Rainfall Forecast and Real-Time Control: Balancing Accuracy and Performance in Urban Flood Mitigation

Simen Hatlen

TU Delft

Rainfall Forecasts and Real-Time Control: Balancing Accuracy and Performance in Urban Flood Mitigation

By

Simen Hatlen

Master of science

In Civil Engineering

at the Delft University of Technology,

to be defended publicly on December 20th, 2024

Student name:	Simen Hatlen	
Student number:	5250994	
Thesis committee:	Dr. Ir. Job van der Werf	TU Delft
	Dr. Ir. Arjan Droste	TU Delft

Cover image: Overflow of Westersingel (Rotterdam) during excessive rain (Chris Zevenbergen, UNESCO-IHE)

An electronic version of this thesis is available at https://repository.tudelft.nl/

Acknowledgement

My time as a student at TU Delft has been an exciting and enlightening experience, providing new insights and knowledge related to water management issues and solutions. Throughout this journey, I have been privileged to learn from remarkable mentors, both during my thesis work and at the university in general. As Charles Dickens wrote[1], "*It was the best of times, it was the worst of times*" – a sentiment that aptly describes the process of working on this thesis. The frustration and hopelessness of troubleshooting a Python script that didn't work as intended stood in stark contrast to the sense of accomplishment, euphoria, and triumph when it finally did. These moments have made the experience not only intellectually enriching but also deeply transformative.

First and foremost, I would like to express my deepest gratitude to my supervisor, Job van der Werf, for his unwavering support, astute feedback, and invaluable guidance throughout the course of this thesis. I extend my thanks to Arjan Droste for the insights, contributions, and suggestions regarding the rainfall forecasts, which greatly improved the quality of my work. Their expertise and encouragement have been crucial in shaping this thesis.

I would also like to thank my mother and father for their unconditional support, motivation, and belief in me throughout this journey. Their patience in listening to my complaints about the programming struggles, always responding with compassion and the same reassuring words[2] – "*Du har aldri vært nærmere slutten enn nå*" – was a source of strength. Thanks to the rest of my family and friends, especially Endre and Martine, for their help with proofreading, suggestions, and grammatical corrections.

This thesis marks the conclusion of my time at TU Delft and my journey as a full-time student – a period of growth, learning, and transformation that I will carry with me into the future. Thank you to everyone who has made this experience both memorable and meaningful.

[1] Dickens, C (1859/2003). *A Tale of Two Cities*. Penguin Books

[2] You have never been closer to the finish line than now

Abstract

Urban stormwater management is put under pressure by increasing urbanisation and climate change leading to more frequent urban flooding. Such flooding can introduce ecological and public health issues by releasing contaminated water into the environment. Upgrading traditional grey infrastructure and implementing blue-green infrastructure (BGI) can be limited in their applicability for flood mitigation making Real-Time Control (RTC) an increasingly popular alternative due to its cost-effectiveness and potential performance benefits.

This thesis investigated the accuracy of rainfall forecasts and examined how forecast-informed RTC procedures could reduce flooding frequency and overflow depth while minimising negative side effects. The rainfall forecast accuracy was assessed based on three key properties: rainfall depth, forecast horizon, and mean forecast intensity. Using the insights from this assessment, three forecast-informed RTC procedures were designed to enhance the flood mitigation performance of a reactive RTC.

With perfect forecast data, flooding was completely prevented for two of the procedures, which used the longest but distinct horizons. The procedure using the shortest horizon was not able to prevent every flooding event due to the restriction in pre-emptive water release ahead of the forecasted rainfall events. The forecast accuracy analysis showed that the accuracy declined with increasing rainfall depth and lengthening horizon. Furthermore, the results indicated a shift from underestimation to overestimation of the rainfall depth as the mean forecast intensity increased. Applying the real forecast data, all designed forecast-informed procedures demonstrated reduction in total overflow depth of up to 70%, with the reduction linked to pump operation. However, the flood mitigation performance did not align with the expected results based on forecast accuracy, indicating that the procedures' logic and implementation played a significant role in determining their effectiveness.

The findings highlight the trade-offs inherent in using forecast-informed RTC procedures, particularly the balance between the uncertainties of longer forecast horizons and the need for sufficient lead time to take preventive action. By addressing these challenges, this thesis provides practical insights to inform the design and implementation of advanced RTC systems, marking a critical step toward more sustainable and resilient urban water management.

Table of contents

1 Introduction

1.1 Problem Analysis

Events where rainwater cannot enter the drainage system and pond on the streets are defined as *urban pluvial flooding*, simply called *pluvial flooding* since urban areas and cities are particularly susceptible to these situations (Jha et al., 2012; Palla et al., 2018; Rosenzweig et al., 2018). Many cities are already affected and are experiencing negative consequences of pluvial flooding. The risk and vulnerability of people and places are expected to get worse, according to the current projections of climate change, owing to increased peak runoff events due to changing rainfall regime (Schmitt & Scheid, 2020) and urbanisation (Ashley, 2007; Azizi et al., 2022).

Effects of climate change are already being experienced and will continue to cause significant stress in the future. The Netherlands is sensitive to climate change, and knowledge about the future climate is of great importance. Issues related to climate changes are further exacerbated in urban areas by the lack of permeable surfaces, limited drainage capacity, and ageing infrastructure. Uncertainties and variations are present in different climate scenarios; however, the same tendencies are observed (EEA, 2020; KNMI, 2014; Pörtner et al., 2022):

- Increase in mean temperature;
- Average precipitation and extreme precipitation will increase during the winter;
- Mean precipitation will decrease while the intensity of extreme precipitation will increase during summer;
- Sea level will continue to rise, and the rate will increase.

In addition to climate change, urbanisation intensifies the pressure on urban water management by reducing the infiltration potential due to construction of impermeable surfaces and straining the sewer network capacity. Changes in land use and land cover alter the natural water cycle through impervious surfaces such as asphalt, concrete, and brick, increasing the runoff volumes and decreasing the response time. Urbanisation and the concentration of population and wealth put pressure on existing drainage systems and increase the likelihood of them being overwhelmed. The urbanisation has been rapid and extensive, causing an unsustainable urban development that greatly exacerbates the susceptibility to flood risks (Acosta-Coll et al., 2018;

Jian et al., 2021). Therefore, inhabitants are more often exposed to floodwater inflicting health and safety issues, while buildings and properties can be subjected to significant damage.

Notwithstanding the evident negative effects of urban pluvial flooding and expected exacerbation, less attention has been given to pluvial flooding in research compared to other types of flooding. Reasons for the lack of focus are expanded on by Rosenzweig et al. (2018). Firstly, urban pluvial flooding has been taken for granted and viewed as a problem with pre-existing solutions and has been assumed to occur due to failure in the drainage system. Secondly, difficulty in predicting urban areas' hydrological response to precipitation events and lack of observational data has restricted the focus. The reliability and accuracy of available data is questionable, making it difficult to assess the impacts and occurrence of pluvial flooding. Thirdly, urban pluvial flooding is assumed to only induce limited impacts, fostering little attention. These repeated and systematic flooding events are often left out of the assessments despite the importance of the cumulative impacts. However, there are multiple examples of how urban flooding can cause more severe impacts such as contaminant and pathogen exposure, property damage, and disruption of transportation networks (Falconer et al., 2009).

Although there has been a lack of recognition in the scientific literature, in recent years urban pluvial flooding and available mitigation methods have started to receive more attention. One of the most common methods for mitigating urban flooding is upgrading traditional grey infrastructure (Chen et al., 2021), which includes updating storage facilities, pipe networks, and pump stations. The performance can be further improved by supplementing grey infrastructure with additional measures, resulting in more sustainable practices. One of these practices is blue-green infrastructure (BGI), such as water retention basins (Robinson et al., 2010) and green roofs (Stovin et al., 2012). BGI aims at restoring the natural water cycle and is contributing to address flooding challenges (Casares et al., 2024).

However, financial constraints, including costs related to implementation and upgrading the system, but also growing operational and maintenance costs, can make these two methods inapplicable. Similarly, the methods can be unsuitable due to spatial constraints, either not being sufficient space available or not being feasible to implement at the specific location. Moreover, the static control grey infrastructure and BGI offers may become insufficient and reduce the flood mitigation performance due to the anticipated increased pressure on urban water systems due to climate change and urbanisation. Therefore, there are clear challenges related to the use and improvements of grey infrastructure and BGI in existing drainage systems. Butler and

Davies (2011) emphasise that the methods need to be cost-effective and acceptable technical improvements. In addition, the impacts must be assessed while continuing to search for sustainable solutions. This highlights the necessity for other methods, with Real-Time Control (RTC) becoming a new, sustainable, and alternative solution for flood mitigation.

Generally, in stormwater engineering, adaptation to mitigate disturbances and limit loss of functionality is done through resilience design aimed at minimising flood duration, magnitude, and impacts (Li & Burian, 2023). RTC is one method for improving the resilience of infrastructure against the uncertain future conditions, and in recent years, it has gained traction as a method to prevent urban pluvial flooding (Kändler et al., 2020). Additional benefits of implementing RTC can include reduction of combined sewer overflows (CSOs), reduction of peak flows, diminishing urban flood volume, and improving water and stormwater quality (Beeneken et al., 2013; Borsányi et al., 2008). These advantages emphasise RTC's potential as a transformative approach for enhancing urban water systems and addressing the challenges posed by increasingly variable conditions.

An RTC approach enhances flexibility and adaptability, by treating the urban drainage system (UDS) as a dynamic entity, capable of handling the highly variable loading conditions it faces. An RTC system utilises real-time data about the urban water system and possibly rainfall predictions to optimally operate existing infrastructure: ensuring adequate capacity within the system by proactively managing space within the infrastructure. Given that RTC does not require additional static investments and only entails minor expansions of the pre-existing sensor network within an urban water system, it can be considered a cost-effective solution compared with more traditional methods (Sun et al., 2023; Xu et al., 2022). By implementing and operating RTC strategies effectively, both the storage capacity of the UDS and urban surface water alike can be utilised more efficiently. For heavy rainfall events, RTC can potentially mitigate urban flooding by selective discharge of water before the onset of forecasted rainfall, thereby increasing the urban water system's water retention capacity for the event. This form of proactive RTC relies on information about the future storm events, provided through rainfall forecasts (Stinson, 2005; Sun et al., 2023). Combining rainfall predictions with real-time observations of the urban water system enables informed control strategies to be implemented.

Although combining real-time observations of the urban water system with rainfall forecasts can theoretically enhance the performance, the accuracy of precipitation forecasts can have significant impact. Parameters such as localisation, timing, and intensity play crucial roles in influencing the effectiveness of RTC relying on forecast data (Xu et al., 2020). Weather forecast's ability to accurately predict future weather states decreases with an increasing forecast horizon (Imhoff et al., 2022). Consequently, finding a balance between acting with ample lead time (using longer forecast horizons) and accurately predicting rainfall (using shorter horizons) becomes necessary. Understanding this trade-off is contingent on the specific case study, yet a systematic evaluation of these trade-offs is currently lacking in the existing scientific literature.

1.2 Objective

The aim of this thesis is twofold: (1) to assess the quality and accuracy of rainfall forecasts; and (2) to explore the potential of mitigating urban pluvial flooding using forecast-informed RTC. Forecast-informed RTC strategies are developed based on the quality and accuracy of the rainfall forecasts, with the objective of temporarily reducing the surface water levels to accommodate larger rainfall depths and mitigate urban pluvial flooding. The inherent uncertainties of the forecast data are expected to influence the effectiveness of RTC systems. Therefore, a balance between the forecast-related uncertainties, the flood mitigation performance of RTC, and potential negative side effects is critical. To address these aims, the following research question and sub-questions are formulated:

Can a Real-Time Control strategy be designed to mitigate urban flooding and minimise negative side effects?

1. What is the current frequency of flooding?
2. Can a Real-Time Control strategy mitigate urban flooding?
3. How does forecast accuracy affect the Real-Time Control strategy performance?
4. Can the Real-Time Control strategy be adjusted to minimise the risk while maximising the benefits?

1.3 Report Structure

The thesis begins with Chapter 2, which provides a literature review on urban pluvial flooding, Numerical Weather Prediction (NWP), as well as RTC, including design options and reported performance. Chapter 3 outlines the study area, and the rainfall datasets used in the analyses.

The thesis is divided into two distinct parts. (1) PART I focuses on the rainfall forecast assessment and includes Chapters 4, 5, and 6. Chapter 4 describes the forecast properties assessed and the evaluation metrics applied. Chapter 5 presents the results of the forecast quality assessment. Chapter 6 concludes PART I by discussing the results and implications of incorporating forecasts into an RTC strategy. (2) PART II focuses on the development and evaluation of RTC strategies and begins with Chapter 7, which introduces the RTC strategies and the methods for performance assessment. Chapter 8 presents the results of the RTC strategy evaluation. Chapter 9 concludes PART II by discussing the RTC performance and influencing factors.

Chapter 10 presents an overall discussion of the project, highlighting the limitations and weaknesses of the developed methodology. Finally, Chapter 11 summarises the findings from both PART I and PART II, answering key research questions, and provides a discussion of the results.

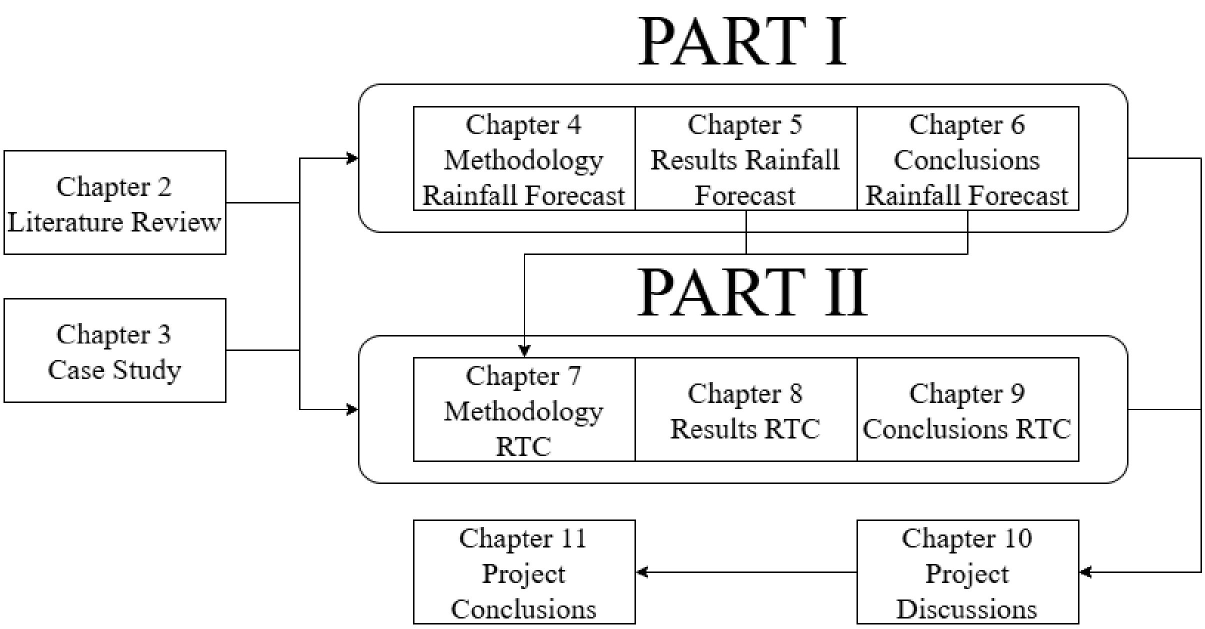

Figure 1.1: Overview of thesis structure. Chapter 2 and 3 creates the foundation for both PART I and PART II. The results and conclusions of PART I are utilised in Chapter 7. Thesis discussions and conclusions combining PART I and PART II are presented in Chapter 10 and 11.

2 Literature Review

This chapter reviews the relevant literature to provide insight into the principal concepts, methods, and challenges associated with applying forecast-informed RTC for flood mitigation. The review is structured to address three primary areas: (1) urban pluvial flooding, (2) NWP, and (3) RTC systems. These areas are examined to identify knowledge gaps, establish the theoretical context, and explore improvements that inform this thesis.

Section 2.1 presents urban pluvial flooding and its associated impacts, emphasising challenges in determining and comparing severity and impacts. Section 2.2 examines NWP as a forecasting technique, with particular attention on accuracy and events that are more difficult to forecast. Lastly, Section 2.3 explores RTC systems, evaluating their design, implementation, and reported performance in urban water management.

The review aims to establish a comprehensive understanding of the existing body of work while identifying critical areas where further research is needed. This forms the basis for the methodology and analyses presented in subsequent chapters.

2.1 Urban Pluvial Flooding

In most cities and urban areas, the frequency and severity of flooding has increased over the past century. Urban flooding can be the result of several sources. Extreme precipitation localised over parts of the urban area and subsequent runoff or high flows in major neighbouring rivers generating urban flooding are common (Ashley, 2007). This thesis will focus on the urban flooding events triggered due to local precipitation that exceeds the urban surface water system's capacity.

Slightly different terms have been used to describe and characterise flooding in urban areas due to precipitation. Examples of the differences can be seen in Falconer et al. (2009), Rosenzweig et al. (2018), and Brendel et al. (2020). Despite the discrepancies in literature, usual characteristics include precipitation being the cause, a lack of capacity in the drainage system for rainwater to enter, and subsequent ponding on the streets. However, some literature differentiates between pluvial flooding and surface water flooding, with the latter including flooding from open-channels and urban watercourses. In this thesis, the term urban pluvial flooding will refer to flooding caused by overflowing canals. Moreover, attempts to categorise and differentiate flooding events in the Netherlands based on the potential impacts and severity has been conducted by RIONED (2006). Three severity degrees of pluvial flooding are used:

1. Hindrance: small quantities of water on the street with a duration of 15-30 minutes with no or minimal property damage.
2. Severe hindrance: large volumes of water on the street. Can cause tunnels to flood and manhole covers to lift, 30-120 minutes duration.
3. Nuisance: water on the street on a larger scale and longer duration. Risks include flooding in shops, damage to property and possibly serious disturbance to (economic) infrastructure.

The first two categories primarily cause inconvenience and traffic disruption, as the water remains confined to streets without causing significant property damage. Disruption of day-to-day activities due to flooding of roads, pavements, cycle paths, and railways falls under the hindrance and severe hindrance categories. Nuisance, defined as the most severe category, includes property damage beyond what is accounted for in the less severe categories. However, there is no clear quantitative difference between the suggested categories, raising questions about their practical utility. Additionally, they focus predominantly on small, frequent events, overlooking the increasing frequency and intensity of extreme rainfall projected by climate models (EEA, 2020). This focus may be inadequate for addressing future flood risks. Consequently, more severe impacts on property, the environment, and public health are anticipated, particularly in urban areas with aging infrastructure.

2.1.1 Pluvial Flood Impacts

Multiple methods to distinguish and differentiate between flood impacts have previously been utilised. Firstly, impacts can be classified as either tangible or intangible (Velasco et al., 2016). This is related to whether the impact can be expressed in monetary value (tangible) or not (intangible). Secondly, differentiation between direct and indirect impacts is utilised relating to how a flooding event is experienced. Direct impacts are caused by immediate physical contact of flood water, whilst indirect impacts relate to losses outside the flooded area (Martínez-Gomariz et al., 2020). Thirdly, impacts can be distinguished into exposure and vulnerability. Exposure includes the amount of people and assets directly impacted by the event, while vulnerability can be described as the severity of the experienced impacts (Rosenzweig et al., 2018). Combining the probability of flooding and potential consequences results in a flood risk that can be assessed (European Commission, 2007; Vergouwe, 2016).

There are three primary ways in which pluvial flooding affects cities: through its impact on the population, on the economic sector, and on physical infrastructure (Zevenbergen et al., 2011). Pluvial flood impacts caused by infrequent, extreme local precipitation events were investigated by van Riel (2011), leading to the creation and definition of several categories of pluvial flood impacts, which were later adapted. Similar attempts to categorise flood impacts have been made (e.g., Moftakhari et al., 2018). Common for the different attempts is that they all include economic losses due to damage to infrastructure, property, lower productivity, failure of services, threats to public health, and general discomfort and nuisance for citizens. The distinction between the flood impact categories and their definitions, from van Riel (2011), are used, and the relationship between impacts and stakeholders is shown (Table 2.1).

- Material impacts: Damage to physical objects triggered by direct contact with rainwater.
- Economic impacts: Induced costs due to changes or interruptions in economic activities or productive capacity.
- Health impacts: Include impacts to physical health from either direct contact with floodwater or the clean-up process. Can also encompass mental health impacts resulting from the flooding experience.
- Discomfort: The total inconvenience and nuisance resulting from a combination of multiple pluvial flood impacts.

Stakeholder	*Impact categories*
Population	Material impacts, health impacts, and inhabitants' discomfort
Economic sector	Material impacts, economic impacts
Physical infrastructure	Material impacts, economic impacts, and inhabitants' discomfort

Table 2.1: Stakeholders and associated impact categories.

2.1.1.1 Material Impact

Pluvial flood is characterised by small water depth and small direct damage. Although pluvial floods usually have minimal material impacts, the cumulative material risk and cost can become large over time. Due to the frequent nature of these events the costs can exceed those of less frequency with more severe impacts (Rosenzweig et al., 2018). Material impacts caused by pluvial flood in lowland areas mainly involve cleaning expenses and the occasional replacement of carpeting on the ground flood. ten Veldhuis (2011) reported that in cases where the water depth reaches 10 cm, the material damage per flooded building can range from €1000 to €30,000 for residential, commercial, and public buildings, but seldom surpasses €3500/household. This information was derived from the low ranges of stage-damage functions from previous studies in Germany and the Netherlands. The cost relates to cleaning and repair costs. Additionally, a slightly higher minimum (€2000) was assumed for commercial and public buildings due to higher cleaning costs. Nevertheless, other impacts, like temporary road closure, disruption in transit service, and inconvenience for pedestrians, are more common and are important for small events (Moftakhari et al., 2018).

2.1.1.2 Economic Impact

Economic risks are considered as indirect impacts, as they can represent lost opportunities. Examples of economic risks include interruption of business activity, and disruption and delay in traffic. Pluvial flooding can heavily impact communities, businesses, and citizens. Several studies estimating economic risks have been conducted. An attempt to quantify damages in monetary values and number of affected people was carried out by ten Veldhuis (2011), demonstrating that the economic risk can be significant. Whilst the monetary value of flooding and ponding of roads are significantly lower than the material risk, a very high number of people are affected. Similarly, Martínez-Gomariz et al. (2019) assessed the economic risk through business interruptions to the economic sector. It was found that economic risks were approximately 30% of the material risks in terms of monetary value.

2.1.1.3 Health Impact

The health impact category can be divided into physical and mental health effects. Physical effects include mortality, injuries, and illness related to the event itself or the clean-up process. Mental health impacts, which occur due to the experience of being flooded, can include stress, depression, and PTSD (Fewtrell et al., 2008; Hammond et al., 2015). Physical injuries, whether minor or major, can occur from stumbling or falling due to objects hidden in the water. Additionally, when the capacity of combined sewer systems is exceeded, contaminated water can overflow onto the surface. This overflow, a mixture of waste and stormwater, might contain bacteria and contaminants that can cause illness (Moftakhari et al., 2018; Sterk et al., 2008).

2.1.1.4 Discomfort

In addition to material, economic, and health impacts, pluvial flooding can cause significant discomfort for inhabitants. This type of impact is individual and intangible, making it challenging to specify and quantify. Discomfort encompasses the overall perceived and experienced stress, inconvenience, hindrance, and nuisance affecting cars, cyclists, and pedestrians. Factors such as age, gender, education, knowledge, and personal experience play crucial roles in shaping the perceived and experienced discomfort. Furthermore, the level of discomfort can be influenced by public warnings, as well as the communicated probability and severity of the flood (Netzel et al., 2021).

While the severity is categorised based on duration and water quantities, impacts are differentiated by whom they affect within the cities and on the type of consequences they cause. However, the challenges of determining and comparing the severity and impacts of floods remain an issue. This thesis focuses on small and frequent events, which are typically characterised by small depths, thereby limiting the tangible impacts. Despite this, the intangible impacts can be significant, especially for low-return-period rainfall events, as they contribute to the overall inconvenience and discomfort.

Quantification of flood-related impacts and severity is beyond the scope of this thesis. Nevertheless, the projected increase in flooding emphasises the need to consider associated risks and implement potential mitigation measures. The literature stresses that while high-intensity, low-frequency rainfall events result in more severe impacts, it is the low-intensity, high-frequency events that often lead to considerable cumulative costs and negative consequences over time. Public acceptance of these frequent flooding events depends not only on their frequency but also on the ability to provide adequate warnings. Therefore, for maintaining public trust and confidence in governmental actions, improving and implementing rainfall forecasts for reducing the flooding frequency through mitigation measures are essential.

2.2 Numerical Weather Prediction

NWP is one of the most common forecast methods available. The method computes and predicts the future weather using the current atmospheric state. Using partial differential equations (PDEs) describing the conservation laws, the future characteristic values of the atmosphere can be estimated. Two conditions are necessary to successfully predict the upcoming weather (Coiffier, 2011; Pu & Kalnay, 2018):

I. Accurate characterisation and knowledge of the atmosphere at the initial state.
II. Sufficient and accurate knowledge of the physical laws from which the state of the atmosphere develops.

The future state of the atmosphere is calculated in two separate stages using a numerical model. The continuous behaviour of the atmosphere is regulated by establishing a system of PDEs. The equations are discretised before they are analytically solved with an appropriate algorithm (Coiffier, 2011). The discretisation of PDEs introduces some error in space and time, a truncation error. This truncation error combined with the chaotic nature of the atmosphere can make small initial errors grow rapidly and deteriorate the forecast quality in a very short time (Buizza, 2019).

However, advances in NWP and the introduction of ensemble prediction have improved the quality and accuracy of forecasting (Buizza, 2019). Forecast ensembles provide a range of possible future scenarios instead of a single deterministic forecast, offering probabilistic estimates into event occurrence. While large variations among ensemble members complicate categorical forecasting, high agreement allows for reasonably confident predictions. This approach provides the user with more comprehensive and valuable information, as the users are supplied with the outcome ranges and associated probabilities. By enabling confidence-based responses, binary decisions which rely on thresholds can support adaptive management strategies and help mitigate risks while avoiding unnecessary measures during lower-probability events.

Since the introduction of ensemble prediction, the forecast skills have increased. This can be proved objectively and quantitatively by comparison of the forecasted weather and the actual weather that occurs (Bauer et al., 2015). Forecast skill is measured using the anomaly correlation coefficient (ACC), which quantifies deviations of forecasts from the climatological average state of the atmosphere. The historical mean values are subtracted from both the forecasted and observed data, and the correlation formula determines how closely the anomalies align. This calculation is performed at the 500 hPa level. Higher ACC values indicate better performance, with values above 60% typically considered skilful and outperforming random predictions (Simmons & Hollingsworth, 2002).

However, the overall forecast skill is dependent on both the forecast horizon, shorter horizons increasing the accuracy (Figure 2.1), and the type and scale of the event, with small and intense events (convective rainfall) difficult to predict due to the necessary resolution (Bauer et al., 2015; Imhoff et al., 2022).

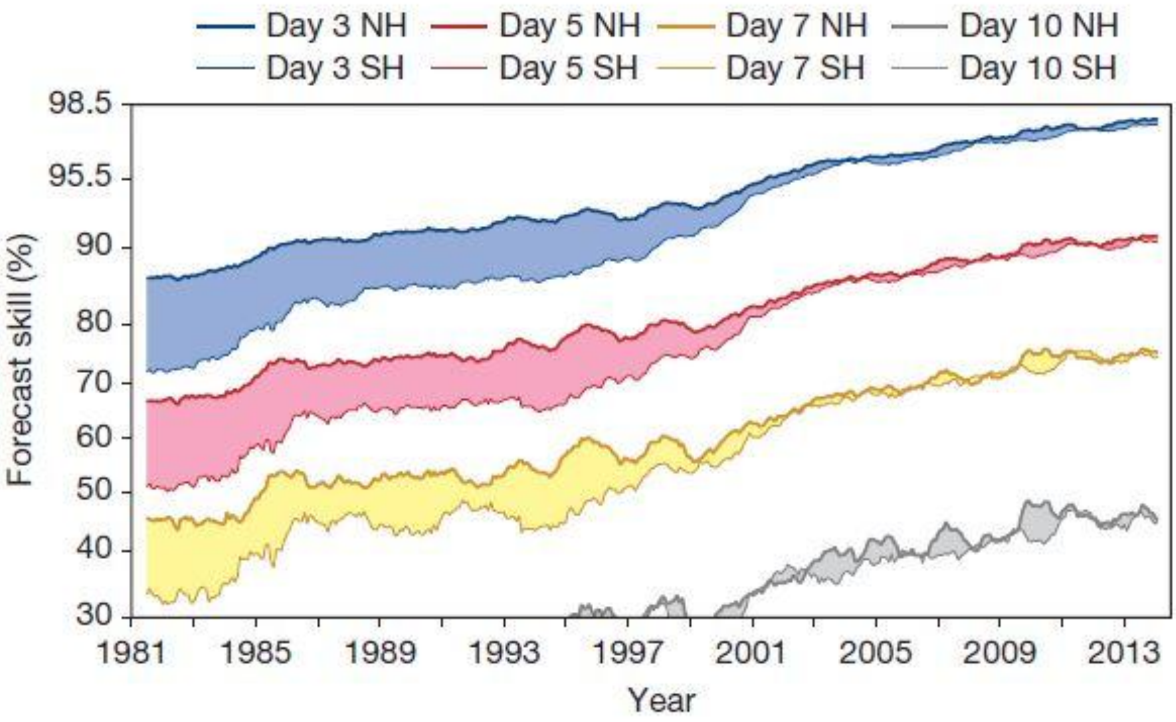

Figure 2.1: Evolution of forecast skill at 3-, 5-, 7- and 10-day ranges for northern and southern hemisphere (Bauer et al., 2015)

Urban flooding is most often caused by these convective rainfall events (Walczykiewicz & Skonieczna, 2020). The inferior forecast skill of these events may limit the potential for flood mitigation when incorporated into the decision-making process. As the horizon lengthens the uncertainty increases, and it becomes important to determine the optimal forecast horizon before including forecasts in flood mitigation strategies. Ashok and Pekkat (2022) reported that the optimal forecast horizon for many flood warning and forecasting studies is between 6 and 48 hours. Imhoff et al. (2022) noted that most early warning systems for floods use NWP models with 12- to 72-hour horizons. While improvements in rainfall forecasts can lead to better hydrological predictions and early warning systems, Jabbari et al. (2020) found that the forecast performance does not significantly depend on the horizon when it is below 36 hours. However, despite the inherent forecast errors and horizon-dependence, inclusion and consideration can reinforce the management strategies (Gaborit et al., 2013).

2.3 Real-Time Control

Changes in rainfall patterns and urbanisation magnify the vulnerability of infrastructure, people, the environment, and property. The aim of water management is to minimise negative impacts in terms of both quantity and quality (Beeneken et al., 2013; Jacobs, 2012). Improvements and incorporation of rainfall forecasting techniques into management strategies may negate impacts associated with increased flooding. Vis et al. (2003) compared raising the dikes with two alternative resilience strategies for flood prevention in the Netherlands and found that the alternative strategies offered greater flexibility. One example of an alternative and resilient strategy is RTC, which has gained popularity for flood protection, due to its improved adaptability, flexibility, and cost advantages with minimal efforts (Kändler et al., 2020).

The literature of RTC mostly is focused on only wastewater volumes in UDS (García et al., 2015). However, there are differences between RTC in UDS and in surface water systems, which can include system complexity and infrastructure constraints. UDS face challenges managing both stormwater and wastewater, balancing flood prevention and overflows to the environment. Research (Bilodeau et al., 2018; Gaborit et al., 2013) have shown that implementing RTC for outflow from detention basins increased performance, and it can be assumed that RTC applied to an urban surface water system may exhibit somewhat similar behaviour in terms of their functioning in flood management. Although most studies focus on RTC implemented in UDS, the same principles can be adapted for surface water control, offering flexibility and efficiency in flood prevention strategies. Consequently, the following subsections primarily address RTC in UDS, reflecting the dominant area of research, while recognising its broader potential.

2.3.1 RTC Design Options

It has been demonstrated that RTC does work in practice through many large-scale studies (Schütze et al., 2004). RTC utilises current information about the UDS to optimally operate existing infrastructure to reach the established objective. However, the performance and efficacy are influenced by the implementation. Implementation of RTC can be divided into three steps (van der Werf et al., 2022). The first step involves establishing the overarching RTC strategy, which includes defining the objective function, designing the architecture, and determining the potential positioning of new actuators. The second step focuses on the RTC procedure, which identifies the optimal settings within the system. Finally, the third step concerns the RTC algorithm, which dictates how the optimal settings are implemented by the actuators.

The RTC performance refers to its capacity to enhance system functionality in accordance with the established objective. The RTC performance can be evaluated using either a data or model driven approach (van Daal et al., 2017). A model-based method entails running simulations of the system using different rainfall events with and without implemented RTC rules. For the data-driven method, datasets with sufficiently comparable characteristics are gathered with and without RTC implemented to determine the impact on the system's functioning. Moreover, the RTC performance can be divided into three types (van der Werf et al., 2022). The actual improvement possible for an implemented RTC strategy, which can only be assessed using data-driven methods, defined as the *Implemented RTC Performance.* The *Theoretical RTC Performance* can be calculated using a model-based method. The theoretical RTC performance assumes ideal functioning of the RTC and can be regarded as the upper-bound of the practical performance. Last is the *Maximum Potential Performance* defining the absolute upper limit of what could be achieved with any RTC strategy for the studied catchment. There can be considerable difference between the theoretical and practical performance, and studies should explicitly mention what type of performance was assessed.

The architecture of the RTC defines how the actuators are controlled and implemented and can be designed with local, central, or distributed control. The difference relates to how the information is processed, handled, and actions determined. The simplest form is local control where the measurements are obtained directly at the actuator site. There is no communication between the actuators, and each is responsible for optimising its relevant part of the UDS and consequently the whole system. Because of its simplicity, local control may represent a suitable solution in cases with few actuators in the system. For more complex systems or if all actuators must be operated jointly, a centralised control is required. The measurement data from local sensors are communicated to a central control room and actuators are operated in a coordinated manner (Schütze et al., 2004). However, the largest potential for centralised control comes from large and complicated UDS (van der Werf et al., 2022). For large-scale and complex systems, distributed control, a combination of local and central control, is common. Communication between the actuators is incorporated, rather than a central controlling agent. Groups of actuators find the best solution for each actuator. For each subsection throughout the system the actuators are operated utilising only the information in this section (García et al., 2015). The circumstances, benefits, and drawbacks of the different control architectures should determine which one to use.

Selection of an appropriate control procedure is essential for developing an RTC for a wastewater system. The RTC procedure refers to how the system determines the optimal settings or setpoints for the actuators (Schütze et al., 2004). Procedures can be categorised into heuristic and optimisation-based control (García et al., 2015). A heuristic procedure relies on extensive knowledge and experience of the system's behaviour. Potential procedures and setpoints are specified and refined through an iterative process involving a simulation model of the system. In contrast, optimisation-based procedures evaluate the impacts of potential control actions and compute the best settings for the actuators at each specified control time interval. This approach allows for high accuracy in predicting the future state of the system due to model updates based on current measurements.

A distinction can be made between reactive and predictive control based on how an RTC strategy utilises information and determines the actuators' setpoints. Predictive control has several possible approaches, ranging from simple heuristic controls incorporating rainfall forecasts (van der Werf et al., 2023) to more complex techniques like Model Predictive Control (MPC). These advanced systems integrate multiple features, including online models and weather forecasts, to adapt water systems based on real-time and predicted conditions (Meneses et al., 2018). Rainfall forecasts can provide valuable information assisting in improving system performance by emptying water from storage facilities before a storm (Sun et al., 2023). However, the RTC performance can be affected by the forecast accuracy and reliability, which tend to deteriorate with increased lead time (García et al., 2015). In contrast, reactive control responds only to current external events. These systems are simpler, and the benefits of upgrading from reactive to predictive control need to be clearly identified to justify the added complexity and costs.

Studies comparing different RTC designs and approaches have shown that both simple and complex procedures can achieve the desired objective. When both types of procedures are effective, preference is frequently given to the simpler approach (Kroll et al., 2018). Performance evaluation typically involves optimisation to identify the best possible strategy and procedure. Although an optimal solution may not be achievable within the given constraints, a suboptimal control decision can still be effective RTC. The key is to ensure that the chosen control decision does not cause worse performance than the no-control scenario

(Schütze et al., 2004). Despite the challenges associated with RTC, it can be a valuable method for enhancing the performance of sewer and water systems.

2.3.2 RTC Benefits

RTC can improve the performance of existing UDS. Despite its usefulness and cost-efficiency, implementing RTC procedures remains challenging for wastewater operators (Kroll et al., 2018). This difficulty arises because RTC is often perceived as too complex and expensive to implement without case-specific investigation. However, compared to conventional systems, the additional expenses of RTC are relatively limited, and it offers considerable benefits (Beeneken et al., 2013). These benefits typically fall into one of three categories: volume-based, pollution-based, or impact-based (van der Werf et al., 2022). A volume-based RTC strategy, for instance, aims to reduce the total CSO or flooding volume, while pollution-based and impact-based strategies incorporate contaminants and their associated impacts, respectively.

Currently, the configuration of UDSs limits operators' control during rainfall and subsequent flooding events. Significant storage capacity might exist within the system, but poor utilisation and management often lead to urban pluvial flooding. RTC has been shown to reduce urban flood risk with various designs, as demonstrated by several studies. For example, Mounce et al. (2019) investigated an RTC based on fuzzy logic to mitigate urban flooding by utilising the existing spare capacity in urban drainage networks. The study employed expert knowledge to develop fuzzy logic rules, expressed as *if-then* statements. Frequent water-level measurements were used to adjust a flow control gate, thereby minimising local flooding and optimising network storage during rainfall events. The results indicated a reduction in local flooding, with an average reduction in flood volume of 66% after optimisation compared to cases without gate control, and a 25% reduction for unseen test rainfall events.

Similarly, the beneficial effects of RTC on urban flood risk mitigation using movable gates have been evaluated (Maiolo et al., 2020). A previously validated distributed real-time system was applied to a highly urbanised catchment with a combined sewer system that suffers from undersized pipes, low slopes, and frequent rainwater discharge. The study considered various scenarios based on the positions and number of movable gates, including the current uncontrolled situation, to evaluate the RTC performance. Findings showed that implementing the RTC strategy improved the system performance by reducing total flood volume and hours flooded. The storage capacity of the conduits was better utilised, demonstrating that RTC can be a viable solution for minimising urban flood risk.

Additionally, RTC strategies involving pumping stations to reduce CSO spills and mitigate flooding have been developed. Nielsen et al. (2010) evaluated a global control RTC strategy by calculating the reduction in overflow volumes. This strategy used a simple rule-based control to determine when to switch upstream pumps on or off as a storage basin filled up. Control functions were based on total inflow to a wastewater treatment plant, water levels in basins and critical parts of the system. The implementation of this RTC at nine pumping stations and the introduction of seven gates led to a 40% reduction in overflow volume without increasing the flood frequency. Moreover, the obtained results from the project only represents the absolute minimum for the control potential.

In a similar vein, Tian et al. (2022) explored the flood mitigation capabilities for an optimisation-based RTC approach. This predictive approach utilised rainfall forecast and real-time data of the system to determine whether control actions were necessary. When action was deemed necessary, pumps were activated to balance the water volume both upstream and downstream, minimising CSO and flooding. The optimisation model showed a reduction in flood volumes compared to the currently used rule-based water-level RTC system. The improved performance was partly attributed to using the entire rainfall events rather than only current information. However, the study emphasises that the optimisation model is an ideal method and used to produce an upper-bound performance estimate in the study.

2.3.3 RTC Related Risks

Regardless of the benefits, the implementation of RTC can introduce associated risks that must be considered. Information failure, actuator failure, and forecast error are among the main risk factors in an RTC (van der Werf et al., 2022). Potential failures in the system can increase flooding frequency and volumes, in addition to negatively impacting public health, biodiversity, and the environment. To minimise these consequences and improve system reliability, fault diagnosis can be incorporated to detect potential failures as they occur.

Information failure occurs when decisions are made based on inaccurate, incomplete, or uncertain data. Sensors collect and communicate information about the system’s current state, including actuator setpoints and predicted rainfall, to determine whether actions are necessary in accordance with the established RTC objectives. The RTC performance can be compromised by poor data quality or a failure to communicate essential information to the control agent (Campisano et al., 2013). Actuator failures can occur despite high-quality information and may deteriorate system performance. Ensuring system safety and reliability requires that actuator failures are properly addressed. These failures are often difficult, if not impossible, to predict due to uncertainty regarding when, what type, and which actuator might fail (Tao, 2004). In the event of an actuator failure, increased spills and flooding may occur, even during dry weather flows, with risk escalating as the number of actuators increases. Although such failures have not yet been reported, they remain a significant threat (van der Werf et al., 2022).

For an RTC with a predictive approach, rainfall forecasting is crucial. Forecast errors have the potential to result in performance loss for the RTC. Typically, forecast errors are classified into three types (Habets et al., 2004): localisation, timing, and intensity. Localisation errors occur when the rainfall is predicted for one area but falls in another, potentially causing uncontrolled overflows or unnecessary actions. The response of a catchment is influenced by both preceding events and the precise timing of the current event, emphasising the critical role of timing errors. Timing is especially important when managing storage capacity in advance of rainfall. If a rainfall event occurs earlier than predicted, the RTC might underperform, though long horizons can mitigate timing errors. Finally, intensity errors can impact flood mitigation efforts. Underestimation of intensity can reduce flood mitigation performance, especially for large events, while overprediction can result in unnecessary actions. The choice of forecast horizon can help minimise the effects of these errors, ensuring that actions are taken neither too early nor too late. Despite the influence of forecast accuracy and errors, RTC generally outperforms conventional systems (Xu et al., 2020). Continued monitoring can help prevent negative side effects and ensure optimal RTC performance.

RTC is recognised as a cost-effective and alternative method for flood mitigation. However, its effectiveness depends heavily on implementation and strategy design. Approaches range from simple designs, such as operating one actuator based on measurement of the existing system (e.g., water level) and predefined thresholds informed by expert knowledge, to more complex systems involving multiple actuators that rely on interdependencies and forecast data. The perceived effectiveness of RTC can also be influenced by the method of performance assessment, as the maximum theoretical performance potential may appear significantly higher than the practical potential if the tested approach is not feasible due to costs, complexity, or data inaccuracy. Therefore, adequate time and consideration must be devoted to the design and implementation of RTC systems, with particular attention to how the performance is assessed, ensuring fair comparisons with alternative methods and procedures. The reviewed literature focused primarily on RTC applications for flood mitigation and CSO reduction implemented in UDSs. Despite operational and structural differences, implementation of RTC in surface water systems is also expected to enhance the performance and reduce flooding.

Projected climate change and urbanisation are expected to increase the frequency of flooding in urban areas, affecting a growing number of people with consequences ranging from material damages to health impacts and discomfort. Improvements in weather forecasting and the integration of this information into innovative strategies offer opportunities to prevent flooding and minimise negative side effects. One such strategy becoming increasingly popular is RTC, particularly predictive RTC, which combines rainfall forecasts with real-time observations of water systems and has demonstrated effectiveness in mitigating flooding and reducing CSOs. Most research on volume-based control has been applied to UDS, while studies on RTC for receiving water bodies and surface waters primarily focus on pollution-based and impact-based control (e.g., Meng et al., 2017; Xu et al., 2010). However, urban canals (so-called *singles*) provide storage capacity that may be utilised more efficiently to mitigate flooding through RTC strategies, comparable to how detention ponds have been employed in prior studies. Nonetheless, the uncertainties related to RTC design and implementation emphasise the importance of case-specific investigations to evaluate practicality in terms of benefits, risks, and costs.

3 Case Study

This chapter introduces the study area where the designed RTC procedures were implemented and investigated. The first section covers the area's location and details the responsibilities for water management. The second section presents the rainfall datasets, highlights their differences, and explains how they were discretised for use in various parts of the thesis.

3.1 Study Area

Rotterdam, one of the largest cities in the Netherlands, is located in the delta where the rivers Rhine, Waal, and Meuse meet the North Sea. The city is divided into a northern and a southern section by Nieuwe Maas. Additionally, Rotterdam has several urban canals that receive runoff from rainfall and, at times, untreated wastewater. Water quality in and around Rotterdam is managed by the waterboards *Hoogheemraadschap van Delfland*, *Hoogheemraadschap van Schieland en de Krimpenerwaard*, and *Waterschap Hollandse Delta*. The city of Rotterdam is responsible for maintaining and operating the open water bodies and sewer systems, while the Dutch federal government (*Rijkswaterstaat*) oversees the Nieuwe Maas (Arup, 2019; van der Werf, 2023).

This mix of stakeholders can present challenges for implementing optimised control strategies, particularly when changes in discharge affect catchments managed by different institutions. However, this issue falls outside the scope of this thesis. van der Werf et al. (2023) has explored heuristic RTC strategies for Rotterdam's UDS, providing a foundation for further research. In this thesis, the focus is on the northern part of the city, centred around Rotterdam Centraal (Figure 3.1). The exact study area was selected to align with the forecast data's location and resolution while also covering Rotterdam's city centre.

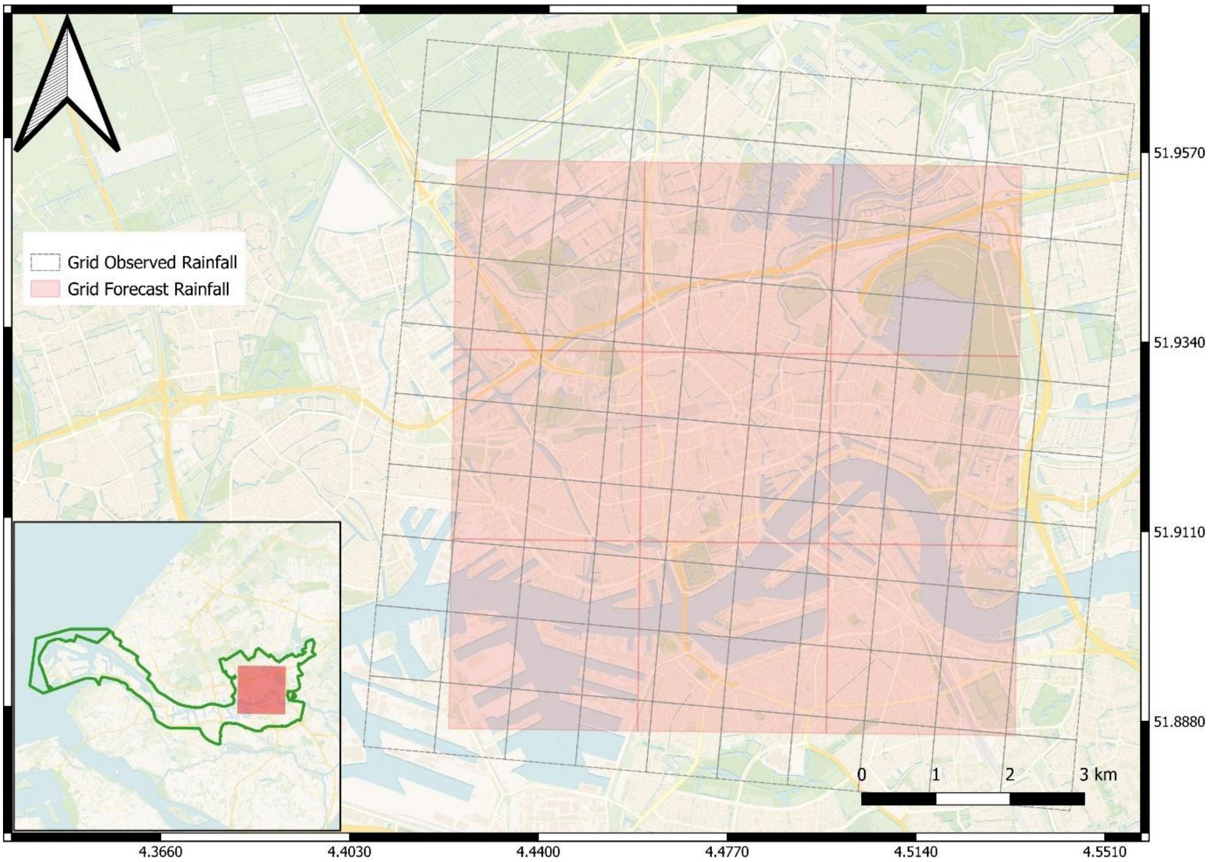

Figure 3.1: Placement of study area within the municipality of Rotterdam. Study area is made up of 9 forecast pixels and 83 pixels for the observed data to cover the same area.

3.2 Rainfall Data

In this case study, two types of precipitation data were used: (1) a rain-gauge adjusted radar dataset; and (2) the HARMONIE-AROME forecast dataset provided by the Royal Netherlands Meteorological Institute (KNMI; see dataplatform.knmi.nl for online access). HARMONIE-AROME is an NWP model that provides a 48-hour forecast, updated every 6 hours. However, the data becomes available only after approximately 3 hours due to calculation time. This forecast data has a spatial resolution of 2.5 km, and 9 pixels were used to cover the study area. In contrast, the radar rainfall data, adjusted using validated rain-gauge data from the KNMI network, is updated every 5 minutes but becomes available with a one-month delay. This dataset has a finer resolution of 1 km per pixel. To cover the same area as the forecast data, 83 pixels were used due to differences in spatial resolution and coordinate reference systems (CRS) between the datasets.

The rainfall data spans three years, from 2019 to 2021. The rain-gauge adjusted dataset is continuous for the entire period, whilst the forecast data has some missing days throughout these years. Additionally, forecasts before February 2019 and after September 2021 are excluded from the analysis due to the unavailability of the data for these periods. In total, 24,070 forecasts were analysed across the study area. The exact distribution of forecast data for the nine forecast pixels is provided (Table 3.1).

Location in study area	*Pixel number*	*Centre Coordinate (EPSG:4326)*	*Number of forecasts*
Lower Left	1	51.8980, 4.4400	2682
Lower Middle	2	51.8980, 4.4770	2673
Lower Right	3	51.8980, 4.5140	2665
Centre Left	4	51.9210, 4.4400	2679
Centre	5	51.9210, 4.4770	2674
Centre Right	6	51.9210, 4.5140	2669
Upper Left	7	51.9440, 4.4400	2681
Upper Middle	8	51.9440, 4.4770	2674
Upper Right	9	51.9440, 4.5140	2673

Table 3.1: Total number of rainfall forecasts evaluated and their distribution over the study area.

3.2.1 Rainfall Discretisation

The rainfall data are used in two different ways in this case study, one for the forecast quality assessment and the other for assessing the flooding frequency, the reactive RTC, and the forecast-informed RTC strategies. For the forecast assessment, the rain-gauge adjusted radar product is discretised to match the forecasts. This first discretisation method does not consider whether it is raining or not, the only objective is to match the events to correspond to one forecast, so it has the same duration, start, and end. Consequently, because of the forecast properties, overlapping and representation of the same hours in different events is present. The second discretisation method is based on the rain-gauge adjusted radar dataset. The rainfall is separated into individual events to allow for an event-based assessment. This is done by discretising the dataset using a minimum inter-event time (MIT) of 12 hours, corresponding to the approximate emptying time of the system (van der Werf et al., 2023).

The MIT discretisation simplifies model simulations by ensuring that the available water storage capacity remains unaffected by preceding events. Over the three-year period, the rainfall was discretised into 1,036 events where both forms of rainfall data were available, which were used to assess whether RTC can mitigate urban pluvial flooding. This number of events was further reduced to 877 by excluding those where discrepancies in rainfall depth occurred due to gaps in the forecast dataset, as these inconsistencies impacted the calculation of rainfall depth. These remaining events have a return period of up to 2 years (Figure 3.2). However, the events are not evenly distributed across the study area, showing significant variations in the number of events, rainfall depth, and event intensity (Table 3.2).

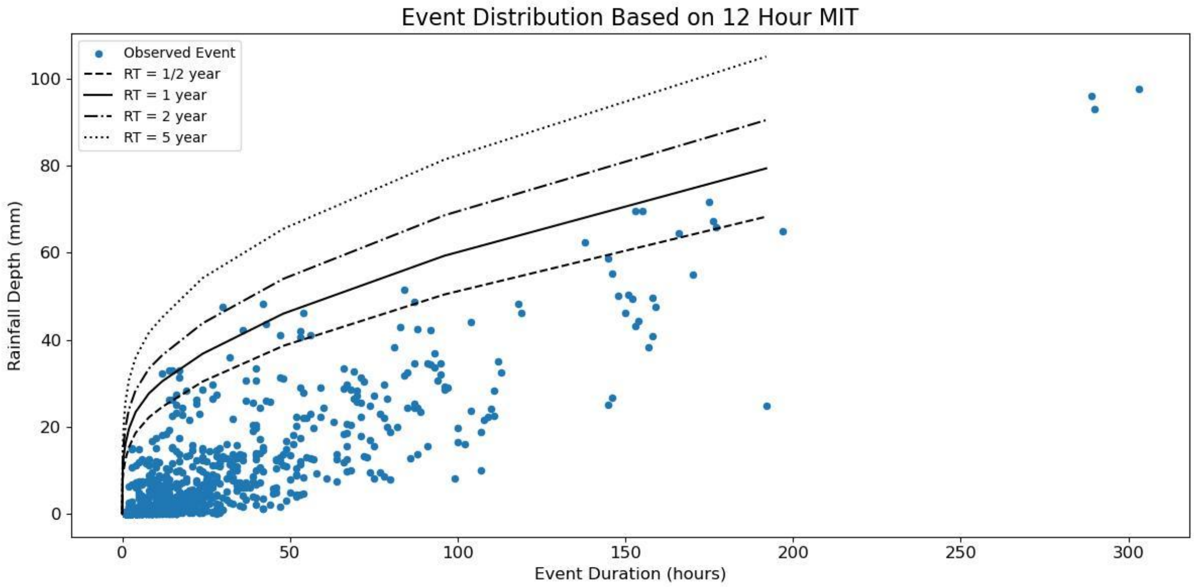

Figure 3.2: Event distribution of evaluated and relevant rainfall statistics (from Beersma et al., 2019). Each event is separated by at least 12 consecutive hours without rainfall.

Pixel Number	Number of Events	Total Rainfall (mm)	Mean Rainfall (mm)	Max Depth (mm)	Max Mean Intensity (mm/hr)
1	262	2142.26	8.18	69.71	3.54
2	120	859.76	7.16	67.21	2.05
3	99	900.58	9.10	69.71	4.97
4	85	774.62	9.11	97.77	2.99
5	36	358.88	9.97	65.07	2.96
6	44	575.86	13.09	71.73	4.99
7	111	1142.11	10.29	96.09	1.93
8	52	582.78	11.21	92.93	1.87
9	68	692.83	10.19	51.55	2.35

Table 3.2: Event distribution across the study area detailing key statistics, including total rainfall depth, the average rainfall depth per event, the event with the largest total depth, and the event with the highest average intensity. Only events where the rainfall depth, duration, and start and end times align between the different approaches are included.

PART I: RAINFALL FORECAST

This first part of the thesis focuses on the quality of the rainfall forecasts. Chapter 4 explains the developed methodology for assessing the accuracy of the rainfall forecasts, including the assessed forecast properties and evaluation metrics. Chapter 5 presents and evaluates the obtained results. Chapter 6 concludes this part with a discussion of the results, focusing on how forecast accuracy is influenced by the forecast properties, and exploring the implications for incorporating these insights into forecast-informed Real-Time Control (RTC) strategies.

4 Methodology Rainfall Forecast

This chapter describes the forecast quality assessment methodology, beginning with the metrics used to evaluate errors and uncertainties in rainfall forecasts, as well as typical factors contributing to these issues. Three key forecast aspects – rainfall depth, forecast horizon, and forecast intensity – are selected for evaluation, with a detailed explanation of why and how each aspect is assessed.

4.1 Quality Assessment

The literature review revealed that forecast errors are commonly separated into three types (location, timing, and intensity). To address these errors, several forecast properties can be evaluated to assess the accuracy and performance of rainfall forecasts. In this thesis, the forecast errors associated with the intensity category are evaluated. Intensity errors are particularly important as they directly influence the overestimation or underestimation of total rainfall depth, with accumulated depth being a critical factor in determining whether flooding occurs (Di Matteo et al., 2019).

In contrast, timing errors are more critical for systems using short horizons but become less significant with increasing forecast horizons (Xu et al., 2020). Similarly, location errors pose challenges, especially for convective rainfall events, where accurately predicting the correct location is difficult, often resulting in misses in smaller catchments (Imhoff et al., 2022). However, these issues are less central in this evaluation. Instead, intensity-related forecast errors emerge as the main contributor to forecast uncertainty on both hourly and daily scales (Shahrban et al., 2016). For this reason, intensity errors are prioritised in this thesis.

Key aspects of the forecast accuracy that are evaluated in this thesis include total rainfall depth, forecast horizon, and forecast intensity, as these factors significantly influence the magnitude and variability of intensity errors. The forecast quality of a Numerical Weather Prediction (NWP) model cannot be sufficiently assessed using only one metric (Shrestha et al., 2012). To achieve a robust evaluation, multiple commonly used metrics are applied. Three binary measures – probability of detection (POD), false alarm ratio (FAR), and specificity (SPC) – are used to assess the forecasts' quality. POD represents the fraction of events correctly predicted as exceeding the threshold. FAR indicates the proportion of false events where rainfall is forecasted but does not occur. SPC measures the fraction of non-events accurately predicted, where both the forecasted and observed rainfall are below the threshold. Additionally, root-mean-square error (RMSE) is employed to assess the average magnitude of forecast errors, while depth difference is used to identify whether forecasts overestimate or underestimate observed rainfall depth. The different evaluation methods used are listed and described (Table 4.1).

Acronym	**Name**	**Description**
POD	Probability of Detection	Fraction of events correctly forecasted
FAR	False Alarm Ratio	Fraction of forecasted events that are actually non-events
SPC	Specificity	Fraction of non-events correctly forecasted
RMSE	Root-Mean-Square Error	Average error magnitude
	Depth Difference	Measure between observed and forecast rainfall depth

Table 4.1: Evaluation metrics and their descriptions as applied in this thesis (based on Ashok & Pekkat, 2022).

4.1.1 Rainfall Depth

The forecast's ability to correctly predict rainfall depth is investigated using the method described by van der Werf et al. (2023). The datasets are transformed into binary values using various thresholds of rainfall depth. For every rainfall forecast, the rain-gauge adjusted radar rainfall for the same period is considered. Each updated prediction compares the accumulated rainfall depth for the forecast and observed rainfall to the threshold value, with exceedance corresponding to a *True* and non-exceedance to a *False*. By applying the method to both datasets, it becomes possible to examine whether the predictability of events varies with rainfall depth. The effectiveness of binary predictions is assessed by calculating the POD, SPC, and FAR as follows:

$$POD = \frac{TP}{TP + FN} \tag{4.1}$$

$$SPC = \frac{TN}{TN + FP} \tag{4.2}$$

$$FAR = \frac{FP}{FP + TP} \tag{4.3}$$

TP (true positives) is the number of correctly predicted events, *FN* (false negatives) is the number of wrongly predicted events, *TN* (true negatives) is the number of correctly predicted non-events and *FP* (false positives) is the number of wrongly predicted non-events.

4.1.2 Forecast Horizon Influence

Insight from the literature suggested that typical forecast horizons for early warning systems range between 12 and 72 hours. Additionally, horizons shorter than 48 hours, but longer than 6 hours was considered optimal, despite being reported insignificant performance improvements when the horizon was reduced from 36 hours. However, as the forecast horizon increases, larger uncertainties are expected. Understanding the influence of forecast horizon accuracy can impact the choice of horizon to be used within the RTC and is valuable information for the user (Imhoff et al., 2020). To assess and quantify the difference between the forecast estimates and observed rainfall depths over the entire horizon, the following metrics are applied: RMSE and depth difference.

$$RMSE(h) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i(h) - \hat{y}_i(h))^2} \tag{4.4}$$

$$Depth\ difference(h) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i(h) - \hat{y}_i(h)\right) \tag{4.5}$$

Where N is the total number of samples in the dataset, $y_i(h)$ is the rain-gauge adjusted radar rainfall for the i^{th} observation at hour h, and $\hat{y}_i(h)$ is the forecast prediction for the i^{th} observation at hour h. For the calculations, the cumulative rainfall for the congruent hours of the forecast and observed rainfall are utilised. The timing becomes less important when using the accumulated rainfall, as the focus shifts to the cumulative rainfall depth over the forecast horizon. The horizon is determined by considering the time difference between the evaluated hours and the start of the forecast.

From a user perspective, it is beneficial to detect whether rainfall will occur as early as possible. To achieve this, the influence of forecast horizons on POD, FAR, and SPC are evaluated, as these categorical verification scores are less sensitive to large errors compared to continuous verification scores (e.g., RMSE) (Shrestha et al., 2012). Combining POD, FAR, and SPC with different forecast horizons might provide insight into whether lead time impacts the probability of detecting rainfall events of different magnitudes. Equations 4.1-4.3 are modified to include the different horizons, which correspond to the forecast update frequency of 6 hours. However, an intensity threshold is now utilised instead of rainfall depth. This change is done because altering the forecast horizon changes the period over which rainfall accumulation is measured. Additionally, earlier studies (e.g., Jee & Kim, 2017) have employed intensity-based thresholds to compare various forecast horizons. As a result, the same depth can represent different levels of event intensity depending on the accumulation period associated with each forecast horizon. In other words, because the accumulation period varies with different horizons, the depth threshold values might be interpreted differently.

4.1.3 Forecast Intensity

Research (Fu et al., 2011) has shown that rainfall characteristics have a significant effect on flood risk management in conventional drainage systems, with average intensity being one of the most important characteristics. The rainfall intensity decreases as the duration of the event increases (Figure 4.1). Consequently, heavy storms are associated with a short duration, whilst drizzle can last for a long time (Butler & Davies, 2011). The mean forecast intensity, the depth difference, and their correlation are used to evaluate forecast accuracy. The depth difference, calculated using Equation 4.5 over the entire horizon, is paired with the associated mean forecast intensity. By relating the depth difference to forecast intensity rather than the observed intensity, knowledge can be gained into whether the forecasts better predict high- or low-intensity events. Additionally, the impacts of forecast intensity on underestimation and overestimation can be assessed. Based on the equation, underestimation is associated with a positive depth difference, while overestimation is exhibited as a negative depth difference.

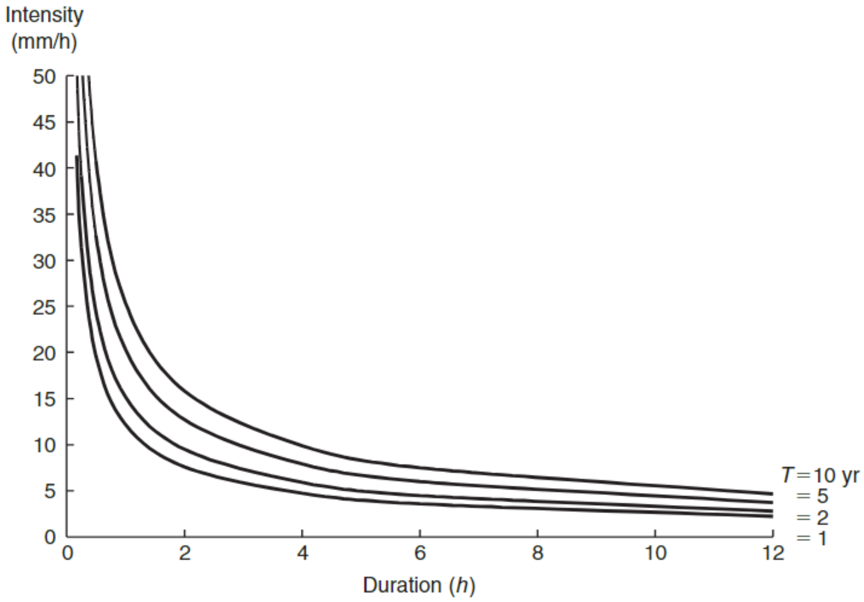

Figure 4.1: Typical intensity-duration-frequency curves from Butler and Davies (2011).

5 Results Rainfall Forecast

This chapter presents the results of the rainfall forecast assessment, with each of the subsections focusing on a key aspect of the analysis. The first subsection presents the results on the forecast's predictability based on cumulative rainfall depth. The second subsection focuses on how the forecast horizon influences predictability, in addition with the impact on forecast errors. The final subsection presents the relationship between the mean forecast intensity and the difference between predicted and observed rainfall depth.

5.1 Quality Assessment

5.1.1 Rainfall Depth

The ability of forecasts to correctly predict rainfall events, based on exceedance of threshold values, is visualised (Figure 5.1). The results show a clear trend: the forecast is significantly better at predicting smaller rainfall depths compared to larger rainfall events. The POD for an event with a cumulative rainfall depth of 1 mm was 0.85, while this POD value declined to 0.1 for a 40 mm threshold. The SPC and FAR exhibit different behaviour; they both increased as the threshold rose. When the threshold was 1 mm, SPC and FAR were 0.82 and 0.09, respectively, but at 40 mm, SPC rose to 0.995 and FAR increased to 0.93. The histogram shows the distribution of the observed rainfall, with most events having a total rainfall depth of 1 mm or less. This distribution may be affected by the discretisation method used for the rainfall events. The observed events were discretised to match the forecasts, resulting in precipitation being represented across multiple events due to overlapping hours and the updating frequency of the forecasts.

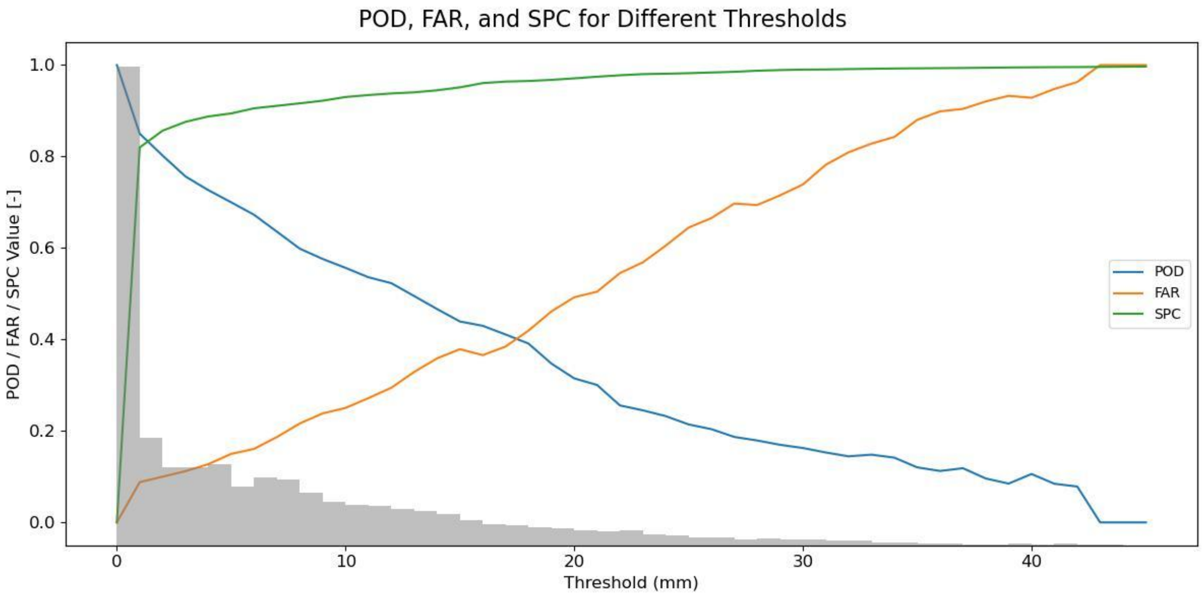

Figure 5.1: POD, FAR, and SPC values for different threshold rainfall depths with a set horizon of 48 hours. Rainfall histogram shows relative abundance of data points per 1 mm threshold interval.

The results show a decline in predictability for larger events, as indicated by the drop in POD with increasing thresholds. While POD decreases, SPC increases, reflecting an improved ability to correctly identify non-events. This is accompanied by an increase in false alarms as the threshold rises. Typically, SPC and FAR have an inverse relationship: false positives lead to a rise in FAR and a drop in SPC. For both SPC and FAR to increase, there must be a substantial increase in the correct identification of non-events (true negatives). This situation occurs when most of the data comprises non-events, and the forecast becomes more effective at identifying them as the threshold increases. The histogram reveals that a large portion of the rainfall events has depths smaller than 1 mm. Consequently, as the threshold increases, the number of actual events decreases, while the number of non-events increases (Table 5.1). Therefore, while false alarms increase, the correct identification of non-events increases even more significantly. Thus, although the forecast is less effective at detecting actual events (lower POD) and results in more false alarms (higher FAR), it identifies non-events more accurately (higher SPC). This suggests an overestimation of occurrence and possibly an underestimation of depth for higher thresholds. Results from higher threshold values should be used with care, as only a few cases occur.

Threshold [mm]	*Observed Exceedances*	*Forecast Exceedances*	*POD*	*FAR*	*SPC*
0.0	24070	24070	1.0	0.0	0.0
1.0	16534	15411	0.85	0.09	0.82
5.0	11103	9142	0.70	0.15	0.89
10.0	6568	4877	0.56	0.25	0.93
15.0	3703	2619	0.44	0.38	0.95
20.0	2092	1298	0.32	0.49	0.97
25.0	1050	633	0.21	0.64	0.98
30.0	504	314	0.16	0.74	0.99
35.0	191	192	0.12	0.88	0.99
40.0	85	126	0.11	0.93	1.0
45.0	17	75	0.0	1.0	1.0

Table 5.1: Number of observed and forecasted events equal or greater than the specific threshold and associated POD, FAR, and SPC.

The obtained results are consistent with previous research. The ability of nowcasting (2-hour horizon) to accurately predict rainfall depths has been studied (van der Werf et al., 2023). Despite the significantly shorter horizon, POD and SPC exhibit the same trend, decreasing and increasing with thresholds, respectively. Longer horizons, up to 9 days, have also been evaluated (Shrestha et al., 2012). In these cases, POD still decreased for higher thresholds, while FAR increased. Therefore, accumulated rainfall depth can be a useful metric for forecasts and their application in an RTC strategy, irrespective of the forecast horizon considered.

5.1.2 Forecast Horizon Influence

Further investigation into the predictability of cumulative precipitation depth across different forecast horizons was carried out (Figure 5.2). The same trend is observed: POD declined, while FAR and SPC increased with higher intensity thresholds across all the considered forecast horizons. Similarly to the results of the entire horizon in subsection 5.1.1, this behaviour may be explained by the significant increase in true negatives as higher thresholds are considered. Upon examining the results, longer horizons perform better in predicting events and not issuing false alarms at lower thresholds. However, SPC is lower, indicating worse performance in predicting non-events. As the threshold increases, shorter horizons become more advantageous: POD becomes higher and FAR lower. The opposite is true for SPC, where long horizons are better at predicting non-events for high intensities, and short horizons better for low intensities, though the difference is negligible. The figure shows that the point at which shorter horizons outperform longer ones occurs when the intensity is just above 0.2 mm/hr. As the intensity increases further, the difference in POD and FAR between the forecast horizons become more pronounced, while the difference in SPC decreases. Moreover, the difference in predictability is relatively small for 48- to 30-hour horizons but becomes larger for 24- to 6-hour horizons.

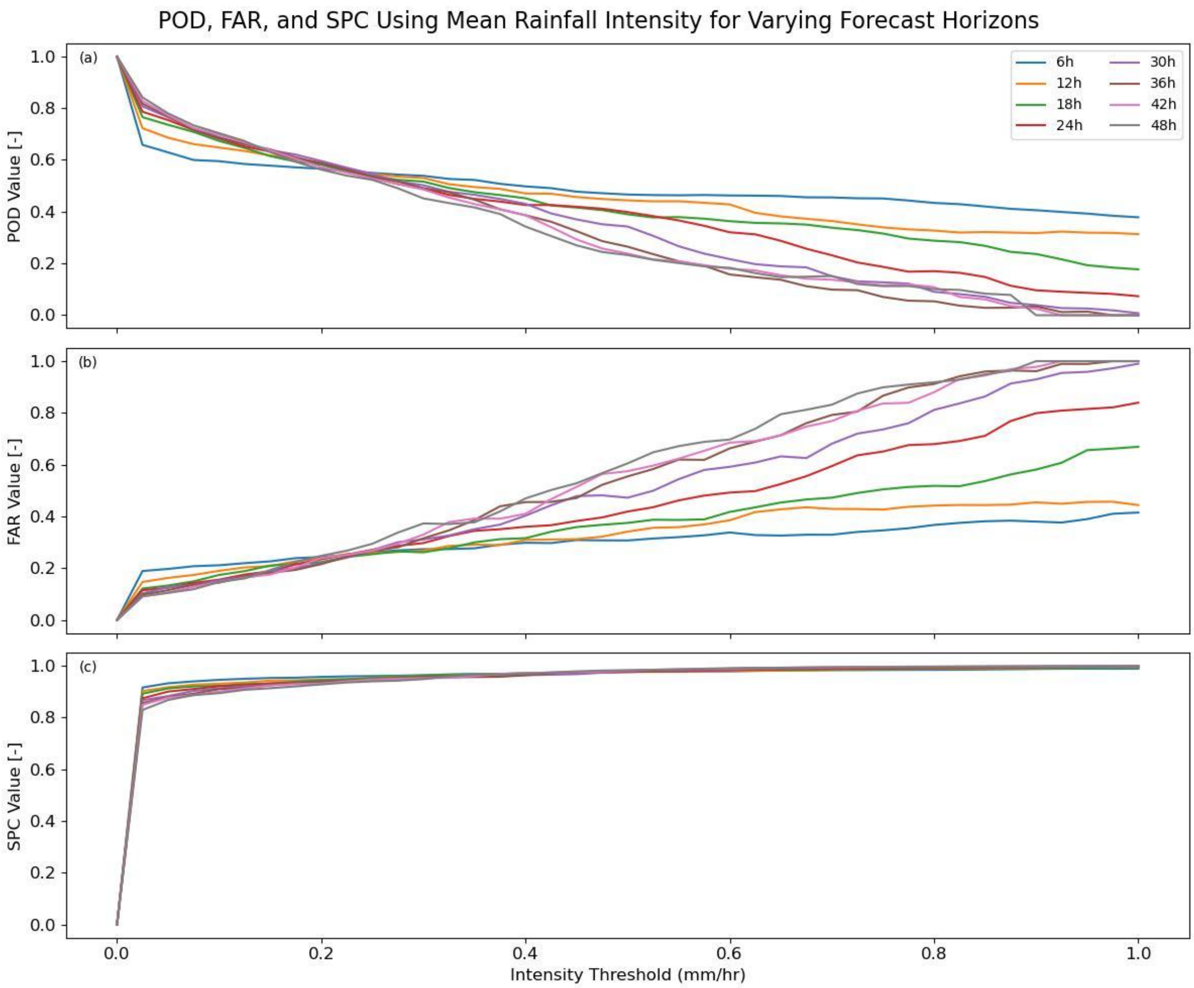

Figure 5.2: POD, FAR, and SPC across different intensity thresholds for forecast horizons corresponding to the forecast update frequency.

When different rainfall thresholds are considered, the predictability varies between forecast horizons depending on the size (depth/intensity) of the rainfall event. Predictability decreases with increasing thresholds across all horizons, but shorter horizons perform better for intense rainfall. This shift in effectiveness has been similarly observed in prior studies (Jee & Kim, 2017). For a static threshold of 0.5 mm/hr, POD decreased and FAR increased for lead times of 18, 12, and 6 hours, while another study (Shrestha et al., 2012) found that FAR decreased with increasing thresholds for 1- and 2-day horizons, before rising again for longer lead times. At lower intensities, the FAR values for the 1- and 2-day horizons in that study are comparable to those in this study. However, for larger thresholds, their results demonstrated superior performance compared to those presented in this study, with significantly lower FAR values. This is desirable as it helps limit false warnings, thereby maintaining public trust in weather services and civil protection authorities (Pirone et al., 2023).

The observed dynamics can partially and possibly be explained by the used datasets. The events were discretised independently of observed rainfall and were only adjusted to match the forecast duration, start, and end. As a result, events without any rainfall are included in the evaluation, increasing the number of observed true negatives and the SPC. Similarly, POD can be artificially improved by issuing false alarms, which increases the number of hits (Yang et al., 2016). For longer horizons, timing becomes less important, as these provide more time to detect the rainfall (Xu et al., 2020), while for shorter horizons, a single hour with rainfall could potentially exceed the assessed threshold. These differences regarding how different intensities and horizons evaluate various rainfall depths can potentially explain the observed differences in POD and FAR. However, it should be noted that if the thresholds were to be transformed to depths, longer horizons would always show preferable performance purely because of the increased rainfall accumulation period. Therefore, the data, type of threshold, and horizon must be considered when assessing the results and their usefulness.

The influence of the forecast horizon can also be observed using the metrics (1) depth difference; and (2) RMSE between the forecasted and observed precipitation depths. For the initial hours, the forecasts are relatively accurate with respect to cumulative rainfall depth, but uncertainty steadily increases as the horizon lengthens. The mean RMSE remains below 4 mm for the entire horizon; however, significant variations between forecasted and observed rainfall are evident from the 75% and 95% confidence intervals (CIs). Assessing the depth difference (Figure 5.3a), minor underestimation is indicated by the positive mean depth difference of the cumulative depth. The large variation in the CIs suggests lower precision as the forecast horizon lengthens. Moreover, the 75% and 95% CIs show that underestimation tends to have a greater absolute value compared to overestimation. Although both over- and underestimation occur, the magnitude of error is larger for underestimation. This large range of error may limit the effectiveness of using cumulative rainfall depth alone in a control strategy, although the mean is relatively close to zero.

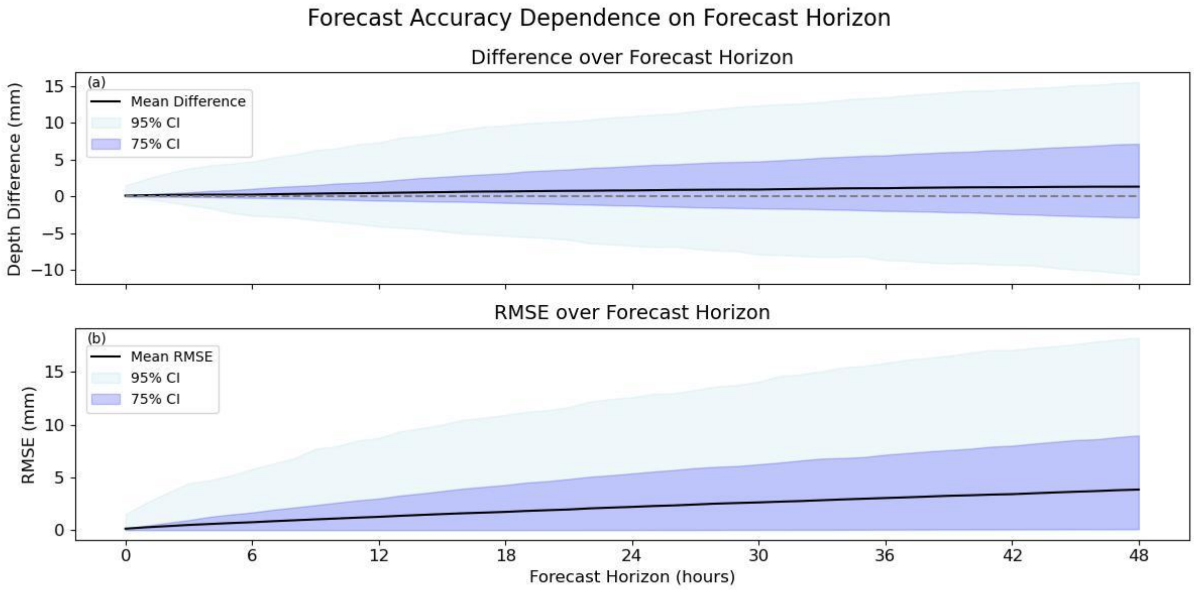

Figure 5.3: RMSE and depth difference, defined as observed minus forecasted rainfall, for cumulative rainfall depth over the entire forecast horizon.

The results confirm that forecast accuracy depends on both the forecast horizon and rainfall depth. As the forecast horizon extends, the magnitude of error in accumulated rainfall depth increases, aligning with previous research (Jang & Hong, 2014; van der Werf et al., 2023; Yang et al., 2021), which demonstrated increased uncertainty with longer lead times. As the horizon increases, the range of error grows due to reduced similarity between forecasted and observed events. For shorter horizons, underestimation of events predominates, but as the horizon lengthens, both over- and underestimation may occur, consistent with earlier studies (Heuvelink et al., 2020; Jabbari et al., 2020; Pirone et al., 2023).

The differences between the obtained results and those from previous research may be attributed to factors such as model resolution, catchment size, forecast horizon, and event types. Generally, forecast performance improves with finer resolution, though this can also lead to overestimation of rainfall, reducing the overall forecast skill (Ashok & Pekkat, 2022). Lower resolution models show a more rapid decline in heavy rainfall prediction accuracy as forecast horizon increases. Additionally, small catchments are more sensitive to rainfall system location, increasing the relative error (Heuvelink et al., 2020). Therefore, it is important to consider the combined effects of resolution, catchment size, and forecast horizon when assessing forecast accuracy. Nonetheless, real-time forecasting systems appear to benefit from shorter horizons, as uncertainty tends to increase with longer lead times.

5.1.3 Forecast Intensity

The relationship between mean forecast intensity and accumulated depth difference has been investigated (Figure 5.4). Since mean intensity is a product of duration, it also provides insight into forecast depth. Both underestimation (positive depth difference) and overestimation (negative depth difference) of cumulative rainfall depth are observed. A clear trend emerges: underestimation occurs with low forecast intensities, while overestimation is prevalent when the forecast has a high mean intensity. As the forecast predicts larger events, both the frequency of overestimation and the depth of overestimation increase. The box plot, where the box represents the interquartile range (IQR) and the whiskers indicate the 95th percentile, shows that the range of error also expands with increasing intensity. When the mean forecast intensity is low, the data exhibit low variability and depth difference compared to the observed events. This pattern can be explained by the distribution of forecasts intensities (Table 5.2), where more than half of the forecasts have intensities below 0.1 mm/hr.

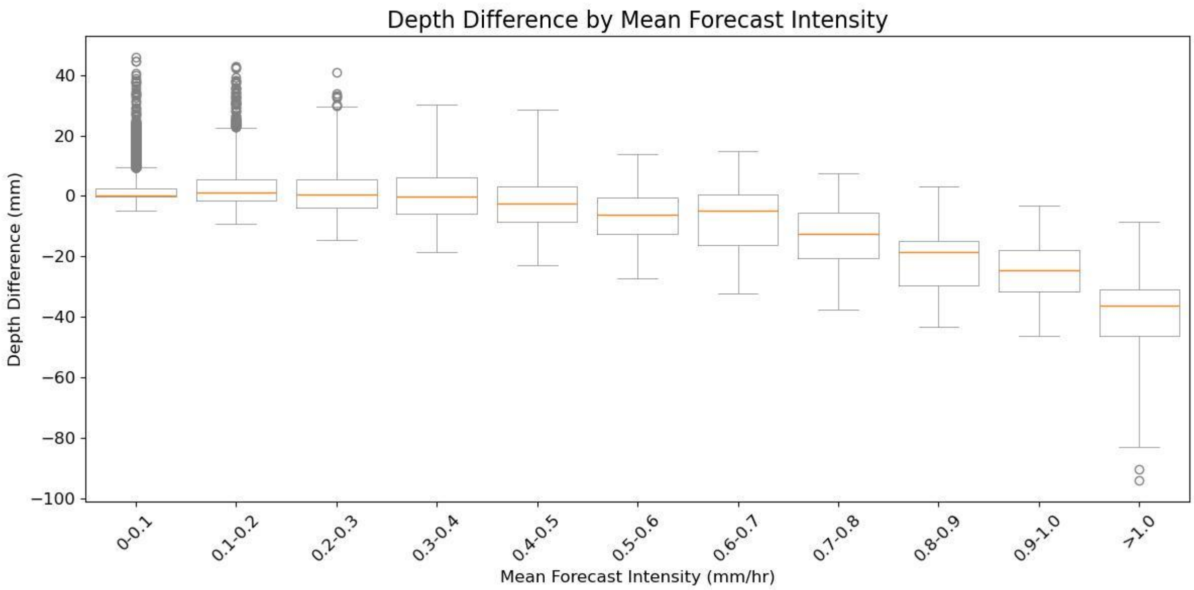

Figure 5.4: Depth difference between forecast and observed cumulative rainfall depth in relation to mean forecast intensities. Depth difference is calculated as observed minus forecast rainfall depth.

For this dataset, the forecasts demonstrate to be more accurate in predicting smaller rainfall events. Low intensities show a smaller depth difference and range of error between the forecasts and observed events, despite the presence of more outliers. However, forecast intensities below 0.9 mm/hr exhibit both overestimation and underestimation of the cumulative depth. On average, the forecasts tend to overestimate the depth when the intensity reaches 0.5 mm/hr.

Previous studies, such as Van Steenbergen and Willems (2014), have reported increased forecast rainfall error with larger rainfall depths, which is consistent with the results obtained in this study. Specifically, large depths are overestimated, and small depths are underestimated by the forecasts. However, other studies, such as Jabbari et al. (2020), have reported different results. Their comparison of NWP forecasts with rain-gauge observations for three heavy rainfall events revealed a tendency for the forecasts to underestimate the precipitation.

The findings suggest that the model does not accurately predict intense rainfall but can be advantageous for less intense events. The relationship between depth difference and mean forecast intensity might be particularly useful in an RTC strategy. By focusing on the forecasts rather than the observed events, decision-makers can gain valuable insights into the uncertainty associated with the forecasts. This approach allows for better-informed decision-making, enabling the implementation of timely actions or the decisions to wait until forecasts become more accurate.

Intensity Bin [mm/hr]	*Number of Forecasts [-]*	*Mean Depth Difference [mm]*	*Min/Max Observed Depth [mm]*	*Max over-/ underestimatio*n *[mm]*
0-0.1	14823	1.813946	0.00 / 48.42	4.79 / 45.88
0.1-0.2	4254	2.780237	0.00 / 51.33	9.30 / 42.88
0.2-0.3	2258	1.567630	0.00 / 54.35	14.45 / 41.10
0.3-0.4	1344	0.396940	0.04 / 49.86	18.59 / 30.40
0.4-0.5	721	-2.767065	0.38 / 48.82	22.86 / 28.56
0.5-0.6	329	-6.816935	0.10 / 41.76	27.41 / 13.85
0.6-0.7	137	-7.769724	0.52 / 47.67	32.29 / 14.95
0.7-0.8	73	-12.817205	0.72 / 43.62	37.60 / 7.57
0.8-0.9	45	-21.653702	0.56 / 43.83	43.26 / 3.23
0.9-1.0	31	-25.921589	0.86 / 42.53	46.27 / -3.27
>1.0	55	-39.668071	1.05 / 42.42	94.26 / -8.64

Table 5.2: Number of forecasts and the differences between forecasted and observed values for specific intensity ranges. The negative values for underestimation actually indicate overestimation.

The focus of Part I was assessing the inaccuracy, uncertainty, and errors existing in the forecast data. To evaluate the accuracy of the NWP model, the forecast data was compared against observed rainfall derived from the rain-gauge adjusted radar dataset. It was assumed that the radar dataset is correct and representative of the actual events. However, radar-based rainfall estimates are affected by issues such as beam blockage, attenuation, and errors introduced during gauge adjustments (Overeem, Buishand, et al., 2009; Overeem, Holleman, et al., 2009). Therefore, it is important to consider these issues when interpreting the results, as they can contribute to the observed differences between the forecasted and actual rainfall.

Additionally, discretising the events to match the forecast data can potentially influence the results as events without predicted or observed rainfall are included in the evaluation. This approach ensures consistency in event comparisons, although it may amplify the perceived forecast inaccuracy. Furthermore, rainfall was considered to occur when the depth exceeded 0 mm and does not account for uncertainties such as measurement noise or model limitations. For comparison, Shrestha et al. (2012) applied a higher threshold (0.1 mm/3hr), accounting for detection limitations. Consequently, the utilised discretisation and threshold, and their combination, may inflate the observed true negatives and skew the performance metrics.

While the methods used provide a practical framework and valuable insight into forecast performance, it is important to acknowledge the limitations, especially when comparing the results with other studies. Therefore, future studies should consider using other discretisation methods and applying different rainfall thresholds to better capture the true performance of rainfall forecasts.

6 Conclusions Rainfall Forecast

It was demonstrated that the overall forecast accuracy was dependent on the rainfall depth, the forecast horizon and the forecast intensity. The ability to correctly predict rainfall decreased while there was a rise in false alarms as the rainfall depth increased. This may present a problem when incorporating forecast data into the decision-making process of an RTC designed to mitigate flooding, as total rainfall depth is an important factor influencing the occurrence of flooding. The RTC performance may be limited when relying on forecasts with low accuracy and large errors, making inclusion of forecasts questionable in these instances.

Similar behaviour with decreasing POD and increasing FAR was observed for all considered forecast horizons. By using intensity as thresholds, rather than depth, events with similar characteristics were compared. Events with low intensities were more easily predicted with long horizons while shorter horizons were suggested to be preferable when the intensity increased. This may limit the potential benefits of including rainfall forecasts in RTC strategies as the short horizon necessary for correct predictions of intense events, most often causing flooding, restricts time for preventive action. However, the combination between horizon and intensity can represent quite small and normal events which have limited impacts on the surface water level considered in this thesis. Additionally, by transforming the intensity thresholds to depth thresholds, longer horizons were significantly better at predicting depths large enough to cause flooding. This is likely due to the decreased importance of timing between the forecasted and the observed rainfall, indicating that longer horizons may be preferable in predicting flood causing depths.

The assessment of the forecast horizon also revealed that longer horizons decreased the forecast skill. The RMSE between the forecast and observed dataset increased almost linearly over the horizon and there could be significant difference between the two datasets. A small underestimation of the cumulative rainfall depth was also shown, despite the depth difference being relatively close to zero. However, there were large variations in the two datasets, with both overestimation and underestimation observed. This increases the difficulty of including rainfall forecasts in an RTC strategy as large overestimation and underestimation errors can be expected when considering the entire horizon. Furthermore, it was shown that the depth was underestimated when the mean forecast intensity was low and overestimated with high forecast intensities. These results were only based on the full horizon of the forecasts, shorter horizons may perform differently and should be considered in future research. Underestimation of the depth may prevent suitable actions to be carried out while overestimation of the depth may result in unnecessary action. Therefore, moderate intensities should be considered as the median depth difference was close to zero despite demonstrating both over- and underestimation. Additionally, these event intensities correspond to cumulative depths sufficient to cause flooding.

Despite the limitations and drawbacks found in the assessment of the forecast data, the results can be used to determine how the forecasts can best be included in a forecast-informed RTC. Firstly, Yang et al. (2016) reported that a false alarm rate below 0.5 is generally tolerable when using forecast data issuing urban inundation alarms. As the FAR increases with the rainfall depth and forecast horizons, this restricts the use of very large depths in combination with long horizons. Secondly, the increased uncertainty with longer horizons suggests that shorter horizons would be more effective for implemented RTCs, provided they are still long enough for the proactive measure to have its full effect. Thirdly, it has been reported (Walczykiewicz & Skonieczna, 2020) that the intensity of events has a large influence on whether urban flooding occurs. Therefore, forecasts with reasonable intensities have been used in the developed RTC procedures. These procedures aim to achieve a compromise between the forecast accuracy, the possibility for flood occurrence and providing sufficient lead time for preventive action.

PART II: REAL-TIME CONTROL

This part of this thesis focuses on developing and evaluating forecast-informed Real-Time Control (RTC) strategies. Chapter 7 explains the RTC methodology, which includes the design of the surface water system model, the reactive RTC, and the forecast-informed RTC procedures. In the design of the forecast-informed RTCs, the results of the rainfall forecast analysis have been incorporated into the methodology, guiding the creation of the pumping rules. Chapter 8 presents the different RTCs' performance and improvement potential. Chapter 9 concludes this part by discussing the benefits and drawbacks of RTC implementation, along with the factors influencing its performance.

7 Methodology RTC

This chapter presents the developed methodology used for the RTC procedures. The first section outlines the design of the conceptual surface water system model, its operation, key parameters, and how the results will be utilised. Section 7.2 builds on the surface water model by introducing a reactive RTC strategy, explaining the modifications made to the model, and presenting the performance metrics used for evaluating the RTC performance. The final section introduces the design of the forecast-informed RTC procedures, which are based on the results of the forecast assessment, along with the parameters used to assess their performance.

7.1 Surface Water Model

To examine the frequency and total overflow depth of urban pluvial flooding, a conceptual bucket model representing the urban canals of Rotterdam is designed. This model operates on a mass balance principle, where the volume of the bucket depends on the inflow and outflow. The mass balance is simplified to consider only rainfall as the inflow, while the outflow includes overflow and the water pumped out of the system, as these are the most important contributing factors during a rainfall event (Equation 7.1).

$$\Delta V(t) = R_{in}(t) - P_{out}(t) - Overflow(t) \tag{7.1}$$

Groundwater infiltration, seepage, and evaporation are not considered, as they are assumed to be negligible. While these factors are not always negligible, their significance depends on the specific system and conditions. During the evaluated events, which involve short-term, intense rainfall causing flooding, the rates of infiltration, seepage, and evaporation are much smaller compared to the inflow and outflow, limiting their impact on urban pluvial flooding dynamics. Additionally, infiltration is reduced in urban contexts due to impermeable surfaces. Seepage is also limited by the short durations of high water levels and insufficient time for prolonged seepage to occur. Furthermore, well-lined or impermeable canals can further restrict seepage.

Evaporation is influenced by factors such as high temperatures, solar radiation, and prolonged dry spells, which are typically absent during rainfall events. Event-based assessment does not account for the time between the events, eliminating the influence of extended dry spells. Cooler weather and cloud cover during the events also reduce evaporation rates. However, for the continuous evaluation, these factors could play a more significant role and may need to be considered.

The total water depth in the model is 100 mm, with an initial water level of 80 mm, and a surface water storage capacity of 20 mm. When the accumulated rainfall exceeds this capacity, flooding and overflow occur. The entire rainfall depth over a pixel is assumed to contribute to the urban canal, meaning that any rainfall event with a depth greater than 20 mm results in flooding. To assess the flooding frequency without any measures in place, an event-based assessment is conducted using rainfall events obtained from the minimum inter-event time (MIT) discretisation. The frequency of flooding is determined by counting the number of events where rainfall depth exceeds 20 mm. Additionally, the total overflow depth from these events is aggregated to determine the average overflow depth. Furthermore, the analysis explores the relationship between event intensity, duration, and the likelihood of flooding.

7.2 Reactive RTC

Building on the flooding frequency and overflow depth assessment, the simplified bucket model of the urban canals is further utilised to evaluate whether a reactive RTC strategy can mitigate urban pluvial flooding. The primary objective of the RTC approach is to prevent flooding by maintaining a constant surface water level. This is achieved by activating a pump as soon as the water level exceeds the initial 80 mm threshold, which corresponds to 20 mm of remaining storage. Essentially, water is pumped out of the system when the available storage falls below 20 mm. The pump operates in a binary fashion – either off or on – at a capacity of 0.7 mm/hr. This pump capacity is large for a surface water pump, which typically has a capacity of 0.4 mm/hr. A higher pump capacity increases the rate at which water is removed, reducing the likelihood of flooding. The higher capacity allows for faster water removal, reducing the risk of overflow, particularly when the rainfall intensity exceeds the capacity of smaller pumps. This factor should be considered when assessing the results.

To assess the performance of the reactive RTC, which will act as a baseline for later comparisons and analyses, three key factors are considered: flooding frequency, total overflow depth, and water level fluctuations. The frequency is evaluated by counting the number of overflow events across the study area using the reactive RTC strategy. This is then compared to the number of rainfall events with depths greater than 20 mm. Total overflow depth is assessed by summing the overflow for each event and comparing these values across scenarios. Finally, the water level is examined by analysing the overall variation in water level throughout the three-year simulation period.

7.3 Forecast-Informed RTC

To assess whether the theoretical RTC performance can be improved by incorporating rainfall forecasts, the previously described model of the urban canals and reactive RTC are utilised. The control strategies are designed to reduce the frequency and volume of flooding caused by precipitation. The forecast-informed RTCs employ a heuristic-predictive approach, utilising forecast properties in the decision-making process to determine whether to activate the pump and mitigate potential urban pluvial flooding. A decision is made for each forecast using a rolling horizon approach (Shishegar et al., 2019). This method allows for dynamic scheduling, as the decision is re-evaluated whenever a new prediction becomes available, thus creating the possibility to account for forecast errors, which decreases with shorter forecast horizons (Xu et al., 2022) (see also subsection 5.1.2). The concept is illustrated schematically (Figure 7.1).

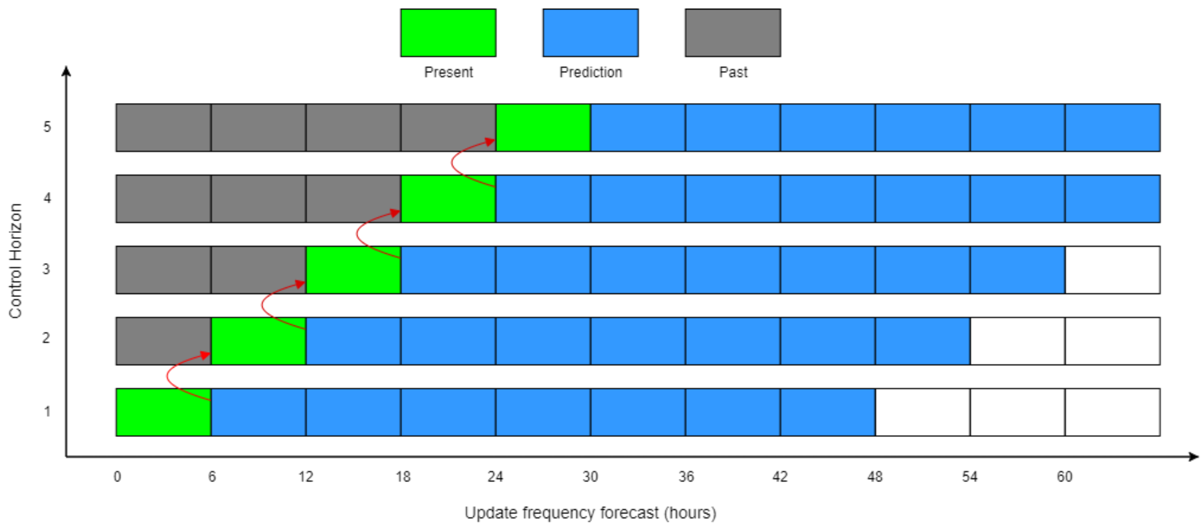

Figure 7.1: Concept of rolling horizon adapted from Shishegar et al. (2019).

To implement this strategy effectively, three forecast-informed procedures have been developed based on the obtained results in Part I (Table 7.1). The evaluated forecasts demonstrated increased uncertainty with longer horizons and larger depths (Chapter 5). The procedures have been designed to account for this uncertainty and inaccuracy by utilising different forecasted rainfall depths and forecast horizons while keeping the false alarm rate at an acceptable level below 0.5.

For Procedure 1 and 2, an approximate intensity of 0.42 mm/hr is selected, corresponding to 20 mm of rainfall over a 48-hour horizon and 10 mm over a 24-hour horizon. Procedure 3 uses a lower rainfall depth of 8.4 mm with a shorter horizon of 12 hours, matching the pump capacity of 0.7 mm/hr. Higher mean forecast intensities were not used because of the limited number of forecasts with these intensities (see Table 5.2). Additionally, these high forecast intensities were shown to be poor in predicting correct rainfall depth with large overestimations.

In all procedures, the pump is activated if the rainfall threshold is exceeded. Additionally, the procedures allow for pump activation even if the rainfall threshold is not reached, but the water level exceeds the initial level of 80 mm as a precautionary measure, identical to the reactive RTC strategy. When the pump is activated based on the water level, it is evaluated every hour, unlike the 6-hour interval used for rainfall forecast evaluations. With the procedures in place, the next step involves applying them to the rainfall forecasts over the study period to assess their performance.

Procedure ID	*Forecast properties*	*Description*
1	Total predicted depth over entire forecast horizon	Pump activated if predicted rainfall ≥ 20 mm or water level ≥ 80 mm.
2	Total predicted depth over first half of forecast horizon	Pump activated if predicted rainfall ≥ 10 mm or water level ≥ 80 mm.
3	Total predicted depth over first quarter of forecast horizon	Pump activated if predicted rainfall ≥ 8.4 mm or water level ≥ 80 mm.

Table 7.1: Description of developed forecast-informed rules used by the RTCs.

In the performance evaluation of the forecast-informed RTC procedures, the reactive RTC serves as the baseline. The evaluation is conducted in two ways: (1) an event-based assessment, utilising the MIT-discretised rainfall events; and (2) a continuous evaluation over the entire period. The event-based evaluation focuses on flood mitigation performance by comparing the flooding frequency and overflow depths of the individual events. Events that cause flooding are further investigated by examining their characteristics. However, discretising the rainfall into individual events may influence the results and misrepresent reality. For instance, pumping may occur at the end of one event in anticipation of future rainfall, resulting in a lower water level that is not accounted for in the following event, where the water level is reset to 80 mm. This issue is addressed by the continuous evaluation, which spans over the entire three-year period. When forecast data is missing (e.g., after September 2021) the reactive RTC logic is applied. The continuous evaluation includes total overflow depth, changes in water level over time, and the percentage of time the water level remains below the target. Additionally, it examines the pump's operation, such as number of activations and total hours of pumping. Lastly, the RTC procedures are evaluated under the assumption of a perfect forecast to determine their maximum theoretical potential.

8 Results RTC

This chapter presents the results related to the forecast-informed RTC, reactive RTC, and the conceptual model of the urban canals of Rotterdam. It begins with section 8.1 presenting rainfall events with depths sufficient to cause flooding, their frequency, and relationship between duration and intensity. Following this, the results of the reactive RTC are presented, highlighting its potential flood mitigation capability. The final section presents the performance of the forecast-informed RTC procedures and compares them to the reactive RTC strategy.

8.1 Flooding Frequency

The datasets included 877 events for which both forecast data and rain-gauge adjusted radar data were available. These events were used to assess the theoretical flooding frequency. Flooding occurred during 139 of these events, while for the remainder, the rainfall depth was insufficient to cause overflow when no measures were implemented (Figure 8.1). Over the three-year period, the total rainfall depth across these 877 events was 8,029.67 mm, yielding an average event depth of 9.16 mm. When focusing exclusively on the theoretical flooding events, which must have a rainfall depth greater than 20 mm, the total rainfall depth was 4,887.17 mm, with an average event depth of 35.16 mm. Of this total, the overflow depth was 2,107.17 mm, averaging 15.16 mm of overflow per flooding event.

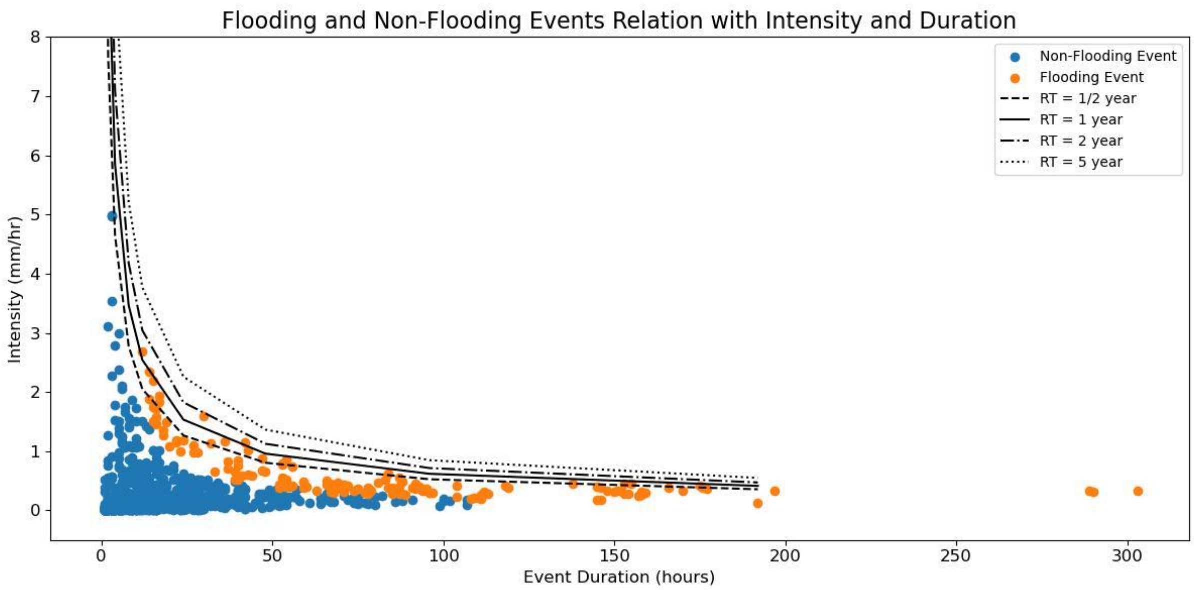

Figure 8.1: Non-flooding and flooding events based on intensity-duration relationship. The rainfall events cover the three-year period and have been discretised using an MIT of 12 hours.

The study area consists of multiple pixels, each representing one data collection point, and a single event may be observed in multiple pixels. These events exhibit minor variations in rainfall depth, duration, and timing as the event moves across the study area. This is particularly noticeable for the three events with the longest duration, each approaching 100 mm in depth. Consequently, an event that covers more than one of the pixels may be regarded as the same event but at different locations. Therefore, due to the division of the study area, a single event might be recorded as up to nine separate events, exaggerating the observed flooding frequency in this thesis.

Flooding can occur during both long-duration, low-intensity rainfall and short-duration, high-intensity events. However, it has been reported (Di Matteo et al., 2019) that events with long durations may limit flood control due to storage capacity being exceeded partway through the event. As the event duration increases, a greater proportion leads to flooding when no control measures are in place. Conversely, shorter duration events often do not cause flooding, as their total rainfall depth remains insufficient to overwhelm the system. Additionally, several of the observed rainfall events potentially causing flooding had a return period of less than twice a year, highlighting the need for measures to reduce the frequency of flooding.

8.2 Reactive RTC Performance

The reactive RTC strategy, which activates the pump as soon as the water level in the urban canals rises, demonstrated its effectiveness by reducing the theoretical flooding frequency and total overflow depth. Using this strategy, 18 of the 877 evaluated rainfall events resulted in overflow, with a cumulative depth of 89.87 mm (Figure 8.2). However, the distribution of events and depths varies across the study area. The number of overflow events ranges from 0 to 3, and the depths from 0 mm to 23.17 mm, for pixel 4 and 9, respectively (see Chapter 3). Despite this spatial variation, a significant number of rainfall events with the potential to cause overflow were prevented. Furthermore, this approach provides a more realistic representation of current practices in Rotterdam and highlights RTC's effectiveness in mitigating urban pluvial flooding.

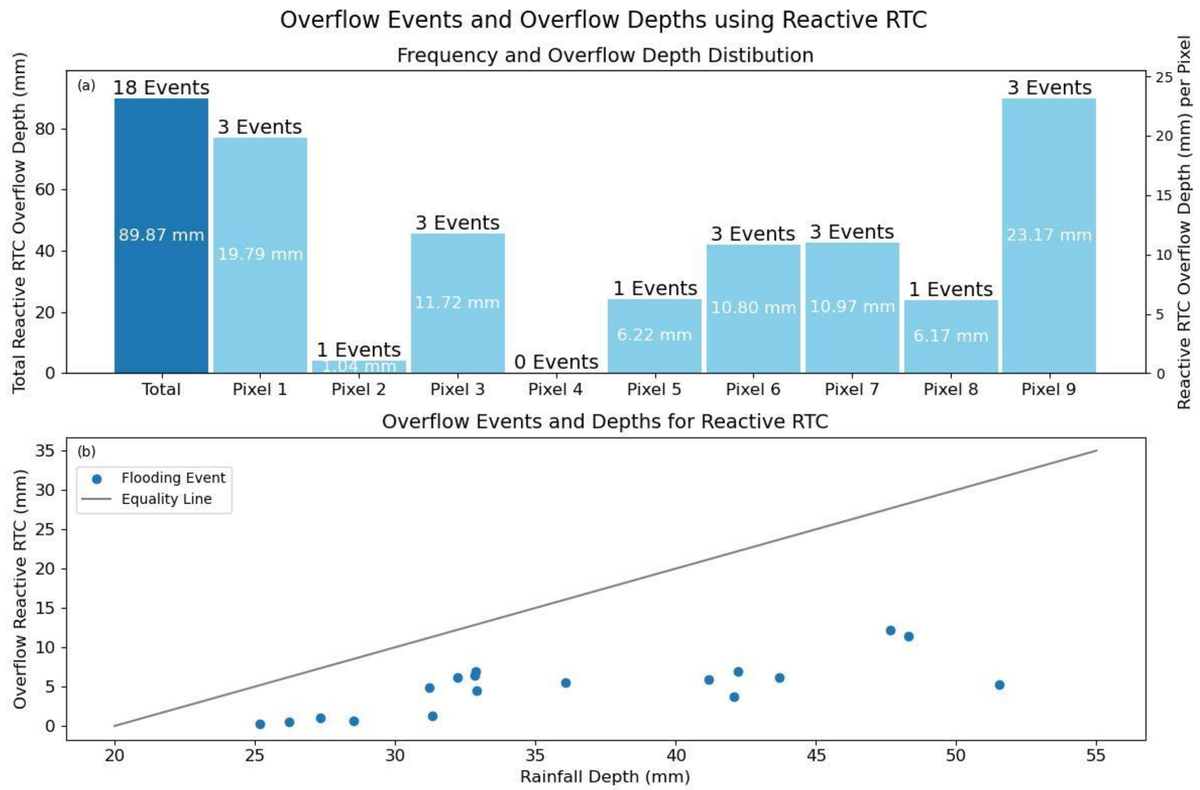

Figure 8.2: Rainfall events where overflow was not prevented using the reactive RTC. Bar chart shows the number of flooding events and combined overflow depth for each pixel. (b) shows overflow depth for each flooding event when using the reactive RTC against rainfall depth. Equality line represents overflow without pumping.

The events in which flooding was not prevented had rainfall depths between 25 mm and 52 mm. Events with larger rainfall depths were successfully managed, making it evident that total rainfall depth alone does not determine whether flooding is prevented. Additionally, the flooding events that were not prevented had durations varying significantly, from 12 to 84 hours. The flooding events were typically characterised by high mean rainfall intensity, high maximum rainfall intensity, or a combination of both, indicating that the pump capacity was insufficient to manage the inflow (Figure 8.3 a). While both mean and peak rainfall intensity influence flooding, the results suggest that peak rainfall intensity plays a more significant role. For instance, the mean rainfall intensity varied from 0.44 mm/hr to 2.69 mm/hr, whereas maximum rainfall intensity ranged from 6.47 mm/hr to 18.04 mm/hr. Moreover, flooding events with longer duration and lower mean intensity generally exhibited a higher maximum intensity, highlighting the importance of peak intensity in determining flooding outcomes.

The reduction in flooding frequency and overflow depth is attributed to the pump's operation. Over the study period, the pump was operational for a total of 10,119 hours, during which it was turned on and off 2,646 times. Although operating the pump incurs costs, it provides additional benefits, such as maintaining a stable water level. Continuous monitoring revealed that the water level never dropped significantly below the target. However, during large or intense rainfall events that exceeded the pump's capacity, temporary rises in water level occurred (Figure 8.3b). Despite these occurrences, the reactive RTC effectively reduced both the frequency of flooding and the depths of overflows, while stabilising the water level.

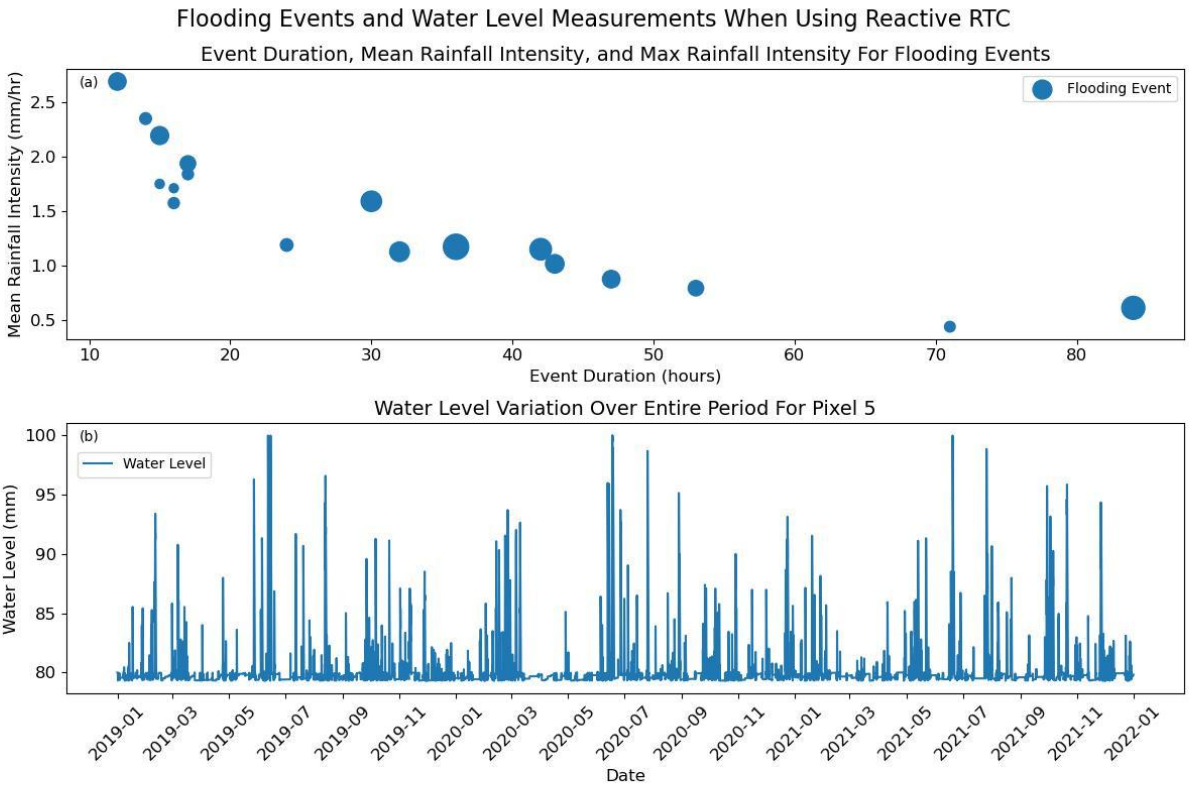

Figure 8.3: (a) Event duration vs. mean rainfall intensity for flooding events under the reactive RTC scenario. Marker size represents maximum rainfall intensity, with larger markers indicating higher intensities. (b) Continuous water level over study period for a single forecast pixel. Note that the number of flooding events differ between panels, as (a) includes all forecast pixels.

8.3 Forecast-Informed RTC Performance

The flood mitigation performance was improved when perfect forecasts were used in the forecast-informed RTC strategies (Figure 8.4). Procedure 1 and 2 demonstrated the best potential and did not have any flooding events. On the other hand, Procedure 3 was not able to prevent every flooding event, with 7 resulting in overflow and a cumulative overflow depth of 25.08 mm. It is evident that the design of the procedures influences the flood mitigation performance, with longer horizons and lower intensity thresholds appearing more effective. The results suggest that the methodology for Procedure 3, with a 12-hour horizon and an intensity threshold of 0.7 mm/hr, is too strict to prevent every overflow event. However, for the overflow events that did occur, the depths were reduced, demonstrating potential while also indicating that the issue lies with the procedure itself.

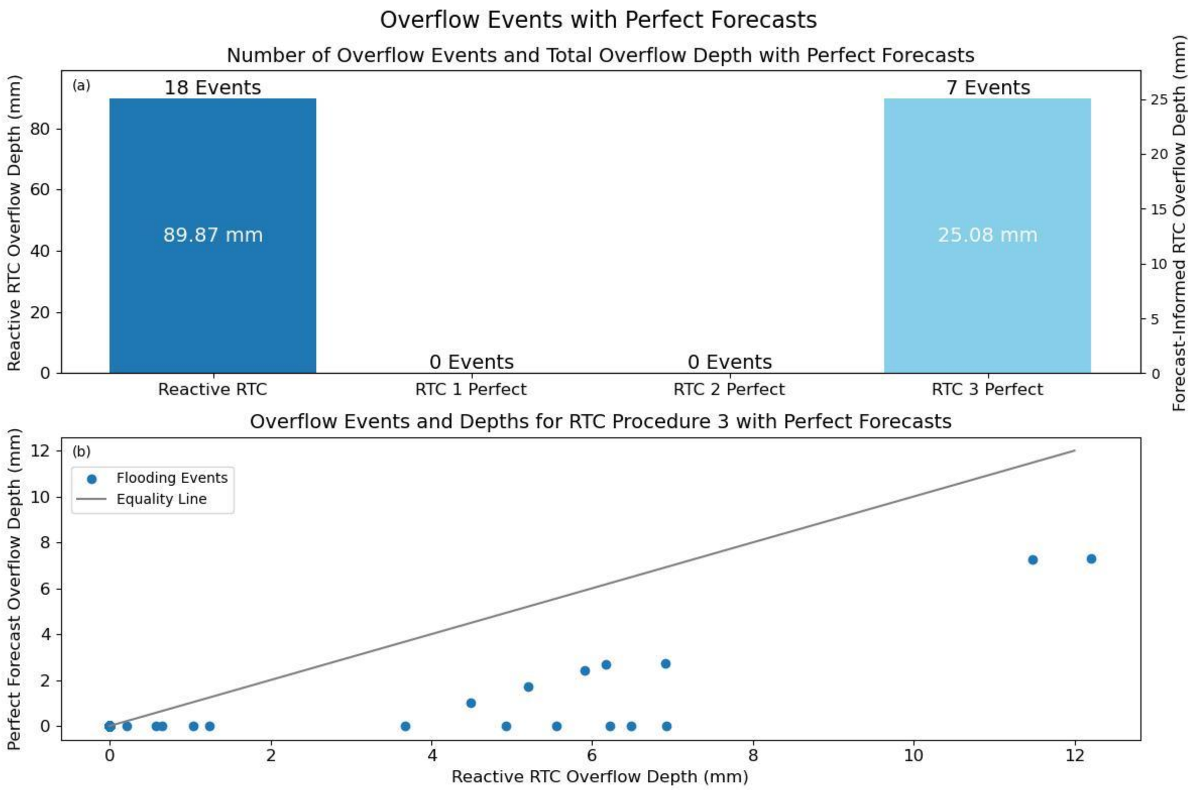

Figure 8.4: (a) Number of overflow events and cumulative depth for an event-based assessment with perfect forecast data. (b) Comparison of overflow depths between reactive RTC and RTC Procedure 3 using perfect forecast data.

The events that were not prevented and caused overflow in Procedure 3 were characterised by a high peak intensity, ranging from 11.06 mm/hr to 18.04 mm/hr. Additionally, the mean intensity of the events varied from 0.61 mm/hr to 1.93 mm/hr, with the total depth ultimately reaching between 32.89 mm and 51.55 mm. The combination between precipitation peaks, the pump's capacity and forecast horizon, restricted the flood prevention potential despite using perfect forecast data.

When using real rainfall forecasts, all three forecast-informed RTC procedures demonstrated improvements in reducing both flooding frequency and overflow depth compared to the reactive RTC (Figure 8.5). Among the procedures, Procedure 2 showed the best overall performance, with 6 flooding events and a total overflow depth of 26.93 mm. Procedure 1 followed next, with 8 flooding events resulting in a cumulative overflow depth of 32.33 mm. Procedure 3 showed the least improvement, with 14 flooding events and a total overflow depth of 54.87 mm. Despite having the most flooding events, Procedure 3 resulted in the lowest average overflow depth per event (3.92 mm) among the three procedures.

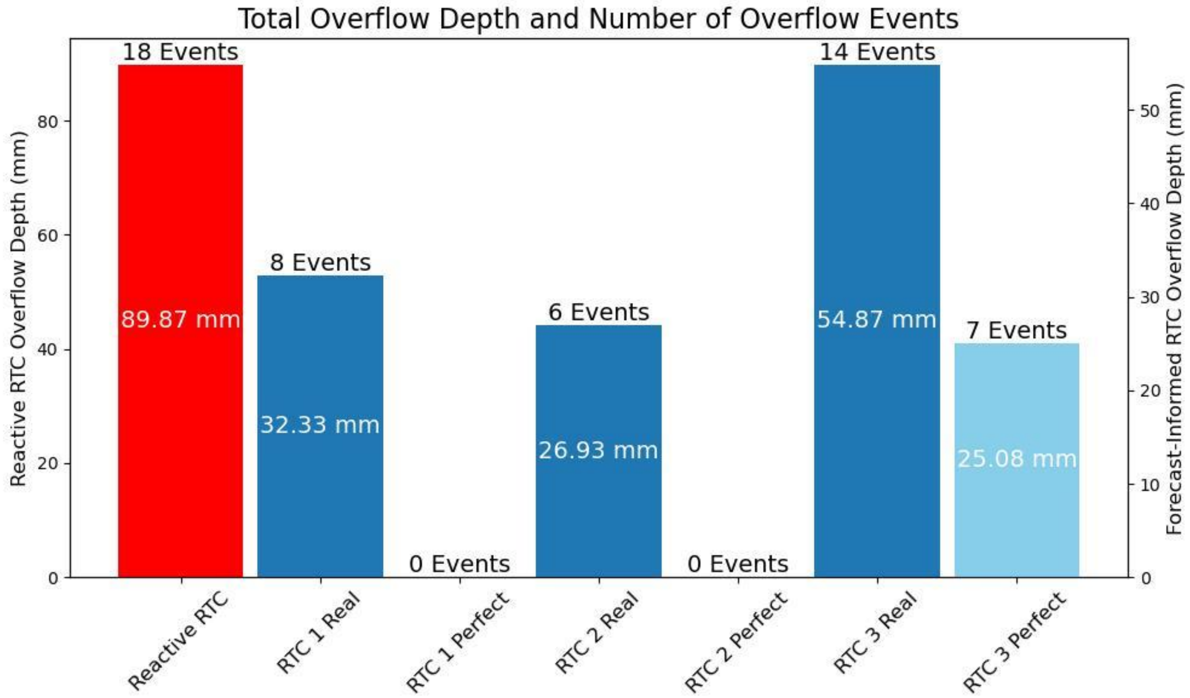

Figure 8.5: Total overflow depth and number of overflow events when using different RTC procedures. The forecast-informed RTC procedures are shown with both real and perfect forecast data.

The failure to eliminate overflow events with Procedure 1 and 2 demonstrated that the flooding events occurred due to forecast uncertainty as they were prevented with perfect forecast data. However, for Procedure 3, some flooding events were due to forecast uncertainty, while others resulted from the designed procedure, or a combination of both the procedure and forecast inaccuracy. The combination between a relatively short forecast horizon and high rainfall threshold limited the time available to sufficiently reduce the water level to mitigate flooding, with the pump's capacity too small for an optimal solution.

Examining the events that caused flooding, it is observed that the forecast-informed RTC procedures never deteriorated the performance in terms of overflow depth when using real forecast data (Figure 8.6). This was because the reactive RTC acted as a precautionary measure within the designed forecast-informed RTC procedures. However, in some cases, the overflow depth was not reduced. Specifically, Procedure 1 had the same overflow depth for 3 events, Procedure 3 for 6 events, while Procedure 2 reduced the depth for every flooding event. The events that the reactive RTC failed to prevent but were mitigated by the forecast-informed RTC procedures were characterised by high maximum rainfall intensity, ranging from 7.90 mm/hr to 18.04 mm/hr. Mean rainfall intensity was shown to be less important, as the reactive RTC already prevented the events with the largest mean rainfall intensity. Despite this, improvements were observed for moderate (0.44 mm/hr) and high mean intensities (2.69 mm/hr).

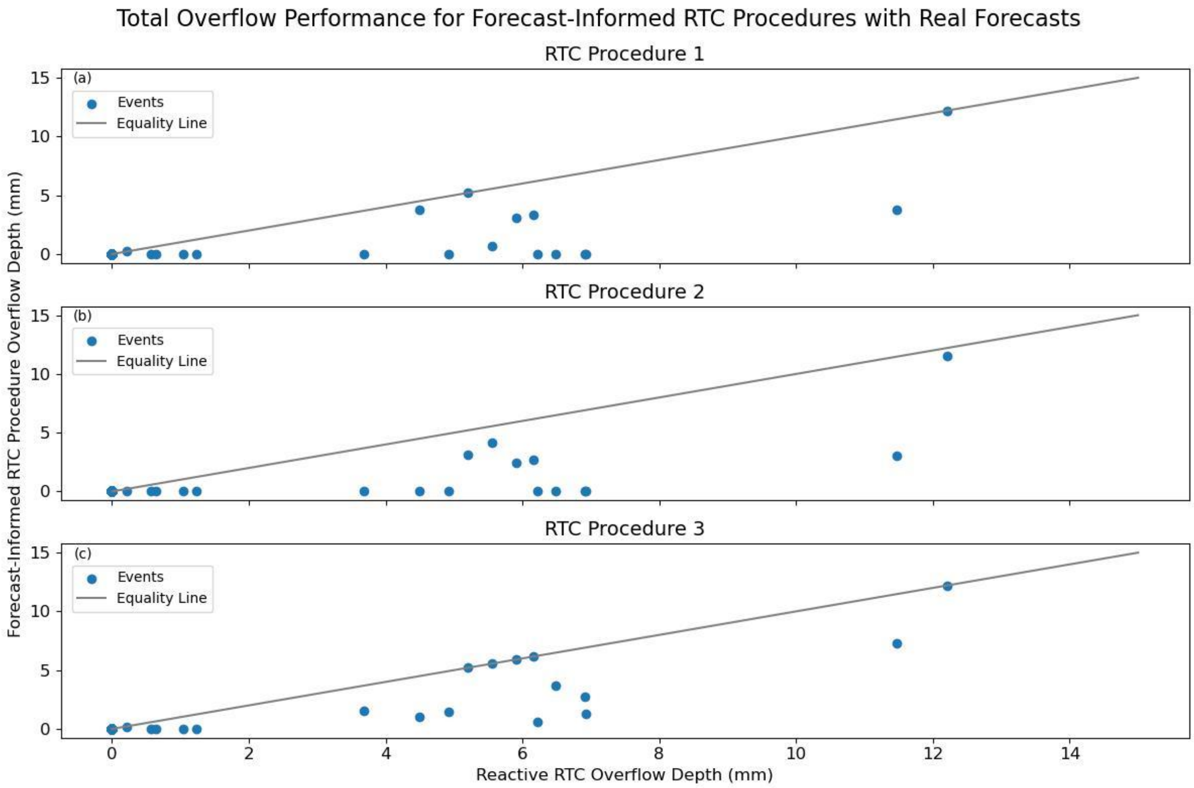

Figure 8.6: Each panel shows a comparison of overflow depths when using the reactive RTC versus a forecast-informed RTC. The equality line represents the overflow depth for an event when the reactive RTC is applied.

As an example, the event with the highest maximum intensity (18.04 mm/hr) is examined (Figure 8.7). This event demonstrated that flooding was prevented by Procedure 1 and Procedure 2, whereas Procedure 3 only reduced the total overflow depth. Procedure 1 started pumping first, as expected, due to its use of the longest forecast horizon. Additionally, the pump remained active for 12 hours, indicating that two consecutive forecasts exceeded the rainfall threshold. Both Procedure 2 and Procedure 3 started pumping at the same time, 6 hours after Procedure 1, despite using different forecast horizons. As the event drew nearer, Procedure 2 carried out more pumping, while Procedure 3 did not. This behaviour corresponds to the observed rainfall data, as no rainfall is observed within a 12-hour window from this point. Just as it started raining, the forecast exceeded the rainfall threshold for Procedure 2 and 3, activating the pump. The forecast uncertainty resulted in Procedure 1 not being triggered here. It can also be observed that most of the rainfall is predicted at the start of the event, with very little difference between the 12-hour and 24-hour horizons in cases where both procedures activate the pump simultaneously. The differences in pumping duration and timing were reflected in the water levels: only Procedure 3 resulted in flooding, whereas Procedure 1 and Procedure 2 successfully lowered the water level sufficiently in advance, completely preventing flooding.

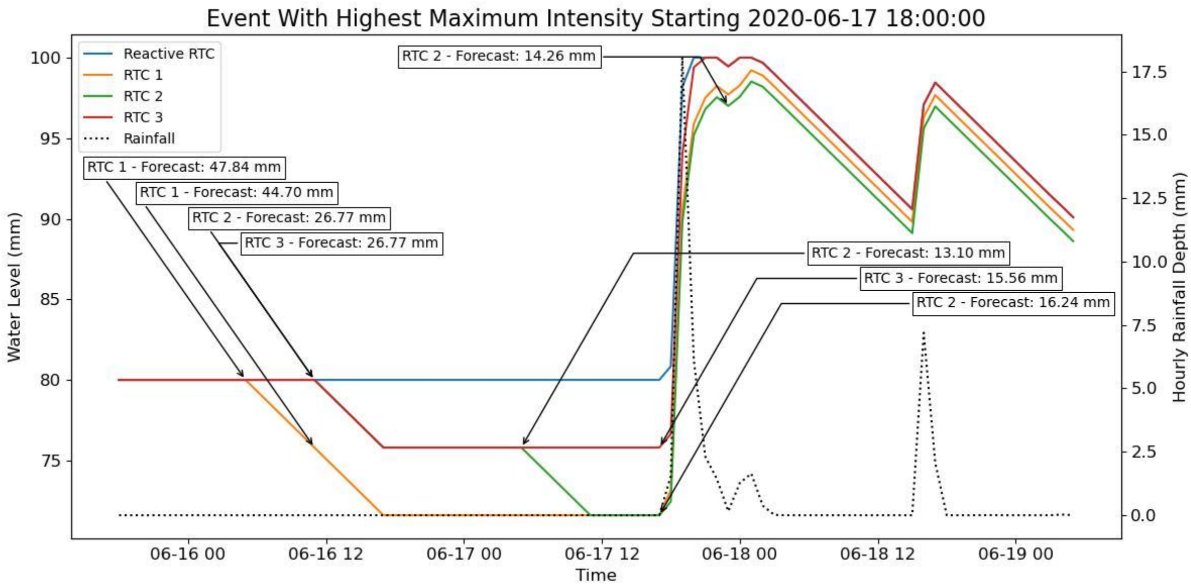

Figure 8.7: Performance of forecast-informed RTC in mitigating flooding for the event with the highest maximum intensity that caused flooding using the reactive RTC. Annotations indicate where the pump is activated by the forecast for the different procedures and predicted rainfall depth.

Uncertainty and inaccuracy in the rainfall forecasts can deteriorate the performance of the forecast-informed RTC procedures. A lengthening forecast horizon increases the root-mean-square error (RMSE), while overestimating rainfall depth commonly occurs when the forecast predicts a high mean intensity. Overestimations can activate the pump unnecessarily, reducing the water level when it is not required. This is evident in cases where the water level falls notably below the initial level. An example of this is shown (Figure 8.8), where all procedures overestimated the total rainfall depth, despite variations in forecast horizons and rainfall depth thresholds, and in this case the reactive RTC operation was most desirable. At the event's end, all forecast-informed procedures triggered the pump again, even though there was no rainfall for the next 12 hours (based on the MIT definition). This pump activation can be due to the used forecast horizons in Procedure 1 and 2, which are longer than the MIT. Other reasons may be overestimation, poor timing, or poor location with the precipitation falling in a different forecast pixel, as specifically represented with Procedure 3.

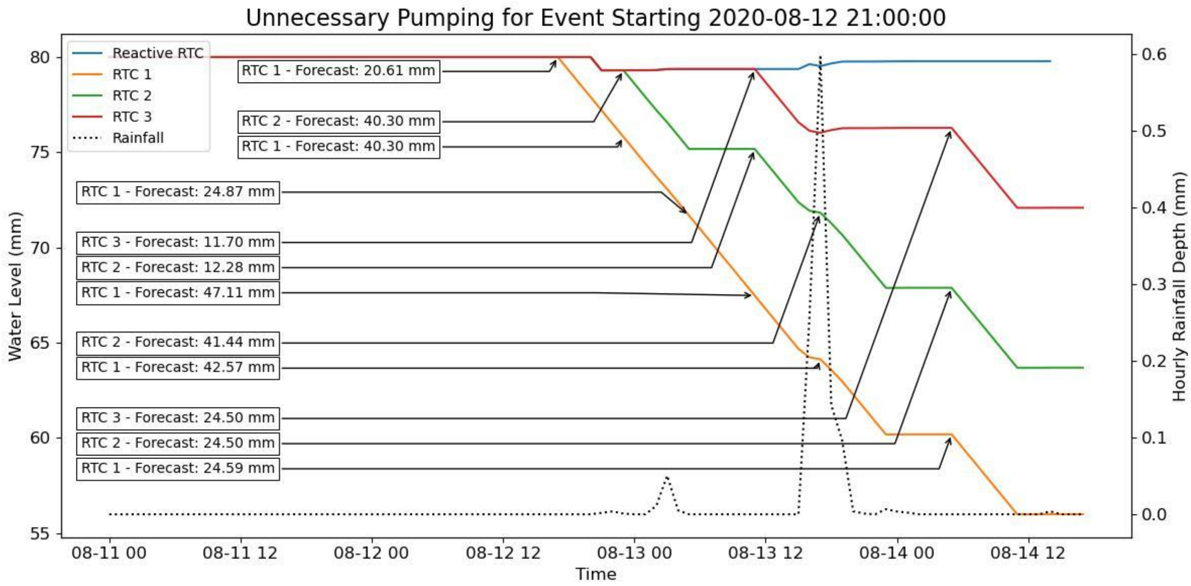

Figure 8.8: Example of an event where forecast-informed RTC procedures resulted in excessive pumping compared to the reactive RTC. Annotations indicate where the pump is activated by the forecast for the different RTC procedures and predicted rainfall depth.

A common factor in events where the water level dropped more than, e.g., 10 mm below the initial water level was small rainfall depth and low maximum intensity, with a few exceptions. This aligns with previous research, indicating that forecasts with higher mean intensities tend to overestimate the rainfall depth. Procedure 3, which used the shortest horizon and the highest intensity threshold, was the most conservative in pumping. Compared to the other forecast-informed RTC procedures, the pump was activated more frequently but for fewer hours and prevented fewer flooding events. Consequently, Procedure 3 also resulted in the fewest instances of unnecessary water level reduction. Procedure 1, with its 48-hour horizon, followed, while Procedure 2, although most effective at preventing flooding, exhibited the highest occurrence of unnecessary pumping.

The results from the continuous evaluation matched the event-based findings, showing that all forecast-informed RTC procedures reduced overflow depth compared to the reactive RTC. Specifically, the reactive RTC recorded an overflow depth of 183.61 mm, while Procedure 2, which had the best flood mitigation performance, reduced this to 55.22 mm. Procedure 1 followed with an overflow depth of 77.14 mm, and Procedure 3 had the highest overflow among the forecast-informed RTCs at 130.41 mm. These outcomes are also reflected in the pump operation of the different RTCs. Longer pumping durations resulted in better flood mitigation, while the frequency of on/off switches declined with longer forecast horizons. For the reactive RTC, the pump operated for 38,399 hours with 7,857 activations. Among the forecast-informed RTCs, Procedure 2 had the longest pump operation at 38,582 hours and 7,262 activations, followed by Procedure 1 with 38,551 hours and 7,539 activations. Procedure 3 had the shortest pump duration, operating for 38,474 hours but with the highest number of activations at 7,680.

Incorporation of forecast data, their uncertainty and increased pumping resulted in longer periods where the water level remained below the target of 80 mm, as well as lower minimum water levels (Figure 8.9). The reactive RTC was below the target 88.43% of the time, followed by Procedure 3 at 89.19%, Procedure 1 at 90.78%, and Procedure 2 at 91.34%. Although the water level fell below the target for a significant portion of time in all cases, it was typically only slightly lower due to the pump's capacity and activation thresholds. The reactive RTC's lowest possible water level of 79.3 mm was used as a new threshold to assess how much time each procedure spent below this level. This change reduced the percentage of time below the target to 3.31% for Procedure 3, 10.38% for Procedure 1, and 11.88% for Procedure 2.

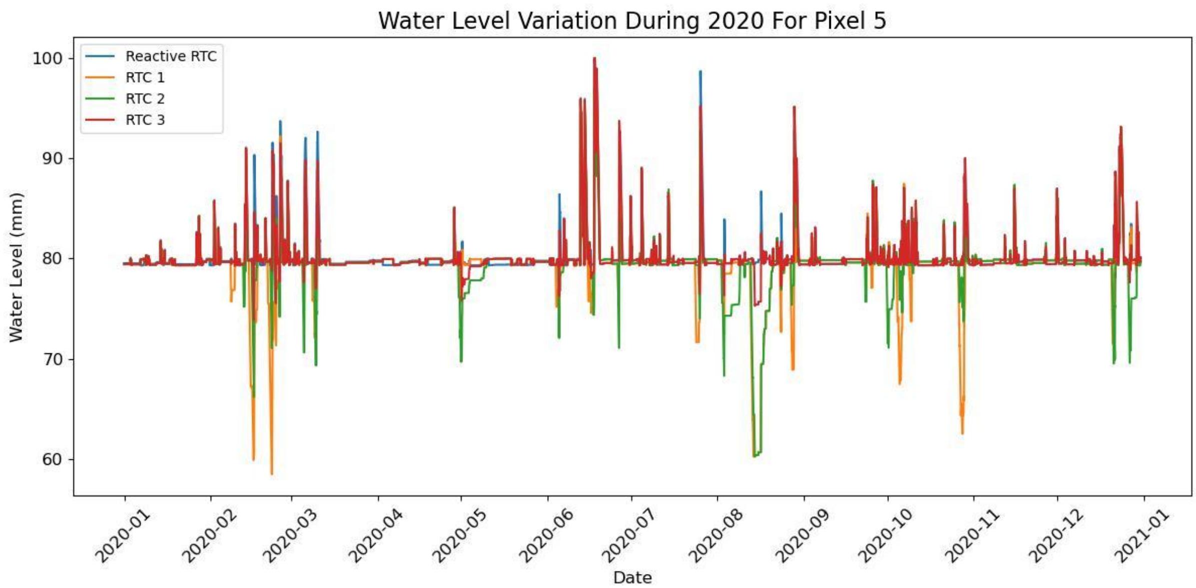

Figure 8.9: Changes in water level for the different procedures. Where forecast data is missing, the reactive RTC has been used.

9 Conclusions RTC

Overall, the efficiency in mitigating urban pluvial flooding relied on the choice of threshold and forecast horizon when using forecast-informed RTC procedures, as well as the forecast accuracy. The first part of the evaluation demonstrated that a reactive RTC can be applied to urban canals to effectively prevent flooding and the second part that incorporating forecast data into an RTC increased this potential. However, uncertainties and inaccuracies in rainfall forecasts limited the potential. A shorter horizon for the same intensity threshold improved flood mitigation performance, as seen with Procedure 1 and 2. This is consistent with previous findings (subsection 5.1.2), which revealed decreased predictability for moderate- to high-intensity rainfall events when considering longer horizons. When the intensity threshold was raised and horizon shortened (as in Procedure 3), fewer flooding events were prevented. This decline may be attributed to (1) forecast uncertainty limiting the number of events exceeding the threshold within the set horizon; or (2) insufficient time to adequately lower the water level.

While incorporating rainfall forecasts into the RTC facilitated in mitigating urban pluvial flooding, forecast errors and uncertainties restricted the performance. With perfect rainfall forecasts, the reactive RTC was improved, with forecast-informed Procedure 1 and 2 eliminating pluvial flooding. These results also indicated that Procedure 3, with a 12-hour horizon, did not prevent all flooding events even with perfect forecast data. Therefore, while forecast accuracy increases with shorter horizons, a strategy's ability to prevent pluvial flooding also depends on the lead time available before an event to take adequate action.

Using real rainfall forecasts, Procedure 2 demonstrated the best flood mitigation performance, as expected based on the forecast assessment. Its combination of rainfall depth threshold and forecast horizon achieved the highest probability of detection (POD) and lowest false alarm ratio (FAR) among the developed procedures. While the RMSE and depth difference of Procedure 2 exceeded the values of Procedure 3, which operated with a 12-hour horizon, the difference was smaller between these two horizons, than between the 24- and 48-hour horizons. Procedure 1 ranked second best in flood mitigation, despite having the lowest POD and highest FAR, in addition to the largest RMSE and depth difference. This emphasises the importance of adequate lead time in performance. Procedure 3 had the most flooding events; however, 7 of 14 were due to limited lead time, making it the second-best when only considering flooding events due to inaccuracy in the forecasts.

Despite the forecast errors and uncertainties, improvements in flood prevention were demonstrated when forecasts were integrated into the reactive RTC strategy. However, the effectiveness of these strategies was limited by the design of the new procedures, data quality, and pump operation. The frequency of pump on/off switches decreased whilst the pumping hours increased in the forecast-informed RTC strategies compared to the reactive RTC. While increased pumping reduces flooding frequency and total overflow depth, it is associated with higher maintenance and energy costs, and frequent pump activation may damage the pumping facilities (Jafari et al., 2018). This increased pumping activity was also reflected in the bucket model's water levels, where lower levels were observed alongside a higher percentage of time below the target. Although the water level generally remained only slightly below the target, larger deviations were also observed with the forecast-informed RTCs. While a lower water level is beneficial in flood mitigation, it can also have potential negative impacts. Fluctuations in water level in lakes and rivers are shown to affect the physical environment, biota, and ecosystems (Leira & Cantonati, 2008), and similar effects may be expected in urban water canals. Additionally, public perception of the area's attractiveness (Stroble & Taylor, 2020) may be influenced by visibly empty or almost empty canals. Therefore, a balance between the benefits and drawbacks is necessary before implementation.

10 Project Discussions

The reported outcomes from this thesis have been influenced by the developed methodology and decisions. In the discretisation of the rainfall, based on the rain-gauge adjusted radar dataset, an event was defined as having a depth larger than 0 mm. By using this threshold value, an excessive number of minor or negligible rainfall events have been included in the analyses. The inclusion of these may have skewed the results towards insignificant events, with the impacts of these very small events on the operation being irrelevant and making the data less meaningful for decision-making. This could have been prevented by filtrating out negligible events by first determining a reasonable threshold. For comparison, Shrestha et al. (2012) defined a rainfall event as any event exceeding 0.1 mm/3hr, excluding trivial events from their evaluation.

A similar approach could have been beneficial in this case, where the chosen rainfall definition, combined with using every forecast and observed event pairs – even when both reflect no rainfall – may have influenced the outcome of the forecast assessment. By including non-events in the assessment, the chosen performance metrics could be improved, as non-events are easier to match. Furthermore, there could be a dominance of these non-events, potentially masking issues and troubles in predicting actual rainfall events. The reliability of the forecasts may be overestimated by reporting higher accuracy. As a result, the true value and limitations of the forecasts become obscured, making it increasingly difficult to evaluate their performance in critical, real-world applications. This can hinder the development of effective decision-making strategies based on these forecasts. However, by using an increasing threshold to determine rainfall events in the forecast assessment, these problems can be mitigated, especially for critical, flood-inducing events.

In the evaluation of the forecast horizon's influence, RMSE and depth difference was calculated for every hour. This provided a continuous measurement of the changes in accuracy over the entire forecast horizon and variations in trends and performance could be identified at specific hours. Consequently, insights for different forecast horizons were offered, considered to be valuable for tailoring operational decisions. However, POD, SPC, and FAR were only evaluated for 6-hour intervals, corresponding to the forecast's updating frequency. This decision was made since the developed mitigation procedures were re-evaluated for each new forecast. In the assessment, intensity thresholds were used rather than cumulative depth, to reduce the impacts of the horizon lengths on the result interpretation. Nonetheless, mean intensities that appear similar may reflect very different hydrological impacts due to the variations of the considered horizons. As such, extreme short-term rainfall may be diluted when averaged over a longer forecast horizon, potentially underestimating their significance and obscuring details critical to specific applications.

The mean forecast intensity was also applied to assess the correlation between the forecasted and observed depth. This analysis was only carried out for the entire forecast horizon, possibly losing valuable information. The obtained results may not be a true representation of the relationship for shorter horizons. This approach does not capture the differences and nuances of shorter and longer horizons and misses possible trends or shifts in performance. Additionally, only considering the entire horizon may cause challenges in operational decisions as they rely on specific forecast horizons, which have not been assessed by analysing only the entire horizon.

Using NWP may have some limitations affecting the feasibility of designing RTC procedures based on these predictions. In addition to the uncertainty and inaccuracies related to the forecasts themselves, the output frequency may constitute a problem. The predictions are updated every 6 hours, however; the necessary computational time is 3 hours. During the time between the initiation of the forecast and the distribution to the users, significant changes to the weather can occur, reducing the reliability. This discrepancy between the start of a forecast and when it becomes available has not been considered in the second part of this thesis, consequently leading to actions being taken at least 3 hours before it is possible.

The rainfall used for the event-based assessment of the RTC procedures are the result of using an MIT of 12 hours, replicating van der Werf et al. (2023). This was recommended as it was the approximate emptying time of the system. However, that study focused on RTC in an UDS with a storage capacity of 10.13 mm. The approximate emptying time of the considered urban canals in this thesis, based on the storage capacity and pump capacity, is 30 hours. Consequently, there may have been an increase of smaller, separate events with reduced rainfall depth. Discretising the events may have led to unnecessary splits (see Figure 8.8) affecting the RTC performance based on the evaluation metrics. However, some of these errors were mitigated in the continuous evaluation.

When developing the forecast-informed RTC procedures, the forecasts were assumed to be correct and were directly used. However, the performance may have been improved with more complex procedures including the rainfall probability. This way actions could be delayed for uncertain forecasts and events with low probability, possibly reducing unnecessary pumping and percentage of time with water below the target level. Additionally, the model has been simplified by assuming that the entire rainfall depth contributes immediately to the water level without any water loss or time lag. This assumption neglects the natural delays and complex processes, such as infiltration, runoff, and storage, that affect how rainfall influences water levels. Typically, real-world systems show time lags in water flows, meaning that the water level may rise more gradually and less than in these simulations. By adding the rainfall depth instantly, the flood risk may be exacerbated. These weaknesses could be reduced by applying a coupled model, integrating more components of the hydrological system and providing more realistic predictions.

Furthermore, in the evaluation, infiltration, seepage, and evaporation have been excluded. This oversimplification of the hydrological process might influence the flooding frequency and overflow depths. By excluding these factors in the assessment, the model becomes more computationally feasible and easier to interpret. For example, evaporation plays a role in reducing water levels, especially during dry spells. By not considering this, the model may overestimate water accumulation and peak levels during prolonged periods without rainfall. This could lead to inaccurate flood forecasts, making them less reliable for decision-making in real-word applications.

Developing forecast-informed RTC procedures based on the forecast assessment and its uncertainty could lead to overly conservative or ineffective actions, such as premature or unnecessary interventions. To mitigate this risk, having a backup measure – such as reactive RTC – is essential to prevent performance deterioration. Furthermore, the outcomes obtained in this thesis have been influenced by a combination of uncertainty and errors in both the forecasts and the developed RTC procedures.

11 Project Conclusions

Urbanisation, changes in land use, and shifting rainfall patterns intensifies the pressure on urban water management and increases the risk of flooding. Conventional mitigation measures are often limited by financial and spatial constraints which can be circumvented with RTC systems. These systems require only minor expansions of the existing system and aims to utilise the storage capacity of both the UDS and urban surface water system more efficiently. This research investigated the errors and uncertainties related to long-term rainfall forecasts and their accuracy’s effect on forecast-informed RTCs’ ability to mitigate urban pluvial flooding.

The first step in this thesis involved assessing and evaluating the accuracy and uncertainties in rainfall forecasts using cumulative rainfall depth, forecast horizon, and forecast intensity. The evaluation process required discretising events to align with the forecast data. Since each forecast event had a fixed duration, multiple forecasts were sometimes needed to cover longer observed events. By focusing solely on matching hours and areas, potential errors in timing may have influenced the results. It was found that longer forecast horizons and greater rainfall depths deteriorated the forecast accuracy. Larger depths were specifically associated with an increase in false alarms and a decrease in the probability of correct predictions.

Although these trends were consistent across all evaluated forecast horizons, forecast uncertainty intensified with lengthening horizons. Forecast accuracy was higher for small rainfall depths in longer horizons, likely due to decreased importance of timing. However, since these small depths were insufficient to cause flooding based on this case study’s setpoints, their practical utility for flood mitigation is limited. Therefore, a shorter forecast horizon may be more effective for flood mitigation. This can be supported by prior findings, with forecast horizons shorter than 48 hours reported (Ashok & Pekkat, 2022) to be optimal. However, it has also been demonstrated (Jabbari et al., 2020) that lead time dependency is nearly negligible for forecast horizons under 36 hours, suggesting that using the entire 48-hour forecast horizon examined in this study could be appropriate in real-world applications.

Additionally, the evaluation of forecast accuracy demonstrated a tendency to overestimate rainfall depth when the mean forecast intensity was high, whereas low-intensity forecasts generally underestimated the cumulative rainfall depth. Forecast accuracy is influenced by characteristics such as rainfall intensity and event duration. Similar to Fabry and Seed (2009), rainfall events with higher intensities were overestimated, while longer horizons were more skilful for longer event durations consistent with Imhoff et al. (2020). In retrospect, it would be advisable to investigate the relationship between maximum forecast intensity and forecast uncertainty. This is recommended because the flooding events that could occur were characterised by high maximum intensities that significantly exceeded the pump capacity. Additionally, Wasko and Nathan (2019) noted that large floods were a result of high rainfall peaks. Furthermore, the results may have been influenced by the distribution of rainfall events, with relatively few high-intensity events amplifying their effect. As such, these results should be interpreted with caution.

The second part of this case study evaluated RTCs' ability to mitigate urban pluvial flooding. The reactive RTC, relying solely on *in-situ* measurement of the water level, proved effective for the moderate-intensity rainfall events but showed limitations under more extreme precipitation. Prior research (Di Matteo et al., 2019) has indicated that shorter-duration events are generally less critical than longer events, as they tend to produce smaller rainfall depths. This discrepancy in the impact of rainfall events may be influenced by available storage capacity, which limits the system's ability to handle longer, more intense storms. Similarly, Liang et al. (2021) found that the critical event duration depended on the storage capacity of the system. However, these studies relied on estimated design peak flows for given durations and annual exceedance probabilities (AEPs) by taking the average peak flow from multiple different temporal patterns, potentially obscuring the characteristics of individual rainfall events.

Forecast-informed RTC solutions have been shown to enhance performance by incorporating rainfall forecasts to overcome limitations of storage capacity (Sun et al., 2023). However, similar to findings by Galelli et al. (2014) and Sun et al. (2024), it is evident from the developed methodology and analyses that forecast-informed RTCs introduce additional complexity into the decision-making, suggesting a trade-off between enhanced flood mitigation and increased operational complexity. By incorporating rainfall forecasts, the RTC procedures enabled proactive capacity management, consistent with Xu et al. (2020), who found that pre-emptive water release ahead of forecasted rainfall events reduced uncontrolled overflows and lowered the flood frequency.

The maximum potential overflow reduction achieved in this case study (70% for Procedure 2) is in the upper region of previously reported (Li, 2020) reductions, which vary from 40% to 70% in flooding magnitude and peak water levels with RTC implementation. However, these results are influenced by the chosen control procedure and the system's complexity. For instance, more complex Model Predictive Control (MPC) systems have shown (Sadler et al., 2019) flood volume reductions up to 78%, calculated by comparing total flood volume under MPC and rule-based controls, which corresponds to the reactive RTC used in this case study. Shishegar et al. (2019) also reported improved performance with 73 to 95% reductions in peak flows across four different scenarios using an optimisation-rule-based model compared to a static approach. The comparison with a static approach may have increased the reported performance, as it represents a forecast-informed RTC-to-no-control comparison. Moreover, the latter study employed 1 control point, and the former used 2, comparable to the single control utilised in this analysis. This similarity in number of control points, may contribute to the modest variations in reported performance despite increased system complexity, as controllability is reported (van der Werf et al., 2022) to be a key variable in RTC efficacy.

Another reason for the discrepancy between observed and reported performance may stem from the type of rainfall forecast employed, as the effectiveness of forecast-informed RTCs is highly dependent on forecast accuracy. Numerous studies (e.g., Di Matteo et al., 2019; Sun et al., 2023) use ideal rainfall forecasts to put emphasis on the RTC methodology, overlooking forecast error, which constitute a significant source of uncertainty. As a result, the performance metrics reported in these studies may appear inflated when compared to the realistic improvements achievable through the use of actual forecast data and outcomes in this thesis. Therefore, it is essential to utilise real forecast models in conjunction with real observed data. Assessing the accuracy of the forecast data is critical prior to designing and implementing forecast-informed RTC systems across various contexts to ensure comparable dynamics. Incorporating these assessments into the methodology will enhance the reliability of results and ensure that findings are relevant and applicable in real-word scenarios (van der Werf et al., 2023).

This thesis aimed at investigating RTC systems' ability to mitigate urban pluvial flooding and minimise negative side effects by incorporating rainfall forecasts into the decision-making, thereby enhancing the urban water management strategies. The results indicated that forecast-informed RTCs significantly improved the flood mitigation performance, achieving reductions in overflow depth and flooding frequency compared to a less complex reactive RTC approach. Moreover, the analyses revealed that the accuracy of rainfall forecasts plays a crucial role in the effectiveness of RTC, despite the RTC procedures not being optimised.

These findings underscore the importance of accurate rainfall forecasts in forecast-informed RTC procedures to improve flood management. The insight from this research can inform future policy decisions and guide the implementation of advanced RTC systems in urban settings. Despite the promising results, this study faced limitations related to the simplified bucket model of the urban canals among others. Future research should consider focusing on more complex UDS and surface water systems and explore the application in diverse geographic contexts. In conclusion, forecast-informed RTC strategies not only enhances the flood mitigation efforts but also represents a critical step toward sustainable water management in the face of increasing urbanisation and climate change.

References

Acosta-Coll, M., Ballester-Merelo, F., Martinez-Peiró, M., & De la Hoz-Franco, E. (2018). Real-Time Early Warning System Design for Pluvial Flash Floods—A Review. *Sensors*, *18*(7). https://doi.org/10.3390/s18072255

Arup. (2019). *The City Water Resilience Approach - Rotterdam City Characterisation Report*.

Ashley, R. M. (2007). *Advances in urban flood management* (R. Ashley, S. Garvin, E. Pasche, A. Vassilopoulos, & C. Zevenbergen, Eds.). Taylor & Francis. https://doi.org/10.1201/9780203945988

Ashok, S. P., & Pekkat, S. (2022). A systematic quantitative review on the performance of some of the recent short-term rainfall forecasting techniques. *Journal of Water and Climate Change*, *13*(8), 3004-3029. https://doi.org/10.2166/wcc.2022.302

Azizi, K., Diko, S. K., Saija, L., Zamani, M. G., & Meier, C. I. (2022). Integrated community-based approaches to urban pluvial flooding research, trends and future directions: A review. *Urban Climate*, *44*, 101237. https://doi.org/10.1016/j.uclim.2022.101237

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47-55. https://doi.org/10.1038/nature14956

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

Beersma, J., Hakvoort, H., Jilderda, R., Overeem, A., & Versteeg, R. (2019). *Neerslagstatistiek en -reeksen voor het waterbeheer 2019*. KNMI.

Bilodeau, Pelletier, & Duchesne. (2018). Real-time control of stormwater detention basins as an adaptation measure in mid-size cities. *Urban Water Journal*, *15*(9), 858-867. https://doi.org/10.1080/1573062X.2019.1574844

Borsányi, P., Benedetti, L., Dirckx, G., De Keyser, W., Muschalla, D., Solvi, A.-M., Vandenberghe, V., Weyand, M., & Vanrolleghem, P. A. (2008). Modelling real-time control options on virtual sewer systems. *Journal of Environmental Engineering and Science*, *7*(4), 395-410. https://doi.org/10.1139/s08-004

Brendel, C. E., Dymond, R. L., & Aguilar, M. F. (2020). Integration of quantitative precipitation forecasts with real-time hydrology and hydraulics modeling towards probabilistic forecasting of urban flooding. *Environmental Modelling & Software*, *134*, 104864. https://doi.org/10.1016/j.envsoft.2020.104864

Buizza, R. (2019). Introduction to the special issue on “25 years of ensemble forecasting”. *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 1-11. https://doi.org/10.1002/qj.3370

Butler, D., & Davies, J. W. (2011). *Urban drainage* (3rd ed.). Spon Press.

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

Casares, P. S., Schofield, R., DePippo, K., & Sewell, N. (2024). Combining small-scale sustainable drainage systems with real-time control systems. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, *177*(5), 11-19. https://doi.org/10.1680/jcien.23.00123

Chen, W., Wang, W., Huang, G., Wang, Z., Lai, C., & Yang, Z. (2021). The capacity of grey infrastructure in urban flood management: A comprehensive analysis of grey infrastructure and the green-grey approach. *International Journal of Disaster Risk Reduction*, *54*, 102045. https://doi.org/10.1016/j.ijdrr.2021.102045

Coiffier, J. (2011). *Fundamentals of numerical weather prediction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511734458

Di Matteo, M., Liang, R., Maier, H. R., Thyer, M. A., Simpson, A. R., Dandy, G. C., & Ernst, B. (2019). Controlling rainwater storage as a system: An opportunity to reduce urban flood peaks for rare, long duration storms. *Environmental Modelling & Software*, *111*, 34-41. https://doi.org/doi.org/10.1016/j.envsoft.2018.09.020

EEA. (2020). *Urban adaptation in Europe: how cities and towns respond to climate change* (12/2020).

European Commission. (2007). *Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks* (Official Journal L 288). http://data.europa.eu/eli/dir/2007/60/oj

Fabry, F., & Seed, A. W. (2009). Quantifying and predicting the accuracy of radar-based quantitative precipitation forecasts. *Advances in Water Resources*, *32*(7), 1043-1049. https://doi.org/doi.org/10.1016/j.advwatres.2008.10.001

Falconer, R. H., Cobby, D., Smyth, P., Astle, G., Dent, J., & Golding, B. (2009). Pluvial flooding: new approaches in flood warning, mapping and risk management. *Journal of Flood Risk Management*, *2*(3), 198-208. https://doi.org/10.1111/j.1753-318X.2009.01034.x

Fewtrell, L., Kay, D., & Ashley, R. (2008). Flooding and health - an evaluation of the health impacts of urban pluvial flooding in the UK. In L. Fewtrell & D. Kay (Eds.), *Health Impact Assessment for Sustainable Water Management* (pp. 121-154). IWA Publishing. https://doi.org/10.2166/9781780401874

Fu, G., Butler, D., Khu, S.-T., & Sun, S. A. (2011). Imprecise probabilistic evaluation of sewer flooding in urban drainage systems using random set theory. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2009WR008944

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

Galelli, S., Goedbloed, A., Schwanenberg, D., & van Overloop, P.-J. (2014). Optimal Real-Time Operation of Multipurpose Urban Reservoirs: Case Study in Singapore. *Journal of Water Resources Planning and Management*, *140*(4), 511-523. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000342

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

Habets, F., LeMoigne, P., & Noilhan, J. (2004). On the utility of operational precipitation forecasts to served as input for streamflow forecasting. *Journal of Hydrology*, *293*(1), 270-288. https://doi.org/10.1016/j.jhydrol.2004.02.004

Hammond, M. J., Chen, A. S., Djordjević, S., Butler, D., & Mark, O. (2015). Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, *12*(1), 14-29. https://doi.org/10.1080/1573062X.2013.857421

Heuvelink, D., Berenguer, M., Brauer, C. C., & Uijlenhoet, R. (2020). Hydrological application of radar rainfall nowcasting in the Netherlands. *Environment International*, *136*, 105431. https://doi.org/10.1016/j.envint.2019.105431

Imhoff, R. O., Brauer, C. C., Overeem, A., Weerts, A. H., & Uijlenhoet, R. (2020). Spatial and Temporal Evaluation of Radar Rainfall Nowcasting Techniques on 1,533 Events. *Water*

Resources Research, *56*(8), e2019WR026723. https://doi.org/doi.org/10.1029/2019WR026723

Imhoff, R. O., Brauer, C. C., van Heeringen, K. J., Uijlenhoet, R., & Weerts, A. H. (2022). Large-Sample Evaluation of Radar Rainfall Nowcasting for Flood Early Warning. *Water Resources Research*, *58*(3), e2021WR031591. https://doi.org/https://doi.org/10.1029/2021WR031591

Jabbari, A., So, J.-M., & Bae, D.-H. (2020). Precipitation Forecast Contribution Assessment in the Coupled Meteo-Hydrological Models. *Atmosphere*, *11*(1). https://doi.org/10.3390/atmos11010034

Jacobs, J. C. J. (2012). The Rotterdam approach: connecting water with opportunities. In C. Howe & C. Mitchell (Eds.), *Water sensitive cities* (pp. 251-263). IWA Publishing. https://doi.org/10.2166/9781780400921

Jafari, F., Jamshid Mousavi, S., Yazdi, J., & Kim, J. H. (2018). Real-Time Operation of Pumping Systems for Urban Flood Mitigation: Single-Period vs. Multi-Period Optimization. *Water Resources Management*, *32*(14), 4643-4660. https://doi.org/10.1007/s11269-018-2076-4

Jang, J., & Hong, S.-Y. (2014). Quantitative forecast experiment of a heavy rainfall event over Korea in a global model: horizontal resolution versus lead time issues. *Meteorology and Atmospheric Physics*, *124*(3), 113-127. https://doi.org/10.1007/s00703-014-0312-x

Jee, J.-B., & Kim, S. (2017). Sensitivity Study on High-Resolution WRF Precipitation Forecast for a Heavy Rainfall Event. *Atmosphere*, *8*(6). https://doi.org/10.3390/atmos8060096

Jha, A. K., Bloch, R., & Lamond, J. (2012). *Cities and flooding : a guide to integrated urban flood risk management for the 21st century*. World Bank. https://hdl.handle.net/10986/2241

Jian, W., Li, S., Lai, C., Wang, Z., Cheng, X., Lo, E. Y.-M., & Pan, T.-C. (2021). Evaluating pluvial flood hazard for highly urbanised cities: a case study of the Pearl River Delta Region in China. *Natural Hazards*, *105*(2), 1691-1719. https://doi.org/10.1007/s11069-020-04372-3

KNMI. (2014). *KNMI'14: Climate Change scenarios for the 21st Century – A Netherlands perspective* (Scientific Report WR2014-01). KNMI. www.climatescenarios.nl

Kroll, S., Weemaes, M., Van Impe, J., & Willems, P. (2018). A Methodology for the Design of RTC Strategies for Combined Sewer Networks. *Water*, *10*(11). https://doi.org/10.3390/w10111675

Kändler, N., Annus, I., Vassiljev, A., & Puust, R. (2020). Peak flow reduction from small catchments using smart inlets. *Urban Water Journal*, *17*(7), 577-586. https://doi.org/10.1080/1573062X.2019.1611888

Leira, M., & Cantonati, M. (2008). Effects of water-level fluctuations on lakes: an annotated bibliography. In K. M. Wantzen, K.-O. Rothhaupt, M. Mörtl, M. Cantonati, L. G. Tóth, & P. Fischer (Eds.), *Ecological Effects of Water-Level Fluctuations in Lakes* (pp. 171-184). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9192-6_16

Li, J. (2020). A data-driven improved fuzzy logic control optimization-simulation tool for reducing flooding volume at downstream urban drainage systems. *Science of The Total Environment*, *732*, 138931. https://doi.org/10.1016/j.scitotenv.2020.138931

Li, J., & Burian, S. J. (2023). Evaluating real-time control of stormwater drainage network and green stormwater infrastructure for enhancing flooding resilience under future rainfall projections. *Resources, Conservation and Recycling*, *198*, 107123. https://doi.org/10.1016/j.resconrec.2023.107123

Liang, R., Thyer, M. A., Maier, H. R., Dandy, G. C., & Di Matteo, M. (2021). Optimising the design and real-time operation of systems of distributed stormwater storages to reduce urban flooding at the catchment scale. *Journal of Hydrology*, *602*, 126787. https://doi.org/10.1016/j.jhydrol.2021.126787

Maiolo, M., Palermo, S. A., Brusco, A. C., Pirouz, B., Turco, M., Vinci, A., Spezzano, G., & Piro, P. (2020). On the Use of a Real-Time Control Approach for Urban Stormwater Management. *Water*, *12*(10). https://doi.org/10.3390/w12102842

Martínez-Gomariz, E., Forero-Ortiz, E., Guerrero-Hidalga, M., Castán, S., & Gómez, M. (2020). Flood Depth–Damage Curves for Spanish Urban Areas. *Sustainability*, *12*(7). https://doi.org/10.3390/su12072666

Martínez-Gomariz, E., Locatelli, L., Guerrero, M., Russo, B., & Martínez, M. (2019). Socio-Economic Potential Impacts Due to Urban Pluvial Floods in Badalona (Spain) in a Context of Climate Change. *Water*, *11*(12). https://doi.org/10.3390/w11122658

Meneses, E. J., Gaussens, M., Jakobsen, C., Mikkelsen, P. S., Grum, M., & Vezzaro, L. (2018). Coordinating Rule-Based and System-Wide Model Predictive Control Strategies to Reduce Storage Expansion of Combined Urban Drainage Systems: The Case Study of Lundtofte, Denmark. *Water*, *10*(1). https://doi.org/10.3390/w10010076

Meng, F., Fu, G., & Butler, D. (2017). Cost-Effective River Water Quality Management using Integrated Real-Time Control Technology. *Environmental Science & Technology*, *51*(17), 9876-9886. https://doi.org/10.1021/acs.est.7b01727

Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Allaire, M., & Matthew, R. A. (2018). What Is Nuisance Flooding? Defining and Monitoring an Emerging Challenge. *Water Resources Research*, *54*(7), 4218-4227. https://doi.org/10.1029/2018WR022828

Mounce, S. R., Shepherd, W., Ostojin, S., Abdel-Aal, M., Schellart, A. N. A., Shucksmith, J. D., & Tait, S. J. (2019). Optimisation of a fuzzy logic-based local real-time control system for mitigation of sewer flooding using genetic algorithms. *Journal of Hydroinformatics*, *22*(2), 281-295. https://doi.org/10.2166/hydro.2019.058

Netzel, L. M., Heldt, S., Engler, S., & Denecke, M. (2021). The importance of public risk perception for the effective management of pluvial floods in urban areas: A case study from Germany. *Journal of Flood Risk Management*, *14*(2), e12688. https://doi.org/10.1111/jfr3.12688

Nielsen, N. H., Ravn, C., & Mølbye, N. (2010, Jun). Implementation and design of a flexible RTC strategy in the sewage system in Kolding, Denmark. Novatech 2010 - 7ème Conférence internationale sur les techniques et stratégies durables pour la gestion des eaux urbaines par temps de pluie / 7th International Conference on sustainable techniques and strategies for urban water management, Lyon, France.

Overeem, A., Buishand, T. A., & Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, *45*(10). https://doi.org/10.1029/2009WR007869

Overeem, A., Holleman, I., & Buishand, T. A. (2009). Derivation of a 10-Year Radar-Based Climatology of Rainfall. *Journal of Applied Meteorology and Climatology*, *48*(7), 1448-1463. https://doi.org/10.1175/2009JAMC1954.1

Palla, A., Colli, M., Candela, A., Aronica, G. T., & Lanza, L. G. (2018). Pluvial flooding in urban areas: the role of surface drainage efficiency. *Journal of Flood Risk Management*, *11*(S2), S663-S676. https://doi.org/10.1111/jfr3.12246

Pirone, D., Cimorelli, L., Del Giudice, G., & Pianese, D. (2023). Short-term rainfall forecasting using cumulative precipitation fields from station data: a probabilistic machine learning approach. *Journal of Hydrology*, *617*, 128949. https://doi.org/10.1016/j.jhydrol.2022.128949

Pu, Z., & Kalnay, E. (2018). Numerical Weather Prediction Basics: Models, Numerical Methods, and Data Assimilation. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble*

Forecasting (pp. 1-31). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_11-1

Pörtner, H.-O., Roberts, D. C., Tignor, M. M. B., Poloczanska, E., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., & Rama, B. (2022). *IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (H.-O. Pörtner, D. C. Roberts, M. M. B. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama, Eds.). Cambridge University Press. https://doi.org/10.1017/9781009325844

RIONED. (2006). *Stedelijke Wateropgave: Vergelijking normen voor water op straat en inundatie*.

Robinson, M., Scholz, M., Bastien, N., & Carfrae, J. (2010). Classification of different sustainable flood retention basin types. *Journal of Environmental Sciences*, *22*(6), 898-903. https://doi.org/10.1016/S1001-0742(09)60195-9

Rosenzweig, B. R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., Iwaniec, D., & Davidson, C. I. (2018). Pluvial flood risk and opportunities for resilience. *WIREs Water*, *5*(6), e1302. https://doi.org/10.1002/wat2.1302

Sadler, J. M., Goodall, J. L., Behl, M., Morsy, M. M., Culver, T. B., & Bowes, B. D. (2019). Leveraging open source software and parallel computing for model predictive control of urban drainage systems using EPA-SWMM5. *Environmental Modelling & Software*, *120*, 104484. https://doi.org/10.1016/j.envsoft.2019.07.009

Schmitt, T. G., & Scheid, C. (2020). Evaluation and communication of pluvial flood risks in urban areas. *WIREs Water*, *7*(1), e1401. https://doi.org/10.1002/wat2.1401

Schütze, M., Campisano, A., Colas, H., Schilling, W., & Vanrolleghem, P. A. (2004). Real time control of urban wastewater systems—where do we stand today? *Journal of Hydrology*, *299*(3), 335-348. https://doi.org/10.1016/j.jhydrol.2004.08.010

Shahrban, M., Walker, J. P., Wang, Q. J., Seed, A., & Steinle, P. (2016). An evaluation of numerical weather prediction based rainfall forecasts. *Hydrological Sciences Journal*, *61*(15), 2704-2717. https://doi.org/10.1080/02626667.2016.1170131

Shishegar, S., Duchesne, S., & Pelletier, G. (2019). An integrated optimization and rule-based approach for predictive real time control of urban stormwater management systems. *Journal of Hydrology*, *577*, 124000. https://doi.org/10.1016/j.jhydrol.2019.124000

Shrestha, D. L., Robertson, D. E., Wang, Q. J., Pagano, T. C., & Hapuarachchi, P. (2012). Evaluation of numerical weather prediction model precipitation forecasts for use in short-term streamflow forecasting. *Hydrology & Earth System Sciences Discussions*, *9*(11), 12563–12611. https://doi.org/10.5194/hessd-9-12563-2012

Simmons, A. J., & Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *128*(580), 647-677. https://doi.org/10.1256/003590002321042135

Sterk, G., ten Veldhuis, J. A. E., Clemens, F. H. L. R., & Berends, B. R. (2008, 31 August - 5 September). *Microbial risk assessment for urban pluvial flooding* [Conference contribution]. 11th International Conference on Urban Drainage, Edinburgh, UK. http://resolver.tudelft.nl/uuid:92b635b7-4ca0-4b47-8926-4ae170a9e3ad

Stinson, M. K. (2005). Benefits of Sewerage System Real-Time Control. In *Impacts of Global Climate Change* (pp. 1-11). American Society of Civil Engineers. https://doi.org/10.1061/40792(173)160

Stovin, V., Vesuviano, G., & Kasmin, H. (2012). The hydrological performance of a green roof test bed under UK climatic conditions. *Journal of Hydrology*, *414-415*, 148-161. https://doi.org/10.1016/j.jhydrol.2011.10.022

Stroble, C., & Taylor, S. M. (2020). History of Stormwater, Watershed, and Receiving Water Quality Modeling. In Water Environment Federation (Ed.), *Stormwater, Watershed, and Receiving Water Quality Modeling* (pp. 11-24). Water Environment Federation (WEF). https://app.knovel.com/kn/resources/kpSWRWQM01/toc

Sun, L., Xia, J., She, D., Ding, W., Jiang, J., Liu, B., & Zhao, F. (2024). A predictive fuzzy logic and rule-based control approach for practical real-time operation of urban stormwater storage system. *Water Research*, *266*, 122437. https://doi.org/10.1016/j.watres.2024.122437

Sun, L., Xia, J., She, D., Guo, Q., Su, Y., & Wang, W. (2023). Integrated intra-storm predictive analysis and real-time control for urban stormwater storage to reduce flooding risk in cities. *Sustainable Cities and Society*, *92*, 104506. https://doi.org/10.1016/j.scs.2023.104506

Tao, G. (2004). *Adaptive control of systems with actuator failures*. Springer. https://doi.org/10.1007/978-1-4471-3758-0

ten Veldhuis, J. A. E. (2011). How the choice of flood damage metrics influences urban flood risk assessment. *Journal of Flood Risk Management*, *4*(4), 281-287. https://doi.org/10.1111/j.1753-318X.2011.01112.x

Tian, W., Liao, Z., Zhi, G., Zhang, Z., & Wang, X. (2022). Combined Sewer Overflow and Flooding Mitigation Through a Reliable Real-Time Control Based on Multi-Reinforcement Learning and Model Predictive Control. *Water Resources Research*, *58*(7), e2021WR030703. https://doi.org/10.1029/2021WR030703 van der Werf, J. A. (2023). *The effect of uncertainties on the performance of real-time control of urban drainage systems* [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:1ebb628d-ecbe-49d9-b132-3a8440119f69

van der Werf, J. A., Kapelan, Z., & Langeveld, J. (2022). Towards the long term implementation of real time control of combined sewer systems: a review of performance and influencing factors. *Water Science and Technology*, *85*(4), 1295-1320. https://doi.org/10.2166/wst.2022.038

van der Werf, J. A., Kapelan, Z., & Langeveld, J. G. (2023). Predictive heuristic control: Inferring risks from heterogeneous nowcast accuracy. *Water Science and Technology*, *87*(4), 1009-1028. https://doi.org/10.2166/wst.2023.027

van Daal, P., Gruber, G., Langeveld, J., Muschalla, D., & Clemens, F. (2017). Performance evaluation of real time control in urban wastewater systems in practice: Review and perspective. *Environmental Modelling & Software*, *95*, 90-101. https://doi.org/10.1016/j.envsoft.2017.06.015

van Riel, W. (2011). *Exploratory study of pluvial flood impact in Dutch urban areas*.

Van Steenbergen, N., & Willems, P. (2014). Rainfall Uncertainty in Flood Forecasting: Belgian Case Study of Rivierbeek. *Journal of Hydrologic Engineering*, *6*(3–4). http://doi.org/10.1061/(ASCE)HE.1943-5584.0001004

Velasco, M., Cabello, À., & Russo, B. (2016). Flood damage assessment in urban areas. Application to the Raval district of Barcelona using synthetic depth damage curves. *Urban Water Journal*, *13*(4), 426-440. https://doi.org/10.1080/1573062X.2014.994005

Vergouwe, R. (2016). *The National Flood Risk Analysis for the Netherlands*. R. V. P. Office.

Vis, M., Klijn, F., De Bruijn, K. M., & Van Buuren, M. (2003). Resilience strategies for flood risk management in the Netherlands. *International Journal of River Basin Management*, *1*(1), 33-40. https://doi.org/10.1080/15715124.2003.9635190

Walczykiewicz, T., & Skonieczna, M. (2020). Rainfall Flooding in Urban Areas in the Context of Geomorphological Aspects. *Geosciences*, *10*(11). https://doi.org/10.3390/geosciences10110457

Wasko, C., & Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology*, *575*, 432-441. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.05.054

Xu, M., van Overloop, P. J., van de Giesen, N. C., & Stelling, G. S. (2010). Real-time control of combined surface water quantity and quality: polder flushing. *Water Science and Technology*, *61*(4), 869-878. https://doi.org/10.2166/wst.2010.847

Xu, W. D., Burns, M. J., Cherqui, F., Duchesne, S., Pelletier, G., & Fletcher, T. D. (2022). Real-time controlled rainwater harvesting systems can improve the performance of stormwater networks. *Journal of Hydrology*, *614*, 128503. https://doi.org/10.1016/j.jhydrol.2022.128503

Xu, W. D., Fletcher, T. D., Burns, M. J., & Cherqui, F. (2020). Real Time Control of Rainwater Harvesting Systems: The Benefits of Increasing Rainfall Forecast Window. *Water Resources Research*, *56*(9), e2020WR027856. https://doi.org/10.1029/2020WR027856

Yang, F., Cerrai, D., & Anagnostou, E. N. (2021). The Effect of Lead-Time Weather Forecast Uncertainty on Outage Prediction Modeling. *Forecasting*, *3*(3), 501-516. https://doi.org/10.3390/forecast3030031

Yang, T. H., Hwang, G. D., Tsai, C. C., & Ho, J. Y. (2016). Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts. *Hydrol. Earth Syst. Sci.*, *20*(12), 4731-4745. https://doi.org/10.5194/hess-20-4731-2016

Zevenbergen, C., Cashman, A., Evelpidou, N., Pasche, E., Garvin, S., & Ashley, R. (2011). *Urban flood management*. CRC.

TUDelft