturbances, e.g., household wastewater discharge and rain precipitation in wastewater networks. Consequently, a simplified model is presented where the transport delays are known. We adapt a multi-scenario approach where multiple scenarios are generated based on the uncertainty of rain forecasts.

The optimization is formulated as a multi-objective convex problem and we use a high-fidelity simulator to evaluate our results based on real forecast data.

The numerical results show that the controller can avoid worst-case situations, typically not considered in a standard predictive control setup. However, the approach is quite conservative when an improbable no-rain scenario is forecasted, restricting the control action while more likely scenarios would require emptying the retention tanks. The paper also highlights the scalability issues which arise when several forecast scenarios are considered in networks with several pumping stations.

Paper E

K. M. Balla, D. Eringis, M. Al Ahdab, J. D. Bendtsen, C. S. Kallesøe, and C. Ocampo-Martínez. "Learning-Based Predictive Control with Gaussian Processes:An Application to Urban Drainage Networks". In *2022 American Control Conference (ACC)*, pages 1–7, Atlanta, 2022b. IEEE.

Paper E is the first publication where the nonlinear effect of rain, wastewater infiltration, and the transport dynamics are combined by a non-physical model. We integrate the known dynamics, i.e., pumping and tank storage, with an additive, unknown part modelled via Gaussian processes. The Gaussian processes provide the residual uncertainties trained with the level sensors and capture the effect of the load on the transport dynamics.

The controller is implemented on an experimental setup that resembles the topology of a real network. We assess the closed-loop performance of the controller through a reference tracking problem, where pumping stations aim to track certain water levels in storage tanks.

The experimental tests show that the controller can recognize the uncertainty over the flow forecasts, hence the tracking of the reference signal is often violated in case the chance for overflow increases.

Paper F

K. M. Balla, J. D. Bendtsen, C. S. Kallesøe, and C. Ocampo-Martínez. "A learning-based approach towards the data-driven predictive control of combined wastewater networks – An experimental study". *Water Research*, pages 1–15, 2022a. doi:10.1016/j.watres.2022.118782.

Paper F provides an extensive study solving the shortcomings of the initial experimental work in Paper E. Similarly to Paper E, we propose to learn the effect of dry- and wet-weather flows through the variation of water level sensors. However, we extend the modelling by providing time as an input to learn the inherent periodicity and dynamic behavior of the dry-weather wastewater flows. Thereby, it becomes possible to provide only the rain intensities as the forecasts to the Gaussian process-based controller. The optimization behind the controller is extended and reformulated to a disturbance rejection problem, where safety bounds in storage elements are utilized to allow the controller to find solutions without any penalties within the safety range. In this work, we also provide key performance indicators to support the idea of a practical implementation as a decision-support tool.

To make a fair evaluation, the controller is implemented on an experimental setup representing a segment of the wastewater utility in Gram, Denmark.

Overall, the system identification and the closed-loop performance show that the wastewater loads and the temporal dynamics of the network are learned well. The controller significantly outperforms the currently deployed rule-based setup, especially during high-intensity rainfalls.

Paper G

K. M. Balla, H. Lemée, C. Schou, and C. S. Kallesøe. "A Toolchain for the Data-driven Decision Support in Waste Water Networks – A Level-based Approach". In *2022 International Water Association (IWA) World Water Congress and Exhibition*, pages 1–4, Copenhagen, 2022c. IWA.

Paper G is a case study, relying on the results provided in Paper D and Paper E. In this work, the stochastic modelling tool based on Gaussian process regression has been tested on a stormwater network in the municipality of Ishøj, Denmark. Five level sensors have been deployed in the network for a period spanning five months (between 16. June 2020 to 27. October 2020) to investigate the effect of rain infiltrating into the system through the level variations in the pipes.

From the collected level and rain datasets, 12 events have been chosen to evaluate the prediction capabilities of the proposed approach. The current solution can predict the levels and the corresponding uncertainty reasonably well for one hour ahead via the level sensors and local rain intensity feeds[3].

[3]The rain intensity used as a forecast is the rain intensity measured over the considered period.

3 Case Studies

The methods used in the thesis have been verified through simulational, experimental, and real-world case studies. This chapter provides an overview of the layout and configuration of all wastewater networks used throughout the study.

3.1 Simulation Studies

Simulation environments allow the use of high-fidelity models and simulate both hydraulic and hydrologic processes in sewer systems, from the drop of rain to the flow leaving the sewer pipes. In our tests, we simulated the runoff due to rainfall and the evolution of flows, volumes, and water levels in hydraulic structures by solving the full Saint-Venant equations. In all simulation environments, the entire network is defined by the true physical parameters, hence we considered these tools as virtual reality for our control and modelling tests.

3.1.1 Simulation Study A

Simulation Study A comprises of a network artificially created in the high-fidelity simulation software `EPA SWMM`, as depicted in Figure 3.1.

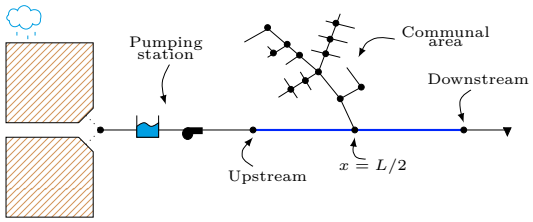


Figure 3.1: Conceptual combined wastewater network simulation model in `EPA SWMM`. Source:[Balla et al., 2020a].

In this network, a single sewer line has been created transporting the sewage from an upstream pumping station to an outlet point through a 7 kilometer-long gravity sewer pipe. Besides, a communal area discharges to the main transport line in the middle of the pipeline. The data generated for the tests are based on the threshold-based operation of the pumps at the pumping station. We attempted to mimic a realistic scenario, where the wastewater pit collects the runoff water induced by a variety of rain events, thereby forcing the pump to turn on for different time durations. To create a realistic scenario, the gravity sewer line is designed such that there are several sections of pipes with different slopes and different diameters. Further details and dimensions of the network are given in Paper A.

3.1.2 Simulation Study B

Simulation Study B comprises of a network in the `MIKE Urban` simulation software and interfaced with the `MIKE1D` [DHI] Application Programming Interface (API). The system is depicted in Figure 3.2.

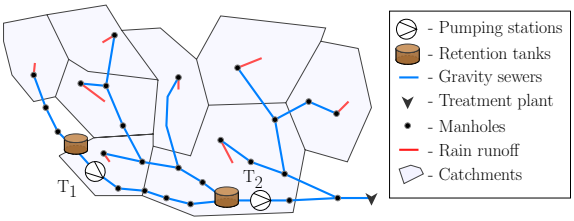


Figure 3.2: Conceptual combined wastewater network simulation model in `MIKE Urban`. Source:[Balla et al., 2020b]

The network consists of two pumping stations T_1 and T_2, equipped with retention tanks of storage capacity 30 (m^3). The pumps are operated by local PID controllers and references are given to the actuators from outside through the API. There is one outlet point in the network, furthermore several catchment areas discharging to the pipe network through manholes. Further details and dimensions of the high-fidelity model are given in Paper D.

3.1.3 Simulation Study C

Simulation Study C is an extended network, artificially created in the high-fidelity simulation software `MIKE Urban` and interfaced with the `MIKE1D` API. The system is depicted in Figure 3.3.

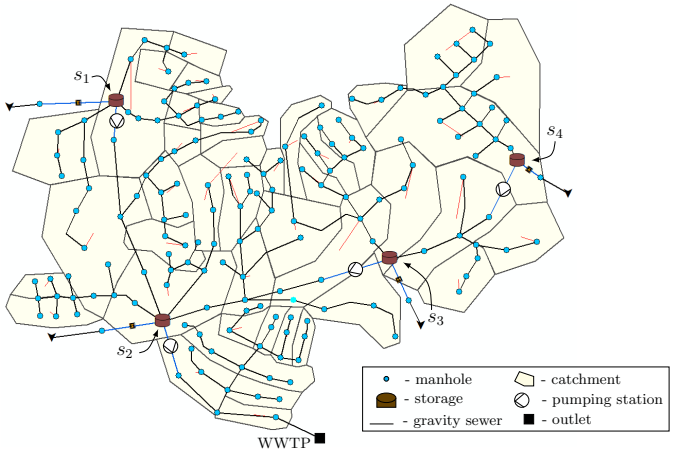


Figure 3.3: Conceptual combined wastewater network simulation model in `MIKE Urban`. Source:[Balla et al., 2022d]

The network consists of four pumping stations, where the collected sewage is transported to $\mathcal{S}_2$ pumping station through two branches; from station $\mathcal{S}_1$ and from the interconnected line of $\mathcal{S}_3$, $\mathcal{S}_4$. The network consists of 170 manholes, 170 segments of gravity pipes and three wet pit at the pumping stations, whereas at the station $\mathcal{S}_2$ there is a retention tank to hold back volume from discharging downstream. Similarly to Simulation Study B, flow and level sensors are read through virtual sensors and control inputs are provided to the pumps through the `MIKE1D` API to provide flow reference to local PID controllers. Further details and dimensions of the network are given in Paper C.

3.2 Experimental Study

The Experimental Study has been conducted on a modular test facility that can be configured to emulate the hydraulic transport processes in wastewater collection systems. The setup and the schematics of the network configuration are shown in Figure 3.4.

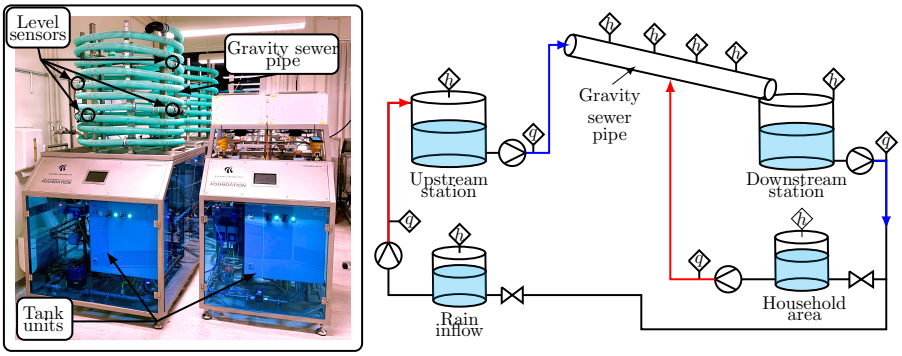


Figure 3.4: Experimental sewer network setup of the Smart Water Infrastructures Laboratory (SWIL) at Aalborg University. Left: Pumping station and pipe modules in the laboratory. Right: Schematics of the network topology. Source: [Balla et al., 2022b], [Balla et al., 2022a]

Throughout the thesis, the laboratory setup enabled us to prototype control algorithms without the risk of compromising the operation of real-world infrastructure. The setup replicates a segment of a combined sewer network, where sewage is transported from an upstream pumping station to a downstream station through a gravity pipe. The disturbances enter the system at two locations: combined rain and household sewage enters at the upstream station, while household sewage is entering in the middle of the gravity pipe. The instrumentation of the network consists of level sensors distributed along the pipelines and inside the storage tanks. Moreover, flow sensors are placed at each pumping unit.

The experiments carried out on a setup represent a $1:80$ scale reality of a wastewater network, meaning that the typical resolution and time scale

of the disturbances, i.e., the one day periodicity of household discharge, are scaled down accordingly. Besides, the sensor measurements are obtained and locally managed at each laboratory unit with a `Codesys` [3S-Smart Software Solutions GmbH] soft-PLC in real time. The data acquisition is done at every 0.5 seconds. The detailed description of the experimental setup is given in [Val Ledesma et al., 2021]. The results of the research verified on this setup are detailed in Paper B, Paper E and Paper F.

3.3 Field Study

The field studies conducted in the thesis are restricted to the analysis and modelling based on the collected sensor data from real-world sewer operation. This is partly because of the safety-critical operation and the lack of controllable assets available in both storm- and wastewater networks included in our study.

3.3.1 Haderslev Case Study

The Haderslev Case Study is a combined sewer network, operated by the Provas Wastewater Utility in Haderslev, Denmark. The network in consideration comprises several wet pits equipped with pumping stations to collect and transport the sewage from the communal areas to a treatment plant outside the city. The layout of the network is depicted in Figure 3.5 below.

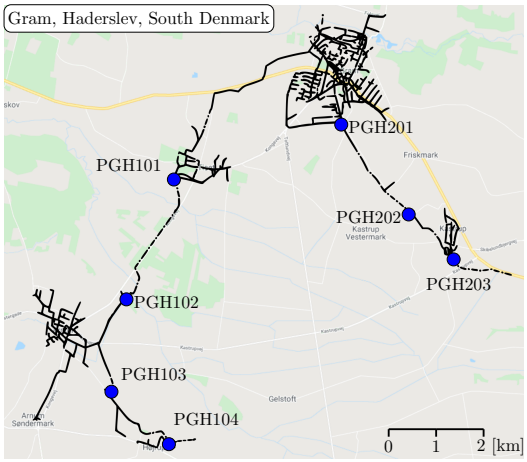


Figure 3.5: A segment of a combined wastewater network in Gram, Haderslev, South Denmark, operated by the Provas Wastewater Utility. Source: [Balla et al., 2020a]

This particular segment of the network consists of seven stations, operated by threshold-based pumping rules, locally managed by the company Provas. The data collection from this network has been carried out through a collaboration between Provas Wastewater Utility and Grundfos Holding A/S. Furthermore, Grundfos has deployed a flow estimation algorithm for real-time

estimation of both the flow provided by the pumps and the discharged flows downstream. As depicted in Figure 3.5, the discharged flows are propagated downstream via open-surface flow. Detailed description of the data and the network is found in Paper A.

3.3.2 Ishøj Case Study

The Ishøj Case Study has been carried out in a pilot project under the collaboration of Grundfos Holding A/S[1], Aalborg University[2], and Ishøj Spildevand[3] in Ishøj Denmark. The network in consideration is a stormwater system, operated by the Ishøj Wastewater Utility. The main operation task of the network is to collect the rain runoff and transport the water volumes towards the sea. The network is depicted in Figure 3.6.

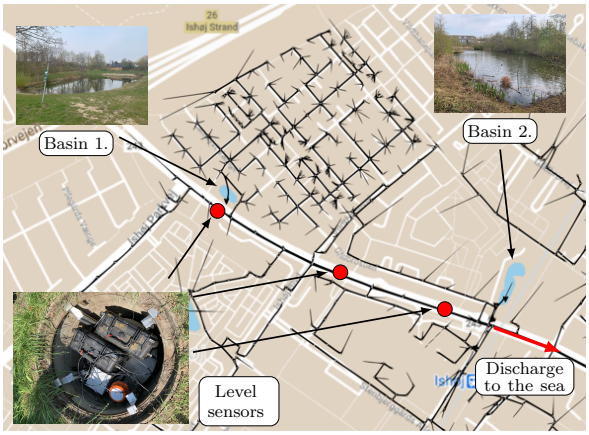


Figure 3.6: A segment of a stormwater network in Ishøj, Denmark, operated by the Ishøj Wastewater Utility. Source: [Balla et al., 2022c].

As seen in Figure 3.6, the network consists of two ponds with high volume capacity both upstream and downstream. These ponds are connected by a main transport line running below the main road called "Stationsvej". Through a period spanning five months, from the 16^{th} of June 2020 to the 27^{th} of October 2020, five level sensors have been deployed, providing 30 seconds measurement resolution to collect in-sewer level data. 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. Detailed description of the data and the case study is provided in Paper G.

[1]Represented by Christian Schou, Carsten Skovmose Kallesøe and Krisztian Mark Balla
[2]Represented by Carsten Skovmose Kallesøe and Krisztian Mark Balla
[3]Represented by Henrik Lemée

3.4 Benchmark Controller

In this work, we build upon the experience collected in the review paper [Lund et al., 2018] for predictive control in urban sewer networks. According to [Lund et al., 2018], the current state-of-the-art typically compares controller performance to rule-based control rules used by the practitioners. Hence, the most common baseline controller is a simple rule based on the on/off switching of the actuators according to the water levels in storage elements. The switching can either be done via a pump turning to full speed and shutting down or a gate opening and closing. In this thesis, we restrict ourselves to pumps as actuators for the baseline control, however, gate operation can equivalently be used for the benchmark.

The switching rule of a pumping station along with the aggregated flow provided by the pumps under the rule-based control is given by

$$Q(t_k) = \begin{cases} \overline{Q}, & \text{if} \quad h(t_k) \geq \overline{h}, \quad \forall t_k, \\ \underline{Q}, & \text{if} \quad h(t_k) \leq \underline{h}, \quad \forall t_k, \\ Q(t_{k-1}), & \text{otherwise}, \quad \forall t_k, \end{cases} \tag{3.1}$$

where $h(t_k)$ is the measured water level in the storage element. Upper and lower bounds of the inlet flow $\overline{Q}, \underline{Q}$ correspond to the maximum and minimum flow capacity of the pumps [Balla et al., 2022d]. Threshold values $\overline{h}$, $\underline{h}$ define the maximum and minimum allowed water levels in the storage element, respectively.

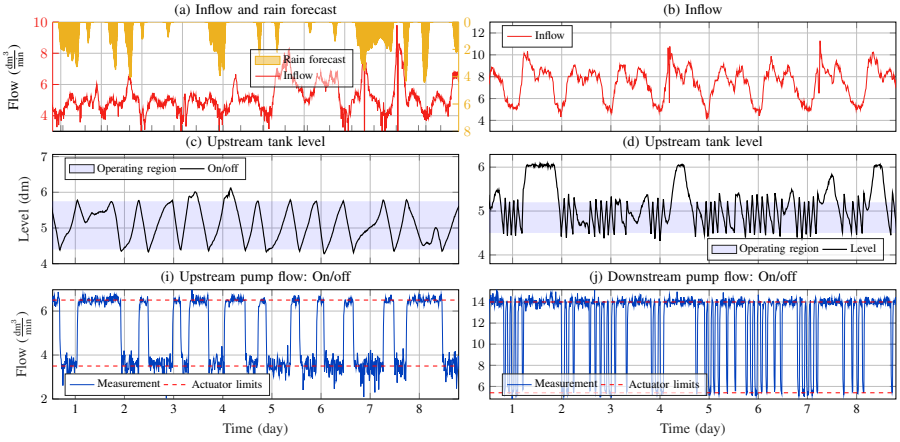


Figure 3.7: Rule-based control implemented in real-time on the Experimental Study setup. Source: [Balla et al., 2022a]

As depicted in Figure 3.7, the pumps turn on and off according to the levels reaching the higher and lower bounds of the operating region. Violation of the upper bound indicates high load on the system, hence the longer pump cycles.

4 Sewer Dynamics Modelling

This chapter provides an overview of the modelling approaches introduced in the research papers.

4.1 Modelling Goals

The approaches proposed in this chapter are developed to obtain network models incorporating only physically measurable water levels as the system states. Consequently, the overall concept of the presented approaches relies on deploying in-sewer level sensors at critical locations in the sewer network. Our modelling purpose is to consider both the temporal and spatial extents of the water network while using the simplified topology to describe the piping layout. Specifically, we utilize the full scale of the network by modelling only the main transport lines between the actuators and the main hydraulic elements. The overall requirements for the modelling developed throughout the thesis can be summarized as follows

- The network topology of the control model is obtained directly from the high-level pipe layout of the real network, considering the hydraulic elements forming a tree-based graph structure,
- Our main purpose with the level sensors is to capture the effect of disturbances through the variation of level signals, hence the sensors are placed at the vicinity of pipes joining the main transport lines,
- The collected level observations and the estimated flow data are aimed for online or offline system parameter identification with the proposed model structures,
- The level-to-flow conversion is either partly or fully bypassed by establishing relation between the in-sewer level states and the level variation in hydraulic storage elements.

In the following, we summarize the main modelling approaches detailed in the published research papers included in this thesis. These approaches are categorized into two groups, namely; parametric (or physically-based) and non-parametric (or conceptual) modelling techniques. The parametric modelling is detailed in Section 4.2 (Parametric Modelling of Sewer Hydraulics) and mainly concerns the hydraulic processes, while the non-parametric modelling is detailed in Section 4.4 (Non-parametric Modelling of Sewer Hydrology) and concerns the combination of hydraulic and hydrologic processes. Note that the presented methods are developed to enable system identification where data is used to find the required model parameters.

4.2 Parametric Modelling of Sewer Hydraulics

In this section, we propose modelling methods, assuming that the mathematical structure of the underlying model is partly known. The parameters of these models are optimised to get the best fit of the input-output dynamics. These approaches are often termed grey-box, for the reason that both theoretical and experimental considerations are taken into account when building the final models for prediction and control. Detailed description of the methods and outcomes regarding parametric modelling are provided in Paper A, Paper B, Paper C and Paper D.

4.2.1 Volume Conservation in Tanks

Storage of wastewater is represented by storage tanks or wet pits, among which some tanks are also equipped with a retention area to hold back a large amount of waste and rainwater in high-load periods. An illustration of such structure is shown in Figure 4.1.

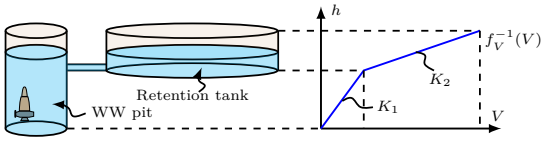


Figure 4.1: Level to volume conversion for a waste water tank equipped with retention area. Source: [Balla et al., 2022d]

Hydraulic structures describing storage are typically linear tanks, setting relation between the volumetric flow to the level change. In the case of hydraulic structures where retention spaces are present, the non-linearity between level and flow is handled by the following assumption.

Assumption 1. *Nonlinear storage elements are approximated with a piecewise linear, strictly monotonic increasing function $f_V(\cdot)$, parameterized by the level to flow constant of the tanks, i.e., the area of the tanks [Balla et al., 2020b].*

Then the infinitesimal level change in storage elements is simply described by the volume conservation, given by eq.(4.1).

$$\frac{df_V(h(t))}{dt} = q(t) - Q_0(t), \tag{4.1}$$

where $f_V(\cdot)$ is defined by the size of the tanks. Besides, $Q_0(t)$ is the outflow controlled by actuators, and $q(t)$ is the inflow to the storage element. Note that soft sensing techniques exist (see [Kallesøe and Knudsen, 2016]) to estimate both in and outflows from pumping stations in real-time. However, predicting the inflow to hydraulic elements is highly influenced by the stochastic nature of the disturbances, i.e., the rain and wastewater infiltration.

4.2.2 Volume Conservation in Pipes

The free-surface motion of flow in wastewater networks is based on the 1-dimensional Saint-Venant hyperbolic Partial Differential Equations (PDEs) for small channel slope assuming constant cross-sectional area. These equations relate to the water level and the flow propagating in open channels.

4.2.2.1 Saint-Venant Partial Differential Equations

The Saint-Venant PDEs describe both the volume and momentum conservation of the fluid, given by eq.(4.2a) and eq.(4.2b), respectively.

$$\frac{\partial A(x,t)}{\partial t} + \frac{\partial Q(x,t)}{\partial x} = \tilde{q}(x,t), \tag{4.2a}$$

$$\underbrace{\frac{1}{gA(x,t)}\left(\frac{\partial Q(x,t)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q(x,t)^2}{A(x,t)}\right)\right) + \underbrace{\frac{\partial h(x,t)}{\partial x} + \overbrace{S_f(x,t) - S_b}^{\text{Kinematic wave}}}_{\text{Diffusion wave}}}_{\text{Dynamic wave}} = 0, \tag{4.2b}$$

where $Q(x,t)$ is the flow propagating inside the pipe and $\tilde{q}(x,t) = q(x,t)/\delta x$ represents the lateral inflows per unit length, where $q(x,t)$ is the lateral inflow, i.e., the disturbance. $A(x,t)$ is the wetted pipe area, $h(x,t)$ represents the water level, furthermore $Q(x,t), q(x,t), A(x,t)$ and $h(x,t)$ are functions from $(0,L) \times \mathbb{R}_+ \to \mathbb{R}_+$, where L is the total length of the gravity pipe [Balla et al., 2020a]. The gravitational acceleration is denoted with g. Moreover, the slope term $S_b \in \mathbb{R}_+$ and friction term $S_f \in \mathbb{R}_+$ are considered to be independent of x and t, i.e., all pipe segments along the gravity pipe are modelled with assuming identical physical attributes [Balla et al., 2020a].

As seen in eq.(4.2), the independent variables describing the full dynamic motion of fluid are the water level $h(x,t)$, the flow inside the channel $Q(x,t)$ and the wetted area $A(x,t)$. Using these variables without a very precise physical parameter description of the network is unrealistic in practice, as the geometry and the boundary conditions coupling these equations are spatially varying. Hence, the problem of using the full dynamic wave equation of the Saint-Venant PDEs for prediction is computationally demanding and requires heavy commissioning and calibration in real-time control [Balla et al., 2020b].

4.2.3 Model simplifications

To overcome the issues with the complex flow routing proposed by the full dynamic Saint-Venant PDEs, several simplifications of the original model are proposed in the literature [Schütze et al., 2002]. The type of simplification depends on which terms are neglected in the momentum equation in eq.(4.2b). As indicated in eq.(4.2b), we consider two approximations, namely the Kinematic wave (KW) and Diffusion wave (DW) approaches. By approximating

the original physically-based PDEs, a trade-off is made by neglecting physical phenomena to the benefit of obtaining a simpler model structure and better suitability for system identification. The comparison of the proposed approximations is shown in Table 4.1.

	Kinematic wave	Diffusion wave	Dynamic wave
Backwater effect	✗	✓	✓
Wave attenuation	✗	✓	✓
Flow acceleration	✗	✗	✓

Table 4.1: Physical phenomena described by the Saint-Venant partial differential equations describing open-channel flow, and their approximations. Source: [Schütze et al., 2002]

Although the KW approximation of the full dynamic Saint-Venant PDEs neglects the wave attenuation phenomena, artificial distortion is obtained via the time and spatial discretization of the model. In this way, the wave attenuation can be numerically recreated.

To obtain a simple model structure for both the KW- and DW-based model approximations, further hydraulic assumptions are introduced. These assumptions regard the geometry of the pipelines and the relation between the independent variables in the original full dynamic equations.

Assumption 2. *A linear relation between water level $h(x,t)$ and wetted pipe area $A(x,t)$ is assumed. Semi-filled pipe segments regardless of their geometry are approximated with a rectangular pipe shape, as illustrated in Figure 4.2.*

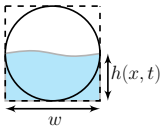


Figure 4.2: Semi-filled circular pipe approximated with a rectangular geometry, parametrized by the width w. Source: [Balla et al., 2022d]

The wetted perimeter $P(x,t)$ and pipe area $A(x,t)$ are formulated, i.e.,

$$A\big(h(x,t)\big) \approx wh(x,t), \tag{4.3a}$$

$$P\big(h(x,t)\big) \approx w + 2h(x,t), \tag{4.3b}$$

where $h(x,t), x \in (0,L)$ is the water level and the cross-section is described by the width parameter w. By using the linear area-level relation, the number of independent variables can be reduced to the levels $h(x,t)$ and flows $Q(x,t)$. Besides, the accuracy of this approximation is highly-dependent on the fullness of the pipe and the original geometry of the cross-section. For instance, circular pipes are approximated well with rectangular shape as shown in Figure 4.2.

Assumption 3. *In the case of the KW-based approximation of the full dynamic PDEs, we relate the friction $S_f(x,t)$ inside the pipes to the flow $Q(x,t)$ and water level $h(x,t)$ by utilizing the empirical Manning formula [Szymkiewicz, 2010], given by*

$$S_f(x,t) = n^2 \frac{P(x,t)^{4/3}}{A(x,t)^{10/3}} Q(x,t)^2, \tag{4.4}$$

where n is the Manning friction factor representing the roughness of the pipe.

Assumption 4. *In the case of the DW-based approximation of the full dynamic PDEs, we relate the friction $S_f(x,t)$ inside the pipes to the flow $Q(x,t)$ and water level $h(x,t)$ by utilizing the empirical Darcy-Weisbach formula [Szymkiewicz, 2010], given by*

$$S_f(x,t) = k \frac{P(x,t)}{8A(x,t)^3 g} Q(x,t)^2, \tag{4.5}$$

where k is the Darcy-Weisbach friction factor representing the friction losses in the pipe and g is the gravitational acceleration.

4.2.4 Kinematic Wave Model

When approximating the momentum equation of the Saint-Venant PDEs with kinematic waves, we assume quasi-steady flow and neglect the phenomena of flow attenuation, acceleration, and backwater by keeping only the friction $S_f(x,t)$ and slope S_b terms in eq.(4.2b). This assumption inherently means that the flow in sewer pipes propagates such that the gravitational and friction forces acting on the fluid are equivalent. This results in the following expression for the momentum of the fluid, i.e.,

$$S_b = S_f\big(Q(x,t), h(x,t)\big), \tag{4.6}$$

where the pipe geometry is simplified to the linear area-level relation, and the friction term is expressed by the Manning formula described by Assumption 2 and Assumption 3, respectively.

To present the transport dynamics in a form more amenable to system identification, the simplified PDEs are reduced to ordinary differential equations by spatially discretizing them. The pipe is divided into N_x non-overlapping δx pipe segments and the signals $h(x,t)$, $Q(x,t)$ and $q(x,t)$ are approximated as piecewise constant functions of the spatial coordinate x [Balla et al., 2020b]. The volume balance inside the pipe is depicted in Figure 4.3.

As shown in Figure 4.3, volume conservation applies for each segment of pipes, holding a balance between the section flows $Q(x,t)$ and lateral inflows $q(x,t)$. When considering the KW approximation, the section flows $Q(x,t)$ are generated due to the water level in each pipe segment. Note that the detailed derivation of the proposed KW-based model approximation is out of the scope of the summary, however, detailed in Paper A, Paper B and Paper C.

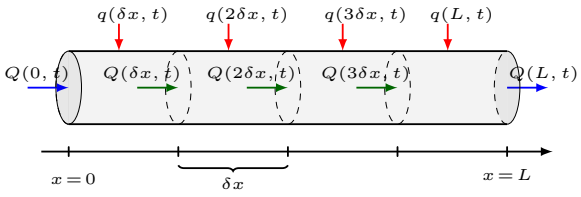


Figure 4.3: Volume balance for a pipe discretized into $N_x = 4$ sections for the KW approximation. Source: [Balla et al., 2021].

Applying the assumptions and the spatial discretization with the backward Euler method, the final form of the KW-based pipe model for the water level as the state is given by

$$\frac{dh_1(t)}{dt} = \theta_1\big(Q_0(t) + q_1(t)\big) - \theta_1\theta_2 f(h_1(t), \theta_3), \tag{4.7a}$$

$$\vdots$$

$$\frac{dh_n(t)}{dt} = \theta_1 q_n(t) + \theta_1\theta_2\big(f(h_{n-1}(t), \theta_3) - f(h_n(t), \theta_3)\big), \tag{4.7b}$$

$$\vdots$$

$$\frac{dh_{N_x}(t)}{dt} = \theta_1(q_{N_x}(t) - Q_{N_x}(t)) + \theta_1\theta_2 f(h_{N_x-1}(t), \theta_3), \tag{4.7c}$$

where $h_1(t), h_2(t), ..., h_{N_x}(t)$ represent the water level in each segment of pipe, while the nonlinear map $f : \mathbb{R}_+ \to \mathbb{R}_+$ is given by

$$f : (h(x,t), \theta_3) \mapsto \frac{h(x,t)^{5/3}}{(h(x,t) + \theta_3)^{2/3}}, \quad \forall\, x \in (0, L), \tag{4.7d}$$

and the physical constants along with the spatial step δx are lumped into the following parameters

$$\theta_1 = \frac{1}{w\delta x}, \quad \theta_2 = \frac{\sqrt{S_b}w^{5/3}}{2^{2/3}n}, \quad \theta_3 = \frac{w}{2}. \tag{4.7e}$$

Besides, the flow $Q_0(t)$ at $x = 0$ is considered as an input provided by controllable actuators. Since the section flows are generated by the water levels in each segment, the discharged flow at the downstream end of the channel is calculated directly from the water level in the last segment, i.e.,

$$Q_{N_x}(t) = \theta_2 f(h_{N_x}(t), \theta_3), \tag{4.8}$$

where $h_{N_x}(t)$ is the water level in the last pipe segment. The resulting KW-based model has a bi-linear structure in its parameters and the physical constants describing the original equations are lumped into three independent parameters. Note that the physical insight of the parameters allows to constrain them, i.e., all θ_1, θ_2, and θ_3 are positive, given that the physical constants and the spatial step δx are positive [Balla et al., 2021].

4.2.5 Diffusion Wave Model

When approximating the momentum equation of the Saint-Venant PDEs with diffusion waves, we do not neglect the diffusion term $\partial h(x,t)/\partial x$ in eq.(4.2b) [Balla et al., 2021]. Hence, the momentum equation for the DW-based model is given by

$$\frac{\partial h(x,t)}{\partial x} = S_b - S_f(x,t), \tag{4.9}$$

where similarly to the KW-based model, the pipe geometry is simplified to the linear area-level relation, and the friction term is expressed by the Darcy-Weisbach formula described by Assumption 2 and Assumption 4, respectively.

To present the transport dynamics in a form more amenable to system identification, the simplified PDEs are reduced to ordinary differential equations via spatial discretization. Similarly to the KW case, the pipe segment is divided into N_x non-overlapping δx pipe segments and the signals $h(x,t)$, $Q(x,t)$ and $q(x,t)$ are approximated as piecewise constant functions of the spatial coordinate x [Balla et al., 2020b]. The volume balance inside the pipe is depicted in Figure 4.4.

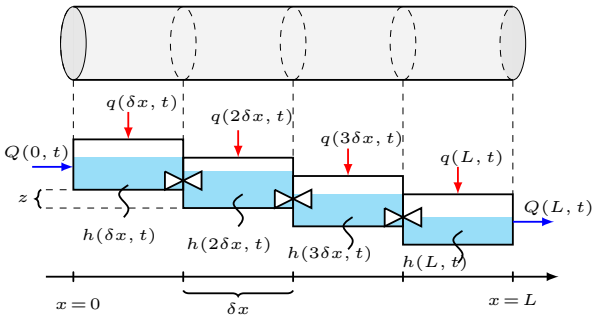


Figure 4.4: Volume balance for a pipe discretized into $N_x = 4$ sections for the DW approximation. Source: [Balla et al., 2021].

As shown in Figure 4.4, similarly to the KW model, the volume conservation applies for each segment of a pipe section. In contrast to the KW model, the relation between the section flows $Q(x,t)$ and the water levels $h(x,t)$ is not one-to-one, meaning that in this case the section flows are generated by the level difference between the interconnected pipe segments. This is depicted in Figure 4.4 by the elevation parameter $z = \delta x S_b$. As opposed to the KW-based model, the section flows are not only dependent on the local water levels, but also the water levels one spatial step forward [Balla et al., 2021]. This feature allows the model to inherently model the accumulation of volumes inside the channel, often occurring in gravity-driven sewer pipes with small slopes.

The detailed derivation of the DW-based model approximation is (likewise the KW-based model) out of the scope of the summary, however, detailed in

Paper B. To obtain the final DW-based model, we apply again the assumptions and the spatial discretization of the original model. In this case we use the backward Euler discretization for the term $\partial Q(x,t)/\partial x$ in eq.(4.2a) and the forward Euler discretization for $\partial h(x,t)/\partial x$ in eq.(4.2b). The final form of the DW-based pipe model for the water levels is given by eq.(4.10).

$$\begin{aligned}
\frac{dh_1(t)}{dt} =& \lambda_1(Q_0(t)+q_1(t)) - \lambda_1\lambda_2\big((h_1(t)-h_2(t)+z)g(h_1(t),\lambda_3)\big)^{\frac{1}{2}}, && \text{(4.10a)}\\
&\vdots\\
\frac{dh_n(t)}{dt} =& \lambda_1 q_n(t) + \lambda_1\lambda_2\Big[\big((h_n(t)-h_n(t)+z)g(h_{n-1}(t),\lambda_3)\big)^{\frac{1}{2}} && \text{(4.10b)}\\
&-\big((h_n(t)-h_{n+1}(t)+z)g(h_n(t),\lambda_3)\big)^{\frac{1}{2}}\Big],\\
&\vdots\\
\frac{dh_{N_x}(t)}{dt} =& \lambda_1(q_{N_x}(t)-Q_{N_x}(t)) && \text{(4.10c)}\\
&+\lambda_1\lambda_2\big((h_{N_x-1}(t)-h_{N_x}(t)+z)g(h_{N_x-1}(t),\lambda_3)\big)^{\frac{1}{2}},
\end{aligned}$$

where the nonlinear map $g:\mathbb{R}_+\to\mathbb{R}_+$ is given by

$$g:\ (h(x,t),\lambda_3)\mapsto \frac{h(x,t)^3}{h(x,t)+\lambda_3},\quad \forall\ x\in(0,L), \tag{4.11}$$

and the physical constants together with the spatial step δx are lumped into the following parameters

$$\lambda_1=\frac{1}{w\delta x},\quad \lambda_2=\left(\frac{4gw^3}{k\delta x}\right)^{1/2},\quad \lambda_3=\frac{w}{2}. \tag{4.12}$$

Note that in the case of the DW-based model, the elevation z is an extra model parameter, which is handled the same way as the lumped parameters, i.e., generally unknown. Similarly to the KW-based model, the parameters are all positive, i.e., $\lambda_1,\lambda_2,\lambda_3\in\mathbb{R}_+$. Note that the parameters for the KW- and DW-based models only differ in θ_2 and λ_2. Furthermore, the number of independent parameters has been extended with the elevation z [Balla et al., 2021].

4.2.6 Conceptual Alternatives

Conceptual modelling of volume conservation in gravity pipes arises from simplified mathematical descriptions, typically not related to physical models such as the Saint-Venant PDEs. The most relevant dynamics governing free-surface flow propagating in sewers are the time delay and the attenuation of the wave. Conceptual models describing hydraulic processes in sewers provide approximations to the flow propagating in the channels. Conceptualizing the physical models is often attractive from the complexity and computation point of view. Throughout the thesis, two types of conceptual modelling have been utilized; a single linear delay translational model and a simple linear version of the KW-based pipe model described in Section 4.2.4 (Kinematic Wave Model).

Modelling entire pipelines as pure delay elements is a computationally simple method where outflows $Q_N(t)$ from a gravity pipe are the delayed sums of the controlled inflows $Q_q(t)$ and the uncontrolled lateral inflows $q(t)$, i.e.,

$$Q_N(t) = Q_0(t - \tau) + q(t - \tau_{lat}), \tag{4.13}$$

where τ and τ_{lat} are time lags measured from the point where the controlled inflows and lateral inflows enter the sewer pipe, respectively. Note that the time delays are parameters to be estimated experimentally in the model. Although physical phenomena such as wave attenuation, backwater, and flow-dependent delays cannot be modelled with this technique, it is considered computationally beneficial for real-time control. From the practical point of view, the delay parameters in this model are difficult to find, especially if the disturbances $q(x,t)$ are spatially distributed along the channel. Such model has been utilized in the Simulation Study B and further detailed in Paper D.

To further simplify the KW-based approximation of the Saint-Venant PDEs, we assume that the relation between the friction $S_f(x,t)$ inside the pipe to the flow $Q(x,t)$ is linear, meaning that the empirical Manning formula introduced in Assumption 3 is a linear expression. Although such model does not have direct mathematical relation to the original Saint-Venant PDEs, it simplifies the KW-based model to a linear alternative, i.e.,

$$\frac{dh_1(t)}{dt} = \theta_1\big(Q_0(t) + q_1(t)\big) - \theta_1\theta_2 h_1(t), \tag{4.14a}$$

$$\vdots$$

$$\frac{dh_n(t)}{dt} = \theta_1 q_n(t) + \theta_1\theta_2\big(h_{n-1}(t) - h_n(t)\big), \tag{4.14b}$$

$$\vdots$$

$$\frac{dh_{N_x}(t)}{dt} = \theta_1\big(q_{N_x}(t) - Q_{N_x}(t)\big) + \theta_1\theta_2 h_{N_x-1}(t), \tag{4.14c}$$

and the discharged flow at the downstream end of the pipe simplifies as shown in eq.(4.15) below.

$$Q_{N_x}(t) = \theta_2 h_{N_x}(t), \tag{4.15}$$

where $h_{N_x}(t)$ is the water level in the last section of the pipe. Note that the above model is bi-linear in the parameters but linear in the control variables $Q_0(t)$ and states $h(t)$. Besides, the two parameters θ_1 and θ_2 are sufficient to describe the delay and flow attenuation of the traveling wave in the channel. As a consequence of neglecting the nonlinearities, this model introduces offset error via the linear flow to level conversion in contrast to the empirical Manning formula. This type of conceptualization has been used as an addition to Paper F, further detailed in Section 4.5.2 (System Identification Results).

4.2.7 Disturbance Model

Until this point, we considered the propagation of the controlled inflows $Q_0(t)$ coming from the upstream end of gravity pipes and the in-sewer level evolution based on physically-based and conceptual sewer models. However, all models are influenced by the disturbances $q(t)$ entering the pipes at multiple locations along the sewer. Due to the spatial discretization of the pipe models, the disturbances are aggregated at each section of the pipeline. The typical sources of disturbances are considered additive in terms of the inflows, i.e.,

$$q(x,t) = q^r(x,t) + q^h(x,t) + q^g(x,t), \quad \forall\ x \in (0,L), \tag{4.16}$$

where $q^r(x,t)$ denotes the rain runoff, $q^h(x,t)$ is the domestic or household sewage and $q^g(x,t)$ is the groundwater infiltrating into the system. The disturbances governing household sewage and groundwater are considered in each of the proposed physically-based models by introducing a disturbance model describing the signals. To set up an adequate disturbance model, the following assumption is introduced.

Assumption 5. *The disturbance flows induced by household sewage have an inherent periodicity such that $q^h(x,t) = q^h(x,t+T)$, where T corresponds to one day. Besides, the disturbance flows due to groundwater infiltration fulfill the constraint $\sum_{i=1}^{N_x} q^g(i,t) = N_x q^g(j,t)$, $\forall\ j \in \{1,2,...,N_x\}$, i.e. uniformly distributed along the whole length of the gravity sewer pipe [Balla et al., 2020a].*

To model the periodicity of the disturbance signal corresponding to the household and groundwater infiltration, we introduce Fourier series. For simplicity let us assume dry-weather flow, hence $q^r(x,t) = 0$, $\forall\ x \in (0,L)$. Then the disturbances describing dry-weather flow are given by

$$q(x,t) = \gamma_0 + \sum_{j=1}^{k} \big(\gamma_{1j}\cos(j\omega t) + \gamma_{2j}\sin(j\omega t)\big), \tag{4.17}$$

where the parameters are $\gamma = (\gamma_0, \gamma_{11}, \gamma_{21}, ..., \gamma_{1k}, \gamma_{2k})^\top \in \mathbb{R}^{2k+1}$. The angular frequency ω corresponds to one day and $k \geq 2$ is the number of frequency terms [Balla et al., 2020a]. The transport models proposed in this section in combination of this disturbance model are then used to find the network parameters θ, λ and the disturbance parameters γ.

Note that the disturbance $q^r(x,t)$ induced by the rain runoff dynamics is not handled by the parametric, physically-based models introduced in this section. However, handling rain runoff is of paramount importance as rain is typically the main driving force of flow propagation and overflows. Besides, the uncertainty present in modelling is not only due to the modelling of the runoff dynamics but also due to the high uncertainty governing the weather forecast of rain intensities. In the following sections, we provide some results of the physically-based modelling presented in this section.

4.3 Parametric Modelling Results

The parametric modelling results presented in this section are partially based on Paper A, Paper B and Paper C, where the KW- and DW-based dynamic models are utilized to estimate the parameters based on the water level sensor data $h(t)$ and the estimated flow of the actuators $Q(t)$.

4.3.1 System Identification Problem

The system identification problems concerning the physically-based parametric modelling in the thesis have been developed for the KW- and DW-based models. In each case, the underlying dynamic models have been discretized and the discrete-time model structure has been used for the system identification with sampled time series data. The dynamic equations of the KW-based model for a single pipeline are given in the implicit form, i.e.,

$$\hat{h}(t_{k+1}) = F_{\theta}^{\mathrm{KW}}\big(\hat{h}(t_k), Q_0(t_k), q(t_k)\big), \tag{4.18a}$$

$$Q_{N_x}(t_k) = \theta_2 f(h_{N_x}(t_k), \theta_3), \tag{4.18b}$$

where the numerical integration from t_k to t_{k+1} is done by the fixed-step 4^{th} order Runge-Kutta method [Balla et al., 2020a]. The predicted water levels are $\hat{h}(t_k) \in \mathbb{R}^{N_x}$, representing the states in each N_x section of pipes [Balla et al., 2021]. Furthermore, the disturbance vector $q(t_k) \in \mathbb{R}^{N_x}$ represents the lateral inflows at each segments of the pipe. The dynamics are given implicitly by $F_{\theta}^{KW} : \mathbb{R}^{N_x} \times \mathbb{R}_+ \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$, where the function is parametrized by the lumped pipeline parameters collected in the parameter vector $\theta = (\theta_1, \theta_2, \theta_3)^\top$. Note that the disturbances $q(t_k)$ in the argument of $F^{\mathrm{KW}}(\cdot)$ correspond to the groundwater and wastewater inflows. Furthermore, the outflow equation corresponding to the KW-based model is given by eq.(4.18b), where the function $f(\cdot)$ is equivalent to the function introduced in eq.(4.7d).

The DW-based model dynamics are given by eq.(4.19) for one pipeline

$$\hat{h}(t_{k+1}) = F_{\lambda,z}^{\mathrm{DW}}\big(\hat{h}(t_k), Q_0(t_k), Q_{N_x}(t_k), q(t_k)\big), \tag{4.19}$$

where the dynamics are discretized in time the same way as for the KW-based model and given implicitly by $F_{\lambda,z}^{DW} : \mathbb{R}^{N_x} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$, parametrized by the elevation z between the pipe segments and the lumped parameter vector $(\lambda_1, \lambda_2, \lambda_3)^\top$ [Balla et al., 2021]. Equivalently to the KW-based model, the modelled disturbances include ground- and wastewater infiltration. Note that we do not provide the outflow equation, as in the case of the DW-based model the flow is generated due to the water levels in the current and the neighboring downstream pipe sections. The downstream pipe section typically represents the receiving storage tank or pumping station. To provide such outflow relation, an extra structure needs to be introduced, depending on the regime of the flow, i.e., whether it is free or submerged flow. Instead, we provide the estimated

outflows $Q_{N_x}(t_k)$ as time series inputs in the argument. For more details, consult Paper B.

The collected measurement data consists of the water levels in pipe sections measured by the sensors. The measurement vector is given as follows

$$v(t_k) = Ch(t_k) + \nu, \tag{4.20}$$

where $C \in \mathbb{R}^{N_0 \times N_x}$ is picking out all measured states $v(t_k)$ from the state vector $h(t_k)$. N_0 denotes the number of level sensors deployed in the system and $\nu \in NID(0, \sigma^2)$ is white Gaussian measurement noise [Balla et al., 2021].

4.3.2 KW-based System Identification

The system identification problem for the KW-based model is given as follows

$$\begin{pmatrix} \theta^* \\ \hat{h}(t_0)^* \end{pmatrix} = \underset{\theta, \hat{h}(t_0)}{\operatorname{argmin}} \sum_{i=0}^{N_t} \left(Q_{N_x}(t_i) - \hat{Q}_{N_x}(t_i) \right)^2 + \Omega \| v(t_i) - \hat{v}(t_i) \|^2 \tag{4.21a}$$

subject to the dynamics in eq.(4.18) and to inequality constraints, i.e.,

$$0 < \quad \hat{h}(t_i) \leq \quad \overline{h}, \tag{4.21b}$$

$$0 < \hat{Q}_{N_x}(t_i) \leq \overline{Q}_{N_x}, \tag{4.21c}$$

$$0 < \quad \theta \quad \leq \quad \overline{\theta}, \tag{4.21d}$$

where eq.(4.21b), eq.(4.21c) and eq.(4.21d) impose bounds on states, outputs and parameters, respectively. Note that we minimize the squared errors between the measured water levels $h(t_k)$ (states) and the estimated outlet flows $Q(t_k)$ (outputs). Moreover, the upper and lower bounds relate to physically meaningful values. For instance, the model parameters are positive and their maximum value can be bounded based on a rough estimate of the slope, channel width, and length of the channel. Besides, Ω is a weighing constant scaling the water levels to the magnitude of the discharged flows [Balla et al., 2020a].

4.3.3 DW-based System Identification

Unlike the KW-based model, the system identification problem for the DW-based scenario is constructed only for the vector of water level measurements $v(t_k)$ as outputs, and the estimated discharge flow $Q_{N_x}(t_k)$ is rather provided as an input to the model, i.e.,

$$\begin{pmatrix} \lambda^* \\ z^* \\ \hat{h}(t_0)^* \end{pmatrix} = \underset{\lambda, z, \hat{h}(t_0)}{\operatorname{argmin}} \sum_{i=0}^{N_t} \|v(t_i) - \hat{v}(t_i)\|^2 \tag{4.22a}$$

subject to the dynamics in eq.(4.19) and to inequality constraints, i.e.,

$$0 < \hat{h}(t_i) \leq \overline{h}, \tag{4.22b}$$

$$0 < \lambda \leq \overline{\lambda}, \tag{4.22c}$$

$$0 < z \leq \overline{z}, \tag{4.22d}$$

where similarly to the KW case, eq.(4.22b) imposes bounds on the states, eq.(4.22c) on the parameters and eq.(4.22d) on the parameter corresponding to elevation, respectively.

The system identification problems in both cases are solved with a Gauss-Newton gradient method in the `nlgreyest` toolbox in `MATLAB`, furthermore, the auxiliary variable N_x are fixed for both problems. Selecting the number of discretized sections of pipelines is further detailed in Paper A.

4.3.4 Parametric System Identification Results

The results presented here are related to the flow propagation in gravity sewer lines regarding the parametric modelling introduced in the previous sections. The methods have been verified with the use of data extracted from high-fidelity simulations, from real world utilitz, and from our experimental laboratory setup.

The flow fitting and disturbance decoupling features of the KW-based model have been verified on data. Results of the flow discharge validation with the KW-based model in our Simulation Study A is shown in Figure 4.5.

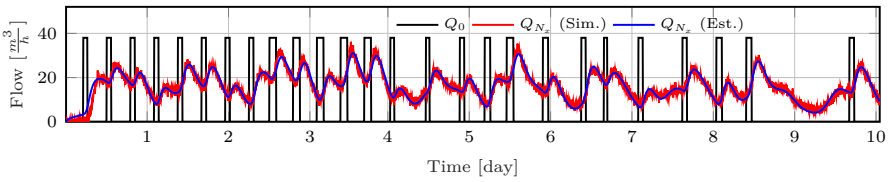


Figure 4.5: Discharged downstream flow prediction under on/off pump operation. Source: [Balla et al., 2020a].

As shown in the figure above, the discharged flow $Q_{N_x}(t_k)$ is composed of the pump flow cycles $Q_0(t_k)$, combined with ground and periodic wastewater patterns. To achieve such smooth model response, $N_x = 4$ sections with $N_0 = 4$ level sensors have been utilized. Moreover, in Simulation Study A we also visualized the disturbance decoupling features of the model. In the simulation, we can show the actual and the discharged disturbances by simply simulating

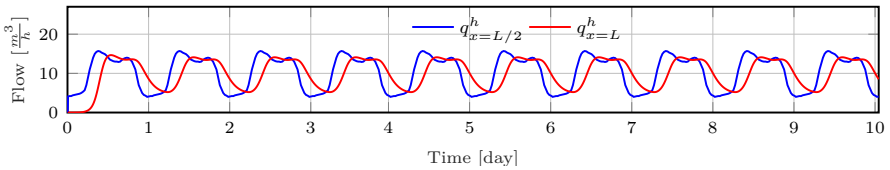


Figure 4.6: Disturbance infiltrated at the middle of the pipe and discharged downstream. Source: [Balla et al., 2020a].

the high-fidelity models without the actuation of the pumps. The flow discharge inside the empty channel in the simulation study is shown in Figure 4.6..
As seen, the wastewater discharged in the middle of the pipeline ($q^h_{x=L/2}$) is delayed and attenuated at the downstream end ($q^h_{N_x}$) where it is discharged. Using the estimated parameters θ of the KW-based model together with the disturbance model presented in Section 4.2.7 (Disturbance Model), the decoupled wastewater disturbance flow can be calculated at the point where it enters the channel. This is shown in Figure 4.7.

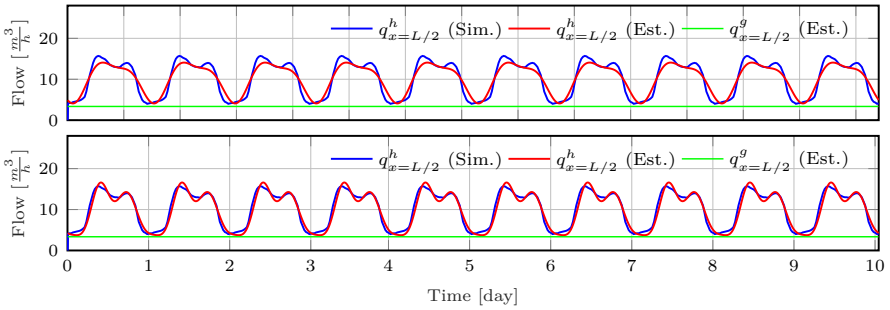


Figure 4.7: Disturbance decopling with 2^{nd} order (above) and 4^{th} order Fourier series. Source: [Balla et al., 2020a].

The higher-order the Fourier series, the better decoupling performance our model yields to the cost of increasing the number of parameters in the combined hydraulic and disturbance model. Note that the model parameters θ and disturbance parameters λ in the KW-based model allow detecting the offset on the disturbance signals, indicating groundwater infiltration into the channel.

Similar tests have been carried out on our Haderslev Case Study where real flow data pairs of the pumped $Q_0(t_k)$ and discharged $Q_{N_x}(t_k)$ flow time series have been utilized. Note that in this study level sensors have not been available, however, our model allows fitting to the input-output flows. Although the tests are from the operation of a real combined sewer system, wastewater infiltration is not present in the data. The resulting validation of the Haderslev Case Study for flow prediction in two different branches of the wastewater network is shown in Figure 4.8.

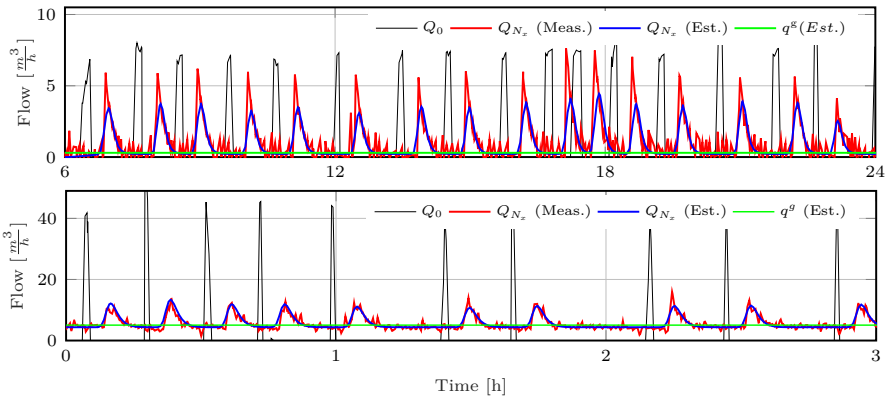


Figure 4.8: Model validation on experimental data of the gravity sewer line between $PGH103$ and $PGH104$ pumping stations (above) and between the $PGH202$ and $PGH203$ pumping stations (below). Source: [Balla et al., 2020a].

The estimation data used in this study are pumped and discharged flows in both cases. As depicted in the figures, the groundwater infiltration in the graph above is nearly zero, meaning that the channel dries out between two pumping cycles. Nevertheless, the groundwater infiltration is significant in the figure below, as the discharged flow to the downstream pumping stations is constant, even if pumps are not operating. The model fitted to the data above was constructed with $N_x = 7$ pipe sections. For the sake of curiosity, it would be interesting to deploy level sensors in the Haderslev Case Study network to detect where the groundwater is infiltrating the sewer pipes.

Following the KW- and DW-based modelling methodologies in Section 4.2.4 and Section 4.2.5, a comparison has been made in our Experimental Study. To validate the results of both methods, we tested the KW and DW models against the same experimental data. To point out the main feature differences between the two methods, we placed a level sensor upstream to the point where lateral inflow disturbance enters the gravity channel, creating backflow in the sewer. The validation data and the results are shown in Figure 4.9. The pump and lateral inflows are shown in Figure 4.9(a-b) while the level sensor measurement data are depicted in the rest of the graphs. Note that the upstream level measurement (h_1) highly resembles the pumping cycles not being affected by the disturbance flows. The disturbance flow enters the pipeline before the third sensor (h_3), however, as seen in Figure 4.9(e-f), the second sensor h_2 is affected by the disturbance in the form of backwater inside the pipe. Both models have been fit to the level data from time 0−1250 (2s), and the remaining data set was used for validation (yellow area). The two models are validated in free-running mode, meaning that an initial state and the input flow Q_0 sequence are applied for prediction in closed-loop. As seen in Figure 4.9(e-f), the KW-based model cannot model the water volumes accumulating according to the sensor at the second position h_2. In contrast to the KW-based model, the DW-based model finds the additional parameter z and captures the backwater effect.

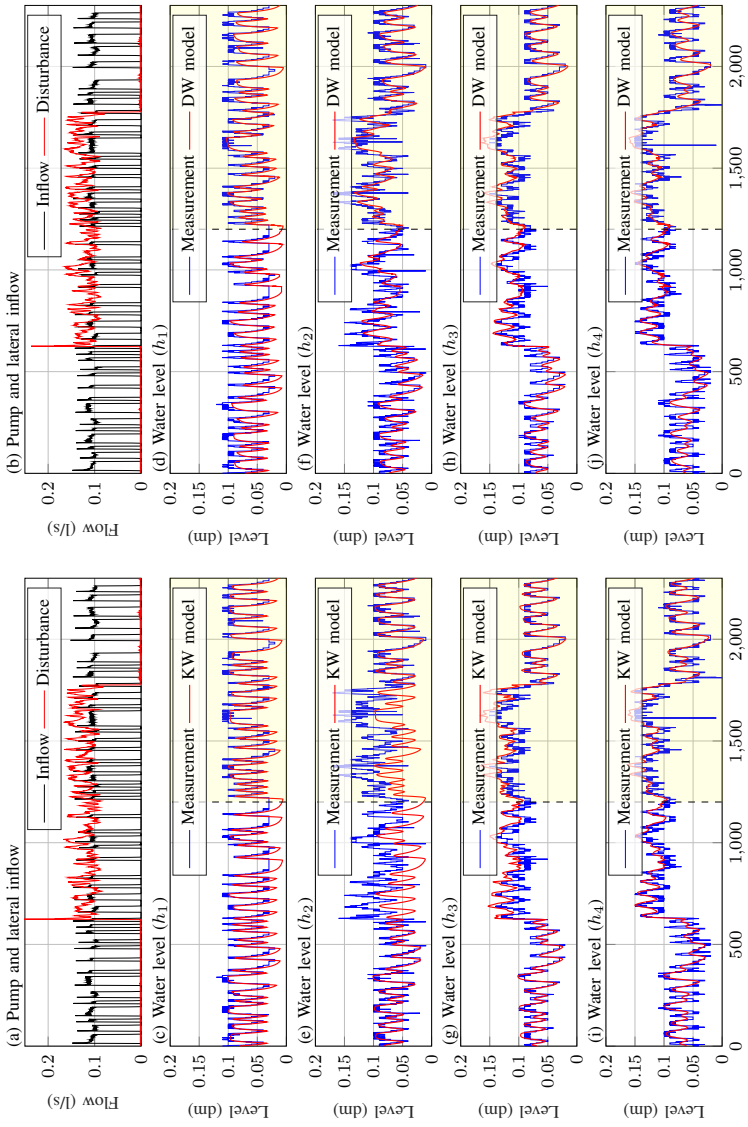


Figure 4.9: System identification and model valuidation comparison between the KW- (left column) and DW-based (right column) modelling approaches. Source:[Balla et al., 2021]

It is important to note that the training data includes half of the backflow period, so the model can adjust and find a good estimate for the z elevation parameter. As shown in the level sensor readings in Figure 4.9(g-j), the models are equally good at predicting the level downstream.

The results above are further verified by showing the histogram of the error signals between the measured data points and the predicted model outcome for both the KW- and DW-based model structures, respectively. The histograms are shown in Figure 4.10.

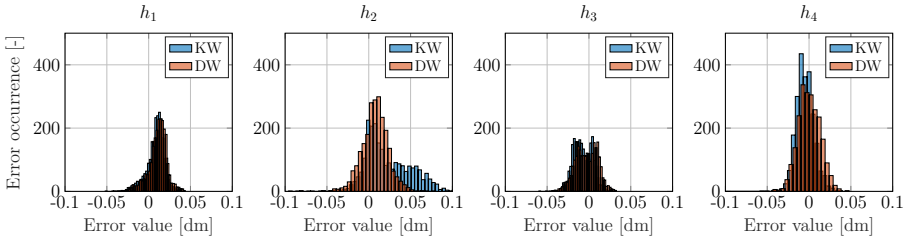


Figure 4.10: Histogram of the error residuals for the KW- and DW-based model predictions.

As shown, the model performance is similar in the two cases expect for h_2 where the backwater lifts the volume in the upstream pipe channel. Note that the histogram for the KW-based model is skewed to the positive error values, as the model does not capture the backwater in the second state. Moreover, the histogram is also slightly skewed in the first state h_1 to the positive values. This can be explained by the physical fact that the channel at the upstrem does not dry out completely due to its small slope. This is depicted in Figure 4.9, where the model predicts that the level attenuates slowly to zero, while there is no input applied to the system. In contrast to that, the measured level saturates at a lower water level.

One of the advantages of the DW-based model over the KW-based version is its capability of keeping track of the accumulated volumes in the pipes, hence accounting for the stored volume (level). This opens the possibility to use large sewer pipes in a smart way to store excess volumes in high-intensity rain periods. However, this feature comes with increased model complexity.

Additional results of the parametric modelling in gravity-driven sewers can be found in Paper C where the KW-based model and the disturbance model are combined in a moving horizon estimation setup to estimate both states and parameters based on historical data batches.

It is important to note that the presented results do not provide a solution to rain infiltration into the network. To predict with the proposed models, the rain disturbance contribution needs to be provided in the form of flow data. In the following sections, we attempt to solve this issue by combining the parametric methods with non-parametric ones to include the hydrologic processes in the model.

4.4 Non-parametric Modelling of Sewer Hydrology

The modelling of hydraulic and hydrologic processes are typically treated as two separate problems in the state-of-art concerned with physically-based sewer modelling. Opposed to the state-of-art, in this section we propose to combine the parametric modelling of the system hydraulics introduced in Section 4.2 (Parametric Modelling of Sewer Hydraulics) with non-parametric approaches to incorporate the dynamic effect of the rain disturbances affecting the sewers.

In Section 4.2 (Parametric Modelling of Sewer Hydraulics), we argued that grey-box modelling is a reasonable approach since the hydraulic processes inside the sewer are governed by well-established white-box models. In this section, we focus on the hydrologic processes and argue that their distributed and highly-stochastic nature makes it very difficult to model them precisely using physical laws. Hence, in the following sections, we turn our attention to the available real-time sensor data to build prediction models.

4.4.1 Gaussian Process Modelling

A powerful way to represent input-output relations is through Gaussian Processes (GPs). GP models belong to a class of black-box modelling, which does not attempt to approximate the modelled system by finding parameters of the corresponding mathematical model structure but rather searches for the input-output relationship in the available data. The predicted output of a GP is a Gaussian distribution parameterized by the mean and the variance of the process, i.e., the mean value representing the most likely outcome of output and the variance representing the measure of confidence. The variance, therefore, depends strongly on the quality and amount of available data.

As opposed to the previously described approaches regarding the volume conservation inside sewers in Section 4.2 (Parametric Modelling of Sewer Hydraulics), instead of claiming that the input-output dynamics belong to a specific mathematical structure, GPs are non-parametric, probabilistic models based on the data. Let us denote the input-output function describing the unknown dynamics as $g(\cdot)$. A GP characterizing the distribution of all possible functions of $g(\cdot)$ of the dynamic process can be given as

$$g(z) \sim GP(m(z), \Sigma_{GP} + I\sigma_n^2), \tag{4.23}$$

where $z \in \mathbb{R}^{N_z \times M}$ is the set of inputs, N_z denoting the number of input dimensions and M the number of data points used for prediction. Besides, $m(\cdot)$ is the mean function and $\Sigma_{GP} \in \mathbb{R}^{M \times M}$ is the covariance matrix. The noise on the data is represented by σ_n^2 and I is the identity matrix of suitable dimension. The mean and covariance functions are defined by

$$m(z(i)) = \mathbb{E}\{g(z(i))\}, \tag{4.24a}$$

$$\Sigma_{GP}(i,j) = \operatorname{cov}\Big(g(z(i)), g(z(j))\Big) \approx k\Big(z(i), z(j)\Big), \tag{4.24b}$$

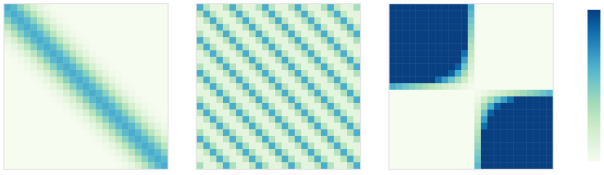
where the mean m and the covariance matrix Σ_{GP} are obtained via evaluating all available data pairs z. Note that $z(i)$ and $z(j)$ are time series at the time instances i and j, respectively. Specifically, the values of the covariance matrix corresponding to the covariances between the function values of $g\big(z(i)\big)$ and $g\big(z(j)\big)$, regarding the arguments of the input set $z(i)$ and $z(j)$. Besides, $\mathbb{E}\{\}$ is the expected value operator. Note that in eq.(4.24b), instead of evaluating the covariance function explicitly, the covariance matrix is approximated with a kernel function k. The purpose of this kernel is to establish a measure of similarity between the function values of g, i.e., to capture the correlation between different training data observations. By choosing a kernel function, we give a structure to our model, based on some prior information or belief governing our system dynamics. Some kernel functions are shown in eq.(4.25).

$$k\Big(z(i), z(j)\Big) = (z(i) - c)^\top \Omega (z(j) - c), \tag{4.25a}$$

$$k\Big(z(i), z(j)\Big) = \sigma_f^2 \exp\left(-\frac{2}{\sigma_L^2} \sin^2\left(\frac{\pi}{T}|z(i) - z(j)|\right)\right), \tag{4.25b}$$

$$k\Big(z(i), z(j)\Big) = \sigma_f^2 \exp\left(-\frac{1}{2}\Big(z(i) - z(j)\Big)^\top \Lambda^{-1}\Big(z(i) - z(j)\Big)\right), \tag{4.25c}$$

where eq.(4.25a) is a linear, eq.(4.25b) is a periodic, and eq.(4.25c) is a squared exponential kernel. Furthermore, the hyper-parameter σ_f^2 describes the variance of the signal, $\Lambda^{-1} = \text{diag}(\sigma_{L,1}^{-2}, ..., \sigma_{L,N_z}^{-2})$ denotes the length scale matrix, where σ_L is the length scale for each input dimension, and $\Omega = \text{diag}(\sigma_{f,1}^2, ..., \sigma_{f,N_z}^2)$ is the signal variance matrix for the linear kernel. Furthermore, T corresponds to the periodicity of the periodic kernel, which can either be pre-defined or found through the optimization while training the model. Finally, the constant c represents an offset in the linear kernel. The visualization of the kernel functions is shown in Figure 4.11.



(a) Squared exponential **(b)** Periodic **(c)** Linear

Figure 4.11: Heatmaps of the squared exponential, periodic and linear kernel approximations, created from 25, equally-spaced values from -5 to 5. Source: [Görtler et al., 2019].

The choice of the kernel structure is of fundamental importance when training a model with GPs. We provide the following remarks for our prior belief and structure on the GP and the kernel and mean functions.

Remark 1. *We utilize the squared exponential kernel defined in eq.(4.25c). We argue that the runoff dynamics and the sewer processes in general exhibit smooth behavior, and therefore the smooth properties of such kernel is sufficient to approximate the covariance function [Rasmussen and Williams, 2018].*

Remark 2. *The mean function utilized in our work is assumed to be constant and deterministic. The reasoning behind choosing a constant mean function is twofold; Firstly, in combined wastewater networks the measured states, i.e., the in-sewer water levels are rarely zero due to the dry-weather flow activity induced by groundwater or the domestic household discharges. Hence the pipes rarely dry out. Secondly, level sensors in practice are difficult to calibrate to zero offset errors. Therefore, we aim to learn the offset on the measurements by choosing a constant mean function.*

Note that zero mean function is typically assumed as the prior mean function for GPs. In that case, we need to make sure that the data is pre-processed such that the training set is zero mean.

Having the data and our prior belief on the structure of the mean and covariance functions, we seek the posterior distribution of $g(\cdot)$. It can be shown that the posterior of the GP using Bayes' Rule is [Kocijan, 2016]

$$\mathbb{P}\{g \,|\, z, y\} = \frac{\mathbb{P}\{g\}\mathbb{P}\{y \,|\, z, g\}}{\mathbb{P}\{y \,|\, z\}}, \tag{4.26}$$

where $\mathbb{P}\{y \,|\, z, g\}$ is the likelihood, $\mathbb{P}\{g\}$ is the function prior, $\mathbb{P}\{y \,|\, z\}$ is the evidence and $\mathbb{P}\{g \,|\, z, y\}$ is the posterior distribution over $g(\cdot)$. Moreover, $\mathbb{P}\{\}$ is the probability operator. Then, the posterior distribution simplifies to

$$\mathbb{P}\{g \,|\, z, y\} \sim GP(m(z), \Sigma_{GP} + I\sigma_n^2). \tag{4.27}$$

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

Once the hyper-parameters are learned from the data, the GP model can be used to predict an output point y^* using the testing point z^*, such that $y^* = g(z^*)$. The problem of predicting one step with the GP model corresponds to finding the probability distribution of $\mathbb{P}\{y^* \,|\, z, y, z^*\}$, given the training input-output set $\{z, y\}$, the testing point z^* and the hyper-parameters. Note that the detailed derivation of how the predictive mean and variance of a GP are obtained is out of the scope of this summary, however, detailed in [Rasmussen and Williams, 2018]. The predicted mean and variance are given by eq.(4.28).

$$\mu_{GP}(z^*) = m(z^*) + K_{z^*z}\left(K_{zz} + I\sigma_n^2\right)^{-1}(y - m(z)), \tag{4.28a}$$

$$\Sigma_{GP}(z^*) = K_{z^*z^*} - K_{z^*z}\left(K_{zz} + I\sigma_n^2\right)^{-1}K_{zz^*}, \tag{4.28b}$$

where $K_{zz^*} = k(z, z^*)$ and $K_{z^*z} = K_{zz^*}^\top$ are the covariances between the training and testing points, furthermore $K_{z^*z^*}$ is the autocovariance of the testing point.

4.4.2 Residual Approach

As stated in Section 4.2 (Parametric Modelling of Sewer Hydraulics), the disturbances in the proposed grey-box model for the gravity flow do not concern the rain infiltrating into the sewers and the uncertainty characterizing the household sewage production. We propose to exclude the exogenous effect of dry- and wet-weather disturbances from our nominal sewer models describing the pipe and storage elements. Our main goal is to deploy the level sensors at critical locations in the wastewater network, where we can capture their effect through the in-sewer level variations. A visualization for the instrumentation of such sewer network is shown in Figure 4.12 below.

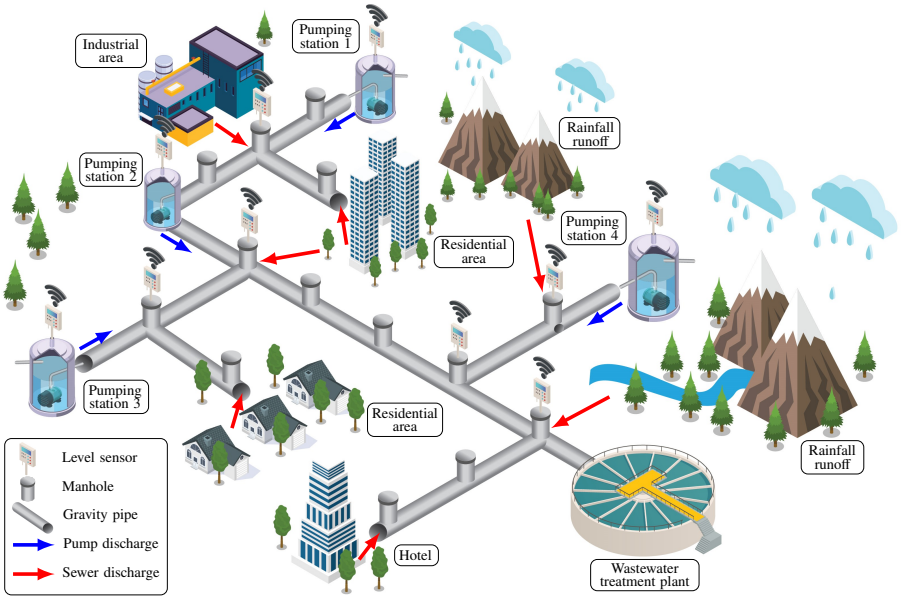


Figure 4.12: An illustration of a sewer network. Sensors are deployed to learn the effect of disturbances through in-sewer variations of the water level. Source: [Balla et al., 2022a].

As shown in Figure 4.12, level sensors are deployed mainly in the vicinity of lateral inflow points where pipes join the main sewer lines connecting the controllable assets in the network, e.g., in our case the pumping stations.

Assumption 6. *The nominal model considered for residual generation is linear (or linearized) with a standard state-space model structure, i.e.,*

$$h(t_{k+1}) = f\big(h(t_k), Q_0(t_k)\big) = Ah(t_k) + BQ_0(t_k), \tag{4.29}$$

where $f(\cdot)$ *denotes the nominal dynamics,* $Q_0(t_k) \in \mathbb{R}^{N_u}$ *is the vector of controllable flows,* N_u *denoting the number of controllable assets. Besides,* $h(t_k) \in \mathbb{R}^{N_p+N_t}$ *is the state vector,* N_p *denoting the number of pipe states and* N_t *the*

number of storage tank states. Furthermore, A and B are the state-space matrices of suitable dimensions.

In eq.(4.29), the state vector includes all available sensor measurements, including the level in both basins and sewer pipes corresponding to the tank and pipe volume conservation dynamics. It is important to note that we slightly modify the parametric dynamics model in comparison to what we presented in Section 4.2 (Parametric Modelling of Sewer Hydraulics). For instance, the storage tank dynamics describe the variation of level but only consider the controlled flow removing volume from the tank. This can be also read from eq.(4.29), where a complete volume conservation model would also include the volumes arriving at the storage tank, i.e., the disturbances q induced by wastewater and rain runoff. However, these disturbances are generally unknown and therefore not considered in this part of the model. Hereinafter we refer to the volume conservation models as nominal dynamics.

Given all the deployed level sensor data h, the measured or estimated flow input of the actuators (pumps), and some weather forecasts in the form of rain intensity measure, we form our modelling problem by the need to learn the unknown and unmodelled dynamics to complement the nominal dynamics described in eq.(4.29). For this reason, we assume a structure for the network model where the dynamics are composed of a nominal and an additive unknown part. The former describes the hydraulics in the network, attempted to model through parametric, grey-box modelling techniques. The latter represents the dynamics governing the disturbances, i.e., the rain runoff dynamics and wastewater inflows. (Note that in some of our research we also considered the flow propagation dynamics in sewer pipes, leaving only the integrator states, i.e., the tank dynamics for the nominal parts.)

Similarly to [Hewing et al., 2020], the combined network model in discrete time is given in the form shown below

$$h(t_{k+1}) = f\Big(h(t_k), Q_0(t_k)\Big) + B_p g\Big(h(t_k), Q_0(t_k), d(t_k), t_k\Big) + w(t_k), \qquad (4.30)$$

where g is a nonlinear, unknown vector function representing the unknown dynamics, modelled by the GPs. The function g is equivalent to what we presented in Section 4.4.1 (Gaussian Process Modelling) for modelling the input-output relation of unknown dynamics. Moreover, $d \in \mathbb{R}^{N_d}$ is the vector of rain intensity forecasts, N_d representing the number of locations where we have forecasts available for predicting the rainfall-runoff. Besides, B_p is a matrix mapping the outputs of the unknown function g to the full state vector h. Simply stated: if there are states not affected by the disturbances or the unmodelled dynamics, then B_p maps zeros as the contribution from the additive GP compensation. The process noise $w \sim \mathcal{N}(0, \Sigma_w)$ follows a Gaussian white noise distribution.

For generality, we collect the data used for generating the residuals in the following vector referred to as input data vector, i.e.,

$$z(t_k) = [h(t_k)^\top, Q_0(t_k)^\top, d(t_k)^\top, t_k]^\top. \tag{4.31}$$

where $h(t_k)$, $Q_0(t_k)$ and $d(t_k)$ are the state, input and disturbance vectors describing the state-input space and the disturbances. Since the dry-weather flows typically follow a diurnal cycle[1] we assume that these disturbances generated by the household sewage correlate with time. For this reason, we provide time t_k as an additional input, assuming that flow patterns are similar at the same time of the day on different days.

To generate the residuals, we subtract the output of the nominal model from the measured states, hence the output data vector becomes

$$y(t_k) = g\big(z(t_k)\big) + w(t_k) = B_p^\dagger\Big(h(t_{k+1}) - f\big(h(t_k), Q_0(t_k)\big)\Big), \tag{4.32}$$

where $y \in \mathbb{R}^{N_y}$ is the vector of residuals, N_y denoting the dimensions of states compensated by the GPs. Besides, the mapping matrix B_p is inverted with the Moore-Penrose pseudo-inverse. We aim to generate residuals to remove the effect of known dynamics from our measurements, which we later use for modelling the input-output relation with the nonparametric GP models. Simply stated: we attempt to model a small subpart of the dynamics with the GPs and leave the grey-box, physically-based nominal dynamics as the backbone of the network model.

The full training set used for training the GPs are then constructed by collecting the input-output data vectors in the following set

$$\mathcal{D} = \Big\{\Big(z(i), y(i)\Big) \mid i = 1, ..., M\Big\}, \tag{4.33}$$

where M corresponds to the number of collected data used for training.

4.4.3 Prediction with Residuals

Predicting multiple-step with our nominal model is straightforward, especially if the sewer dynamics are governed by linear dynamics and are deterministic. However, this is not the case after combining our physically-based model with the Gaussian processes, as the output of the GPs are stochastic variables following a Gaussian distribution. Predicting multiple-step with such models inherently means that the mean and variance of the previously predicted states are used to predict the next state values in time. Hence, we feedback normally distributed variables, which in general does not result in normally distributed outputs, as our kernel stated in eq.(4.25c) is nonlinear.

To resolve this problem, we present the combined dynamics of the nominal and GP models, such that at each prediction step t_k the states h and the GP dynamics are approximated as jointly Gaussian, i.e.,

[1] A diurnal cycle is a pattern that repeats at every 24 hour in accordance to one full rotation of the Earth. In our application, the activity of human behavior plays a role in the daily wastewater generation discharged to the sewer network.

$$\begin{pmatrix} h(t_k) \\ GP(t_k) \end{pmatrix} \sim \mathcal{N}\Big(\mu(t_k), \Sigma(t_k)\Big) = \left(\begin{bmatrix} \mu_h(t_k) \\ \mu_{GP}(t_k) \end{bmatrix}, \begin{bmatrix} \Sigma_h(t_k) & \Sigma_{h,GP}(t_k) \\ \Sigma_{GP,h}(t_k) & \Sigma_{GP}(t_k) \end{bmatrix}\right), \quad (4.34)$$

where $\Sigma_{h,GP} = (\Sigma_{GP,h})^\top$ are the cross-covariances between the water levels and the GPs, $\mu_h(t_k)$ is the vector of mean levels and $\Sigma_h(t_k)$ is the covariance matrix of the water levels at time t_k. The inputs $Q_0(t_k)$ are considered as deterministic variables. (Otherwise it is necessary to consider the cross-correlation between the inputs and the states, and between the GPs and the inputs.)

To obtain the transition probability of the full state-space, similarly to [Hewing et al., 2020], we apply the first-order Taylor expansion of the joint mean-variance dynamics shown in eq.(4.34). Then, the linearized mean and variance dynamics become

$$\mu_h(t_{k+1}) = f\Big(\mu_h(t_k), Q_0(t_k)\Big) + B_p \mu_{GP}(t_k), \quad (4.35a)$$

$$\Sigma_h(t_{k+1}) = \Big[\nabla_h f\Big(\mu_h(t_k), Q_0(t_k)\Big), B_p\Big] \Sigma(t_k) \Big[\nabla_h f\Big(\mu_h(t_k), Q_0(t_k)\Big), B_p\Big]^\top, \quad (4.35b)$$

where Σ is the joint covariance matrix given in eq.(4.34) and the operator ∇_h denotes the first-order partial derivative with respect to the states.

The final mean-variance dynamics of the combined nominal and GP model is given by expressing the state-space structure of the nominal dynamics, i.e.,

$$\mu_h(t_{k+1}) = A\mu_h(t_k) + BQ_0(t_k) + B_p \mu_{GP}(t_k), \quad (4.36a)$$

$$\begin{aligned} \Sigma_h(t_{k+1}) = {} & A\Sigma_h(t_k)A^\top + B_p \Sigma_{GP,h}(t_k) A^\top \\ & + A\Sigma_{h,GP}(t_k) B_p^\top + B_p \Sigma_{GP}(t_k) B_p^\top, \end{aligned} \quad (4.36b)$$

where the covariance and cross-covariance updates are calculated by

$$\mu_{GP}(t_k) = \mu_{GP}(\tilde{z}(t_k)), \quad (4.37a)$$

$$\Sigma_{h,GP}(t_k) = \Sigma_h\Big(\nabla_h \mu_{GP}(\tilde{z}(t_k))\Big)^\top, \quad (4.37b)$$

$$\Sigma_{GP}(t_k) = \Sigma_{GP}\Big(\tilde{z}(t_k)\Big) + \nabla_h \mu_{GP}\Big(\tilde{z}(t_k)\Big) \Sigma_h(t_k) \Big(\mu_{GP}(\tilde{z}(t_k))\Big)^\top, \quad (4.37c)$$

where the input vector is $\tilde{z}(t_k) = [\mu_h^\top(t_k), {Q_0}^\top(t_k), d^\top(t_k), t_k]^\top$ with the measured state being the mean $\mu_h(t_k)$. Note that in eq.(4.37) we need to calculate the derivative of the posterior mean $\nabla_h \mu_{GP}(\tilde{z}(t_k))$. As seen from the expression of the GP mean in eq.(4.28), only the $K_{z^* z}$ term depends on the testing point z^*, since we chose the mean of the GP $m(\cdot)$ to be constant. Therefore, to calculate the first derivative of the posterior mean $\mu_{GP}(\tilde{z}(t_k))$, only the kernel

$K_{z^*z} \approx k(z^*, z)$ needs to be derivated. The derivative of the posterior mean with squared exponential kernel is given in [Kocijan, 2016].

The algorithm to predict N-step ahead, i.e., to propagate the uncertainty with the final model is shown in Algorithm 1 below.

Algorithm 1 State uncertainty propagation with Taylor approximation

Input: $\mu_h(t_k)$, $\Sigma_h(t_k)$, $Q_0(t_k)$, $d(t_k)$, t_k, and $\mathcal{D}$
Construct testing vector $z^*(t_k) = [\mu_h^\top(t_k), {Q_0}^\top(t_k), d^\top(t_k), t_k]^\top$
Construct training vector $z = [\mu_h^\top, {Q_0}^\top, d^\top, t]^\top$
Approximate kernel $K_{z,z*} \approx k(z, z^*)$
Calculate posterior mean $\mu_{GP}(z^*) = m + K_{zz^*}(K_{zz} + I\sigma_n^2)^{-1}(y - m)$
Calculate cross-covariance $\Sigma_{h,GP}(t_k) = \Sigma_h\Big(\nabla_h \mu_{GP}(z^*(t_k))\Big)^\top$
Calc. cov. $\Sigma_{GP}(t_k) = \Sigma_{GP}(z^*(t_k)) + \nabla_h \mu_{GP}\Big(z^*(t_k)\Big)\Sigma_h(t_k)\Big(\mu_{GP}(z^*(t_k))\Big)^\top$
Construct joint covariance $\Sigma(t_k) = \begin{bmatrix} \Sigma_h(t_k) & \Sigma_{h,GP}(t_k) \\ \Sigma_{GP,h}(t_k) & \Sigma_{GP}(t_k) \end{bmatrix}$
Calculate predicted state mean $\mu_h(t_{k+1}) = f\Big(z^*(t_k)\Big) + B_p \mu_{GP}(t_k)$
Calculate predicted state covariance $\Sigma_h(t_{k+1}) =$
$\Big[\nabla_h f\Big(z^*(t_k)\Big), B_p\Big]\Sigma(t_k)\Big[\nabla_h f\Big(z^*(t_k)\Big), B_p\Big]^\top$
Return: $\mu_h(t_{k+1})$ and $\Sigma_h(t_{k+1})$

In Algorithm 1 the inputs are the measured states $\mu_h(t_k)$, the initial covariance $\Sigma_h(t_k)$, the controlled flow input $Q_0(t_k)$, the forecasted rain intensity d(t), the time of the day t, and the historical training data vector $\mathcal{D}$, consisting of M datapoints of the input vector z and the residual vector y.

By including the variance of the states in our model, we propagate the variances representing the uncertainty around our mean predictions for predicting N-step ahead in time. The prediction is therefore largely dependent on the number of datapoints M and their quality, as the variance provides us confidence regarding our historical observations of the process. In our application, this confidence is influenced mainly by the rain disturbances, i.e., states under previously observed storm events are likely to be predicted with high confidence.

4.5 Non-parametric Modelling Results

The non-parametric modelling results presented in this section are partially based on Paper E, Paper F and Paper G, where the GP models combined with the tank model (i) and the GP models combined with both the tank and simplified pipe models (ii) are trained with the use of the water level sensor data h, the estimated flow of the actuators Q_0 and with either the intensity of rain r or the disturbance flow appearing due to the rain q^r.

4.5.1 System Identification Problem

The system identification problem based on the non-parametric modelling and the residual generation has been carried out via the volume conservation in tanks (Section 4.2.1 - Volume Conservation in Tanks), the conceptualization of the KW-based pipe model with the assumption that the flow to level relation is linear (Section 4.2.6 - Conceptual Alternatives), and the Gaussian processes (Section 4.4.1 - Gaussian Process Modelling). The system identification problem has been tested on the Experimental Study, recreating a subpart of a real-world wastewater infrastructure, further detailed in Paper E and Paper F. For the sake of clarity, the schematics of the experimental setup with the corresponding sensor utilization is shown again in Figure 4.13 below.

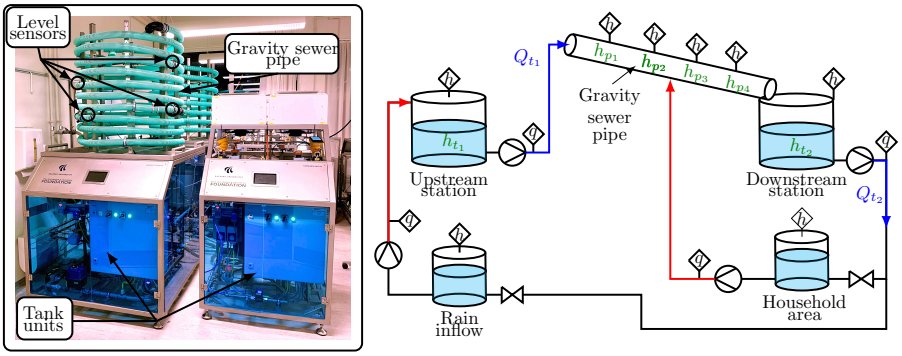


Figure 4.13: Experimental sewer network setup of the Smart Water Infrastructures Laboratory (SWIL) at Aalborg University. Left: Pumping station and pipe modules in the laboratory. Right: Schematics of the network topology, where q and h denote flow and level sensors, respectively. Source: [Balla et al., 2022b] and [Balla et al., 2022a].

Following the methodology in Section 4.4 (Non-parametric Modelling of Sewer Hydrology), our aim is to generate the residuals between measurements and the nominal model, i.e.,

$$y(t_k) = B_p^{\dagger}\Big(h(t_{k+1}) - f\big(h(t_k), Q_0(t_k)\big)\Big), \tag{4.38}$$

where $f(\cdot)$ describes all the known dynamics regarding the tank and pipe states, whereas $h(t_k)$ depicts the vector of water levels. The tank dynamics are considered to be fully defined by knowing the diameter and geometry properties of the tank, while the linear pipe dynamics described in Section 4.2.6 (Conceptual Alternatives) only consider the actuated inflow propagating from upstream to downstream. The latter is identified on level data as described in Section 4.3.2 (KW-based System Identification). Besides, the training data array for identifying the hyperparameters of the GPs is constructed such that

$$z(t_k) = [h^{\top}(t_k), Q_0^{\top}(t_k), d(t_k), t_k]^{\top}, \tag{4.39}$$

where $d(t_k)$ denotes the forecast of the disturbance and t_k is the time. Besides, the state $h(t_k)$ and input $Q_0(t_k)$ vectors are written specifically for our Experimental Study, i.e.,

$$h(t_k) = [h_{t_1}(t_k), h_{t_2}(t_k), h_{p_3}(t_k)]^\top, \tag{4.40a}$$
$$Q_0(t_k) = [Q_{t_1}(t_k), Q_{t_2}(t_k)]^\top, \tag{4.40b}$$

where $h_{t_1}(t_k)$ and $h_{t_2}(t_k)$ correspond to the integrator states and $h_{p_3}(t_k)$ is the level in the gravity pipe at location p_3. Note that out of the four available pipe state measurements we utilize only one for the residual generation. Furthermore, $Q_{t_1}(t_k)$ and $Q_{t_2}(t_k)$ correspond to the actuated flows at the two pumping stations, respectively. In this specific case study, we argue that using only one level sensor inside the pipe for the GP-based part of the model is sufficient enough as the sensor located at the downstream end of the pipeline captures the information to model how the $Q_{t_1}(t_k)$ pumpflows and the disturbance flows $q_{p_3}(t_k)$ enter and propagate in the system. Note that if the pipe dynamics are incorporated in the nominal model, a minimum of two sensors are required. This is because the PDE-based models need state measurements at two different locations to capture the attenuation and the delay inside the pipes.

Given the nominal model pre-identified on the level and estimated flow datasets, the contribution of the physically-based tank (and pipe model if included) can be subtracted from the state measurements. Hence the residuals $y \in \mathbb{R}^3$ can be constructed. Beyond the sensor availability, knowledge of the physical system plays a role in how efficiently we can train the non-parametric GP models with the generated residuals. To find the hyper-parameters of each GP model, it is unnecessary to use the entire training input vector defined in eq.(4.39), as this can lead to high dimensionality resulting in large computation.

Remark 3. *The dimension of each training input vector z is reduced via using the high-level layout information of the network. We introduce slicing matrices to map the training input vector z, such that each mapping defines which dimensions of the original training vector z influence the given residual, i.e.,*

$$S^1_{i_1,\dots,i_{n_1}} \in \mathbb{R}^{n_1 \times N_z}, \quad i \in \{4, 6, 7\}, \tag{4.41a}$$
$$S^2_{i_1,\dots,i_{n_2}} \in \mathbb{R}^{n_2 \times N_z}, \quad i \in \{3, 5, 7\}, \tag{4.41b}$$
$$S^3_{i_1,\dots,i_{n_3}} \in \mathbb{R}^{n_3 \times N_z}, \quad i \in \{4, 7\}, \tag{4.41c}$$

where i is the index of the predictor in the input vector z.

The feature selection from the original training set is visualized in Figure 4.14.

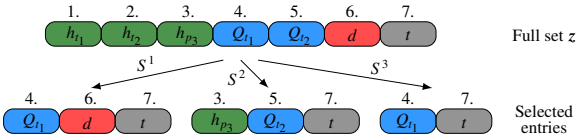


Figure 4.14: Feature selection with the slicing matrices. Source: [Balla et al., 2022c].

As an example, the first residual corresponding to the upstream tank state t_1 uses the predictors Q_{t_1} ($i = 4$), the rain forecasts d ($i = 6$) and the time t ($i = 7$). The selection of these predictors is well-aligned with the physical structure of the Experimental Study, as the water level variation in the upstream tank is influenced by the dry- and wet-weather disturbances (d accounting for rain forecasts and t for the periodic wastewater), and the corresponding pumpflows Q_{t_1} pumping out the volume from the tank. The GPs are then trained on data collected from operation under an on/off threshold-based controller using the squared exponential kernel.

The system identification problem is solved with the use of the `fitrgp` toolbox in `MATLAB`. The formulation of the GP-based system identification problem is further detailed in Paper E and Paper F.

4.5.2 Non-parametric System Identification Results

The results presented here are related to the modelling of wet- and dry-weather disturbances infiltrating into the network, and the volume and flow conservation dynamics inside the sewers. The presented results here describe real data extracted from our Experimental Study setup. To test the modelling capabilities of the combined parametric and non-parametric network model, the collected experimental data has been divided into training and validation sets. The GPs have been trained on 80% of the collected data while the rest was used for validation. Besides, two different tests have been carried out; where the simplified pipe dynamics are either included or excluded from the nominal system dynamics.

A particular case is when the simplified pipe model is included in the nominal dynamics $f(\cdot)$. In this case, the nominal pipe dynamics are pre-identified on the measured level and flow data mainly to detect the delays and approximate the propagated volumes induced by the pumping upstream. Figure 4.15 shows the system identification results for the simplified KW-based pipe model on data collected under nominal sewer operation, i.e., on/off. The results below are shown for the first (h_{p_1}) and last (h_{p_4}) sensor locations along the pipeline. As seen in Figure 4.15, the level sensor measurement at the upstream of the gravity sewer resembles the flow cycles due to pumps turning on and off. This is expected as there is no disturbance entering the sewer at the upstream location except the flow provided by the pumps.

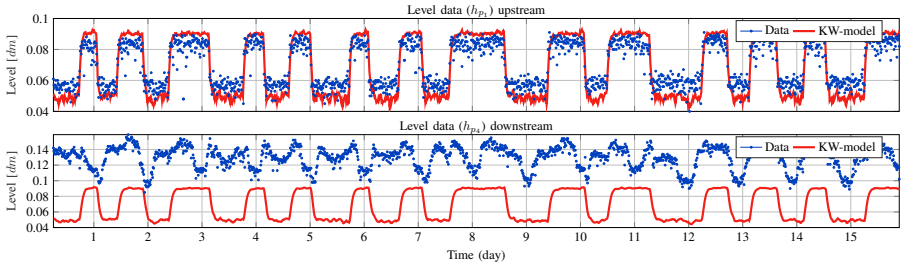


Figure 4.15: System identification results of water levels corresponding to the simplified KW-based model in Section 4.2.6 (Conceptual Alternatives).

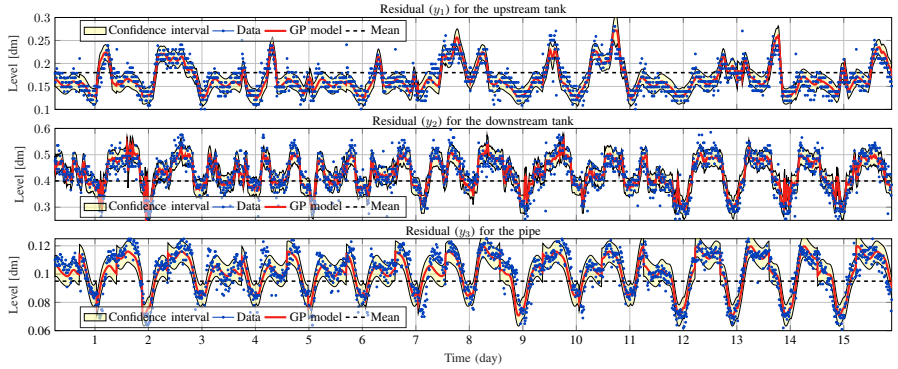
Nevertheless, the level sensor data shown in Figure 4.15 at the downstream end of the channel are lifted by the disturbance entering the sewer in the middle of the pipe. Note that since we only model the nominal dynamics governing the pumped flow propagating inside the pipe, the model is not expected to fit the downstream level data disturbed by the lateral inflows. However, we expect that the model output of the simplified KW-based model resembles the delayed and attenuated level variations induced by the pumpflows.

We provide the validation results of the residual prediction for the two cases, i.e., when the simplified KW-based pipe dynamics are included and excluded from the nominal model $f(\cdot)$. The results for both cases regarding the two residual for the upstream and downstream tanks (y_1 and y_2), and the water level in the pipe (y_3) are depicted in Figure 4.16.

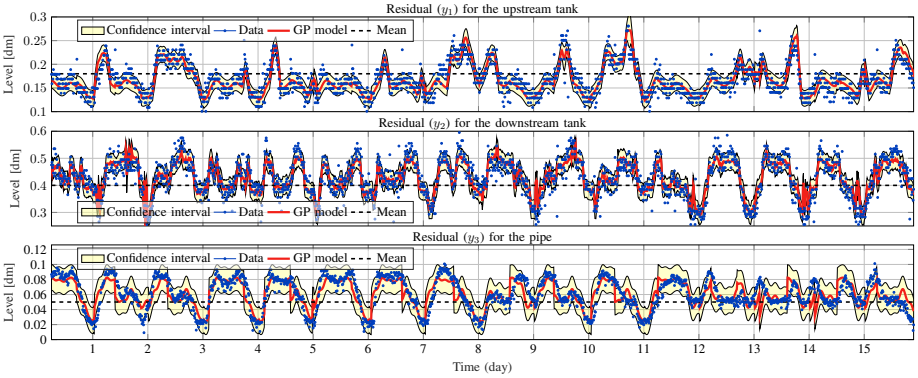
As seen from both figures, the GP model outputs match the observed data points. Moreover, the confidence interval characterized by the GP variance covers the distribution of the data points. (The variance in the data is primarily due to measurement noise and the precision of the level sensors.) As seen in residual y_1 corresponding to the level variations in storage tank t_1, subtracting the output of the nominal tank dynamics results in level variations following the combination of dry- and wet-weather disturbance flows entering the pipes. The diurnal level variation in y_1 corresponds to dry-weather, while the level peaks are due to the wet-weather flows. Residual y_1 are equal in both cases, as the upstream tank levels are not influenced by the nominal pipe dynamics.

The data describing residual y_2 varies due to the discharged pump flow upstream and the dry-weather flow discharged in the gravity pipe and propagating down to the t_2 downstream tank. Note that removing volume from the tank due to pumping is incorporated in the nominal dynamics, hence its effect is not visible on the residual data. Likewise, the first residual, residual y_2 is the same for the two cases.

Lastly, residual y_3 differs for the two cases. Note that in Figure 4.16a the level variation depicts the exact level measurements inside the sewer, while for the case in Figure 4.16b, the nominal pipe dynamics governed by the KW-based level propagation are subtracted from the measurement. The resulting



(a) Residual level validation with the GP model without the simplified KW-based pipe dynamics. Source: [Balla et al., 2022c].



(b) Residual level validation with the GP model with the simplified KW-based pipe dynamics.

Figure 4.16: Residual validation with the GP models.

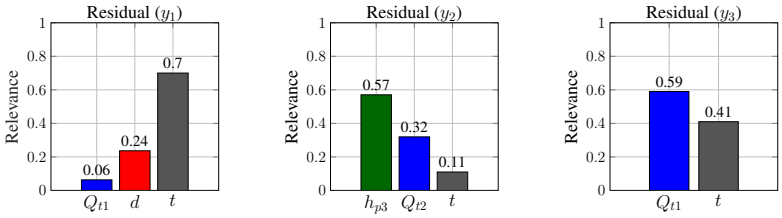
residual for the case depicted in Figure 4.16b has a lower mean because the volume added by the pumping is removed from the signal.

It is important to make sure that our non-parametric model captures the correlation between the selected features (inputs) and the generated residuals. To provide a measure of the correlation between the inputs and outputs, the relevance of the regressors is introduced, specifically for squared exponential kernels, i.e.,

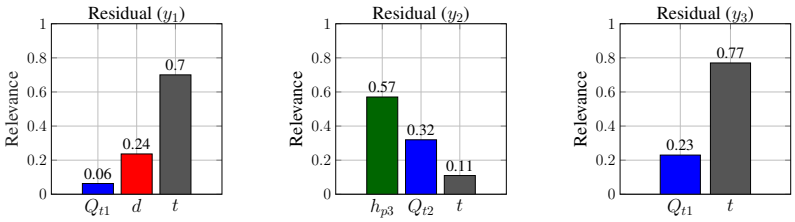
$$\overline{r}_i = \frac{\exp(-\sigma_{L,i})}{\sum_{j=1}^{N_L} \sigma_{L,j}}, \tag{4.42}$$

where $\overline{r}_i$ is the normalized relevance of the i^{th} predictor mapped by the slicing matrices defined in eq.(4.41) and N_L is the number of length-scale hyperparameters matching the dimension of the feature vector. The relevant input dimensions receive positive values between one and zero, while a value close to

zero indicates irrelevant input data. Simply stated, the length-scale parameter in each GP scales the input dimensions according to their relevance, which is quantified by the $\overline{r}$ normalized relevance vector. The relevance comparison for the two cases are shown in Figure 4.17.



(a) Input relevance for the GP models without the simplified KW-based pipe dynamics. Source: [Balla et al., 2022c].



(b) Input relevance for the GP models with the simplified KW-based pipe dynamics.

Figure 4.17: Input relevance for the GP models, where Q_t denotes the aggregated pump flow corresponding to the pumping stations, h_p denotes the pipe states, d stands for rain forecasts and t is the time recurring at every 24 hours sampled according to the control intervals.

As seen, the two tank states are equivalent in the two test cases, as the nominal tank dynamics are included in both. The relevance bars corresponding to residual y_1 show that time t has a very high relevance and the disturbance forecast indicates the rain precipitation has the second-highest relevance. This is in line with our expectation, as the majority of the residual y_1 incorporates information about the dry-weather flows discharging to the tank, while wet-weather infiltration occurs less frequently. As expected, the relevance of the pumps removing volume from the tank is negligible since our nominal integrator dynamics describe the level variation in tanks quite precisely.

The relevance bars corresponding to residual y_2 show that the level variations in the sewer pipe h_{p_4} have the highest impact on the residual created from the measured downstream tank level. This is expected, as the only source of discharge to the tank is the pipe connecting it with the upstream pumping stations, where lateral disturbance inflows are infiltrating into the pipe. Despite the nominal tank dynamics, our model shows some impact from the pump flows Q_{t_2} removing volume from the tank, which is partly because under high-load periods, the downstream pumping station turns on resulting in

similar characteristics as in h_{p_3} and Q_{t_2}.

Lastly, the relevance of the input data differs regarding residual y_3 corresponding to the level measured in the sewer pipe. As seen in Figure 4.17a, the relevance of the upstream pump flows propagating down to the channel is high in comparison to the second case depicted in Figure 4.17b, where the simplified KW-based pipe dynamics are included in the nominal system model. Hence, predicting the residuals with the latter GP model, the resulting pipe level variations are less influenced by inlet flows of the upstream pumps, as a large part of the volume prediction is taken care of by the nominal dynamics. Furthermore, we do not expect the relevance to be zero as the pre-identified KW-based depicted in Figure 4.15 shows that the resulting nominal model cannot capture the level propagation with the simplified model perfectly, as compared to the original non-linear grey-box models introduced in Section 4.2.4 (Kinematic Wave Model).

4.5.2.1 Real-world Tests

The real-world evaluation of the proposed modelling methodology described in Section 4.5 (Non-parametric Modelling Results) has been carried out in a pilot project in collaboration with Ishøj Wastewater Utility in Ishøj, Denmark. The Ishøj Case Study is a stormwater network, where five level sensors have been deployed to detect the effect of rain on the level variations in the system. A deployed level sensor is shown in Figure 4.18.

Figure 4.18: Level sensor deployed inside a manhole. The data in the Ishøj Case Study have been obtained through level sensors equipped with battery and wireless communication, sending data at every 2 minutes to a our data base.

Unfortunately, this network has neither active control, nor the water ponds are utilized to retent water upstream in ther gravitational sewer network[2], therefore our methods have been limited to predicting the water levels based on the rain intensities detected through the test period. Because of that, we

[2]The lack of utilization of storage capacity has been one of the key results during this pilot study, providing insightful and cruical information for the utility where and how to improve the operation of their storm water network.

tested the Gaussian process regression without generating the residuals, as we could not subtract any nominal dynamics in this specific case study. The validation results and the data for the level prediction is shown in Figure 4.19.

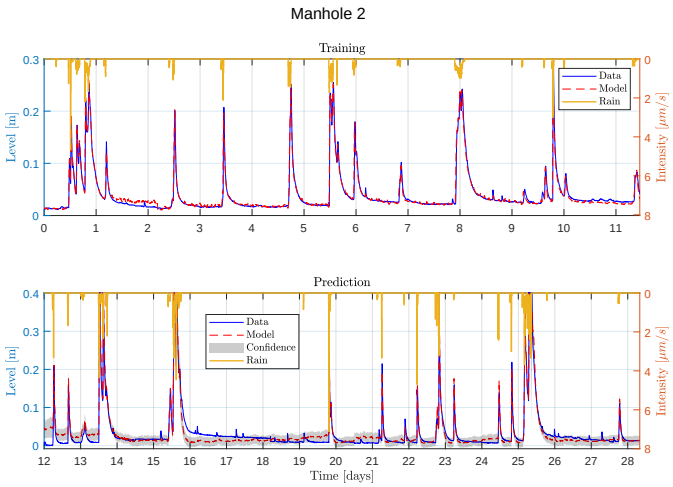


Figure 4.19: Training (above) and validation (below) for the level variations induced by rainfall in Manhole 2.

Note that in the case above we did not use the entire dataset for training as we deliberately wanted to show the predictions under unexplored data, i.e., under rain events which have significantly different nature, hence uncovered in the training data set. A situation of such happens between Day 25 and 26, which is visualized in Figure 4.20.

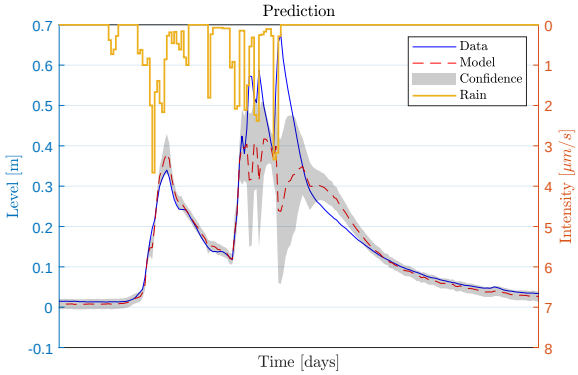


Figure 4.20: Prediction of a long and high-intensity rain event showing the prediction deficiency due to unexplored data.

As shown in Figure 4.20, due to the length of the rain event the channel does not have time to dry out, hence in the middle of the day even short peaks of high-intensity rainfall triggers a 0.6 meter water level in the pipe. As seen in Figure 4.19, the training data only covers events covering the $0 - 0.3$ meter water level range. Hence, predicting with the historical data results in degraded validation performance and wider confidence intervals, as the data is unexplored.

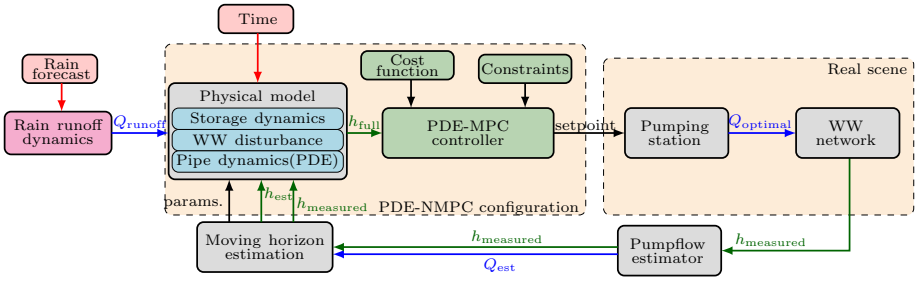
5 Sewer Volume Control

This chapter provides an overview of volume optimization in sewer networks based on the controller architectures proposed in Paper C, Paper D, Paper E and Paper F.

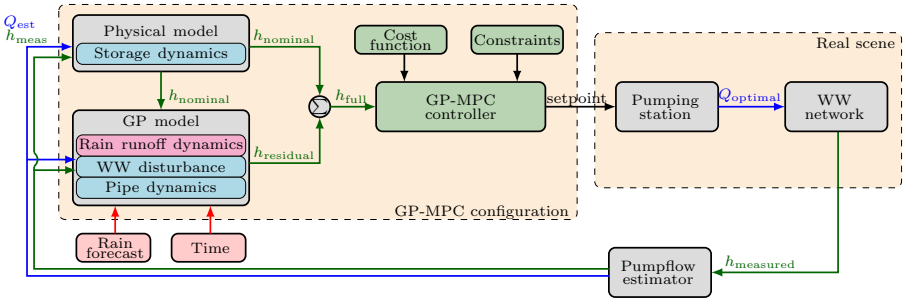
5.1 Control Architectures

The modelling approaches presented in Chapter 4 (Sewer Dynamics Modelling) serve as the baseline for the model-based predictive controllers developed in this chapter. Since this study is focused on the control-oriented modelling of both the hydraulic and the hydrologic processes in sewer networks, several architectures for predictive control have been explored throughout the thesis. Specifically, Paper C focused on the physically-based derivation of in-sewer flow propagation by capturing the effect of the actuator, waste, and groundwater disturbance flow via level sensors deployed in the system. This was mainly detailed in Section 4.2.2 (Volume Conservation in Pipes). The control architecture corresponding to Paper C is shown in Figure 5.1a, hereinafter referred to as the PDE-based MPC controller. The part of the study covering this type of control architecture focuses on a physically-based model structure for both the sewer flows and the volume conservation in storage elements. In this architecture, however, attention has not been given to the hydrologic modelling for the rain infiltrating the system. Consequently, the *Rain runoff dynamics* block is not part of the controller, as the runoff dynamics engine of `MIKE Urban` is utilized as a disturbance generator to provide flow inputs based in rain forecasts to the control-oriented models governing the sewer processes.

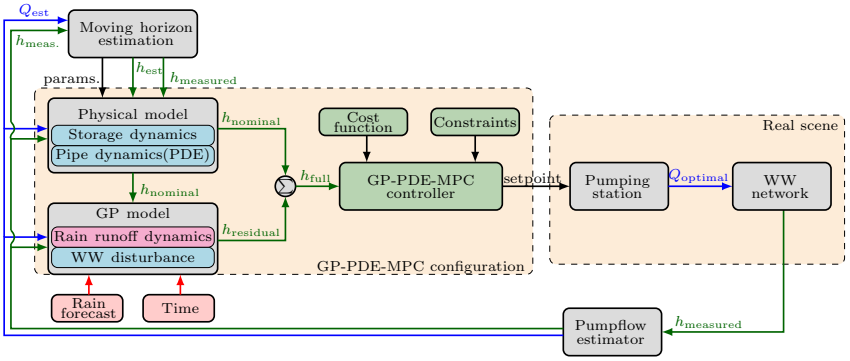
Opposed to the PDE-based MPC controller, a slightly different architecture is proposed in Figure 5.1b. In this architecture, we utilize the non-parametric modelling methods detailed in Section 4.4 (Non-parametric Modelling of Sewer Hydrology) and the controller corresponding to Paper F. Hereinafter we refer to it as the GP-based MPC controller. In this setup, the rain runoff dynamics along with the pipe and wastewater disturbances have been collectively modelled by nonparametric Gaussian processes acting as additive dynamics to the physical volume conservation models. In this approach, the focus has been shifted from the physically-based model structure to the data obtained through our level measurements. The main improvement illustrated in Figure 5.1b is regarding the placement of the *Rain runoff dynamics* box inside the nonparametric model part of the controller and thereby making the configuration independent from any high-fidelity runoff engines. The benefit of fitting the combination of physical and nonparametric models to residual data is the *Plug and Play* nature of this control architecture, while the main drawback is the loss of physical model insight and modelling structure.



(a) Predictive control architecture based on the Saint-Venant PDE pipe, volume-based storage and high-fidelity rain runoff models, combined with moving horizon parameter and state estimation.



(b) Predictive control architecture based on Gaussian process and volume-based storage models.



(c) Predictive control architecture combining the Saint-Venant PDE-based pipe, volume-based storage and Gaussian process models using moving horizon estimation for updating the pipe model.

Figure 5.1: Control architecture comparison.

Lastly, the control architecture depicted in Figure 5.1c is the combination of the former two approaches. The main goal with the combined GP-PDE-based MPC controller approach is to utilize the model structure of all the hydraulics describing the flow propagation and volume conservation inside the storage elements and conduits. On the other hand, we restrict ourselves to only using the nonparametric Gaussian processes to describe the dynamics of rain and wastewater infiltrating into the system. Consequently, the resulting control architecture depends less on the forecast data quality as the nominal model parts are robust enough to handle the dry-weather flow disturbances and the flows created by the controllable assets in the system. Hence, the residuals created between the output of the volume conservation dynamics and the level measurement resemble the hydrologic effects imposed on the sewer and the unmodelled dynamics due to the simplicity of the physically-based part of the model. A drawback of the method includes the fact that physically-based flow propagation in pipe models imposes chain model-type dynamics with a large number of states, meaning prohibitively many sensor installations. (In case all states are measured, which is typically not a requirement.) Besides, the purpose of including the pipe model is to predict the pumpflow propagation travelling down the channel, as it is depicted in Figure 4.15. For this reason, an observer is necessary to recreate the pipe model states at locations where not only pumped flow but disturbance flow is present on the measured water level signals.[1] Therefore, similarly as for the PDE-based MPC controller, the estimation of the pipe states needs to be considered.

In the following, the aim is to introduce the above-mentioned control architectures, present the objectives of the control task, and show the resulting closed-loop control performances.[2]

5.2 Predictive Control in Sewers

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 simple to incorporate into the optimization problem. In MPC, we formulate an open-loop finite-horizon optimization problem, which we solve in a receding horizon fashion. The objectives of the control problem and the operational and physical constraints need to be carefully chosen to obtain the desired control performance. According to the model dynamics formulated as the constraints and the disturbance forecasts, the MPC algorithm optimizes the input variables over a given prediction horizon H_p of chosen length. In general, at time step t_k the following steps are undertaken

[1] Note that measuring water levels resembling the inlet flow provided by the actuators is only possible if the sensor is deployed very close to the discharge point without any lateral inflows in its vicinity.

[2] The predicting capabilities and residual generation with the GP-PDE-based MPC controller has been tested in Section 4.2.6 (Non-parametric System Identification Results), however, the closed-loop control implementation is reserved for future studies.

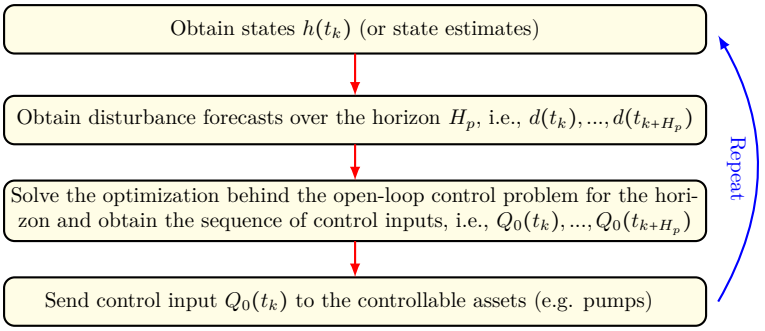


Figure 5.2: Solving the optimization in a receding horizon fashion.

As seen in Figure 5.2, the receding horizon strategy takes full advantage of the real-time measurements and the forecasted disturbances. As visualized in the third step, after solving the optimization problem, only the first step of the obtained control sequence is applied to the network. After letting the system respond to this action, the entire process repeats itself, i.e., new feedback measurements are taken, and the forecasts are updated based on the latest information.

In wastewater applications, constraints are typically formed on the physical limits of hydraulic storage elements and the physical capacity of the controllable assets, e.g., the maximum obtainable flow by a pumping station. The disturbances are considered as the dry- and wet-weather flow infiltration, among which the latter is highly stochastic. The forecasting capabilities of the rain intensity at a specific location is a research topic on its own, hence not detailed in this thesis.

In this work, we formulate constraints on the states, i.e., the water levels, and on the input variables, i.e., the controllable flow provided by the actuators. In all our tests and experiments, the actuators are considered to be pumps connected in parallel. Besides, in Paper F we introduced an operational constraint for keeping safety levels in storage tanks instead of simply minimizing the levels to a reference at all times. Keeping the water level in storage elements is particularly important as before a storm event, the storage capacity needs to be utilized to its full extent. Due to the stochastic nature of rain forecasts, uncertainty is inherently present in the evolution of the governing water levels. Therefore, the controller aims to reject the dry- and wet-weather disturbance flows and attempts to keep the water levels within a safety region in tanks so that the system is always prepared for unexpected storm events. In this study, we adopt some ideas from [Grosso et al., 2014] and introduce the previously-mentioned operational constraint for safety water levels. The state constraints utilized in our study are illustrated in Figure 5.3.

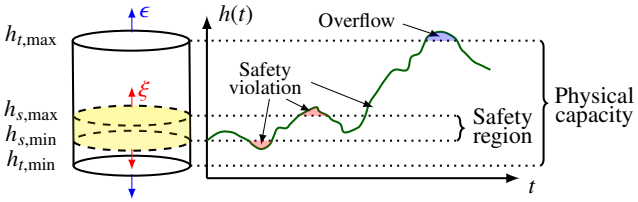


Figure 5.3: Safety and capacity constraints, where the blue arrows denote constraint relaxation for overflow (ϵ) and red arrows are constraint relaxations for safety violation (ξ). Source: [Balla et al., 2022a].

As shown in the figure above, while the physical limitations of, e.g., storage tanks are evident, the determination of the safety bounds is less clear. In our work, we argue that the placement of the safety bounds is best in the lower regions of the storage tanks, as the system remains emptied and prepared in case of unexpected disturbances.

Besides the constraints, the key component of the control architectures proposed in Section 5.1 (Control Architectures) is the formulation of the cost function. The formulation of the control problem in all our studies relates to the manipulation of water levels (volumes) to avoid undesirable overflows and water surges outside the sewer pipes. Our main focus is on rejecting the highly stochastic wet-weather (rain-runoff) and dry-weather (household and groundwater) loads while aiming to avoid constraint violations. In this work, we focused mainly on the following operational and management criteria while designing closed-loop control (listed in decreasing order of priority)

I. Minimize overflow in storage elements,

II. Minimize the variation on the inlet flow to the treatment plant,

III. Minimize the safety volume violation in storage elements,

IV. Minimize the water level in storage elements and manholes,

V. Minimize the control action of pumps.

Note that the above list is not limited to the proposed objectives. For instance, the objectives for a specific case study might vary according to the infrastructure design. To solve the control problem, we combine the objectives and solve a multi-objective optimization problem. Throughout the thesis, the overall cost defined in MPC has been formulated as the linearly weighted sum of quadratic terms, i.e.,

$$\mathcal{L}(t_k) = \underbrace{w_1 \|\epsilon(t_k)\|^2_{\Omega_1}}_{\mathcal{L}_{I.}} + \underbrace{w_2 \|\xi(t_k)\|^2_{\Omega_2}}_{\mathcal{L}_{III.}} + \underbrace{w_3 \|h(t_k)\|^2_{\Omega_3}}_{\mathcal{L}_{IV.}} + \underbrace{w_4 \|\Delta Q_0(t_k)\|^2_{\Omega_4}}_{\mathcal{L}_{V.}} \tag{5.1}$$
$$+ \underbrace{w_5 \Big(Q_w(t_k) - \frac{1}{H_p} \sum_{k=0}^{H_p - 1} Q_w(t_k) \Big)}_{\mathcal{L}_{II.}}$$

where the specific terms relate to the pre-defined control criteria above, i.e., $\mathcal{L}_I$ describes the overflow penalty, $\mathcal{L}_{II}$ relates to the variation of inlet flow to the treatment plant over the H_p prediction horizon, $\mathcal{L}_{III}$ is the penalty on the safety bound violations, $\mathcal{L}_{IV}$ is a state penalty, and $\mathcal{L}_V$ is the penalty on the input change, accounting for smooth and slow system response.[3] Moreover, $\epsilon(t_k) \in \mathbb{R}^{2N_t}$ is the vector of slack variables measuring overflow, $\xi(t_k) \in \mathbb{R}^{2N_t}$ is the vector of variables measuring the violation of the safety constraint, $h(t_k) \in \mathbb{R}^{N_t+N_p}$ is the vector of states representing water level in pipes and storage tanks, $Q_0(t_k) \in \mathbb{R}^{N_Q}$ is the vector of manipulated flow variables, and lastly $Q_w(t_k) \in \mathbb{R}$ is the discharged flow to the treatment plant. The weighting constants denoted by w are prioritizing the weight of the different objectives and Ω is diagonal and $0 \leq \Omega \leq I$, where I is the identity matrix of suitable dimensions.

5.3 PDE-based Predictive Control

The PDE-based MPC controller uses the KW-based pipe model and the storage model based on simple volume conservation, detailed in Section 4.2.2 (Volume Conservation in Pipes) and Section 4.2.1 (Volume Conservation in Tanks), respectively. The study has been carried out on the Simulation Study C with the use of four pumping stations equipped with storage tanks, discharging to a wastewater treatment plant. The specific dimensions of the study are the following; $N_Q = 4$, $N_t = 4$, and $N_p = 17$, denoting the number of pumping stations, storage tanks, tank and pipe states, respectively.

In any optimization problem, the initial states are either obtained through measurements or state estimation. As the PDE-based MPC controller uses the KW-based pipe models, typically more states are considered than it is economically feasible to measure. The reasoning for this is mainly to obtain a good flow attenuation and transport-delay modelling performance, given the spatial and time discretization properties of the model. Therefore, state estimation is necessary to reconstruct the entire state vector h out of a few measurement points. The study has been carried out by utilizing online moving horizon estimation for both the states and the parameters regarding the pipes. To this

[3] Note that smooth and slow system response in our application is desired, as sudden changes in the control action can degrade the lifetime of, e.g., pumps.

end, we solve the state and parameter estimation problems jointly with a fixed-size horizon at certain time intervals. The development of this estimator is not detailed in this summary, however, the interested reader may consult Paper C for further details.

5.3.1 Optimization Problem

Let us distinguish between the states governing pipes and tanks and introduce $h(t_k) = [h_t^\intercal(t_k) \quad h_p^\intercal(t_k)]^\intercal$ as the full state vector. Besides, let the vectors $q_t(t_k) \in \mathbb{R}^{N_{q_t}}$ and $q_p(t_k) \in \mathbb{R}^{N_{q_p}}$ denote the disturbances corresponding to tank and pipe states, respectively. Moreover, the prediction horizon for the optimization problem is divided into two parts, where the near-future is predicted with a 10 minutes accuracy for up to two hours, and the far-future is predicted at each hour summing up the entire horizon to an entire day.[4] Note that for the sake of simplicity, we present the problem with the combined prediction horizon H_p. The pipe model parameters are re-identified every six hours with a data batch of the last two days, while the states are estimated at every $T_s = 10$ minutes control step for providing the initial conditions for the optimization problem based on a two days time window. The nonlinear optimization problem behind the PDE-based MPC controller strategy is given as

$$\underset{\substack{\Delta Q_0(0),\ldots,\Delta Q_0(H_p-1)\\ \epsilon(0),\ldots,\epsilon(H_p-1)}}{\text{Minimize}} \sum_{k=0}^{H_p-1} \mathcal{L}_I\big(\epsilon(t_k)\big)+\mathcal{L}_{II}\big(Q_w(t_k)\big) + \mathcal{L}_{IV}\big(h(t_k)\big) + \mathcal{L}_V\big(\Delta Q_0(t_k)\big) \tag{5.2a}$$

subject to the KW-based pipe and Fourier disturbance dynamics

$$h_p(t_{k+1}) = F_{\theta,\lambda}^{\mathrm{KW}}\big(h_p(t_k), Q_0(t_k), q_p(t_k)\big), \tag{5.2b}$$

$$Q_{N_x}(t_k) = \theta_2 f(h_{p,N_x}(t_k), \theta_3), \tag{5.2c}$$

to the storage dynamics

$$h_t(t_{k+1}) = H\big(h_t(t_k), Q_0(t_k), q_t(t_k), Q_{N_x}(t_k)\big), \tag{5.2d}$$

to integral action

$$Q_0(t_{k+1}) = Q_0(t_k) + \Delta Q_0(t_k), \tag{5.2e}$$

and to state, input and output constraints

$$f_V(\underline{h}_t) + \epsilon(t_k) \le f_V\big(h_t(t_k)\big) \le f_V(\overline{h}_t) + \epsilon(t_k), \tag{5.2f}$$

$$0 \le h_p(t_k) \le \overline{h}_g, \tag{5.2g}$$

$$\underline{Q}_0 \le Q_0(t_k) \le \overline{Q}_0, \tag{5.2h}$$

$$0 \le Q_w(t_k) \le \overline{Q}_w, \tag{5.2i}$$

[4] The prediction horizon of one day is relevant when the objective $\mathcal{L}_{II}$ is considered for smoothing the inflow to the treatment plant. Hence, the daily variations of wastewater flow can be treated, especially in dry-weather periods.

where the pipe dynamics in eq.(5.2b) and eq.(5.2c) are equivalent to the ones presented in Section 4.2.2 (Volume Conservation in Pipes). Besides, the storage dynamics are equivalent to the tank model presented in and Section 4.2.1 (Volume Conservation in Tanks). Note that the presented optimization problem used in Paper C does not use the objective $\mathcal{L}_{III}$ governing the safety volume violations, however, it is expected to improve the network operation, especially under dry-weather periods. Furthermore, $Q_w(t_k)$ in eq.(5.2i) denotes the inflow to the treatment plant, which is equivalent to the outflow Q_{N_x} in the pipe link discharging to the root of the network. The nonlinear level to volume conversion is handled outside of the optimization, where $f_V(\cdot)$ in eq.(5.2f) is a piecewise linear function that allows to track the overflow slack ϵ in terms of volumes, i.e., the actual overflow [Balla et al., 2022d]. Note that the overflow slack lifts the upper and lower bounds of the physical volume capacities in storage tanks, thereby keeping track of the excess storage.

Remark 4. *In case of overflows the excess volume in our optimization for the PDE-based MPC controller strategy escapes the system, i.e., it is considered as spilled after the overflow. It is ensured via constraining the slack variables such that* $0 \leq \epsilon(t_k)$ *at all time steps [Balla et al., 2022d].*

In the next section, we provide some closed-loop control results of the PDE-based MPC controller strategy.

5.3.2 PDE-based Control Results

The closed-loop control results presented in this subsection are partially based on Paper C. The results of implementing the PDE-based MPC controller strategy aim to show the benefits of using a physically-based network model, where the assimilation of water level data is used to obtain the model parameters and to keep the model updated. The proposed method is compared to the baseline controller introduced in Section 3.4 (Benchmark Controller), which is based on a two-point control method turning the pumps on and off, often used in industry. Furthermore, to evaluate the performance of the controller under both dry- and wet-weather, two days are selected under which overflows are triggered due to the insufficient capacity of the sewer network and the storage tanks. In our results, all Ω scaling values are set equal in the individual terms of the objective function, meaning that none of the four pumping stations are prioritized over each other. For instance, the filling sensitivity of each storage tank at the pumping stations and the penalty on the overflows are equal at each control node of the network. As the overflows are not fully avoidable in the selected period, the overall goal of the controller is to reduce the aggregated volume spilled at all pumping stations together. The overflow comparison between the baseline and the PDE-based MPC controller strategy is shown in Figure 5.4 and the numerical results in the high-fidelity simulator of the Simulation Study C is shown in Figure 5.5 and Figure 5.6, respectively. Note that each column of the results in Figure 5.5 and Figure 5.6 correspond to the disturbance, overflow,

state and pump flow of each separate pumping station s_1, s_2, s_3 and s_4, among which s_2 is the one equipped with large retention capabilities. Furthermore, s_2 is the pumping station discharging the collected sewage to the treatment plant.

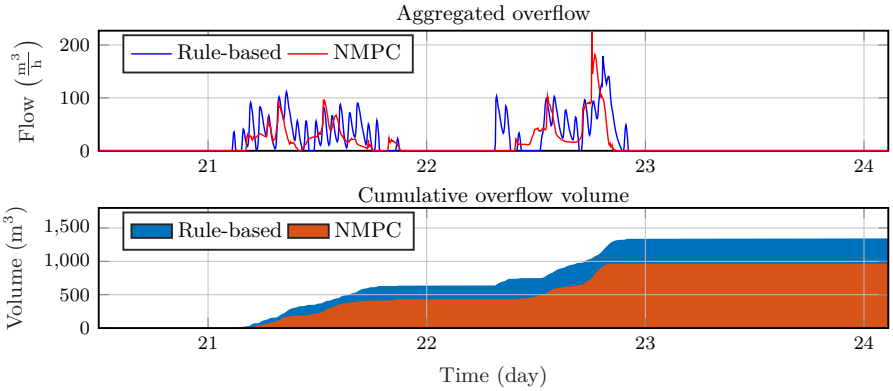


Figure 5.4: Overflow comparison in the full-scale network for the PDE-based predictive controller in closed-loop. Source: [Balla et al., 2022d].

As shown in Figure 5.4, the proposed control strategy results in approximately 28% cumulative overflow volume decrease compared to the baseline on/off controller under the same meteorological load and the same time period.

As depicted in Figure 5.5 and Figure 5.6, the disturbance signals used in the HiFi simulator are historic rain and wastewater inflows. Although we do not consider the stochastic evaluation of the states in this case study, uncertainty is introduced in the disturbance signals. To that end, we generate 10 different disturbance scenarios by adding normally distributed random data on top of the historic events [Balla et al., 2022d]. As shown in Figure 5.5(a,b) and Figure 5.6(i,j), the forecast provided to the predictive controller can vary between the values shaded by the ensemble of different rain and wastewater flow forecasts.

To show the deviation between the controller and the actual states retrieved by the HiFi simulator, we indicated the predictions by dashed red line in Figure 5.5(a,b) and Figure 5.6(i,j). Note that the upper constraints corresponding to the physical limits of the tanks are violated under overflow events, as the slack variables ϵ are approximating the overflows by increasing the volume in the storage tanks. Besides, the lower bounds can also be violated due to the uncertainty introduced in the disturbance, typically in cases when the actual inflow is higher than indicated by the forecasts. This could potentially result in the dry-run of the sewer pumps.

The main improvement achieved by our controller in contrast to the simple rule-based strategy is best illustrated on the states and the mitigation of overflows in Figure 5.5(c,d) and Figure 5.6(k,l).

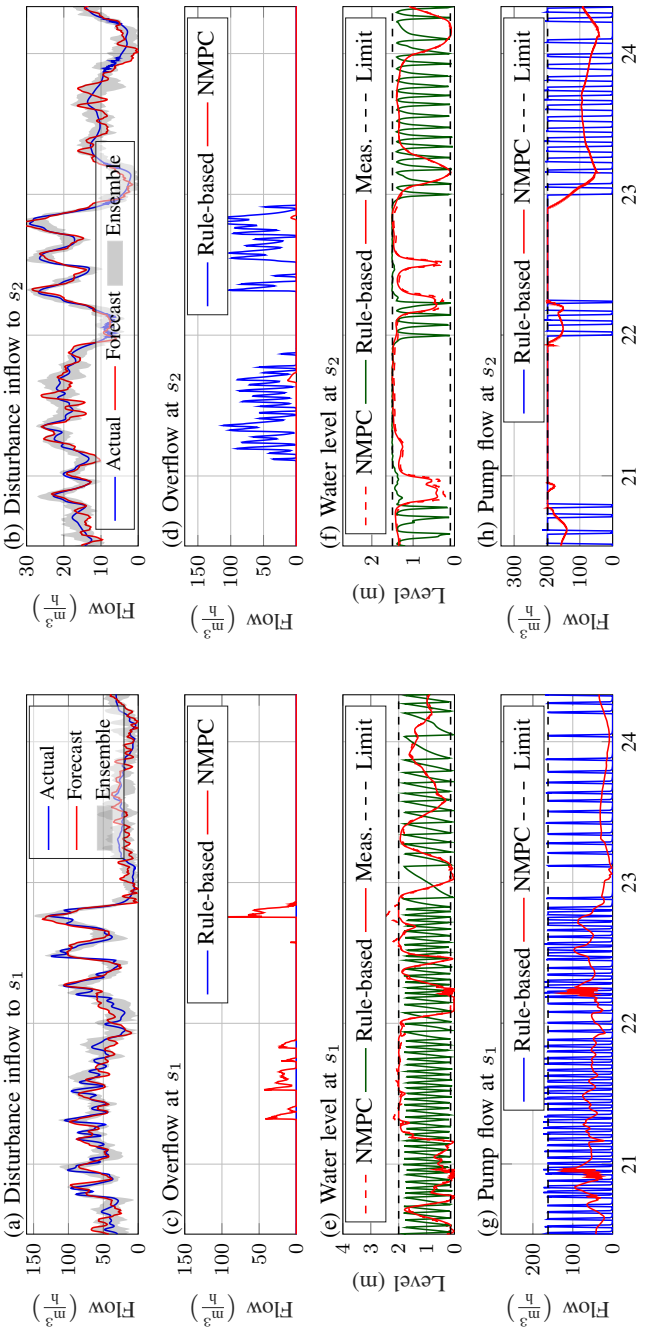


Figure 5.5: PDE-based MPC results with moving horizon estimation compared with rule-based control. (a), (b) show disturbance inflows $q_{s,i}$, (c), (d) show overflows $\epsilon_{s,i}$, (e), (f) show states (water levels), while (f), (h) show pump flows for each $i \in \{1, ..., N_s\}$ station, respectively. Source: [Balla et al., 2022d].

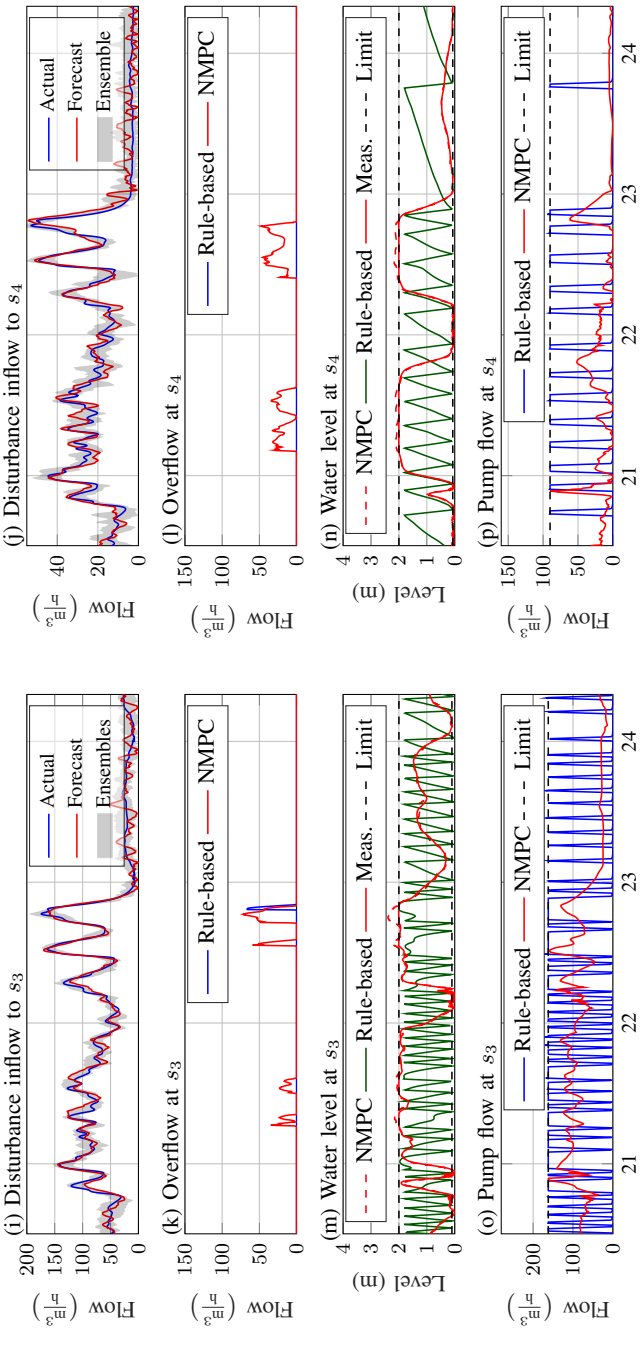


Figure 5.6: PDE-based MPC results with moving horizon estimation compared with rule-based control. (i), (j) show disturbance inflows $q_{s,i}$, (k), (l) show overflows $\epsilon_{s,i}$, (m), (n) show states (water levels), while (o), (p) show pump flows for each $i \in \{1, ..., N_s\}$ station, respectively. Source: [Balla et al., 2022d].

As seen, the predictive controller overflows the upstream tanks (s_1, s_2, s_3 and s_4) rather than overloading the downstream collection tank at s_2. Under these periods, all storage nodes are prepared by being emptied before the load on the system increases and therefore the overflows in comparison to the rule-based case are intentional and coordinated between the pumping stations [Balla et al., 2022d]. As shown in Figure 5.5(c,d) and Figure 5.6(k,l), the timing of the overflows is also shifted in time as the upstream storage nodes attempt to hold back the volume until their capacity allows [Balla et al., 2022d].

5.4 GP-based Predictive Control

The GP-based MPC controller uses the Gaussian process regression and storage model based on the simple volume conservation, detailed in Section 4.4.1 (Gaussian Process Modelling) and Section 4.2.1 (Volume Conservation in Tanks), respectively. The study has been carried out on the Experimental Study, emulating a combined wastewater network with pumping stations interconnected by gravity-driven sewer pipes. The laboratory setup is equipped with level sensors in both the gravity lines and the storage tanks. Specifically, the dimensions of the given study are the following; $N_Q = 2$, $N_t = 2$, and $N_p = 1$, meaning that we have two states corresponding to the storage tanks and one state to the water level in the gravity pipe. Although four level sensors are available on our setup, only one of them is utilized in control. Specifically, we use the level sensor placed downstream to the lateral inflow point since we aim to learn the hydrological inflows and their impact on the sewer dynamics by observing the variations of the water levels in the pipe. In contrast to the PDE-based MPC controller, this implementation does not consider PDE-type chain dynamics for the sewer pipes. The attention is rather shifted to the data and the selection of the feature vectors, as described in Section 4.5 (Non-parametric Modelling Results). The setup of the modelling and data structure is otherwise equivalent as described in Section 4.5 (Non-parametric Modelling Results).

5.4.1 Optimization Problem

Similarly to the PDE-based MPC controller, the slack variables ϵ representing overflows for the state constraint relaxation are decision variables as well as the input, i.e., the flow provided by the pumps. Such optimization problem is treated in Paper E, and the improved version in Paper F in which an additional decision variable ξ is introduced, which is penalized when the safety volume in the storage tanks is violated. Besides, in this study, we did not consider the operational objective $\mathcal{L}_{II}$ for the penalization of the inflow variation to the treatment plant.

Introducing the non-linear kernel to approximate the covariance of the Gaussian Processes, furthermore, the propagation of uncertainty results in stochastic system states, i.e., normally distributed states. As the decision

variables Q_0, ϵ and ξ are treated deterministic and the water treatment objective $\mathcal{L}_{II}$ are excluded from this study, the only stochastic term remaining in eq.(5.1) is $\mathcal{L}_{IV}$ regarding the system states. We take the expected value of the stochastic variables, i.e.,

$$\mathbb{E}\Big\{\mathcal{L}_{IV}(h(t_k))\Big\} = \mathbb{E}\Big\{w_3||h(t_k)||^2_{\Omega_3}\Big\} = w_3\big[||\mu_h(t_k)||^2_{\Omega_3} + \mathrm{tr}\{\Omega_3\Sigma_h(t_k)\}\big], \quad (5.3)$$

where tr{} is the trace operator and Σ_h is the co-variance of the states. As seen from eq.(5.3), the variance of the states becomes a state variable in our objective function and, therefore, in our optimization problem.

Bringing together the approximation via the GPs, taking the expected value of the objective function and treating the state constraints probabilistic, the optimization problem behind the GP-based MPC controller is given by

$$\underset{\substack{\Delta Q_0(0),\dots,\Delta Q_0(H_p-1)\\ \epsilon(0),\dots,\epsilon(H_p-1)\\ \xi(0),\dots,\xi(H_p-1)}}{\text{Minimize}} \sum_{k=0}^{Hp-1} \mathcal{L}_I\big(\epsilon(t_k)\big) + \mathcal{L}_{III}\big(\xi(t_k)\big) + \mathcal{L}_V\big(\Delta Q_0(t_k)\big) \quad (5.4a)$$
$$+ w_3\big[||\mu_h(t_k)||^2_{\Omega_3} + \mathrm{tr}\{\Omega_3\Sigma_h(t_k)\}\big],$$

subject to the nominal dynamics combined with the GPs

$$\mu_h(t_{k+1}) = f\Big(\mu_h(t_k), Q_0(t_k)\Big) + B_p\mu_{GP}(t_k), \quad (5.4b)$$

$$\Sigma_h(t_{k+1}) = \Big[\nabla_h f\Big(\mu_h(t_k), Q_0(t_k)\Big), B_p\Big]\Sigma(t_k)\Big[\nabla_h f\Big(\mu_h(t_k), Q_0(t_k)\Big), B_p\Big]^\top, \quad (5.4c)$$

where the GP properties are given by

$$\mu_{GP}(t_k) \text{ and } \Sigma_{GP}(t_k) \text{ are according to eq.(4.28)}, \quad (5.4d)$$
$$\Sigma(t_k) \text{ is according to eq.(4.34)}, \quad (5.4e)$$
$$\mu_h(t_0) = h(t_k), \quad \Sigma_h(t_0) = 0, \quad (5.4f)$$

moreover subject to integral action

$$Q_0(t_{k+1}) = Q_0(t_k) + \Delta Q_0(t_k), \quad (5.4g)$$

and to state and input constraints

$$\mathbb{P}\big\{\underline{h} + \epsilon(t_k) \le h_t(t_k) \le \overline{h} + \epsilon(t_k)\big\} \ge \alpha, \quad (5.4h)$$
$$\mathbb{P}\big\{\underline{h}_s - \xi(t_k) \le h_t(t_k) \le \overline{h}_s + \xi(t_k)\big\} \ge \alpha_s, \quad (5.4i)$$
$$\underline{Q}_0 \le Q_0(t_k) \le \overline{Q}_0, \quad (5.4j)$$
$$\epsilon(t_k) \ge 0, \quad \xi(t_k) \ge 0, \quad (5.4k)$$

where the equality constraints in eq.(5.4b) and eq.(5.4c) imposed by the combined nominal and GP dynamics are provided in the *mean-variance* dynamics

form, where μ_h and Σ_h are the mean and variance describing the evolution of the system states and characterizing the uncertainty, respectively. The equality constraints representing the network dynamics are equivalent to the state transition dynamics introduced in Section 4.2 (Parametric Modelling of Sewer Hydraulics) and Section 4.4 (Non-parametric Modelling of Sewer Hydrology). Besides, when we solve the above optimization problem, we assume that the variance at the initial step is zero as indicated by eq.(5.4f).

As shown in eq.(5.4h) and eq.(5.4i), the box constraints for the states are stochastic, meaning that there is a probability measure describing whether the constraint is fulfilled or not. The state constraints are fulfilled with a confidence level of α and α_s. Since the evolution of the states follows a Gaussian distribution, the above state constraints can be re-written in a tractable form to solve the optimization. In our work, we utilize chance constraints, which are further detailed in Paper E and Paper F.

In the next section, we provide some closed-loop control results of the GP-based MPC controller strategy.

5.4.2 GP-based Control Results

The closed-loop control results presented in this subsection are partially based on Paper E and Paper F. The evaluation of the controller has been carried out in the Experimental Study, where the corresponding time steps, disturbances, and sampling times represent a $1:80$ scale of a real-world combined sewer network. The controller has been implemented with an $H_p = 20$ step prediction horizon, corresponding to a four-hour-long prediction in real life. From the implementation point of view, the execution of the *mean-variance* dynamics in eq.(5.4b) and eq.(5.4c) depends on the number of data points used for the prediction. This is because μ_h and Σ_h are conditioned on the observed data, and therefore evaluating eq.(5.4) has a growing computational cost with the number of points. In this experimental study, we utilized $M = 80$ points for the predictions, using a simple point selection algorithm that chooses points close to the previously predicted horizon. The selection criteria and the algorithm is further detailed in Paper E and Paper F. Besides, the presented controller is launched for online tests after the GP-based model is pre-trained on 60 days of data obtained via operation of the baseline controller, i.e., the nominal operation. (The pre-training results have been illustrated in Section 4.5.2 (Non-parametric System Identification Results).) In this way, the point selection utilized in this study already has a wide feature space to select from. The results of the experiment are shown in Figure 8.

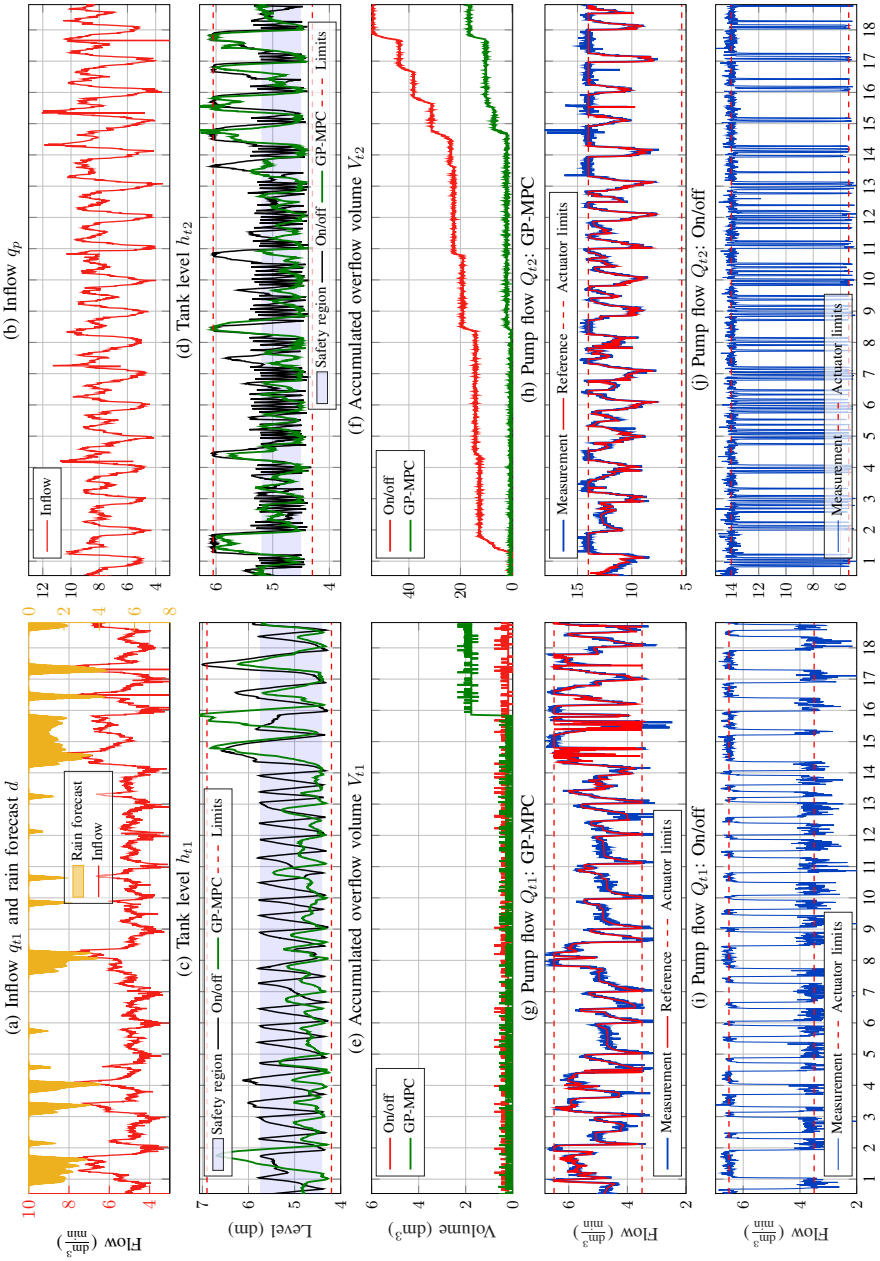


Figure 5.7: Performance comparison between the GP-MPC and On/off controller, operating the experimental setup representing the case study area during an 18-days period with heavy rain-events. Inflows, water levels in the basins, accumulated overflow and the control actions are shown regarding the upstream pumping station (left column) and the downstream pumping station (right column). (Source: [Balla et al., 2022a].)

The results of implementing the GP-based MPC controller strategy aim to show the benefits of deploying several level sensors in the sewer network, and to use the GP-based models to learn the unknown dynamics and thereby complement the nominal model part of the sewer system. Similarly, to the PDE-based MPC controller strategy, to assess the performance of the GP-based MPC controller, the methods are compared to the baseline controller described previously in Section 3.4 (Benchmark Controller). To stretch both controllers to their limits, a period of 18 days has been chosen with intensive rain events, forcing the network to overflow on several occasions. In our results, all Ω scaling values are set equal in the individual terms of the objective function, meaning that none of the two pumping stations in the experimental setup are prioritized over each other.

The results in Figure 8 compare the baseline and GP-based 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 GP-MPC (g,h) and by the on/off baseline controller (i,j). Overflows are triggered several times under the baseline controller, especially at the downstream location since the pumping stations are not collaborating to mitigate the water volumes under heavy rain periods. Note that the upstream pumping station (left column) rarely reaches its upper pumping capacity and thereby saves the capacity downstream. The controller shifts the time of overflows, by delaying the surges downstream so the downstream tank spends less time overflowing. This is observed between days $1-2$ and days $14-17$. In the latter period, the controller is exposed to a high-intensity rain event. Note that during this event, the control action oscillates at the upstream station when the safety bounds are violated and the controller realizes that using the slack variables ϵ for the overflows is necessary to reduce the overall accumulated spilled volumes. A possible explanation for this performance degradation can be that the forecasted rain is not represented well in our data batch, and additionally, that the point selection behind the algorithm is not performing well. Key performance indicators regarding the control actions, the level of uncertainty in the state propagation, and the overflow risks are presented and further detailed in Paper F.

Note that between days $14-16$, there is a relatively dry period. During this period, the pumps at the stations mimic the dry-weather flow variation (diurnal pattern), indicating that the GP correction predicts an average wastewater inflow with an uncertainty that fits quite well with the actual inflows.

The results shown in our experimental test show several benefits and challenges to using the GP-based MPC controller strategy. To get an overview of these benefits and drawbacks, in the next section we compare the results obtained in this section to the results obtained via the PDE-based MPC controller.

5.5 Control Results Comparison

To get an overview of the benefits and drawbacks of the control architectures proposed in this chapter, here we provide some discussions about the PDE-based MPC controller and the GP-based MPC controller. It is important to note that the GP-PDE-based MPC controller depicted in Figure 5.1c is an improved version of the GP-based MPC controller. The former architecture builds on similar methodological approaches and incorporates more physical knowledge into the controller to ease the influence of the dynamic error corrections provided by the GPs. Therefore, we do not consider it explicitly in this comparison discussion, rather consider it along with the GP-based MPC controller

In the following, we discuss and compare the two methods based on their complexity and practicability.

5.5.1 Complexity Comparison

We consider two aspects of the complexity, i.e., the computation complexity of solving the optimization problem and the complexity of the objective function proposed for the two controllers.

For the PDE-based MPC controller, the optimization problem has been solved for three different configurations, including 2, 3 and 4 pumping stations in the Simulation Study C high-fidelity environment. The solving times and the size of the optimization problem for the three different cases are shown in the table below.

Num. of stations	Avg. CPU time (s)	Max CPU time (s)	Decision var.	Constr.	Param.
4 ($s_{1,2,3,4}$)	2.14	8.37	5361	8385	1216
3 ($s_{1,2,3}$)	1.8	4.36	3912	6072	912
2 ($s_{1,2}$)	1.15	1.86	2608	4192	610

Table 5.1: CPU times for solving the optimization problem behind the PDE-based MPC controller strategy with different number of pumping stations stations. Source: [Balla et al., 2022d].

As depicted, the size of the optimization problem increases by including more pumping stations in the control problem, however, the computation times remain low as all constraints can be cast as linear equality and inequality constraints. The average and maximum CPU times for solving the full-scale problem are 2.14 and 8.37 seconds, respectively [Balla et al., 2022d]. The maximum CPU time corresponding to 8.37 seconds is acceptable in practice considering that the worst-case computation (under overflows when the slacks are activated) is less than 2% of the control intervals. (This study has been carried out on a 2.6 GHz, Intel Core i7 machine with 16-GB RAM [Balla et al., 2022d].)

Opposed to the PDE-based MPC controller, in case of the experimental testing of the GP-based MPC controller we could not conduct tests including more than two pumping stations due to the limitations on laboratory infrastructure. Besides, the two optimization problems have been carried out on two different machines and therefore we cannot provide a comprehensive study on the computation comparison. However, the solving time for the GP-based controller for two stations with $M = 80$ points resulted in an average CPU time of 1.7 and maximum CPU time of 4 seconds on a machine with stronger resources. Most importantly, the GP-based solution has been solved with a prediction horizon of 4-hours, while the PDE-based solution provided the same solving time for two stations but a prediction horizon of 24 hours. Hence, there is a significant difference between the computation burdens of the two problems. The increase of the optimization problem and the number of data points M is expected to drastically increase the computation burden for the GP-based implementation, meaning that a more efficient point selection and the implementation of sparse methods [Snelson and Ghahramani, 2006] are required to run the problem feasibly in a real-world implementation. (The GP-based study has been carried out on a 3.6 GHz, Intel Xeon machine with 64-GB RAM.)

The objective function of the PDE-based MPC controller is relatively simple, given that we solve a deterministic control problem and both the storage and pipe dynamics are represented by physically measurable state and optimization variables. Although the number of states used in the PDE-based optimization is high due to the inclusion of the discretized PDE-based pipe dynamics, these states are updated either by measurement or estimation. The complexity is increased significantly in the GP-based implementation as we introduce an extra state variable to characterize the uncertainty propagating in the states. This inclusion also requires solving the optimization problem in a stochastic manner, treating the constraint regarding state variables as stochastic (as we assumed that the inputs are deterministic).

5.5.2 Practicability Comparison

Practical implementation of both the PDE-based MPC controller and the GP-based MPC controller methods require to deploy water level sensors in the network to identify the transport dynamics and the disturbances infiltrating into the system. In both cases, the high-level network layout is utilized to find the link connections between the controllable assets. By using that knowledge, the graph of the network can be generated upon which the control-oriented transport and storage volume conservation dynamics are built.

To carry out the experimental implementation of the PDE-based MPC controller at water utilities, a reliable mapping between the forecasted rain intensities and the actual flow appearing in the system is required [Balla et al., 2022d]. This strategy does not solve this problem, as the underlying prediction model is physically-based on the volume conservation in the network. Therefore, the optimization problem we solve behind the controller is operating with

actual flow and level (volume) values, providing a high level of robustness in the control.

In contrast to the PDE-based MPC controller, the GP-based MPC controller strategy learns the dry- and wet-weather flows from the deployed in-sewer water level sensors. Arguably, the major benefit of learning the effect of disturbances on the level data is the ability to launch the controller without physically meaningful calibration of the sensors, furthermore physically meaningful calibration of the level-to-flow conversion. The level-to-flow conversion is partly bypassed since the mapping between the predicted upstream levels and the change of level in storage tanks is solved via utilizing the Gaussian process-based non-parametric model corrections. However, as the experimental tests have shown, the adoption of purely data-driven methods is underpinned by some practical issues. For instance, data-driven control cannot be easily explained and explicit guarantees for safe operation cannot be given. Moreover, it was shown that the selection of points used for prediction is critical and therefore our decision-making is highly reliant not only on the data quality but on the excitation of the system.

Although we lose insight and robustness by using data-driven methods without strong mathematical structures, the practical benefits gained with the GP-based MPC controller provide a feasible solution for implementation, while PDE-based MPC controller lacks some of the plug and play features that are highly desired by practitioners working in the water sector.

6 Concluding Remarks

This chapter concludes on the research objectives and provides future perspectives and reflections.

6.1 Results Evaluation

With a data-driven and *Plug and Play* emphasis, the thesis contributes to the broader real-time modelling and control of wastewater networks. The proposed modelling and control solutions aim to contribute with an online (or offline) management tool to control the actuators and give decision support to operators of wastewater systems. By working in the context of a real-world system process, our work has also focused on the practicability and feasibility of the solutions to support data-driven modelling and predictive control as feasible solutions in the water sector. In the following, we evaluate the results obtained in this work based on the two hypotheses we stated at the beginning of the thesis.

Hypothesis 1: *Structure preserving reduced models can represent the network dynamics such that the main sewer processes are captured, relying solely on in-sewer water level sensors, the topological network layout, and flow estimation data regarding pump operation from the pumping stations.*

Hypothesis 1 regards the modelling of the sewer processes through the hydraulic and hydrologic parts of the process chain. The major part of our work was dedicated to the development of control-oriented models behind the real-time predictive control algorithms. Arguably, this is a critical part of our work since the models are used to obtain the future states of the wastewater system. Chapter 4 (Sewer Dynamics Modelling) highlights the modelling approaches and the results by predicting with the models under different types of system loads.

To introduce a physically-based, yet data-driven solution for predicting water levels inside the sewer infrastructure, we developed our models by considering the basic conservation of volumes in the network. To this end, we used simple integrator-type dynamics for storage elements and simplified time and spatial-discretized partial differential equations to provide a tank-in-series representation for the volume transport in pipes. While our proposed grey-box network model is highly conceptualized, it allows us to assimilate data and use system identification by utilizing the high-level piping layout of any sewer system with controllable elements. By utilizing the links connecting the controllable elements, we proposed to generate a control model that discards the low-level details of the piping infrastructure. To compensate for these losses on

the modelling side, we used the data to capture the disturbances through the variation of our level sensor measurements. The system identification studies presented in Paper A and Paper B showed promising results to validate our hypothesis, i.e., to identify a low number of parameters of the obtained model to predict the level N-step ahead in time. Besides, to learn and distinguish between the domestic and groundwater infiltration into the sewers. Although we did not quantify the identifiability in this work, the two main modelling techniques, i.e., the kinematic and diffusion wave dynamics for volume transport in pipes, have been compared in Paper B. Using the diffusion wave model holds the major benefit of keeping track of accumulated water in the pipes. It makes possible to use the pipe infrastructure as controllable storage elements under high-load periods. Unfortunately, with the capability of tracking the backwater comes considerably high modelling and identification complexity, making the method less attractive compared to the simple kinematic wave pipe models.

To provide a solution that is easy to set into operation, obtaining a physically-based model where the lumped parameters are identified through the water level variations inside the manholes has been crucial. Besides, the assimilation of data has allowed us to shift our focus from calibration of the sensors to rather continuously updating the scaling factor from the measured level to the estimated discharge flow. To this end, we tested online system identification in the form of moving horizon estimation for both the system states and parameters of the network. The results in Paper C supported the feasibility of the idea that the structure of the model is robust enough to estimate unmeasured system states and the three parameters used for the model predictions.

One caveat of the modelling applied for the hydraulic part of the sewer processes is the lack of considering the rain load on the sewer system. Out of the three disturbances, rain is considered to have the largest uncertainty in both the modelling and control of wastewater networks. To consider the effect of rain infiltration while using the level-based framework for predict the volume inside the sewers, we proposed to use a residual approach combined with a non-parametric stochastic model in the form of Gaussian processes. To this end, Paper E and Paper F explored the possibility to model the hydrologic effect and the governing uncertainty of the rain by creating residuals with the use of the deployed level sensors and the output of the physical models developed solely for the sewer processes. In this way, we established a framework where the effect of all unknown dynamics is concentrated on locations where the states are measured, i.e., where level sensors are deployed. Although the proposed non-parametric framework is a powerful tool, its robustness for real-time implementation is debatable. Since we predict with the data itself, the demand for high quality and a large amount of data batches makes the approach vulnerable to data deficiency. Besides, propagating the uncertainty is a computationally complex task, especially if the uncertainty is explicitly treated as a state variable during the optimization. To partly overcome some of these issues, we showed some results of how we can incorporate more and more nominal knowledge into the combined network model, and therefore re-

duce the mean of the residuals representing the unknown system dynamics and the uncertainty. The results detailed in the summary implied that the more system knowledge we can incorporate into our combined models, the more likely that the non-parametric part of the model finds the correct relevances between the effect of disturbances and system states, and the decision variables and the system states.

Overall, the complete network modell proposed in this work provides an easy-commissionable solution based on in-sewer level measurements and rain forecast feeds. Besides, the framework can learn by re-identifying the nominal parts in run-time and adding newly discovered points to the data batches used for the prediction. The performance of the proposed solution is especially high if the nominal dynamics are pre-identified and the uncertainty and unknown parts are reduced to be learned by the Gaussian processes. Solving the control problem is however a more difficult task than providing open-loop predictions ahead of time, leading to the issues raised by the second hypothesis of this work.

Hypothesis 2: *A physically-based yet data-driven model relying on the available system knowledge and the topological network layout can be used to predict the effect of hydraulic and hydrologic processes and their corresponding uncertainties in an integrated predictive controller.*

Hypothesis 2 regards the control of the sewer processes using predictive control with the in-sewer level measurements, the high-level layout of the piping network, and rain forecast feeds. The research covering this part of the hypothesis strongly builds on the modelling results obtained in the first half of this work. Chapter 5 (Sewer Volume Control) highlights the control approaches and the results by applying the research to high-fidelity simulation studies and also providing proof-of-concept in experimental laboratory tests.

Paper C implemented the kinematic wave approximation based predictive controller, i.e., the PDE-based MPC controller using online moving horizon estimation. The obtained results showed that even though the number of states is high, the model complexity is kept within practically achievable limits, suitable for nonlinear predictive control [Balla et al., 2022d]. Besides, the effectiveness of the control and estimation method has been demonstrated on a network with four pumping stations, interconnected in a tree graph topology. The robustness of the closed-loop control scheme proved to be efficient both computationally and towards the uncertainty in the inflow forecasts. This is in line with the prediction capabilities of the PDE-based models, as in this case, we provided a strong mathematical structure to describe the hydraulic processes occurring in the sewers. Nevertheless, the performance shown in Paper C is underpinned by the fact that the forecast of all disturbances has been considered in terms of actual flows entering the sewer system. While this is unrealistic, the modularity of the approach has shown great potential to use in a real-time modelling scheme.

To overcome the issues with the disturbance forecasting in Paper C, the works in Paper E and Paper F attempted to solve the combined hydraulic and hydrologic problem by considering the rain forecast feeds in terms of the infiltration induced by the intensity of rain. To this end, we used the combined GP-based model with assuming the storage in the network via simple volume conservation. When solving the optimization problem behind the GP-based controller, we considered the states being stochastic, hence introducing the mean and variance of the states to fully describe the water level evolution in the predictions. Even though the results in Paper E and Paper F are promising to learn the wet- and dry-weather disturbances, the real-time implementation in our experimental work has raised some concerns about the complexity and practicability of the approach. Firstly, the size of the data batches used for the predictions has been kept minimal such that our optimization executes in $1-2$ seconds intervals within the 10 seconds control step. This allowed us to solve the optimization in our scaled-down tests but also pointed out the degradation of the control performance. Secondly, the simple point selection algorithm that we used in our tests resulted in low performance when unexplored state-input and disturbance pairs have been obtained, and in these cases, the minimization of the variance has become the dominant term in our objective function.

Overall, the residual-based modelling with the parametric and non-parametric model parts solved in real-time optimization showed promising results to tackle the complete control problem in wastewater networks without employing flow sensors in the network. Although the two types of control problems are underpinned by practical and complexity issues, the combination of the two and the incorporation of more physical knowledge in the modelling (i.e. simplified pipe model, simplified runoff dynamics model) is expected to increase the robustness and performance of the proposed control architectures.

6.2 Future Perspectives

In the following, some future research directions are given based on the author's recommendations in the field of wastewater control. We aim to list some of the practical, field- and method-specific challenges that arise with the proposed control architectures, however, have not been treated in the thesis.

1. The baseline controllers in wastewater networks typically operate via on/off schedules according to pre-defined maximum and minimum levels in the wastewater storage tanks. Our current implementation of the proposed control architecture communicates a flow setpoint to the local pumping units translated to, e.g., the speed of a pump via a frequency drive. Since the proposed architecture requires frequency drive for the local pumping units, it becomes laborious and complex to build in the fallback functionality to the baseline controller in case our control algorithm breaks down or the operators want to disconnect our control. Hence, it is interesting to research whether we can translate the flow reference from

our optimization to level-based threshold values. An illustration of such communication architecture is shown in Figure 6.1 below.

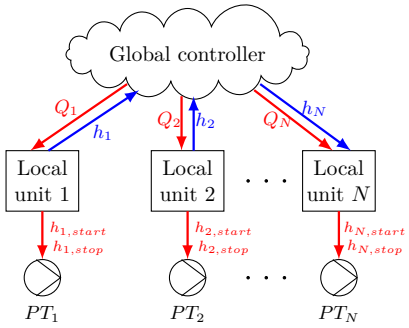


Figure 6.1: Flow reference translation to local level-based pump control.

Deploying such a solution would inherently mean that advanced control methods such as the one proposed here, could be deployed without any modification in the locally managed hierarchy level.

2. The estimation of the discharged flow arriving from an upstream station or infiltrated flow to a downstream tank has been investigated in [Kallesøe and Knudsen, 2016]. In this study, the methods have been successfully applied in networks where the downstream nodes have only one branch discharging to them. While most of the smaller wastewater networks inhabit such tree-based structures, there can be cases when two branches discharge to the same downstream station. An investigation of whether the discharge from several branches can be separated and identified is a matter of future work.

3. The Diffusion wave-based pipe model discussed in Section 4.2.5 (Diffusion Wave Model) and developed in Paper B has not been considered to estimate the discharged flow at downstream collection points. It is partly because the boundary condition of free-fall and submerged flow inhabit different types of mathematical structures, too complex to incorporate into system-identification-based modelling. A future study will be interesting to see whether the further conceptualization of the downstream boundary conditions makes it possible to find model parameters from the data.

4. It is important to note that the grey-box models developed for the volume propagation in the sewer pipes require that either the estimated flow or the level is precise during our system identification. This is mainly because we model the conservation of the volumes, and therefore when we model the discharged flow from the level measurements, we use the discharged flow to calculate the change of level in the storage elements.

When combining the physical models with the Gaussian Process non-parametric models, we argued that the data-driven part can compensate for these modelling and offset errors on the sensors. However, to achieve a robust operation, especially during real-time control, it will be interesting to investigate the effect of the flow estimation error on our algorithms.

5. The mean function selected for the Gaussian processes in our work has been chosen constant. In the scientific literature, there are several ways of estimating model parameters of a deterministic regression model combined with the hyper-parameters related to the Gaussian process [Rasmussen and Williams, 2018]. Our simplified Kinematic wave pipe model proposed in Section 4.2.6 (Conceptual Alternatives) can in fact be written in the following form: $\phi(z)^\top \beta$, where the coefficients β correspond to the θ lumped parameters in the pipe model and ϕ is a basis function fitting the simplified Kinematic wave pipe model. Besides, our storage tank model based on volume conservation is already linear in its parameters. Identifying such model would mean that the pre-processing spent on residual generation can be omitted when deploying our proposed control algorithms for wastewater networks.

6. When the parametric models corresponding to sewer pipes are identified, the data is pre-processed such that periods with dry weather are selected. It is not an unrealistic assumption, as there are periods with no rain, where mainly wastewater and groundwater infiltrate into the channels. (However, this is dependent on the specific climate and weather.) It would be interesting to re-do the tests without separating dry- and wet-weather periods and investigate whether the parametric models can find the parameters regarding the pump operation and the dry-weather flows.

7. The operation of combined wastewater networks is typically twofold; they are operated nominally during dry weather, and extreme or high-load operation is utilized during wet-weather. The operational goals might be prioritized differently during these two types of operating modes, e.g., during wet-weather the main operational objective is to avoid overflow, while during dry weather the equal mitigation and constant feed to the treatment plant are prioritized. Varying the objective function, or at least the prioritization of the terms in the objective function is expected to result in improved control performance. Therefore, an investigation into the adaptive tuning of the objectives in the optimization is a particularly relevant research topic for wastewater networks.

8. Tuning the objective function for the stochastic optimization problem in the case of the GP-based MPC controller has been challenging during our experimental tests. It has partly been due to the uncertainty propagation utilized for the predictions. The spread of the disturbance variance has been uncoordinated in our closed-loop controller, however, as proposed in

[Hewing et al., 2020], including an ancillary Linear Quadratic Regulator (LQR) might resolve this issue.

9. As mentioned during the discussion of the results governing the GP-based MPC controller, the uncertainty propagation provides a measure of the uncertainty predicted in the future. However, by utilizing such complex constraints with introducing the variance as a state variable, we increase the optimization complexity significantly. It would be interesting to test a so-called "naive" implementation approach [Kocijan, 2016], where the Gaussian process-based compensation only considers the prediction of the mean and not the variance.

6.3 Summary

One of the main contributions of the thesis is to bridge the gap between hydraulic and hydrologic modelling, often used individually as two separate problems for predictive control in wastewater networks. Two data-driven frameworks have been developed (parametric, and non-parametric) to complement each other based on hydraulic and hydrologic disciplines. Another main contribution of the works is to meet realistic implementation requirements, often neglected by the state-of-art, however crucial for practitioners and operators at the water utilities. It has been achieved by focusing on a framework that is easy to commission, i.e., the instrumentation and maintenance of the control problem do not require hard-to-maintain flow sensors spatially distributed in the sewer network.

While the data-driven and non-parametric parts of the control solution lack insight into the model, the methods have been successfully applied in practical research of this work, considering a laboratory environment, yet with the assumptions on instrumentation and data availability that is likely to meet when one aims to deploy such advanced control algorithms. Understanding the robustness of the proposed control strategy is critical to ensure proper functionality and thereby shift the rule-based paradigm of wastewater control to a more advanced and data-driven system control approach.

References

3S-Smart Software Solutions GmbH. “Codesys”. URL: https://www.codesys.com.

M. F. Aguilar, W. M. McDonald, and R. L. Dymond. “Benchmarking laboratory observation uncertainty for in-pipe storm sewer discharge measurements”. *Journal of Hydrology*, 534:73–86, 2016. ISSN 00221694. doi:10.1016/j.jhydrol.2015.12.052.

M. Ahm, S. Thorndahl, J. E. Nielsen, and M. R. Rasmussen. “Estimation of combined sewer overflow discharge: A software sensor approach based on local water level measurements”. *Water Science and Technology*, 74(11):2683–2696, 2016. ISSN 02731223. doi:10.2166/wst.2016.361.

T. Ahonen, J. Tamminen, J. Ahola, J. Viholainen, N. Aranto, and J. Kestilä. “Estimation of pump operational state with model-based methods”. *Energy Conversion and Management*, 51:1319–1325, 2010. ISSN 01968904. doi:10.1016/j.enconman.2010.01.009.

Ö. Akgiray. “Simple Formulae for Velocity, Depth of Flow, and Slope Calculations in Partially Filled Circular Pipes”. *Environmental Engineering Science*, 21(3):371–385, 2004. ISSN 10928758. doi:10.1089/109287504323067012.

F. Allgöwer, T. A. Badgwell, J. S. Qin, J. B. Rawlings, and S. J. Wright. “Nonlinear Predictive Control and Moving Horizon Estimation — An Introductory Overview”. In *Advances in Control*, pages 391–449. Springer, London, 1999. doi:10.1007/978-1-4471-0853-5_19.

J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. “CasADi: a software framework for nonlinear optimization and optimal control”. *Mathematical Programming Computation*, 11(1):1–36, 2019. ISSN 18672957. doi:10.1007/s12532-018-0139-4.

K. M. Balla, C. S. Kallesøe, C. Schou, and J. D. Bendtsen. “Nonlinear grey-box identification with inflow decoupling in gravity sewers”. *IFAC-PapersOnLine*, 53 (2):1065–1070, 2020a. ISSN 24058963. doi:10.1016/j.ifacol.2020.12.1295.

K. M. Balla, C. Schou, J. Dimon Bendtsen, and C. S. Kallesøe. “Multi-scenario model predictive control of combined sewer overflows in urban drainage networks”. In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1042–1047, Montréal, 2020b. IEEE. ISBN 9781728171401. doi:10.1109/CCTA41146.2020.9206362.

K. M. Balla, C. H. Knudsen, A. Hodzic, J. D. Bendtsen, and C. S. Kallesøe. “Nonlinear Grey-box Identification of Sewer Networks with the Backwater Effect: An Experimental Study”. In *2021 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1202–1207, San Diego, 2021. IEEE. doi:10.1109/CCTA48906.2021.9658864.

K. M. Balla, J. D. Bendtsen, C. S. Kallesøe, and C. Ocampo-Martínez. “A learning-based approach towards the data-driven predictive control of combined wastewater networks – An experimental study”. *Water Research*, pages 1–15, 2022a. doi:10.1016/j.watres.2022.118782.

K. M. Balla, D. Eringis, M. Al Ahdab, J. D. Bendtsen, C. S. Kallesøe, and C. Ocampo-Martínez. "Learning-Based Predictive Control with Gaussian Processes:An Application to Urban Drainage Networks". In *2022 American Control Conference (ACC)*, pages 1–7, Atlanta, 2022b. IEEE.

K. M. Balla, H. Lemée, C. Schou, and C. S. Kallesøe. "A Toolchain for the Data-driven Decision Support in Waste Water Networks – A Level-based Approach". In *2022 International Water Association (IWA) World Water Congress and Exhibition*, pages 1–4, Copenhagen, 2022c. IWA.

K. M. Balla, C. Schou, J. D. Bendtsen, C. Ocampo-Martínez, and C. S. Kallesøe. "A Nonlinear Predictive Control Approach for Urban Drainage Networks Using Data-Driven Models and Moving Horizon Estimation". *IEEE Transactions on Control Systems Technology (Early Access)*, pages 1–16, 2022d. doi:10.1109/TCST.2021.3137712.

B. K. Banik, L. Alfonso, C. Di Cristo, A. Leopardi, and A. Mynett. "Evaluation of different formulations to optimally locate sensors in sewer systems". *Journal of Water Resources Planning and Management*, 143(7):04017026, 2017. ISSN 0733-9496. doi:10.1061/(asce)wr.1943-5452.0000778.

M. Bartos and B. Kerkez. "Pipedream: An interactive digital twin model for natural and urban drainage systems". *Environmental Modelling and Software*, 144:105120, 2021. ISSN 13648152. doi:10.1016/j.envsoft.2021.105120.

L. Bassø, M. R. Rasmussen, S. Lindberg, D. K. Bak, O. H. Pjengaard, C. Schou, R. Pourmoayed, P. D. Simonsen, L. L. Kraglund, and S. B. Ploug. "DRAINMAN - Uvedkommende vand og det intelligente spildevandssystem". Technical report, Copenhagen, 2020.

H. Bechmann, H. Madsen, N. K. Poulsen, and M. K. Nielsen. "Grey box modeling of first flush and incoming wastewater at a wastewater treatment plant". *Environmetrics*, 11(1):1–12, 2000. ISSN 11804009. doi:10.1002/(SICI)1099-095X(200001/02)11:1<1::AID-ENV377>3.0.CO;2-N.

T. Beeneken, V. Erbe, A. Messmer, C. Reder, R. Rohlfing, M. Scheer, M. Schuetze, B. Schumacher, M. Weilandt, and M. Weyand. "Real time control (RTC) of urban drainage systems - A discussion of the additional efforts compared to conventionally operated systems". *Urban Water Journal*, 10(5):293–299, 2013. ISSN 17449006. doi:10.1080/1573062X.2013.790980.

S. Berkhahn, L. Fuchs, and I. Neuweiler. "An ensemble neural network model for real-time prediction of urban floods". *Journal of Hydrology*, 575:743–754, 2019. ISSN 00221694. doi:10.1016/j.jhydrol.2019.05.066.

S. J. Birkinshaw, G. O'Donnell, V. Glenis, and C. Kilsby. "Improved hydrological modelling of urban catchments using runoff coefficients". *Journal of Hydrology*, 594:125884, 2021. ISSN 00221694. doi:10.1016/j.jhydrol.2020.125884.

S. Bjorklund and L. Ljung. "A review of time-delay estimation techniques". In *Proceedings of the 42nd IEEE International Conference on Decision and Control(CDC)*, pages 2502–2507, Maui, USA, 2003. doi:10.1109/cdc.2003.1272997.

T. Bohlin. "A case study of grey box identification". *Automatica*, 30(2):307–318, 1994. ISSN 00051098. doi:10.1016/0005-1098(94)90032-9.

M. Borup, M. Grum, J. J. Linde, and P. S. Mikkelsen. "Dynamic gauge adjustment of high-resolution X-band radar data for convective rain storms: Model-based evaluation against measured combined sewer overflow". *Journal of Hydrology*, 539: 687–699, 2016. ISSN 00221694. doi:10.1016/j.jhydrol.2016.05.002.

A. Breinholt, F. Ö. Thordarson, J. K. Møller, M. Grum, P. S. Mikkelsen, and H. Madsen. "Grey-box modelling of flow in sewer systems with state-dependent diffusion". *Environmetrics*, 22(8):946–9610–, 2011. ISSN 11804009. doi:10.1002/env.1135.

D. Butler and J. W. Davies. *Urban Drainage*. Spon Press, 2006. doi:10.1016/s1462-0758(00)00017-0.

A. Campisano, J. Cabot Ple, D. Muschalla, M. Pleau, and P. A. Vanrolleghem. "Potential and limitations of modern equipment for real time control of urban wastewater systems". *Urban Water Journal*, 10(5):300–311, 2013. ISSN 1573062X. doi:10.1080/1573062X.2013.763996.

A. Carron, E. Arcari, M. Wermelinger, L. Hewing, M. Hutter, and M. N. Zeilinger. "Data-driven model predictive control for trajectory tracking with a robotic arm". *IEEE Robotics and Automation Letters*, 4(4):3758–3765, 2019.

G. Cembrano, J. Quevedo, M. Salamero, V. Puig, J. Figueras, and J. Martí. "Optimal control of urban drainage systems. A case study". *Control Engineering Practice*, 12(1):1–9, 2004. ISSN 09670661. doi:10.1016/S0967-0661(02)00280-0.

K. Chalupka, C. K. Williams, and I. Murray. "A framework for evaluating approximation methods for Gaussian process regression". *Journal of Machine Learning Research*, 14(1):330–350, 2013. ISSN 15324435.

F. J. Chang, J. M. Liang, and Y. C. Chen. "Flood forecasting using radial basis function neural networks". *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 31(4):530–535, 2001. ISSN 10946977. doi:10.1109/5326.983936.

J. Chen, R. Ganigué, Y. Liu, and Z. Yuan. "Real-Time Multistep Prediction of Sewer Flow for Online Chemical Dosing Control". *Journal of Environmental Engineering*, 140(11):040140371–040140379, 2014. ISSN 0733-9372. doi:10.1061/(asce)ee.1943-7870.0000860.

V. T. Chow. *Open-Channel Hydraulics*. 2009. doi:10.1126/science.131.3408.1215.b.

G. T. Costanzo, F. Sossan, M. Marinelli, P. Bacher, and H. Madsen. "Grey-box modeling for system identification of household refrigerators: A step toward smart appliances". In *IYCE 2013 - 4th International Youth Conference on Energy*, pages 1–5, 2013. ISBN 9781467355568. doi:10.1109/IYCE.2013.6604197.

V. Courdent, L. Vezzaro, P. S. Mikkelsen, A. L. Mollerup, and M. Grum. "Using ensemble weather forecast in a risk based real time optimization of urban drainage systems". *Houille Blanche*, 2:101–107, 2015. ISSN 00186368. doi:10.1051/lhb/20150025.

V. Courdent, M. Grum, and P. S. Mikkelsen. "Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles". *Journal of Hydrology*, 556:1013–1025, 2018. ISSN 00221694. doi:10.1016/j.jhydrol.2016.08.015.

Y. Cui, M. Jin, D. Li, Y. Xi, and L. Cen. "Iterative learning predictive control for urban drainage systems". In *Chinese Control Conference, CCC*, pages 4107–4112, 2015. ISBN 9789881563897. doi:10.1109/ChiCC.2015.7260272.

V. Dalmas, G. Robert, G. Besançon, and D. Georges. "Simplified Non-Uniform Models for Various Flow Configurations in Open Channels". *IFAC-PapersOnLine*, 50(1):12320–12325, 2017. ISSN 24058963. doi:10.1016/j.ifacol.2017.08.2159.

C. W. Dawson and R. L. Wilby. "Hydrological modelling using artificial neural networks". *Progress in Physical Geography*, 25(1):80–108, 2001. ISSN 03091333. doi:10.1177/030913330102500104.

Den Danske Vandklynge. "Integrated Control of Wastewater and Rainwater Systems Workshop". Technical report, Den Danske Vandklynge, Vejle, 2022.

S. Dey. "Free overfall in open channels: State-of-the-art review". *Flow Measurement and Instrumentation*, 13(5-6):247–264, 2002. ISSN 09555986. doi:10.1016/S0955-5986(02)00055-9.

DHI. "MIKE for developers". Technical report. URL: `https://docs.mikepoweredbydhi.com/engine_libraries/mike1d/mike1d_api/`.

DHI. "MIKE URBAN: Collection System Modelling of storm water drainage networks and sewer collection systems", 2016.

DHI. "Collection System - Modelling of storm water drainage networks and sewer collection systems". Technical report, Hørsholm, 2017.

DHI. "MIKE 1D: DHI Simulation Engine for 1D river and urban modelling - Reference Manual". Technical report, DHI, 2019.

S. Duchesne, A. Mailhot, E. Dequidt, and J. P. Villeneuve. "Mathematical modeling of sewers under surcharge for real time control of combined sewer overflows". *Urban Water*, 3:241–252, 2001. ISSN 14620758. doi:10.1016/S1462-0758(01)00037-1.

S. Duchesne, A. Mailhot, and J. P. Villeneuve. "Predictive real time control of surcharged interceptors: Impact of several control parameters". *Journal of the American Water Resources Association*, 39(1):125–135, 2003. ISSN 1093474X. doi:10.1111/j.1752-1688.2003.tb01566.x.

S. Duchesne, A. Mailhot, and J.-P. Villeneuve. "Global Predictive Real-Time Control of Sewers Allowing Surcharged Flows". *Journal of Environmental Engineering*, 130 (5):526–534, 2004. ISSN 0733-9372. doi:10.1061/(asce)0733-9372(2004)130:5(526).

A. P. Duncan, A. S. Chen, E. C. Keedwell, S. Djordjević, and D. A. Savić. "Urban flood prediction in real-time from weather radar and rainfall data using artificial neural networks". In *Weather Radar and Hydrology: IAHS Red Book Proceedings.*, pages 568–573, Exeter, 2012. ISBN 9781907161261.

S. Eggimann, L. Mutzner, O. Wani, M. Y. Schneider, D. Spuhler, M. Moy De Vitry, P. Beutler, and M. Maurer. "The potential of knowing more: A review of data-driven urban water management". *Environmental Science and Technology*, 51(5): 2538–2553, 2017. ISSN 15205851. doi:10.1021/acs.est.6b04267.

A. G. El-Din and D. W. Smith. "A neural network model to predict the wastewater inflow incorporating rainfall events". *Water Research*, 36:1115–1126, 2002. ISSN 00431354. doi:10.1016/S0043-1354(01)00287-1.

K. El Ghazouli, J. El Khatabi, A. Soulhi, and I. Shahrour. "Model predictive control based on artificial intelligence and EPA-SWMM model to reduce CSOs impacts in sewer systems". *Water Science and Technology*, 85(1):398–408, 2022. ISSN 0273-1223. doi:10.2166/wst.2021.511.

G. Evans, J. Blackledge, and P. Yardley. *Numerical methods for partial differential equations*. Springer Science & Business Media, 2012.

M. Farina, L. Giulioni, and R. Scattolini. "Stochastic linear Model Predictive Control with chance constraints - A review". *Journal of Process Control*, 44:53–67, 2016. ISSN 09591524. doi:10.1016/j.jprocont.2016.03.005.

D. Fiorelli and G. Schutz. "Real-time control of a sewer network using a multi-goal objective function". In *the 17th Mediterranean Conference on Control & Automation*, pages 676–681, Thessaloniki, 2009. doi:10.1109/med.2009.5164621.

T. D. Fletcher, H. Andrieu, and P. Hamel. "Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art". *Advances in Water Resources*, 51:261.279, 2013. ISSN 03091708. doi:10.1016/j.advwatres.2012.09.001.

X. Fu, H. Goddard, X. Wang, and M. E. Hopton. "Development of a scenario-based stormwater management planning support system for reducing combined sewer overflows (CSOs)". *Journal of Environmental Management*, 236:571–580, 2019. ISSN 10958630. doi:10.1016/j.jenvman.2018.12.089.

E. Gaborit, D. Muschalla, B. Vallet, P. A. Vanrolleghem, and F. Anctil. "Improving the performance of stormwater detention basins by real-time control using rainfall forecasts". *Urban Water Journal*, 10(4):230–246, 2013. ISSN 1573062X. doi:10.1080/1573062X.2012.726229.

L. García, J. Barreiro-Gomez, E. Escobar, D. Téllez, N. Quijano, and C. Ocampo-Martinez. "Modeling and real-time control of urban drainage systems: A review". *Advances in Water Resources*, 85:120–132, 2015. ISSN 03091708. doi:10.1016/j.advwatres.2015.08.007.

M. S. Gelormino and N. L. Ricker. "Model-predictive control of a combined sewer system". *International Journal of Control*, 59(3):793–816, 1994. ISSN 13665820. doi:10.1080/00207179408923105.

S. Gillé, D. Fiorelli, E. Henry, and K. Klepiszewski. "Optimal Operation of a Sewer Network Using a Simplified Hydraulic Model". In *11th International Conference on Urban Drainage*, Edinburgh, 2008.

A. Girard, C. E. Rasmussen, J. Quinonero-Candela, and R. Murray-Smith. "Gaussian process priors with uncertain inputs: Application to multiple-step ahead time series forecasting". 2003.

Grafana Labs. "Grafana documentation", 2018. URL: https://grafana.com/docs/.

F. Granata and G. de Marinis. "Machine learning methods for wastewater hydraulics". *Flow Measurement and Instrumentation*, 57:1–9, 2017. ISSN 09555986. doi:10.1016/j.flowmeasinst.2017.08.004.

M. Grant and S. Boyd. "CVX: Matlab software for disciplined convex programming, version 2.1", 2013. URL: http://cvxr.com/cvx.

J. M. Grosso, C. Ocampo-Martínez, and V. Puig. "Learning-based tuning of supervisory model predictive control for drinking water networks". *Engineering Applications of Artificial Intelligence*, 26:1741–1750, 2013. ISSN 09521976. doi:10.1016/j.engappai.2013.03.003.

J. M. Grosso, C. Ocampo-Martínez, V. Puig, and B. Joseph. "Chance-constrained model predictive control for drinking water networks". *Journal of Process Control*, 24:504–516, 2014. ISSN 09591524. doi:10.1016/j.jprocont.2014.01.010.

J. M. Grosso, P. Velarde, C. Ocampo-Martinez, J. M. Maestre, and V. Puig. "Stochastic model predictive control approaches applied to drinking water networks". *Optimal Control Applications and Methods*, 38(4):541–558, 2017. ISSN 10991514. doi:10.1002/oca.2269.

J. Görtler, R. Kehlbeck, and O. Deussen. "A visual exploration of gaussian processes". *Distill*, 2019. doi:10.23915/distill.00017. https://distill.pub/2019/visual-exploration-gaussian-processes.

H. Gudaparthi, R. Johnson, H. Challa, and N. Niu. "Deep Learning for Smart Sewer Systems: Assessing Nonfunctional Requirements". In *Proceedings - 2020 ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society, ICSE-SEIS 2020*, pages 35–38, 2020. ISBN 9781450371254.

L. Guo, S. Tik, J. M. Ledergerber, D. Santoro, E. Elbeshbishy, and P. A. Vanrolleghem. "Conceptualizing the sewage collection system for integrated sewer-WWTP modelling and optimization". *Journal of Hydrology*, 573:710–716, 2019. ISSN 00221694. doi:10.1016/j.jhydrol.2019.04.012.

M. Häck and J. Wiese. "Trends in instrumentation, control and automation and the consequences on urban water systems". *Water Science and Technology*, 54(11-12): 265–272, 2006. ISSN 02731223. doi:10.2166/wst.2006.797.

R. Halvgaard and A. K. V. Falk. "Water system overflow modeling for Model Predictive Control". In *the 12th IWA Specialised Conference on Instrumentation, Control and Automation*, Québec City, 2017.

R. Halvgaard, A. K. V. Falk, N. S. V. Lund, M. Borup, P. S. Mikkelsen, and H. Madsen. "Model predictive control for urban drainage: testing with a nonlinear hydrodynamic model". In *the 14th IWA international conference on urban drainage*, pages 2471–2474, Prague, 2017.

M. Henze and K. Arnbjerg-Nielsen. "Modernizing sewers and wastewater systems with new technologies: The Danish action plan for promotion of eco-efficient technologies - Danish lessons". Technical report, 2008. URL: http://www.ecoinnovation.dk/NR/rdonlyres/8A77E57C-FEB6-404B-9F53-4B827EFDB1B8/0/spildevand{_}baggrundsartikel{_}1.pdf.

L. Hewing, J. Kabzan, and M. N. Zeilinger. "Cautious Model Predictive Control Using Gaussian Process Regression". *IEEE Transactions on Control Systems Technology*, 28(6):2736–2744, 2020. ISSN 15580865. doi:10.1109/TCST.2019.2949757.

D. Hill, B. Kerkez, A. Rasekh, A. Ostfeld, B. Minsker, and M. K. Banks. "Sensing and Cyberinfrastructure for Smarter Water Management: The Promise and Challenge of Ubiquity". *Journal of Water Resources Planning and Management*, 140(7):018140021–018140023, 2014. ISSN 0733-9496. doi:10.1061/(asce)wr.1943-5452.0000449.

Y. Hu, D. Scavia, and B. Kerkez. "Are all data useful? Inferring causality to predict flows across sewer and drainage systems using directed information and boosted regression trees". *Water Research*, 145:697–706, 2018. ISSN 18792448. doi:10.1016/j.watres.2018.09.009.

C. Hutton, L. Vamvakeridou-Lyroudia, Z. Kapelan, and D. Savic. "Real-time Modelling and Data Assimilation Techniques for Improving the Accuracy of Model Predictions". Technical report, 2011.

C. Jacopin, E. Lucas, M. Desbordes, and P. Bourgogne. "Optimisation of operational management practices for the detention basins". *Water Science and Technology*, 44 (2-3):277–285, 2001. ISSN 02731223. doi:10.2166/wst.2001.0780.

M. È. Jean, S. Duchesne, G. Pelletier, and M. Pleau. "Selection of rainfall information as input data for the design of combined sewer overflow solutions". *Journal of Hydrology*, 565:559–569, 2018. ISSN 00221694. doi:10.1016/j.jhydrol.2018.08.064.

T. N. Jensen, C. S. Kallesøe, J. D. Bendtsen, and R. Wisniewski. "Iterative learning pressure control in water distribution networks". In *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pages 583–588. IEEE, 2018.

Y. Jia, F. Zheng, H. R. Maier, A. Ostfeld, E. Creaco, D. Savic, J. Langeveld, and Z. Kapelan. "Water quality modeling in sewer networks: Review and future research directions". *Water Research*, 202:117419, 2021. ISSN 18792448. doi:10.1016/j.watres.2021.117419.

E. R. Jones, M. T. Van Vliet, M. Qadir, and M. F. Bierkens. "Country-level and gridded estimates of wastewater production, collection, treatment and reuse". *Earth System Science Data*, 13:237–254, 2021. ISSN 18663516. doi:10.5194/essd-13-237-2021.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Output-feedback model predictive control of sewer networks through moving horizon estimation". In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 1061–1066, Los Angeles, 2014a. doi:10.1109/CDC.2014.7039522.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Hybrid modeling and receding horizon control of sewer networks". *Water Resources Research*, 50(11): 8497 – 8514, 2014b. ISSN 19447973. doi:10.1002/2013WR015119.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Output-feedback control of combined sewer networks through receding horizon control with moving horizon estimation". *Water Resources Research*, 51(10):8129–8145, 2015. ISSN 19447973. doi:10.1002/2014WR016696.

J. Kabzan, L. Hewing, A. Liniger, and M. N. Zeilinger. "Learning-based model predictive control for autonomous racing". *IEEE Robotics and Automation Letters*, 4(4):3363–3370, 2019.

P. Kadlec, B. Gabrys, and S. Strandt. "Data-driven Soft Sensors in the process industry". *Computers and Chemical Engineering*, 33:795–814, 2009. ISSN 00981354. doi:10.1016/j.compchemeng.2008.12.012.

C. S. Kallesøe and M. Eriksen. "Supervision of Pumps and their Operating Conditions in Sewage Pumping Stations". *Water Practice and Technology*, 5(2):1–11, 2010. ISSN 1751-231X. doi:10.2166/wpt.2010.040.

C. S. Kallesøe and T. Knudsen. "Self calibrating flow estimation in waste water pumping stations". In *2016 European Control Conference, ECC 2016*, pages 55–60, 2016. ISBN 9781509025916. doi:10.1109/ECC.2016.7810263.

A. Kenward, N. Zenes, J. Bronzan, J. Brady, and K. Shah. "Overflow: Climate Change, Heavy Rain, and Sewage". Technical report, Climate Central, 2016. URL: `http://assets.climatecentral.org/pdfs/Overflow{_}sewagereport{_}update.pdf`.

B. Kerkez, C. Gruden, M. Lewis, L. Montestruque, M. Quigley, B. Wong, A. Bedig, R. Kertesz, T. Braun, O. Cadwalader, A. Poresky, and C. Pak. "Smarter stormwater systems". *Environmental Science and Technology*, 50:7267–7273, 2016. ISSN 15205851. doi:10.1021/acs.est.5b05870.

R. Kicsiny. "Grey-box model for pipe temperature based on linear regression". *International Journal of Heat and Mass Transfer*, 107:13–20, 2017. ISSN 00179310. doi:10.1016/j.ijheatmasstransfer.2016.11.033.

O. Kisi, J. Shiri, and M. Tombul. "Modeling rainfall-runoff process using soft computing techniques". *Computers and Geosciences*, 51:108–117, 2013. ISSN 00983004. doi:10.1016/j.cageo.2012.07.001.

R. Kitchin. "The real-time city? Big data and smart urbanism". *GeoJournal*, 79(1): 1–14, 2014. ISSN 03432521. doi:10.1007/s10708-013-9516-8.

E. D. Klenske, M. N. Zeilinger, B. Schölkopf, and P. Hennig. "Gaussian Process-Based Predictive Control for Periodic Error Correction". *IEEE Transactions on Control Systems Technology*, 24(1):110–121, 2016. ISSN 10636536. doi:10.1109/TCST.2015.2420629.

K. Klepiszewski and T. G. Schmitt. "Comparison of conventional rule based flow control with control processes based on fuzzy logic in a combined sewer system". *Water Science and Technology*, 46(6-7):77–84, 2002. ISSN 02731223. doi:10.2166/wst.2002.0665.

J. Kocijan. *Modelling and control of dynamic systems using Gaussian process models*. Springer, 2016.

J. W. Labadie. "Advances in water resources systems engineering: Applications of machine learning". In *Modern Water Resources Engineering*. 2014. ISBN 9781627035958. doi:10.1007/978-1-62703-595-8_10.

S. Leirens, J. M. Giraldo, R. R. Negenborn, and B. De Schutter. "A pattern search method for improving the operation of sewer systems". In *IFAC Proceedings Volumes (IFAC-PapersOnline)*, pages 591–596, 2010. ISBN 9783902661913. doi:10.3182/20100712-3-fr-2020.00096.

J. P. Leitão, J. P. Carbajal, J. Rieckermann, N. E. Simões, A. Sá Marques, and L. M. de Sousa. "Identifying the best locations to install flow control devices in sewer networks to enable in-sewer storage". *Journal of Hydrology*, 556:371–383, 2018. ISSN 00221694. doi:10.1016/j.jhydrol.2017.11.020.

G. Li, X. Zhou, et al. "Iterative learning control of dam-river channel irrigation systems.". *Journal of Guangxi Normal University-Natural Science Edition*, 36(1): 53–60, 2018.

J. Li, K. Sharma, Y. Liu, G. Jiang, and Z. Yuan. "Real-time prediction of rain-impacted sewage flow for on-line control of chemical dosing in sewers". *Water Research*, 149:311–321, 2019. ISSN 18792448. doi:10.1016/j.watres.2018.11.021.

J. Li, X. Yang, and R. Sitzenfrei. "Rethinking the Framework of Smart Water System: A Review". *MDPI Water*, 12(2):412, 2020. doi:10.3390/w12020412. URL: `https://www.mdpi.com/2073-4441/12/2/412`.

J. Li, K. Sharma, W. Li, and Z. Yuan. "Swift Hydraulic Models for Real-time Control Applications in Sewer Networks". *Water Research*, 213:118141, 2022. doi:https://doi.org/10.1016/j.watres.2022.118141.

W.-W. Li, G.-P. Sheng, R. J. Zeng, X.-W. Liu, and H.-Q. Yu. "China's wastewater discharge standards in urbanization". *Environmental Science and Pollution Research*, 19:1422–1431, 2012. ISSN 0944-1344. doi:10.1007/s11356-011-0572-7.

X. Litrico and V. Fromion. "Simplified Modeling of Irrigation Canals for Controller Design". *Journal of Irrigation and Drainage Engineering*, 130(5):373–383, 2004. ISSN 0733-9437. doi:10.1061/(asce)0733-9437(2004)130:5(373).

X. Litrico and V. Fromion. "Boundary control of linearized Saint-Venant equations oscillating modes". *Automatica*, 42:967–972, 2006. ISSN 00051098. doi:10.1016/j.automatica.2006.02.002.

X. Litrico and V. Fromion. *Modeling and control of hydrosystems*. Springer, 2009. ISBN 9781848826236. doi:10.1007/978-1-84882-624-3.

T. Liu, J. E. Ramirez-Marquez, S. C. Jagupilla, and V. Prigiobbe. "Combining a statistical model with machine learning to predict groundwater flooding (or infiltration) into sewer networks". *Journal of Hydrology*, 2021. ISSN 00221694. doi:10.1016/j.jhydrol.2021.126916.

A. M. D. Livera, R. J. Hyndman, and R. D. Snyder. "Forecasting time series with complex seasonal patterns using exponential smoothing". *Journal of the American Statistical Association*, 106(496):1513–1527, 2011. doi:10.1198/jasa.2011.tm09771.

R. Löwe, S. Thorndahl, P. S. Mikkelsen, M. R. Rasmussen, and H. Madsen. "Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar". *Journal of Hydrology*, 512:397–407, 2014. ISSN 00221694. doi:10.1016/j.jhydrol.2014.03.027.

R. Löwe, L. Vezzaro, P. S. Mikkelsen, M. Grum, and H. Madsen. "Probabilistic runoff volume forecasting in risk-based optimization for RTC of urban drainage systems". *Environmental Modelling and Software*, 80:143–158, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.02.027.

N. S. V. Lund, A. K. V. Falk, M. Borup, H. Madsen, and P. Steen Mikkelsen. "Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management". *Critical Reviews in Environmental Science and Technology*, 48(3):279–339, 2018. ISSN 15476537. doi:10.1080/10643389.2018.1455484.

H. Luo, N. Oberg, B. J. Landry, and M. H. García. "Assessing the system performance of an evolving and integrated urban drainage system to control combined sewer overflows using a multiple-layer based coupled modeling approach". *Journal of Hydrology*, 603:127130, 2021. ISSN 00221694. doi:10.1016/j.jhydrol.2021.127130.

J. Ma, W. Sun, G. Yang, and D. Zhang. "Hydrological Analysis Using Satellite Remote Sensing Big Data and CREST Model". *IEEE Open Access Journals and Conferences*, 6:9006–9016, 2018. ISSN 21693536. doi:10.1109/ACCESS.2018.2810252.

H. Madsen, A. K. Falk, and R. Halvgaard. "A Model Predictive Control Framework for Real-Time Optimisation of Water System Operations". In *the 13th International Conference on Hydroinformatics (HIC)*, pages 1281–1288, Palermo, 2018. EPiC Engineering. doi:10.29007/fg7g.

L. W. Mays. *Stormwater Collection Systems Design Handbook*. McGraw-Hill Education, 1st edition, 2001. ISBN 0071354719.

B. McDonnell, K. Ratliff, M. Tryby, J. Wu, and A. Mullapudi. "PySWMM: The Python Interface to Stormwater Management Model (SWMM)". *Journal of Open Source Software*, 5(52):2292–2295, 2020. ISSN 2475-9066. doi:10.21105/joss.02292.

E. Mignot, H. Bonakdari, P. Knothe, G. Lipeme Kouyi, A. Bessette, N. Rivière, and J. L. Bertrand-Krajewski. "Experiments and 3D simulations of flow structures in junctions and their influence on location of flowmeters". *Water Science and Technology*, 66(6):1325–1332, 2012. ISSN 02731223. doi:10.2166/wst.2012.319.

A. Molini, L. G. Lanza, and P. La Barbera. "The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications". *Hydrological Processes*, 19:1073–1088, 2005. ISSN 08856087. doi:10.1002/hyp.5646.

A. L. Mollerup, P. S. Mikkelsen, and G. Sin. "A methodological approach to the design of optimising control strategies for sewer systems". *Environmental Modelling and Software*, 83:103–115, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.05.004.

S. R. Mounce, W. Shepherd, G. Sailor, J. Shucksmith, and A. J. Saul. "Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data". *Water Science and Technology*, 69(6):1326–1333, 2014. ISSN 02731223. doi:10.2166/wst.2014.024.

A. Mullapudi, M. J. Lewis, C. L. Gruden, and B. Kerkez. "Deep reinforcement learning for the real time control of stormwater systems". *Advances in Water Resources*, 140:103600, 2020. ISSN 03091708. doi:10.1016/j.advwatres.2020.103600.

S. Munier, X. Litrico, G. Belaud, and P. O. Malaterre. "Distributed approximation of open-channel flow routing accounting for backwater effects". *Advances in Water Resources*, 31(12):1590–1602, 2008. ISSN 03091708. doi:10.1016/j.advwatres.2008.07.007.

H. A. Nasir, M. Cantoni, Y. Li, and E. Weyer. "Stochastic model predictive control based reference planning for automated open-water channels". *IEEE Transactions on Control Systems Technology*, 29(2):607–619, 2019.

NCEI. "Extremes in 1-day Precipitation", 2021. URL: `https://www.ncdc.noaa.gov/extremes/cei/graph/us/01-12/4`.

M. Niazi, C. Nietch, M. Maghrebi, N. Jackson, B. R. Bennett, M. Tryby, and A. Massoudieh. "Storm Water Management Model: Performance Review and Gap Analysis". *Journal of Sustainable Water in the Built Environment*, 3(2):040170021–0401700232, 2017. ISSN 2379-6111. doi:10.1061/jswbay.0000817.

H. A. Nielsen and H. Madsen. "Modelling the heat consumption in district heating systems using a grey-box approach". *Energy and Buildings*, 38:63–71, 2006. ISSN 03787788. doi:10.1016/j.enbuild.2005.05.002.

K. M. Nielsen, T. S. Pedersen, C. Kallesøe, and P. Andersen. "Control of cod flow to a waste water treatment plant". In *International Conference on Informatics in Control, Automation and Robotics*, pages 83–104. Springer, 2020a.

K. M. Nielsen, T. S. Pedersen, C. Kallesøe, P. Andersen, L. S. Mestre, and P. K. Murugesan. "Control of sewer flow using a buffer tank". In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics*, pages 63–70. SCITEPRESS Digital Library, 2020b.

C. Ocampo-Martinez. *Model Predictive Control of Wastewater Systems*. Springer, Barcelona, 1st edition, 2010. ISBN 978-1-84996-352-7. doi:10.1007/978-1-84996-353-4.

C. Ocampo-Martinez and V. Puig. "On modelling approaches for receding-horizon control design applied to large-scale sewage systems". In *Proceedings of the IEEE Conference on Decision and Control*, pages 8052–8058, 2009. ISBN 9781424438716. doi:10.1109/CDC.2009.5400216.

C. Ocampo-Martinez, V. Puig, G. Cembrano, and J. Quevedo. "Application of predictive control strategies to the management of complex networks in the urban water cycle". *IEEE Control Systems*, 33(1):15–41, 2013. ISSN 1941000X. doi:10.1109/MCS.2012.2225919.

D. Ochoa, G. Riano-Briceno, N. Quijano, and C. Ocampo-Martinez. "Control of urban drainage systems: Optimal flow control and deep learning in action". In *Proceedings of the American Control Conference*, pages 4826–4831, 2019. ISBN 9781538679265. doi:10.23919/acc.2019.8814958.

R. Palmitessa, P. S. Mikkelsen, M. Borup, and A. W. Law. "Soft sensing of water depth in combined sewers using LSTM neural networks with missing observations". *Journal of Hydro-Environment Research*, 38:106–116, 2021. ISSN 15706443. doi:10.1016/j.jher.2021.01.006.

A. Pedersen, J. Pedersen, M. Borup, A. Brink-Kjær, L. Christiansen, and P. Mikkelsen. "Using multi-event hydrologic and hydraulic signatures from water level sensors to diagnose locations of uncertainty in integrated urban drainage models used in living digital twins". *Water Science and Technology*, 2021.

G. Petrucci and B. Tassin. "A simple model of flow-rate attenuation in sewer systems. Application to urban stormwater source control". *Journal of Hydrology*, 522:534–543, 2015. ISSN 00221694. doi:10.1016/j.jhydrol.2015.01.012.

M. Poehler and P. Uwe Thamsen. "Design of a decentralised intelligent network for five wet pit pumping stations". *Czasopismo Techniczne*, 114(2):129–138, 2017. ISSN 0011-4561. doi:10.4467/2353737xct.17.023.6216.

V. Puig, G. Cembrano, J. Romera, J. Quevedo, B. Aznar, G. Ramón, and J. Cabot. "Predictive optimal control of sewer networks using CORAL tool: Application to Riera Blanca catchment in Barcelona". *Water Science and Technology*, 60(4): 869–878, 2009. ISSN 02731223. doi:10.2166/wst.2009.424.

J. Quinonero-Candela and C. E. Rasmussen. "A unifying view of sparse approximate gaussian process regression". *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA, 2018. doi:10.7551/mitpress/3206.001.0001.

L. Raso, D. Schwanenberg, N. C. van de Giesen, and P. J. van Overloop. "Short-term optimal operation of water systems using ensemble forecasts". *Advances in Water Resources*, 71:200–208, 2014. ISSN 03091708. doi:10.1016/j.advwatres.2014.06.009.

J. C. Refsgaard. "Validation and intercomparison of different updating procedures for real-time forecasting". *Nordic Hydrology*, 28:65–84, 1997. ISSN 00291277. doi:10.2166/nh.1997.0005.

G. Riaño-Briceño, J. Barreiro-Gomez, A. Ramirez-Jaime, N. Quijano, and C. Ocampo-Martinez. "MatSWMM - An open-source toolbox for designing real-time control of urban drainage systems". *Environmental Modelling and Software*, 83:143–154, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.05.009.

Y. A. Rjeily, O. Abbas, M. Sadek, I. Shahrour, and F. H. Chehade. "Flood forecasting within urban drainage systems using NARX neural network". *Water Science and Technology*, 76(9):2401–2412, 2017. ISSN 02731223. doi:10.2166/wst.2017.409.

J. A. Roberson and C. T. Crowe. *Engineering Fluid Mechanics*. Houghton Mifflin Company, Boston, 5th edition, 1993. ISBN 0-395-66161-7. doi:10.1007/978-1-4020-6742-6.

T. J. Rogers, G. R. Holmes, E. J. Cross, and K. Worden. *On a Grey Box Modelling Framework for Nonlinear System Identification*. Springer International Publishing, 2017. ISBN 978-3-319-53841-9.

L. Romero, B. Joseph-Duran, C. Sun, J. Meseguer, G. Cembrano, R. Guasch, M. Martínez, E. Muñoz, and V. Puig. "An integrated software architecture for the pollution-based real-time control of urban drainage systems". *Journal of Hydroinformatics*, 23(3):671–687, 2021. ISSN 14651734. doi:10.2166/HYDRO.2021.149.

L. Rossman. "Storm Water Management Model User's Manual". *US Environment Protection Agency*, 2015.

M. Schütze, A. Campisano, H. Colas, W. Schilling, and P. A. Vanrolleghem. "Real time control of urban wastewater systems - Where do we stand today?". *Journal of Hydrology*, 299:335–248, 2004. ISSN 00221694. doi:10.1016/j.jhydrol.2004.08.010.

M. R. Schütze, D. Butler, and M. B. Beck. *Modelling, Simulation and Control of Urban Wastewater Systems*. Springer, 2002. doi:10.1007/978-1-4471-0157-4.

J. Schuurmans, O. H. Bosgra, and R. Brouwer. "Open-channel flow model approximation for controller design". *Applied Mathematical Modelling*, 19(9):525–530, 1995. ISSN 0307904X. doi:10.1016/0307-904X(95)00053-M.

D. Schwanenberg, S. Galelli, and I. Sheret. "Nonlinear Model Predictive Control for heterogeneous process models in water resources". In *IFAC Proceedings Volumes (IFAC-PapersOnline)*, pages 10565–10570, 2011. ISBN 9783902661937. doi:10.3182/20110828-6-IT-1002.02858.

P. Segovia, L. Rajaoarisoa, F. Nejjari, E. Duviella, and V. Puig. "Input-Delay Model Predictive Control of Inland Waterways Considering the Backwater Effect". In *2018 IEEE Conference on Control Technology and Applications*, pages 589–594, 2018. ISBN 9781538676981. doi:10.1109/CCTA.2018.8511553.

V. P. Singh. "Kinematic wave modelling in water resources: A historical perspective". *Hydrological Processes*, 15:671–706, 2001. ISSN 08856087. doi:10.1002/hyp.99.

E. Snelson and Z. Ghahramani. "Sparse gaussian processes using pseudo-inputs". *Advances in Neural Information Processing Systems*, 18:1259–1266, 2006.

J. F. Sturm. "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones". *Optimization Methods and Software*, 11(1-4):625–653, 1999. doi:10.1080/10556789908805766. URL: https://doi.org/10.1080/10556789908805766.

Su Ki Ooi and E. Weyer. "Closed loop identification of an irrigation channel". In *Proceedings of the 40th IEEE Conference on Decision and Control*, pages 4338–4343, Orlando, Florida, 2003. doi:10.1109/cdc.2001.980883.

H. Sufi Karimi, B. Natarajan, C. L. Ramsey, J. Henson, J. L. Tedder, and E. Kemper. "Comparison of learning-based wastewater flow prediction methodologies for smart sewer management". *Journal of Hydrology*, 577:123977, 2019. ISSN 00221694. doi:10.1016/j.jhydrol.2019.123977.

C. Sun, B. Joseph-Duran, T. Maruejouls, G. Cembrano, E. Munoz, J. Meseguer, A. Montserrat, S. Sampe, V. Puig Cayuela, and X. Litrico. "Conceptual Quality Modelling and Integrated Control of Combined Urban Drainage System". In *12th IWA Conference on Instrumentation, Control and Automation*, pages 141–148, Quebec, 2017.

K. Sundar and A. Zlotnik. "Dynamic State and Parameter Estimation for Natural Gas Networks using Real Pipeline". In *CCTA 2019 - 3rd IEEE Conference on Control Technology and Applications*, 2019. ISBN 9781728127675. doi:10.1109/CCTA.2019.8920430.

N. S.V. Lund, H. Madsen, M. Mazzoleni, D. Solomatine, and M. Borup. "Assimilating flow and level data into an urban drainage surrogate model for forecasting flows and overflows". *Journal of Environmental Management*, 248:109052, 2019. ISSN 10958630. doi:10.1016/j.jenvman.2019.05.110.

J. L. Svensen, H. H. Niemann, A. K. V. Falk, and N. K. Poulsen. "Chance-constrained model predictive control a reformulated approach suitable for sewer networks". *Advanced Control for Applications*, 3(4):1–21, 2021. ISSN 2578-0727. doi:10.1002/adc2.94.

R. Szymkiewicz. *Numerical modeling in open channel hydraulics*, volume 83. Springer Science & Business Media, 2010.

R. Szymkiewicz and D. Gasiorowski. "Simulation of unsteady flow over floodplain using the diffusive wave equation and the modified finite element method". *Journal of Hydrology*, 464-465:165–175, 2012. ISSN 00221694. doi:10.1016/j.jhydrol.2012.07.009.

F. Ö. Thordarson, A. Breinholt, J. K. Møller, P. S. Mikkelsen, M. Grum, and H. Madsen. "Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion". *Stochastic Environmental Research and Risk Assessment*, 26:1151–1162, 2012. ISSN 14363259. doi:10.1007/s00477-012-0563-3.

S. Thorndahl, T. S. Poulsen, T. Bøvith, M. Borup, M. Ahm, J. E. Nielsen, M. Grum, M. R. Rasmussen, R. Gill, and P. S. Mikkelsen. "Comparison of short-term rainfall forecasts for modelbased flow prediction in urban drainage systems". *Water Science and Technology*, 68(2):472–480, 2013. ISSN 02731223. doi:10.2166/wst.2013.274.

S. Thorndahl, T. Einfalt, P. Willems, J. Ellerbæk Nielsen, M. C. Ten Veldhuis, K. Arnbjerg-Nielsen, M. R. Rasmussen, and P. Molnar. "Weather radar rainfall data in urban hydrology". *Hydrology and Earth System Sciences*, 21:1359–1380, 2017. ISSN 16077938. doi:10.5194/hess-21-1359-2017.

C. Thrysøe, K. Arnbjerg-Nielsen, and M. Borup. "Identifying fit-for-purpose lumped surrogate models for large urban drainage systems using GLUE". *Journal of Hydrology*, 568:517–533, 2019. ISSN 00221694. doi:10.1016/j.jhydrol.2018.11.005.

X. Tian, R. R. Negenborn, P. J. van Overloop, J. María Maestre, A. Sadowska, and N. van de Giesen. "Efficient multi-scenario Model Predictive Control for water resources management with ensemble streamflow forecasts". *Advances in Water Resources*, 109:58–68, 2017. ISSN 03091708. doi:10.1016/j.advwatres.2017.08.015.

X. Tian, Y. Guo, R. R. Negenborn, L. Wei, N. M. Lin, and J. M. Maestre. "Multi-Scenario Model Predictive Control Based on Genetic Algorithms for Level Regulation of Open Water Systems under Ensemble Forecasts". *Water Resources Management*, 33(9):3025–3040, 2019. ISSN 15731650. doi:10.1007/s11269-019-02284-x.

E. Todini. "Hydrological catchment modelling: Past, present and future". *Hydrology and Earth System Sciences*, 11(1):468–482, 2007. ISSN 16077938. doi:10.5194/hess-11-468-2007.

S. C. Troutman, N. Schambach, N. G. Love, and B. Kerkez. "An automated toolchain for the data-driven and dynamical modeling of combined sewer systems". *Water Research*, 126:88–100, 2017. ISSN 18792448. doi:10.1016/j.watres.2017.08.065.

H. J. Tulleken. "Grey-box modelling and identification using physical knowledge and bayesian techniques". *Automatica*, 29(2):285–308, 1993. ISSN 00051098. doi:10.1016/0005-1098(93)90124-C.

UN DESA. "World Urbanization Prospects: The 2018 Revision". Technical report, United Nations, New York,NY,USA, 2019. URL: `https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf`.

UNESCO World Water Assessment Programme. *Wastwater The Untapped Resource*. UN-Water, 2017. ISBN 9789231002014.

US EPA. "Clean Watersheds Needs Survey 2012". *United States Environmental Protection Agency*, 2016. ISSN 1098-6596.

US EPA. "Seasonality and Climate Change: A Review of Observed Evidence in the United States. U.S. Environmental Protection Agency". Technical report, U.S. Environmental Protection Agency, 2021. URL: `https://www.epa.gov/climate-indicators/seasonality-and-climate-change`.

J. Val Ledesma, R. Wisniewski, and C. S. Kallesøe. "Smart water infrastructures laboratory: Reconfigurable test-beds for research in water infrastructures management". *Water (Switzerland)*, 13(13):1875, 2021. ISSN 20734441. doi:10.3390/w13131875.

J. Val Ledesma, R. Wisniewski, and C. S. Kallesøe. "Real-time reinforcement learning control in poor experimental conditions". In *EUCA Conference Editorial Board*, volume 978. European Union Control Association, 2021a.

J. Val Ledesma, R. Wisniewski, and C. S. Kallesøe. "Reinforcement learning control for water distribution networks with periodic disturbances". In *2021 Annual American Control Conference, ACC 2021*, 2021b.

J. A. van der Werf, Z. Kapelan, and J. Langeveld. "Quantifying the true potential of Real Time Control in urban drainage systems". *Urban Water Journal*, 18(10): 873–884, 2021. ISSN 17449006. doi:10.1080/1573062X.2021.1943460.

K.-J. van Heeringen, R. van Nooijen, K. Kooij, and B. Postma. "Real-time Control of sewer pumps by using ControlNEXT to smooth inflow at Waste Water Treatment Plant Garmerwolde". In *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, pages EPSC2016–1412, Apr. 2016.

R. R. van Nooijen and A. Kolechkina. "A controlled sewer system should be treated as a sampled data system with events". *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 51(16):61–66, 2018. ISSN 24058963. doi:10.1016/j.ifacol.2018.08.011.

P. J. van Overloop, S. Weijs, and S. Dijkstra. "Multiple Model Predictive Control on a drainage canal system". *Control Engineering Practice*, 16(5):531–540, 2008. ISSN 09670661. doi:10.1016/j.conengprac.2007.06.002.

P. Velarde, J. M. Maestre, C. Ocampo-Martínez, and C. Bordons. "Application of robust model predictive control to a renewable hydrogen-based microgrid". In *2016 European Control Conference (ECC)*, pages 1209–1214. IEEE, 2016.

P. Velarde, X. Tian, A. D. Sadowska, and J. M. Maestre. "Scenario-Based Hierarchical and Distributed MPC for Water Resources Management with Dynamical Uncertainty". *Water Resources Management*, 33:677–696, 2019. ISSN 15731650. doi:10.1007/s11269-018-2130-2.

L. Vezzaro and M. Grum. "A generalised Dynamic Overflow Risk Assessment (DORA) for Real Time Control of urban drainage systems". In *the 9th International Conference on Urban Drainage Modelling*, Belgrade, 2014. doi:10.1016/j.jhydrol.2014.05.019.

V. K. Vidyarthi, A. Jain, and S. Chourasiya. "Modeling rainfall-runoff process using artificial neural network with emphasis on parameter sensitivity". *Modeling Earth Systems and Environment*, 6:2177–2188, 2020. ISSN 23636211. doi:10.1007/s40808-020-00833-7.

A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". *Mathematical Programming*, 106(1):25–57, 2006. ISSN 00255610. doi:10.1007/s10107-004-0559-y.

A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". *Mathematical programming*, 106(1):25–57, 2006.

B. Wang, P. Dehghanian, and D. Zhao. "Chance-constrained energy management system for power grids with high proliferation of renewables and electric vehicles". *IEEE Transactions on Smart Grid*, 11(3):2324–2336, 2019.

Y. Wang, C. Ocampo-Martínez, V. Puig, and J. Quevedo. "Gaussian-process-based demand forecasting for predictive control of drinking water networks". In *Proceedings of the 9th International Conference on Critical Information InfrastructuresSecurity (CRITIS)*, pages 1–12, 2014. ISBN 9783319316635. doi:10.1007/978-3-319-31664-2_8.

Y. Wang, C. Ocampo-Martinez, and V. Puig. "Stochastic model predictive control based on Gaussian processes applied to drinking water networks". *IET Control Theory and Applications*, 10(8):947–955, 2016. ISSN 17518652. doi:10.1049/iet-cta.2015.0657.

E. Weyer. "System identification of an open water channel". *Control Engineering Practice*, 33(15):265–270, 2001. ISSN 09670661. doi:10.1016/S0967-0661(01)00099-5.

C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

A. Wills and B. Ninness. "On gradient-based search for multivariable system estimates". *IEEE Transactions on Automatic Control*, 53(1):298–306, 2008. ISSN 00189286. doi:10.1109/TAC.2007.914953.

V. Wolfs, M. F. Villazon, and P. Willems. "Development of a semi-automated model identification and calibration tool for conceptual modelling of sewer systems". *Water Science and Technology*, 68(1):167–175, 2013. ISSN 02731223. doi:10.2166/wst.2013.237.

V. Wolfs, P. Meert, and P. Willems. "Modular conceptual modelling approach and software for river hydraulic simulations". *Environmental Modelling and Software*, 71:60–77, 2015. ISSN 13648152. doi:10.1016/j.envsoft.2015.05.010.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow". *Advances in Water Resources*, 34:282–290, 2011. ISSN 03091708. doi:10.1016/j.advwatres.2010.11.009.

M. Xu, R. R. Negenborn, P. J. van Overloop, and N. C. van de Giesen. "De Saint-Venant equations-based model assessment in model predictive control of open channel flow". *Advances in Water Resources*, 49:37–45, 2012. ISSN 03091708. doi:10.1016/j.advwatres.2012.07.004.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "Model reduction in model predictive control of combined water quantity and quality in open channels". *Environmental Modelling and Software*, 42:72–87, 2013. ISSN 13648152. doi:10.1016/j.envsoft.2012.12.008.

Z. Yuan, G. Olsson, R. Cardell-Oliver, K. van Schagen, A. Marchi, A. Deletic, C. Urich, W. Rauch, Y. Liu, and G. Jiang. "Sweating the assets – The role of instrumentation, control and automation in urban water systems". *Water Research*, 155(1):381–402, 2019. ISSN 18792448. doi:10.1016/j.watres.2019.02.034.

D. Zhang, E. S. Hølland, G. Lindholm, and H. Ratnaweera. "Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer". *Journal of Hydrology*, 567:792–802, 2018a. ISSN 00221694. doi:10.1016/j.jhydrol.2017.11.029.

D. Zhang, G. Lindholm, and H. Ratnaweera. "Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring". *Journal of Hydrology*, 556:409–418, 2018b. ISSN 00221694. doi:10.1016/j.jhydrol.2017.11.018.

M. Zhang, Y. Liu, X. Cheng, D. Z. Zhu, H. Shi, and Z. Yuan. "Quantifying rainfall-derived inflow and infiltration in sanitary sewer systems based on conductivity monitoring". *Journal of Hydrology*, 558:174–183, 2018c. ISSN 00221694. doi:10.1016/j.jhydrol.2018.01.002.

Q. Zhang, F. Zheng, Y. Jia, D. Savic, and Z. Kapelan. "Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems". *Water Research*, 188(1):116544, 2021. ISSN 18792448. doi:10.1016/j.watres.2020.116544.

Y. Zou, L. Cen, D. Li, and X. He. "Simplified state-space model and validation of irrigation canal systems". In *Chinese Control Conference, CCC*, pages 2002–2007, Hangzhou, 2015. ISBN 9789881563897. doi:10.1109/ChiCC.2015.7259938.

Part II

Papers

Paper A

Nonlinear Grey-Box Identification with Inflow Decoupling in Gravity Sewers

Krisztian Mark Balla[a,b], Carsten Skovmose Kallesøe[a,b], Christian Schou[a] and Jan Dimon Bendtsen[b]

[a]Grundfos Holding A/S, Poul Due Jensens Vej 7, DK-8850, Bjerringbro, Denmark
[b]Section for Automation and Control, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark

Abstract—*Knowing where wastewater is flowing in drainage networks is essential to utilize system storage, predict overflows and to optimize system operation. Unfortunately, flow in gravity-driven sewers is subject to transport delays, and typically influenced by significant disturbances entering the sewer pipes in the form of domestic, ground and rain inflows. Model-based optimal control of urban drainage requires knowledge about these inflows, even though it is often not feasible in operational setups. To this end, we propose a lumped-parameter hydrodynamic model with a bi-linear structure for identifying the transport delays, decouple periodic disturbances and to predict the discharged flow. Pumped inlet and discharged dry-weather flow is used to find the model parameters. Under mild assumptions on the domestic and groundwater inflows, i.e. disturbances, the decoupling capabilities of the identified model are presented. A numerical case study on an EPA Storm Water Management Model (EPA SWMM) and experimental results on a real network demonstrate the proposed method.*

Keywords—*Process identification; Transport delay; Disturbance parameters; Open hydraulics*

Published in the *Proceedings of World Congress of the International Federation of Automatic Control*, 2020.

DOI: https://doi.org/10.1016/j.ifacol.2020.12.1295

Contents

1 Introduction

In sewers, waste water is collected and transported towards treatment plants, where contaminants are removed before releasing the water back to the environment [Butler and Davies, 2006]. Flow routing in sewers is a complex task, since the network is characterized by large spatial dimensions, nonlinear dynamics, large flow variations and significant time delays. In this work, a nonlinear system identification approach is proposed to predict flow and delays based on the simplified Saint-Venant (SV) equations.

Gravity-driven flow in open channels is represented by a system of coupled partial differential equations (PDEs). Due to the complexity of these PDEs, simplified and linearized models are typically used in optimal control design. In [Xu et al., 2011], the control effectiveness of reduced SV models in Model Predictive Control (MPC) has been studied. In [Leirens et al., 2010], a linearized SV model has been proposed for pumped sewer networks. Moreover, linear cascade modelling is a common approach in open water systems, e.g. in irrigation canals [Litrico and Fromion, 2004], and in inland waterways [Segovia et al., 2018]. Linearization, however, does not allow flow-dependent delays and the maximum allowed flow deviation from the steady-state solution is restricted.

In a previous paper, [Kallesøe and Knudsen, 2016], self-calibrating flow estimation has been developed for tracking the pump flow at the inlet, and for the discharged flow at the outlet of gravity pipes. This algorithm utilized information about wastewater pits and pump operation, hence applicable in any system configuration. However, prediction of flow dependant delays considering different disturbance inflows (i.e. domestic- and groundwater) has not been encountered yet. The current work utilizes this previously-established flow estimation algorithm as the source of training data, and proposes a modeling method for flow and delay prediction in long gravity sewers. While we do not explicitly address control strategies, the proposed identification approach is dedicated for the internal model of a predictive controller. The main compromise of establishing such a model is typically between complexity, accuracy and the computational burden [Lund et al., 2018]. In recent years, identification-based modeling has gained more attention, as data has become widely available at utilities. Yet, reports on Grey-Box modeling of open-channel water systems are relatively few [Su Ki Ooi and Weyer, 2003], [Weyer, 2001]. Moreover, research on Grey-Box modelling in different domains, e.g. [Sundar and Zlotnik, 2019], considered large-scale natural gas networks where state and parameter estimation have been developed using data and the underlying network graph.

In many applications in this framework (e.g. irrigation canals, sewage networks, etc.), linear physical models are used. In contrast to these approaches, we propose a nonlinear model structure which can describe a wide range of flows. The task of finding the correct physical parameters (e.g. length, shape, slope and friction) disappears with the proposed data-driven approach, thus enabling scalability to large systems with arbitrary structures.

Our approach is data-driven, yet we establish our proposed model struc-

ture based on physical considerations, familiar to those working in the water domain. In contrast to methods utilizing cross-correlation analysis e.g. for delay detection [Bjorklund and Ljung, 2003], and statistical blackbox models for disturbance prediction [Troutman et al., 2017], we aim to preserve intuition by giving physical interpretation of the flows in the system.

The remainder of the paper is structured as follows. Section 2 provides an overview of the system, whereupon in Section 3 we review the PDE model for gravity sewers and propose a model for the disturbance flows. In Section 4, we formulate the estimation as a least squares problem in the form of a Nonlinear Program (NLP), while in Section 5, we present numerical and experimental results. Finally, Section 6 sums up the contributions of the work. Throughout the paper, all quantities mentioned are real. We use boldface letters for sets, such as $\boldsymbol{s} = \{s_1, ..., s_n\}$, as well as for vectors $\boldsymbol{x} = [x_1, ..., x_n]^T \in \mathbb{R}^n$.

2 Sewer systems overview

We consider networks with transport lines over long distances. The layout of such network is shown in Figure 1.

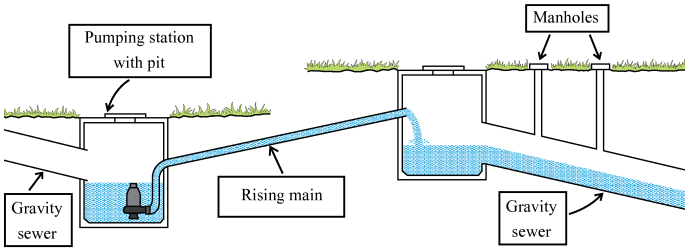


Figure 1: Topology of a pumped sewer network.

The sewage is first collected at a pumping station. The pumping station consists of a small storage tank (pit) and one or more pumping units. The collected water is then pumped through a rising main, whereupon it enters a long gravity-driven sewer channel. Typically, pumps operate in combination and deliver flow at a fixed rate governed by local on/off controllers (see [Schütze et al., 2002]). When the pump state is off and the pit volume reaches the maximum threshold, the pumps turn on. Then, the pumped flow arrives to the next pit in line with a delay, thereby forcing the next station to turn on, and so forth. Exogenous inflows, i.e. disturbances, enter gravity sewers in the form of domestic waste, rainfall run-off and through leakages allowing groundwater to infiltrate into the channel.These disturbances are characterized by specific flow patterns, shown in Figure 2.

Households are common sources of inflows in sewer. Due to the large spatial dimension of sewers, domestic waste collected from urban areas may enter the network along the gravity pipes. Domestic waste is typically characterized by a

diurnal pattern which has an inherent periodicity of 24 hours with peak points in the morning and in the afternoon (see Pattern *a* in Figure 2).

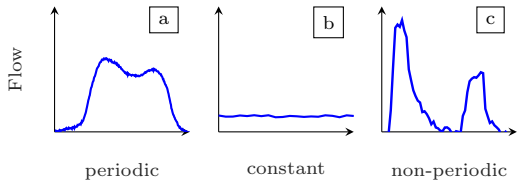


Figure 2: Disturbances occurring in sewer networks.

Groundwater is present when rain runs off slowly and thereby water accumulates in the ground. In this case, water may infiltrate gravity channels through leakages with a nearly constant flow (see Pattern *b* in Figure 2).

The third type of disturbance flow is due to rainfall run-off. These discharges are disregarded and we rather focus on domestic and groundwater infiltration. Indeed, this is a common practice (see [Courdent et al., 2018]), as typically the domestic and groundwater flows are estimated first, using inflow measurements from dry-weather periods.

3 Modeling

3.1 Flow model

Unsteady, one-dimensional water flow in gravity sewers can be computed accurately by the well-known Saint-Venant PDEs [Schütze et al., 2002]

$$\frac{\partial A_{x,t}}{\partial t} + \frac{\partial Q_{x,t}}{\partial x} = q_{x,t}, \tag{1a}$$

$$\frac{\partial Q_{x,t}}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q_{x,t}^2}{A_{x,t}}\right) + gA_{x,t}\left(\frac{\partial h_{x,t}}{\partial x} + S_f + S_b\right) = 0, \tag{1b}$$

where $Q_{x,t}$ denotes the flow inside the channel, $q_{x,t}$ is the disturbance inflow, $A_{x,t}$ is the wetted channel area and $h_{x,t}$ is the water level. Moreover, $Q_{x,t}, q_{x,t}, A_{x,t}$ and $h_{x,t}$ are functions from $(0, L) \times \mathbb{R}_+$ to $\mathbb{R}_+$. The gravitational acceleration is $g \in \mathbb{R}_+$, furthermore we assume that $S_b \in \mathbb{R}_+$ bed slope and $S_f \in \mathbb{R}_+$ friction slope parameters are independent of x and t, which is a fair assumption if the slope variance is small [Schütze et al., 2002].

For simplicity, we assume kinematic waves, meaning that we neglect the first three terms in eq.(1b) [Singh, 2001], which results in a balance between friction and gravitational forces. For determining S_f in eq.(1b), we utilize Manning's equation [Schütze et al., 2002]. Then, eq.(1b) simplifies to

$$S_b = S_f(Q_{x,t}, h_{x,t}) = \frac{n^2 Q_{x,t}^2}{A_{x,t}^2 R_{x,t}^{\frac{4}{3}}}, \tag{2}$$

where $R = \frac{A}{P}$ is the hydraulic radius, $P \in \mathbb{R}_+$ is the wetted perimeter and $n \in \mathbb{R}_+$ is the Manning coefficient. Simplification with the Manning formula restricts the flows to be one-directional, meaning that the phenomena of backwater effect is not considered. Backwater occurs when the channel is overloaded, thus water surcharges. This is typically negligible in large gravity lines, transporting waste water over long distances Singh [2001].

We argue that semi-filled circular sewers are well-approximated by a rectangular shape, shown in Figure 3.

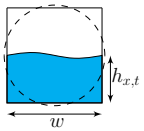


Figure 3: Rectangular channel, where $w \in \mathbb{R}_+$ is the width.

The hydraulic radius of a channel is parametrized by w and approximated by

$$R_{x,t} = \frac{A_{x,t}}{P_{x,t}} \approx \frac{wh_{x,t}}{2h_{x,t} + w}, \tag{3}$$

where a linear area-level relation is used. The independent variables remaining in the SV equations are flow and level on the domain $(0, L) \times \mathbb{R}_+$, given by

$$w\frac{\partial h_{x,t}}{\partial t} + \frac{\partial Q_{x,t}}{\partial x} = q_{x,t}, \tag{4a}$$

$$Q_{x,t} = \frac{\sqrt{S_b}}{n}\frac{(wh_{x,t})^{\frac{5}{3}}}{(2h_{x,t} + w)^{\frac{2}{3}}}. \tag{4b}$$

The equation in eq.(4b) comes from eq.(2) and eq.(3). The coupled PDEs in equations eq.(4a) and eq.(4b) describe kinematic waves traveling through open channels. Note, that semi-fillness of the channel is an assumption which does not always hold. The formulation presented here does not hold for fully-filled flow conditions.

3.2 Discretized model

We formulate the physical model in a form more amenable to system identification. The channel is partitioned into N_x equal-sized, non-overlapping δx segments of length, while $h_{x,t}$, $Q_{x,t}$ and $q_{x,t}$ are approximated as piece-wise constant functions of x. The spatial discretization is shown in Figure 4. We use backward Euler discretization for the spatial and forward Euler for the time coordinate. The left boundary is defined at $x = 0$ and the right boundary at $x = L$. Discretizing equations eq.(4a) and eq.(4b), the model may be recast as

$$h_{x,t+\delta t} = h_{x,t} + \alpha q_{x,t} + \beta_1(Q_{x-\delta x,t} - Q_{x,t}), \tag{5a}$$

$$Q_{x,t} = \beta_2 \frac{h_{x,t}^{\frac{5}{3}}}{(h_{x,t} + \gamma)^{\frac{2}{3}}}, \quad \forall\, x \in (0, L) \tag{5b}$$

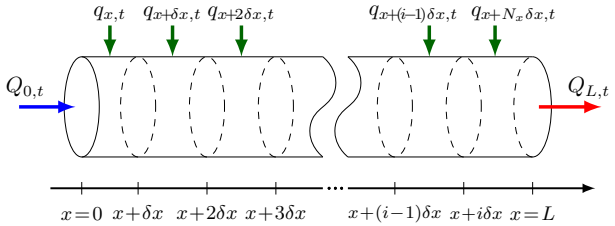


Figure 4: Channel divided into N_x sections, where $i \in \{1, 2, ..., N_x\}$.

where δt is the sampling time and we define the parameters

$$\alpha \triangleq \frac{\delta t}{w}, \quad \beta_1 \triangleq \frac{\delta t}{w \delta x}, \quad \beta_2 \triangleq \frac{\sqrt{S_b} w^{\frac{5}{3}}}{2^{\frac{2}{3}} n}, \quad \gamma \triangleq \frac{w}{2}, \tag{6}$$

where $\alpha, \beta_1, \beta_2, \gamma \in \mathbb{R}_+$. Note that the time and spatial steps, δt and δx, are part of the parameters α and β_1. This implies that eq.(5a) and eq.(5b) are affected by the choice of the sampling time and the section size. Indeed, δt and δx affects the dynamics by introducing distortion in the traveling wave (see [Singh, 2001]). Next, we insert the section flows $Q_{x,t}$ from eq.(5b) into eq.(5a), so we get

$$h_{x,t+\delta t} = h_{x,t} + \alpha q_{x,t} + \beta_1 \beta_2 \left[\frac{h_{x-\delta x,t}^{\frac{5}{3}}}{(h_{x-\delta x,t} + \gamma)^{\frac{2}{3}}} - \frac{h_{x,t}^{\frac{5}{3}}}{(h_{x,t} + \gamma)^{\frac{2}{3}}} \right] \tag{7}$$

The state equation in eq.(7) is parametrized by α, β_1, β_2 and γ. In order to reduce the number of parameters, and thus avoid non-identifiability, we attempt to restructure eq.(7) by removing γ from the denominator of the nonlinear expression. Therefore, we define new states such that

$$h_{x,t} \triangleq \gamma z_{x,t}, \quad \forall\ x \in (0, L), \tag{8}$$

where $z_{x,t}$ are the scaled equivalents of the physically measurable water levels $h_{x,t}$, i.e. the transformed states. Now let us define a mapping $g : \mathbb{R}_+ \to \mathbb{R}_+$

$$g : (z_{x,t}) \mapsto \frac{z_{x,t}^{\frac{5}{3}}}{(z_{x,t} + 1)^{\frac{2}{3}}}, \quad \forall\ x \in (0, L). \tag{9}$$

Utilizing the state transformation and the nonlinear mapping g, we recast the state equation in eq.(7).

This yields as a set of coupled bi-linear difference equations, describing the states in the N_x partitioned sections along the channel.

$$\begin{aligned} z_{0,t+\delta t} &= z_{0,t} + \tilde{q}_{0,t} + \theta_1 Q_{0,t} - \theta_1\theta_2 g(z_{\delta x,t}), \\ &\vdots \\ z_{x,t+\delta t} &= z_{x,t} + \tilde{q}_{x,t} + \theta_1\theta_2\big(g(z_{x-\delta x,t}) - g(z_{x,t})\big), \\ &\vdots \\ z_{L,t+\delta t} &= z_{L,t} + \tilde{q}_{L,t} + \theta_1\theta_2\big(g(z_{L-\delta x,t}) - g(z_{L,t})\big), \end{aligned} \tag{10}$$

where we defined $\tilde{q}_{x,t} \triangleq \frac{\alpha}{\gamma} q_{x,t}$ as scaled disturbance flows. Furthermore, the task of parameter estimation has been reduced to find the parameters $\theta_1, \theta_2 \in \mathbb{R}_+$ and the unknown disturbances $\tilde{q}_{x,t}$. The parameters are given by

$$\theta_1 \triangleq \frac{\beta_1}{\gamma}, \quad \theta_2 \triangleq \beta_2\gamma. \tag{11}$$

We consider the upstream boundary flow $Q_{0,t}$ (hereinafter Q_{in}), as the control input. The output of the control model is the downstream flow at the boundary $x = L$, which we hereinafter call Q_{out}. The output equation is then

$$Q_{\text{L,t}} = \theta_2 g(z_{L,t}), \tag{12}$$

which is the reformulated Manning equation in eq.(5b).

3.3 Disturbance model

In this application, we consider periodic domestic waste flows coming from urban areas and groundwater infiltration which we assume to be constant in time. Here, a Fourier series model is presented to estimate the periodic disturbance signals $\tilde{q}_{x,t}$ entering into the dynamic flow model in eq.(10). It is assumed that the infiltration of groundwater is uniformly distributed along the partitioned channel sections, i.e. groundwater enters each section with the same amplitude. Moreover, it is assumed that we know where residential areas are connected by pipelines, i.e. where domestic waste water enters the channel. This is a fair assumption, as typically the piping layout of the infrastructure is stored in, e.g. a GIS (Geographic Information System) database at most utilities. Hence, the scaled disturbances entering the i^{th} section at time t are given by

$$\begin{aligned} \tilde{q}_{i,t}(\lambda_0, \boldsymbol{\lambda}) &\triangleq \tilde{q}_i^{\text{gnd}}(\lambda_0) + \tilde{q}_{i,t}^{\text{dom}}(\boldsymbol{\lambda}) \\ &\triangleq \underbrace{\lambda_0 + \lambda_1 + \sum_{j=1}^{k} \big(\lambda_{1j} cos(j\omega t) + \lambda_{2j} sin(j\omega t)\big)}_{f_{i,t}(\boldsymbol{\lambda})} \end{aligned} \tag{13}$$

where the set of parameters regarding domestic flows are $\boldsymbol{\lambda} \triangleq \{\lambda_1, \lambda_{11}, \lambda_{21}, ..., \lambda_{1k}, \lambda_{2k}\} \in \mathbb{R}$, and $\lambda_0 \in \mathbb{R}_+$ represents constant groundwater flows.

The angular frequency ω corresponds to a period of one day and $k \geq 2$ is the number of frequency terms in the truncated Fourier series. Furthermore, $f_{i,t}(\boldsymbol{\lambda})$ is a family of functions parametrised by $\boldsymbol{\lambda}$, and for each $\boldsymbol{\lambda}$, $f_{i,t} : (1, N_x) \times \mathbb{R}_+ \to \mathbb{R}$.

Note, that the model describing the scaled disturbances does not correspond to the real domestic and groundwater flows. In order to calculate the real disturbances $q_{i,t}$ from the scaled estimates $\tilde{q}_{i,t}$, recall that the disturbance flows are scaled by the model parameters, such that

$$\tilde{q}_{i,t} \triangleq \frac{\alpha}{\gamma} q_{i,t}, \quad \forall\ i \in (0, N_x). \tag{14}$$

Then, using the estimated parameters defined in eq.(11) and the physical model parameters in eq.(6), it is seen that

$$\sum_{i=1}^{N_x} q_{i,t} \delta x = \sum_{i=1}^{N_x} \frac{\tilde{q}_{i,t}}{\theta_1} = \sum_{i=1}^{N_x} \frac{\alpha}{\beta_1} q_{i,t} = q_{i,t} \delta x N_x, \tag{15}$$

where we use θ_1 to calculate $q_{i,t}$. The last term in eq.(15) represents all disturbance flows along the total pipe length.

Now, let us consider the function $f_{i,t}(\boldsymbol{\lambda})$ in eq.(13). From this, we form the vector $\boldsymbol{f}_t(\boldsymbol{\lambda})$, such that the function at index i is $f_{i,t}(\boldsymbol{\lambda})$, i.e. $\boldsymbol{f}_t : \mathbb{R}_+ \to \mathbb{R}^{N_x}$. Using the vector representation and the relation shown in eq.(15), we decompose the domestic and groundwater flows such that

$$\sum_{i=1}^{N_x} q_{i,t} \delta x = \underbrace{\frac{\lambda_0}{\theta_1} N_x}_{q^{\mathrm{gnd}}(\lambda_0, \theta_1)} + \underbrace{\frac{1}{\theta_1} \boldsymbol{b}^T \boldsymbol{f}_t(\boldsymbol{\lambda})}_{q^{\mathrm{dom}}(\boldsymbol{\lambda}, \theta_1)}, \tag{16}$$

where q^{gnd} is the ground infiltration flow summed over all pipe sections and q^{dom} is the total domestic inflow. The vector $\boldsymbol{b} \in \{0, 1\}^{N_x}$ has ones in its entries regarding section points where urban areas are connected, and zeros with no connection. The number of parameters that we need to identify is $4 + 2k$, depending on order of the Fourier series.

4 System identification

The system identification is given as a constrained nonlinear least squares problem, where samples, at time t_i, $i = \{0, ..., N_t\}$, of the pumped inlet flows Q_{in} and the discharged Q_{out} flows are known and estimated a priori. Let $\boldsymbol{\theta} \triangleq \{\theta_1, \theta_2\} \in \mathbb{R}_+$ denote the set of system parameters and $\boldsymbol{\Lambda} \triangleq \{\lambda_0, \boldsymbol{\lambda}\} \in \mathbb{R}$ denote the parameters corresponding to the disturbance flows.

The parameters $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$ and the initial states $\boldsymbol{z}(t_0)$ are found by solving the NLP

$$\begin{pmatrix} \boldsymbol{\theta}^* \\ \boldsymbol{\Lambda}^* \\ \boldsymbol{z}^*(t_0) \end{pmatrix} = \underset{\boldsymbol{\theta},\boldsymbol{\Lambda},\boldsymbol{z}(t_0)}{\operatorname{argmin}} \sum_{i=0}^{N_t} \left(Q_{out}(t_i) - \hat{Q}_{out}(t_i)\right)^2 \tag{17a}$$

s.t.

$$\boldsymbol{z}(t_{i+1}) = \boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\Lambda}}(\boldsymbol{z}(t_i), Q_{in}(t_i)), \tag{17b}$$

$$\hat{Q}_{out}(t_i) = H_{\theta_2}(z_L(t_i)), \tag{17c}$$

$$\mathbf{0} \le \boldsymbol{z}(t_i) \le \overline{\boldsymbol{z}}, \tag{17d}$$

$$\mathbf{0} \le \boldsymbol{\theta} \le \overline{\boldsymbol{\theta}}, \tag{17e}$$

where $\boldsymbol{z}(t_i) \in \mathbb{R}^{N_x}$ is the vector of states in eq.(10) and the dynamics in eq.(17b) are represented by $\boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\Lambda}}(\boldsymbol{z}(t_i), Q_{in}(t_i)) : \mathbb{R}_+ \to \mathbb{R}^{N_x}$. The function $H_{\theta_2}(z_L(t_i)) : \mathbb{R}_+ \to \mathbb{R}_+$ in equation eq.(17c) represents the scalar output where z_L corresponds to the downstream boundary state. The constraints in eq.(17d) and eq.(17e) impose bounds on the transformed state variables and parameters, respectively.

In the above NLP, we assumed that the number of states are fixed, i.e. N_x is given. Instead, we introduce N_x as an auxiliary variable in the model. Hence, we carry out estimations multiple times on equivalent models but with different grid sizes as explained in the algorithm below

Algorithm 2 Model evaluation for different N_x

Input: Q_{in}, Q_{out}, $\overline{\boldsymbol{z}}$, $\overline{\boldsymbol{\theta}}$
1: **repeat** at every iteration $k = 1,2,\ldots$ N_x
2: Initialize: $\boldsymbol{\theta}$, $\boldsymbol{\Lambda}$, $\boldsymbol{z}(t_0)$
3: Solve: $\min_{\boldsymbol{\theta},\boldsymbol{\Lambda},\boldsymbol{z}(t_0)} \sum_{i=0}^{N_t} \left(Q_{out}(t_i) - \hat{Q}_{out}(t_i)\right)^2$
4: s.t. constraints
5: **until** $RMSE(\hat{Q}_{out}^{k-1}) < RMSE(\hat{Q}_{out}^{k})$
Output: $\boldsymbol{\theta}^*$,$\boldsymbol{\Lambda}^*$,$\boldsymbol{z}^*(t_0)$, N_x

The model is evaluated for each trial of N_x using Root Mean Squared Errors (RMSE) and the algorithm is terminated at the lowest N_x. The estimation accuracy is shown in Figure 5 for a selection of k iteration steps.

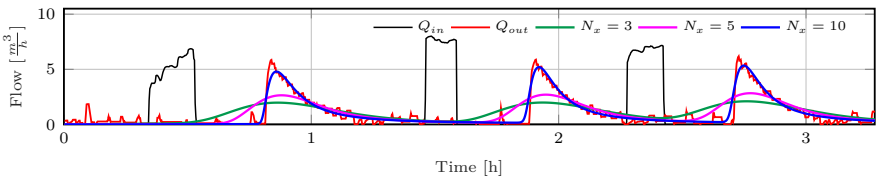


Figure 5: Model accuracy for different N_x tested on real data.

As seen, increasing N_x above a threshold does not increase accuracy.

To evaluate the convergence of Algorithm 1, the RMSE at each iteration is calculated.

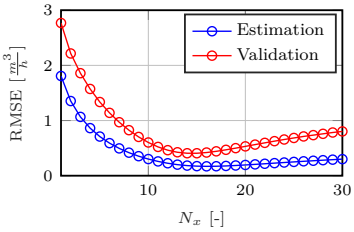


Figure 6: Estimation and validation evaluated against N_x.

As seen in Figure 6, there is an optimal selection for N_x where model accuracy is the highest for the provided training and validation datasets.

5 Results

5.1 Numerical results

Results of applying the method are first presented on a numerical case study in the EPA SWMM[Rossman, 2015] simulation software. The test model is depicted in Figure 7.

In this network, a single sewer line is considered, transporting the sewage from a pumping station to an outlet point, representing either the next pumping station or the treatment plant. We consider an urban area discharging to the transportation line at $x = L/2$. The pumped inlet flows Q_{in} enter the sewer at the upstream and we observe the discharged flows Q_{out} at the downstream, indicated in Figure 7. Moreover, measurement noise is added to the simulated $Q_{out}(t_i)$ flows with the property of $n \sim \mathcal{N}(0, 0.2)$.

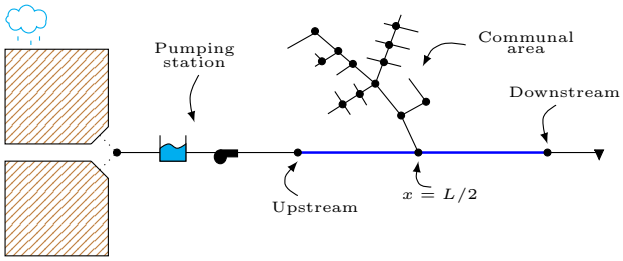


Figure 7: Schematics of the EPA SWMM simulation model.

In simulation, we attempt to mimic the behavior of a real scenario, where the wastewater pit collects non-periodic runoff water with a variety of rain intensity, forcing the pump to turn on for different time duration.

The validation of the identified model is shown in Figure 8.

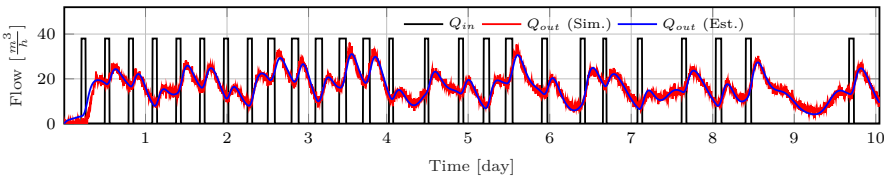


Figure 8: Discharged downstream flow prediction.

$N_x = 4$ sections resulted in good model accuracy. Note, that the discharged flow shown in Figure 8 consists of the delayed non-periodic pumped flows Q_{in} and the periodic disturbance inflows q. In EPA SWMM, we can access the disturbances q for validating our results. The disturbance q entering the network at $x = L/2$ and discharged at the end of the channel is shown in Figure 9.

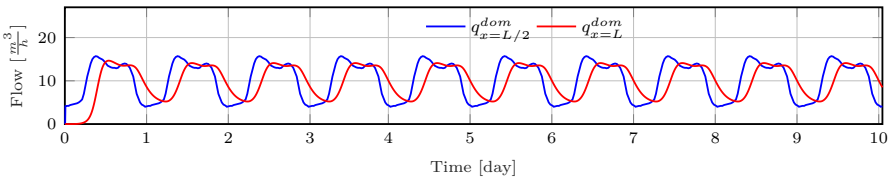


Figure 9: Upstream and discharged domestic waste flows.

The domestic waste flow arrives to the downstream with a flow-dependent delay. Using the estimated parameter θ_1 and the disturbance model defined in Section 4, the decoupled domestic flow $q^{dom}_{x=L/2}$ yields as shown in Figure 10.

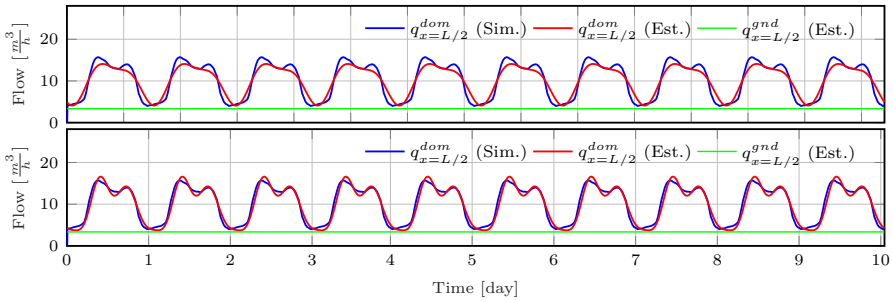


Figure 10: Disturbance decoupling with $2nd$ (above) and $4th$ order Fourier series (below).

Figure 10 shows that we are able to estimate the periodic domestic waste flows and the constant groundwater infiltration at the point where they enter the system, by using the disturbance model and the identified model parameters. The disturbances q^{dom} and q^{gnd} in Figure 10 correspond to the decoupled flows in eq.(16). With higher order Fourier series, the estimation is more precise, however at the cost of increasing the number of parameters.

5.2 Experimental results

We also present results of applying the system identification method on a real world case study. The available data is flow estimation, extracted from a sewer network, operated by Provas A/S, located in Gram, Denmark. The pipe layout of the drainage network is shown in Figure 11.

This particular segment of the network consists of seven pits with corresponding pumping stations. The estimation data has been sampled at 1 Hz and gathered from the gravity sewers connected by PGH103-104 and PGH202-203 pumping stations. (For detailed explanation of the flow estimation method utilized in this work, consult [Kallesøe and Knudsen, 2016]). In the two test scenarios, urban areas are not connected, therefore our tests have been restricted to groundwater detection.

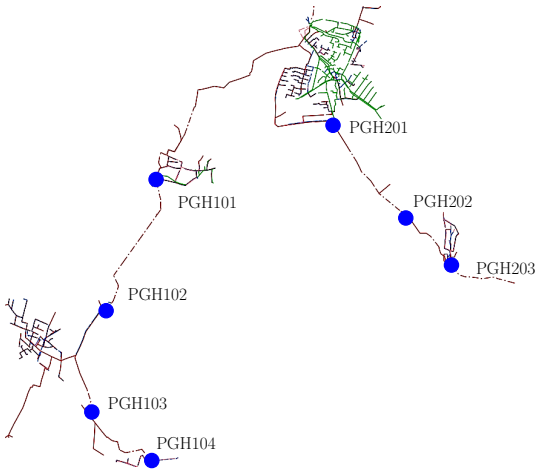


Figure 11: A segment of a combined sewer network, where blue dots denote waste water pumping stations.

The model validations are shown for the two tests in Figure 12. The estimation data covers two days in both cases. In the graph above, groundwater infiltration is approximately zero, meaning that between each pump cycle the discharged flow becomes zero, thus the channel dries out. However in the graph below, groundwater infiltration is significant, meaning that initial water level estimation is necessary. The model error yielded sufficiently small with $N_x = 7$ channel sections in both cases. As shown, the model describes the flow-dependent delays accurately under significant ground water infiltration and under a large variety of pumped input flows.

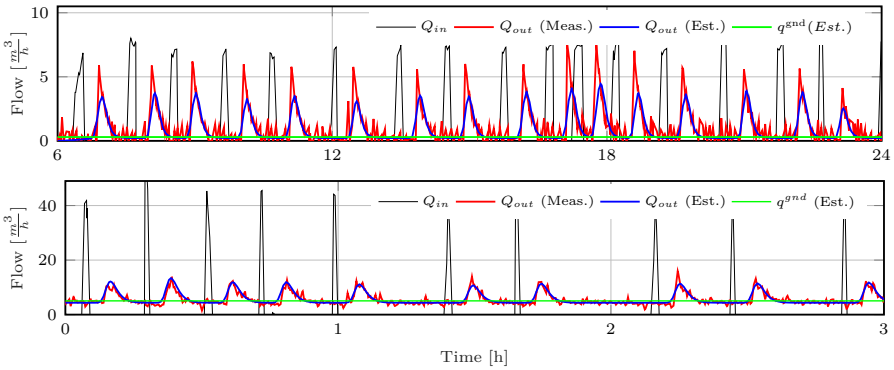


Figure 12: Model validation on experimental data. Gravity sewer between PGH103-104 stations above, and between PGH202-203 stations below.

6 Conclusion

The presented paper focused on detecting and decoupling periodic and constant disturbance flows from the total discharge in gravity sewers. To this end, a data-driven identification method has been proposed based on physical models. The method has been tested in simulation and on data from a real network. The implementation has shown that the identification is feasible and that the estimated models predict flow and transport delays with high accuracy. The main advantage of the data-driven aspect of the modeling is that the method becomes scalable to a variety of networks, having different structures and physical parameters. Additionally, using a physical model carries an advantage of restricting the parameter space.

In our future work, we focus on stability and identifiability. Note, that the bound of the physical parameters have been chosen in a heuristic fashion. Furthermore, the state transformation resulted in loss of insight regarding water levels. Including the above considerations, we consider utilizing the models in an MPC framework.

Acknowledgements

The authors would like to thank to Provas A/S for letting us use data from their network. This project was supported by Innovation Fund Denmark (Ref. *9065-00018A*).

References

S. Bjorklund and L. Ljung. "A review of time-delay estimation techniques". In *Proceedings of the 42nd IEEE International Conference on Decision and Control(CDC)*, pages 2502–2507, Maui, USA, 2003. doi:10.1109/cdc.2003.1272997.

D. Butler and J. W. Davies. *Urban Drainage*. Spon Press, 2006. doi:10.1016/s1462-0758(00)00017-0.

V. Courdent, M. Grum, and P. S. Mikkelsen. "Distinguishing high and low flow domains in urban drainage systems 2 days ahead using numerical weather prediction ensembles". *Journal of Hydrology*, 556:1013–1025, 2018. ISSN 00221694. doi:10.1016/j.jhydrol.2016.08.015.

C. S. Kallesøe and T. Knudsen. "Self calibrating flow estimation in waste water pumping stations". In *2016 European Control Conference, ECC 2016*, pages 55–60, 2016. ISBN 9781509025916. doi:10.1109/ECC.2016.7810263.

S. Leirens, J. M. Giraldo, R. R. Negenborn, and B. De Schutter. "A pattern search method for improving the operation of sewer systems". In *IFAC Proceedings Volumes (IFAC-PapersOnline)*, pages 591–596, 2010. ISBN 9783902661913. doi:10.3182/20100712-3-fr-2020.00096.

X. Litrico and V. Fromion. "Simplified Modeling of Irrigation Canals for Controller Design". *Journal of Irrigation and Drainage Engineering*, 130(5):373–383, 2004. ISSN 0733-9437. doi:10.1061/(asce)0733-9437(2004)130:5(373).

N. S. V. Lund, A. K. V. Falk, M. Borup, H. Madsen, and P. Steen Mikkelsen. "Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management". *Critical Reviews in Environmental Science and Technology*, 48(3):279–339, 2018. ISSN 15476537. doi:10.1080/10643389.2018.1455484.

L. Rossman. "Storm Water Management Model User's Manual". *US Environment Protection Agency*, 2015.

M. R. Schütze, D. Butler, and M. B. Beck. *Modelling, Simulation and Control of Urban Wastewater Systems*. Springer, 2002. doi:10.1007/978-1-4471-0157-4.

P. Segovia, L. Rajaoarisoa, F. Nejjari, E. Duviella, and V. Puig. "Input-Delay Model Predictive Control of Inland Waterways Considering the Backwater Effect". In *2018 IEEE Conference on Control Technology and Applications*, pages 589–594, 2018. ISBN 9781538676981. doi:10.1109/CCTA.2018.8511553.

V. P. Singh. "Kinematic wave modelling in water resources: A historical perspective". *Hydrological Processes*, 15:671–706, 2001. ISSN 08856087. doi:10.1002/hyp.99.

Su Ki Ooi and E. Weyer. "Closed loop identification of an irrigation channel". In *Proceedings of the 40th IEEE Conference on Decision and Control*, pages 4338–4343, Orlando, Florida, 2003. doi:10.1109/cdc.2001.980883.

K. Sundar and A. Zlotnik. "Dynamic State and Parameter Estimation for Natural Gas Networks using Real Pipeline". In *CCTA 2019 - 3rd IEEE Conference on Control Technology and Applications*, 2019. ISBN 9781728127675. doi:10.1109/CCTA.2019.8920430.

S. C. Troutman, N. Schambach, N. G. Love, and B. Kerkez. "An automated toolchain for the data-driven and dynamical modeling of combined sewer systems". *Water Research*, 126:88–100, 2017. ISSN 18792448. doi:10.1016/j.watres.2017.08.065.

E. Weyer. "System identification of an open water channel". *Control Engineering Practice*, 33(15):265–270, 2001. ISSN 09670661. doi:10.1016/S0967-0661(01)00099-5.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow". *Advances in Water Resources*, 34:282–290, 2011. ISSN 03091708. doi:10.1016/j.advwatres.2010.11.009.

Paper B

Nonlinear Grey-box Identification of Sewer Networks with the Backwater Effect: An Experimental Study

Krisztian Mark Balla[a,b], Casper Houtved Knudsen[b], Adis Hodzic[b] and Jan Dimon Bendtsen[b] Carsten Skovmose Kallesøe[a,b]

[a]Grundfos Holding A/S, Poul Due Jensens Vej 7, DK-8850, Bjerringbro, Denmark
[b]Section for Automation and Control, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark

Abstract—*Real-time control of urban drainage networks requires knowledge about stored volumes and flows in order to predict overflows and optimize system operation. However, using flow sensors inside the pipelines means prohibitively high installation and maintenance costs. In this article, we formulate two nonlinear, constrained estimation problems for identifying the open-channel flow in urban drainage networks. To this end, we distribute cost-efficient level sensors along the pipelines and formulate the estimation problems based on the spatially-discretized kinematic and diffusion wave approximations of the full Saint-Venant partial differential equations. To evaluate the capabilities of the two models, the two approaches are compared and evaluated on modeling a typical phenomenon occurring in drainage systems: the backwater effect. An extensive real-world experiment demonstrates the effectiveness of the two approaches in obtaining the model parameters on a scaled water laboratory setup, in the presence of measurement noise.*

Keywords—*System identification; Backwater effect; Kinematic wave, Diffusion Wave*

Published in the *Proceedings of Conference on Control Technology and Applications*, 2021.

DOI: https://doi.org/10.1109/CCTA48906.2021.9658864

Contents

1 Introduction

Urban Drainage Networks (UDNs) are large-scale systems where sewage is transported in open-channel conduits towards the root of the network, where it is treated before being discharged to the environment. In this article, we focus on pumped systems where the sewage is pumped to overcome the elevation and levelness of the landscape, then allowed to flow by gravity towards the next collection pit in line [Schütze et al., 2002]. In order to control the volumes in UDNs, the flows and stored volumes are essential to know to schedule unavoidable overflows and regulate the inlet to the treatment plant [Gelormino and Ricker, 1994].

Transport in UDNs is a complex process due to its nonlinear nature and to the large time-delays imposed by long travel times between wastewater pumping stations. The transport processes are modeled by Partial Differential Equations (PDEs), where the level and flow of water appear as independent variables. However, these PDEs are often too complex to solve in real-time applications and require the precise network dimensions or a High Fidelity (HiFi) model of the UDN [Ocampo-Martinez, 2010], [Litrico and Fromion, 2009].

Some papers report on using the full dynamic PDEs in control of UDNs [Xu et al., 2012] and modeling of open-channel water infrastructures, e.g., irrigation canals and river systems. However, using full-PDE models requires a HiFi simulator or installing several flow sensors along the pipelines, resulting in high installation and maintenance costs. To overcome the difficulties with model complexity, some research proposes to use reduced sewer models, relying on the physical attributes (pipe dimensions, friction, slope) of sewer pipes [Xu et al., 2011]. The most common approximations of the original PDEs are the Kinematic Wave (KW) and Diffusion Wave (DW) methods [Singh, 2001], where the original model is simplified by omitting several physical phenomena in the model [Evans et al., 2012]. The most common phenomena of this type is the backwater effect, which is a local flow reversal that occurs inside the pipelines when the capacity is overloaded and water volumes are accumulating [Litrico and Fromion, 2009].

Linearizing PDEs around operating points has been extensively used in slowly-varying water applications such as river control, but also in UDNs by means of transfer functions capturing the backwater effect in [Litrico and Fromion, 2006], and by state-space models in [Zou et al., 2015]. Furthermore, due to its simplicity, delay-models have been used in predictive control in many UDN applications as well [Balla et al., 2020b]. Data-driven modelling has been reported in [Balla et al., 2020a] and [Troutman et al., 2017], where KW approximation-based, grey-box modelling and black-box approaches have been extensively used to model the gravity-driven sewer flows.

In this article, experiments are carried out for both the KW and DW approximations of the full PDEs to compare the model performance in system identification. In contrast to methods relying on flow measurements and HiFi models, we consider level sensors distributed along the network and utilize flow

estimation techniques to show the identifiability of the KW- and DW-based models. Our approach is data-driven, yet the models rely on the hydraulics of the network, familiar to operators in the open-channel, water infrastructures industry. Furthermore, the proposed technique using the physically-based process models carries an advantage that it only requires data collection under nominal operation, unlike conventional data-driven methods.

The remainder of the paper is structured as follows. Section 2 introduces the KW and DW approximation principles of the original PDEs. In Section 3, we formulate the system identification problem for both cases, whereupon Section 4 describes the case study laboratory setup. The results are provided in Section 5, where sensor and estimation data from the experiments are utilized. This is followed by Section 6, where conclusions and future research directions are provided.

1.1 Nomenclature

Throughout the paper, all quantities mentioned are real values. Boldface letters are used for sets, such as $\mathbf{s} = \{s_1, ...s_n\}$ as well as for vectors $\mathbf{x} = (x_1, ...x_n)^\top \in \mathbb{R}^n$. In case of vectors, $<, \leq, =, >, \geq$ denote element-wise relations. Moreover, for a vector $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$ denotes the Euclidean norm.

2 Modeling

Open-channel flow in UDNs is typically modelled by the shallow water equations, which are given in uni-directional form by the Saint-Venant (SV) PDEs [Schütze et al., 2002]. These PDEs describe the mass balance and the momentum conservation of the fluid, shown in eq.(1a) and eq.(1b), respectively:

$$\frac{\partial A_{x,t}}{\partial t} + \frac{\partial q_{x,t}}{\partial x} = \tilde{d}_{x,t}, \tag{1a}$$

$$\frac{1}{gA_{x,t}}\left(\frac{\partial q_{x,t}}{\partial t} + \frac{\partial}{\partial x}\left(\frac{q_{x,t}^2}{A_{x,t}}\right)\right) + \underbrace{\frac{\partial h_{x,t}}{\partial x} + \overbrace{S_{f|x,t} - S_b}^{\text{Kinematic wave}}}_{\text{Diffusion wave}} = 0, \tag{1b}$$

where $A_{x,t}$ is the wetted area, $q_{x,t}$ is the open-channel flow inside the pipes, $h_{x,t}$ is the water level and $\tilde{d}_{x,t} = d_{x,t}/\delta x$ represent lateral inflows per unit length, where $d_{x,t}$ is the lateral inflow, hereinafter referred to as disturbance. Moreover, all variables mentioned are functions mapping from $(0, L) \times \mathbb{R}_+ \to \mathbb{R}_+$, where L represents the total length of the pipe. The gravity constant is denoted by g and the full dynamic SV-based PDEs are parametrized by the S_b slope and the $S_{f|x,t}$ friction terms. Note, that S_b is independent of the spatial and temporal coordinates, as we assume that the slopes of the modelled pipe segments are close to being constant throughout the pipe length. Moreover, different approximations to the SV-based PDEs exist, depending on

which terms are neglected in eq.(1b). In this study, we focus on the Kinematic and Diffusion wave approximations.

Assumption B1. *It is assumed that semi-filled pipe segments with a given geometry are sufficiently well-approximated with a rectangular pipe shape. Hence, the wetted area $A_{x,t}$ and perimeter $P_{x,t}$ are approximated as:*

$$A_{x,t}(h_{x,t}) \approx w h_{x,t}, \tag{2a}$$
$$P_{x,t}(h_{x,t}) \approx w + 2h_{x,t}, \tag{2b}$$

where w is the width of the pipe. Note, that the accuracy of the linear approximation varies according to the operating level inside the pipes, i.e., how full the pipe segment is.

In the following, the KW and DW reduction of the SV-based PDEs and the geometry simplifications are considered.

2.1 Kinematic wave approximation

When approximating the full SV-based PDEs with kinematic waves, we assume uniform, quasi-steady flow and neglect the physical phenomena such as the backwater effect by keeping only the $S_{f|x,t}$ friction and S_b slope terms in eq.(1b). Hence, the gravitational and friction forces acting on the fluid are equal.

In order to relate the friction to flows $q_{x,t}$ and water levels $h_{x,t}, x \in (0, L)$, we utilize the Manning formula [Roberson and Crowe, 1993]:

$$S_{f|x,t} = n^2 \frac{P_{x,t}^{4/3}}{A_{x,t}^{10/3}} q_{x,t}^2, \tag{3}$$

where $P_{x,t}$ is the wetted perimeter and n is the Manning friction factor. Using Assumption 1, the friction term in eq.(3) is rewritten as:

$$S_{f|x,t} = n^2 \frac{(w + 2h_{x,t})^{4/3}}{(w h_{x,t})^{10/3}} q_{x,t}^2. \tag{4}$$

Then, isolating $q_{x,t}$ and inserting back into the SV-based PDEs for a square pipe, the KW approximation yields PDEs of first-order in the form:

$$w \frac{\partial h_{x,t}}{\partial t} + \frac{\partial q_{x,t}}{\partial x} = \frac{d_{x,t}}{\delta x}, \tag{5a}$$

$$q_{x,t} = \frac{\sqrt{S_b}}{n} \frac{(w h_{x,t})^{5/3}}{(2h_{x,t} + w)^{2/3}}. \tag{5b}$$

Note that $q_{x,t}$ only depends on $h_{x,t}$, implying that the mapping from level to flow is injective.

To obtain a model structure more amenable to system identification, the PDEs in eq.(5a) and eq.(5b) are reduced to a system of finite-dimensional Ordinary Differential Equations (ODEs) by discretizing each pipe spatially as shown in Figure 1.

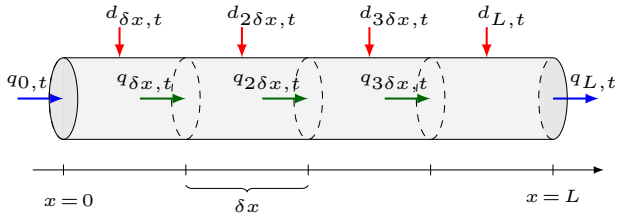


Figure 1: Flow balances for a pipe discretized into $N_x = 4$ sections.

Note that for each segment, the flow balance incorporates the lateral inflows $d_{x,t}$, the boundary flows $q_{0,t}$, the discharge flow $q_{L,t}$ and the section flows $q_{x,t}$. The section flows are generated by the water level in each pipe segment.

For the spatial discretization, we apply the backward Euler method with a spatial step size of δx, such that:

$$\frac{dh_{x,t}}{dt} = \theta_1(q_{x-\delta x,t} - q_{x,t} + d_{x,t}), \tag{6a}$$

$$q_{x,t} = \theta_2 f(h_{x,t}, \theta_3), \tag{6b}$$

where the nonlinear map $f : \mathbb{R}_+ \to \mathbb{R}_+$ is given by:

$$f : (h_{x,t}, \theta_3) \mapsto \frac{h_{x,t}^{5/3}}{(h_{x,t} + \theta_3)^{2/3}}, \quad \forall\, x \in (0, L), \tag{7}$$

and the physical constants along with the spatial step δx are gathered in parameters, where:

$$\theta_1 \triangleq \frac{1}{w\delta x}, \quad \theta_2 \triangleq \frac{\sqrt{S_b} w^{5/3}}{2^{2/3} n}, \quad \theta_3 \triangleq \frac{w}{2}. \tag{8}$$

Note that the parameters $\theta_1, \theta_2, \theta_3$ are positive, given that the physical constants and the spatial steps are positive.

In order to obtain a model with water levels as states, the $q_{x,t}$ flows in eq.(6a) are substituted with water levels from eq.(6b). For ease of notation the time index t is omitted.

Then, the reduced KW-based model is given by a set of N_x ODEs, each representing a section of length of the pipe:

$$\frac{dh_1}{dt} = \theta_1(q_0 + d_1) - \theta_1\theta_2 f(h_1, \theta_3), \tag{9a}$$

$$\vdots$$

$$\frac{dh_n}{dt} = \theta_1 d_n + \theta_1\theta_2\big(f(h_{n-1}, \theta_3) - f(h_n, \theta_3)\big), \tag{9b}$$

$$\vdots$$

$$\frac{dh_{N_x}}{dt} = \theta_1(d_{N_x} - q_{N_x}) + \theta_1\theta_2 f(h_{N_x-1}, \theta_3), \tag{9c}$$

where the flow q_0 at $x = 0$ is considered as a controlled input. As the section flows depend only on the water levels in the corresponding segments, the discharged flow is calculated directly from the level in the last segment:

$$q_{N_x} = \theta_2 f(h_{N_x}, \theta_3), \tag{10}$$

where h_{N_x} is the water level in the last pipe segment.

2.2 Diffusion wave approximation

Unlike the KW-based model, the DW approximation does not neglect the term $\partial h_{x,t}/\partial x$ in eq.(1b). Hence, the momentum equation of the SV-based PDEs becomes:

$$\frac{\partial h_{x,t}}{\partial x} = S_b - S_{f|x,t}. \tag{11}$$

In case of the DW approximation, the friction term is given by the open-channel Darcy-Weisbach formula [Schütze et al., 2002]:

$$S_{f|x,t} = k\frac{P_{x,t}}{8A_{x,t}^3 g}q_{x,t}^2, \tag{12}$$

where k is the Darcy-Weisbach friction factor and g is the gravitational acceleration.

To show the structure of the DW-based model, first we spatially discretize the SV-based PDEs in eq.(1). We use the backward Euler method for discretizing $\partial q_{x,t}/\partial x$ in eq.(1a) (as for the KW-based model) and the forward Euler method for $\partial h_{x,t}/\partial x$ in eq.(1b). Then, the SV-based PDEs are reduced to a system of finite-dimensional, first-order ODEs:

$$\delta x w\frac{dh_{x,t}}{dt} = q_{x-\delta x,t} - q_{x,t} + d_{x,t}, \tag{13a}$$

$$h_{x+\delta x,t} - h_{x,t} = z - r(q_{x,t}, h_{x,t}), \tag{13b}$$

where $z \triangleq \delta x S_b$ defines the elevation difference between the equal-sized,

non-overlapping segments δx and the pipe resistance $r_x(q_{x,t}, h_{x,t}) \triangleq \delta x S_{f|x,t}$ due to friction. Applying Assumption 1, the resistance function becomes:

$$r_x(q_{x,t}, h_{x,t}) = k \frac{\delta x}{8g} \frac{w + 2h_{x,t}}{w^3 h_{x,t}^3} q_{x,t}^2. \tag{14}$$

The flows $q_{x,t}$ between pipe segments are expressed similarly to the section flows in the KW-based model. The inverse of the resistance function $r(q_{x,t}, h_{x,t})$ is used to solve the momentum equation in eq.(13b), such that:

$$q_{x,t} = r_x^{-1}(h_{x,t} + z - h_{x+\delta x,t}). \tag{15}$$

As opposed to the KW-based model, the pipe flows $q_{x,t}$ are not only a function of the local water levels $h_{x,t}$, but also the water level one spatial step forward and the elevation between the neighboring sections, as depicted in Figure 2.

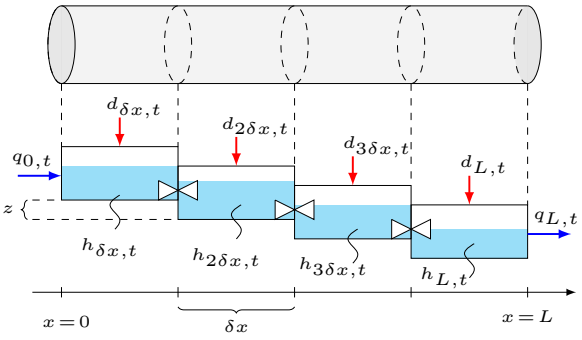


Figure 2: Discretized pipe, indicating level differences generating the flow.

Note that the relation between section flows $q_{x,t}$ and water levels $h_{x,t}$ is not one-to-one, as it was in the KW-based model. Furthermore, the section flows are generated by the elevation difference due to the pipe slope, as well as by the level difference between the interconnected segments.

The spatial discretization of the DW-based model yields:

$$\frac{dh_{x,t}}{dt} = \lambda_1(q_{x-\delta x,t} - q_{x,t} + d_{x,t}), \tag{16a}$$

$$q_{x,t} = \lambda_2\big((h_{x,t} + z - h_{x-\delta x,t}) g(h_{x,t}, \lambda_3)\big)^{1/2}, \tag{16b}$$

where the nonlinear map $g : \mathbb{R}_+ \to \mathbb{R}_+$ is given by:

$$g : (h_{x,t}, \lambda_3) \mapsto \frac{h_{x,t}^3}{h_{x,t} + \lambda_3}, \quad \forall\, x \in (0, L). \tag{17}$$

Here, z is an extra model parameter and the physical constants are collected in parameters, where:

$$\lambda_1 \triangleq \frac{1}{w\delta x}, \quad \lambda_2 \triangleq \left(\frac{4gw^3}{k\delta x}\right)^{1/2}, \quad \lambda_3 \triangleq \frac{w}{2}. \tag{18}$$

Similarly to the KW-based model, the parameters $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}_+$. Note that the parameters for the KW and DW case only differ in θ_2 and λ_2.

To obtain a model with water levels as states, the $q_{x,t}$ flows in eq.(16a) are substituted with water levels from eq.(16b). For ease of notation the time index t is omitted. Then, the reduced DW-based model is given by a set of N_x ODEs, each representing a section of length of the pipeline:

$$\frac{dh_1}{dt} = \lambda_1(q_0 + d_1) - \lambda_1\lambda_2\Big((h_1 - h_2 + z)g(h_1, \lambda_3)\Big)^{\frac{1}{2}}, \tag{19a}$$

$$\vdots$$

$$\frac{dh_n}{dt} = \lambda_1 d_n + \lambda_1\lambda_2\Big[\Big((h_n - h_n + z)g(h_{n-1}, \lambda_3)\Big)^{\frac{1}{2}} - \Big((h_n - h_{n+1} + z)g(h_n, \lambda_3)\Big)^{\frac{1}{2}}\Big] \tag{19b}$$

$$\vdots$$

$$\frac{dh_{N_x}}{dt} = \lambda_1(d_{N_x} - q_{N_x}) + \lambda_1\lambda_2\Big((h_{N_x-1} - h_{N_x} + z)g(h_{N_x-1}, \lambda_3)\Big)^{\frac{1}{2}}, \tag{19c}$$

where in a similar manner to the KW-based model, the boundary flows at the upstream and downstream end of the channel are given by q_0 and q_{N_x}, respectively. Unlike the KW-based model, the boundary flow downstream cannot be directly calculated from the water level in the last segment. The DW-based model inherently incorporates the internal connections between the connected elements at the N_x and N_{x+1} spatial steps, where N_{x+1} corresponds to the connected structure.

Remark B1. *Hydraulic structures define the level-flow relation at the boundary points of pipelines. These structures are typically wastewater basins, weirs, gates or the receiving water body, e.g., the sea.*

The mathematical description of hydraulic structures differs for free [Dey, 2002] and submerged flow [Litrico and Fromion, 2009], hence the model structure and parameters differ too. The outflow-level relation for the two different cases are given by the function:

$$q_{N_x} = G_{\boldsymbol{\mu}_f}(h_{N_x}) \qquad \text{for free flow} \tag{20a}$$

$$q_{N_x} = G_{\boldsymbol{\mu}_s}(h_{N_x}, h_{N_x+1}) \qquad \text{for submerged flow} \tag{20b}$$

where h_{N_x} is the water level in the last segment of the pipeline and h_{N_x+1} in the hydraulic structure. Moreover, $\boldsymbol{\mu}_f$ and $\boldsymbol{\mu}_s$ are vectors of structure parameters corresponding to free and submerged flows, respectively. Models of hydraulic structures corresponding to the most common elements in UDNs are reported in [Litrico and Fromion, 2009].

2.3 Discrete system model

In this study, discrete-time system dynamics are utilized for solving the system identification problems for both the KW- and DW-based models. The dynamics of the KW-based model described in eq.(9) and eq.(10) are given for one pipeline:

$$\hat{\boldsymbol{h}}(t_{k+1}) = \boldsymbol{F}_{\boldsymbol{\theta}}^{\mathrm{KW}}\big(\hat{\boldsymbol{h}}(t_k), q_0(t_k), \boldsymbol{d}(t_k)\big), \tag{21a}$$

$$q_{N_x}(t_k) = \theta_2 f(h_{N_x}(t_k), \theta_3), \tag{21b}$$

where the numerical integration from t_k to t_{k+1} is done by the fixed-step 4^{th} order Runge-Kutta method. Furthermore, $\hat{\boldsymbol{h}}(t_k) \in \mathbb{R}^{N_x}$ is the vector of states, representing the water levels in each of the N_x sections. The vector $\boldsymbol{d}(t_k)$ represents lateral inflow disturbances along the pipe line in each segment. The dynamics are given by $\boldsymbol{F}_{\boldsymbol{\theta}}^{KW} : \mathbb{R}^{N_x} \times \mathbb{R}_+ \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$. The outlet flows are given by eq.(21b).

The dynamics of the DW-based model described in eq.(19a) are given for one pipeline:

$$\hat{\boldsymbol{h}}(t_{k+1}) = \boldsymbol{F}_{\boldsymbol{\lambda},z}^{\mathrm{DW}}\big(\hat{\boldsymbol{h}}(t_k), q_0(t_k), q_{N_x}(t_k), \boldsymbol{d}(t_k)\big), \tag{22}$$

where the dynamics are given by $\boldsymbol{F}_{\boldsymbol{\lambda},z}^{DW} : \mathbb{R}^{N_x} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$. The discharged flow is given by eq.(20a).

3 System Identification

The system identification problem in both cases is given by a finite-dimensional constrained Nonlinear Programming (NLP) problem, where the boundary flows q_0, q_{N_x} and the disturbances $\boldsymbol{d}$ at t_i, $i = \{0, ..., N_t\}$ are known a priori.

Remark B2. *The pumped inlet flows are estimated by the polynomial expression of fixed-speed wastewater pumps [Kallesøe and Knudsen, 2016], and the outlet flows are estimated by mass conservation. The flow estimation algorithm suitable for the application and methodologies presented in this study is detailed in [Kallesøe and Knudsen, 2016], which the interested reader may refer to for more details.*

Water levels in each sections of the gravity pipe are also known a priori by means of sensor measurements, given by:

$$\boldsymbol{v} = \boldsymbol{C}\boldsymbol{h} + \nu, \tag{23}$$

where $\boldsymbol{C} \in \mathbb{R}^{N_0 \times N_x}$ is picking out all measured states $\boldsymbol{v}$ from the states $\boldsymbol{h}$. N_0 denotes the number of level sensors and $\nu \in NID(0, \sigma^2)$ is white Gaussian measurement noise.

The input vector for the KW-based identification at time instance t_i is given by $\boldsymbol{u}(t_i) \triangleq \big(q_0(t_i), \boldsymbol{d}^{\mathsf{T}}(t_i)\big)^{\mathsf{T}}$, and the corresponding output vector by

$\boldsymbol{y}(t_i) \triangleq \left(\boldsymbol{v}^\top(t_i), q_{N_x}(t_i)\right)^\top$. The model parameters are given by $\boldsymbol{\theta} \triangleq (\theta_1, \theta_2, \theta_3)^\top$. Then, the NLP problem for the KW-based model is given by:

$$\begin{pmatrix} \boldsymbol{\theta}^* \\ \hat{\boldsymbol{h}}(t_0)^* \end{pmatrix} = \underset{\boldsymbol{\theta}, \hat{\boldsymbol{h}}(t_0)}{\operatorname{argmin}} \sum_{i=0}^{N_t} \left(q_{N_x}(t_i) - \hat{q}_{N_x}(t_i)\right)^2 + \Omega \|\boldsymbol{v}(t_i) - \hat{\boldsymbol{v}}(t_i)\|^2 \tag{24a}$$

subject to dynamics in eq.(21) and to inequality constraints:

$$\mathbf{0} < \hat{\boldsymbol{h}}(t_i) \leq \overline{\boldsymbol{h}}, \tag{24b}$$

$$0 < \hat{q}_{N_x}(t_i) \leq \overline{q}_{N_x}, \tag{24c}$$

$$\mathbf{0} < \boldsymbol{\theta} \leq \overline{\boldsymbol{\theta}}, \tag{24d}$$

where eq.(24b), eq.(24c) and eq.(24d) impose bounds on states, outputs and parameters, respectively. Note that the upper and lower bounds in each constraint are based on meaningful physical values, e.g., the slope and width of the channel must be within meaningful physical ranges and the water levels and flows cannot be negative. Besides, Ω is a weighing constant in eq.(24a), scaling the water levels to flows.

Unlike the KW-based model identification, the inputs in the DW-based model at time instance t_i are defined by $\boldsymbol{u}(t_i) \triangleq \left(q_0(t_i), q_{N_x}(t_i), \boldsymbol{d}^\top(t_i)\right)^\top$, while the outputs are $\boldsymbol{y}(t_i) \triangleq \left(\boldsymbol{v}^\top(t_i)\right)^\top$. Furthermore, we define the parameter vector by $\boldsymbol{\lambda} \triangleq (\lambda_1, \lambda_2, \lambda_3)^\top$. Note that instead of using the outflow model defined at the downstream boundary of pipelines, described in eq.(20a), the estimated discharged flow q_{N_x} is used as an input. This is done in order to avoid introducing extra μ parameters for hydraulic structures, hence restrict the parameter space. Instead, the flow q_{N_x} is estimated as stated in Remark 2. Then, the NLP problem for the DW-based model is given by:

$$\begin{pmatrix} \boldsymbol{\lambda}^* \\ z^* \\ \hat{\boldsymbol{h}}(t_0)^* \end{pmatrix} = \underset{\boldsymbol{\lambda}, z, \hat{\boldsymbol{h}}(t_0)}{\operatorname{argmin}} \sum_{i=0}^{N_t} \|\boldsymbol{v}(t_i) - \hat{\boldsymbol{v}}(t_i)\|^2 \tag{25a}$$

subject to dynamics in eq.(22) and to inequality constraints:

$$\mathbf{0} < \hat{\boldsymbol{h}}(t_i) \leq \overline{\boldsymbol{h}}, \tag{25b}$$

$$\mathbf{0} < \boldsymbol{\lambda} \leq \overline{\boldsymbol{\lambda}}, \tag{25c}$$

$$0 < z \leq \overline{z}, \tag{25d}$$

where similarly to the KW case, eq.(25b) imposes bounds on the states, and eq.(25c) and eq.(25d) on the parameters, respectively.

The NLP problems in both cases are solved by using a Gauss-Newton gradient-based method, detailed in [Wills and Ninness, 2008]. Furthermore, the auxiliary variable N_x, i.e., the number of sections into which each pipeline is discretized into are fixed in both NLP problems. Grid size selection can be evaluated on Monte Carlo simulations for varying step sizes [Balla et al., 2020a].

4 Case study

The experimental setup for validating the proposed methodologies is shown in Fig. Figure 3.

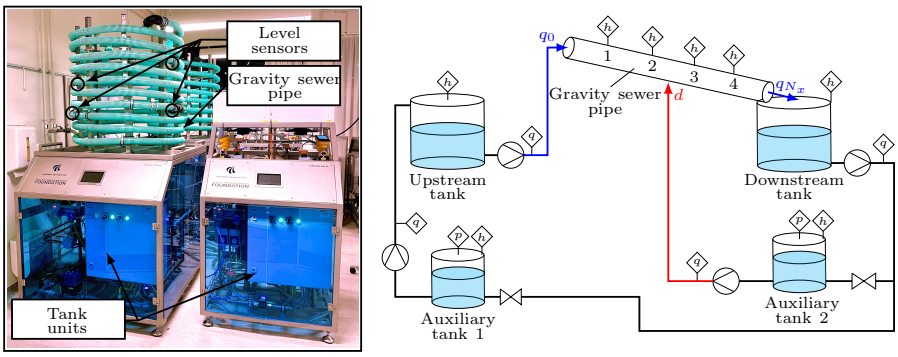


Figure 3: Sewer modules of the Smart Water Laboratory setup at Aalborg university (left). Schematics of the experimental setup (right). The sensor placements are indicated for each individual module with pressure sensor (p), level sensor (h) and flow sensor (q).

The setup represents a scaled version of a typical gravity-driven sewer pipeline most commonly found in real-life infrastructures. The proposed setup consists of an open-channel pipeline, along which $N_0 = 4$ level sensors are installed. The inlet flow q_0 is pumped from an upstream tank, while the discharged flow q_{N_x} is calculated from the mass balances of a downstream tank. Auxiliary tanks are utilized as flow sources to pump the disturbance in the middle point of the open-channel sewer pipe. Note, that the disturbance enters the pipeline between the second and third level sensor. Hence, the designed flow event allows to create backflow in the middle of the channel, which is captured by the second level sensor and then propagates to the downstream tank. The level and flow measurement data are obtained and locally managed at each laboratory unit with a Codesys soft-PLC [3S-Smart Software Solutions GmbH], and all local modules are interfaced in real-time. The data used for system identification is gathered over a three-hours long experiment and sampled at $T = 2[s]$. Besides, both the KW and DW models utilized in the system identification are built up of a pipeline discretized into $N_x = 8$ sections.

5 Experimental results

Following the KW and DW model methodologies discussed in Section 2, the two problems have been verified on data extracted from the same experiment. The inlet flow q_0 and lateral inflow d is shown in Figure 5 (a-b), where the inlet pumps turn on and off with a fixed speed, while there is a consecutive inflow event in the middle of the experiment.

In case of the KW-based model, the output vector $\boldsymbol{y}$ includes the discharged flow q_{N_x}, hence the flow prediction is tested against data. As shown in Figure 4, the one-step prediction of the KW-based model produces accurate flows compared to the estimated discharge flow data.

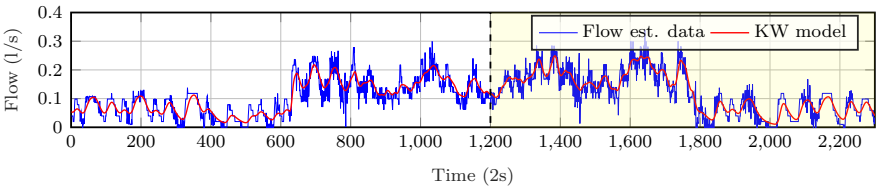


Figure 4: Identification and validation of discharged flow in the KW model.

The validation results show that the model is predicting precisely the flow-dependant process delays both in the presence and without the lateral inflows. Note that the discharged flow increases, when lateral inflow is present along the pipe. This is due to mass conservation, as extra volume is propagating down the channel.

The comparison of the one-step state prediction for the KW- and DW-based models are shown in Figure 5. Note that the first sensor measurement (h_1) is not affected by the backflow, however, as the disturbance is applied to the system, the sensor at the second position (h_2) captures the water volumes accumulating inside the pipes. This is shown in (e-f), where the KW-based model assumes the downstream propagation of the disturbance without affecting the upstream state, i.e. h_2. In contrast to that, the DW-based model finds the correct z parameters, and due to the level difference in the segments, accounts for the backflow. As shown in Figure 5 (g-h-i-j), both models are equally good at state prediction after the location where the disturbance enters the channel.

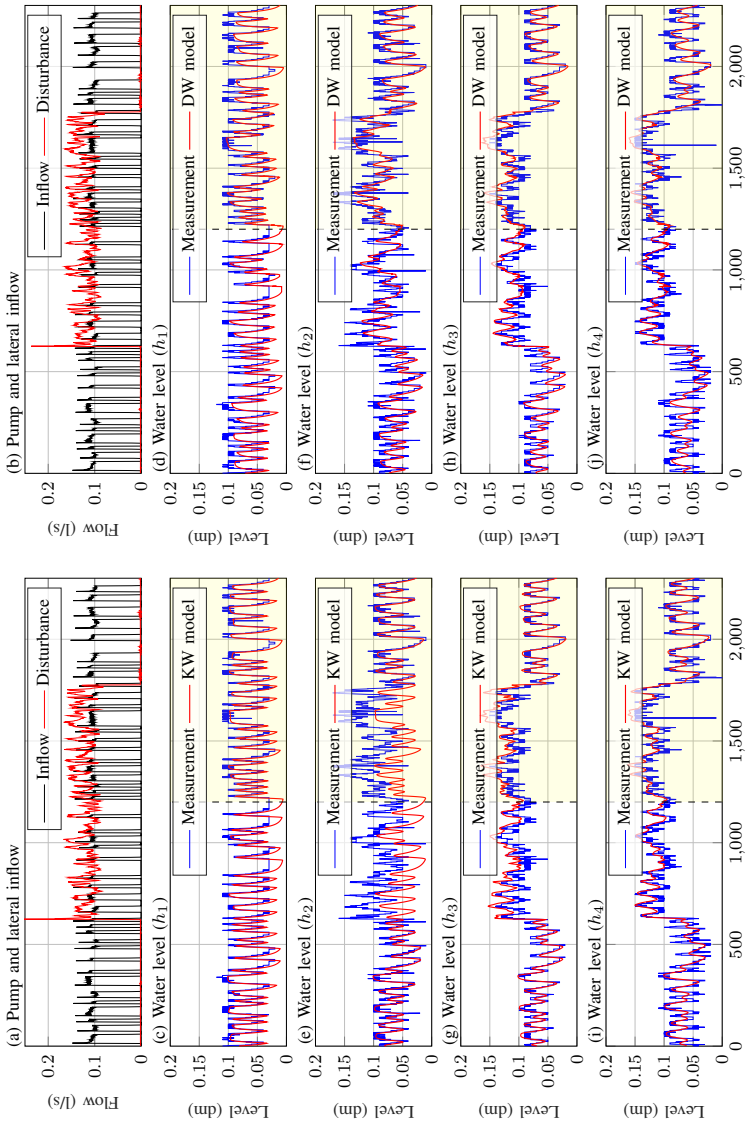


Figure 5: System identification and validation results for the KW- (left) and the DW-based (right) modelling approaches.

6 Conclusion

In this article, a comparison of two model structures reduced from the Saint-Venant partial differential equations has been presented to model open-channel, gravity-driven flow processes with the backwater phenomena. To this end, a grey-box approach has been proposed using level sensors distributed along the pipeline and utilizing the spatially discretized kinematic and diffusion wave approximations of the full dynamic Saint-Venant equations. The constrained nonlinear system identification problem has been solved for both approaches, where data has been extracted from a scaled laboratory setup built for control of water infrastructures. The experimental results corroborate the feasibility of both approaches and point out the capabilities of the diffusion wave approach in capturing backflow inside the pipes.

Acknowledgement

The authors would like to thank the Poul Due Jensen Foundation for providing the Smart Water Lab at Aalborg University. The project was supported by Innovation Fund Denmark and Grundfos Holding A/S as part of a Danish Industrial Ph.D. Project [Application number: *9065-00018A*].

References

3S-Smart Software Solutions GmbH. "Codesys". URL: `https://www.codesys.com`.

K. M. Balla, C. S. Kallesøe, C. Schou, and J. D. Bendtsen. "Nonlinear grey-box identification with inflow decoupling in gravity sewers". *IFAC-PapersOnLine*, 53 (2):1065–1070, 2020a. ISSN 24058963. doi:10.1016/j.ifacol.2020.12.1295.

K. M. Balla, C. Schou, J. Dimon Bendtsen, and C. S. Kallesøe. "Multi-scenario model predictive control of combined sewer overflows in urban drainage networks". In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1042–1047, Montréal, 2020b. IEEE. ISBN 9781728171401. doi:10.1109/CCTA41146.2020.9206362.

S. Dey. "Free overfall in open channels: State-of-the-art review". *Flow Measurement and Instrumentation*, 13(5-6):247–264, 2002. ISSN 09555986. doi:10.1016/S0955-5986(02)00055-9.

G. Evans, J. Blackledge, and P. Yardley. *Numerical methods for partial differential equations*. Springer Science & Business Media, 2012.

M. S. Gelormino and N. L. Ricker. "Model-predictive control of a combined sewer system". *International Journal of Control*, 59(3):793–816, 1994. ISSN 13665820. doi:10.1080/00207179408923105.

C. S. Kallesøe and T. Knudsen. "Self calibrating flow estimation in waste water pumping stations". In *2016 European Control Conference, ECC 2016*, pages 55–60, 2016. ISBN 9781509025916. doi:10.1109/ECC.2016.7810263.

X. Litrico and V. Fromion. "Boundary control of linearized Saint-Venant equations oscillating modes". *Automatica*, 42:967–972, 2006. ISSN 00051098. doi:10.1016/j.automatica.2006.02.002.

X. Litrico and V. Fromion. *Modeling and control of hydrosystems*. Springer, 2009. ISBN 9781848826236. doi:10.1007/978-1-84882-624-3.

C. Ocampo-Martinez. *Model Predictive Control of Wastewater Systems*. Springer, Barcelona, 1st edition, 2010. ISBN 978-1-84996-352-7. doi:10.1007/978-1-84996-353-4.

J. A. Roberson and C. T. Crowe. *Engineering Fluid Mechanics*. Houghton Mifflin Company, Boston, 5th edition, 1993. ISBN 0-395-66161-7. doi:10.1007/978-1-4020-6742-6.

M. R. Schütze, D. Butler, and M. B. Beck. *Modelling, Simulation and Control of Urban Wastewater Systems*. Springer, 2002. doi:10.1007/978-1-4471-0157-4.

V. P. Singh. "Kinematic wave modelling in water resources: A historical perspective". *Hydrological Processes*, 15:671–706, 2001. ISSN 08856087. doi:10.1002/hyp.99.

S. C. Troutman, N. Schambach, N. G. Love, and B. Kerkez. "An automated toolchain for the data-driven and dynamical modeling of combined sewer systems". *Water Research*, 126:88–100, 2017. ISSN 18792448. doi:10.1016/j.watres.2017.08.065.

A. Wills and B. Ninness. "On gradient-based search for multivariable system estimates". *IEEE Transactions on Automatic Control*, 53(1):298–306, 2008. ISSN 00189286. doi:10.1109/TAC.2007.914953.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow". *Advances in Water Resources*, 34:282–290, 2011. ISSN 03091708. doi:10.1016/j.advwatres.2010.11.009.

M. Xu, R. R. Negenborn, P. J. van Overloop, and N. C. van de Giesen. "De Saint-Venant equations-based model assessment in model predictive control of open channel flow". *Advances in Water Resources*, 49:37–45, 2012. ISSN 03091708. doi:10.1016/j.advwatres.2012.07.004.

Y. Zou, L. Cen, D. Li, and X. He. "Simplified state-space model and validation of irrigation canal systems". In *Chinese Control Conference, CCC*, pages 2002–2007, Hangzhou, 2015. ISBN 9789881563897. doi:10.1109/ChiCC.2015.7259938.

Paper C

A Nonlinear Predictive Control Approach for Urban Drainage Networks Using Data-Driven Models and Moving Horizon Estimation

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

[a]Grundfos Holding A/S, Poul Due Jensens Vej 7, DK-8850, Bjerringbro, Denmark
[b]Section for Automation and Control, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark
[c]Automatic Control Department, Institut de Robotica i Informatica Industrial (CSIC-UPC), Universitat Politècnica de Catalunya, Llorens i Artigas 4-6, Planta 2, 08028, Barcelona, Spain

Abstract—*Real-time control of urban drainage networks is a complex task where transport flows are non-pressurized and therefore impose flow-dependent time delays in the system. Unfortunately, the installation of flow sensors is economically out of reach at most utilities, although knowing volumes and flows are essential to optimize system operation. In this article, we formulate joint parameter and state estimation based on level sensors deployed inside manholes and basins in the network. We describe the flow dynamics on the main pipelines by the level variations inside manholes, characterized by a system of coupled partial differential equations. These dynamics are approximated with kinematic waves where the network model is established with the water levels being the system states. Moving horizon estimation is developed where the states and parameters are obtained via the levels and estimated flow data, utilizing the topological layout of the network. The obtained model complexity is kept within practically achievable limits, suitable for nonlinear predictive control. The effectiveness of the control and estimation method is demonstrated on a high-fidelity model of a drainage network, acting as virtual reality. We use real rain and wastewater flow data and test the controller against the uncertainty in the disturbance forecasts.*

Keywords—*Receding horizon control; Transport delay; Partial differential equation; Urban drainage network*

Published in the *IEEE Transactions on Control Systems Technology.*

DOI: https://doi.org/10.1109/TCST.2021.3137712

Contents

1 Introduction

Open - channel hydraulic systems are large-scale networks where water is transported with a free surface in pipes or conduits [Schütze et al., 2002]. In this work, we focus on Urban Drainage Networks (UDNs), more specifically, on systems where rain and wastewater are combined, and pumped in open channels. Pumped UDNs are typical in areas where the gravitation of water is limited due to the flatness of the landscape [García et al., 2015]. Moreover, combined sewers carry both domestic and stormwater towards treatment plants, where the sewage is treated before being released to the environment [Butler and Davies, 2006]. Combined sewers are present in many large cities and they are often overloaded due to the under-dimensioned capacity of the infrastructure induced by fast urbanization and the growing number of end-users [Lund et al., 2018]. Besides, UDNs are increasingly being pushed to their limits due to changing weather conditions, resulting in more frequent Combined Sewer Overflows (CSOs) [Lund et al., 2018]. The changing conditions challenge flow prediction and raise the question of how to handle the increased load on these systems.

In the last few decades, several Real-Time Control (RTC) techniques have been developed for UDN applications. These techniques typically exploit the available sensor measurements, rain forecasts, and the available physical description of the network. Many of the applied methods for evaluating the network capacities and solving optimization problems are typically predictive model-based control techniques [Mays, 2001; Ocampo-Martinez et al., 2013; Ocampo-Martinez, 2010]. However, transport flows in open-channel hydraulics are governed by a set of Partial Differential Equations (PDEs), too complex to identify with data and often infeasible to adapt to RTC applications in large-scale problems. Several methods in the literature typically propose the use of reduced PDE-based models in Model Predictive Control (MPC). These methods rely on the physical properties available (e.g., pipe dimensions, friction, and slope parameters) for model calibration with HiFi (High Fidelity) model simulators [Xu et al., 2011], [Xu et al., 2012]. Simulating gravity-driven flow with full PDE-based models in large-scale UDNs requires either a HiFi simulation environment or the placement of several flow and level sensors along the pipelines, meaning prohibitively expensive installation and maintenance costs.

PDEs linearized around an operating point have been used in UDN applications, where transfer functions [Dalmas et al., 2017], [Litrico and Fromion, 2006] and state-space models [Zou et al., 2015] have been developed. Due to the complexity of PDE-based control, conceptual models are also used in the state-of-art, for instance, [Balla et al., 2020b] used algebraic models with a single delay parameter, while [Ocampo-Martinez, 2010], [Gelormino and Ricker, 1994], [Mollerup et al., 2016], [Ocampo-Martinez et al., 2013], [Joseph-Duran et al., 2015] used a dynamic control model where the available capacity of pipes and tanks have been collectively modelled as virtual buffers. However, linearized and conceptual internal models do not allow flow-dependent time

delays, conceptualize the physically measurable levels and flows, furthermore restrict the flow deviation from steady-state solutions. Data-driven modeling has been reported in [Balla et al., 2020a] and in [Troutman et al., 2017], where grey-box and black-box identification have been used, respectively.

In this article, we propose a PDE-based modeling framework, where the system of PDEs is approximated to obtain a simple representation of the network, preserving the main system dynamics for control. We report on the modularity of the framework by arguing that using the network topology and water level sensors, a model suitable for control is obtained. Opposed to the current state-of-art, the proposed modeling framework captures the inflows to the UDN through water level measurements. In this way, we disregard the use of HiFi simulation models for model calibration.

Moreover, a new Nonlinear Model Predictive Control (NMPC) approach is proposed, based on a data-driven model, reduced from PDEs. In our approach, a Moving Horizon Estimation (MHE) method is used for constrained parameter and state estimation governing the PDEs, which we spatially discretize to Ordinary Differential Equations (ODEs). Time periodicity conditions are imposed on disturbance inflows, generated by household activity, to incorporate additional structure in the model dynamics used for predictions in the NMPC. The proposed control architecture is shown in Figure 1. The Moving Horizon Parameter Estimation (MHPE) along with the Moving Horizon State Estimation (MHSE) is carried out using easy-accessible level sensors distributed and placed inside manholes along the main sewer lines. Besides, we utilize flow estimation techniques which allow us to use pumped inlet and gravitated discharge flows, further detailed in [Kallesøe and Knudsen, 2016]. By using MHPE with NMPC, the system can re-identify the slowly changing pipe dynamics due to accumulated sludge in the bottom of sewer pipes. Besides, the NMPC can adapt to varying flow conditions imposed by the changing rain infiltration due to seasonality. The MHPE and MHSE problems, similarly to [Joseph-Duran et al., 2015], [Joseph-Duran et al., 2014], are both formulated as nonlinear least-squares problems, subject to state and parameter constraints, further detailed in Section 4. As shown in Figure 1, the NMPC is utilized as a global controller, solving a multi-criteria optimization problem and thereby providing references to the pumps at the local pumping stations. The proposed control and estimation methods are demonstrated on a HiFi network, simulated in the Mike Urban (MU)[1] simulation software where we use the catchment dynamics and the MU runoff engine for generating rain-runoff appearing as the load on the network. Finding the rain-runoff based on rain intensity forecasts by radars and numerical weather predictions is an active field of research, which has been extensively studied in [Ma et al., 2018], [Chang et al., 2001] and its effect of uncertainty on UDNs in [Löwe et al., 2016], [Löwe et al., 2014]. Moreover, several works in the literature report on how to handle rain forecast uncertainty, e.g.,

[1]MIKE Urban is a standard hydraulic simulation and planning tool, used as a planning tool by many operators at water utilities. The MU simulation environment solves the full dynamic PDEs for open-channel flow [DHI, 2017].

in UDNs in [Balla et al., 2020b] and in river applications in [Tian et al., 2017]. In this work, historical events of rain and wastewater are utilized in terms of real measurements, representing the imperfect weather forecasts.

The proposed data-driven method using the reduced network model and MHE has two clear benefits:

- First, it is a data-driven method that does not require heavy computation and difficult calibration procedures opposed to HiFi models, used at many utilities.
- Second, it is only required to collect data from periods under normal operational behavior, opposed to conventional data-driven methods where historical data is required for all the abnormal system behaviors.

The rest of the article is organized as follows. In Section 2, a preliminary overview of the operation of UDNs is presented. Section 3 first presents the PDE-based model for open-channel flow, followed by the reduced, data-driven system of the nonlinear ODE model obtained via spatial discretization. Then, the model of storage elements and the time-periodicity assumption on the disturbance signals are presented with the description of the system as a directed tree graph. In Section 4, the MHPE and MHSE techniques are detailed, whereupon Section 5 introduces the NMPC design and establishes the main control objectives. In Section 6, the results using data from a real-world network are presented. This is followed by Section 7 and Section 8, where a discussion, conclusions and future research directions are provided.

1.1 Nomenclature

Let $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m\times n}$, denote the field of real numbers, the set of real column vectors of length n and the set of m by n real matrices, respectively. Throughout the paper, all quantities mentioned are real. We use boldface letters for sets, such as $\boldsymbol{s} = \{s_1, ..., s_n\}$, as well as for vectors $\boldsymbol{x} = [x_1, ..., x_n]^T \in \mathbb{R}^n$. The superscript $^\top$ denotes transposition, and the operators $<, \leq, =, >, \geq$ denote element-wise relations of vectors. Moreover, for a vector $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$ denotes the Euclidean norm.

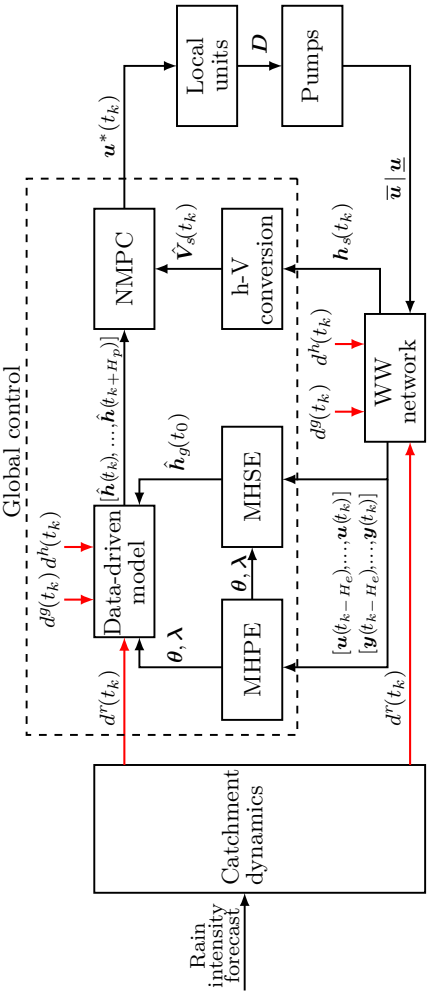


Figure 1: The proposed closed-loop control architecture, where rain intensities are known by means of weather forecasts and the transformation between the intensities and the runoff flow appearing in the sewers is characterized by the catchment dynamics. Red signals denoted with d represent rain, household and groundwater inflow disturbances, h and V are the system states representing water level and water volume, respectively. Moreover, u denotes the input of the aggregated flows which are delivered by locally-controlled pumps. The pipe network (plant) is represented by the WW (Waste Water) network block.

2 Drainage System overview

UDNs contain several elements, including gravitation pipes, manholes, pits, and in some cases, retention tanks. The most widely used actuators in pressurized sewer networks are pumps, typically installed inside wastewater pits where the sewage and rain are collected [Schütze et al., 2002]. These units often consist of one or several pumps in parallel, controlling the transport of the sewage from pit to pit. First, the water is pumped through a rising main, whereupon it gravitates through sewer pipes towards a downstream station, shown in Figure 2.

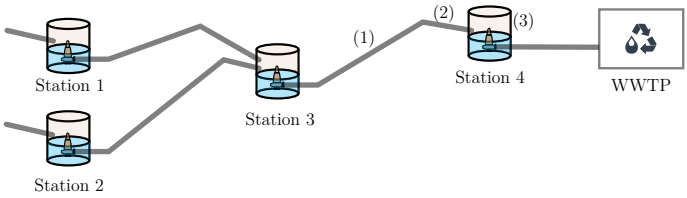


Figure 2: Tree topology of a pumped sewer network, where (1) illustrates rising mains, (2) gravity sewer pipes and (3) pumping stations [Butler and Davies, 2006].

UDNs typically have a tree structure, where the Waste Water Treatment Plant (WWTP) represents the root of the network.

3 System model

The modelling based on physics is introduced to show how the reduced model is obtained considering simple mass conservation rules and assumptions on the geometry of hydraulic structures. We aim to obtain a model structure with a low number of lumped parameters, where the system states are expressed by water levels. Besides, we show that the proposed internal model structure allows us to make assumptions on the initial parameters and their upper and lower bounds.

3.1 Physical transport model

Flow propagation in UDNs can be accurately computed by the full Saint-Venant (SV) equations, which are non-linear hyperbolic PDEs describing the mass and momentum of fluid:

$$\frac{\partial A_{x,t}}{\partial t} + \frac{\partial q_{x,t}}{\partial x} = \tilde{d}_{x,t}, \tag{1a}$$

$$\frac{\partial q_{x,t}}{\partial t} + \frac{\partial}{\partial x}\left(\frac{q_{x,t}^2}{A_{x,t}}\right) + gA_{x,t}\Big(\overbrace{\frac{\partial h_{x,t}}{\partial x} + \underbrace{S_f - S_b}_{\text{Kinematic wave}}}^{\text{Diffusion wave}}\Big) = 0, \tag{1b}$$

where $q_{x,t}$ is the flow in the pipe and $\tilde{d}_{x,t} = d_{x,t}/dx$ represents lateral inflows per unit length, where $d_{x,t}$ is the lateral inflow hereinafter referred to as disturbance. $A_{x,t}$ is the wetted pipe area, $h_{x,t}$ represents the water level, furthermore $q_{x,t}, d_{x,t}, A_{x,t}$ and $h_{x,t}$ are functions from $(0, L) \times \mathbb{R}_+ \to \mathbb{R}_+$, where L is the total length of the gravity pipe. The gravitational acceleration is denoted with g, moreover the slope term $S_b \in \mathbb{R}_+$ and friction term $S_f \in \mathbb{R}_+$ are assumed to be independent of x and t, i.e. all pipe segments along the gravity pipe are modelled with assuming identical physical attributes.

The dynamics of each transport pipe in eq.(1a) and eq.(1b) are coupled through boundary conditions, hence the problem can become computationally demanding to solve in the case of complex networks [Xu et al., 2012]. Assumptions on the flow characteristics can lead to loss of dynamics, however, can lead to significant simplifications to the model structure. In this work, we utilize the Kinematic Wave approximation of the SV equations, thereby removing the left-hand-side terms of eq.(1b). In this way, we omit the phenomena of wave attenuation, flow acceleration, and the phenomena of backwater effect[2]. These simplifications inherently mean that the considered flow characteristics are uniform and hence quasi-steady flow is assumed at all $x \in (0, L)$. The momentum equation in eq.(1b) only considers two terms, i.e.,

$$S_b = S_f(q_{x,t}, h_{x,t}), \tag{2}$$

where the friction term S_f is obtained from the Manning equation, which is an empirical formula for energy balance between gravity and friction, expressed by the level h and flow q variables [Schütze et al., 2002] as

$$S_b = \frac{n^2 q_{x,t}^2}{A_{x,t}^2 R_{x,t}^{4/3}}, \tag{3}$$

where $R = \frac{A}{P}$ is the hydraulic radius, $P \in \mathbb{R}_+$ is the wetted perimeter and $n \in \mathbb{R}_+$ is the Manning coefficient. Note that by knowing a map $f : A_{x,t} \mapsto h_{x,t}$, an expression between $q_{x,t}$ flow and $h_{x,t}$ level is constructed.

Assumption C1. *We assume a linear map f between the wetted-area $A_{x,t}$ and water level $h_{x,t}$. It is assumed that semi-filled circular sewers are reasonably well-approximated by rectangular pipe shapes, i.e.,*

$$R_{x,t} = \frac{A_{x,t}}{P_{x,t}} \approx \frac{w h_{x,t}}{2h_{x,t} + w}, \tag{4}$$

where $h_{x,t}$, $x \in (0, L)$ is the water level and the cross section is parametrized by the w channel width shown in Figure 3.

[2]Backwater occurs in sewers when the receiving water body becomes overloaded and therefore water volumes are accumulating at downstream of the connected hydraulic structure [Munier et al., 2008].

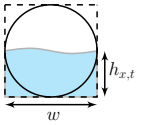


Figure 3: Semi-filled circular pipe approximated with rectangular geometry.

The independent variables remaining in the simplified SV PDEs in eq.(1a) and eq.(1b) are reduced to $q_{x,t}$ flow and $h_{x,t}$ level distributions on the domain $(0, L) \times \mathbb{R}_+$, given by:

$$w\frac{\partial h_{x,t}}{\partial t} + \frac{\partial q_{x,t}}{\partial x} = \frac{d_{x,t}}{\delta x}, \tag{5a}$$

$$q_{x,t} = \frac{\sqrt{S_b}}{n}\frac{(wh_{x,t})^{5/3}}{(2h_{x,t} + w)^{2/3}}, \tag{5b}$$

which is an approximation of the full dynamic SV-PDEs. Note that Assumption 1 on the pipe geometry means a linear scaling from levels $h_{x,t}$ to flows $q_{x,t}$, which leads to inaccuracy for circular pipe profiles. However, the assumption on the linear geometry profile keeps the model complexity low.

3.2 Reduced, data-driven transport model

In order to formulate the transport dynamics in a form more amenable to system identification, the spatial discretization of the approximated SV-PDEs in eq.(5a) and eq.(5b) is considered. The gravity pipes are partitioned into N_x non-overlapping δx segments of length, while the signals $h_{x,t}$, $q_{x,t}$ and $d_{x,t}$ are approximated as piece-wise constant functions of the spatial coordinate x, as shown in Figure 4.

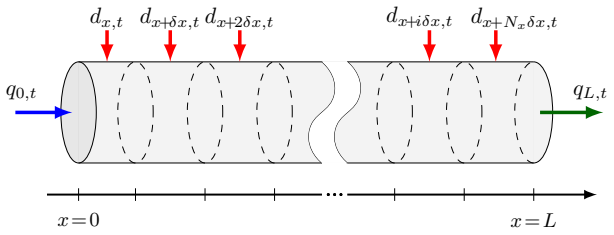


Figure 4: Gravity pipe divided into N_x, equal-sized, non-overlapping segments.

In Figure 4, $q_{0,t}$ and $q_{L,t}$ denote the flows corresponding to the upstream and downstream boundaries, respectively. Furthermore, $d_{x+i\delta x,t}$ represents the lateral inflows (disturbances) entering the i^{th} pipe section, where $i \in \{1, 2, ..., N_x\}$.

Remark C1. *It is not necessary to partition gravity pipes into equal-sized δx sections. The length of the spatial step δx can be defined by the placement of*

manholes along the sewer pipes, among which some may be equipped with level sensors.

Remark C2. *Close to the downstream end of gravity pipes ($x = L$), the discharge conditions of $q_{L,t}$ are influenced by the receiving hydraulic structure and the corresponding water levels [Roberson and Crowe, 1993], [Dey, 2002]. This relation imposes dynamics governing the water level $h_{L,t}$ in the last section. In this study, the effect of these types of dynamics are excluded, hence the positioning of water level sensors close to $x = L$ are chosen such that*

$$L - x \geq s_{min}, \tag{6}$$

where $s_{min} \in \mathbb{R}_+$ denotes the minimal distance from the end of the channel where level sensors at position x should be placed. The criteria of choosing s_{min} for the free fall condition of fluids, based on the diameter of open-channel pipes is detailed in [Roberson and Crowe, 1993, pp.698-699].

The spatial discretization of eq.(5a) and eq.(5b) is done by the backward Euler method. The left boundary (upstream) is defined at $x = 0$ and the right boundary (downstream) at $x = L$. Then, the SV-PDEs are reduced to the following system of finite dimensional ODEs:

$$\frac{dh_{x,t}}{dt} = \theta_1(q_{x-\delta x,t} - q_{x,t} + d_{x,t}), \quad \forall\ x \in (0, L), \tag{7a}$$

$$q_{x,t} = \theta_2 \frac{h_{x,t}^{5/3}}{(h_{x,t} + \theta_3)^{2/3}}, \quad \forall\ x \in (0, L), \tag{7b}$$

where the physical constants and the spatial time step are lumped into the parameters

$$\theta_1 \triangleq \frac{1}{w\delta x}, \quad \theta_2 \triangleq \frac{\sqrt{S_b} w^{\frac{5}{3}}}{2^{\frac{2}{3}} n}, \quad \theta_3 \triangleq \frac{w}{2}, \tag{8}$$

where $\theta_1, \theta_2, \theta_3 \in \mathbb{R}_+$. Note that θ_3 is directly related to the width parameter w and θ_1 would change along the pipe in case of non-equal spatial steps δx. For the sake of simplicity, the model is presented with fixed δx spatial steps.

Remark C3. *Due to the spatial discretization, numerical distortion is introduced in the traveling wave [Xu et al., 2012], which compensates for the flow attenuation phenomena in gravity pipes. This artificial attenuation vanishes as $\delta x \to 0$.*

In order to obtain the state equation with water levels as states, the section flow distribution $q_{x,t}$ in eq.(7a) is substituted with water levels from eq.(7b), which yields

$$\frac{dh_{x,t}}{dt} = \theta_1\theta_2\left(\frac{h_{x-\delta x,t}^{5/3}}{(h_{x-\delta x,t} + \theta_3)^{2/3}} - \frac{h_{x,t}^{5/3}}{(h_{x,t} + \theta_3)^{2/3}}\right) + \theta_1 d_{x,t}, \tag{9}$$

where, opposed to previous work in [Balla et al., 2020a], the flow balance in the SV equations is reformulated with physically measurable water levels. For ease of notation, let us define a non-linear map $g : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$g : (h_{x,t}, \theta_3) \mapsto \frac{h_{x,t}^{5/3}}{(h_{x,t} + \theta_3)^{2/3}}, \quad \forall\, x \in (0, L). \tag{10}$$

Then, the transport flow model is completed by introducing the boundary conditions into the N_x coupled ODEs, i.e.,

$$\begin{aligned}
\frac{dh_{0,t}}{dt} &= \theta_1(q_{0,t} + d_{0,t}) - \theta_1\theta_2 g(h_{\delta x,t}, \theta_3), \\
&\vdots \\
\frac{dh_{x,t}}{dt} &= \theta_1\theta_2\big(g(h_{x-\delta x,t}, \theta_3) - g(h_{x,t}, \theta_3)\big) + \theta_1 d_{x,t}, \\
&\vdots \\
\frac{dh_{L,t}}{dt} &= \theta_1\theta_2 g(h_{L-\delta x,t}, \theta_3) + \theta_1(d_{L,t} - q_{L,t}),
\end{aligned} \tag{11}$$

where $d_{x,t}$ is the unknown disturbances in form of lateral inflows. Besides, the upstream boundary flow $q_{0,t}$ is subject to control and hereinafter denoted as u. The downstream boundary flow $q_{L,t}$ is the discharged output, which we consider as the controlled model output, hereinafter denoted as

$$y = \theta_2 g(h_{L,t}, \theta_3), \tag{12}$$

where $h_{L,t}$ represents the water level at the downstream boundary $x = L$. Note that eq.(12) is the parametric form of eq.(3), relating the level to flow.

3.3 Disturbance model

The proposed model described in eq.(11) aggregates the unknown, scaled disturbances in $d_{x,t}$. The disturbances are typically composed of several different types of inflows:

$$d_{x,t} \triangleq d^r_{x,t} + d^h_{x,t} + d^g_{x,t}, \quad \forall\, x \in (0, L), \tag{13}$$

where $d^r_{x,t}$ denotes rain runoff, $d^h_{x,t}$ is the household flow due to human activity and $d^g_{x,t}$ stands for groundwater.

Assumption C2. *The disturbance flow generated by households has an inherent periodicity, such that $d^h_{x,t} = d^h_{x,t+T}$, where T typically corresponds to one day. Moreover, disturbances generated by groundwater infiltration fulfill the constraint $\sum_{i=1}^{N_x} d^g_{i,t} = N_x d^g_{j,t}$, $\forall\, j \in \{1, 2, ..., N_x\}$, i.e. uniformly distributed along the whole length of gravity sewer pipes.*

Remark C4. *Seasonality with different time periodicity have been considered (e.g. weeks, months) in [Livera et al., 2011], where methodologies such as Fourier models have been used to decompose seasonal components in a broad range of applications.*

Besides, the rain runoff $d^r_{x,t}$ is generated by the dynamics of catchments where the intensity of rain precipitation is typically provided by weather forecasts. Several works have been done on relating rain radar forecasts to actual runoff flow in UDNs, e.g., [Löwe et al., 2016], [Löwe et al., 2014], [Ma et al., 2018], [Chang et al., 2001].

For modelling the periodic household flows $d^h_{x,t}$ and the constant groundwater $d^g_{x,t}$ flows, Fourier series are utilized. For simplicity, let us assume that $d^r_{x,t} = 0,\ \forall\ x \in (0, L)$, i.e., assuming a dry-weather period. Then, the scaled disturbances in eq.(11) are defined as:

$$\begin{aligned} \tilde{d}_{x,t} &\triangleq \tilde{d}^g_{x,t} + \tilde{d}^h_{x,t} \\ &\triangleq \lambda_0 + \sum_{j=1}^{k} \big(\lambda_{1j} \cos(j\omega t) + \lambda_{2j} \sin(j\omega t)\big), \end{aligned} \tag{14}$$

where the set of disturbance parameters is $\boldsymbol{\lambda} \triangleq \{\lambda_0, \lambda_{11}, \lambda_{21}, ..., \lambda_{1k}, \lambda_{2k}\} \in \mathbb{R}^{2k+1}$. The angular frequency ω corresponds to a period of one day and $k \geq 2$ is the number of frequency terms in the truncated Fourier series. The transport model in eq.(11) and eq.(12), in combination with the disturbance model in eq.(14) are used to find parameters $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$.

3.4 Storage model

Stored volume within the network is represented through wastewater pits, among which some are specifically constructed to retent extreme peak flows caused by sudden rainfall runoffs. An example for such storage structure is shown in Figure 5. These pits are distinguished from single wastewater pits due to their large capacity and therefore referred to as retention pits in the rest of this paper. For each storage element at pumping station $i \in \{1, ..., N_s\}$, the infinitesimal level change is computed as the sum of all in- and outflows as

$$\frac{df_V(h_{s,i})}{dt} = d_{s,i} + \sum_{j=1}^{N_y} y_j - u_i, \tag{15}$$

where d_s denotes disturbance inflows to storage tanks, h_s is the water level in storage units and u is the sum of controlled pump flows moving the water towards the next pumping station in line. Note that u is equivalent to the inlet flow $q_{0,t}$ of a gravity pipe located between interconnected storage units, described in eq.(11). Besides, N_y is the number of gravity-driven transport links discharging to the i^{th} storage unit and y_j is the arriving discharge from the j^{th} upstream pumping station, defined in eq.(12). N_s denotes the overall number of pumping stations in the UDN. Moreover, let us consider a map $f_V : \mathbb{R}_+ \to \mathbb{R}_+$ from water level h to water volume V, where f_V is strictly monotonic increasing.

Assumption C3. *For storage elements, $f_V(h)$ in eq.(15) is approximated with a piece-wise linear, strictly monotonic increasing function, parameterized by the level-flow constant of storage tanks.*

Retention pits classify for the assumption on piece-wise linear behavior, while the relation between level and volume simplifies to linear in case of single pits. The piece-wise linear relation along with the hydraulic structure is shown in Figure 5.

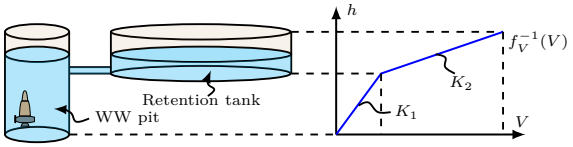


Figure 5: Level-Volume conversion for waste water pit with retention tank.

Tank constants K_1 and K_2 correspond to the slope of the h–V conversion curve, where K_2 is only relevant if pits are equipped with retention tanks. Note that in dry weather the storage capacity of pits is sufficient, hence wastewater flow is typically bypassing the retention tank, acting as a single pit.

3.5 Network description

The links between system components define the topology of the network. The topology considered is a directed tree-graph with nodes representing storage (except the root) and edges transporting flow in between the nodes towards the root. The root of the graph is an outlet point, where the flow is discharged to the receiving environment, e.g., to the WWTP. The tree structure topology is shown in Figure 6.

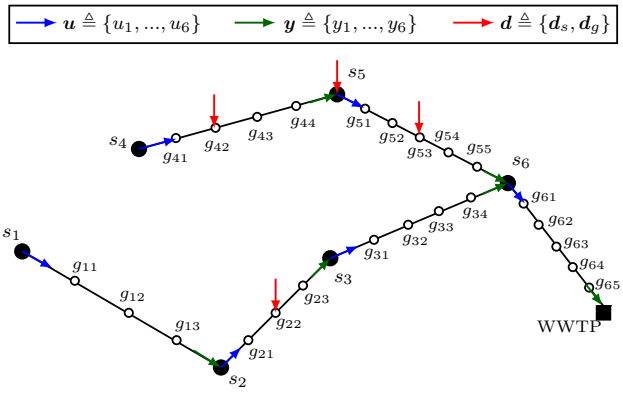


Figure 6: Graph representation where the filled nodes are pumping stations, empty nodes are manholes, whereas edges represent transport pipe segments.

Let us denote the set of nodes corresponding to tanks and pits at the pumping stations with $\mathcal{S} \triangleq \{s_i = (h_{s,i}, d_{s,i}, u_i) \,|\, i \in \{1, ..., N_s\}\}$, where $h_{s,i}$ is the water level, $d_{s,i}$ is the unknown flow disturbance defined in eq.(13) and u_i is

the controlled flow of the i^{th} pumping station. The remaining nodes represent manholes along the gravity sewer transport links. These set of nodes are denoted with $\mathcal{G} \triangleq \{g_{ij} = (h_{g,ij}, d_{g,ij}) \,|\, i \in \{1,...,N_s\}, j \in \{1,...,N_g\}\}$, where $h_{g,ij}$ are the water levels in the j^{th} segment of the gravity pipe rooting from the i^{th} upstream station. Furthermore, $d_{g,ij}$ are lateral inflows along the i^{th} gravity pipe, entering through the j^{th} manhole. These disturbance components are given in eq.(13). The numbering of manholes along the gravity pipes denotes the upstream pumping station (first digit) from which they are numbered in an increasing order towards the downstream station (second digit). The connections between the storage and junction nodes are defined by the piping layout.

Note that the set of $\mathcal{G}$ junction points also represents storage by means of the volume of pipe sections, as the spatially discretized and reduced SV-based model is equivalent to volumes connected in series, where the set of water levels $\{h_{g,i1},...,h_{g,iNg}\}, i \in \{1,...,N_s\}$ relate to water volumes stored in each segment. However, we distinguish between $\mathcal{S}$ and $\mathcal{G}$ for the reason that nodes in $\mathcal{S}$ are subject to control u.

3.6 Discrete network model

In this study, discrete-time network dynamics are utilized for solving the MHE and NMPC problems. The transport and storage dynamics, described in Section 3.2 and Section 3.4, are given for each individual network element, respectively, as

$$\hat{\boldsymbol{h}}_g(t_{k+1}) = \boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\lambda}}\big(u(t_k), \hat{\boldsymbol{h}}_g(t_k), \boldsymbol{d}_g(t_k)\big), \tag{16a}$$

$$\hat{h}_s(t_{k+1}) = \boldsymbol{H}\big(u(t_k), \hat{h}_s(t_k), d_s(t_k), \hat{\boldsymbol{y}}(t_k)\big), \tag{16b}$$

$$\hat{y}(t_k) = \boldsymbol{G}_{\boldsymbol{\theta}}\big(\hat{\boldsymbol{h}}_g(t_k)\big), \tag{16c}$$

where the numerical integration from t_k to t_{k+1} is done by the fixed step, 4^{th} order Runge-Kutta method. Moreover, $\hat{\boldsymbol{h}}_g(t_k) \in \mathbb{R}^{N_x}$ is the vector of water levels along a transport link between two pumping stations. The system dynamics corresponding to transport flows in eq.(11) are defined by $\boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\lambda}} : \mathbb{R}_+ \times \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$. The discrete storage dynamics are given by $\boldsymbol{H} : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^{N_y} \to \mathbb{R}_+$, where N_y is the number of transport links discharging to the specific storage node. The outputs are represented by $\boldsymbol{G}_{\boldsymbol{\theta}} : \mathbb{R}^{N_x} \to \mathbb{R}$, corresponding to the discharged gravity flow previously described in eq.(12).

4 Moving Horizon Estimation

In order to incorporate system knowledge in the state and parameter estimation in form of constraints, a MHE approach is utilized in this paper. Past data samples of the inputs, i.e., pump flows $\{u(t_{k-H_e}), u(t_{k-H_e+1}), ..., u(t_k)\}$ and the outputs, i.e., discharged gravity pipe flows $\{y(t_{k-H_e}), y(t_{k-H_e+1}), ..., y(t_k)\}$ are used up to the current time sample t_k, where H_e is the estimation horizon.

Moreover, for each transport link $i \in \{1, ..., N_s\}$, let us define $\boldsymbol{h}_{g,i} \in \mathbb{R}^{N_x}$ as the vector of water levels in all $j \in \{1, ..., N_g\}$ pipe segments. The MHE problem regarding states (MHSE) and parameters (MHPE) is solved for each transport link i individually. Therefore, we ease the notation and discard the i and j indices and we present the parameter and state estimation for a single transport link.

Additional outputs may be available by means of water level sensor measurements, placed in manholes along the main transporting sewers. Hence, we define $\boldsymbol{C} \in \mathbb{R}^{N_0 \times N_x}$ matrix, associated with a linear mapping which picks all the measured states. N_0 is the number of water level sensors along the transport link. Then, the output vector is given by

$$\boldsymbol{z}_g = \boldsymbol{C}\boldsymbol{h}_g + \nu, \tag{17}$$

where $\nu \in NID(0, \sigma^2)$ is white Gaussian noise accounting for measurement corruption, and $\boldsymbol{z}_g \in \mathbb{R}^{N_0}$. Past data samples $\{\boldsymbol{z}_g(t_{k-H_e}), \boldsymbol{z}_g(t_{k-H_e+1}), ..., \boldsymbol{z}_g(t_k)\}$ of these outputs are utilized together with the input u and output y flow data.

The main purpose of the MHPE is to identify the unknown dynamics of each transport link without using information about the physical properties of sewer pipes, such as pipe diameters, length, slope or roughness. Due to the linearized level-flow scaling introduced by Assumption 1 in Section 3.1, fixed model parameters might result in inaccurate flow predictions, based on whether the pipes are close to being filled or semi-filled. These characteristics can change over time due to seasonality, hence we utilize the MHPE method, attempting to adapt the model parameters to varying flow conditions. Moreover, the dynamics might change over time due to sludge accumulating within certain sections of the pipes, for which the proposed MHPE method is also able to account. As a natural extension, the states are also estimated in a moving horizon fashion (MHSE).

In the following, we distinguish between the horizons of parameter and state estimations. For parameter estimation, we denote the length of the horizon with H_{pe} and for state estimation with H_{se}. Due to the slowly changing dynamics of sewer pipes, we argue that the MHPE is sufficient to carry out above the frequency of the NMPC, having at least one day up to a week long H_{pe} horizon. However, the MHSE problem is executed with a minimum of one day long horizon and with the same frequency as the NMPC. The one day long MHSE horizon is due to the inherent periodicity of the waste water disturbance inflows d^h. Moreover, by calling the MHPE less frequent than the NMPC, we lower the typically high computation demand of MHE algorithms, where state and parameter estimations are carried out simultaneously [Allgöwer et al., 1999].

4.1 Parameter estimation

The MHPE problem of transport flows is formulated as a constrained, least-squares nonlinear minimization problem.

Remark C5. *The control inputs u, depicted in Fig. Figure 6, are estimated considering the polynomial expression of fixed-speed wastewater pumps in the form*

$$\hat{u} = s\mu_0 + s\mu_1 \Delta p + s\mu_2 P_p, \tag{18}$$

where s is the number of running pumps at the pumping station, Δp is the differential pressure and P_p is the power consumption of the pumps [Kallesøe and Knudsen, 2016]. Besides, the outputs y corresponding to discharged flows in Fig. Figure 6 are estimated using mass conservation, detailed in [Kallesøe and Knudsen, 2016]. In this work, we use the outcome of the referenced flow estimation algorithm to provide outputs for the MHE problem.

Let $\boldsymbol{\theta} \triangleq \{\theta_1, \theta_2, \theta_3\} \in \mathbb{R}_+$ denote the set of bounded system parameters and $\boldsymbol{\lambda} \in \mathbb{R}$ denote the parameters corresponding to the Fourier disturbance model. Then, for each transport link, the initial states $\hat{\boldsymbol{h}}_g(t_0)$, the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are found by solving the following finite-dimensional constrained Nonlinear Programming (NLP) problem at time t_k:

$$\begin{aligned}\begin{pmatrix} \boldsymbol{\theta}^* \\ \boldsymbol{\lambda}^* \\ \hat{\boldsymbol{h}}_g^*(t_0) \end{pmatrix} = \underset{\boldsymbol{\theta},\boldsymbol{\lambda},\hat{\boldsymbol{h}}_g(t_0)}{\operatorname{argmin}} & \sum_{i=k-H_{pe}}^{k} \big(y(t_i) - \hat{y}(t_i)\big)^2 \\ & + W_1 \|\boldsymbol{z}_g(t_i) - \hat{\boldsymbol{z}}_g(t_i)\|^2,\end{aligned} \tag{19a}$$

subject to sewer dynamics:

$$\hat{\boldsymbol{h}}_g(t_{i+1}) = \boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\lambda}}\big(u(t_i), \hat{\boldsymbol{h}}_g(t_i), \boldsymbol{d}_g(t_i)\big), \tag{19b}$$

$$\hat{y}(t_i) = \boldsymbol{G}_{\boldsymbol{\theta}}\big(\hat{\boldsymbol{h}}_g(t_i)\big), \tag{19c}$$

$$\hat{\boldsymbol{z}}_g(t_i) = \boldsymbol{C}\hat{\boldsymbol{h}}_g(t_i), \tag{19d}$$

and inequality constraints:

$$\mathbf{0} \leq \hat{\boldsymbol{h}}_g(t_i) \leq \overline{\boldsymbol{h}}_g, \tag{19e}$$

$$\mathbf{0} < \boldsymbol{\theta} \leq \overline{\boldsymbol{\theta}}, \tag{19f}$$

$$0 \leq \hat{y}(t_i) \leq \overline{y}, \tag{19g}$$

where $\hat{\boldsymbol{h}}_g(t_i) \in \mathbb{R}^{N_x}$ is the vector of states corresponding to a transport link. Note that y represents the discharged flow, while $\boldsymbol{h}_g$ represents the vector of water levels in the manholes. Therefore, we use W_1 as a weighing constant in eq.(19a), scaling the water levels to the magnitudes of the discharged flows. The constraints in eq.(19e), eq.(19f) and eq.(19g) impose bounds on state variables, parameters and the output, respectively. Note that the states $\hat{\boldsymbol{h}}_g$ and the output variable y correspond to physically measurable water levels and the discharged flow in sewer pipes, respectively. The water level measurements addressed in eq.(17) are denoted by $\boldsymbol{z}_g$. The state and output bounds are

chosen consistent with physically meaningful values, such as water levels and flows are never negative inside the pipes.

Moreover, the upper bound on the states $\overline{\boldsymbol{h}}_g$ is the maximum allowed water level in manholes defined by the physical height from the bottom to the surface. From eq.(8), we know that the pipe parameters $\boldsymbol{\theta}$ are positive. Besides, using eq.(8), we approximate a physically meaningful maximum value for the pipe diameters, spatial step, time steps, and friction values. In eq.(19g), the flow is assumed to be non-negative inside the pipe and the maximum is defined as the physically possible full-pipe flow. The bound constrained nonlinear minimization problem in eq.(19) is then solved via a gradient descent algorithm.

Note that the number of sections N_x, illustrated in Figure 4, is treated as an auxiliary variable in the NLP, meaning that the MHPE problem can be carried out multiple times with different grid sizes to find the optimal number of sections regarding some performance index, e.g., Root Mean Squared Error (RMSE). This procedure is not detailed here, as the reader may consult a previous study focusing on how to choose grid size for a flow-based SV-PDE model in [Balla et al., 2020a].

4.2 State estimation

Full state measurement in the proposed sewer system model requires sensor installation inside all available manholes within the network. This is neither economically feasible, nor required by the control point of view. However, it is assumed that there is a subset of states $\boldsymbol{z}_g$ which are measured. Similarly to the MHPE problem in eq.(19a), the full system states, i.e., $\boldsymbol{h}_g$ water levels are being reconstructed out of a few output measurements by means of the MHSE. However opposed to the MHPE problem in eq.(19a), the state estimation is solved at each control time step t_k, thus providing initial state estimates for the NMPC. The MHSE reconstructs $\boldsymbol{h}_g(t_{k-H_{se}}), ..., \boldsymbol{h}_g(t_k)$ states, based on the measured inputs $\boldsymbol{u}(t_{k-H_{se}}), ..., \boldsymbol{u}(t_k)$, measured outputs $y(t_{k-H_{se}}), ..., y(t_k)$ and $\boldsymbol{z}_g(t_{k-H_{se}}), ..., \boldsymbol{z}_g(t_k)$ over the horizon H_{se}, while the dynamics are provided as constraints. The MHSE is defined by the following optimization problem:

$$\begin{pmatrix} \hat{\boldsymbol{h}}_g^*(t_{k-H_{se}}) \\ \vdots \\ \hat{\boldsymbol{h}}_g^*(t_k) \end{pmatrix} = \underset{\hat{\boldsymbol{h}}_g(t_{k-H_{se}}),\ldots,\hat{\boldsymbol{h}}_g(t_k)}{\operatorname{argmin}} \sum_{i=k-H_{se}}^{k} \left(y(t_i) - \hat{y}(t_i)\right)^2 + W_2 \|\boldsymbol{z}_g(t_i) - \hat{\boldsymbol{z}}_g(t_i)\|^2, \tag{20}$$

subject to the dynamics in eq.(19b) to eq.(19d), the state constraint in eq.(19e) and the output constraint in eq.(19g). W_2 is a weighing matrix for scaling levels to flows, similarly as in eq.(19a). Note that from the solution of the MHSE problem in eq.(20), the estimated state vector at the current time step $\hat{\boldsymbol{h}}_g^*(t_k)$ is used. The same gradient descent algorithm [Wills and Ninness, 2008] is used to solve the problem in eq.(20), as for the MHPE.

5 Control design

The vector of control variables is defined by $\boldsymbol{u} \in \mathbb{R}^{N_s}$, where all individual pump flows are aggregated at the nodes $s \in \mathcal{S}$, representing the N_s pumping stations in the network. The states correspond to levels along transport links and levels in storage units, e.g., pits. The state vector is defined by:

$$\boldsymbol{h} \triangleq (\boldsymbol{h}_s^\top, \boldsymbol{h}_{g,1}^\top, \boldsymbol{h}_{g,2}^\top, ..., \boldsymbol{h}_{g,N_s}^\top)^\top, \tag{21}$$

where $\boldsymbol{h}_s \in \mathbb{R}^{N_s}$ represents the vector of levels in storage elements and for each $i \in \{1, ..., N_s\}$ transport link $\boldsymbol{h}_{g,i} \in \mathbb{R}^{N_{x,i}}$ consists of $N_{x,i}$ entries depending on how many sections each transport link is discretized into. The outputs, i.e., discharged flows at the end of each transport link, are given by:

$$\boldsymbol{y} \triangleq (y_1, y_2, ..., y_{N_s})^\top, \tag{22}$$

where the last element y_{N_s} is the discharged flow leading to the root of the network, which we hereinafter denote as y_w. Introduced previously, the closed-loop control scheme together with the MHE problem is depicted in Figure 1. Note that the rain run-off dynamics along with the weather forecasts provide flow inputs to the proposed closed-loop control scheme.

5.1 NMPC problem

To account for both the dry- and wet-weather loads in a computationally efficient way, the NMPC problem is formulated over two subsequent prediction horizons. To this end, let H_{p_1} denote the predictions over the near future (nowcasts) and H_{p_2} the predictions further in the future (forecasts), respectively. This formulation of the NMPC problem is motivated by the inherent periodicity of the household disturbances d^h, which typically corresponds to one day. However, the network is exposed to large disturbance loads in terms of the d_r rain run-off, where the so-called nowcasts are reliable only within a short horizon. According to [Löwe et al., 2014], rainfall radars can provide sufficient accuracy of spatial and temporal resolution for urban catchments only up to a 2 (h) horizon. Therefore, computing the decision variables for $T = 24$ (h) is unnecessary, and results in high computational costs. Instead, let $H_p = H_{p_1} + H_{p_2}$ be the entire length of the horizon, T_s the time step and let us define $\boldsymbol{h} \triangleq [\boldsymbol{h}_s^\top, \boldsymbol{h}_g^\top]^\top$ as the entire state vector. Then, the NMPC problem for the entire network is given as

$$\min_{\substack{\Delta \boldsymbol{u}(0),\dots,\Delta \boldsymbol{u}(H_{p_1}) \\ \Delta \boldsymbol{u}(H_{p_1}+1),\dots,\Delta \boldsymbol{u}(H_p-1)}} \sum_{k=0}^{H_p-1} \mathcal{L}\big(\Delta \boldsymbol{u}(t_k), \boldsymbol{h}(t_k), y_w(t_k)\big) + \mathcal{S}\big(\boldsymbol{h}(t_{H_p})\big) \tag{23a}$$

subject to transport link dynamics

$$\boldsymbol{h}_g(t_{k+1}) = \boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\lambda}}\big(\boldsymbol{u}(t_k), \boldsymbol{h}_g(t_k), \boldsymbol{d}_g(t_k)\big), \tag{23b}$$

$$\boldsymbol{u}(t_{k+1}) = \boldsymbol{u}(t_k) + \Delta \boldsymbol{u}(t_k), \tag{23c}$$

$$y_w(t_k) = \boldsymbol{G}_{\boldsymbol{\theta}}\big(\boldsymbol{h}_g(t_k)\big), \tag{23d}$$

storage dynamics

$$\boldsymbol{h}_s(t_{k+1}) = \boldsymbol{H}\big(\boldsymbol{u}(t_k), \boldsymbol{h}_s(t_k), \boldsymbol{d}_s(t_k), \boldsymbol{y}(t_k)\big), \tag{23e}$$

state, input and output constraints

$$\underline{\boldsymbol{V}} + \boldsymbol{V}_{of}(t_k) \le \boldsymbol{f}_V\big(\boldsymbol{h}_s(t_k)\big) \le \overline{\boldsymbol{V}} + \boldsymbol{V}_{of}(t_k), \tag{23f}$$

$$0 \le \boldsymbol{h}_g(t_k) \le \overline{\boldsymbol{h}}_g, \tag{23g}$$

$$\underline{\boldsymbol{u}} \le \boldsymbol{u}(t_k) \le \overline{\boldsymbol{u}}, \tag{23h}$$

$$0 \le y_w(t_k) \le \overline{y}_w, \tag{23i}$$

terminal constraint

$$\underline{\boldsymbol{V}} \le \boldsymbol{f}_V\big(\boldsymbol{h}_s(t_{H_p})\big) \le \overline{\boldsymbol{V}}, \tag{23j}$$

$$0 \le \boldsymbol{h}_g(t_{H_p}) \le \overline{\boldsymbol{h}}_g, \tag{23k}$$

where $\Delta \boldsymbol{u}(t_k) \triangleq \boldsymbol{u}(t_{k+1}) - \boldsymbol{u}(t_k)$ is the input change. The integral action accounts for smooth and slow system response, avoiding sudden jumps in the control action. The proposed optimization problem in eq.(23a) is solved for $[\Delta \boldsymbol{u}^\top(0), ..., \Delta \boldsymbol{u}^\top(H_{p_1})]^\top \in \mathbb{R}^{H_{p_1}}$, whereas the problem of finding the decision variables over H_{p_2} is reduced to finding H_{p2}/τ number of optimization variables, where τ defines how many T_s control steps each decision variable is kept constant. This is due to the fact that the control over H_{p_2} does not require the same precision as for the nowcasts over H_{p_1}. The stage and terminal costs formulated in eq.(23a) are sums of square-type functions, and the multiple operational objectives in the stage cost $\mathcal{L}$ are detailed in Section 5.2.

The dynamics $\boldsymbol{F}_{\boldsymbol{\theta},\boldsymbol{\lambda}}$, $\boldsymbol{H}$ and $\boldsymbol{G}_{\boldsymbol{\theta}}$ are defined in eq.(16a), eq.(16b), eq.(16c) for the entire network and the output equation in eq.(23d) is formulated on the discharged flow y_w arriving to the root of the network. The nonlinear level to volume conversion is kept outside of the optimization, where $\boldsymbol{f}_V$ is a piece-wise linear map from eq.(15). Furthermore, the control is subject to state constraints on pipe states in eq.(23g) and storage states in eq.(23f), where $\boldsymbol{V}_{of} \in \mathbb{R}^{N_s}$ is the vector of slack variables, lifting the upper and lower state bounds. This variable

is considered as a virtual volume triggered at times when the physical limits of storage elements are extended. In case of an overflow, the slack variable lifts both the lower $\underline{\boldsymbol{V}}$ and upper $\overline{\boldsymbol{V}}$ state bounds, thereby keeping track of the excess storage [Gelormino and Ricker, 1994]. The upper bound of states corresponds to the physically maximum volume capacity in the storage nodes. The lower limit is defined by the user with the criteria that a minimum volume of water needs to be kept in the storage tanks at all times to fully cover the wastewater pumps, hence avoiding the dry-run of pumps.

Remark C6. *In case of overflows, the excess water volume leaves the network immediately. This is assured by constraining the slack variables*

$$\mathbf{0} \leq \boldsymbol{V}_{of}(t_k), \tag{24}$$

meaning that spilled sewage escapes all $s \in \mathcal{S}$ storage nodes.

Furthermore, eq.(23h) imposes physical bounds on the minimum and maximum flow capacity of pumps. Equation eq.(23i) formulates a constraint regarding the inflow capacity of the WWTP, where $\overline{y}_w$ is the maximum allowed inflow defined by the size of the WWTP. For closed-loop stability considerations of the NMPC, the terminal constrains in eq.(23j) and eq.(23k) are introduced along with the terminal cost $\mathcal{S}$ in eq.(23a) to enforce stability [Allgöwer et al., 1999]. The formulation in eq.(23a) is solved via a gradient descent algorithm, where the dynamics are discretized according to Section 3.6.

5.2 Objectives

The control problem addressed in eq.(23) has multiple objectives with different priorities. For an extensive analysis on choosing objectives in UDN control, consult [Ocampo-Martinez, 2010; Mollerup et al., 2016]. To prioritize objectives, the stage cost is formulated as a linearly weighted sum and the terminal cost is given as

$$\mathcal{L}\big(\Delta \boldsymbol{u}(t_k), \boldsymbol{h}(t_k), y_w(t_k)\big) \triangleq \sum_{j=1}^{\Gamma} \lambda_j \mu_j \mathcal{F}_j(t_k), \tag{25a}$$

$$\mathcal{S}\big(\boldsymbol{h}(t_{H_p})\big) \triangleq \boldsymbol{h}^{\top}(t_{H_p}) \boldsymbol{P} \boldsymbol{h}(t_{H_p}), \tag{25b}$$

where λ_j denotes the scaling weights among the different objectives and Γ is the total number of the control objectives. The scaling constants μ_j normalize each objective term to dimensionless values, such that water levels and flows become comparable. Furthermore, the terminal cost $\mathcal{S}$ is defined for all states, where the symmetric positive definite matrix $\boldsymbol{P}$ is the solution to the associated Ricatti equation. Note that $\boldsymbol{P}$ is designed based on the weights λ_j on the state and input terms in the stage cost function $\mathcal{L}$. Moreover, the Jacobian linearization of the network model is considered at an operating point, where state values are at their 25 % utilization of their upper limit. Furthermore,

disturbance and input flows are considered at the daily mean of household wastewater production without rain.

The most common control criteria in sewer network control is related to overflow minimization

$$\mathcal{F}_1(t_k) \triangleq \boldsymbol{V}_{of}^{\top}(t_k)\boldsymbol{\Omega}_1\boldsymbol{V}_{of}(t_k), \tag{26}$$

where $\boldsymbol{V}_{of} \in \mathbb{R}^{Ns}$ is the vector of slack variables, representing overflow volumes. The overflows $\boldsymbol{V}_{of}$ between stations are prioritized according to the diagonal $\boldsymbol{\Omega}_1$ matrix, where $\operatorname{diag}(\boldsymbol{\Omega}_1) \in [0, 1]$. Note that the weight corresponding to the overflow objective λ_1 is significantly higher than any other weights, in order to make the use of the overflow slack variables undesirable if possible.

The penalty on water level in storage elements is given by

$$\mathcal{F}_2(t_k) \triangleq \boldsymbol{h}_s^{\top}(t_k)\boldsymbol{\Omega}_2\boldsymbol{h}_s(t_k), \tag{27}$$

where $\boldsymbol{h}_s \in \mathbb{R}^{Ns}$ is the vector of water levels in storage nodes and $\boldsymbol{\Omega}_2$ is the diagonal weighting matrix, where $\operatorname{diag}(\boldsymbol{\Omega}_2) \in [0, 1]$. The level in storage nodes is minimized to avoid long retention times and thus odor problems occurring in the waste water tanks. Moreover, the weight matrix $\boldsymbol{\Omega}_2$ allows to adjust the filling sensitivity of storage elements, meaning that sensitive tanks are filled slower and emptied faster than less sensitive storage tanks.

The inputs are minimized such that

$$\mathcal{F}_3(t_k) \triangleq \Delta\boldsymbol{u}^{\top}(t_k)\boldsymbol{\Omega}_3\Delta\boldsymbol{u}(t_k), \tag{28}$$

where $\Delta\boldsymbol{u} \in \mathbb{R}^{Ns}$ is the vector of input change regarding the aggregated flows delivered by sewer pumps placed at each network node $s \in \mathcal{S}$. Moreover, $\boldsymbol{\Omega}_3$ is the weighting matrix between the network nodes, where $\operatorname{diag}(\boldsymbol{\Omega}_3) \in [0, 1]$.

The system states in any $g \in \mathcal{G}$ nodes, i.e., gravity pipe sections are water levels, representing storage along the edges of the underlying network graph. Hence, we penalize manholes prone to suffer overflows:

$$\mathcal{F}_4(t_k) \triangleq \boldsymbol{h}_g^{\top}(t_k)\boldsymbol{\Omega}_4\boldsymbol{h}_g(t_k), \tag{29}$$

where $\boldsymbol{h}_g$ is the vector consisting of selected network nodes which can overflow under high loads. Similarly to all objectives, $\operatorname{diag}(\boldsymbol{\Omega}_4) \in [0, 1]$ allows to adjust priority of overflows and filling sensitivity of manholes.

In this work, we consider the objective of controlling the inflow to the WWTP, which is formulated as follows

$$\mathcal{F}_5(t_k) \triangleq \Big(y_w(t_k) - \frac{1}{H_p}\sum_{k=0}^{H_p-1} y_w(t_k)\Big)^2, \tag{30}$$

where the inflow variation to the WWTP is minimized. This is achieved, by calculating a reference flow as an average inflow over the same time period as the time periodicity of the d^h houshold disturbances, which typically corresponds

to one day. This objective allows to correct the irregular inflow pattern to the root of the network, which influences negatively the operation of the WWTP. An extensive study on the regulation of inlet flow to the WWTP is detailed in [van Heeringen et al., 2016].

6 Numerical results

We now present the numerical results. The results are related to the closed-loop control scheme performance when both the MHE and NMPC are considered. As presented in Section 3.6, the network model uses the fixed step, 4^{th} order Runge-Kutta method for the finite-difference approximation of the derivative terms. The optimization problem related to the MHSE and MHPE has been solved via a Gauss-Newton gradient-based method. This solver is chosen due to the reliable estimate of the Hessian for least-squares type problems, such as the MHE formulation in this paper [Wills and Ninness, 2008]. Furthermore, the optimization problem related to the NMPC has been solved via direct multiple shooting in the symbolic framework `CasADI` [Andersson et al., 2019]. A primer-dual interior point solver `IPOPT` [Wächter and Biegler, 2006] has been chosen to solve the nonlinear optimization problem in eq.(23), due to its ability to leverage sparse linear algebra computations. Since the sampling interval is significantly short compared to the dynamics and the sampling time of the NMPC, the optimization problem has been solved by warm-starting at each control time step. Error tolerance of 10^{-5} has been chosen in both the MHE and NMPC problems. Moreover, all the numerical experiments have been carried out on a 2.6 (GHz), Intel Core i7 machine with 16-GB RAM.

Following the model methodology discussed in Section 3, the control-oriented model is identified based on measurements extracted from a physically-based HiFi network, shown in Figure 7 (left). The topological representation as a directed graph along with the location of sensors are depicted in Figure 7 (right). To test the NMPC with the MHE strategy, real rain intensity and wastewater flow are utilized starting from 1 September 2019 to 30 September 2019[3]. These data are used as the load to the HiFi case study network.

6.1 Baseline controller

In this work, we follow the guidelines proposed in [Lund et al., 2018] to benchmark the MPC performance, where the current state-of-art uses CSO and flooded volume as an evaluation measure. The proposed NMPC/MHE strategy is tested against an on/off rule-based controller, most commonly used as baseline control in both practice and literature [Lund et al., 2018], [García et al., 2015].

[3]Rain intensity data have been extracted from the weather archive of the Danish Meteorological Institute (DMI), while the domestic wastewater flow measurement data have been obtained and scaled down from the municipality of Fredericia, Denmark.

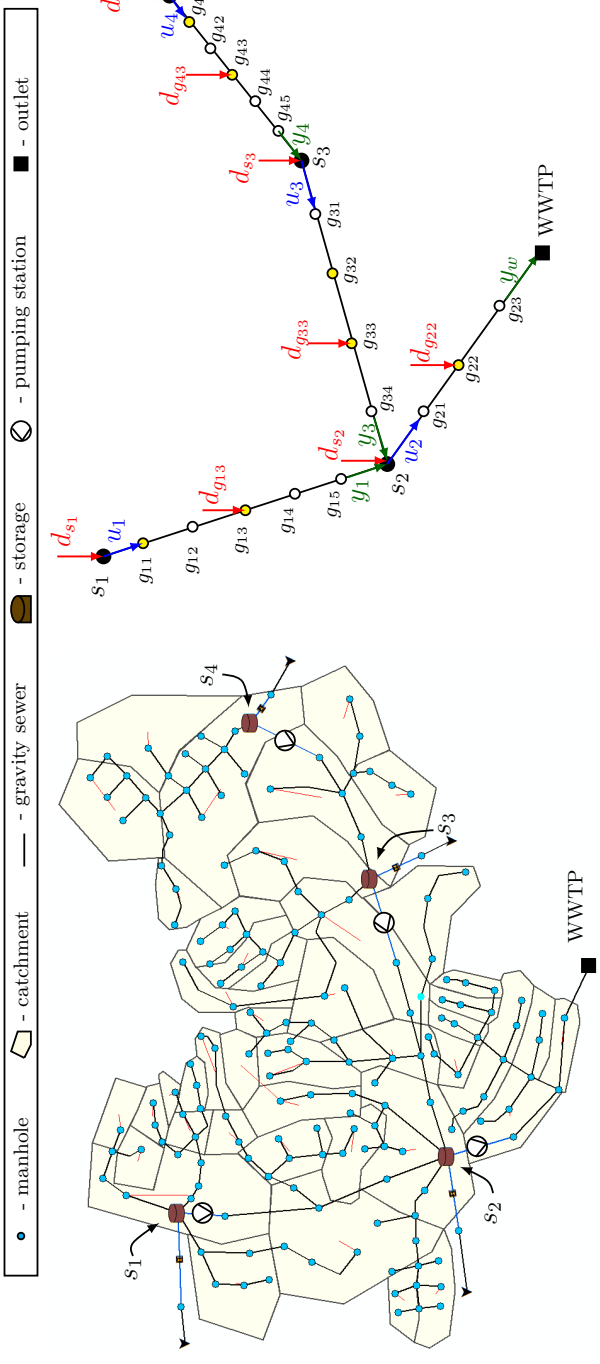


Figure 7: High-fidelity network model in the MIKE Urban simulation software (left), and the graph representation (right), where the number of empty nodes represents the number of discretized sections. The filled, yellow nodes represent level sensors placed in manholes.

The switching rule together with the aggregated flow provided by the pumps at each pumping station under the rule-based control is given by

$$u(t_k) = \begin{cases} \overline{u}, & \text{if} \quad h_s(t_k) \geq \overline{h}_s, \quad \forall t_k, \\ \underline{u}, & \text{if} \quad h_s(t_k) \leq \underline{h}_s, \quad \forall t_k, \\ u(t_{k-1}), & \text{otherwise}, \qquad \forall t_k, \end{cases} \tag{31}$$

where $h_s(t_k)$ is the measured water level in the storage element. Upper and lower bounds of the inlet flow $\overline{u}, \underline{u}$ are equivalent to the bounds in eq.(23h), corresponding to the maximum and minimum flow capacity of the pumps. Threshold values $\overline{h}_s$, $\underline{h}_s$ are equivalent to eq.(23f).

6.2 Case study

The topological properties of the HiFi network shown in Figure 7 are summarized in Table 1. We consider a combined sewer network, where both rain runoff

Attribute	Number	Variable	Unit
Single pits	3	h_s	(m)
Retention pits	1	h_s	(m)
Pumping stations	4	u	(m^3/h) or (m^3/s)
Level sensors in manholes	7	h_g	(m)
Catchment runoff	45	d^r	(m^3/h) or (m^3/s)
Waste water inflow	10	d^h	(m^3/h) or (m^3/s)
Treatment plants	1	y_w	(m^3/h) or (m^3/s)

Table 1: MIKE Urban HiFi simulation properties.

and wastewater enters the sewer via the catchments (yellow areas) and the manholes (junction points), respectively. The network consists of 170 manholes, 170 gravity pipes, moreover three single pits (s_1, s_3 and s_4) and a retention pit (s_2). Using the proposed modelling methodology, the tree graph representation of the UDN and the control variables in the reduced graph representation are given by:

$$\boldsymbol{u} \triangleq (u_1, u_2, u_3, u_4)^\top, \tag{32a}$$
$$\boldsymbol{h} \triangleq (h_{s_1}, h_{s_2}, h_{s_3}, h_{s_4}, \boldsymbol{h}_{g_1}^\top, \boldsymbol{h}_{g_2}^\top, \boldsymbol{h}_{g_3}^\top, \boldsymbol{h}_{g_4}^\top)^\top, \tag{32b}$$
$$\boldsymbol{y} \triangleq (y_1, y_2, y_3, y_w)^\top, \tag{32c}$$
$$\boldsymbol{d} \triangleq (d_{s_1}, d_{s_2}, d_{s_3}, d_{s_4}, d_{g_{13}}, d_{g_{22}}, d_{g_{33}}, d_{g_{43}})^\top, \tag{32d}$$

where the state vector $\boldsymbol{h}$ consists of the gravity pipe subvector states $\boldsymbol{h}_{g_1} \in \mathbb{R}^5$, $\boldsymbol{h}_{g_2} \in \mathbb{R}^3$, $\boldsymbol{h}_{g_3} \in \mathbb{R}^4$, $\boldsymbol{h}_{g_4} \in \mathbb{R}^5$. Moreover, the number of pumping stations is $N_s = 4$ and the rain and domestic wastewater disturbances are concentrated on certain network nodes. The control time step of the NMPC is $T_{\text{NMPC}} =$

10 (min), while the rule-based controllers operate with a $T_{\text{On/Off}}$ = 1 (min) sampling period. The prediction horizon for nowcasts is H_{p_1} = 2 (hours) and for the forecasts H_{p_2} = 22 (hours), summing up to a total of one-day horizon.

The MHPE is carried out with a horizon H_{pe} = 2 (day) and utilized with a T_{pe} = 6 (hours) period time. At every 6 (hours), the MHPE uses data from the past two days and updates the $\boldsymbol{\theta}$ system and $\boldsymbol{\lambda}$ disturbance parameters accordingly. A minimum of two days has been chosen to detect the one-day periodicity of the household wastewater with the Fourier disturbance model. The MHSE is carried out with the same horizon as the MHPE, i.e., H_{se} = 2 (day) and utilized with the same frequency as the NMPC, i.e., T_{se} = 10 (min).

6.3 Simulation environment

To test the NMPC/MHE controller, the MIKE Urban [DHI, 2017] simulation software has been used to simulate the HiFi network model depicted in Figure 7. MIKE Urban allows for the hydraulic and hydrodynamic simulation of flows and water levels by numerically solving the full SV equations in eq.(1). The model of the network in MIKE Urban is defined by the true physical parameters of the hydraulic components. In Figure 7, the catchments (yellow areas) are connected to manholes, hence water volumes enter the pipe network through the network nodes. The simulation is done in two steps: First, the network loads are computed with the catchment dynamics. Then, the rain runoff together with the household waste and groundwater appears as a load (marked with red arrows in Figure 7).

In this work, the NMPC/MHE strategy is used as an upper level controller, where the MIKE Urban model is simulated as a virtual reality. To this end, we utilize the MIKE 1D Application Programming Interface (API) [DHI, 2019], [DHI] which allows us for setting flow references to the pumps and reading flow and level values of hydraulic structures during simulation. These flow references are calculated at every T_{NMPC} time and then used as set-points for local PID controllers based on (virtual) flow sensor measurements placed right after the downstream end of the pumps. The HiFi model runs with a sampling time of $T_{\text{on/off}}$, however the set-points for the PIDs are kept constant during the time interval T_{NMPC}.

6.4 Identification results

To estimate the parameters and the initial states in transport pipes, the measurements $\boldsymbol{z}$ along with the historical data on the estimated inlet and discharged flows $\boldsymbol{u}$ and $\boldsymbol{y}$ are utilized. To show the capabilities of the MHE approach, the initial conditions estimated for the problem in eq.(23) are compared to the measurements in the HiFi simulation, shown in Fig Figure 8.

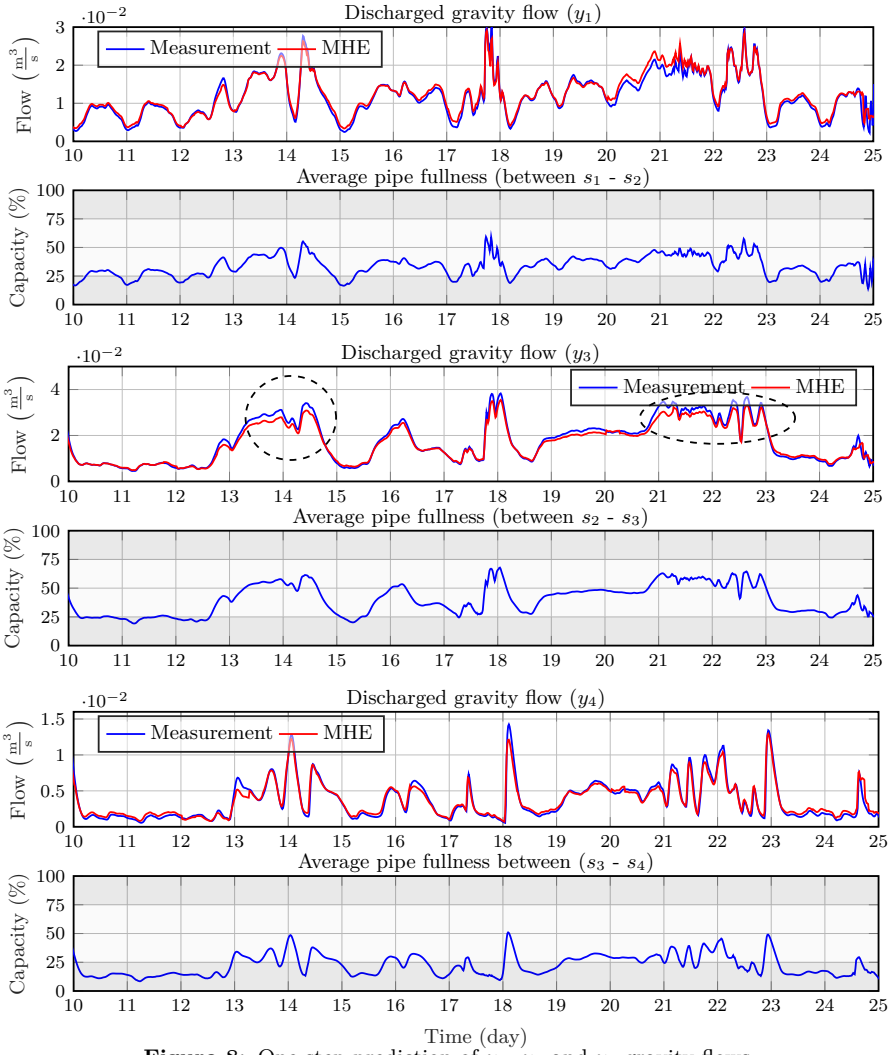


Figure 8: One-step prediction of y_1, y_2 and y_3 gravity flows.

The results show 15 days, where the estimated flow $\hat{y}(t_0)$ is the result of the MHPE and MHSE blocks depicted in Figure 1. Note that instead of showing the estimated states $\hat{h}_g(t_0)$, we rather show the discharged flow $\hat{y}(t_0)$, obtained by eq.(19c). From the application point of view, this is reasonable, since the water level in the last section of a transport pipe does not indicate how the volume is affected in the receiving hydraulic structure (storage tanks), opposed to the volumetric flow rate. Besides, for each transport link g_1, g_2 and g_3, we show the average pipe fullness along $x \in L$ length of the pipes, indicating the capacity of each pipeline.

In the HiFi simulation environment, all sewer pipes are circular, hence the flow-level translation imposed by Assumption 1 (Section 3) is only accurate for small variations of water level. In order to show the variations of water levels inside gravity pipes, we illustrate two different operating regions (shaded areas in Fig. 8.). The middle range of the pipe is defined between 25 – 75 %, moreover the lower and upper regions between 0 – 25 % and 75 – 100 %, respectively. Small level variations within these regions are expected to yield accurate flow estimates based on Assumption 1.

As shown in Fig. Figure 8, the one-step predictions of the MHE strategy produces accurate estimates of the discharged flows $\hat{y}(t_0)$ in comparison with the flow measurements ($y(t_0)$) from the HiFi model. This is achieved without using any flow sensor in the network, however, assuming the linear flow-level relation in the internal model. The prediction results in Figure 8 show inaccurate flow estimates at certain time steps imposed by the simplified pipe geometries. This is because the internal model with the simplified geometry attempts to produce flows close to the ones obtained by the linear flow-level mapping, rather than the actual flow. This is most visible on y_3 at periods encircled with dashed black lines. During both of these periods, the pipes are filled up from 25 % to 50 %, where the previous level with the time window of the MHPE only provides information of low-filled, slow-varying level conditions. Hence, the internal model underestimates the actual flow by calculating lower volume than there is inside the middle operating range of circular pipes.

6.5 Control results

The control results aim to show the benefits of distributing level sensors along the network to obtain the data-driven network model from the full SV-PDEs. The proposed methodology is compared with a traditional, two-point controller detailed in Section 6.1, most commonly used by water utility operators. The NMPC acts as a global controller and computes reference points to local controllers (as depicted in Figure 1). To evaluate the closed-loop performance of the NMPC/MHE strategy, we selected two days with heavy overflows due to the insufficient capacity in the network. The numerical results are shown in Figure 9-A and Figure 9-B for each $i \in \{1, ..., N_s\}$ pumping station, showing the time evaluation of the disturbances, overflows, tank levels and the pumped inlet flows. In the case study, all $\mathbf{\Omega}$ weight values are set equal, hence none of the stations are prioritized over the other. This means that overflow and the filling sensitivities are not prioritized. As the overflows are not avoidable over the selected two days period, the overall goal is to reduce the amount of flooded volume.

The disturbance signals used in the HiFi simulator are historic rain and wastewater flows. To evaluate the NMPC/MHE performance under uncertainty, we generated imperfect forecasts for the internal model of the NMPC. To this end, $n = 10$ different disturbance scenarios have been created by adding normally distributed random data on top of the historic flow signals.

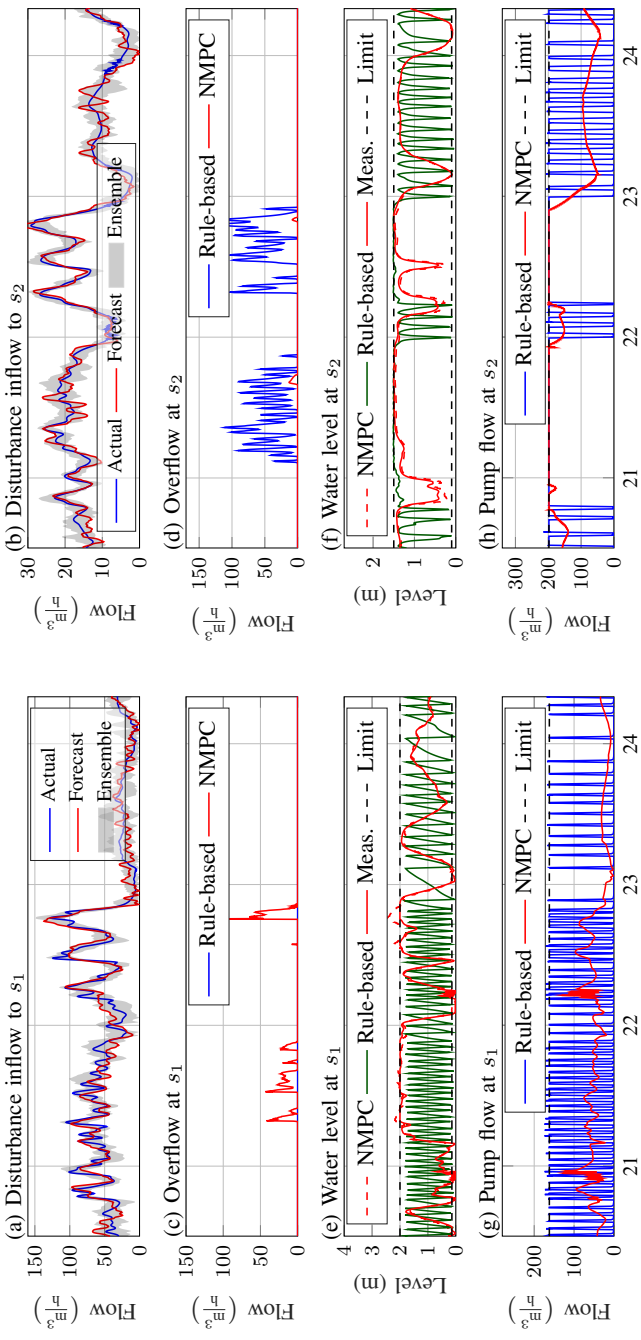


Figure 9-A: NMPC with MHPE and MHSE results compared with rule-based control. (a), (b), (i), (j) show disturbance inflows $d_{s,i}$, (c), (d), (k), (l) show overflows $q^{of}_{s,i}$, (e), (f), (m), (n) show states (water levels), while (f), (h), (o), (p) show inputs (pump flows) for each $i \in \{1, ..., N_s\}$ station, respectively.

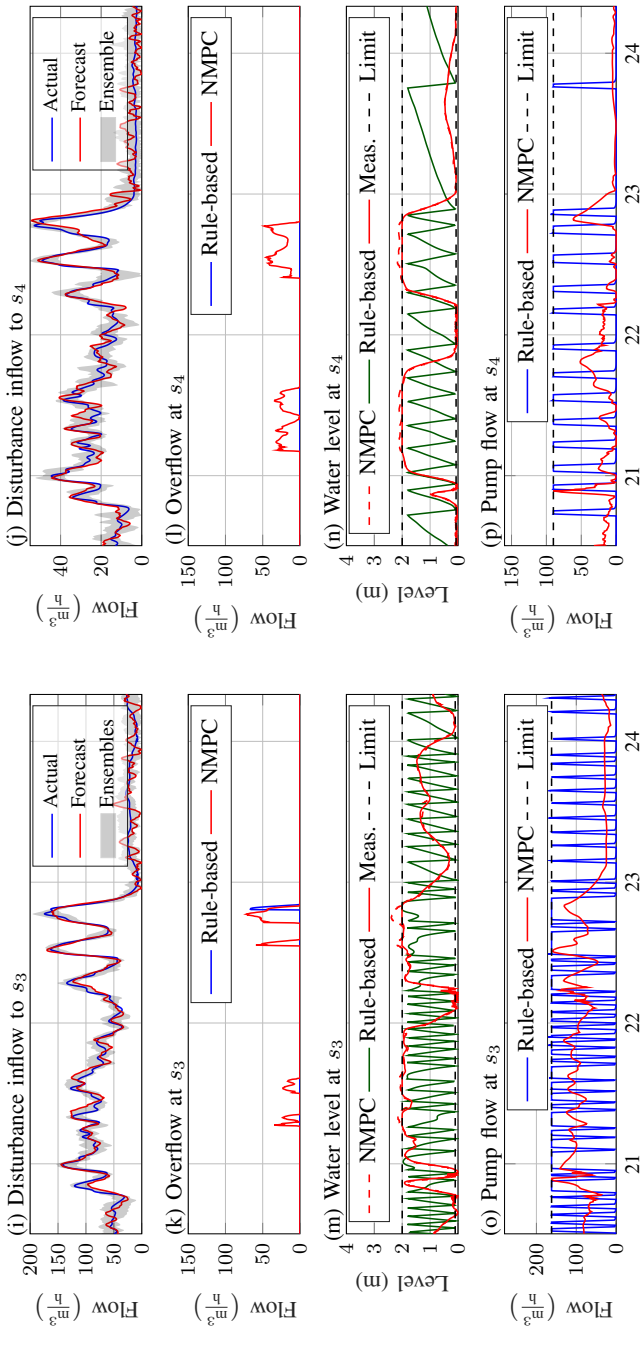


Figure 9-B: NMPC with MHPE and MHSE results compared with rule-based control. (a), (b), (i), (j) show disturbance inflows $d_{s,i}$, (c), (d), (k), (l) show overflows $q^{of}_{s,i}$, (e), (f), (m), (n) show states (water levels), while (f), (h), (o), (p) show inputs (pump flows) for each $i \in \{1, ..., N_s\}$ station, respectively.

As shown in Figure 9-A/B(a), (b), (i), and (j), a set of ensemble of forecasts is produced, indicating a range of possible disturbances. The characterization of the uncertainty for each disturbance is given in the Appendix.

To show the deviation between the prediction by the controller and the state measurement retrieved from the HiFi simulator under uncertainty, we indicated the one-step predictions in Figure 9-A/B(e), (f), (m), and (n) with the dashed red line. Note that the upper constraint is violated under overflow events, due to the slack approximating the volume of overflows. Furthermore, the lower bounds are violated in case the forecasts indicate higher volumes than expected, ending up in the dry-run of the pumps. The NMPC/MHE strategy overflows the upstream tanks (s_1, s_3 and s_4) at times where the rule-based method avoids overflows. This is depicted in Figure 9-A/B(c), (k) and (l), where all storage nodes are prepared by being emptied before the load increases on the network and therefore the controllers distribute the flooded volumes among the corresponding stations, as shown in Figure 9-A/B(e), (m) and (n). Opposed to the rule-based strategy, the overflows are intentional and coordinated, thereby avoiding the overload on the retention tank s_2 during the heavy load period. As the system states (i.e., water levels) show, the overflows are shifted in time as the storage nodes attempt to hold back water until their capacity allows. Note that the precise flow-level translation and the precise discharged flow predictions (y_1, y_3 and y_4) guarantee the proper management of the pits (s_1, s_3, s_3) and the retention pit (s_2), mitigating the overflow volumes optimally. The comparison of overflow reduction between the baseline and NMPC controllers is shown in Figure 10. Applying the proposed NMPC/MHE strategy results in 28 % cumulative overflow volume decrease over the considered period.

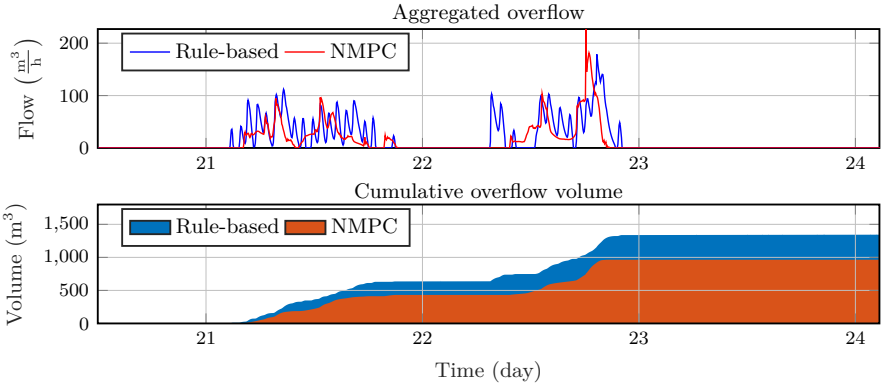


Figure 10: Overflow comparison throughout the entire network.

To comprehensively illustrate the practicability of implementing the NMPC/MHE framework, we report on the computational complexity and the dimensions of the optimization problem. To this end, we reduced the network graph shown in Figure 7, by excluding one and two pumping stations, respectively.

The results, along with the size of the optimization problem, are shown in Table Table 2.

Num. of stations	Avg. CPU time (s)	Max CPU time (s)	Decision var.	Constr.	Param.
4 ($s_{1,2,3,4}$)	2.14	8.37	5361	8385	1216
3 ($s_{1,2,3}$)	1.8	4.36	3912	6072	912
2 ($s_{1,2}$)	1.15	1.86	2608	4192	610

Table 2: Computation complexity with different number of stations.

The optimization problem is carried out on the case study network scaling from two pumping stations to the full extent of the network. As shown, the size of the optimization problem is increasing with including more pumping stations and transport links, however, the computation remains low, as all constraints can be cast as linear equalities and inequalities. Moreover, the average and maximum CPU times for the full scale of the network are only $2.14s$ and $8.37s$. This is acceptable in practice, considering that the worst-case calculations (occurring under overflow events) utilize less than 2 % of the sampling interval $T_{\text{NMPC}} = 10$ (min).

The numerical results carried out on the HiFi network show the feasibility of the proposed data-driven design and provide a basis for onward development. A key outcome of the system identification and control results is that the reduced physically-based SV-PDE flow model can be obtained based on water level measurements, moreover, the discharge predictions are accurately computed via the moving horizon parameter and state estimation.

7 Discussion

Our framework aims to allow operators at wastewater utilities to build internal models of the main transport lines and storage nodes in UDNs based on easy-accessible level measurements. Identifying the internal model parameters automatically from standard measurements is therefore one of our contributions. The proposed NMPC/MHE strategy has comparable performance as standard predictive control strategies reported in the literature, benchmarked with rule-based controllers. For instance, references [Joseph-Duran et al., 2014], [Joseph-Duran et al., 2015] report on a hybrid strategy where the internal MPC models exploit all available knowledge from the HiFi network of the UDN. As opposed to [Joseph-Duran et al., 2014], we report on the modularity of our approach, focusing on an internal model obtained by water level data.

Practical implementation of using the method includes the fact that water level sensors need to be deployed in the network to identify the transport dynamics between stations and the periodic household disturbances. Furthermore, our identification approach exploits knowledge about the high-level lay-

out of the network, which is typically available at water utilities. To carry out the experimental implementation of the work, a reliable mapping between rain intensity and the actual flow appearing in the system is needed. Besides, an implementation of a communication strategy is required, where the calculated flow references are translated to reference signals at the local pumping units.

8 Conclusions and future work

In this article, a new methodology for data-driven predictive control in urban drainage networks has been presented and tested. The proposed data-driven modeling approach is based on the physical characteristics of open-channel unpressurized flow, governed by the reduced Saint-Venant partial differential equations. A modified version of this model has been used for predicting the internal water levels in the sewer network, moreover to predict the discharged flows to the storage units. To update the model from data, level sensors have been distributed in manholes to enhance the internal prediction performance by taking into account periodic and non-periodic lateral inflows along the pipelines. Moving horizon parameter estimation has been proposed to overcome the inaccuracy issues, introduced by the linearization of the pipe geometries and the approximation of the reduced Saint-Venant partial differential equations. To overcome the problem of limited sensor measurements in the network, moving horizon state estimation has been proposed. The nominal nonlinear multi-objective optimization problem has been solved in a receding horizon fashion, along with the proposed state and parameter estimations. The performance of the proposed methodology has been successfully tested on a high-fidelity sewer network simulator with real rain and domestic wastewater inflow measurement data.

In future work, the methodology will be tested on urban drainage networks with different sizes and topologies. Moreover, it will be interesting to investigate the proposed method in different applications, e.g., stormwater collection networks. Also, an investigation into how rain and domestic wastewater uncertainties can be integrated with the current modeling and control methodology is a matter of future work.

Appendix

Simulation parameters

In this appendix we provide the numerical values of the control parameters, the constraint bounds and the main physical attributes of the HiFi simulation network, given in Table 3, Table 4 and Table 5, respectively.

Note that the values for the upper height constraints $\overline{h}$ are equivalent to the diameter of the pipes.

Attributes	Pipe g_1	Pipe g_2	Pipe g_3	Pipe g_4	Unit
Geometry	circular	circular	circular	circular	(-)
Diameter (d)	0.4	0.6	0.45	0.35	(m)
Slope (S_b)	0.03	0.05	0.02	0.02	(-)
Roughness (n)	0.013	0.013	0.013	0.013	(-)
Length (L)	0.9	0.45	2	2.4	(km)

Table 3: Network attributes for pipes.

Attributes	Pit s_1	Pit s_2	Pit s_3	Pit s_4	Unit
Constant (K_1)	21.5	30	30	43	(m^2)
Constant (K_2)	-	130	-	-	(m^2)
Volume ($\overline{V}$)	43	95	60	86	(m^3)
Pump flow ($\overline{u}$)	162	198	162	90	(m^3/h)

Table 4: Network attributes for pits.

T_{NMPC}	$T_{\text{on/off}}$	H_{p1}	H_{p2}	H_{pe}	H_{se}	T_{se}
10 (m)	1 (m)	2 (h)	22 (h)	48 (h)	48 (h)	10 (m)

Table 5: Simulation parameters.

The disturbance model uses $k = 2$ frequency terms and $\omega = 1$ (day) frequency for all pipes. Moreover, the disturbance signal scenarios are characterized by normally distributed, zero mean random uncertainty, where $\sigma^2_{s_1} = 30.6$, $\sigma^2_{s_2} = 5.4$, $\sigma^2_{s_3} = 27$ and $\sigma^2_{s_4} = 10.8$ (m^3/h). The lateral inflows along the gravity pipelines are all characterized by $\sigma^2_g = 15$ (m^3/h).

Acknowledgment

The authors would like to thank Rasmus Halvgaard from DHI Group for helping with setting up the MIKE 1D API to the MIKE URBAN simulation environment. The work of C. Ocampo-Martinez has been supported by the project PID2020-115905RB-C21 (L-BEST) funded by MCIN/ AEI /10.13039/50110001 1033. Furthermore, we thank the wastewater utility in Fredericia, Denmark for providing real flow data for our numerical experiment.

References

F. Allgöwer, T. A. Badgwell, J. S. Qin, J. B. Rawlings, and S. J. Wright. “Nonlinear Predictive Control and Moving Horizon Estimation — An Introductory Overview”. In *Advances in Control*, pages 391–449. Springer, London, 1999. doi:10.1007/978-1-4471-0853-5_19.

J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. "CasADi: a software framework for nonlinear optimization and optimal control". *Mathematical Programming Computation*, 11(1):1–36, 2019. ISSN 18672957. doi:10.1007/s12532-018-0139-4.

K. M. Balla, C. S. Kallesøe, C. Schou, and J. D. Bendtsen. "Nonlinear grey-box identification with inflow decoupling in gravity sewers". *IFAC-PapersOnLine*, 53 (2):1065–1070, 2020a. ISSN 24058963. doi:10.1016/j.ifacol.2020.12.1295.

K. M. Balla, C. Schou, J. Dimon Bendtsen, and C. S. Kallesøe. "Multi-scenario model predictive control of combined sewer overflows in urban drainage networks". In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1042–1047, Montréal, 2020b. IEEE. ISBN 9781728171401. doi:10.1109/CCTA41146.2020.9206362.

D. Butler and J. W. Davies. *Urban Drainage*. Spon Press, 2006. doi:10.1016/s1462-0758(00)00017-0.

F. J. Chang, J. M. Liang, and Y. C. Chen. "Flood forecasting using radial basis function neural networks". *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 31(4):530–535, 2001. ISSN 10946977. doi:10.1109/5326.983936.

V. Dalmas, G. Robert, G. Besançon, and D. Georges. "Simplified Non-Uniform Models for Various Flow Configurations in Open Channels". *IFAC-PapersOnLine*, 50(1):12320–12325, 2017. ISSN 24058963. doi:10.1016/j.ifacol.2017.08.2159.

S. Dey. "Free overfall in open channels: State-of-the-art review". *Flow Measurement and Instrumentation*, 13(5-6):247–264, 2002. ISSN 09555986. doi:10.1016/S0955-5986(02)00055-9.

DHI. "MIKE for developers". Technical report. URL: `https://docs.mikepoweredbydhi.com/engine_libraries/mike1d/mike1d_api/`.

DHI. "Collection System - Modelling of storm water drainage networks and sewer collection systems". Technical report, Hørsholm, 2017.

DHI. "MIKE 1D: DHI Simulation Engine for 1D river and urban modelling - Reference Manual". Technical report, DHI, 2019.

L. García, J. Barreiro-Gomez, E. Escobar, D. Téllez, N. Quijano, and C. Ocampo-Martinez. "Modeling and real-time control of urban drainage systems: A review". *Advances in Water Resources*, 85:120–132, 2015. ISSN 03091708. doi:10.1016/j.advwatres.2015.08.007.

M. S. Gelormino and N. L. Ricker. "Model-predictive control of a combined sewer system". *International Journal of Control*, 59(3):793–816, 1994. ISSN 13665820. doi:10.1080/00207179408923105.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Output-feedback model predictive control of sewer networks through moving horizon estimation". In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 1061–1066, Los Angeles, 2014. doi:10.1109/CDC.2014.7039522.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Output-feedback control of combined sewer networks through receding horizon control with moving horizon estimation". *Water Resources Research*, 51(10):8129–8145, 2015. ISSN 19447973. doi:10.1002/2014WR016696.

C. S. Kallesøe and T. Knudsen. "Self calibrating flow estimation in waste water pumping stations". In *2016 European Control Conference, ECC 2016*, pages 55–60, 2016. ISBN 9781509025916. doi:10.1109/ECC.2016.7810263.

X. Litrico and V. Fromion. "Boundary control of linearized Saint-Venant equations oscillating modes". *Automatica*, 42:967–972, 2006. ISSN 00051098. doi:10.1016/j.automatica.2006.02.002.

A. M. D. Livera, R. J. Hyndman, and R. D. Snyder. "Forecasting time series with complex seasonal patterns using exponential smoothing". *Journal of the American Statistical Association*, 106(496):1513–1527, 2011. doi:10.1198/jasa.2011.tm09771.

R. Löwe, S. Thorndahl, P. S. Mikkelsen, M. R. Rasmussen, and H. Madsen. "Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar". *Journal of Hydrology*, 512:397–407, 2014. ISSN 00221694. doi:10.1016/j.jhydrol.2014.03.027.

R. Löwe, L. Vezzaro, P. S. Mikkelsen, M. Grum, and H. Madsen. "Probabilistic runoff volume forecasting in risk-based optimization for RTC of urban drainage systems". *Environmental Modelling and Software*, 80:143–158, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.02.027.

N. S. V. Lund, A. K. V. Falk, M. Borup, H. Madsen, and P. Steen Mikkelsen. "Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management". *Critical Reviews in Environmental Science and Technology*, 48(3):279–339, 2018. ISSN 15476537. doi:10.1080/10643389.2018.1455484.

J. Ma, W. Sun, G. Yang, and D. Zhang. "Hydrological Analysis Using Satellite Remote Sensing Big Data and CREST Model". *IEEE Open Access Journals and Conferences*, 6:9006–9016, 2018. ISSN 21693536. doi:10.1109/ACCESS.2018.2810252.

L. W. Mays. *Stormwater Collection Systems Design Handbook*. McGraw-Hill Education, 1st edition, 2001. ISBN 0071354719.

A. L. Mollerup, P. S. Mikkelsen, and G. Sin. "A methodological approach to the design of optimising control strategies for sewer systems". *Environmental Modelling and Software*, 83:103–115, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.05.004.

S. Munier, X. Litrico, G. Belaud, and P. O. Malaterre. "Distributed approximation of open-channel flow routing accounting for backwater effects". *Advances in Water Resources*, 31(12):1590–1602, 2008. ISSN 03091708. doi:10.1016/j.advwatres.2008.07.007.

C. Ocampo-Martinez. *Model Predictive Control of Wastewater Systems*. Springer, Barcelona, 1st edition, 2010. ISBN 978-1-84996-352-7. doi:10.1007/978-1-84996-353-4.

C. Ocampo-Martinez, V. Puig, G. Cembrano, and J. Quevedo. "Application of predictive control strategies to the management of complex networks in the urban water cycle". *IEEE Control Systems*, 33(1):15–41, 2013. ISSN 1941000X. doi:10.1109/MCS.2012.2225919.

J. A. Roberson and C. T. Crowe. *Engineering Fluid Mechanics*. Houghton Mifflin Company, Boston, 5th edition, 1993. ISBN 0-395-66161-7. doi:10.1007/978-1-4020-6742-6.

M. R. Schütze, D. Butler, and M. B. Beck. *Modelling, Simulation and Control of Urban Wastewater Systems*. Springer, 2002. doi:10.1007/978-1-4471-0157-4.

X. Tian, R. R. Negenborn, P. J. van Overloop, J. María Maestre, A. Sadowska, and N. van de Giesen. "Efficient multi-scenario Model Predictive Control for water resources management with ensemble streamflow forecasts". *Advances in Water Resources*, 109:58–68, 2017. ISSN 03091708. doi:10.1016/j.advwatres.2017.08.015.

S. C. Troutman, N. Schambach, N. G. Love, and B. Kerkez. "An automated toolchain for the data-driven and dynamical modeling of combined sewer systems". *Water Research*, 126:88–100, 2017. ISSN 18792448. doi:10.1016/j.watres.2017.08.065.

K.-J. van Heeringen, R. van Nooijen, K. Kooij, and B. Postma. "Real-time Control of sewer pumps by using ControlNEXT to smooth inflow at Waste Water Treatment Plant Garmerwolde". In *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, pages EPSC2016–1412, Apr. 2016.

A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". *Mathematical Programming*, 106(1):25–57, 2006. ISSN 00255610. doi:10.1007/s10107-004-0559-y.

A. Wills and B. Ninness. "On gradient-based search for multivariable system estimates". *IEEE Transactions on Automatic Control*, 53(1):298–306, 2008. ISSN 00189286. doi:10.1109/TAC.2007.914953.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow". *Advances in Water Resources*, 34:282–290, 2011. ISSN 03091708. doi:10.1016/j.advwatres.2010.11.009.

M. Xu, R. R. Negenborn, P. J. van Overloop, and N. C. van de Giesen. "De Saint-Venant equations-based model assessment in model predictive control of open channel flow". *Advances in Water Resources*, 49:37–45, 2012. ISSN 03091708. doi:10.1016/j.advwatres.2012.07.004.

Y. Zou, L. Cen, D. Li, and X. He. "Simplified state-space model and validation of irrigation canal systems". In *Chinese Control Conference, CCC*, pages 2002–2007, Hangzhou, 2015. ISBN 9789881563897. doi:10.1109/ChiCC.2015.7259938.

Paper D

Multi-scenario Model Predictive Control of Combined Sewer Overflows in Urban Drainage Networks

Krisztian Mark Balla[a,b], Christian Schou[a] and Jan Dimon Bendtsen[b] Carsten Skovmose Kallesøe[a,b]

[a]Grundfos Holding A/S, Poul Due Jensens Vej 7, DK-8850, Bjerringbro, Denmark
[b]Section for Automation and Control, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark

Abstract—*Urban drainage networks (UDN) are among the most vital infrastructures within the natural water cycle. The most widely applied Real Time Control (RTC) on these systems is Model Predictive Control (MPC), which typically incorporates transport time delays and the effect of disturbances explicitly in the objectives and constraints. One of the greatest challenges in the control of UDNs is to formulate multiple control criteria regarding operational requirements of the network. Furthermore, MPC faces the challenge of handling uncertainty caused by disturbances, e.g. weather predictions. One way to incorporate the uncertainty in the decision making is to consider multiple scenarios, i.e. to generate different ensembles based on rain forecasts. To this end, we propose a Multi-scenario MPC (MS-MPC) approach, that deals with uncertainty in the expected inflow. First, a generic multi-objective MPC is established which deals with the time delays explicitly in the optimization. Then, this framework is extended to our formulation of the multiple scenario problem. The algorithm is verified through a case study by interfacing a high-fidelity simulator model of a sewer network as virtual reality.*

Keywords—*Multi-scenario MPC; Combined sewer; Forecast uncertainty*

Published in the *Proceedings of Conference on Control Technology and Applications*, 2020.

DOI: https://doi.org/10.1109/CCTA41146.2020.9206362

Contents

1 Introduction

Combined sewers carry domestic wastewater and rain runoff towards treatment plants, where the sewage is treated before it is discharged to the environment [Schütze et al., 2002]. Real-Time Control (RTC) of these networks is a challenging task since the system is characterized by large-scale dimensions, nonlinear dynamics, and significant time-delays. Besides, UDNs are increasingly being pushed to their capacity limits due to changing weather conditions, resulting in increased amounts and more frequent Combined Sewer Overflows (CSO).

Constrained optimal control has been done in several works, mainly considering MPC. Due to the complexity and the large-scale nature of drainage networks, typically conceptualized control models are used, considering the available network volumes [García et al., 2015]. In [Puig et al., 2009] and [Schütze et al., 2002], the volumetric storage of pipes, manholes, and retention tanks have been collectively modeled, while in [Gillé et al., 2008] and [Balla et al., 2020], simplified hydraulic models were proposed, considering gravity-driven sewage pipes as simple delay elements without storage. In [Gelormino and Ricker, 1994] and [Fiorelli and Schutz, 2009], the overflows have been conceptualized by introducing an artificial variable, indicating the average overflow over a specific horizon. Extending these previous frameworks, [Halvgaard and Falk, 2017] and [Halvgaard et al., 2017] used an indicator variable which was forecasting overflow only in case of an actual tank overflow. In [Madsen et al., 2018], this previously-established, fast-solvable optimisation model has been successfully utilized in a simulation study, representing a real large-scale drainage network in Denmark.

The control problem in UDNs consists of multiple criteria. For instance, [Fiorelli and Schutz, 2009] investigated various ways of weighting control objectives with regard to different rain conditions. The study in [Mollerup et al., 2016] proposed a systematic control design and focused on the multi-objective control performance regarding the choice of optimisation variables and the formulation of the objective function.

Nonetheless, the majority of the research reporting on MPC of UDNs investigates the performance by considering historical disturbances, i.e. historical rain data. Works on model-based optimization, taking into account the uncertainties, are relatively few. Runoff forecast uncertainties in risk-based optimization have been considered by using stochastic grey-box models in [Löwe et al., 2016] and in [Vezzaro and Grum, 2014]. In [Courdent et al., 2015], an optimization framework has been introduced, which considered the estimated uncertainty of rain runoff forecasts, thereby estimating the risk of overflows based on the stored volumes in the system. This framework has used an optimization strategy with a simplified model, while the transport times have not been considered between pumping stations.

Another way to consider stochastic hydrological processes in optimization is to assume possible scenarios, estimate their likelihood and test the optimization under these assumptions. A flow control problem has been studied

in [Tian et al., 2017], where a Multi-Scenario (MS-MPC) approach has been implemented on a simulation model of a dutch canal system [van Overloop et al., 2008]. In [Grosso et al., 2017], a chance-constrained, tree-based and multi-scenario stochastic MPC approaches have been compared and applied to drinking water networks.

In the present paper, an MS-MPC approach is applied to a high-fidelity simulator model of a UDN, considering a simplified representation of the network. In contrast to [Vezzaro and Grum, 2014], we implement a fast-solvable MPC strategy that considers the network delays in terms of the transport flows between pumping stations. Furthermore, we extend the work in [Puig et al., 2009] by evaluating the performance of MS-MPC, considering multiple operational and control objectives. We combine the results of [Mollerup et al., 2016], where the operational objectives and the tuning of the optimization parameters have been analyzed.

The remainder of the paper is structured as follows: In Section 2, the preliminary introduction of UDNs and the simulation network are presented. Section 3 reviews the simplified network models, whereupon Section 4 introduces the generic MPC and the proposed MS-MPC control approaches. In Section 5 numerical results and the applied scenarios are presented. Finally, Section 6 provides conclusions and sums up the contributions of the work.

1.1 Nomenclature

Throughout the paper, all quantities mentioned are real. Boldface letters are used for sets, such as $\mathbf{s} = \{s_1, ...s_n\}$ as well as for vectors $\mathbf{x} = [x_1, ...x_n]^T \in \mathbb{R}^n$. Time dependent variables are denoted by $x(t)$ or $x(t_k)$, where $t \in \mathbb{R}^+$ and $t_k \in \mathbb{Z}^+$ are the continuous and discrete time variables, respectively.

2 Drainage Systems

Urban drainage systems typically consist of storage elements such as gravity pipes, retention tanks, catchment areas and one or several outlet points leading to the treatment plants. The most common actuators in these networks are pumps and gates. In the present work, networks with multiple retention tanks are considered, where the stored sewage volumes are controlled by pumps. Hence, the regulated variable is flow, provided by local, variable-speed pumps.

In order to make closed-loop control, a high-fidelity model is used in the MIKE URBAN[1] (MU) simulation environment. The network model is shown in Figure 1.

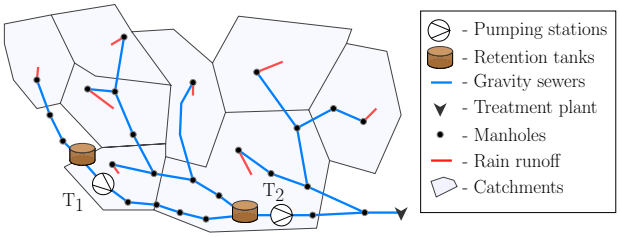


Figure 1: Schematics of the high-fidelity simulator in MIKE URBAN.

The network consists of two pumping stations, equipped with retention tanks with a total storage capacity of about $30[m^3]$. The pumps are operated by local PID controllers. There is one outlet point representing the treatment plant and several catchment areas, where rainfall runs off and enters the system through manholes. The disturbances considered here are domestic sewage and rain infiltration. In the network, rainfall run-off flow enters the network through eight inlet points, distributed over the entire network.

3 Network Model

3.1 Gravity sewers

Gravity-driven flow in sewage pipes can be computed accurately by the well-known Saint-Venant partial differential equations [Schütze et al., 2002]. Due to their computation burden and complexity, these equations are not well-suited for large-scale RTC applications. Instead, similarly to [Gillé et al., 2008] and [Gelormino and Ricker, 1994], the pipes are modelled as pure delay elements.

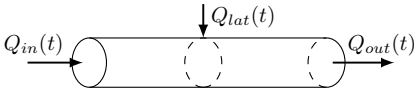


Figure 2: Delay translation model.

Hence, outflows from a gravity pipe section are the delayed sums of controlled pump flows and uncontrolled lateral inflows, as shown in Figure 2. (Lateral inflows are additional flows that enter the pipelines along the length of the channel.)

[1]MIKE URBAN is a standard hydraulic simulation and modeling tool, used by operators at many water utilities. The MU simulation environment solves the full dynamic Saint-Venant equations for open-channel flow [DHI, 2016].

The mass balance relation at time t is formulated as follows:

$$Q_{out}(t) = Q_{in}(t-\tau) + Q_{lat}(t-\tau_{lat}) \tag{1}$$

where $\tau \in \mathbb{R}_+$ and $\tau_{lat} \in \mathbb{R}_+$ are time lags measured from the upstream and from the point where lateral flows enter the pipeline, respectively. After discretization, delays are defined in δt sampling steps, hence the delayed flow is modeled with an augmented state vector consisting of the previous flows. The state equation, assuming $Q_{lat} = 0$ (to ease the notation), is given by:

$$\begin{bmatrix} Q_{out}(t+\delta t) \\ Q_{in}(t-\tau+2\delta t) \\ \vdots \\ \vdots \\ Q_{in}(t) \end{bmatrix} = \mathbf{A} \begin{bmatrix} Q_{out}(t) \\ Q_{in}(t-\tau+\delta t) \\ Q_{in}(t-\tau+2\delta t) \\ \vdots \\ Q_{in}(t-\delta t) \end{bmatrix} + \mathbf{B}_u Q_{in}(t) \tag{2}$$

where Q_{in} inlet flow is subject to control, Q_{out} discharged flow is the output and the system matrices $\mathbf{A}$ and $\mathbf{B}_u$ are given by:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 1 \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{\tau \times \tau}, \quad \mathbf{B}_u = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{\tau}. \tag{3}$$

Note, that in case there are Q_{lat} inflows, the augmented state vectors are stacked together. This simple delay translation model is considered computationally beneficial and realistic enough for system-wide optimization, even though the physical phenomena such as flow attenuation and backwater effect are not incorporated in this formulation.

3.2 Retention tanks

Storage within the network is modeled by conceptual tanks that can account for overflows, as shown in Figure 3.

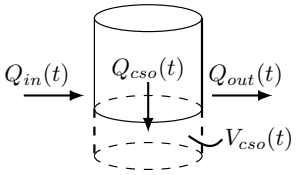


Figure 3: Linear retention tank with V_{CSO} virtual overflow volume.

Flows to retention tanks (Q_{in}) are considered as (i) forecasted disturbances and (ii) controlled flows, coming from an upstream pumping station. The manipulated flow variables are denoted with Q_{out}, furthermore V represents

the stored volume in the tank. The mass balance for each tank is:

$$\frac{dV(t)}{dt} = \sum_{n=1}^{N} Q_{in,n}(t) - \sum_{m=1}^{M} Q_{out,m}(t), \tag{4}$$

where N and M are the number of inlet points and number of pumps, respectively. The translation between volume and level is done by using the constant cross section area A.

In order to model overflows, the formulation in eq.(4) is extended with a virtual volume, similarly as done in [Halvgaard and Falk, 2017] and [Halvgaard et al., 2017]. Hence, as depicted in Figure 3, the storage model considers two joint volume elements:

$$V_t(t) \triangleq V(t) + V_{cso}(t), \tag{5}$$

where V is the physical volume of fluid and V_{cso} is the virtual volume accounting for overflows. To keep track of the physical volumes and to trigger an overflow in the tanks at the time when the physical limits are exceeded, the following restrictions apply to the storage model:

$$A\underline{h} + V_{cso}(t) \leq V_t(t) \leq A\overline{h} + V_{cso}(t), \quad \forall t_k = 1, ..., T, \tag{6}$$

where $\underline{h}$ and $\overline{h}$ are the physical lower and upper level bounds respectively and $V = Ah$. In case of an overflow event, V_{cso} increases both the lower and upper bounds, thereby keeping track of the physical storage and moving the overflow volume into the virtual storage at the bottom of the tank. Furthermore, the excess water leaves the system immediately. We assure this by letting $V_{cso} \geq 0$, meaning that the spilled sewage Q_{cso} spills to the environment, thus never flowing back to the retention tanks.

4 Predictive Control

The terminology used in MPC of UDNs often differs in the literature coming from different backgrounds. For clarity, all the considered variables of UDNs are assigned to control-oriented variables, summarized in Table 1.

Type of variable	Related symbols
System states (x)	V or equivalently h
Virtual states (z)	V_{cso} or equivalently h_{cso}
Control input (u)	Q_{out}
Disturbance (d)	Q_{in}
Output (y)	Q_W

Table 1

Note, that the term disturbance represents the rain-runoff and domestic wastewater entering the network.

4.1 Multi-criteria MPC

In the optimization problem, $Q_{out,i}$ is considered as the decision variable, denoting the pumped flows of the i^{th} pumping station. The physical system states are sewage levels h_j corresponding to the j^{th} retention tank. In this study, similarly to [Puig et al., 2009], the objective function is formulated as a linearly weighted sum. The optimization is given by:

$$\min_{Q_{out}(0),\ldots,Q_{out}(H_p)} \mathcal{L}(h, Q_{out}, Q_{in}, t_k) \triangleq \sum_{t_k=0}^{H_p-1} \sum_{j=1}^{\Gamma} \lambda_j \mathcal{F}_j(t_k), \tag{7}$$

where λ_j denotes the scaling weights, H_p is the prediction horizon, t_k is the discrete time index, and Γ is the number of control objectives. Note, that we write h tank levels as the system states, for the reason that levels are directly measurable in real life.

The first two terms $\mathcal{F}_1$ and $\mathcal{F}_2$ stand for overflow avoidance and tank emptying, respectively:

$$\mathcal{F}_1(t_k) \triangleq \sum_{i=1}^{P} V_{cso,i}(t_k)^2 \quad \text{and} \quad \mathcal{F}_2(t_k) \triangleq \sum_{i=1}^{P} V(t_k)^2, \tag{8}$$

where P is the number of overflow elements, i.e. retention tanks. Recall, that the overflow indicator $V_{cso} \geq 0$ is used to keep track of the water running out of the storage volume, as described in eq.(6). Due to the fact that these physical level boundaries never decrease, V_{cso} has to be reset each time when the problem is resolved over H_p. The weights corresponding to overflows λ_1 are chosen to be significantly higher than the cost of other terms, making the usage of overflows undesirable if possible. Furthermore, we introduce $\mathcal{F}_2$ objective, as emptying the tanks is necessary to avoid odor problems occurring due to long retention times. Moreover, the weights on $\mathcal{F}_2$ allow to include the filling sensitivity of retention tanks, meaning that sensitive tanks are filled slower and emptied faster than less sensitive tanks.

The third objective $\mathcal{F}_3$ stands for minimizing the flow variation of the sewage leading to the treatment plant:

$$\mathcal{F}_3(t_k) \triangleq \left(Q_W(t_k) - \frac{1}{H_p} \sum_{j=0}^{H_p-1} Q_W(j) \right)^2, \tag{9}$$

where Q_W is the sum of controlled and disturbance flows leading to the treatment plant. Furthermore, the second term in eq.(9) is considered as a reference flow, determined by the mean of the H_p-step ahead outlet flows towards the treatment plant. This formulation is inspired by [Nielsen et al., 2020], where the inlet flow variations to the treatment plant has been minimized over a daily horizon, assuming dry-weather conditions.

The fourth sub-goal $\mathcal{F}_4$ relates to the operation of actuators, where the control action is minimized. Hence, $\mathcal{F}_4$ is:

$$\mathcal{F}_4(t_k) \triangleq \sum_{l=1}^{L} Q_{out,l}(t_k)^2, \tag{10}$$

where L is the number of pumping stations and Q_{out} is the accumulated outflow, provided by pumps at the l^{th} station.

The optimization problem in eq.(7) is formulated as a linear program, subject to the flow delays in Equation eq.(1) and to the discretized tank dynamics in eq.(4). Furthermore, the equality constraint introduced in eq.(5) and inequality constraint in eq.(6) apply to the tank model, where V_{cso} is used as a virtual state in the optimization problem. Additionally, the control problem is subject to operational and physical constraints in the form:

$$\underline{Q}_{out} \leq Q_{out}(t_k) \leq \overline{Q}_{out}, \quad \forall t_k = 1, ..., T_k, \tag{11}$$

where $\underline{Q}_{out}$ and $\overline{Q}_{out}$ are the physical lower and upper bounds of the accumulated pump flows, respectively. Moreover, the rate of change of the control variables Q_{out} are constrained, in order to avoid deterioration of the pumps and pressure shocks in the following pressurized rising mains:

$$|Q_{out}(t_{k+1}) - Q_{out}(t_k)| \leq \overline{\Delta Q}_{out}, \quad \forall t_k = 1, ..., T_k, \tag{12}$$

where $\overline{\Delta Q}_{out}$ is the maximum allowed control input change, defined respectively for each pump. The operational constraint regarding the maximum inflow capacity of the treatment plant reads as follows:

$$Q_W(t_k) \leq \overline{Q}_W, \quad \forall t_k = 1, ..., T_k, \tag{13}$$

where $\overline{Q}_W$ is the maximum flow to the treatment plant.

4.2 Multi-Scenario MPC

Disturbances within an urban drainage framework include rainfall precipitation, groundwater infiltration, and domestic household sewage, among which rainfall is a stochastic hydrological process. The usage of various forecasting methods for rainfall infiltration, e.g. numerical weather predictions or radar rainfall estimates [Courdent et al., 2015], implies that uncertainty is implicitly involved in the control of UDNs. For that reason, we extend the generic Multi-Criteria (MC-MPC) formulation and approximate the solution of the stochastic optimization problem with a Multi-Scenario (MS-MPC) approach. The control is then obtained by taking into account several forecasts, thereby making the decision making more robust towards weather prediction inaccuracies. To translate rainfall intensities to runoff flows, the catchment dynamics in the MIKE URBAN runoff environment are utilized. This engine makes several

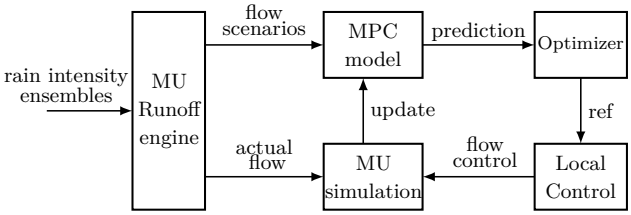


Figure 4: Structure of the implemented MS-MPC approach.

realizations of disturbance inflows based on forecasted rainfall intensities. The hierarchical structure of such control scheme is shown in Figure 4.
The MU runoff engine incorporates the dynamics of catchments and produces a surface runoff hydrograph in response to a rain event, similarly as shown in Figure 5 below.

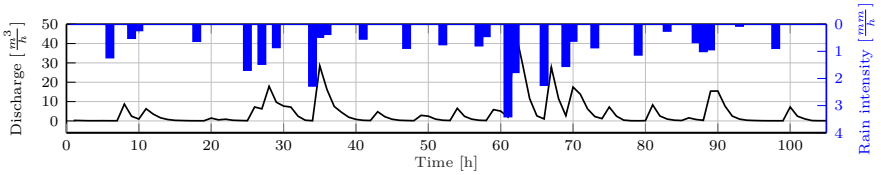


Figure 5: Rain run-off computed by the MU engine. The reversed y-axis on the right denotes the intensity of rain, sampled at an hourly rate.

The optimization is reformulated such that the objectives of all scenarios are summed and weighted by the likelihood of occurrences. Hence, the MS-MPC can be recast as:

$$\min_{Q_{out}(0),\ldots,Q_{out}(H_p)} \sum_{j=1}^{N_s} p_j \mathcal{L}_j(h_j, Q_{out}, Q_{in,j}, t_k), \tag{14}$$

where the subscript j represents the j^{th} scenario, whereas p_j is the likelihood of occurrence. Moreover, $N_s \in \mathbb{Z}_+$ represents the number of scenarios used in the optimization. Note, that the cost functions differ in each scenario, as the different meteorological disturbances create different h_j future trajectories. Thus, there are dynamic and inequality constraints devoted to each scenario. To solve the MS-MPC problem, a common control Q_{out} is computed, which attempts to find the best decision for the most likely future states and prepare the system for possible worst-case events. For solving the problem, CVX is used [Grant and Boyd, 2013] with the *SeDuMi* solver [Sturm, 1999].

An issue with the above formulation occurs in the case when a single ensemble does not predict rain at a certain time t_k, while the rest of the forecasts imply that there is a future storm event for which the system has to be prepared. In this case, the optimizer should act based on the likelihood

of the events. However, the hard inequality constraint formulated in eq.(6), devoted to the no-rain scenario, does not allow to increase the control action, i.e. $Q_{out}(0), ..., Q_{out}(H_p)$. This is also the case if one scenario has significantly smaller likelihood than any of the others, but requires to decrease the lower storage bounds $\underline{V}$ in order to increase the common control actions. We solve this problem by inserting a slack variable into the hard inequality constraint such that:

$$A\underline{h} + V_{cso}(t_k) - \epsilon(t_k) \leq Ah_t(t_k) \leq A\overline{h} + V_{cso}(t_k), \tag{15}$$

$\forall t_k = 1, ..., T_k$, where $\epsilon \geq 0$. The slack variable is penalized and it is activated only when the likelihood of a no-rain event weights significantly less than ensembles predicting overflow. Hence, we avoid that an unlikely no-rain scenario restricts the usage of control actions when likely scenarios require to empty retention tanks due to heavy loads on the system.

5 Numerical results

The performance of the MS-MPC algorithm is assessed based on the high-fidelity model shown in Figure 1. The control algorithm is tested against $N_s = 4$ different weather scenarios, covering a six days long wet-weather period. The test scenarios have been created based on rainfall intensities corresponding to realistic design storm events.[2]

It should be noted, that we do not aim to show how precisely the future is forecasted. Instead, the goal is to present how plausible future forecasts are embedded in a standard MPC problem. The combined run-off and domestic wastewater replicates are depicted in Figure 6.

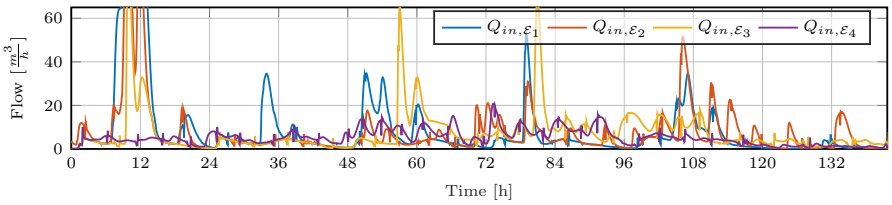


Figure 6: Inflow scenarios computed by MU, using ensemble rain forecasts.

The signals $Q_{in,\mathcal{E}_1}, ..., Q_{in,\mathcal{E}_4}$ represent four scenarios. The length of the simulation is $T = 6$ days.

[2]The data is from https://www.silkeborg-vejret.dk/statistik.php. We deliberately chose rainy periods between 21-27 April, 2018, from Silkeborg, Denmark. The domestic wastewater inflow is artificially created and scaled.

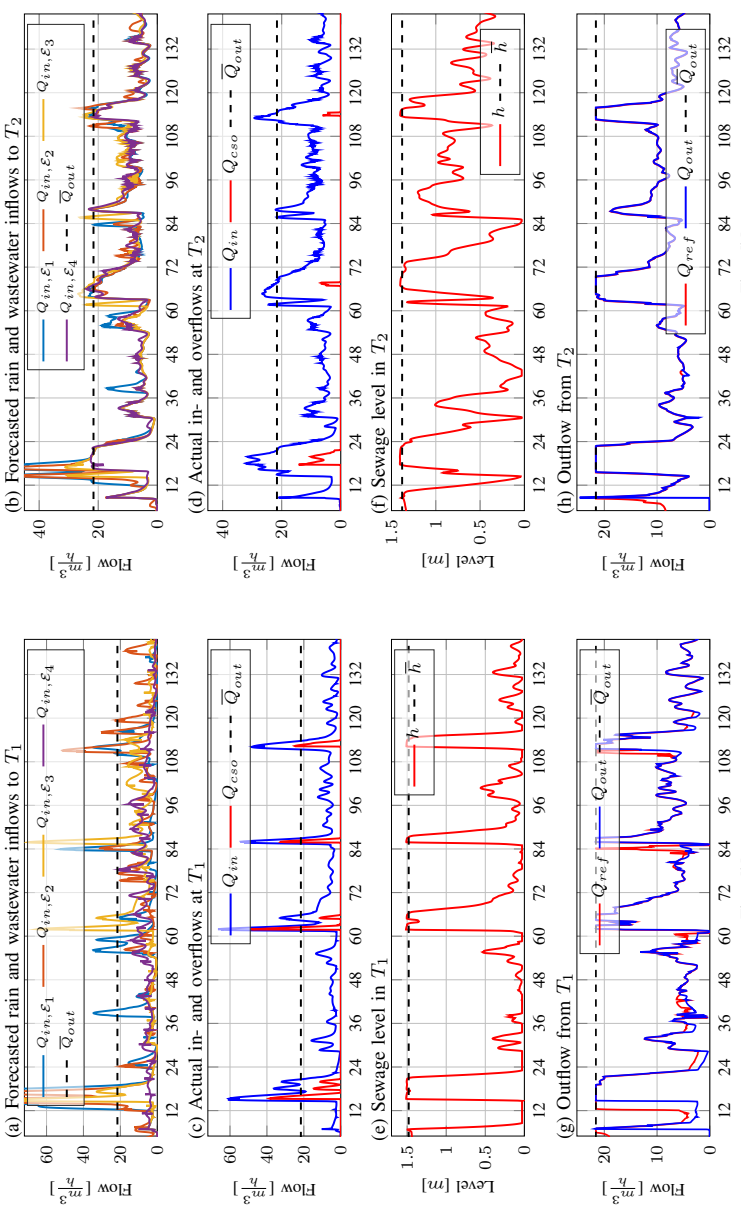


Figure 7: Performance of the MS-MPC with equal likelihoods of four rain scenarios over a six days simulation period with $H_p = 2$ [h] and $\delta t = 5$ [min]. Results are shown in two columns, where signals corresponding to the upstream (T_1) and to the downstream (T_2) pumping stations are shown in the first and second columns, respectively. The plots (a) and (b) show possible inflow sequences, (c) and (d) the actual flows occurring in the system, (e) and (f) the current level in the retention tanks and finally, (g) and (h) the optimal flow set-points Q_{ref} and the actual control flows Q_{out} at the pumping stations.

The longest travel time within the network is related to the connection between T_1 and T_2 stations and is approximately 90 minutes. An $H_p = 2$ [h] prediction horizon is used, with a sampling time and control step $\delta t = 5$ [min]. Moreover, the likelihoods $p_j = 0.25, \forall j = 1, ..., 4$ are set equal. The weight parameters λ_1 corresponding to CSO prevention are equal for both T_1 and T_2 retention tanks, hence we do not prioritize overflows of one tank over another. (The optimization parameters are listed in the appendix.) The test results are shown in Figure 7.

The MS-MPC acts as an upper-level controller that computes optimal set-points (Q_{ref}) to local PID controllers. The signals in Figure 7(c) and Figure 7(d) show the actual accumulated inflows and the amount of overflows at T_1 and at T_2 stations. The control can account for the uncertainties in case the actual disturbances are close or within the range of the possible scenarios. For instance, as seen between 108 and 120 [h], T_2 overflows since the inflow exceeds the pumping capacities more than it was forecasted. Besides, Figure 7(c) and Figure 7(d) show overflows which could not be prevented due to the insufficient storage and pumping capacity of the network. Between 60 and 72 [h], the pumps at T_1 decrease the flow instead of further emptying the tank. This is for the reason that the delayed pumped sewage would arrive at T_2 in a high-inflow period, causing heavier overflows at the downstream.

Moreover, with the proposed method, overflows can be prioritized using λ_1 corresponding to eq.(8). The CSO reductions in two extreme cases are shown in Figure 8.

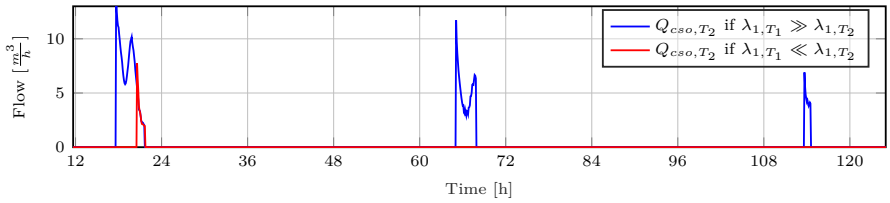


Figure 8: CSO reduction at T_2 in case overflows are prioritized to T_1.

To protect T_2, the pumps at T_1 can hold the sewage back in heavy-load periods. Nevertheless, in the proposed framework, it should not be expected to obtain a universal solution that is optimal for all ensemble weather predictions, especially if there are conflicting objectives. Around 36 [h], for instance, the first ensemble predicts a potential overflow shown in Figure 7(c), for which the pumps try to react by keeping the retention level at a minimum. Even though there is a nearly rain-free period during this time, the flow references to the pumps (red curves in Figure 7(g) and Figure 7(h)) indicate that the MS-MPC attempts to prepare the system for potential overflows. The same situation is observed at 84 [h], where a coupled rain event is expected earlier than it happens.

In addition to CSO prevention, the smoothing of the inflow to the treatment plant has been considered. This has been done by compensating the variances on the disturbance flows leading to the treatment plant over the prediction horizon. The performance is shown in Figure 9.

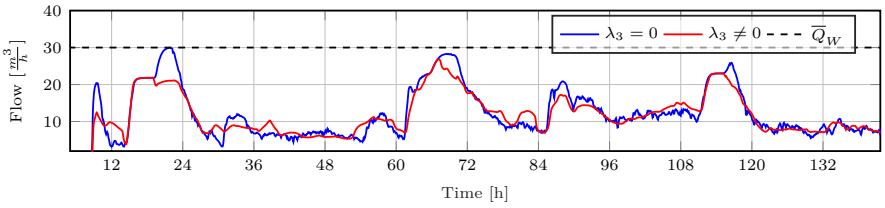


Figure 9: Inlet flow (Q_W) to the treatment plant.

where $\overline{Q}_W$ is the maximum possible inlet flow capacity of the treatment plant and λ_3 is the weighting coefficient from eq.(9). As seen, in contrast to maximizing the inlet flow towards the treatment plant, $\mathcal{F}_3$ attempts to even out the variations. The pumps at T_2 station need to consider the disturbances entering directly to the treatment plant and attempt to compensate for them. The peak flows in Figure 9. correspond to periods when the overflow risk is high, therefore the pumps at T_2 move as much sewage as possible to avoid CSOs. This is for the reason that $\mathcal{F}_1$ and $\mathcal{F}_3$ are conflicting objectives and $\mathcal{F}_1$ is prioritized over the flow smoothing to the treatment plant. Hence, the highest potential for improving the quality of the treatment process is in dry-weather periods when the risk of overflooding is low.

6 Conclusion

The presented paper studied how the hydrological uncertainties can be tackled in an RTC problem regarding the control of urban drainage networks. To this end, we proposed a Multi-Scenario approach as an extension of a standard, fast-solvable MPC framework. The method has been tested on a high-fidelity model of a test network and the implementation showed that both the simplified delay and retention tank models are feasible for on-line storage capacity optimization in UDNs. The MS-MPC has been tested on four different scenarios and the results showed that MS-MPC has a significant advantage over standard MPC methods that neglect uncertain weather forecasts. Although some scenarios can have a low probability of occurrence, the damage may be very high. Moreover, the results showed that the transport delays affect the MPC performance significantly, especially when prioritizing overflows and protecting sensitive waters.

In our future work, we focus on developing a systematic way of tuning the MS-MPC parameters, including the analysis of H_p size and the penalty gains on the objectives. Note, that the weights λ have been chosen based on pre-defined

performance goals in a heuristic fashion. Furthermore, a natural extension of the treatment plant flow variation objective $\mathcal{F}_3$, formulated in eq.(9), is to extend the prediction horizon to a daily scale, where varying sampling times are used. This allows for better optimization in dry-weather periods, where only domestic wastewater is considered with an inherent periodicity of one day.

Appendix

The optimization parameters are shown in Table 2.

$\lambda_{1,T_1} = \lambda_{1,T_2} = 10^5[-]$	$\overline{h}_{T_1} = 1.5[m]$
$\lambda_{2,T_1} = \lambda_{2,T_2} = 10^3[-]$	$\overline{h}_{T_2} = 1.4[m]$
$\lambda_{4,T_1} = \lambda_{4,T_2} = 1[-]$	$\lambda_3 = 10^2[m]$
$\overline{Q}_{out,T_1} = \overline{Q}_{out,T_2} = 21.6[\frac{m^3}{h}]$	$N_s = 4[-]$
$\overline{\Delta Q}_{out,T_1} = \overline{\Delta Q}_{out,T_2} = 10[\frac{m^3}{h}]$	$\overline{Q}_W = 30[\frac{m^3}{h}]$

Table 2: Optimization parameters.

Acknowledgement

The authors would like to thank Rasmus Halvgaard from DHI Group for the discussions and for the help with interfacing the MIKE URBAN simulation software. The project was supported by Innovation Fund Denmark (Ref. *9065-00018A*).

References

K. M. Balla, C. S. Kallesøe, C. Schou, and J. D. Bendtsen. "Nonlinear grey-box identification with inflow decoupling in gravity sewers". *IFAC-PapersOnLine*, 53 (2):1065–1070, 2020. ISSN 24058963. doi:10.1016/j.ifacol.2020.12.1295.

V. Courdent, L. Vezzaro, P. S. Mikkelsen, A. L. Mollerup, and M. Grum. "Using ensemble weather forecast in a risk based real time optimization of urban drainage systems". *Houille Blanche*, 2:101–107, 2015. ISSN 00186368. doi:10.1051/lhb/20150025.

DHI. "MIKE URBAN: Collection System Modelling of storm water drainage networks and sewer collection systems", 2016.

D. Fiorelli and G. Schutz. "Real-time control of a sewer network using a multi-goal objective function". In *the 17th Mediterranean Conference on Control & Automation*, pages 676–681, Thessaloniki, 2009. doi:10.1109/med.2009.5164621.

L. García, J. Barreiro-Gomez, E. Escobar, D. Téllez, N. Quijano, and C. Ocampo-Martinez. "Modeling and real-time control of urban drainage systems: A review". *Advances in Water Resources*, 85:120–132, 2015. ISSN 03091708. doi:10.1016/j.advwatres.2015.08.007.

M. S. Gelormino and N. L. Ricker. "Model-predictive control of a combined sewer system". *International Journal of Control*, 59(3):793–816, 1994. ISSN 13665820. doi:10.1080/00207179408923105.

S. Gillé, D. Fiorelli, E. Henry, and K. Klepiszewski. "Optimal Operation of a Sewer Network Using a Simplified Hydraulic Model". In *11th International Conference on Urban Drainage*, Edinburgh, 2008.

M. Grant and S. Boyd. "CVX: Matlab software for disciplined convex programming, version 2.1", 2013. URL: `http://cvxr.com/cvx`.

J. M. Grosso, P. Velarde, C. Ocampo-Martinez, J. M. Maestre, and V. Puig. "Stochastic model predictive control approaches applied to drinking water networks". *Optimal Control Applications and Methods*, 38(4):541–558, 2017. ISSN 10991514. doi:10.1002/oca.2269.

R. Halvgaard and A. K. V. Falk. "Water system overflow modeling for Model Predictive Control". In *the 12th IWA Specialised Conference on Instrumentation, Control and Automation*, Québec City, 2017.

R. Halvgaard, A. K. V. Falk, N. S. V. Lund, M. Borup, P. S. Mikkelsen, and H. Madsen. "Model predictive control for urban drainage: testing with a nonlinear hydrodynamic model". In *the 14th IWA international conference on urban drainage*, pages 2471–2474, Prague, 2017.

R. Löwe, L. Vezzaro, P. S. Mikkelsen, M. Grum, and H. Madsen. "Probabilistic runoff volume forecasting in risk-based optimization for RTC of urban drainage systems". *Environmental Modelling and Software*, 80:143–158, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.02.027.

H. Madsen, A. K. Falk, and R. Halvgaard. "A Model Predictive Control Framework for Real-Time Optimisation of Water System Operations". In *the 13th International Conference on Hydroinformatics (HIC)*, pages 1281–1288, Palermo, 2018. EPiC Engineering. doi:10.29007/fg7g.

A. L. Mollerup, P. S. Mikkelsen, and G. Sin. "A methodological approach to the design of optimising control strategies for sewer systems". *Environmental Modelling and Software*, 83:103–115, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.05.004.

K. M. Nielsen, T. S. Pedersen, C. Kallesøe, P. Andersen, L. S. Mestre, and P. K. Murugesan. "Control of sewer flow using a buffer tank". In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics*, pages 63–70. SCITEPRESS Digital Library, 2020.

V. Puig, G. Cembrano, J. Romera, J. Quevedo, B. Aznar, G. Ramón, and J. Cabot. "Predictive optimal control of sewer networks using CORAL tool: Application to Riera Blanca catchment in Barcelona". *Water Science and Technology*, 60(4): 869–878, 2009. ISSN 02731223. doi:10.2166/wst.2009.424.

M. R. Schütze, D. Butler, and M. B. Beck. *Modelling, Simulation and Control of Urban Wastewater Systems*. Springer, 2002. doi:10.1007/978-1-4471-0157-4.

J. F. Sturm. "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones". *Optimization Methods and Software*, 11(1-4):625–653, 1999. doi:10.1080/10556789908805766. URL: https://doi.org/10.1080/10556789908805766.

X. Tian, R. R. Negenborn, P. J. van Overloop, J. María Maestre, A. Sadowska, and N. van de Giesen. "Efficient multi-scenario Model Predictive Control for water resources management with ensemble streamflow forecasts". *Advances in Water Resources*, 109:58–68, 2017. ISSN 03091708. doi:10.1016/j.advwatres.2017.08.015.

P. J. van Overloop, S. Weijs, and S. Dijkstra. "Multiple Model Predictive Control on a drainage canal system". *Control Engineering Practice*, 16(5):531–540, 2008. ISSN 09670661. doi:10.1016/j.conengprac.2007.06.002.

L. Vezzaro and M. Grum. "A generalised Dynamic Overflow Risk Assessment (DORA) for Real Time Control of urban drainage systems". In *the 9th International Conference on Urban Drainage Modelling*, Belgrade, 2014. doi:10.1016/j.jhydrol.2014.05.019.

Paper E

Learning-Based Predictive Control with Gaussian Processes:An Application to Urban Drainage Networks

Krisztian Mark Balla[a,b], Deividas Eringis[b], Mohammad Al Ahdab[b], Jan Dimon Bendtsen[b], Carsten Skovmose Kallesøe[a,b] and Carlos Ocampo-Martinez[c]

[a]Technology Innovation, Controls Department , Grundfos Holding A/S, Poul Due Jensens vej 7, DK-8850 Bjerringbro, Denmark
[b]Section for Automation and Control, Department of Electronic Systems, Aalborg University, Fredrik Bajers vej 7c, DK-9220 Aalborg, Denmark
[c]Automatic Control Department, Universitat Politècnica de Catalunya, Llorens i Artigas 4-6, Planta 2, 08028, Barcelona, Spain

Abstract**—***Many traditional control solutions in urban drainage networks suffer from unmodelled nonlinear effects such as rain and wastewater infiltrating the system. These effects are challenging and often too complex to capture through physical modelling without using a high number of flow sensors. In this article, we use level sensors and design a stochastic model predictive controller by combining nominal dynamics (hydraulics) with unknown nonlinearities (hydrology) modelled as Gaussian processes. The Gaussian process model provides residual uncertainties trained via the level measurements and captures the effect of the hydrologic load and the transport dynamics in the network. To show the practical effectiveness of the approach, we present the improvement of the closed-loop control performance on an experimental laboratory setup using real rain and wastewater flow data.*

Keywords**—***Learning-based predictive control, Gaussian Process, Urban drainage network*

Accepted in the *Proceedings of American Control Conference*, 2022.

Contents

1 Introduction

Real-time control in Urban Drainage Networks (UDNs) allows for the systematic mitigation of water volumes, typically exploiting the available sensor measurements, weather forecasts, and, in some cases, the available physical description of the network. In an urban drainage context, sensors typically measure flow and level, furthermore, the most common actuators are pumps and gates. In combined UDNs, the disturbances are considered as the meteorological and human waste water loads. In this article, we focus on combined UDNs, where the actuators are pumps and both rain and wastewater gravitate from station to station in open pipes until reaching the Waste Water Treatment Plant (WWTP). To predict the volumes, e.g., avoid overflows and utilize the system capacity equally, it is essential to measure levels and flows. Here, we distribute only easy-accessible level sensors in the network and use a Gaussian Process (GP)-based system identification. The proposed approach uses water level residuals to capture the nonlinearities coming from the behavior of the pipes and the infiltration of the disturbances. Furthermore, the approach captures the uncertainties in modelling and the disturbance forecasts.

Modelling UDNs for control is a complex task, given the delays and nonlinearities imposed by the flow transport between pumping stations. The flow-to-level translation inside the sewers requires either a High Fidelity (HiFi) model or a large number of flow measurements, often economically out of reach for smaller water utilities. Traditional techniques on conceptual modelling are reported in [Ocampo-Martinez, 2010] and [Litrico and Fromion, 2009], where the capacity of pipes are collectively modelled as virtual buffers, and in [Joseph-Duran et al., 2015], where the flow-to-level translation is modelled by polynomials. Grey-box modelling for level propagation in open pipes has been reported in [Balla et al., 2022], [Balla et al., 2021] and [Xu et al., 2011].

Model Predictive Control (MPC) for UDNs has been reported in [Joseph-Duran et al., 2015], [Litrico and Fromion, 2009], [Ocampo-Martinez, 2010] and [García et al., 2015], where the operational constraints and the weather forecasts have been considered deterministic. Taking into account disturbance uncertainties with ensemble forecasts has been reported in [Balla et al., 2021], while [Grosso et al., 2014] reported on using chance constraints to track the daily demands of drinking water consumption.

GP regression has been widely used in machine learning [Williams and Rasmussen, 2006] for applications where an unknown system is given without structural information. A systematic framework for uncertainty propagation in real-time control of dynamic systems has been proposed in [Hewing et al., 2020], while in [Wang et al., 2016], flow forecasting has been done with the use of GPs in drinking water applications. Learning with data via GPs and using a nominal model allows identifying the nonlinearities and provides a measure of uncertainty without any prior knowledge.

The contribution of the article is the following. A model and control methodology is proposed for UDNs that can characterize the uncertainty along

with the predictions and reject the meteorological disturbances. To this end, we utilize level sensors to generate residuals between the unknown part (pipe dynamics, model, and disturbance uncertainty) and the nominal part (storage tanks). We incorporate the topological structure between the GP outputs and select the regressors based on prior process knowledge. To show the practical feasibility of the problem in UDNs, we carry out experimental tests on a scaled laboratory setup, where the dimensionality of the uncertainty propagation via the GPs is reduced by actively selecting the points used for prediction.

The article is organized as follows. In Section 2, an overview of the operation of UDNs is presented. Section 3 first presents the GP regression, followed by the formulation of the uncertainty propagation. In Section 4, the stochastic MPC design is presented, whereupon the point selection algorithm is introduced. In Section 5, the experiment on the laboratory setup is detailed and results are presented, using disturbance data from a real-world network. This is followed by Section 6, where conclusions, and future research directions are drawn.

1.1 Nomenclature

Let $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{m\times n}$ denote the field of real numbers, the set of real column vectors of length n and the set of $m \times n$ matrices composed of entries in the real numbers, respectively. The superscript $^\top$ denotes transposition and I is the identity matrix of suitable dimensions. A normally distributed vector x with mean μ and variance σ is denoted by $x \sim \mathcal{N}(\mu, \sigma)$, and the expected value of a random variable by $\mathbb{E}\{\}$.

2 UDN model

2.1 Network representation

We consider stations, where pumps are placed in tanks acting as the actuators. Moreover, the main piping layout defines the topology of the network. In UDNs, the tree topology is the most common in practice [Lund et al., 2018], where the collected wastewater, rain, and groundwater are pumped from station to station until they reach the root of the network. The root is an outlet point, where the water is discharged either to a WWTP or to the environment. An illustration of main transport lines in a UDNs are shown in Figure 1.

Disturbances represent the meteorological load on the sewer network. In this work, we use forecasts of runoff flow due to rain combined with the wastewater produced by households. These disturbances enter the piping network through the nodes representing storage tanks or manholes. Note that our representation of the network takes into account only the main pipelines, which connect the pumping stations.

Remark E1. *To measure level variation in both tanks and pipes, we distribute level sensors along the network nodes. The location of the level sensors is*

based on the high-level piping layout, meaning that we aim to deploy sensors at network nodes where the disturbances act on the network, e.g., where urban areas or catchments are discharging.

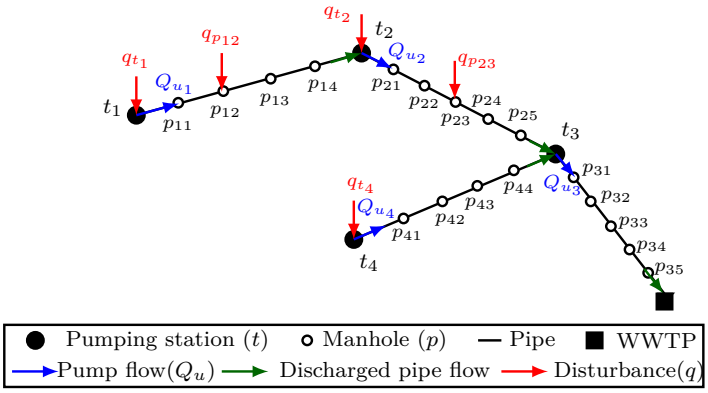


Figure 1: Tree topology of UDNs, where filled nodes represent storage tanks.

2.2 Physical component model

Flow propagating in open-channel pipes is most commonly approximated by the Saint-Venant nonlinear PDEs, describing the mass and the momentum of the fluid [Litrico and Fromion, 2009]:

$$\frac{\partial A(x,t)}{\partial t} + \frac{\partial Q(x,t)}{\partial x} = \tilde{q}(x,t), \tag{1a}$$

$$\frac{\partial Q(x,t)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q(x,t)^2}{A(x,t)}\right) + gA(x,t)\left(\frac{\partial h(x,t)}{\partial x} + S_f - S_b\right) = 0, \tag{1b}$$

where $Q(x,t)$ is the flow propagating inside the pipe at location x and time t and $\tilde{q}(x,t) = q(x,t)/dx$ is the lateral inflow per unit length, where we refer to $q(x,t)$ lateral inflow as a disturbance. $A(x,t)$ is a function describing the wetted pipe area while $h(x,t)$ is the level of water inside the channel. Besides, $Q(x,t), q(x,t), A(x,t)$ and $h(x,t)$ are functions from $(0,L)\times\mathbb{R}_+ \to \mathbb{R}_+$, and L, S_b and S_f are the length, slope and friction parameters, respectively.

The stored volumes in the network are modelled by linear tanks, for which the change in level per time unit is computed as the sum of all in- and outflows, i.e.,

$$\tau\frac{dh_t(t)}{dt} = q_t(t) + Q(t) - Q_u(t), \tag{2}$$

where $Q_u(t)$ denotes the sum of flows generated by pumps, $h_t(t)$ is the level in the tank, $q_t(t)$ is disturbance inflow to storage tanks, while $Q_t(t)$ is the incoming gravitated discharge from upstream stations. Besides, τ is the tank parameter, representing the geometry and size of the tank.

2.3 Data-driven model

The first-principle dynamics of transport pipes in eq.(1) are coupled by means of boundary conditions and the full PDEs can be computationally expensive to solve for complex network topology. As an alternative to [Hewing et al., 2020], we consider a discrete-time representation of the entire UDN in the form,

$$\begin{aligned} x(k+1) = f(x(k), u(k), d(k)) \\ + B_d\big(g(x(k), u(k), d(k)) + w(k)\big), \end{aligned} \tag{3}$$

where the model is composed of a nominal part f, describing integrators, i.e., the discrete-time storage tanks in eq.(2), whereas the remaining part defined by g represents the time and spatially discretized dynamics, i.e., the transport model for pipes in eq.(1). Besides, $x(k) \in \mathbb{R}^{N_x}$, $u(k) \in \mathbb{R}^{N_u}$ and $d(k) \in \mathbb{R}^{N_d}$ are the system state, input and disturbance at time step k, respectively. The process noise $w \sim \mathcal{N}(0, \Sigma^w)$ is considered independent identically distributed, with Σ^w being a diagonal variance matrix. Besides, B_d is a matrix mapping states corresponding to the pipe dynamics from the full state vector x. Furthermore, we consider the nominal dynamics to be linear in the standard state-space form:

$$x(k+1) = f(x(k), u(k), d(k)) = Ax(k) + Bu(k) + Ed(k), \tag{4}$$

where A, B and E are constant matrices of suitable dimensions.

3 Gaussian Process Regression

Similarly to [Hewing et al., 2020], we use GPs to identify the unknown dynamics g and the uncertainty w in eq.(3). A GP model is a probabilistic, non-parametric framework, most commonly used in supervised machine learning for predicting the distribution of output variables [Williams and Rasmussen, 2006]. In order to formulate output data (or target points) for the GPs, we create residuals between our measurements and the output of the nominal model dynamics, using the formulation in eq.(3). We assume that state measurements x_i are available at time step i:

$$y_i \triangleq g(x_i, u_i, \hat{d}_i) + w_i = B_p^{\dagger}\big(x_{i+1} - f(x_i, u_i, \hat{d}_i)\big), \tag{5}$$

where $\hat{d}_i \in \mathbb{R}^{N_d}$ is the vector of forecasted disturbances and $B_p^{\dagger}$ is the Moore-Penrose pseudo-inverse. The training set $\mathcal{D}$ is constructed from the inputs z

and outputs y by collecting data with a nominal controller

$$\begin{aligned}\mathcal{D} = \{&y = (y_1, \dots, y_M)^\top \in \mathbb{R}^{M \times N_y} \\ &z = (z_1, \dots, z_M)^\top \in \mathbb{R}^{M \times N_z}\},\end{aligned} \tag{6}$$

where $z_i \triangleq [x_i^\top \; u_i^\top \; \hat{d}_i^\top]^\top$, $N_z = N_x + N_u + N_d$ and M denotes the number of collected data points. Note that we use the GPs not only for capturing the model uncertainties w but to take care of the error between the forecasted and actual disturbances (d_i and $\hat{d}_i$) as well. Furthermore, we assume that the residuals for each state are independent, and therefore, we perform a GP regression for each residual y^a with $a \in \{1, \dots, N_y\}$. Since each finite collection of y^a is normally distributed, we can write, for each y^a,

$$y^a \sim \mathcal{N}\big(\mu^a(z), K_{zz}^a + I\sigma_a^2\big), \tag{7}$$

where y^a is the a^{th} column of y and σ_a^2 is the process noise variance. Besides, $\mu^a(z)$ is the mean and K_{zz}^a is the Gram matrix such that $K_{ij}^a = k^a(z_i, z_j)$, with $k^a(z_i, z_j)$ being a kernel function [Williams and Rasmussen, 2006]. We use the kernel function to describe the prior of the GP distribution, e.g., the covariance K between the points belonging to set $\mathcal{D}$ in eq.(6). The choice of the kernel function k^a is determined based on knowledge of the physical process. The squared exponential kernel function is chosen assuming dynamics that exhibit smooth and continuous behavior [Kocijan, 2016], i.e.,

$$k^a(z_i, z_j) = \sigma_{f,a}^2 \exp\!\left(-\frac{1}{2}(z_i - z_j)^\top S_a^\top \Lambda_a^{-1} S_a (z_i - z_j)\right), \tag{8}$$

where the hyper-parameter $\sigma_{f,a}$ is the signal variance and $\Lambda_a^{-1} = \operatorname{diag}(\sigma_{L,1}^{-2}, \dots, \sigma_{L,N_z}^{-2})$ is the length scale matrix. Note that different length-scale parameters are used on each dimension of z, thereby determining the relative importance of the contributions made by each input.

Remark E2. *Mapping matrices $S_{i_1,\dots,i_n}^a \in \mathbb{R}^{n \times N_z}$ are introduced for each output dimension a, picking states x_i, inputs u_i and disturbances $\hat{d}_i$ for each residual y_i. The mapping is determined based on the structure of the network, i.e., based on the physical network model in eq.(1) and eq.(2). This allows to reduce the training set for each output dimension, thereby easing the computational cost due to the high dimension of the training set $\mathcal{D}$.*

Given a testing point z_*, we aim to predict the residual $y_* = (g(z_*) + w_*)$ given the training set $\mathcal{D}$, i.e., we aim to find the distribution of $p(y_*^a | y^a)$. The joint distribution is

$$\begin{pmatrix} y^a \\ y_*^a \end{pmatrix} \sim \mathcal{N}\left(\mu^a, \begin{bmatrix} K_{zz}^a + I\sigma_a^2 & K_{zz^*}^a \\ K_{z^*z}^a & K_{z^*z^*}^a \end{bmatrix}\right), \tag{9}$$

where $[K^a_{zz^*}]_i = k^a(z_i, z^*)$, $K^a_{z^*z} = (K^a_{zz^*})^\top$, and $K^a_{z^*z^*} = k^a(z^*, z^*)$. The conditional distribution of the residuals is Gaussian [Girard et al., 2003], where the mean and variance are given as

$$y^a_* | y^a \sim \mathcal{N}(\mu^p_a(z^*), \Sigma^p_a(z^*)), \tag{10a}$$

$$\mu^p_a(z^*) = K^a_{z^*z}\left(K^a_{zz} + I\sigma^2_a\right)^{-1} y^a, \tag{10b}$$

$$\Sigma^p_a(z^*) = K^a_{z^*z^*} - K^a_{z^*z}\left(K^a_{zz} + I\sigma^2_a\right)^{-1} K^a_{zz^*}, \tag{10c}$$

where μ^p_a and Σ^p_a are the mean and variances of the GP for output dimension a, respectively. By stacking the predicted residuals y^a_* in a single vector $p(z^*)$, we can write

$$p(z^*) \sim \mathcal{N}(\mu^p(z^*), \Sigma^p(z^*)), \tag{11}$$

with mean $\mu^p(z^*) = [\mu^p_1(z^*), \ldots, \mu^p_{N_x}(z^*)]^\top$, and variance $\Sigma^p(z^*) = \operatorname{diag}(\Sigma^p_1(z^*), \ldots, \Sigma^p_{N_x}(z^*))$.

3.1 Uncertainty Propagation

The iterative, multi-step ahead prediction with the GP model is done by feeding back the mean and the variance of the predicted states, making each input a Gaussian random variable. Hence, the prediction of the states is in general non-Gaussian [Girard et al., 2003], as the probability distribution of the GP needs to be propagated through the nonlinear kernel function in eq.(8). In this work, the states and the GP are approximated as jointly Gaussians, where the predicted mean μ^x_i and covariance Σ^x_i of the states are given by

$$\mu^x_{i+1} = f(\tilde{z}_i) + B_d \mu^p_i, \tag{12a}$$

$$\Sigma^x_{i+1} = [\nabla_x f(\tilde{z}_i) \quad B_p] \Sigma_i [\nabla_x f(\tilde{z}_i) \quad B_p]^\top, \tag{12b}$$

where $\tilde{z}_i \triangleq (\mu^x_i, u_i, \hat{d}_i)$, the mean μ^p and variance Σ^p are given by eq.(10), and Σ_i is the covariance of the jointly Gaussian approximation of the states and the GP. Note that the inputs are assumed to be known and therefore treated as deterministic variables. The mean μ_i and the covariance Σ_i of the joint distribution are then given by

$$\begin{pmatrix} x_i \\ p_i + w_i \end{pmatrix} \sim \mathcal{N}(\mu_i, \Sigma_i) = \left(\begin{bmatrix} \mu^x_i \\ \mu^p_i \end{bmatrix}, \begin{bmatrix} \Sigma^x_i & \Sigma^{xp}_i \\ \Sigma^{px}_i & \Sigma^p_i + \Sigma^w_i \end{bmatrix} \right), \tag{13}$$

where $\Sigma^{px} = (\Sigma^{px})^\top$ are the cross-covariances between the states and the GP. Due to the linear nominal dynamics in eq.(4), the mean and variance dynamics can be simply written with the cross-covariances

$$\mu^x_{i+1} = A\mu^x_i + Bu_i + E\hat{d}_i + B_p \mu^p_i, \tag{14a}$$

$$\Sigma^x_{i+1} = A\Sigma^x_i A^\top + B_p \Sigma^{px}_i A^\top + A\Sigma^{xp}_i B_p^\top + B_p(\Sigma^p_i + \Sigma^w_i)B_p^\top. \tag{14b}$$

To solve the above approximation of the state distribution in a tractable way, the dynamics of the Gaussian distribution in eq.(12) are found through Taylor series expansion of eq.(10) around the mean μ^x. The covariance matrix Σ_i^p for each GP and the cross-covariances Σ_i^{xp} between the GP and the states are then updated such that

$$\mu_i^p = \mu^p(\tilde{z}_i), \tag{15a}$$

$$\Sigma_i^{xp} = \Sigma_i^x \left(\nabla_x \mu^p(\tilde{z}_i)\right)^\top \tag{15b}$$

$$\Sigma_i^p = \Sigma^p(\tilde{z}_i) + \nabla_x \mu^p(\tilde{z}_i) \Sigma_i^x \left(\nabla_x \mu^p(\tilde{z}_i)\right)^\top, \tag{15c}$$

where $\mu^p(\tilde{z}_i)$ and $\Sigma^p(\tilde{z}_i)$ are for each GP dimension a, as stated in eq.(10). The Taylor approximation used in this work is detailed in [Girard et al., 2003]. Different methods for approximating the posterior of a GP from Gaussian input have been researched, see for example [Williams and Rasmussen, 2006].

4 Stochastic MPC Design

4.1 Tractable GP-MPC

Introducing the nonlinear kernel and propagating the uncertainties with the states following a Gaussian distribution $x_i \sim \mathcal{N}(\mu_i^x, \Sigma_i^x)$ adds complexity to the optimization problem behind the GP-MPC. To solve the optimization problem in a tractable way, we formulate the problem as follows:

$$\mathcal{L}(k) = \mathbb{E}\Big\{ \sum_{i=k}^{k+H_p-1} \|x_i - x_i^r\|_Q^2 + \|\Delta u_i\|_R^2 + \|\epsilon\|_S^2 \Big\}, \tag{16}$$

where H_p denotes the length of the prediction horizon, the Δu term is introduced for integral action on flow control of the pumps and x^r is a reference signal for water level in the tanks. The reference is introduced in the state penalty term to show the effectiveness of the closed-loop behavior. Although this is somewhat unrealistic regarding the application, we artificially create reference scenarios to push the controller to its limits when testing under uncertain forecast signals. The slack variable ϵ is introduced for state constraint relaxation for water level violation in storage tanks, where the amount of level violation is related to the overflow volumes escaping the system. Furthermore, the weights Q, R and S represent a prioritization between the different objectives, and the individual terms in eq.(16) are normalized, such that water level and flow terms are comparable in magnitude. Although the evolution of the states is stochastic, we chose to implement the optimization problem by simply considering deterministic state constraints where the slack variables for overflow provide recursive feasibility to the problem.

Using the approximate distribution of the states and the expected value of the objective function, we formulate the deterministic optimization problem

behind the GP-MPC. The optimization problem in eq.(17) is solved in a receding horizon fashion where the dynamics f are discretized with the fixed step, 4^{th} order Runge-Kutta method. Moreover, the single-shooting method is used in the symbolic framework `CasADI` [Andersson et al., 2019] and a primal-dual interior point solver, `IPOPT` [Wächter and Biegler, 2006] is used to solve the non-convex optimization problem. The optimization is solved with warm start at each time step, given the process is slowly-varying.

$$\min_{\Delta u_i} \sum_{i=k}^{k+Hp-1} \|\mu_i^x - x_i^r\|_Q^2 + \text{tr}(Q\Sigma_i^x) + \|\Delta u_i\|_R^2 + \|\epsilon_i\|_S^2 \tag{17a}$$

$$\begin{aligned}
\text{s.t.} \quad & \mu_{i+1}^x = f(\tilde{z}_i) + B_d \mu^p(\tilde{z}_i), && \text{(17b)}\\
& \Sigma_{i+1}^x = [\nabla_x f(\tilde{z}_i)\ B_p]\Sigma_i[\nabla_x f(\tilde{z}_i)\ B_p]^\intercal, && \text{(17c)}\\
& \Delta u_i = u_i - u_{i-1}, && \text{(17d)}\\
& H_x \mu_i^x \leq b_x + H_\epsilon \epsilon_i, && \text{(17e)}\\
& H_u u_i \leq b_u, && \text{(17f)}\\
& \mu_i^p,\ \Sigma_i \text{ according to eq.(14b), eq.(15)}, && \text{(17g)}\\
& \mu_0^x = x(k), \quad \Sigma_0^x = 0, && \text{(17h)}
\end{aligned}$$

where $i = k, ..., k + H_p - 1$, furthermore eq.(17e) and eq.(17f) are $2N_x$ and $2N_u$ dimensional polytopes representing state and input constraints, respectively, where $H_x \in \mathbb{R}^{2N_x \times Nx}$, $b_x \in \mathbb{R}^{N_x}$, $H_u \in \mathbb{R}^{2N_u \times N_u}$, $b_u \in \mathbb{R}^{N_u}$ and $H_\epsilon \in \mathbb{R}^{2N_x \times Nx}$.

4.2 Subset of data approximation

The computational complexity of solving the optimization problem presented in eq.(17) is highly influenced by the GP model representing the unknown pipe dynamics and the uncertainty of the disturbance load on the network. The computational burden of propagating the uncertainty depends on the number of data points M used in eq.(6). To lower the computational complexity, the data set used for the prediction needs to be used in a computationally efficient way. Several sparse methods exist for approximating the distribution of GPs [Snelson and Ghahramani, 2006], among which we use the Subset of Data (SoD) method, where the computation is reduced to $\mathcal{O}(\tilde{M}^3)$ from $\mathcal{O}(M^3)$, by using a subset of $\tilde{M} < M$ data points [Chalupka et al., 2013]. Similar point selection methods have been used in [Kabzan et al., 2019]. The SoD method is shown in Figure 2.

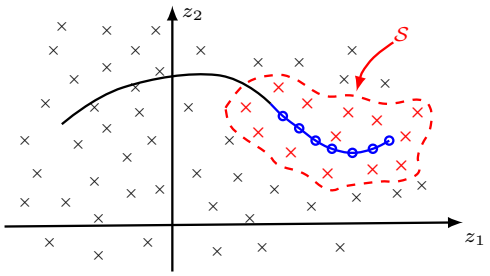


Figure 2: Conceptual point selection scheme, where $\mathcal{S}$ is the selected subset.

As opposed to more advanced sparse approximation methods, the computation with SoD can be reduced drastically, however, to the cost of degrading the prediction quality. Therefore, we spend the extra resources on using a larger subset $\tilde{M}$ and selecting new points at each sampling time [Quinonero-Candela and Rasmussen, 2005]. The proposed approach is detailed in Algorithm 1.

Algorithm 1: Subset of Data point selection

input: $\mathcal{P} = z$, $i = 0$, $\mathcal{I} = \varnothing$
while $i < \tilde{M}$ **do**
  for $j = 0$ to $H_p - 1$ **do**
    i = i + 1
    $r = \text{argmin}_r \, \|\mathcal{P}(r) - \hat{z}_j\|^2$
    $\mathcal{I} = \mathcal{I} \cup r$
    $\mathcal{P} = \mathcal{P} \setminus r$
    if $i = \tilde{M}$ **then**
      break
    end if
  end for
end while
return: $\boldsymbol{\theta}^*$,$\boldsymbol{\Lambda}^*$,$\boldsymbol{z}^*(t_0)$, N_x

At each time step we choose a subset $\mathcal{S}$ of the training set $\mathcal{P}$, such that all points in $\mathcal{S}$ are close to the previously predicted trajectory $\hat{z}$. In Algorithm 1, the index set $\mathcal{I}$ is a set containing the indices r. Here, $\mathcal{I}$ is used to index into the full training data set $\mathcal{P}$, e.g., to find the locations of the data points closest to the previously predicted trajectory $\hat{z}_j$, where $j = (0, ..., H_p - 1)$. The SoD method assumes that the selected points are close enough to the solution trajectory calculated at the previous time step such that the prediction accuracy for the current time step is sufficient. Using the previous solution trajectory as the selection criteria is a fair assumption in case of the control of UDNs, as the dynamics and disturbances are slowly varying with respect to the sampling time of the controller.

5 Case Study

The experimental setup for testing the GP-MPC in UDN control is shown in Figure 3 [Val Ledesma et al., 2021]. This setup of a UDN represents a $1:80$ scale model of a section of a real-life UDN, meaning that the typical resolution and time scale of the disturbances, control steps, tank emptying, and sampling times have been scaled down accordingly. An upstream and downstream pumping station are connected by a gravity-driven sewer line, most commonly found in real-life pumped sewer infrastructures [Butler and Davies, 2006]. The open-channel pipeline is equipped with four level sensors equivalent to water levels measured in manholes in a real network. As shown in Figure 3, auxiliary tanks are utilized to pump the disturbances to the points where they act on the system. Note that in our practical setup we generate these disturbances by pumping them to the inflow locations, while our controller knows only the forecast. The level sensor measurements of the pipe and the two tanks, furthermore the flow of the pumps are obtained and locally managed at each unit with a Codesys soft-PLC in real-time [3S-Smart Software Solutions GmbH]. The data acquisition is done at every 0.5 s, while the control input is applied at every 10 minute. The periodic component of the disturbances is equal to 17 minutes, which corresponds to one day in real life.

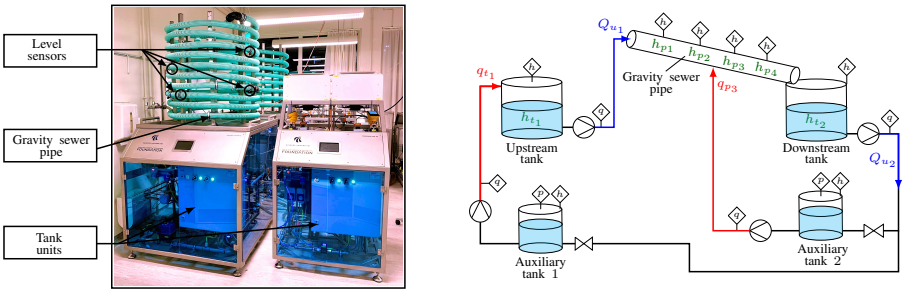


Figure 3: Sewer modules of the Smart Water Laboratory setup at Aalborg university (left). Schematics of the experimental setup (right). The sensor placements are indicated for each individual module with pressure sensor (p), level sensor (h) and flow sensor (q).

5.1 UDN model

Following the methodology in Section 2, the states, inputs and disturbances are given by the physical variables

$$x = \begin{bmatrix} h_{t_1} & h_{t_2} & h_{p_1} & h_{p_4} \end{bmatrix}^\top, \tag{18}$$

$$u = \begin{bmatrix} Q_{u_1} & Q_{u_2} \end{bmatrix}^\top, \tag{19}$$

$$\hat{d} = \begin{bmatrix} q_{t_1} & q_{p_3} \end{bmatrix}^\top, \tag{20}$$

where the input set $z = [x, u, \hat{d}]^\top$ is constructed as in eq.(6), where $N_x = 4$, $N_u = 2$ and $N_d = 2$, following the notation shown in Figure 3. Note that we utilize only h_{p_1} and h_{p_4} at the up- and downstream end of the pipeline, for the reason that the input flow Q_{u_1} from station one and the lateral disturbance flow q_{p_3} in the middle of the pipeline can be captured indirectly on these two level measurements. Hence, excluding h_{p_2} and h_{p_3} eases computation.

The nominal dynamics in eq.(4) are given in the form

$$A = \begin{bmatrix} I_{2\times2} & 0_{2\times2} \\ 0_{2\times2} & 0_{2\times2} \end{bmatrix}, B = \begin{bmatrix} \frac{T_s}{\tau_{t_1}} & 0 \\ 0 & \frac{T_s}{\tau_{t_2}} \\ & 0_{2\times2} \end{bmatrix}, E = \begin{bmatrix} -\frac{T_s}{\tau_{t_1}} & 0 \\ 0 & -\frac{T_s}{\tau_{t_2}} \\ & 0_{2\times2} \end{bmatrix} \tag{21}$$

where $A \in \mathbb{R}^{N_x \times N_x}$, $B \in \mathbb{R}^{N_x \times N_u}$ and $E \in \mathbb{R}^{N_x \times N_d}$. Moreover, T_s is the sampling time of the controller. Note that the nominal dynamics include the two integrator states h_{t_1} and h_{t_2}, whereas the remaining entries in the state vector are zeros, meaning that the nonlinear pipe dynamics are not part of the nominal dynamics. Hence, the GPs are chosen to take into account the pipe dynamics, the model uncertainty of the integrator states and the uncertainty of the forecasted disturbances $\hat{d}$. As such, the matrix B_p in eq.(3) is an identity of suitable dimensions. The uncertainty on the forecasted disturbances $\hat{d}$ is designed as an additive white noise on top of the mean of the forecast. Furthermore, due to the scaled nature of the Smart Water Laboratory setup, uncertainty and additional dynamics are present due to the pumps.

The mapping matrices defined in eq.(8) are given for each GP dimension $a = 1, ..., 4$, respectively:

$$\begin{aligned} S^1_{i_1,\dots,i_{n_1}} &\in \mathbb{R}^{n_1 \times N_z}, \quad i \in \{1, 5, 7\} \\ S^2_{i_1,\dots,i_{n_2}} &\in \mathbb{R}^{n_2 \times N_z}, \quad i \in \{2, 4, 6\} \\ S^3_{i_1,\dots,i_{n_3}} &\in \mathbb{R}^{n_3 \times N_z}, \quad i \in \{3, 5\} \\ S^4_{i_1,\dots,i_{n_4}} &\in \mathbb{R}^{n_4 \times N_z}, \quad i \in \{3, 4, 8\} \end{aligned}$$

where n defines the total number of dimensions we pick from the original N_z dimensional data set. Furthermore, i specifies the number of entry we pick from z for each output dimension a, respectively.

5.2 Experimental results

Following the model and data collection methodology in Section 3, the residuals y are constructed between the level measurements x in eq.(18) and the nominal tank dynamics in eq.(4). The data is collected from the laboratory setup under a nominal, threshold-based level controller, most commonly found in practice [Lund et al., 2018]. The hyper-parameters σ_f^2 and Λ are found for each output dimension $a = 1, ..., N_a$, using Bayesian optimization with the `fitrgp` toolbox in MATLAB. Moreover, $H_p = 25$ steps are used for prediction, corresponding

to 5 hours in real life. This is a reasonable horizon length in UDNs, as rainfall radar predictions can provide sufficient accuracy only up to $2-4$ hours. The point selection is done on a data set, where $\tilde{M} = 80$ points are selected at each time step for the residual predictions. Besides, the full dataset z from which we select the points for prediction is continuously updated.

The experimental results of the optimization problem in eq.(17) are shown in Figure 4-A and Figure 4-B. To evaluate the capabilities of the GP-MPC under uncertainty, $n = 10$ scenarios of imperfect rain and wastewater flow forecasts have been created by adding Gaussian distributed random data on top of the historic real flow data. The imperfect forecast is provided to the GP-MPC, whereas the historical flow data is implemented on the laboratory setup. Note that the pumps at both stations have a lower constraint different from zero, although in reality pumps can be shut down completely. This is for the reason, that below the lower limits the flow-based PI controllers cannot keep the given reference due to the small pressure drop in the test setup. As shown in Figure 4-A (c) and (d), the level references at the two stations are tracked, indicating that the effect of disturbances and the unknown dynamics are learned well. As expected, the water level in t_2 indicates a higher spread of the uncertainty in the predictions. This is partly because we use the GPs to model both the disturbance uncertainty and the nonlinear pipe dynamics providing the transport between the two stations.

In Figure 4-B (g-l), we show an event, which we observe at $t = 1500$ time steps in Figure 4-A (c) and (d). The pumps at the upstream t_1 tank reduce the flow to the lower limit, aiming to retent as much volume upstream as possible. Hence, the load on tank t_2 is eased, where at $t = 1500$ we operate the station close to its upper constraint. Note that even though the pumps at t_2, shown in Figure 4-B (l), run at full speed, the water level h_{t_2} is approaching a level, where constraint violation is likely under the uncertain flow forecasts. The constraint violation of the upper tank level is highly undesirable, as it relates to an overflow event, where the slacks ϵ have to be used in the optimization to lift the physical upper level of the tanks. The controller recognizes that the slacks in eq.(17) need to be used to avoid the infeasibility of the optimization problem, hence the controller rather violates the tracking of the reference in the tank t_1 and shuts down the pump, as shown in Figure 4-B (i) and (k). As shown in the left column of the results in Figure 4-A and Figure 4-B, the GP can learn the uncertainty on the forecasts and thereby control the pumps to the reference.

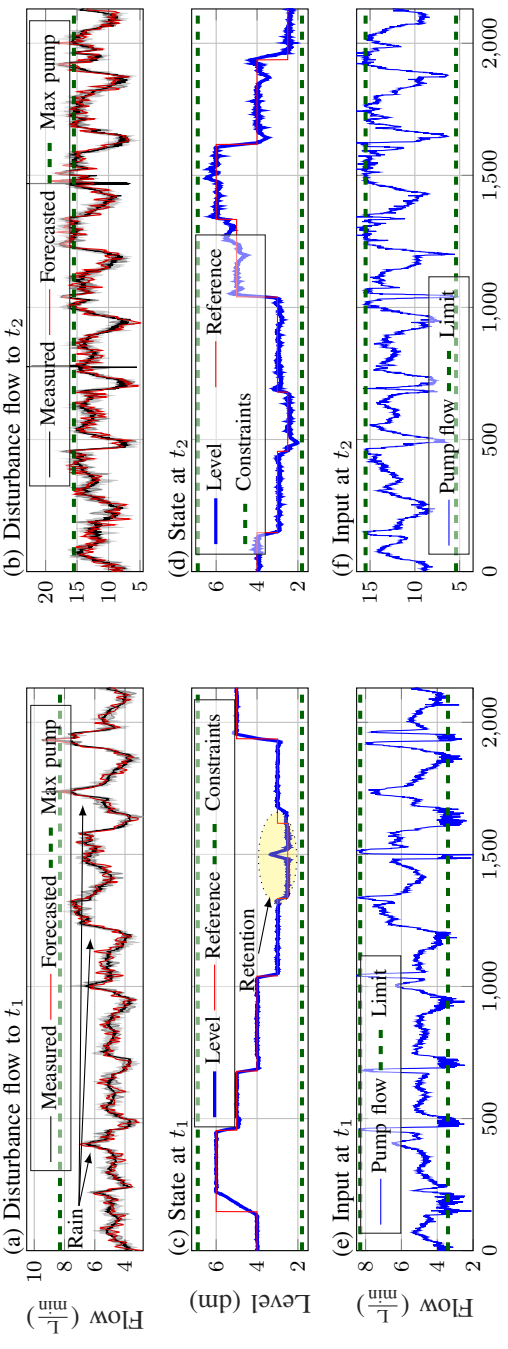


Figure 4-A: Closed-loop control results of the GP-MPC for disturbance rejection with water level reference tracking under uncertain weather forecasts.

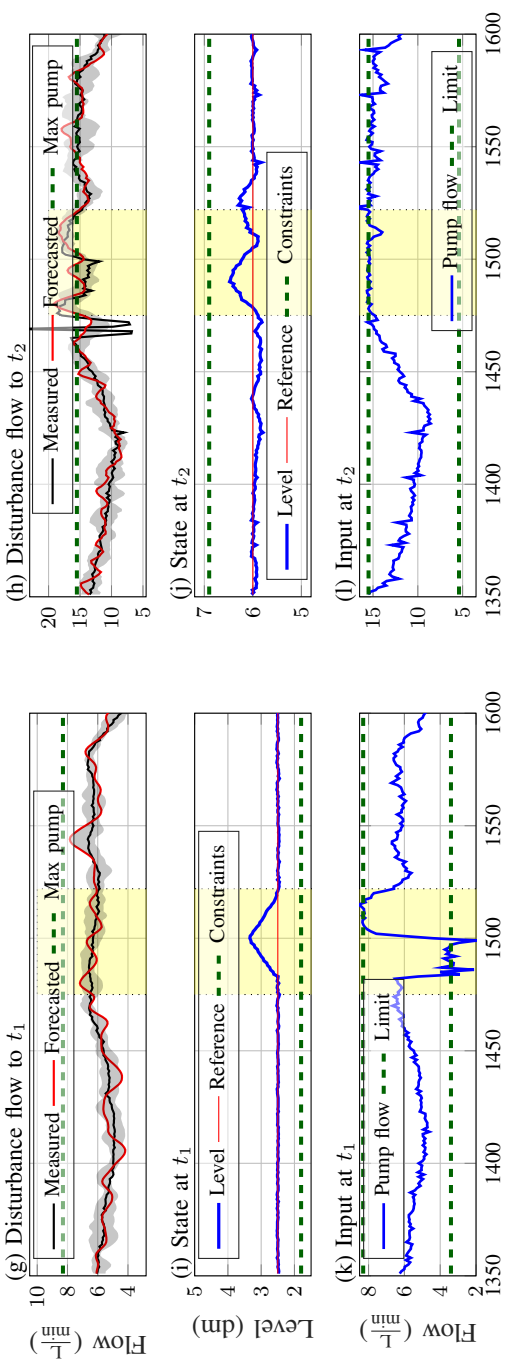


Figure 4-B: Closed-loop control results of the GP-MPC for disturbance rejection with water level reference tracking under uncertain weather forecasts.

6 Conclusions and Future Work

In this paper, we discussed a predictive control approach in urban drainage networks, where Gaussian process regression has been used to model the unknown dynamics and the disturbance uncertainties. For this purpose, we utilized the Gaussian process regression framework to learn only the parts where the traditional predictive control lacks a good first principle modelling approach. To this end, we used level sensors distributed along the network to learn the disturbance uncertainty and the pipe dynamics from the level variation under a rule-based nominal controller. The residual prediction through uncertainty propagation along with the deterministic model predictive controller based on the known dynamics has been solved in a receding horizon fashion on an experimental laboratory setup, using disturbance flow forecast data extracted from a real-world waste water utility. The performance of the reference tracking underpinned the feasibility of the practical utility of the Gaussian process-based predictive controller in sewer networks and showed the robustness capabilities towards uncertain disturbance forecasts.

An investigation into how the periodic behavior of human disturbances can be learned from the level sensor data with Gaussian processes and using rain intensity instead of flow is a matter of future work.

Acknowledgement

The authors would like to thank the Poul Due Jensen Foundation for providing the Smart Water Lab at Aalborg University. The project was supported by Innovation Fund Denmark and Grundfos Holding A/S as part of a Danish Industrial Ph.D. Project [Application number: *9065-00018A*]. The work of C. Ocampo-Martinez has been supported by the L-BEST project (Ref. PID2020-115905RB-C21) from the Spanish Ministry of Science and Innovation.

References

3S-Smart Software Solutions GmbH. "Codesys". URL: `https://www.codesys.com`.

J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. "CasADi: a software framework for nonlinear optimization and optimal control". *Mathematical Programming Computation*, 11(1):1–36, 2019. ISSN 18672957. doi:10.1007/s12532-018-0139-4.

K. M. Balla, C. H. Knudsen, A. Hodzic, J. D. Bendtsen, and C. S. Kallesøe. "Nonlinear Grey-box Identification of Sewer Networks with the Backwater Effect: An Experimental Study". In *2021 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1202–1207, San Diego, 2021. IEEE. doi:10.1109/CCTA48906.2021.9658864.

K. M. Balla, C. Schou, J. D. Bendtsen, C. Ocampo-Martínez, and C. S. Kallesøe. "A Nonlinear Predictive Control Approach for Urban Drainage Networks Using Data-Driven Models and Moving Horizon Estimation". *IEEE Transactions on Control Systems Technology (Early Access)*, pages 1–16, 2022. doi:10.1109/TCST.2021.3137712.

D. Butler and J. W. Davies. *Urban Drainage*. Spon Press, 2006. doi:10.1016/s1462-0758(00)00017-0.

K. Chalupka, C. K. Williams, and I. Murray. "A framework for evaluating approximation methods for Gaussian process regression". *Journal of Machine Learning Research*, 14(1):330–350, 2013. ISSN 15324435.

L. García, J. Barreiro-Gomez, E. Escobar, D. Téllez, N. Quijano, and C. Ocampo-Martinez. "Modeling and real-time control of urban drainage systems: A review". *Advances in Water Resources*, 85:120–132, 2015. ISSN 03091708. doi:10.1016/j.advwatres.2015.08.007.

A. Girard, C. E. Rasmussen, J. Quinonero-Candela, and R. Murray-Smith. "Gaussian process priors with uncertain inputs: Application to multiple-step ahead time series forecasting". 2003.

J. M. Grosso, C. Ocampo-Martínez, V. Puig, and B. Joseph. "Chance-constrained model predictive control for drinking water networks". *Journal of Process Control*, 24:504–516, 2014. ISSN 09591524. doi:10.1016/j.jprocont.2014.01.010.

L. Hewing, J. Kabzan, and M. N. Zeilinger. "Cautious Model Predictive Control Using Gaussian Process Regression". *IEEE Transactions on Control Systems Technology*, 28(6):2736–2744, 2020. ISSN 15580865. doi:10.1109/TCST.2019.2949757.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Output-feedback control of combined sewer networks through receding horizon control with moving horizon estimation". *Water Resources Research*, 51(10):8129–8145, 2015. ISSN 19447973. doi:10.1002/2014WR016696.

J. Kabzan, L. Hewing, A. Liniger, and M. N. Zeilinger. "Learning-based model predictive control for autonomous racing". *IEEE Robotics and Automation Letters*, 4(4):3363–3370, 2019.

J. Kocijan. *Modelling and control of dynamic systems using Gaussian process models*. Springer, 2016.

X. Litrico and V. Fromion. *Modeling and control of hydrosystems*. Springer, 2009. ISBN 9781848826236. doi:10.1007/978-1-84882-624-3.

N. S. V. Lund, A. K. V. Falk, M. Borup, H. Madsen, and P. Steen Mikkelsen. "Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management". *Critical Reviews in Environmental Science and Technology*, 48(3):279–339, 2018. ISSN 15476537. doi:10.1080/10643389.2018.1455484.

C. Ocampo-Martinez. *Model Predictive Control of Wastewater Systems*. Springer, Barcelona, 1st edition, 2010. ISBN 978-1-84996-352-7. doi:10.1007/978-1-84996-353-4.

J. Quinonero-Candela and C. E. Rasmussen. "A unifying view of sparse approximate gaussian process regression". *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

E. Snelson and Z. Ghahramani. "Sparse gaussian processes using pseudo-inputs". *Advances in Neural Information Processing Systems*, 18:1259–1266, 2006.

J. Val Ledesma, R. Wisniewski, and C. S. Kallesøe. "Smart water infrastructures laboratory: Reconfigurable test-beds for research in water infrastructures management". *Water (Switzerland)*, 13(13):1875, 2021. ISSN 20734441. doi:10.3390/w13131875.

A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". *Mathematical programming*, 106(1):25–57, 2006.

Y. Wang, C. Ocampo-Martinez, and V. Puig. "Stochastic model predictive control based on Gaussian processes applied to drinking water networks". *IET Control Theory and Applications*, 10(8):947–955, 2016. ISSN 17518652. doi:10.1049/iet-cta.2015.0657.

C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow". *Advances in Water Resources*, 34:282–290, 2011. ISSN 03091708. doi:10.1016/j.advwatres.2010.11.009.

Paper F

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

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

[a]Technology Innovation, Grundfos, Poul Due Jensens Vej 7, 8850, Bjerringbro, DK
[b]Water Utility, Grundfos, Poul Due Jensens Vej 7, 8850, Bjerringbro, DK
[c]Dept. of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, DK
[d]Automatic Control Department, Universitat Politècnica de Catalunya, Llorens i Artigas 4-6, Planta 2, 08028, Barcelona, ES

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.*

Keywords—*Smart water system, Data-driven modelling, Uncertainty*

Published in the *Water Research Journal*.

DOI: https://doi.org/10.1016/j.watres.2022.118782

Contents

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 [Yuan et al., 2019], [Eggimann et al., 2017]. 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].

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 [Yuan et al., 2019], [Mignot et al., 2012], as well as for modelling the hydrologic processes, e.g., rain running off the catchments [Li et al., 2019]. 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 [Kisi et al., 2013], [Ahm et al., 2016], [Chen et al., 2014], [Kallesøe and Knudsen, 2016], [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., 2011], [Xu et al., 2012], 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 [Zhang et al., 2021], [Thrysøe et al., 2019].

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], [Löwe et al., 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., 2014] 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 an are of research 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 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 the available physical description of the network, we make our method robust towards forecast uncertainties 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.

2 Problem Statement

The overall concept of the method is shown in Figure 1, where in-sewer level sensors are deployed at critical locations in manholes and basins. The network topology is defined by a directed tree graph [Thrysøe et al., 2019], 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 Figure 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}). 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 pump flows in the proposed output-feedback scheme. In the following, we present the control scheme by describing each building block.

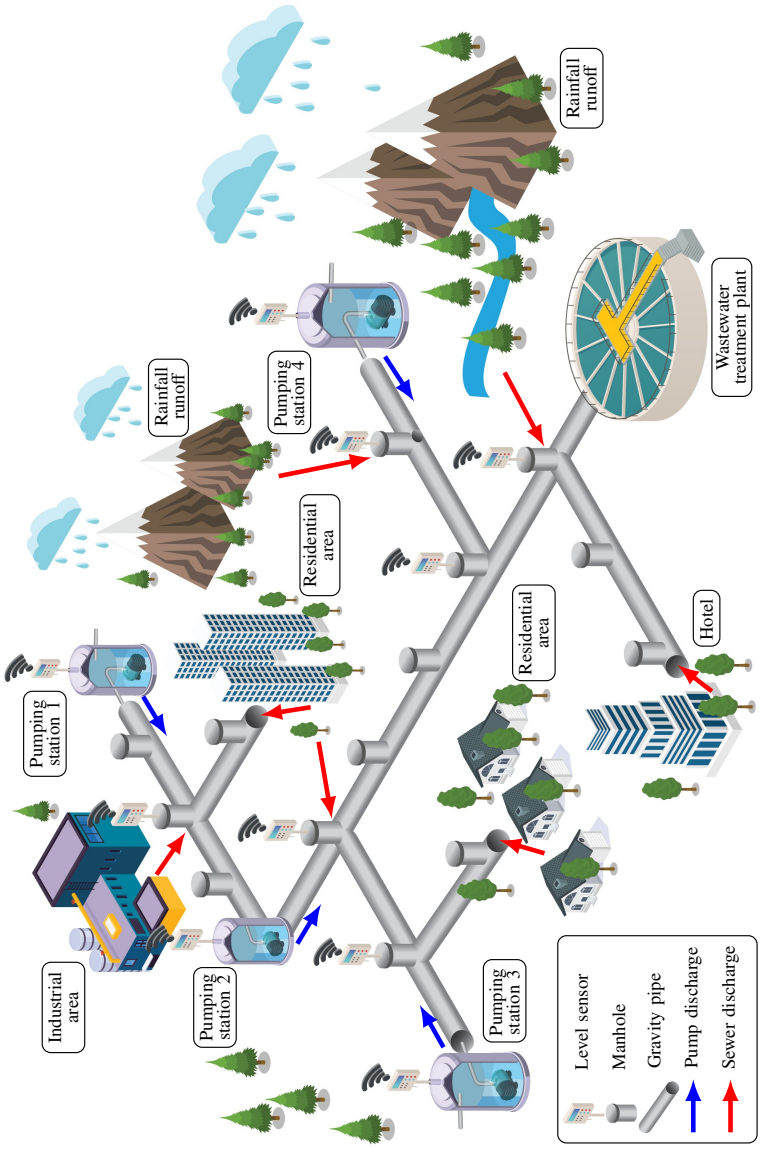


Figure 1: An illustration of a pumped wastewater network, where water level sensors are deployed in critical points.

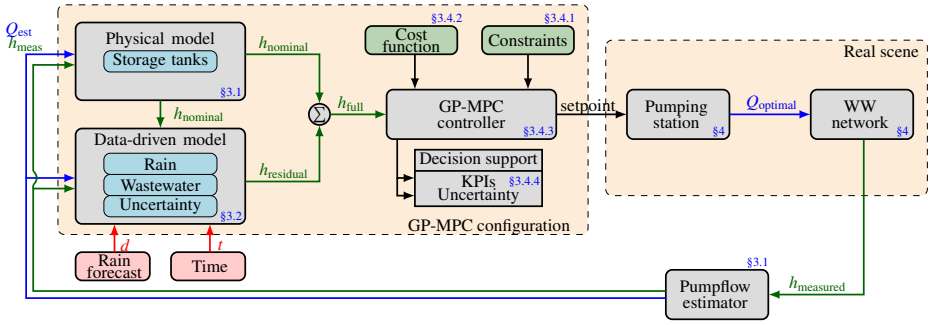


Figure 2: Closed-loop topology of the GP-MPC controller. Signals denoted with blue are flow variables, green signals are water levels and red signals denote the rain forecast and time. The pipe network (plant) is represented by the *WW network* block and each block is labeled by the number of section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

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

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

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 (rainfall-runoff), i.e., $q(t) = q_r(t) + q_{ww}(t)$, where q_r and q_{ww} are the flows generated by rainfall-runoff and domestic wastewater production, respectively, while q is the combined flow. This formulation allows to apply it to both combined and stormwater 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). 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

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

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 $w \sim \mathcal{N}(0, \Sigma_w)$, following Gaussian white noise distribution. Besides, $\boldsymbol{B}_p$ is a matrix mapping the nonlinear dynamics g to the full state vector $\boldsymbol{h}$. Simply stated: if there is a storage tank where the level variation h is not affected by uncertain inflows, $\boldsymbol{B}_p$ maps the lower dimensional outputs of the function 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 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 forecasts indicate how rainfall infiltrates and appears as flow q_r in the sewers. Since the dry-weather flow q_{ww} generated by domestic wastewater production typically follows a diurnal pattern, the cyclical behavior correlates to time, i.e., it is likely that the flow patterns are similar at the same time of the day at a different week. For this reason, when we consider the implicit formulation of $\boldsymbol{g}$, we assume that time t is also an input. In this way, we provide time as an additional input indicating the variation of the dry-weather flow patterns. For ease of notation, let us define the input data as a vector

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

To generate output data for learning, we use the additive feature of the nominal and unknown dynamics defined in eq.(4). To create residuals, we use the data provided by the level sensors and subtract the nominal dynamics, i.e.,

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

where $\boldsymbol{y} \in \mathbb{R}^{N_y}$ is the vector of residuals of size N_y, corresponding to the number of states or 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. The main benefit of constructing the residuals $\boldsymbol{y}$ is that we remove the already known dynamics from the signal while obtaining a more clear picture of the impact of dry- and wet-weather flows on the level variations. Simply stated: by using the level sensors distributed in the network, we aim to capture as much dry- and wet-weather flow dynamics as possible.

The training set is constructed by collecting the input-output data pairs under nominal operation, i.e.,

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

where M is the number of collected data points.

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 parametrizing g, we characterize the distribution of all possible functions of g, i.e., the residuals we generated with eq.(6). Hence the GP model representing one dimension of the residual $\boldsymbol{y}$ is given by

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

where the distribution of the Gaussian process is fully characterized by its mean function $m(\boldsymbol{z})$ and covariance $\boldsymbol{\Sigma}_{GP}$. We consider the mean $m(\boldsymbol{z})$ as a constant function, equivalent to a model bias in case of a zero mean GP. 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

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

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

where the mean $m(\boldsymbol{z})$ and the covariance matrix $\boldsymbol{\Sigma}_{GP}$ are obtained by evaluating the mean and covariance functions given all measured data pairs in $\mathcal{D}$. 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 $\boldsymbol{g}$. 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]. It is however straightforward to use any other differentiable kernels which characterize well our initial knowledge or belief about the function $\boldsymbol{g}$. The squared exponential kernel is given by

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

where the kernel is characterized by its hyper-parameters σ_f^2 and $\Lambda^{-1} = \operatorname{diag}(\sigma_{L,1}^{-2}, ..., \sigma_{L,N_z}^{-2})$ 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 $\boldsymbol{z}$ [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 even if some of the information is irrelevant. Consequently, the mapping matrix $\boldsymbol{S}$ is constructed by using the topology

of the network for each residual. These mapping matrices embed information about how the network components are connected and thereby map only the relevant dimensions of the training set $\mathcal{D}$.

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

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

Given our problem formulation, the posterior distribution simplifies to [Rasmussen and Williams, 2018]

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

The hyper-parameters of the above problem are learned by maximizing the marginal likelihood $\mathbb{P}\{\boldsymbol{y}\,|\,\boldsymbol{z}\}$, 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 residual y^* at a test point $\boldsymbol{z}^*$, using the relation $y^* = g(z^*)$. The problem of predicting the residual corresponds to finding the probability distribution of $\mathbb{P}\{y^*\,|\,\mathcal{D},\boldsymbol{z}^*\}$, given the training data $\mathcal{D}$, a testing input $\boldsymbol{z}^*$ 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.,

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

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

where $\boldsymbol{K}_{\boldsymbol{z}z^*} = \boldsymbol{k}(\boldsymbol{z}, z^*)$ and $\boldsymbol{K}_{z^*\boldsymbol{z}} = \boldsymbol{K}_{\boldsymbol{z}z^*}^\top$ are the covariances between the training and testing points, furthermore $\boldsymbol{K}_{z^*z^*}$ is the autocovariance of the testing point.

3.3 Probabilistic prediction model

Adopting the Gaussian processes, we aim to predict the effect of wet- and dry-weather discharges on the water levels over multiple prediction steps. 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.,

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

where $\boldsymbol{\Sigma}_{h,GP} = (\boldsymbol{\Sigma}_{GP,h})^\top$ are the cross-covariances between the states and the Gaussian process, $\boldsymbol{\mu}_h(t)$ is the vector of mean levels and $\boldsymbol{\Sigma}_h(t)$ is the covariance matrix of the water levels at time t . The pump flows $\boldsymbol{Q}$ are treated as deterministic variables.

To find the transition probability of the system states, 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 states at time step t [Hewing et al., 2020]. Then the mean and variance dynamics of the water levels h result in

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

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

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. Then, the mean and variance dynamics become

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

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

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

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

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

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

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

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. The optimal inputs are computed to obtain a future response of the water system. The calculated 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 $\boldsymbol{b}_Q = [\boldsymbol{Q}_{\max}^\top, \boldsymbol{Q}_{\min}^\top]^\top \in \mathbb{R}^{2N_Q}$ 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 $\boldsymbol{H}_Q = [\boldsymbol{I}_{N_Q}, -\boldsymbol{I}_{N_Q}]^\top$ 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., 2016], 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 Figure 3.

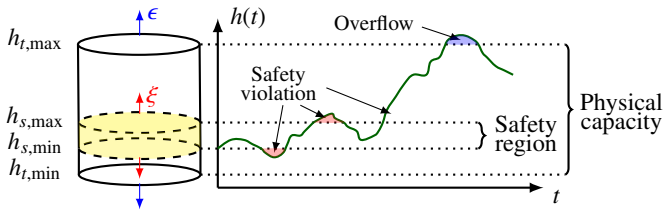


Figure 3: Safety and capacity constraints, where blue and red arrows are constraint relaxations for overflow (ϵ) and safety violation (ξ).

While the minimum and maximum level values of the physical capacity con-

straints 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., 2016], 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 $\boldsymbol{H}_h = [\boldsymbol{I}_{N_h}, -\boldsymbol{I}_{N_h}]^\top$ and $\boldsymbol{H}_s = [\boldsymbol{I}_{N_{h_t}}, -\boldsymbol{I}_{N_{h_t}}]^\top$ map the vector of water levels $\boldsymbol{h}$ to suitable size of $\boldsymbol{b} = [\boldsymbol{h}_{t,\max}^\top, \boldsymbol{h}_{t,\min}^\top]^\top$ and $\boldsymbol{b}_s = [\boldsymbol{h}_{s,\max}^\top, \boldsymbol{h}_{s,\min}^\top]^\top$ water level bounds, respectively.

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 [Wang et al., 2016], [Hewing et al., 2020]. The constraints are given by

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

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

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 $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_{\max}^\top, \boldsymbol{\epsilon}_{\min}^\top]^\top$ and $\boldsymbol{\xi} = [\boldsymbol{\xi}_{\max}^\top, \boldsymbol{\xi}_{\min}^\top]^\top$ 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 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

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

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 $\boldsymbol{\lambda}_1$ matrix, where $\boldsymbol{\lambda}_1$ is diagonal and $\mathbf{0} \leq \boldsymbol{\lambda} \leq \boldsymbol{I}$, similarly to all $\boldsymbol{\lambda}$ 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 $\boldsymbol{\lambda_3}$, 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]:

$$\mathbb{E}\Big\{W_3\|\boldsymbol{h}(t)\|^2_{\boldsymbol{\lambda_3}}\Big\} = W_3\Big[\|\boldsymbol{\mu}_h(t)\|^2_{\boldsymbol{\lambda_3}} + \mathrm{tr}\{\boldsymbol{\lambda}_3\boldsymbol{\Sigma}_h(t)\}\Big], \tag{22}$$

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 Figure 2). The problem is given by

$$\begin{aligned}\underset{\substack{\Delta\boldsymbol{Q}(0),\dots,\Delta\boldsymbol{Q}(H_p-1)\\ \boldsymbol{\epsilon}(0),\dots,\boldsymbol{\epsilon}(H_p-1)\\ \boldsymbol{\xi}(0),\dots,\boldsymbol{\xi}(H_p-1)}}{\text{Minimize}} \quad & \sum_{t=i}^{i+Hp-1} W_1\|\boldsymbol{\epsilon}(i)\|^2_{\boldsymbol{\lambda_1}} + W_2\|\boldsymbol{\xi}(i)\|^2_{\boldsymbol{\lambda_2}} + W_3\Big[\|\boldsymbol{\mu}_h(i)\|^2_{\boldsymbol{\lambda_3}} \\ & + \mathrm{tr}\{\boldsymbol{\lambda}_3\boldsymbol{\Sigma}_h(i)\}\Big] + W_4\|\Delta\boldsymbol{Q}(i)\|^2_{\boldsymbol{\lambda_4}}, \end{aligned} \tag{23a}$$

subject to

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

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

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

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

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

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

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

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

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

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

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 eq.(23b) and eq.(23c), forming equality constraints. After solving the optimization problem in eq.(23) at state $\boldsymbol{h}(0)$, the resulting decision variables form an optimal control sequence of the change in pumpflows $\boldsymbol{u} = [\Delta\boldsymbol{Q}^\top(0), \Delta\boldsymbol{Q}^\top(1), \dots, \Delta\boldsymbol{Q}^\top(H_p - 1)]$, 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 Figure 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

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

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

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

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

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.

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 Figure 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 minutes, corresponding to one day in real life. While the data acquisition is done at every 0.5 second, the control time step is 10 seconds, equivalent to sending a control signal every 12 minutes in real life. Besides, the data used in the experiment are real wastewater and rain precipitation[1].

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 Figure 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 $\boldsymbol{z}(t) = [\boldsymbol{h}^\top(t), \boldsymbol{Q}^\top(t), d(t), t]^\top$, where the level and flow signals at time t are given by

[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.

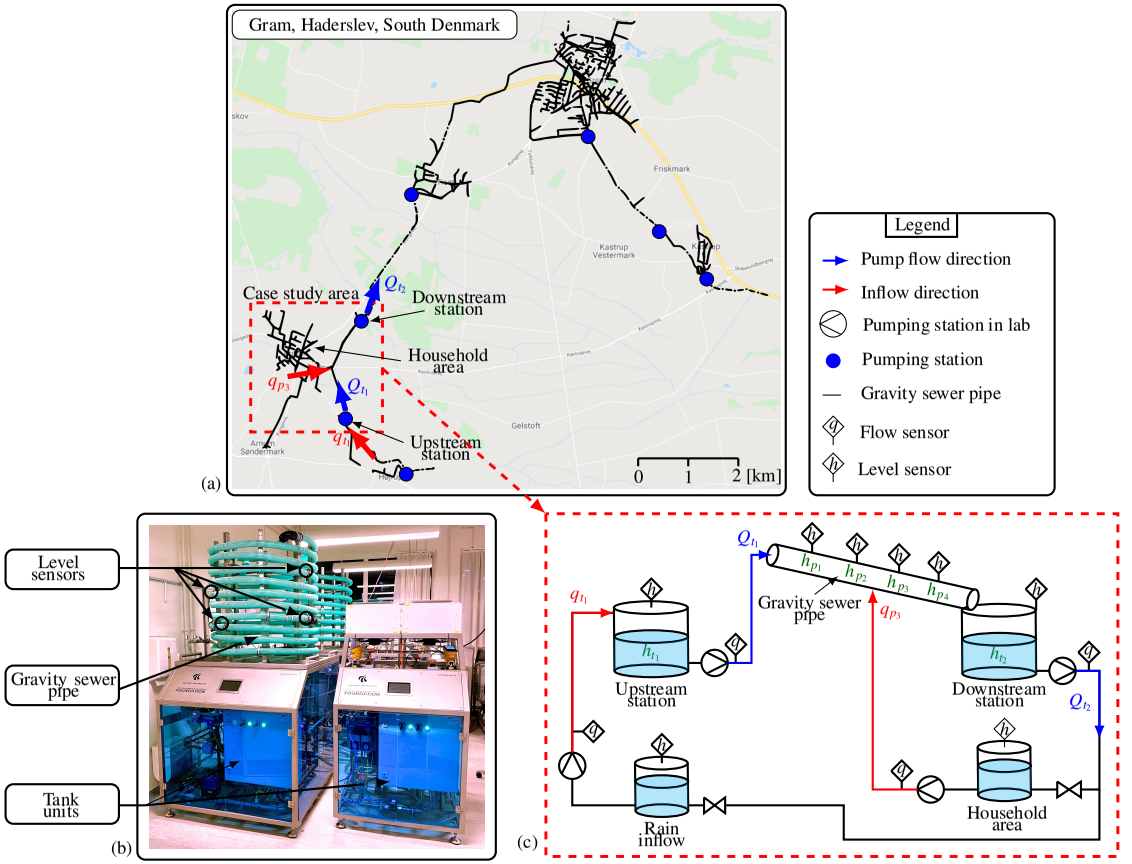


Figure 4: Case study area of a combined wastewater network in Gram, Denmark (a) and the equivalent representation of the considered network segment by the Smart Water Infrastructures Laboratory, where (b) is showing the experimental setup and (c) the schematics of the topology.

$$\boldsymbol{h}(t) = [h_{t_1}(t), h_{t_2}(t), h_{p_3}(t)]^\top, \tag{25a}$$
$$\boldsymbol{Q}(t) = [Q_{t_1}(t), Q_{t_2}(t)]^\top. \tag{25b}$$

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

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_{2\times 2} & \boldsymbol{0}_{2\times 1} \\ \boldsymbol{0}_{1\times 2} & 0 \end{bmatrix}, B = \begin{bmatrix} \frac{T_s}{\tau_{t_1}} & 0 \\ 0 & \frac{T_s}{\tau_{t_2}} \\ \boldsymbol{0}_{1\times 2} & \end{bmatrix}, \tag{26}$$

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 $\boldsymbol{z}$ influence the given residual based on the topological layout of the system, i.e.,

$$S^1_{i_1,\dots,i_{n_1}} \in \mathbb{R}^{n_1\times N_z}, \quad i \in \{4,6,7\}, \tag{27a}$$
$$S^2_{i_1,\dots,i_{n_2}} \in \mathbb{R}^{n_2\times N_z}, \quad i \in \{3,5,7\}, \tag{27b}$$
$$S^3_{i_1,\dots,i_{n_3}} \in \mathbb{R}^{n_3\times N_z}, \quad i \in \{4,7\}, \tag{27c}$$

where i is the index of the predictor in the input set z. As an example, to train the GP on residual y_1 (corresponding to the upstream tank t_1), the predictors $Q_{t1}, (i = 4)$, the rain forecasts $d, (i = 6)$, and the time $t, (i = 7)$ are used. 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 Figure 5.

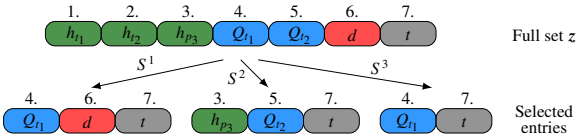


Figure 5: Feature selection with the slicing matrices S.

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. Figure 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 Figure 6).

The data describing residual y_2 varies due to the discharged pump flow coming from the upstream station t_1 and the lateral inflow coming from the household area. However, note that all dynamics due to the pumping have been removed from the signal. 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, while some effects of the gravitated discharge flow from the pump activity break the periodicity of the signal.

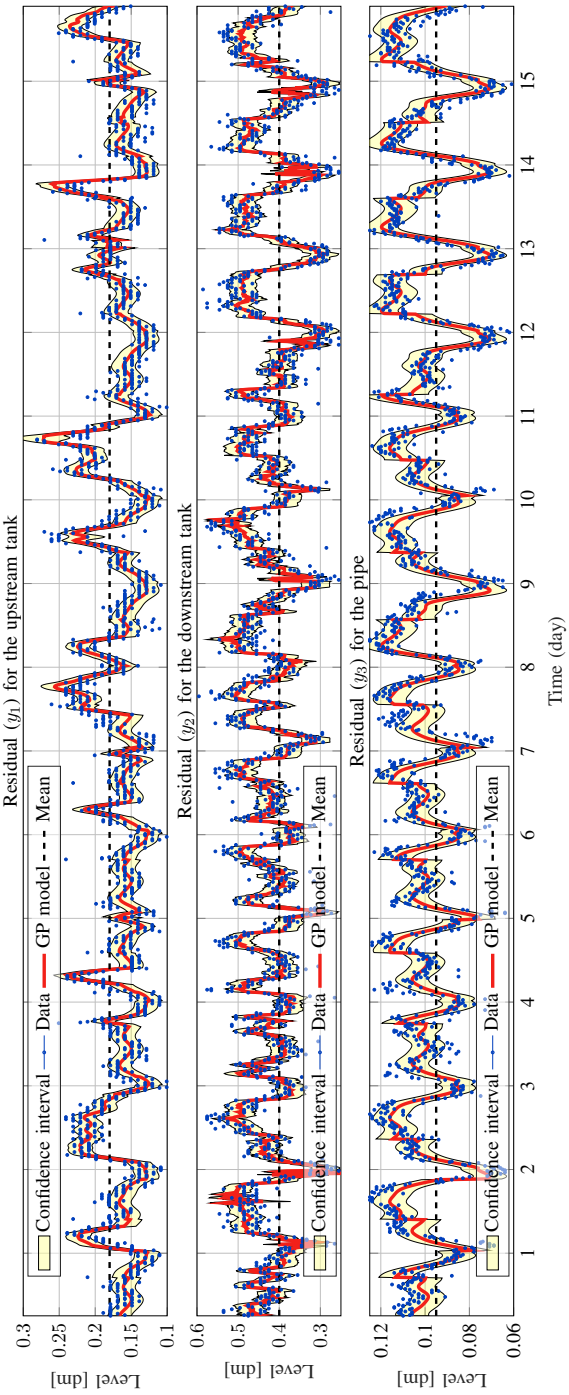


Figure 6: Validation of the GP model with the residuals regarding the level variations in the two tanks and in the sewer pipe, respectively.

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 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:

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

where $\overline{r}_i$ is the normalized relevance of the i^{th} predictor selected with the slicing matrices in eq.(27) 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 Figure 7.

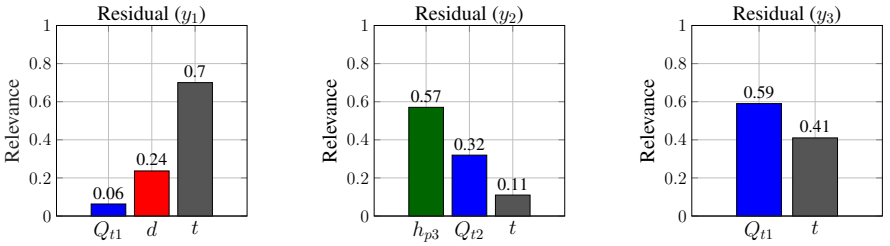


Figure 7: Relevance of the regressors showing the effect of the input data on the residuals.

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 $\mathbf{\Sigma}_{GP}$ 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 $\mathcal{D}$ training set with a criteria that these points need to be close to the previously 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 $\mathcal{D}$, 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. 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 Figure 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 $14-17$.

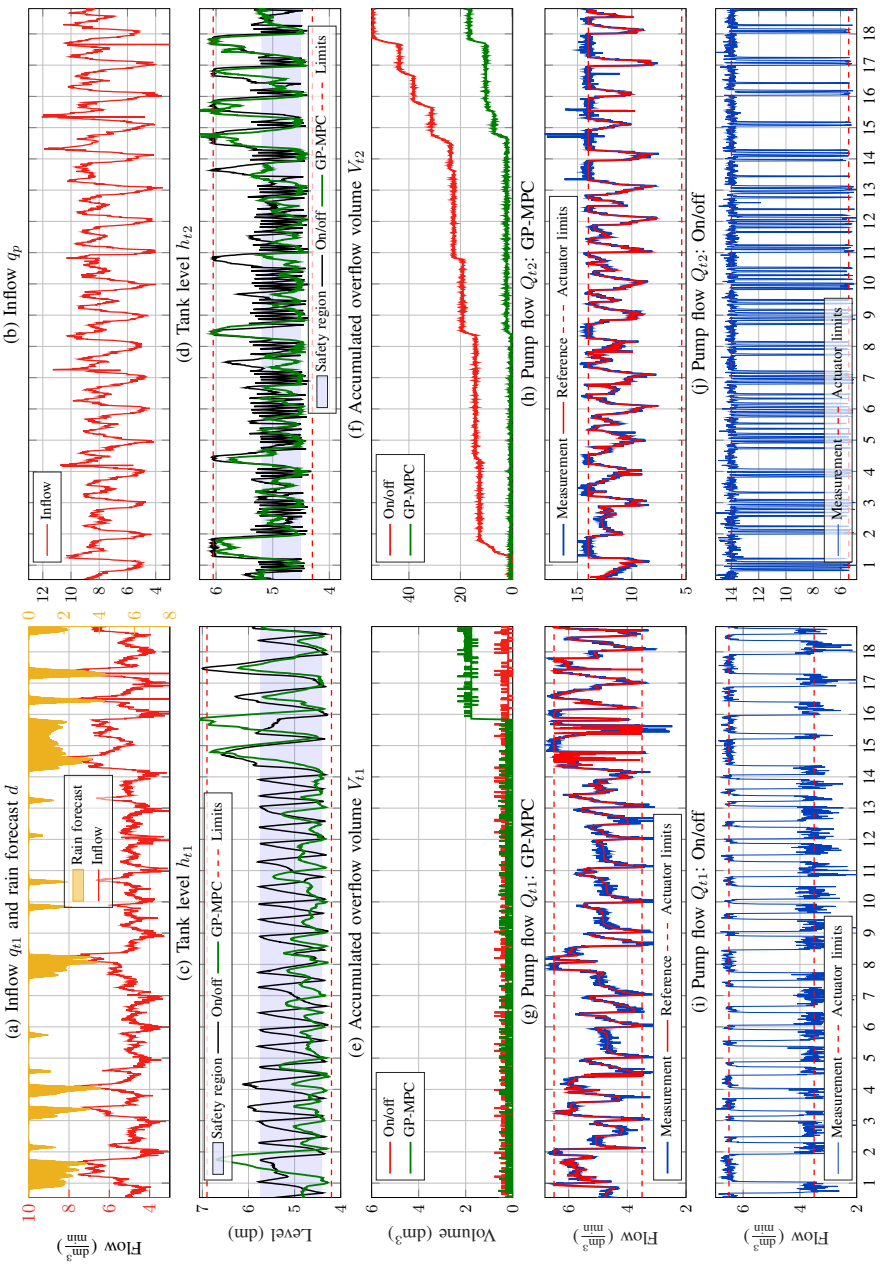


Figure 8: Performance comparison between the GP-MPC and On/off controller, operating the experimental setup representing the case study area during an 18-days period with heavy rain-events. Inflows, water levels in the basins, accumulated overflow and the control actions are shown regarding the upstream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

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 Figure 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 Figure 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 the points selected from the feature space $\mathcal{D}$ 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 indicates that even though our simple point selection with $M = 80$ points makes it possible to solve the optimization problem in under 2 seconds on average, the prediction quality and thereby the smoothness of the control action are significantly degraded. 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 Figure 8 (c-d)).

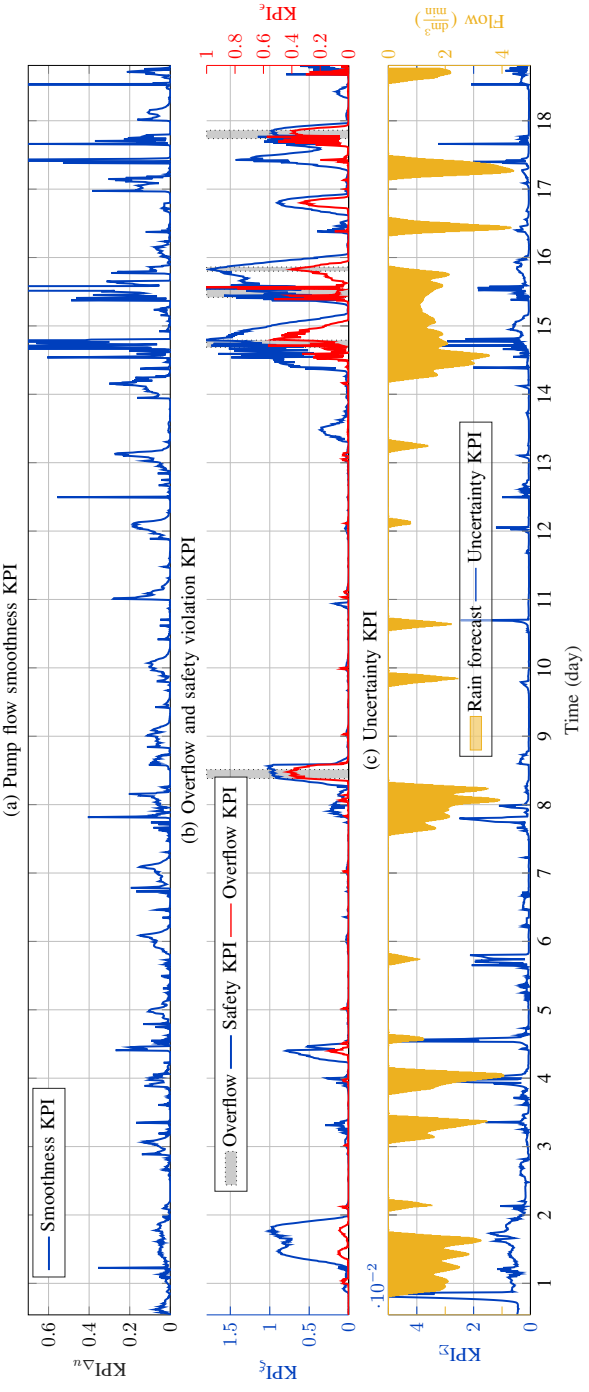


Figure 9: Key performance indicators provided by the user in run time, showing at each control step the smoothness of the control action, overflow, safety violation and uncertainty averaged over the prediction horizon.

The 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.

Acknowledgements

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

3S-Smart Software Solutions GmbH. “Codesys”. URL: https://www.codesys.com.

M. Ahm, S. Thorndahl, J. E. Nielsen, and M. R. Rasmussen. “Estimation of combined sewer overflow discharge: A software sensor approach based on local water level measurements”. *Water Science and Technology*, 74(11):2683–2696, 2016. ISSN 02731223. doi:10.2166/wst.2016.361.

J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. “CasADi: a software framework for nonlinear optimization and optimal control”. *Mathematical Programming Computation*, 11(1):1–36, 2019. ISSN 18672957. doi:10.1007/s12532-018-0139-4.

K. M. Balla, C. Schou, J. Dimon Bendtsen, and C. S. Kallesøe. “Multi-scenario model predictive control of combined sewer overflows in urban drainage networks”. In *2020 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1042–1047, Montréal, 2020. IEEE. ISBN 9781728171401. doi:10.1109/CCTA41146.2020.9206362.

K. M. Balla, C. Schou, J. D. Bendtsen, C. Ocampo-Martínez, and C. S. Kallesøe. “A Nonlinear Predictive Control Approach for Urban Drainage Networks Using Data-Driven Models and Moving Horizon Estimation”. *IEEE Transactions on Control Systems Technology (Early Access)*, pages 1–16, 2022. doi:10.1109/TCST.2021.3137712.

B. K. Banik, L. Alfonso, C. Di Cristo, A. Leopardi, and A. Mynett. “Evaluation of different formulations to optimally locate sensors in sewer systems”. *Journal of Water Resources Planning and Management*, 143(7):04017026, 2017. ISSN 0733-9496. doi:10.1061/(asce)wr.1943-5452.0000778.

T. Beeneken, V. Erbe, A. Messmer, C. Reder, R. Rohlfing, M. Scheer, M. Schuetze, B. Schumacher, M. Weilandt, and M. Weyand. “Real time control (RTC) of urban drainage systems - A discussion of the additional efforts compared to conventionally operated systems”. *Urban Water Journal*, 10(5):293–299, 2013. ISSN 17449006. doi:10.1080/1573062X.2013.790980.

A. Campisano, J. Cabot Ple, D. Muschalla, M. Pleau, and P. A. Vanrolleghem. “Potential and limitations of modern equipment for real time control of urban wastewater systems”. *Urban Water Journal*, 10(5):300–311, 2013. ISSN 1573062X. doi:10.1080/1573062X.2013.763996.

K. Chalupka, C. K. Williams, and I. Murray. “A framework for evaluating approximation methods for Gaussian process regression”. *Journal of Machine Learning Research*, 14(1):330–350, 2013. ISSN 15324435.

F. J. Chang, J. M. Liang, and Y. C. Chen. “Flood forecasting using radial basis function neural networks”. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 31(4):530–535, 2001. ISSN 10946977. doi:10.1109/5326.983936.

J. Chen, R. Ganigué, Y. Liu, and Z. Yuan. "Real-Time Multistep Prediction of Sewer Flow for Online Chemical Dosing Control". *Journal of Environmental Engineering*, 140(11):040140371–040140379, 2014. ISSN 0733-9372. doi:10.1061/(asce)ee.1943-7870.0000860.

Y. Cui, M. Jin, D. Li, Y. Xi, and L. Cen. "Iterative learning predictive control for urban drainage systems". In *Chinese Control Conference, CCC*, pages 4107–4112, 2015. ISBN 9789881563897. doi:10.1109/ChiCC.2015.7260272.

C. W. Dawson and R. L. Wilby. "Hydrological modelling using artificial neural networks". *Progress in Physical Geography*, 25(1):80–108, 2001. ISSN 03091333. doi:10.1177/030913330102500104.

A. P. Duncan, A. S. Chen, E. C. Keedwell, S. Djordjević, and D. A. Savić. "Urban flood prediction in real-time from weather radar and rainfall data using artificial neural networks". In *Weather Radar and Hydrology: IAHS Red Book Proceedings.*, pages 568–573, Exeter, 2012. ISBN 9781907161261.

S. Eggimann, L. Mutzner, O. Wani, M. Y. Schneider, D. Spuhler, M. Moy De Vitry, P. Beutler, and M. Maurer. "The potential of knowing more: A review of data-driven urban water management". *Environmental Science and Technology*, 51(5): 2538–2553, 2017. ISSN 15205851. doi:10.1021/acs.est.6b04267.

J. M. Grosso, C. Ocampo-Martínez, V. Puig, and B. Joseph. "Chance-constrained model predictive control for drinking water networks". *Journal of Process Control*, 24:504–516, 2014. ISSN 09591524. doi:10.1016/j.jprocont.2014.01.010.

L. Hewing, J. Kabzan, and M. N. Zeilinger. "Cautious Model Predictive Control Using Gaussian Process Regression". *IEEE Transactions on Control Systems Technology*, 28(6):2736–2744, 2020. ISSN 15580865. doi:10.1109/TCST.2019.2949757.

B. Joseph-Duran, C. Ocampo-Martinez, and G. Cembrano. "Output-feedback control of combined sewer networks through receding horizon control with moving horizon estimation". *Water Resources Research*, 51(10):8129–8145, 2015. ISSN 19447973. doi:10.1002/2014WR016696.

C. S. Kallesøe and T. Knudsen. "Self calibrating flow estimation in waste water pumping stations". In *2016 European Control Conference, ECC 2016*, pages 55–60, 2016. ISBN 9781509025916. doi:10.1109/ECC.2016.7810263.

O. Kisi, J. Shiri, and M. Tombul. "Modeling rainfall-runoff process using soft computing techniques". *Computers and Geosciences*, 51:108–117, 2013. ISSN 00983004. doi:10.1016/j.cageo.2012.07.001.

R. Kitchin. "The real-time city? Big data and smart urbanism". *GeoJournal*, 79(1): 1–14, 2014. ISSN 03432521. doi:10.1007/s10708-013-9516-8.

J. Li, K. Sharma, Y. Liu, G. Jiang, and Z. Yuan. "Real-time prediction of rain-impacted sewage flow for on-line control of chemical dosing in sewers". *Water Research*, 149:311–321, 2019. ISSN 18792448. doi:10.1016/j.watres.2018.11.021.

R. Löwe, S. Thorndahl, P. S. Mikkelsen, M. R. Rasmussen, and H. Madsen. "Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar". *Journal of Hydrology*, 512:397–407, 2014. ISSN 00221694. doi:10.1016/j.jhydrol.2014.03.027.

R. Löwe, L. Vezzaro, P. S. Mikkelsen, M. Grum, and H. Madsen. "Probabilistic runoff volume forecasting in risk-based optimization for RTC of urban drainage systems". *Environmental Modelling and Software*, 80:143–158, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.02.027.

N. S. V. Lund, A. K. V. Falk, M. Borup, H. Madsen, and P. Steen Mikkelsen. "Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management". *Critical Reviews in Environmental Science and Technology*, 48(3):279–339, 2018. ISSN 15476537. doi:10.1080/10643389.2018.1455484.

E. Mignot, H. Bonakdari, P. Knothe, G. Lipeme Kouyi, A. Bessette, N. Rivière, and J. L. Bertrand-Krajewski. "Experiments and 3D simulations of flow structures in junctions and their influence on location of flowmeters". *Water Science and Technology*, 66(6):1325–1332, 2012. ISSN 02731223. doi:10.2166/wst.2012.319.

A. L. Mollerup, P. S. Mikkelsen, and G. Sin. "A methodological approach to the design of optimising control strategies for sewer systems". *Environmental Modelling and Software*, 83:103–115, 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.05.004.

S. R. Mounce, W. Shepherd, G. Sailor, J. Shucksmith, and A. J. Saul. "Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data". *Water Science and Technology*, 69(6):1326–1333, 2014. ISSN 02731223. doi:10.2166/wst.2014.024.

A. Mullapudi, M. J. Lewis, C. L. Gruden, and B. Kerkez. "Deep reinforcement learning for the real time control of stormwater systems". *Advances in Water Resources*, 140:103600, 2020. ISSN 03091708. doi:10.1016/j.advwatres.2020.103600.

C. Ocampo-Martinez. *Model Predictive Control of Wastewater Systems*. Springer, Barcelona, 1st edition, 2010. ISBN 978-1-84996-352-7. doi:10.1007/978-1-84996-353-4.

D. Ochoa, G. Riano-Briceno, N. Quijano, and C. Ocampo-Martinez. "Control of urban drainage systems: Optimal flow control and deep learning in action". In *Proceedings of the American Control Conference*, pages 4826–4831, 2019. ISBN 9781538679265. doi:10.23919/acc.2019.8814958.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA, 2018. doi:10.7551/mitpress/3206.001.0001.

Y. A. Rjeily, O. Abbas, M. Sadek, I. Shahrour, and F. H. Chehade. "Flood forecasting within urban drainage systems using NARX neural network". *Water Science and Technology*, 76(9):2401–2412, 2017. ISSN 02731223. doi:10.2166/wst.2017.409.

M. R. Schütze, D. Butler, and M. B. Beck. *Modelling, Simulation and Control of Urban Wastewater Systems*. Springer, 2002. doi:10.1007/978-1-4471-0157-4.

C. Thrysøe, K. Arnbjerg-Nielsen, and M. Borup. "Identifying fit-for-purpose lumped surrogate models for large urban drainage systems using GLUE". *Journal of Hydrology*, 568:517–533, 2019. ISSN 00221694. doi:10.1016/j.jhydrol.2018.11.005.

X. Tian, R. R. Negenborn, P. J. van Overloop, J. María Maestre, A. Sadowska, and N. van de Giesen. "Efficient multi-scenario Model Predictive Control for water resources management with ensemble streamflow forecasts". *Advances in Water Resources*, 109:58–68, 2017. ISSN 03091708. doi:10.1016/j.advwatres.2017.08.015.

E. Todini. "Hydrological catchment modelling: Past, present and future". *Hydrology and Earth System Sciences*, 11(1):468–482, 2007. ISSN 16077938. doi:10.5194/hess-11-468-2007.

S. C. Troutman, N. Schambach, N. G. Love, and B. Kerkez. "An automated toolchain for the data-driven and dynamical modeling of combined sewer systems". *Water Research*, 126:88–100, 2017. ISSN 18792448. doi:10.1016/j.watres.2017.08.065.

J. Val Ledesma, R. Wisniewski, and C. S. Kallesøe. "Smart water infrastructures laboratory: Reconfigurable test-beds for research in water infrastructures management". *Water (Switzerland)*, 13(13):1875, 2021. ISSN 20734441. doi:10.3390/w13131875.

L. Vezzaro and M. Grum. "A generalised Dynamic Overflow Risk Assessment (DORA) for Real Time Control of urban drainage systems". In *the 9th International Conference on Urban Drainage Modelling*, Belgrade, 2014. doi:10.1016/j.jhydrol.2014.05.019.

V. K. Vidyarthi, A. Jain, and S. Chourasiya. "Modeling rainfall-runoff process using artificial neural network with emphasis on parameter sensitivity". *Modeling Earth Systems and Environment*, 6:2177–2188, 2020. ISSN 23636211. doi:10.1007/s40808-020-00833-7.

A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". *Mathematical Programming*, 106(1):25–57, 2006. ISSN 00255610. doi:10.1007/s10107-004-0559-y.

Y. Wang, C. Ocampo-Martínez, V. Puig, and J. Quevedo. "Gaussian-process-based demand forecasting for predictive control of drinking water networks". In *Proceedings of the 9th International Conference on Critical Information InfrastructuresSecurity (CRITIS)*, pages 1–12, 2014. ISBN 9783319316635. doi:10.1007/978-3-319-31664-2_8.

Y. Wang, C. Ocampo-Martinez, and V. Puig. "Stochastic model predictive control based on Gaussian processes applied to drinking water networks". *IET Control Theory and Applications*, 10(8):947–955, 2016. ISSN 17518652. doi:10.1049/iet-cta.2015.0657.

M. Xu, P. J. van Overloop, and N. C. van de Giesen. "On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow". *Advances in Water Resources*, 34:282–290, 2011. ISSN 03091708. doi:10.1016/j.advwatres.2010.11.009.

M. Xu, R. R. Negenborn, P. J. van Overloop, and N. C. van de Giesen. "De Saint-Venant equations-based model assessment in model predictive control of open channel flow". *Advances in Water Resources*, 49:37–45, 2012. ISSN 03091708. doi:10.1016/j.advwatres.2012.07.004.

Z. Yuan, G. Olsson, R. Cardell-Oliver, K. van Schagen, A. Marchi, A. Deletic, C. Urich, W. Rauch, Y. Liu, and G. Jiang. "Sweating the assets – The role of instrumentation, control and automation in urban water systems". *Water Research*, 155(1):381–402, 2019. ISSN 18792448. doi:10.1016/j.watres.2019.02.034.

Q. Zhang, F. Zheng, Y. Jia, D. Savic, and Z. Kapelan. "Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems". *Water Research*, 188(1):116544, 2021. ISSN 18792448. doi:10.1016/j.watres.2020.116544.

Paper G

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

Krisztian Mark Balla[a,b], Henrik Lemée[c], Christian Schou[b] and Carsten Skovmose Kallesøe[a,b]

[a]Grundfos Holding A/S, Poul Due Jensens Vej 7, DK-8850, Bjerringbro, Denmark
[b]Section for Automation and Control, Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark
[c]Ishøj Forsyning, Baldersbækvej 6, 2635, Ishøj, Denmark

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 proof of the proposed concept.*

Keywords—*Decision support, Smart Water systems, Data-driven modelling, Climate adaptation*

Accepted in the *Proceedings of IWA World Water Congress and Exhibition 2022*

Contents

1 Introduction

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].

2 Motivation and 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 and Play* solutions have a high impact in practice.

3 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.

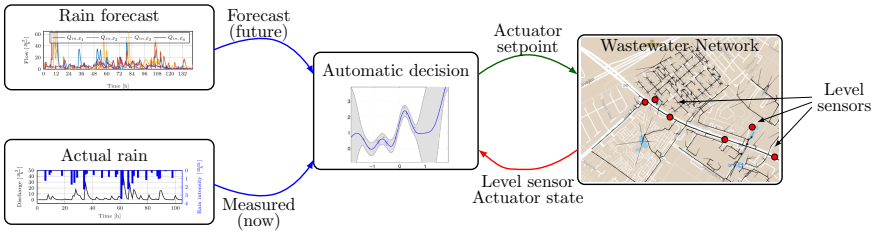


Figure 1: Closed-loop automatic decision support strategy using water level sensors and feedback from actuators.

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.

4 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 seconds measurement resolution, while the rain data has been obtained from the Danish Meteorological Institute's (DMI) service at a 1 minute 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:

- Water volumes accumulate downstream, increasing the risk of flooding in case of high intensity rain events and high sea levels downstream ,
- Without control, the capacity of Basin 1. is not utilized, therefore all volumes are bypassing and propagate downstream.

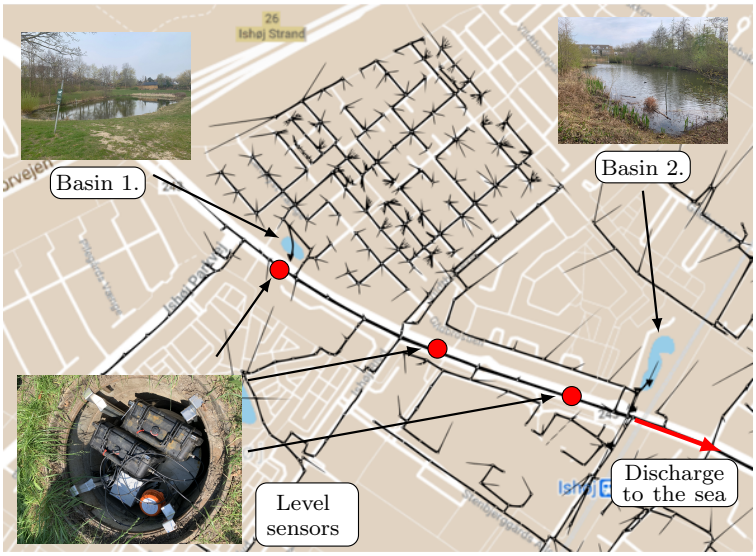


Figure 2: Storm water network in Ishøj, where red dots denote water level sensors.

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.

5 Results

A visualization interface has been developed on top of the Grafana time series visualization package [Grafana Labs, 2018] shown in Figure 3.
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 5. 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.

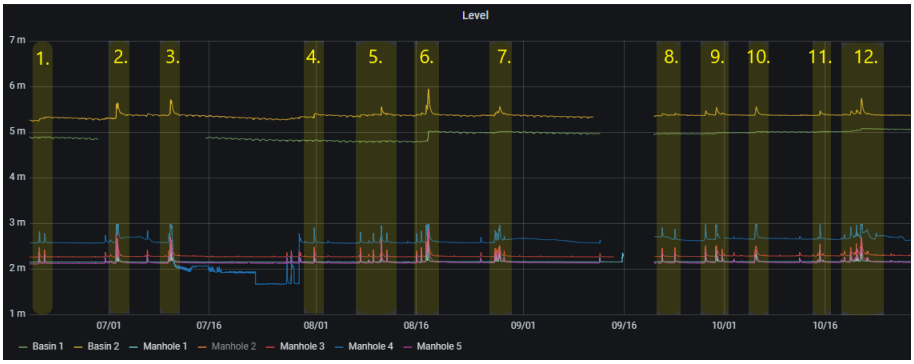


Figure 3: Grafana visualization interface developed for the case study.

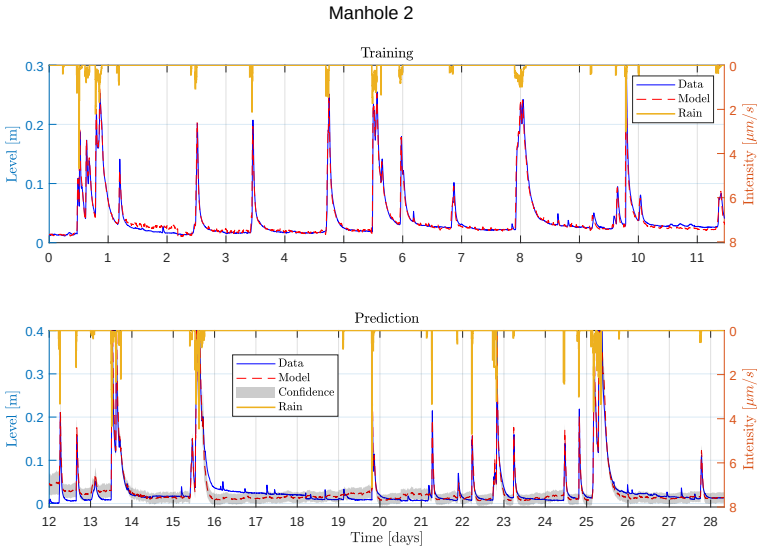


Figure 4: Example of model training and predicted water level response with characterization of the uncertainty based on the 12 rain events over a five months test period from 16-June-2020 to 27 October 2020.

6 Conclusions

The results show that the current solution is capable of predicting reasonable levels and uncertainty measures with solely using water level and historical rain gauge measurements as forecasts. (Note, that testing on real forecasts is anticipated as future work.) 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.

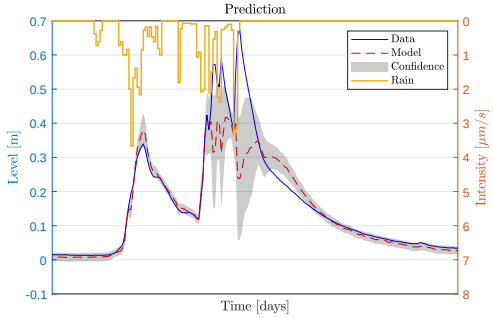


Figure 5: Model validation of the water level response with characterization of the uncertainty based on the 12 rain events over a five months test period from 16-June-2020 to 27 October 2020.

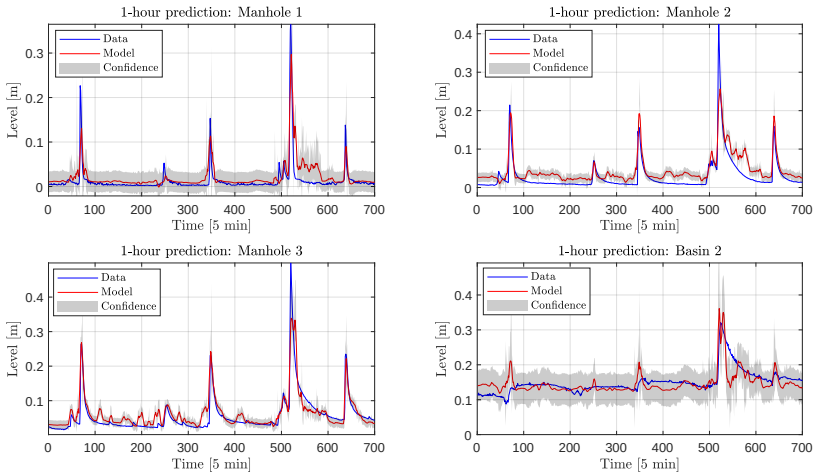


Figure 6: 1-hour predictions verified over the validation set for the rain event between Day 13 and 14.

References

S. Eggimann, L. Mutzner, O. Wani, M. Y. Schneider, D. Spuhler, M. Moy De Vitry, P. Beutler, and M. Maurer. "The potential of knowing more: A review of data-driven urban water management". *Environmental Science and Technology*, 51(5): 2538–2553, 2017. ISSN 15205851. doi:10.1021/acs.est.6b04267.

Grafana Labs. "Grafana documentation", 2018. URL: `https://grafana.com/docs/`.

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-876-2

AALBORG UNIVERSITY PRESS