DTU Human-Centered Artificial Intelligence
Department of Applied Mathematics and Computer Science

The outlook for smart drainage systems

Assessment of data and methods for data-driven urban drainage forecasts in Reykjavik, Iceland

Master Thesis

The outlook for smart drainage systems
Assessment of data and methods for data-driven urban drainage forecasts in Reykjavik, Iceland

Study programme
MSc Eng in Human-Centered Artificial Intelligence

Master Thesis
December, 2021

Written by
Sverrir Heiðar Davíðsson(s200147)

Supervisors
Luca Vezzaro, Associate Professor at DTU Environment
Jonas Wied Pedersen, PhD, Postdoc at DTU Environment
Michael Riis Andersen, Associate Professor at DTU Compute
Hlöðver Stefán Þorgeirsson, Waste Water R&D Lead at Veitur

Cover photo: Vibeke Hempler, 2012
Published by: DTU
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, Building 324
2800 Kgs. Lyngby
Denmark
https://www.compute.dtu.dk/

Abstract

Rainfall-runoff modelling in urban hydrology is a widely studied problem that has mainly been approached with physical and conceptual models. Data-driven models are becoming more widely used in many fields due to increased data and computing power. Many applications of rainfall-runoff modelling require high-quality input data with high temporal and spatial resolution. The Icelandic wastewater utility company Veitur is in the first stages of modernizing its wastewater system and, until now, have primarily used data for monitoring purposes. Three neural network layers were tested in four combinations with three different rainfall data sources for drainage flow simulation and forecasting. These layers were a temporal convolutional layer, a long short-term memory layer, and a fully-connected linear layer. Several rainfall estimates were made using radar data. The accuracy of the radar estimated rainfall and NWP are evaluated using rain-gauges. The difference between pumping station inflow and outflow is analyzed using high-resolution flow and water level data. The thesis results found that the estimates of rainfall produced at a lower altitude can be better but also more sensitive to interference from the ground. Models using rain gauge data and NWP data outperformed those using radar-estimated rainfall as input to drainage flow models. The temporal convolutional layer outperforms all other models, and performance is almost always improved with longer sequences of past observations. It was found that the main application possible with these models, current data and limited control mechanisms available to Veitur is early warning systems based on prediction thresholds. With investments to better control mechanisms, improvements to the drainage flow data and better radar estimated rainfall, further applications of drainage flow forecasts would likely become possible.

Preface

This thesis has been prepared a the Department of Applied Mathematics and Computer Science at the Technical University of Denmark, DTU, in partial fulfilment for the degree Master of Science in Engineering, MSc Eng and accounts for 35 ECTS.

It is assumed that the reader has a basic knowledge in the areas of statistics and machine learning.

The code used in this project can be found at https://github.com/SverrirH/ThesisCode

Acknowledgements

I would like to express my gratitude to my four supervisors who each helped me in different ways: Jonas for introducing me to the fields of urban hydrology, weather radar and numerical weather predictions and always pointing me in the right direction; Luca for helping me stay on track and reminding me of things like the sad but necessary lesson of killing one's darlings, without which I would probably still be trying to perfect my radar nowcast; Michael for helping me stay grounded with my ambitious and usually unrealistic modelling ideas; Hlöðver for helping me get to know and appreciate wastewater systems and jumping at every opportunity to help.

I also want to thank the various people at Veitur for their great support throughout this thesis and the other projects I have done with them. I would also like to thank the people at the Icelandic Meteorological Office for giving their valuable time, letting me use their data and helping me get to know the field of meteorology.

Last but not least, my girlfriend, her parents and my mom who supported me in many ways throughout the pandemic that gave me the space I needed to focus on my studies.

Contents

1 Introduction

Urban hydrology is a subfield of hydrology, which is the study of water's distribution, movement, and interactions with the natural and man-made environment. Urban hydrology looks specifically at these processes in the context of urbanised territory. The systems responsible for the safe handling of rainwater are called drainage systems, and the systems responsible for other wastewater, such as sewage, are known as wastewater systems. Both systems are an integral part of a city's infrastructure.

These systems are divided into two categories: combined or separated sewer systems, depending on whether they handle runoff and sewage together or separately. In combined systems, a failure to handle rainfall-runoff can result in what is known as combined sewer overflow (CSO) discharges. CSO discharges happen when the combined sewage and runoff overwhelms the system, resulting in a release of untreated sewage to the environment. CSO discharges are undesirable but are used as a last resort to prevent flooding, property loss, increased risk of disease, and potential risk to other infrastructure such as domestic water sources.

In hydrology, a drainage basin, also called a catchment area, refers to land from which surface runoff, i.e. the water that runs off the surface after rainfall, drains into a common stream or outlet. Rainfall-runoff models try to predict the flow of water within a catchment or from the common outlet of the catchment using rainfall observations. An urban drainage flow forecast is a specialised case of a rainfall-runoff forecast, and there are many operational strategies for urban drainage systems that require or would benefit significantly from having a drainage flow forecast.

There are many approaches to rainfall-runoff modelling, but data-driven models have gained popularity in many fields due to the increasing availability of data and greater computational power. The conventional approach of rainfall-runoff modelling uses simplified equations based on domain knowledge and experience in hydrology, but the data-driven approach lacks a connection to the physical system. Instead, these models use general equations to simulate the system's processes. One versatile class of data-driven models are called artificial neural networks (ANN) and many variations of these, like long short-term memory (LSTM) and convolutional neural networks (CNN), are especially well suited for modelling Spatio-temporal processes like rainfall-runoff. Like all data-driven models, these models depend on the abundance and quality of the data to make accurate predictions.

Weather radar and rain gauges are the two most commonly used rainfall data sources in urban hydrology. Both methods have some benefits and drawbacks, but the two complement each other well. Rain gauges are usually considered the most reliable measurement of rainfall and are often used as 'ground truth' data in hydrological research, despite their shortcomings. Weather radar transmits electromagnetic waves and uses an antenna to measure the waves reflected off clouds to estimate their densities. Radar data contains much more spatial information than rain gauge data but has many problems of its own. Radar data suffers from uncertainty in rainfall estimates and has a problem with erroneous data like clutter, which is the unwanted reflectivity from static objects or geographical features, such as mountains, that obstruct or reflect the radar beams. Rainfall can be estimated from radar reflectivity, but this estimate is subject to uncertainty. Neither rain gauge data nor radar data are enough to predict rainfall in the future, which is required to forecast drainage flow.

The two most common forms of precipitation forecasts are numerical weather predictions (NWP) and radar nowcasts, also known as radar extrapolation. NWPs are made with mathematical models of the atmosphere and oceans, which are used to forecast pre-

cipitation and many more aspects of weather. Radar nowcasting is a particular type of short-term forecast where a sequence of recent radar observations are used with data-driven models or algorithms to extrapolate the future positions and intensities of clouds which can then be used to predict rainfall. In order to compare the two methods, a different data source is required, one that provides rainfall observations instead of forecasts.

In many countries, the urban drainage and wastewater systems still in use today were designed many decades ago when utility companies did not have access to the vast amount of data, computing power and computational methods that have since come into widespread use. The utility company Veitur is responsible for the operation and development of the drainage and wastewater system in Iceland's capital Reykjavík. Veitur is in the process of modernising many of its utilities, including the wastewater utility. Today, the company does not use any form of simulation or forecasting models in its system but has been taking steps in recent years to bring its wastewater system up to date with current technologies.

1.1 Research objectives

This thesis is about the design, implementation and evaluation of a simple data-driven drainage flow model for use in the Reykjavík wastewater and drainage system, as well as an assessment of the data available for this task. The model will be designed so that it can be used both for simulation and forecasting. A special emphasis is placed on evaluating radar data and numerical weather predictions as they are the main candidates for rainfall input into the model. The goal of the evaluations in this paper is to show if certain applications are possible or give practical recommendations for how to improve the systems. The thesis will address the following questions:

1. How accurately can urban flow be simulated and forecasted in Reykjavík, Iceland, with currently available data and simple machine learning methods.

2. How can simple data-driven simulations and forecasts be used to improve operations of the wastewater system in Reykjavík with currently available data?

3. What are the main improvements to the data or methods that would be required for better performance and additional applications of the simulations and forecasts?

1.2 Thesis outline

Chapter 2 looks at the existing literature on the topics of this thesis. The wastewater system and the available data are introduced in chapter 3. Chapter 4 describes the approaches used to process and evaluate the data as well as the design, implementation and evaluation of the models used for drainage simulations and forecasting. The results of the experiments and evaluations described in chapter 4 are presented and discussed in chapter 5. Then the conclusions can be found in chapter 6.

2 Literature review

The purpose of this literature review is to give an overview of the current literature relating to the problem of data-driven simulation and forecasting of urban drainage flow.

The literature review will begin by looking into rainfall-runoff models used in urban hydrology; this will include the use of rain gauges, radar and radar nowcasts for simulating and forecasting rainfall-runoff and numerical weather prediction for the same objective.

2.1 Rainfall-runoff modelling

A drainage basin is the area from which surface runoff drains into a common stream (Britannica, 2020). Rainfall-runoff models try to predict the flow of water within a catchment using estimated rainfall on that catchment. An urban drainage flow forecast is a specialized type of rainfall-runoff forecast, and there are many operational strategies for urban drainage systems that require or would benefit significantly from having a drainage flow forecast (Thorndahl et al., 2017). Rainfall-runoff models have a long history, dating back at least 170 years when an Irish engineer named Thomas James Mulvaney published a paper with an attempt to model the peak runoff flow in response to rainfall with a simple equation(Dingman, 2015). Since then, thanks to advances in hydrology, more data and greater computational power, much progress has been made in rainfall-runoff modelling.

Rainfall-runoff models can be grouped roughly into three categories: physical, conceptual, and empirical. Physical models use physical laws and equations based on known hydrological responses to estimate the runoff. These can incorporate spatial and temporal variability to a very fine scale and require the calibration of many parameters specific to each catchment. Conceptual models use simplified equations that represent water storage in a catchment and are much easier to calibrate, but these lack the consideration for spatial variability within the catchment. Empirical models take a different approach to the problem, avoiding the need for specialized equations with inherent limitations and instead use more general models that try to describe the relationship between the input, i.e. rainfall data and other relevant measurements, and the output, the stream-flow (Sitterson et al., 2018).

The main requirement for practical drainage flow forecasting is a precipitation forecast, and the two main types of precipitation forecast used in urban hydrology are NWP and radar nowcasts. The former has a considerably lower spatial and temporal resolution but a larger forecast horizon than the latter. Therefore, radar nowcasts are often considered the better choice for applications such as real-time control, which require at most very short-term but precise forecasts. Although, as (Pedersen, 2021) notes, the better option of the two is going to depend on the application in question and its temporal and spatial scale.

2.1.1 Applications of rainfall-runoff forecasts

(Thorndahl et al., 2017) splits applications for radar estimated rainfall in urban hydrology into two categories: online, meaning they must be computed in real-time and updated continuously as new data becomes available, and offline, meaning they may be done at any time using historical data. An example of an offline application is gathering general statistical characteristics of the system, such as evaluating the distributions of rainfall events in terms of extremes to estimate how likely a rainfall event of a given size is to occur. Another type of offline application is the analysis of events after they occur for things like insurance claims, analysis of the system during certain events or modelling for risk assessments. Examples of online applications are warning systems for severe rainfall

that is likely to cause flooding and modelling of scenarios from rainfall forecasts that can be used to evaluate different control strategies for reservoirs or pumping stations.

(AghaKouchak & Habib, 2010) maps out the resolution and temporal requirements for many applications of radar estimated rainfall. In figure 2.1 this map is shown, but other authors suggest some of these applications may be possible with a lower temporal resolution like that of NWP. The figure places control of reservoirs at just below the 15 minutes of temporal resolution, but this of course, will depend greatly on the size of the reservoir.

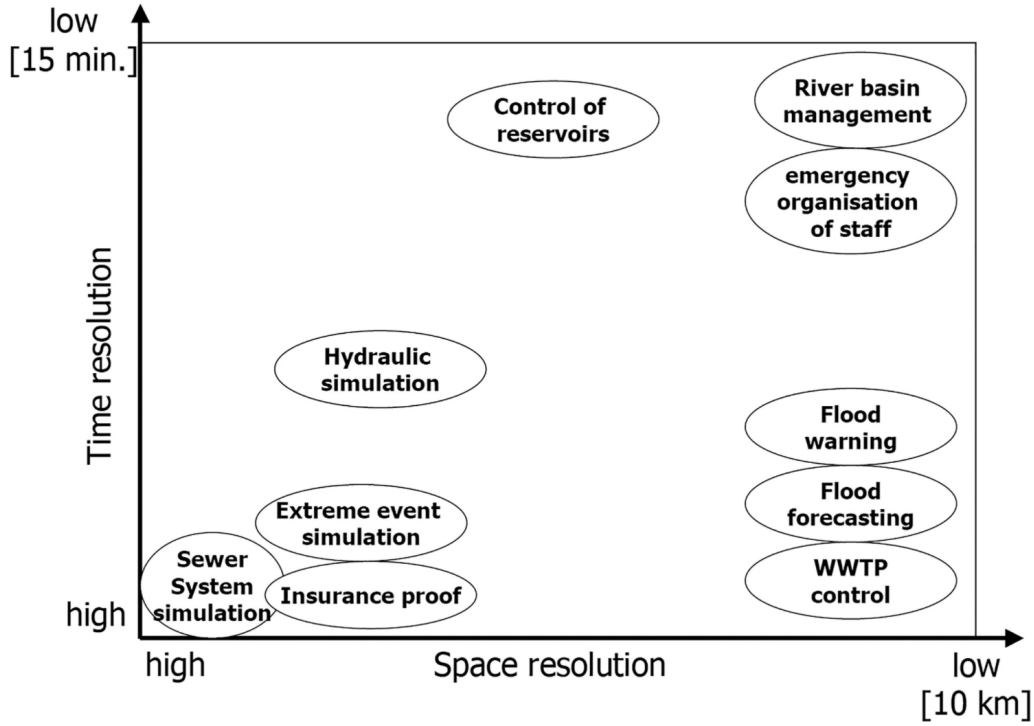

Figure 2.1: Resolution requirements radar for different urban hydrological applications using radar estimated rainfall (AghaKouchak & Habib, 2010)

(Pedersen, 2021) notes that publications on the topic of applications of NWP in urban drainage and wastewater purposes is limited and may be broadly categorized into four topics: Generic rain and flow forecasting, urban pluvial flood forecasting, real-time control and post-processing. The approach used by (Courdent et al., 2018) to distinguish high and low flow domains using NWPs demonstrates a few possible applications of NWPs in urban drainage despite their low temporal and spatial resolution.

2.2 Radar data in urban drainage

Many authors emphasize the value of radar data in urban hydrology, but they have also highlighted the importance of processing the radar data and the need to adjust the radar data with more reliable rainfall estimates.

Radar reflectivity (Z) can be converted to rain rates (R), but this conversion is subject to uncertainty in the relationship between reflectivity and rainfall intensity, known as the Z-R relationship because the distribution of raindrop sizes affects the relationship and is usually not known (Beven, 2011). The Z-R relationship is an exponential equation $Z = aR^b$, where a and b are adjustable parameters that can vary greatly over time and between events. The particular combination of $a = 200$ and $b = 1.6$ is called the Marshall–Palmer relation.

In order to evaluate the coefficients of the Z-R relationship for reliable precipitation estimates, rain gauges or other ground observations of rainfall are required. These ground observations can be used to adjust the radar estimated rainfall (Ochoa-Rodriguez et al., 2019; Thorndahl et al., 2017). This technique is known as radar rainfall adjustment or radar-rain gauge merging. The two main categories of radar rainfall adjustment are static and real-time adjustments. Static adjustments use a single set of coefficients for the Z-R relationship across the whole data set, while real-time adjustments require constant evaluation and updating of these coefficients. notes that real-time adjustment should be made with caution due to its sensitive nature. (Löwe et al., 2014) found a single optimal adjustment to improve the quality of a model more than a real-time adjusted estimate for a model that used radar rainfall to predict runoff. More authors (Vallabhaneni et al., n.d.) suggest caution for real-time adjustment of radar due to its sensitive nature. Some authors also suggest specific combinations of coefficients of the Z-R relationship depending on the type of rainfall event, e.g. tropical storms (Einfalt et al., 2004).

(Germann & Joss, 2004) state that high-quality rain-gauge data from a relatively dense rain-gauge network, 1 per $10 - 30km^2$ is a prerequisite to radar-rain gauge merging. This statement is also supported by (Einfalt et al., 2004) and (Thorndahl et al., 2017). Some note that the required network density for successful adjustment is rarely found (Beven, 2011; Cole & Moore, 2008). In addition to the rarity of these ideal conditions for radar rain gauge merging, rain gauge data is also frequently observed having shortcomings like missing data and erroneous data caused by clogged gauges, evaporation, underestimation of rainfall due to high wind speed, etc. (Goudenhoofdt & Delobbe, 2009; Stránský et al., 2007; Thorndahl et al., 2017).

2.3 Data driven models

One versatile class of data-driven models is called artificial neural networks (ANN). One of the key attributes of neural networks is their modular nature. Conventional ANNs are comprised of units called layers or hidden layers, and each layer contains one or more neurons or nodes (Hopfield, 1982). The nodes are connected in a structured way where the output of the nodes in one layer is connected to the nodes' input in the next layer.

Due to these desirable properties, multiple authors have used neural network-based rainfall-runoff models (Abou Rjeily et al., 2017; Bruen & Yang, 2006; She & You, 2019). One notable study used LSTMs in a rainfall-runoff model and was able to show that the LSTM had learned latent representation of snow accumulation which indicated that the architecture could model rainfall-runoff while including snowmelt (Kratzert et al., 2018). Other types of neural network architectures have also been used more broadly in time series forecasting called convolutional neural networks. These networks exhibit a property called shift-invariance, where weights in the network are applied equally to all-time steps, which can be used to reduce the dimensionality of a sequence of inputs. A 1-dimensional variation of these can be applied to temporal data. These are sometimes called time-delay neural networks (Waibel et al., 1989).

3 Case study

The main purpose of this section is to inform the reader about the state of affairs in the context of the Reykjavík drainage and wastewater system, the relevant data available from the Icelandic Meteorological Office (IMO) and Veitur. This utility company manages the drainage and wastewater system in Reykjavík.

3.1 Reykjavík drainage and wastewater system

The drainage and wastewater system in Reykjavík, Iceland, is made up of two smaller systems, Sundaveita and Skerjafjarðaveita, each one with a wastewater treatment facility, Hreinsistöð Klettagörðum and Hreinsistöð Ánanaustum, respectively. Figure 3.1 shows the two systems, the locations of main wastewater treatment facilities, the various pumping stations that pump wastewater to the water treatment facilities. The two main pumping stations used for the analysis and the positions of the rain gauges are mentioned later in this paper, Boðagrandi and Gelgjutangi. Boðagrandi is on the north-western coast of Reykjavík within the Skerjafjarðaveita system. Gelgjutangi is close to the centre of Sundaveita and is the larger of the two.

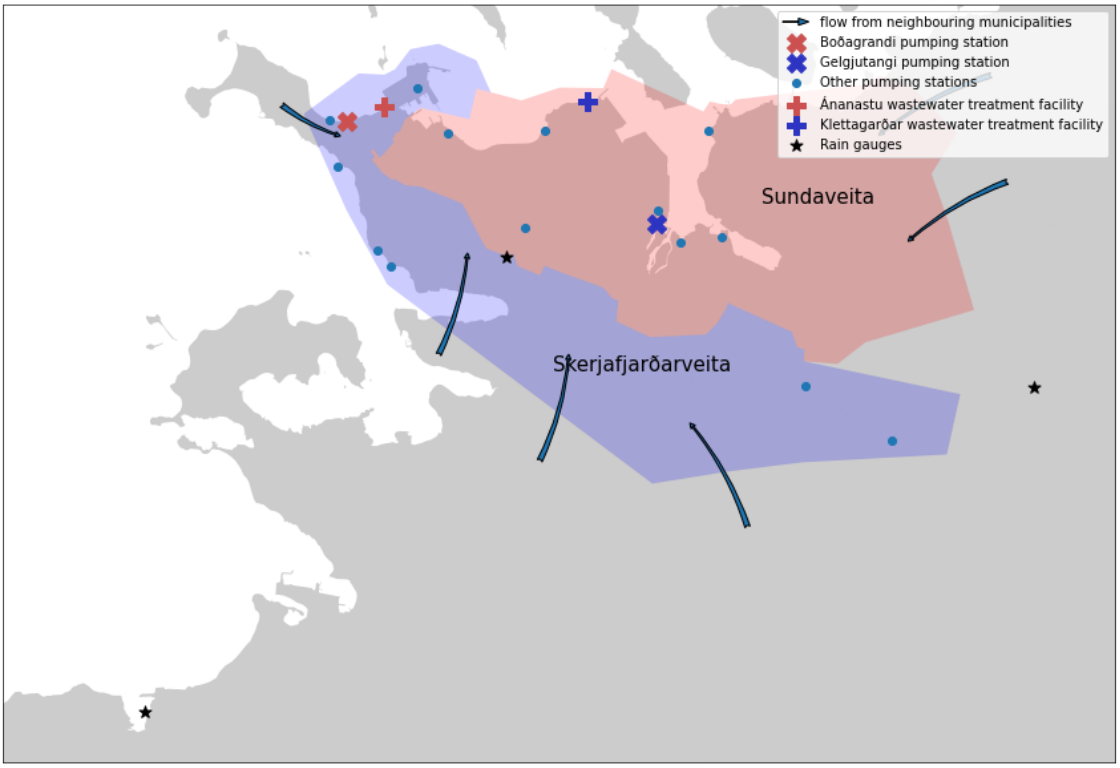

Figure 3.1: The two main systems in the Reykjavík Urban Drainage system, along with some of the main assets of interest in this project

Wastewater systems are generally split into two main categories depending on if they separate sewage from rainwater or combine them in a single pipe. The wastewater system in Reykjavík has both combined and separated sewers. Most combined sewers belong to older parts of the system in the downtown area of Reykjavík. An overview of the types is found in figure 3.2 where green lines represent combined sewer pipes while blue and red lines represent separate runoff and sewage pipes, respectively.

The main challenge of combined sewers is the increased risk of the system being unable to cope with the added flow due to rainfall, compared to separate systems. The combined sewage must pass through a wastewater treatment facility for safe disposal. The two pumping stations highlighted on figures 3.1 and 3.2 have a substantial proportion of their inputs coming from rainfall-runoff.

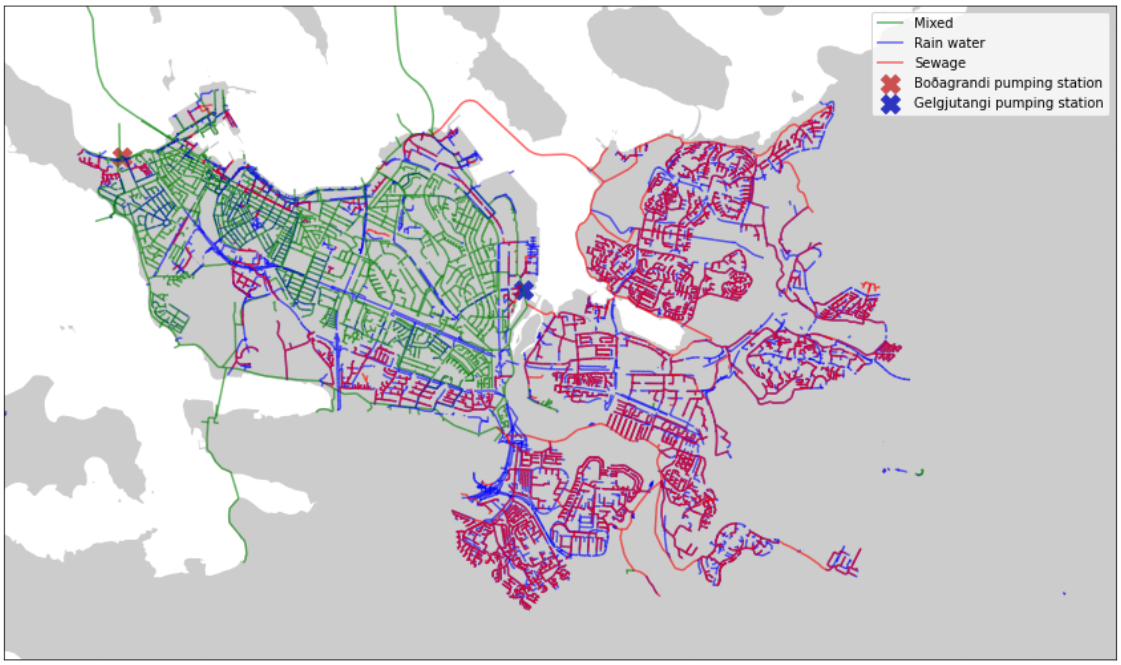

Figure 3.2: An overview of the types of wastewater pipes in the Reykjavík wastweater system. Green lines: combined sewer pipes. Red lines: Sewage pipes. Blue lines: Rainfall-runoff pipes.

3.1.1 Drainage flow data

The drainage data available for this study consists of hourly flow averages from two drainage flow sensors. These sensors belong to separate pumping stations known as Boðagrandi and Gelgjutangi. Gelgjutangi is notable because it is the oldest pumping station still in use in Reykjavík and gets its wastewater from the south-eastern catchment of the two catchments drawn in figure 3.3. This system collects a high proportion of the rainfall-runoff within its catchment and is thus considered a good candidate. The other pumping station, Boðagrandi, is located on the north-eastern tip of Reykjavík and a large portion of the catchment is inside Seltjarnarnes, a neighbouring municipality of Reykjavík.

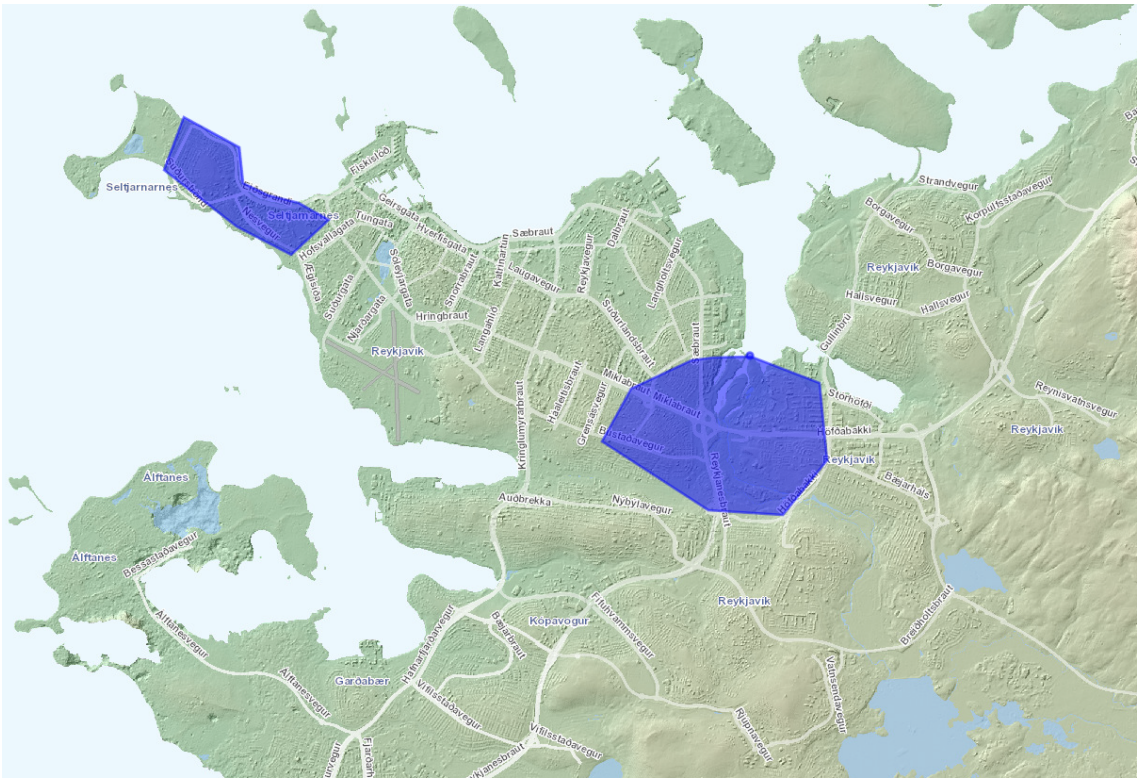

Figure 3.3: A rough outline of the two catchments of the pumping stations studied in this paper. The north-western and smaller of the two catchments is the one that leads to the Boðagrandi pumping station. The bigger catchment is the one that leads to the Gelgjutangi pumping station.

Both of these pumping stations are wet well pumping stations (see figure 4.10), meaning that inside them is a chamber known as a wet well that can hold wastewater temporarily before it gets pumped out. The water level within the chamber is measured, and hourly averages of this value are stored. Due to this chamber, there is often a slight difference in the inflow and outflow from these pumping stations. Section 4.2.3 describes this problem in more detail, as well as a description of the methods used to analyze this difference.

3.1.2 Hot water utility and district heating

Most buildings in Reykjavík are heated using a geothermal district heating system (Gunnlaugsson, 2004). Unlike many other district heating systems, Reykjavík uses a mostly open-loop system where, once the water has passed though a radiator or floor heating, it leaves the system, usually into a combined wastewater system or other drainage solution. Thus predicting the use of district heating becomes a part of the drainage flow forecasting problem. Flow observations in the district heating system are available for many regions of Reykjavík but the wastewater system and drainage systems are not divided in the same way and flow measurements cannot always be directly assigned to a single pumping station. As is further discussed in section 4.1.2, the variable used to account for the inflow of the district heating water in the wastewater system is the outside temperature which, in addition to temperature forecasts, is used by Veitur currently to forecast hot water usage with good results.

3.2 Weather data

3.2.1 Radar data

The radar data used in this project contains radar echo data over a 6 year period starting at the beginning of 2015 to the end of 2020 collected and supplied by the Icelandic Meteorological Office. The radar in question is the ISKEF radar in the Southern Peninsula Region. The radar data has many possible applications in urban hydrology. The main applications come from the ability to estimate rainfall using radar data. More detailed description of the radar data is given in section 4.2.1, where the processing steps to turn the raw data into rainfall estimates is also described.

As is described in the literature review, there is significant uncertainty in radar estimated rainfall and in order to get more accurate estimates, ground observations are required to adjust the radar data. Using ground observations of rainfall to adjust rdara data is called radar-rain gauge merging, and without it, many applications that require relatively high resolution and high accuracy rainfall estimates will not be possible. The application of short-term precipitation forecasts using radar data, called radar nowcasts, was also described in the literature review. These short-term forecasts could allow for high resolution and high accuracy drainage flow forecasts, which are the basis for several applications such as real-time control (RTC) for pumping stations and reservoirs. Currently, the Icelandic Meteorological Office does not perform any radar-rain gauge merging or radar nowcasting, and so all applications that require these cannot be considered by Veitur for the time being.

3.2.2 Rain gauge data

Within reykjavík multiple rain gauges have recently been added, but only one has enough historical data to be used in the analyses later in this paper. Some rain gauges outside Reykjavík could be used for applications within Reykjavík. Three main rain gauges are considered in this report, see table 3.1, these gauges are also shown in multiple figures in this paper, for example see figures 3.1, 3.4 and 4.7. In figure 4.8 these gauges are also show with their respective IDs. These rain gauges have enough historical data to compare them with the NWP and radar data. The rain gauges are treated as 'ground truth' for rainfall observations when evaluating the accuracy of the NWP forecasts and the radar estimated rainfall, but these are of course, not estimates of the same thing. The radar is a volume-based estimate of rainfall, and the NWP are estimates of rainfall over 2x2km area while the rain gauges are effectively point-estimates of rainfall.

As mentioned in the literature review, both the quality and density of rain gauge networks are prerequisites for merging radar and rain gauge data. The three gauges in table 3.1 are each of a slightly different type. According to specialists at the IMO, gauge 1475 is considered the most reliable of the three and the most suited to Icelandic weather conditions and gauge 1473 is considered to be the least reliable since it is both unheated, meaning that measurements of snow will often be erroneous, and is known to underestimate rainfall in certain wind conditions. The data from these gauges is stored in a time series with a 10-minute resolution. The gauges with 'yes' in the 'manually reviewed' column in table 3.1 were also available as hourly values that had been reviewed and processed by staff at IMO.

In the literature review, the density of one gauge per $10 - 30km^2$ is considered good enough for radar rain gauge adjustment. The shortest distance between these gauges is about 10km, the distance between gauge 1475 and 1473. If we assume each gauge covers an area in a circle with a radius equal to half the distance to the closest gauge, then each gauge has to cover an area of roughly $80km^2$, and so the density of these gauges falls wide of the mark. Although, as previously mentioned, there have recently been some additions

to the gauges within and around Reykjavík, so future applications that require radar-rain gauge merging might be possible.

Table 3.1: The three main rain gauges used in this study. The only rain gauge used for rainfall runoff modelling is gauge 1475 since it is located nearest to both catchments

ID	Location	GPS	Type	manually reviewed
1475	Reykjavík	64.1274, -21.9024	Heated weighing bucket	Yes
1473	Straumsvík	64.0438 , -22.0404	Unheated tipping bucket	Yes
1481	Hólmsheiði	64.1085, -21.6864	Heated tipping bucket	No

3.2.3 Numerical weather prediction data

The numerical weather prediction (NWP) data used in these experiments was also provided by the Icelandic Meteorological office and comes from the HARMONIE - numerical weather prediction model, which has been in use since 2011 (Íslands, n.d.). The data comes in the form of cumulative rainfall from the initial time step and has a forecast horizon of 66 hours, temporal resolution of 1 hour and spatial resolution of 2x2km and is updated every 6 hours. In figure 3.4 the positions and sizes of the NWP pixels, which are $2x2km$ in area can be seen over and around the Reykjavík area. The forecast values at these pixels are used as input for the drainage forecast models in chapter 4. For a comparison of scale with radar estimated rainfall pixels, see figure 4.7.

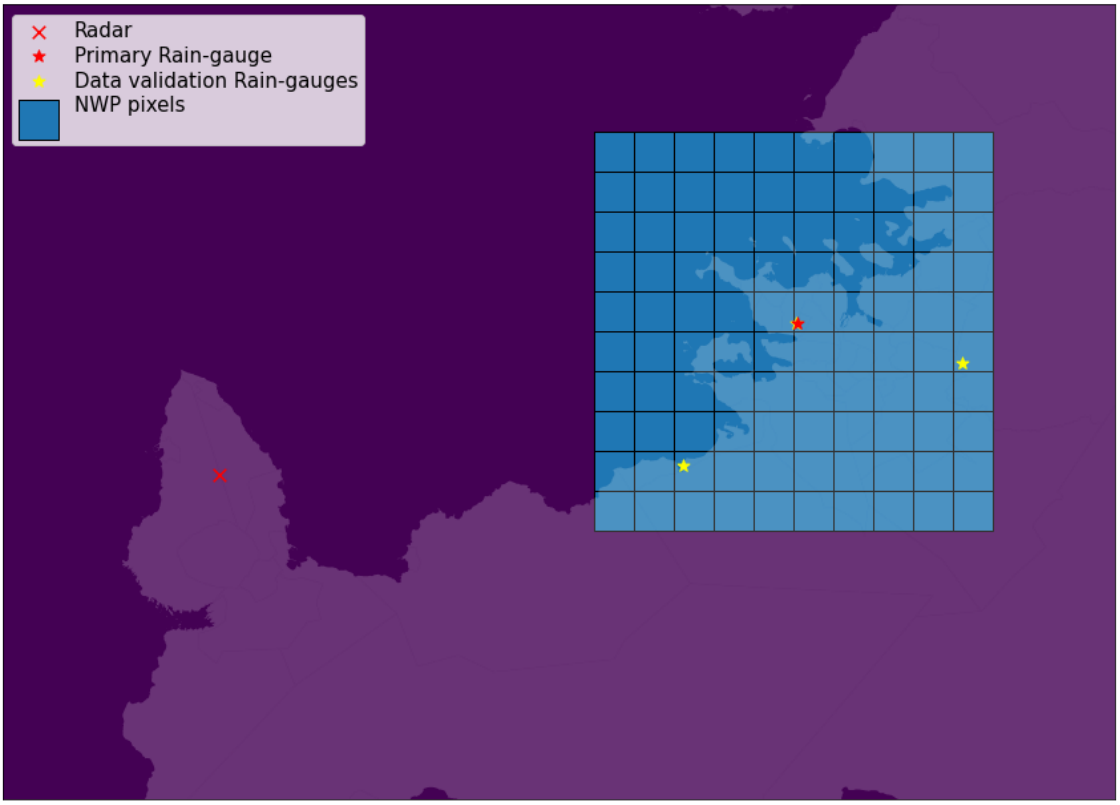

Figure 3.4: 10x10 grid of NWP pixels shown to scale in relation to capital area and the two main rainfall sensors: the radar by Keflavík and the primary rain-gauge by the Meteorological Office in Reykjavík and the two other rain gauge used to evaluate the radar and NWP data.

3.2.4 Temperature forecast data

To perform forecast using a model that depends on the temperature, it is necessary to use temperature forecasts instead of temperature observations to get a realistic view of the performance that would be seen when forecasting. The temperature forecast data used in this project originates from the same NWP model as the precipitation forecast data but is specifically estimated for at the location of the 1475 weather station where the temperature observations are also made.

3.3 Applications for the Reykjavík wastewater system

Applications of a drainage flow forecast are limited by the observations' quality, even for those categorized as 'offline' applications. Some offline applications may be ruled out until a better radar estimated rainfall becomes available like those relating to insurance claims and other analyses relating to extreme rainfall events, since they may be limited by the resolution of the drainage flow data as well as the quality of rainfall estimates. Some applications may only require a long history of data rather than high quality data, such as those relating to useful statistics about the system that may be used as a basis for design decisions in the future. An example of such a statistic is the frequency of overflow events, which may be used to justify investment costs into improved systems and provide a baseline of performance for future changes to the system to evaluate their worth.

The more challenging category of applications given the current constraints of the system and the available data are 'online' applications. The wastewater system in Reykjavík has very few control mechanisms that can be used to take preemptive measures if the risk of heavy rainfall is high. Currently, most pumping stations are equipped with pumps that can only be set to 'on' and 'off' states, and these states are controlled by the water level inside the wet well of the pumping stations. Applications of drainage flow forecasts are mainly relevant for combined sewer systems, and thus, since a large part of Reykjavík has separate sewers, only some of the pumping stations in Reykjavík could make use of such a forecast.

The wet wells inside the pumping stations serve as a relatively small buffer. Here the buffer is considered small because the ratio of 'well volume' over 'total pumping capacity is relatively short. For example, the useable capacity of the Boðagrandi wet well is less than 90′000 litres, and the well has three sewage pumps, each capable of 180 litres per second and 3 CSO pumps each with about 370 litres per second capacity. Just the sewage pumps can empty the wet well in less than 3 minutes, and with the rainwater pumps, it can be emptied in about 55 seconds. Since the duration of rainfall events can be much longer than this time, the wet wells have relatively small buffers for applications relating to rainfall forecasts. To reduce combined sewer overflow with drainage forecasts, the capacity of storage assets would need to be considerably greater than the current wet wells.

Since the only currently available rainfall forecast is the NWP, online applications are limited to those that do not require high spatial or temporal resolution. In addition to the limited control mechanisms, this limitation rules out the possibility of most or all real-time-control applications for small reservoirs and pumping stations. One way to utilize NWP precipitation forecasts is with a similar approach as the one used in (Courdent et al., 2018), where NWP forecasts were not explicitly used to predict the drainage flow but rather distinguish between so-called flow domains. These flow domains are ranges of flow that may have different operational implications that can be used to prepare or send out a warning of increased risk of flooding or CSO discharges. The distinction between flow domains is based on flow thresholds that are used to generate binary outputs, e.g. "greater than 500 l/s" or "less than or equal to 500 l/s". A similar flow threshold-based approach will be used in section 4.3 to evaluate the performance of the different drainage

flow simulation and forecast models.

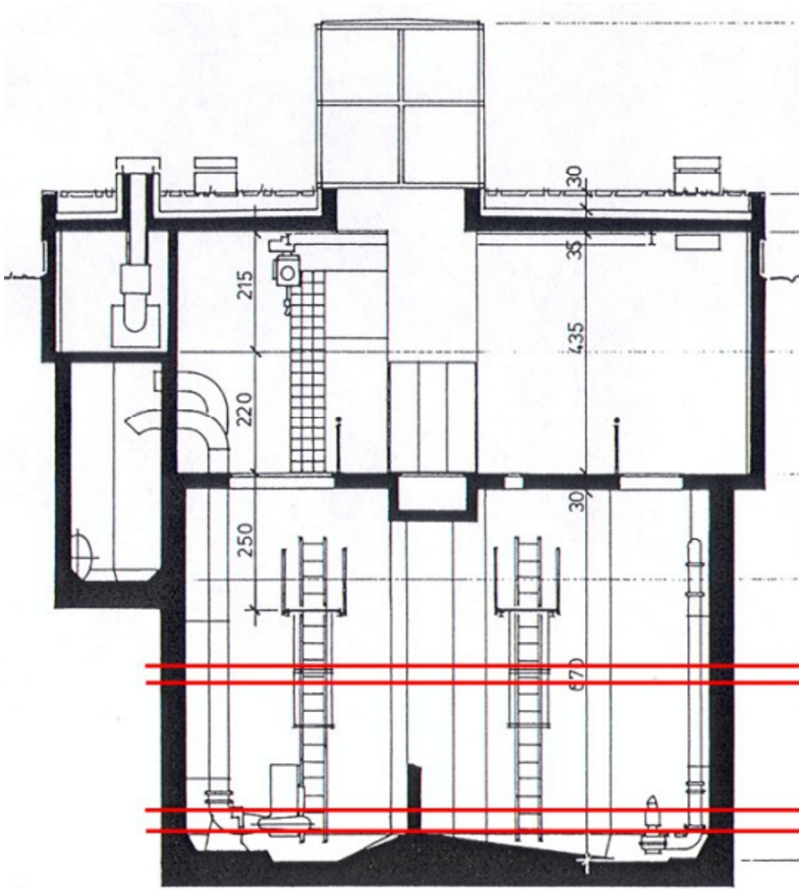

Figure 3.5: A drawing of the Boðagrandi pumping station the water levels relevant to the pump control system. The red lines represent the height that triggers the system to turn the pumps on or off. The highest line is marked at 3.35m and the rest are 3.0m, 0.8m, and 0.5m.

4 Methods

This chapter describes the methods and details the approaches applied in this thesis to assess the data and appropriate methods for drainage forecasting in Reykjavík. The chapter starts with section 4.1 which gives context to the experiments and data in the following sections. Next, section 4.2 describes how the raw radar data was processed and turned into rainfall estimates, as well as how the quality of the radar rainfall and NWP forecasts are evaluated with rain gauges. It also describes an approach to evaluate potential errors and uncertainty in the drainage flow data. Lastly, section 4.3 starts by describing the drainage flow simulation and forecasting models, how the model training was performed and how the best models were selected and evaluated.

4.1 System

This section will describe a simplified version of the wastewater system in Reykjavík as well as give context to the most relevant data available for analysis.

All parts of the wastewater system in Reykjavík get input from one or more of the following sources, also seen in figure 4.1: water used for district heating, sewage, cold water (also called drinking water/potable water), wastewater, and rainfall-runoff. Since the inputs to the system are mostly known, it is possible to examine each source separately and what data is necessary to forecast them. This approach is used to select the input data for the models to simulate and forecast the drainage flow.

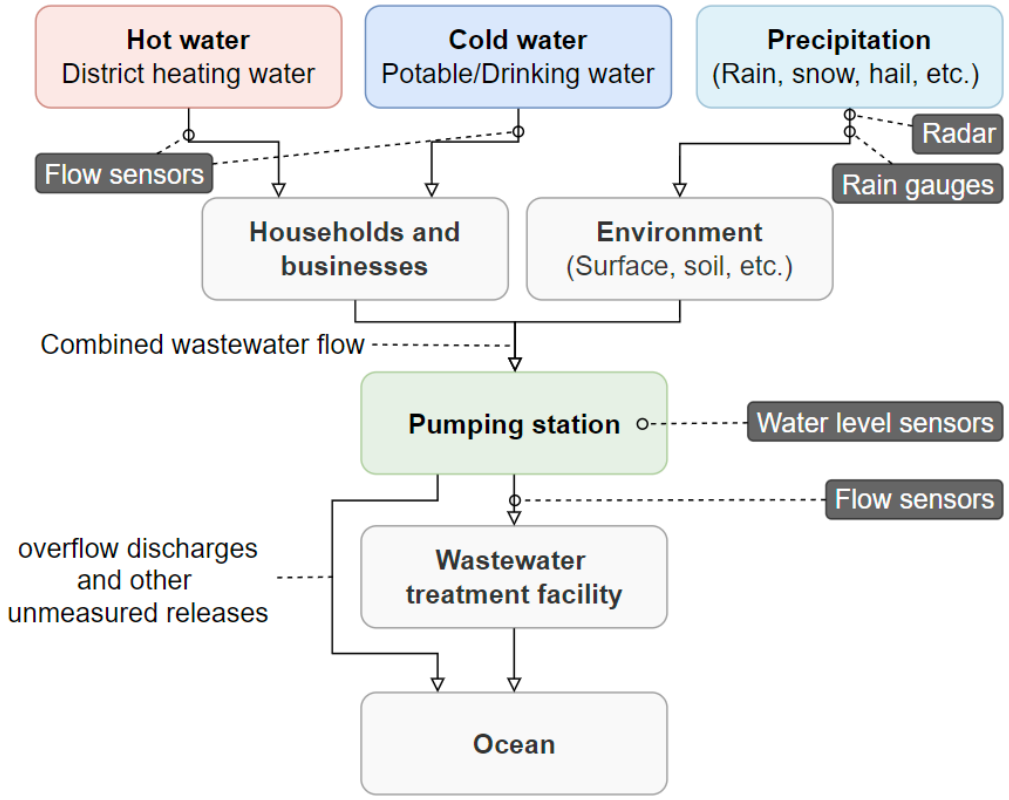

Figure 4.1: The primary inputs into drainage system: hot water, cold water and precipitation, as well as some of the most relevant data sources available within the system. The flow sensors used in this paper measure the outflow from the pumping stations rather than the inflow.

4.1.1 Cold water

Cold water, also known as drinking water and potable water, is used for drinking, toilets, showers and many other household applications. The main data source for the drinking water utility is flow meters in different neighbourhoods. This flow data is not used in the experiments in section 4.3 both because the data has too many problems for convenient modelling purposes and because the modelling approach requires that only data with an accompanying forecast be used so that the simulations and forecast get equivalent data. Some examples of these data problems include non-stationary trends, not representative of the water entering the sewage system caused by leakages and constructions. When water is rerouted through temporary connections due to construction work, it often affects the flow measurements in ways that would indicate increased or decreased usage while, in fact, the amount of water entering the sewage system remains unchanged.

In order to predict cold water usage, the time of day is considered to be the best input variable. Experts at Veitur shared their experience on the subject and noted that cold water is likely the most consistent of all inputs to the sewer system and that it has only slight seasonal variation in some neighbourhoods but is almost completely determined by the time of day and the day of the week. Figure 4.2 demonstrates this consistency across an entire year, indicating that daily and weekly patterns of water usage are the most significant factor in cold-water usage. In order to turn the time of day into a usable variable for the drainage flow models, it is represented by a one-hot encoded representation.

One-hot encoding is a commonly used machine learning technique that transforms categorical variables into numerical variables to make them compatible with regression and classification models. This encoding works by assigning a new variable to all the unique values of the old variable. Since the drainage flow data is hourly, the relevant, unique values are the 24 hours of the day. In one-hot encoded format, the hours of the day look as follows:

$$
\begin{aligned}
\text{hour} &: [\text{One-hot encoded hour}] \\
0-1 &: [1,0,0,...,0,0] \\
1-2 &: [0,1,0,...,0,0] \\
2-3 &: [0,0,1,...,0,0] \\
&\dots \\
22-23 &: [0,0,0,...,1,0] \\
23-0 &: [0,0,0,...,0,1]
\end{aligned}
$$

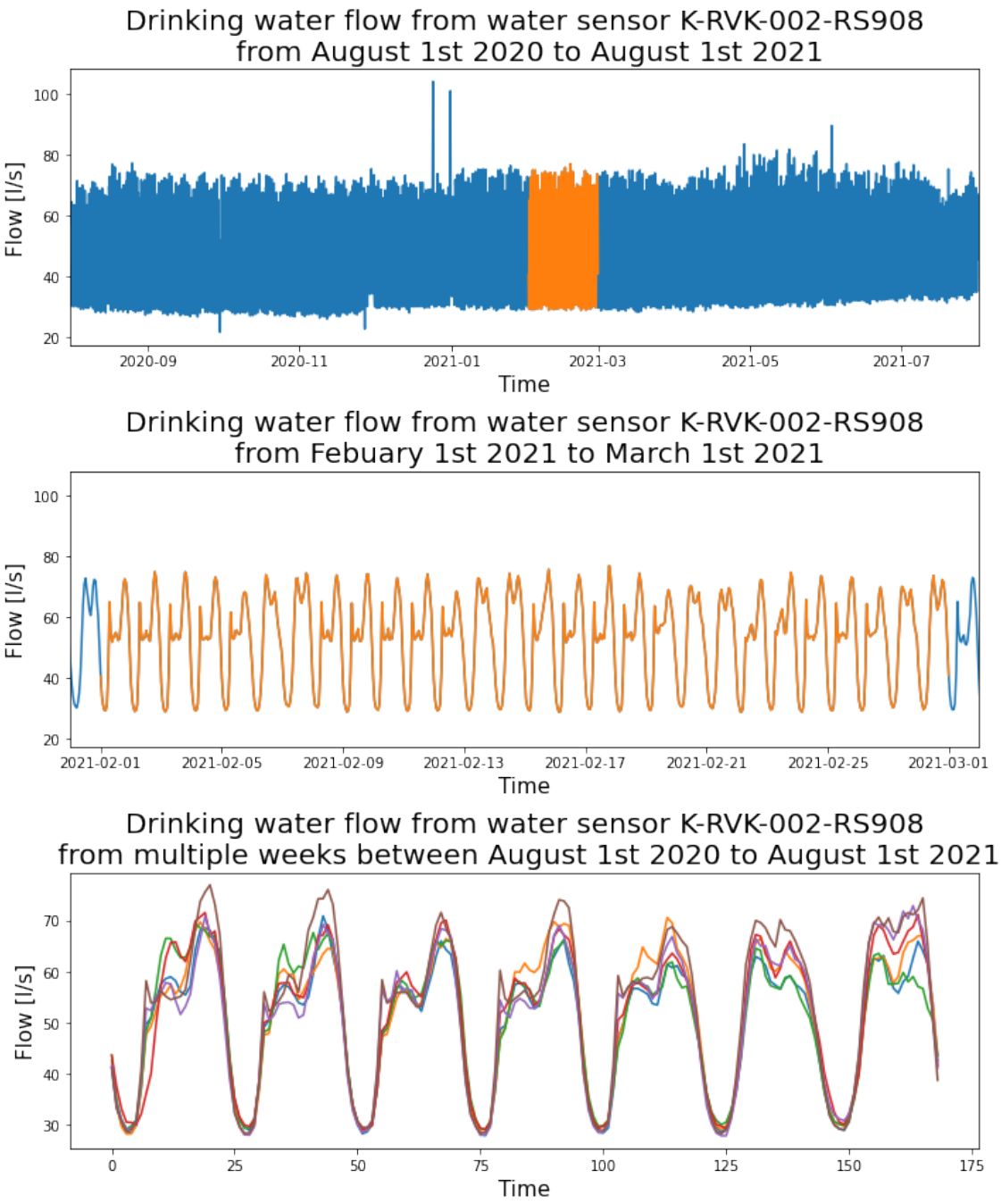

Figure 4.2: (Top) A year of cold water time series data for sensor 'K-RVK-002-RS908', just one of many cold water sensors in Reykjavík. (Middle) A zoomed-in view of the highlighted month from the top graph shows a consistent and repeating daily pattern of cold water usage.(Bottom) Six consecutive weeks superimposed to demonstrate the consistency of cold water usage in Reykjavík, dependent almost exclusively on the time of day

4.1.2 District heating

The district heating in Iceland is geothermally powered and, in contrary to most district heating systems, is an open-loop system that often ends in combined sewers. According to district heating experts at Veitur, they currently employ a data-driven model to forecast district heating flow that uses a sequence of past temperature observations with a sequence of forecasts appended. They combine this forecast with the most recent observations with an auto-regressive model, which is a model that assumes that the output variable is a

function of its own previous values plus some stochastic term, which is an imperfectly predictable term (Shumway et al., 2000).

The district heating usage does not depend directly on the current temperature but rather the need for heating in houses. Other weather variables may also be relevant such as wind speed, moisture and solar radiation but, for simplification purposes and because Veitur has had good results by using just the outside temperature, only the temperature will be used in the models. The models used by Veitur use an exponentially weighted sum of a two-week long sequence of hourly temperature observations in their model, so this model will also contain more than just a single temperature prediction.

For the simulations, only past observations will be used in the sequence included in the input. The forecasts will substitute past observations with forecasted temperature. For example, if a model requires a sequence of 24 temperature observations, the model will use the 24 most recent observations to predict the current output but will use only 12 observations and 12 forecasted values to predict the output 12 hours into the future. This processing is often referred to as a sliding window view of the data. See figure 4.3 for a visual demonstration.

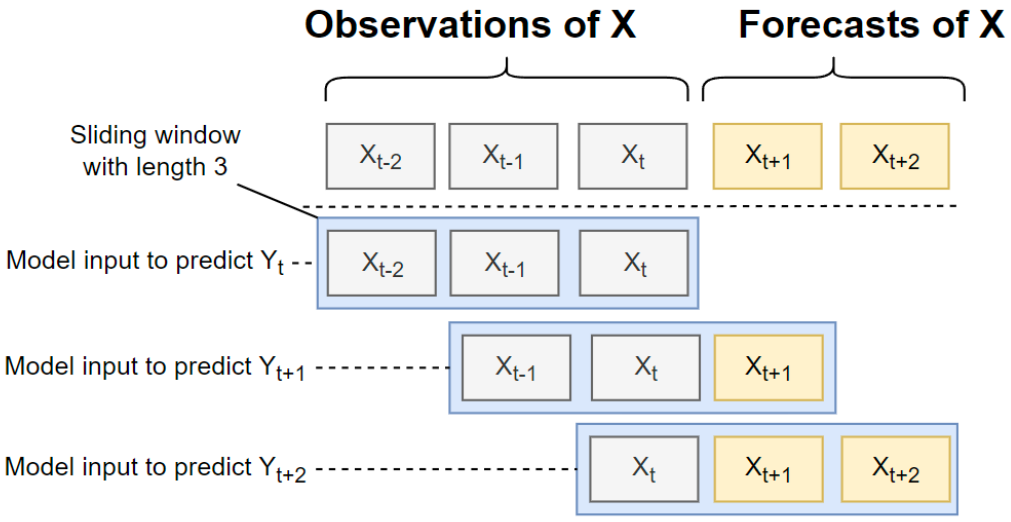

Figure 4.3: A demonstration of a sliding window view applied to a vector with three observations and two forecasts appended. The top row shows all the data in the sequence. The first vector from the sliding window view will include only past observations in the sequence, while the next will include a single forecasted value since the sliding window shifts by a single time step with each consecutive view created.

4.1.3 Rainfall-runoff

Rainfall-runoff modelling in both hydrology and urban hydrology is faced with some of the same problems like a need for accurate rainfall measurements. Urban hydrology faces additional challenges due to the smaller catchments and the different surface characteristics within cities than out in nature. Rainfall within a city runs off streets, sidewalks, houses and other surfaces into the sewer much more quickly than rainfall out in nature can run into streams and rivers. Urban catchments may also be small enough that the closest rain gauge might be too far away in some cases to measure rainfall within the catchment. The two most relevant data sources for simulating rainfall-runoff are:

- Radar

- Rain gauges

The most significant obstacle for using radar data in rainfall-runoff modelling is processing raw radar echo data into usable rainfall estimates and adjusting the rainfall estimates with ground observations like rain gauges. This adjustment is called radar-rain gauge merging. The Icelandic Met. office does not supply any adjusted radar rainfall estimates, and the rain gauges within and around the greater capital region, until now, have been too sparse and some of them of too low quality to use for radar-rain gauge merging, so radar-rain gauge merging will not be performed or explored much more in this paper. Radar-rainfall estimates without adjustments are considered a considerably worse source of rainfall data than rain gauges, but they might still be the best rainfall estimates in regions far from any rain gauge. In section 4.2.1 the steps used to transform the raw radar data into rainfall estimates are described. The best available observational data for rainfall for this study is from rain gauges, and so it will also be used in the experiments later in this chapter to predict rainfall-runoff. One limitation of rain-gauges is that they cannot be used to forecast. The two most commonly studied data precipitation forecasts for use in hydrology and urban hydrology are:

- Numerical weather predictions
- Radar nowcasts

Radar nowcasts are described briefly in the literature review bit since the Icelandic Meteorological Office does not have a radar nowcast, these will not be considered more in this report. However, due to the possibility of future applications requiring the use of radar nowcasts, it is helpful to consider the experiments described in this chapter as setting a baseline of performance for radar nowcasts. The accuracy of a radar nowcast will likely be worse than that of radar observations. The only precipitation forecasts used in this paper are from numerical weather predictions. The data used in this study is described in chapter 3. Section 4.2.2 describes how the accuracy of the radar data and NWP are evaluated using rain gauges. Although these cannot be compared directly since they are not evaluated on the same subset of rain-gauge data and since the resolution of the radar estimated rainfall is different from the NWP resolution.

4.2 Data processing and evaluation

4.2.1 Radar processing

The main objective of the radar data processing described in this section is to turn the raw radar data into a 2D distribution of rainfall over the area of interest. Initially, the radar data in this particular data set comes in a sequence of 420 by 120 matrices of 8-bit unsigned integer values with accompanying metadata. The metadata includes the radar scanning strategy, the timestamp of the scan, and the radar's location.

A radar scanning strategy is composed of the instructions the radar follows like angles above ground (elevation angles) and angles around its spin axis (azimuth angle) and characteristics of the radar beam like the width of the beam, maximum distance and resolution along its length. With all these values, it is possible to calculate the coordinates and altitude that each pixel of the radar image represents. The two most common radar strategies used during the six-year period from which the data was gathered can be seen in figures 4.4 and 4.5, which were generated with functions from the python library Wradlib (Heistermann et al., 2013).

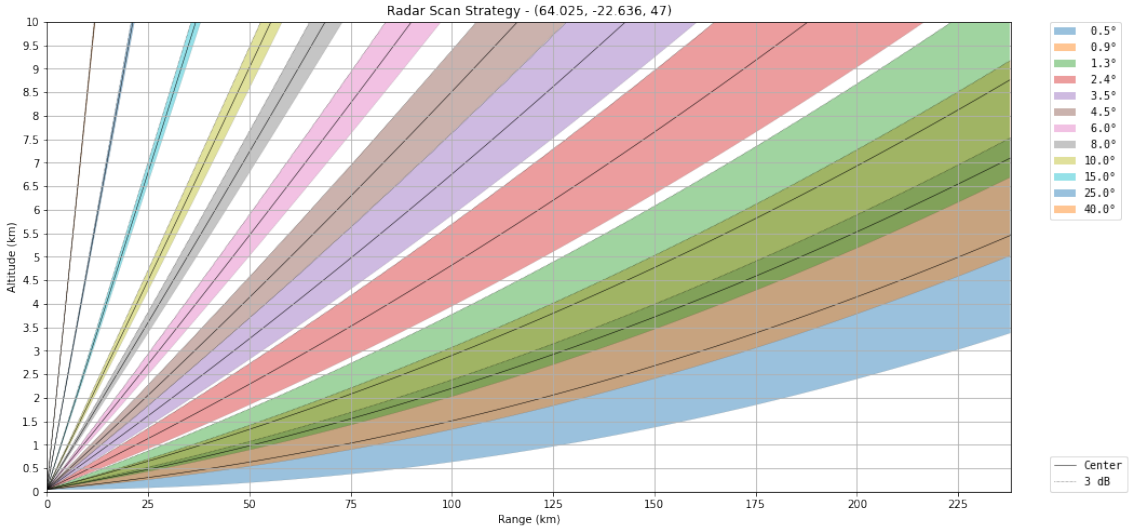

Figure 4.4: Radar strategy 1. This strategy was the primary radar strategy used from the start of 2015 and to the end of 2017. This strategy allowed for a 15 minute frequency of complete scans.

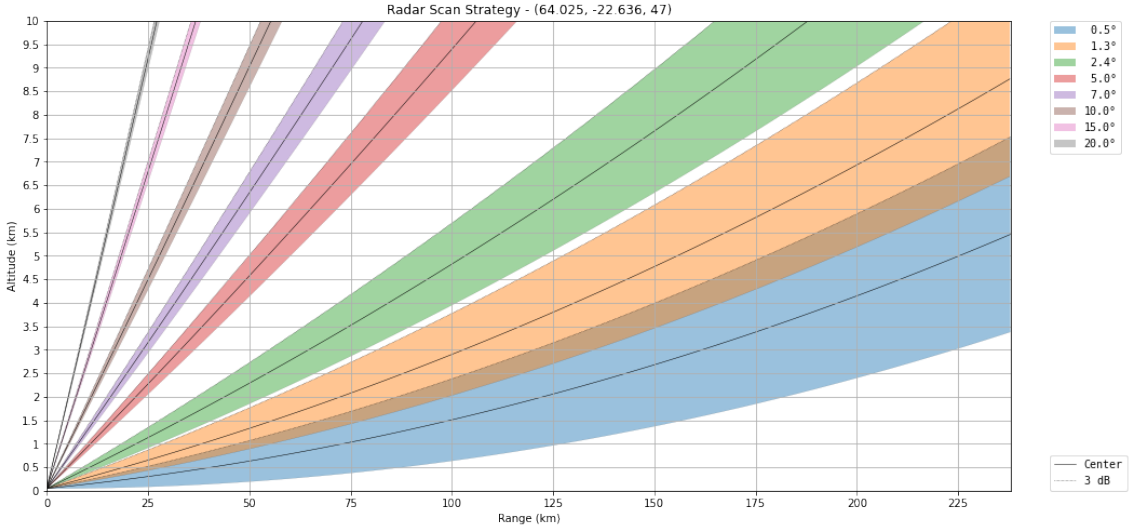

Figure 4.5: Radar strategy 2. This radar strategy was the primary radar strategy used from the start of 2018 to the end of 2020. It has much fewer elevation angles and was thus able to complete each complete scan more quickly, one every 5 minutes

Usually, the data from different strategies would not be compatible with the same model since there might be different numbers of elevation angles and more layers of radar data but also because the elevation angles might not be the same, and so the layers would not contain estimates of the same volume. This incompatibility can be overcome by computing a Constant altitude plan position indicator (CAPPI). The CAPPI involves first computing the position of all the measured volumes in space and then interpolating or using the closest value to some constant altitude. The CAPPI can be computed regardless of the radar strategy. It allows for the generation of a single continuous data set that can be used with the same model. Figure 4.6 gives a good visual description of how the CAPPI and pCAPPI are computed. A slight variation of CAPPI called pseudo-CAPPI or pCAPPI will use the lowest and highest available layers in the 'blind spots' of the radar, below and above the radars scanned volumes. Conventional CAPPI would otherwise remove or ignore

these blind spots. The implementation of the pseudo-CAPPI used in this paper came from the python library *wradlib* (Heistermann et al., 2013).

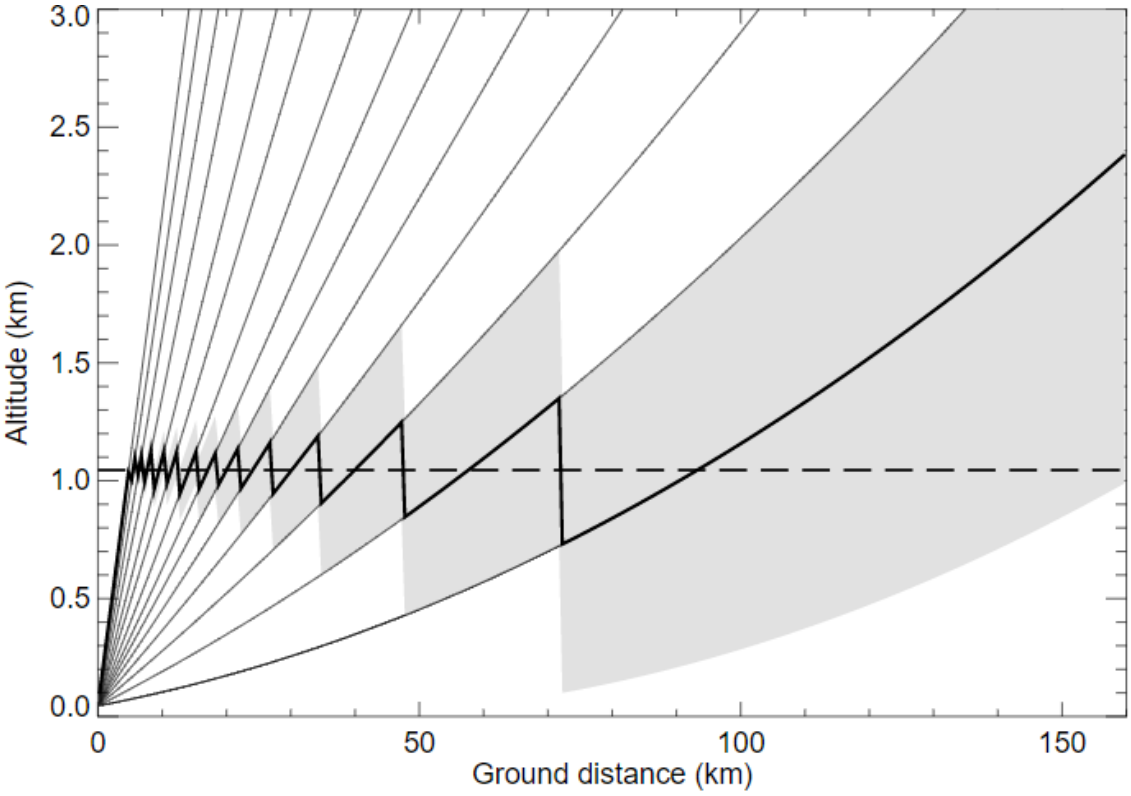

Figure 4.6: (Uijlenhoet et al., 2006) Visualization of how CAPPI is computed at 1 km above the radar. The construction uses 14 elevation angles. The dashed line represents the constant altitude for which the pCAPPI is calculated. In this case, it is 1km above the radar. The solid bold line represents the closest beam centres to the selected height at a given distance from the radar

The pCAPPI images have a resolution of 600x600 and cover a radius of 250km. Two pCAPPI data sets were used. One at 2km, the elevation used by IMO, and 1.25km, which is halfway between 2km and 500m, the lower end of the altitude at which nimbostratus clouds appear (Goodson & Kopnitsky, n.d.). Pixels derived from the data cover a roughly 0.8x0.8km area each. Figure 4.7 shows these pixels along with the position of the weather radar and the three rain gauges used to evaluate the radar estimated rainfall. In order to use the different resolutions and match the resolution to the manually reviewed rain gauge data, the pCAPPI values were first converted to rainfall and then aggregated to hourly mean values of rainfall.

The radar reflectivity is stored in units of *decibel relative to Z* (dBZ). To convert the radar reflectivity to rainfall estimates you first convert the decibel relative to Z to just Z with the equation $f(x) = 10^{\frac{x}{10}}$ and then to turn Z to rainfall (R), the equation $f(z) = (\frac{x}{a})^{\frac{1}{b}}$ is used. The particular values of a and b will vary depending on droplet size distribution, but the combination of $a = 200$ and $b = 1.6$ is called the Marshall–Palmer relation and is often used as a default. The Marshall-Palmer relation is used, but the values for a and b are also computed to minimize the mean-squared error between the rain-gauge values and the radar rainfall estimates closest to the rain gauges. This is described in more detail in the next section.

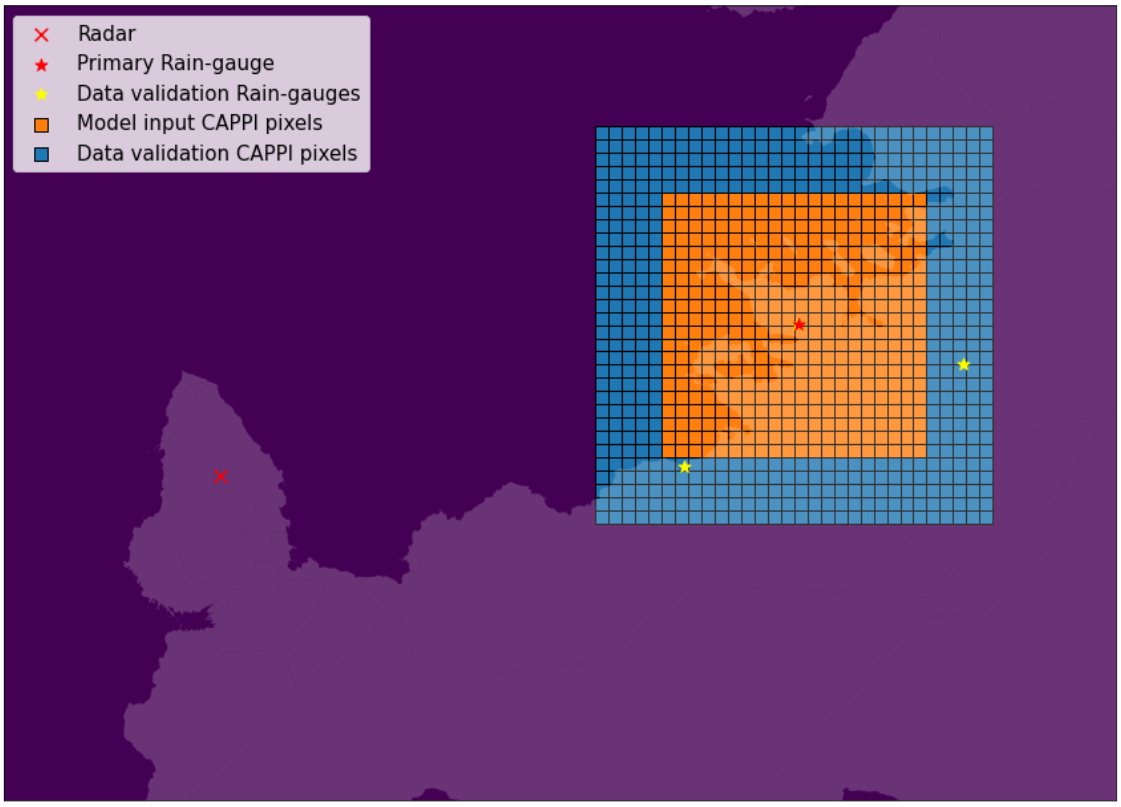

Figure 4.7: Radar and primary rain gauge positions and the pCAPPI pixels overlaid on a map. The 20x20 grid of orange squares that represent the pixels that are fed into the models described in section 4.3. The blue pixels represent the pCAPPI pixels specifically generated added to perform the evaluation described in the next section

4.2.2 Radar and NWP evaluation

Since one of the research objectives is to evaluate the quality of the data available and suggest improvements to the data or data processing, it is important to evaluate one of the primary inputs into the models, rainfall observations and forecasts. The three types of rainfall data used in this paper are rain gauges, radar estimated rainfall, and numerical weather predictions (NWP). Rain gauges are often treated as 'ground truth' for rainfall observations in hydrological research, and so in this section, they will be treated as such. One important thing to note is that rain-gauges don't estimate the same thing as radar and numerical weather predictions. Radar data is volumetric in origin, and NWP estimate rainfall over large areas, so even if all three data sources were perfectly accurate, they could still give different values for rainfall in a particular location.

The three rain gauges introduced in chapter 3 will be used for evaluation. The data used for these evaluations is from 2015 to 2020. January and September of 2015 are not included because changes were made to the model and its forecast horizon in September 2015. Every data source also has gaps, some gaps are only a single missing hour, but the most significant gaps can be days or weeks long. The year 2020 is also left untouched in this analysis and used exclusively as test data input for the experiment described in section 4.3.

The pixel closest to each of the rain gauges is used to evaluate the radar and NWP data with the rain gauges. An overview showing the positions of the radar, the rain gauges and the closest NWP and pCAPPI pixels are shown in figures 4.8 4.9.

The raw radar has a temporal resolution between 15 and 5 minutes depending on

the radar strategy used, and the raw rain gauge data has a 10-minute resolution. The numerical weather predictions have a temporal resolution of 1 hour, and the manually reviewed rain gauge data from the IMO is only available as an hourly time series, so to compare the three, they are all transformed to hourly average rainfall time series. Both the 2km and the 1.25km pCAPPI layers described in section 4.2.1 are compared to see if the altitude has a significant effect on the quality of the estimated rainfall.

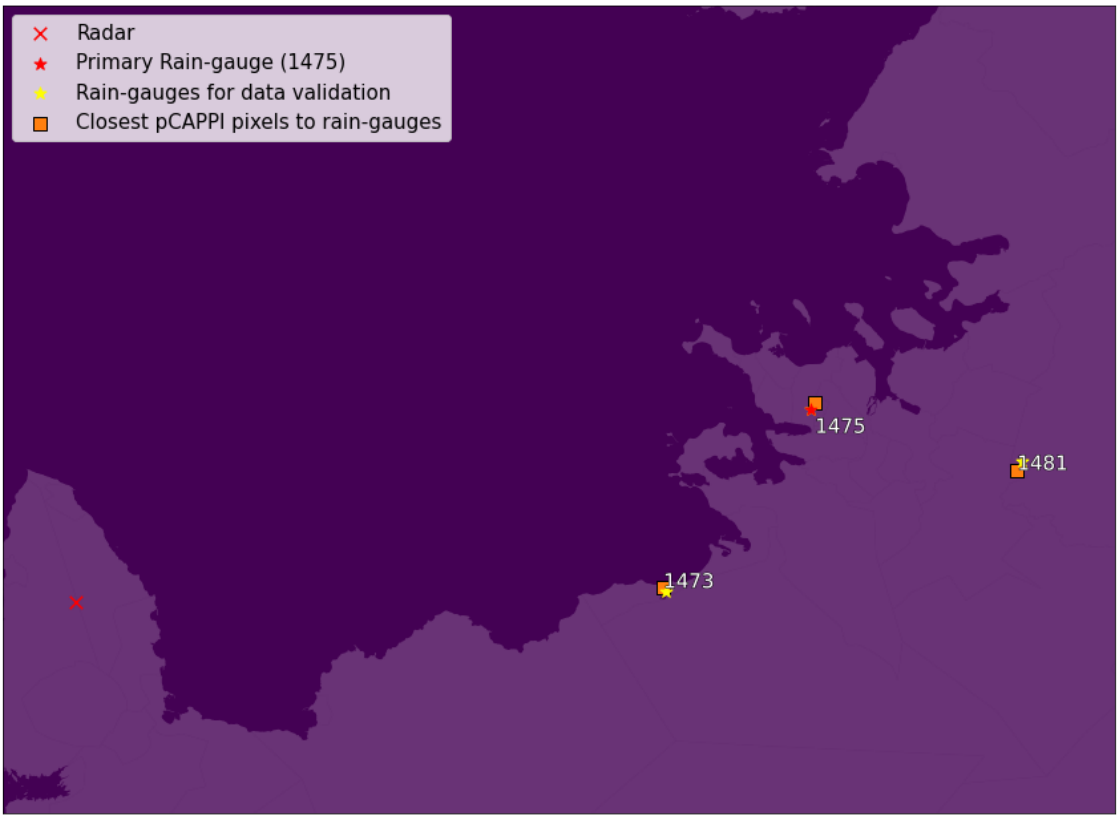

Figure 4.8: The locations of the weather radar, marked with a red 'X' on the Reykjanes peninsula, as well as the three main rain gauges. The primary rain gauge (1475) is marked with a red star, while other rain gauges (1481 and 1473) are marked with yellow stars. The closest available pCAPPI pixel is shown as an orange rectangle

4.2.3 Drainage flow data

The drainage flow data used in this paper is the outflow from the pumping stations, not the inflow. This distinction is important for two reasons, both of which result in a disparity between inflow and outflow. The first reason for this disparity is due to CSO discharges or rainwater pumps since the flow is not measured within these alternative paths. The other reason for the disparity between inflow and outflow is because the chamber within the pumping stations, see figure 4.10, acts as a buffer and the volume of water within the station may be different at the start and end of the hour.

It should be noted that this difference of inflow and outflow is partly due to the type of control system strategy used by Veitur. In figure 4.11 a plot taken from Veitur's Supervisory control and data acquisition (SCADA) system shows a time series with the outflow and water level within the Boðagrandi pumping station. At around 9 AM on the graph, the water level is about 1.7m, and then at 10 AM, the water level is down to 1m. The water level within the pumping stations goes through cycles caused by the control system, which turns the sewage pumps on and off depending on the water level. The

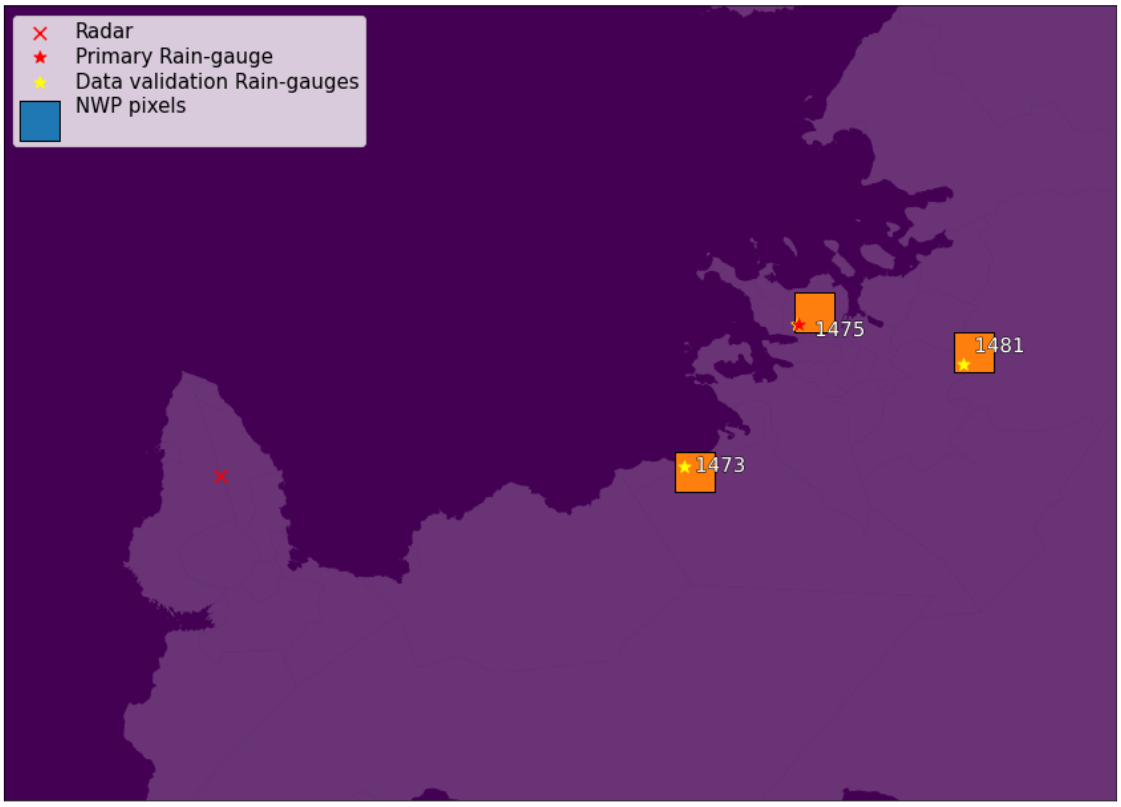

Figure 4.9: The locations of the weather radar, marked with a red 'X' on the Reykjanes peninsula, as well as the three main rain gauges. The primary rain gauge (1475) is marked with a red star, while other rain gauges (1481 and 1473) are marked with yellow stars. The closest available NWP pixel is shown as an orange rectangle

difference in volume within the pumping station at the end of the hour and the start of the hour divided by the size of the time interval (one hour for most historical data) is equal to the difference in average inflow and average outflow.

Estimating or restoring the flow during the extreme rainfall events is impossible with the current data but estimating the inflow from the outflow and water level observations is possible with some minor changes. Between August 13th 2021 and December 2021, the average flow and water level data were logged at minute intervals instead of hourly. This was done in order to try to measure how great the discrepancy is.

The change in water level between hours is used to estimate the inflow for each hour. Although the average water level during the first minute is not exactly equal to the water level at the start of the hour, it is used as a close approximation. This number only gives a change in the height of the water within the station, not the volume. In order to estimate the volume, a heuristic is used to estimate the cross-sectional area of the chamber. In order to proceed with this heuristic, the following assumptions have to be made:

- The area of the horizontal cross-section of the chamber is approximately the same at all heights
- The flow into the station is approximately smooth most of the time (i.e. the rate of flow changes slowly)

In addition to these assumptions, a metric for smoothness is needed. Let's first define a

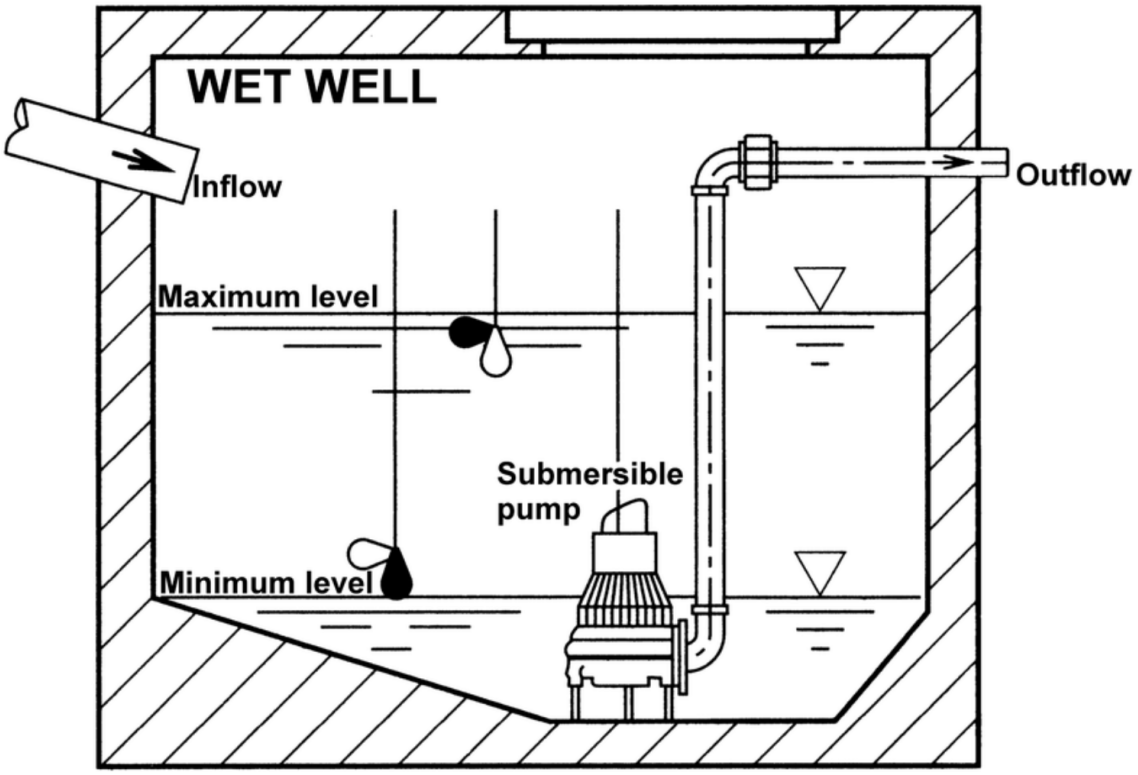

Figure 4.10: Diagram of a pumping station wet well with a single submersible pump (Fecarotta et al., 2018). This diagram shows that the water level can vary substantially from the minimum to the maximum heights, which may cause a difference between the average inflow and outflow during any given hour.

function for the estimated inflow as

$$f_{\text{in}}(\mathbf{t}, A) = f_{\text{out}}(\mathbf{t}) + \Delta h(\mathbf{t}) * A$$

where $f_{in}(\mathbf{t})$ is the flow into the station at time interval $\mathbf{t}$, which is the average flow between some time t_0 and $t_0 + t_1$, $f_{out}(\mathbf{t})$ is the average flow out of the station during time interval $\mathbf{t}$ and $\Delta h(t)$ is the change in water level over the time interval t, and A is the area of the cross section of the chamber within the pumping station. The metric for smoothness across the whole time series is defined as the sum of the differences between all observations. We then define smoothness as the function

$$s(A) = -\sum_{\mathbf{t}} \Delta f_{in}(\mathbf{t}, A)$$

where $\Delta f_{in}(\mathbf{t}, A)$ is the difference between the average flow from the last time interval to the current one. Then finally we maximize s with respect to A, i.e. we solve

$$\underset{A}{argmax}\, s(A)$$

and then we can compare $f_{in}(\mathbf{t}, A)$ with $f_{\text{out}}(\mathbf{t})$ and evaluate it based on the average, maximum and minimum difference or the mean absolute error.

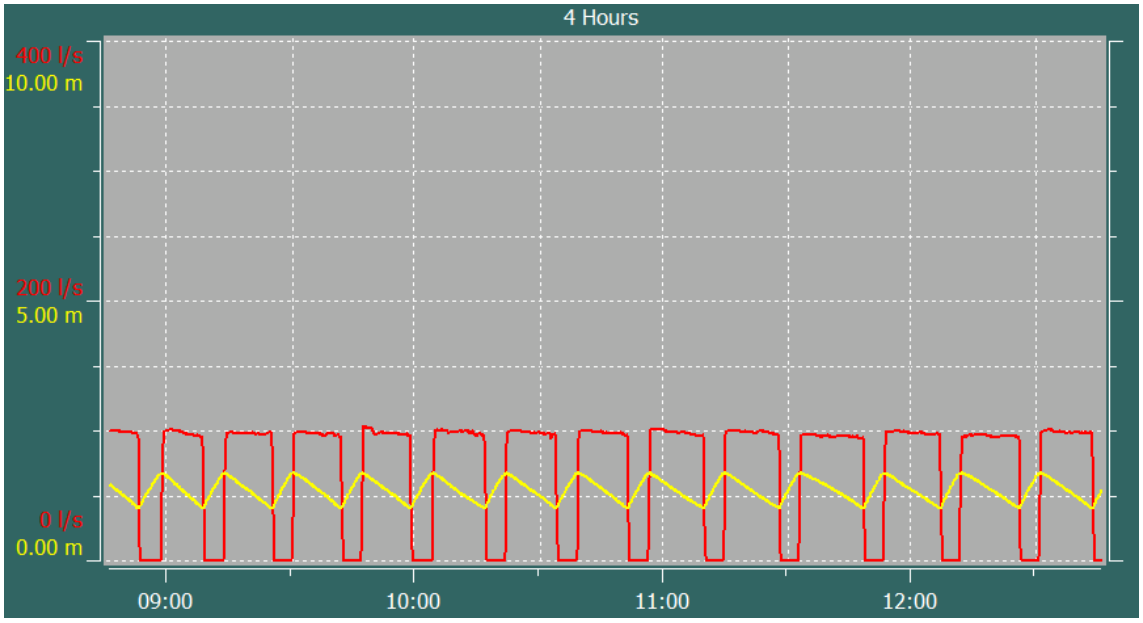

Figure 4.11: Plot taken from the SCADA dashboard used by Veitur around noon on December 4th 2021. The plot shows the flow measured going out of the Boðagrandi pumping station (red) and the water level (yellow) at the same time. The X-axis is time. The periods in both time series are due to the nature of pumps and control systems used by Veitur.

4.3 Drainage flow simulation and forecasts

This chapter will describe the models, model training and evaluation methodology for the drainage flow simulation and forecasting models. The term simulation refers here to predictions made about the system's current state using a sequence of past observations, like temperature and rainfall. The term forecast refers to predicting future states of the system using a sequence of past observations as well as forecast sequences for variables that determine the system's future state, such as temperature and rainfall forecasts. Since one of the objectives of this thesis is to design a single model that can be used for both purposes, any implications relating to one or the other will often apply to both simulations and forecasts. The word prediction may be used to refer to both simulations and forecasts.

In order to gain insight into what factors are most important for the drainage flow models, several variations of a machine learning model were designed. One of the objectives named in the introduction was to design a model that can be used for both simulation and forecasting. This design criterion was used to evaluate the effect of the data source and model design separately. I.e. If the same model performs very differently on two data sources, then it implies that the difference in performance can be associated with the data rather than the model.

4.3.1 Models

The goal of these models is to predict drainage flow in the wastewater system in Reykjavík. Three different sources of rainfall data are tested, two of them for drainage flow simulation and one for drainage flow forecasting. The rain gauge data used in this experiment is a one-dimensional time series with hourly rainfall observations. The radar data is a 400-dimensional time series since the source data is a 20x20 grid. The NWP data is a sequence of 100-dimensional sequences. Each forecast consists consists of 60 time steps. Since one of the goals of the model design is that the same models can be used for both the forecasting and the simulation task without changing the model, a special approach must be used

with the data.

Figure 4.12 shows how the only difference between the forecasting task and the simulation task is how the data shaped and ordered. Every model gets an input in the form of a 3-way tensor or a sequence of matrices. Each individual simulation or forecast is computed from a matrix with the shape (sxp), where p is the number of input variables, which includes the rainfall, temperature and time of day variables. The temperature is a 1-dimensional variable, the time of day is a 24-dimensional variable and the three sources of rainfall are 1-, 400-, 100-dimensional for the rain gauge, radar and NWP respectively. So for example one model could have radar data input and a sequence length of 48, so the input data would be in the shape ($nx48x425$). Both the data source and the length of the sequence are treated as hyperparameters.

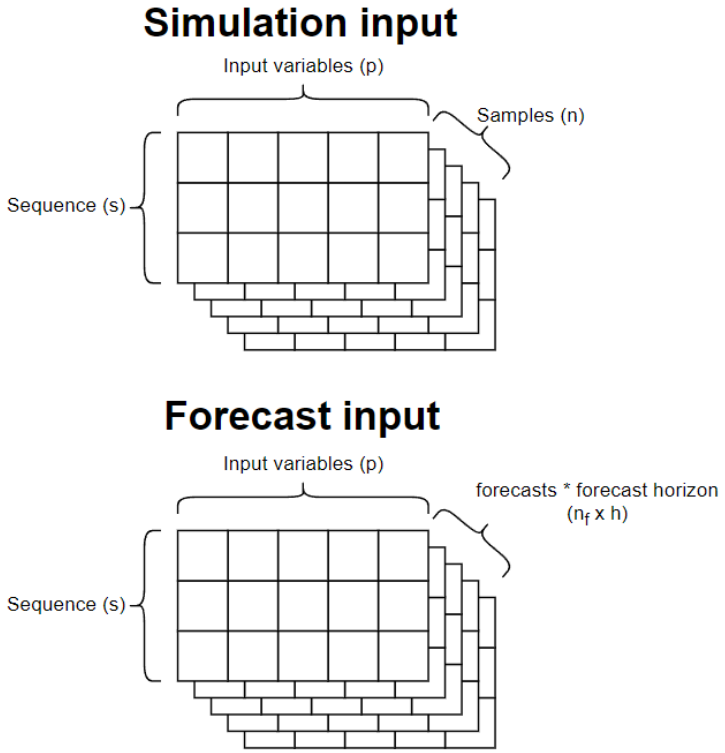

Figure 4.12: The difference between the simulation and forecasting tasks comes down to only the shape of the input data. The sequences for the simulations **(top)** are comprised of past observations, and the number of samples is one per hour since that is how much observational data exists. The sequences for the forecasts (Bottom) are comprised of sequential forecasts, and the number of samples is equal to the number of forecasts made, which is one per 6 hours, times the forecast horizon, 60 hours

Due to the high dimensionality of the input and the potential length of the sequence, the type of model that can be used becomes limited by the computational cost of optimizing the model. If one flattens out the input for a model using the radar data and a sequence length of 48, and uses a linear model, the number of coefficients required would be $20'400$, and the time it would take to optimize this model using the normal equations could be too high for practical model selection.

One way of handling the high dimensionality of the data is by using special variants of artificial neural networks (ANN). Models such as recurrent neural networks (RNN) are generally used to model data with temporal sequences. Convolutional neural networks are

generally used to model data with spatial information or apply a form of dimensionality reduction along a temporal sequence if applied in a certain way.

Artificial neural networks

ANNs are a class of models that all share a few key features that make them particularly useful for all kinds of tasks. One of the key attributes of neural networks is their modular nature. Conventional ANNs are comprised of units called layers or hidden layers, and each layer contains one or more neurons or nodes. The nodes are connected to each other in a structured way where the output of the nodes in one layer is connected to the input of the nodes in the next layer. Each neuron is defined primarily by two aspects, it's connectivity and its activation function. Some types of neurons like those in Recurrent neural networks also have states which are passed between steps along a temporal sequence. The connectivity is defined by which neurons in the previous layer are passed as input into the neuron, and the activation function is a function applied to the weighted sum of the inputs. The weights in this weighted sum are learned by the model during training, and the activation function is usually a simple non-linear function with desirable computational properties.

A 'fully connected' neural network is one where the nodes in every node are connected to every other node in the previous layer. The connections in a fully-connected ANN never form a cycle within the network, and this property makes them feedforward neural networks. A different type of ANN called recurrent neural networks (RNNs) allow for connections to exist along a temporal sequence, or in other words, back to nodes in the same layer at another time step. RNNs can use their internal state, also called memory, to process sequences of inputs. These characteristics allow RNNs to learn temporal relationships. Another variation of ANNs called convolutional neural networks (CNN), also known as shift invariant or space invariant artificial neural networks. These networks, similar to RNNs may be applied along a temporal (or spatial) sequence, but they do not share information or connections along this sequence. Instead, they apply the same set of weights along this dimension. When applied along a temporal sequence, they are sometimes referred to as 'Time delay neural networks' (Abiodun et al., 2018; Egmont-Petersen et al., 2002; Tealab, 2018; Zell, 1994).

Part of the motivation for picking CNNs and RNNs is that the number of parameters stays completely unchanged when a longer sequence of inputs is used with a recurrent model and grows much slower for CNNs than for linear models. The same is not true with linear models where the number of parameters increases for every added variable, and so the number of parameters would scale with the product of the length of the sequence and the number of variables. It is harder to interpret the impact of adding a longer sequence to a model when the number of parameters in the model also increases, which can cause overfitting. The same is true for temporal convolutional neural networks, although their output changes shape when the number of sequences is changed. The implementations of the neural network models used in the paper were all done with the python package Tensorflow (Martín Abadi et al., 2015).

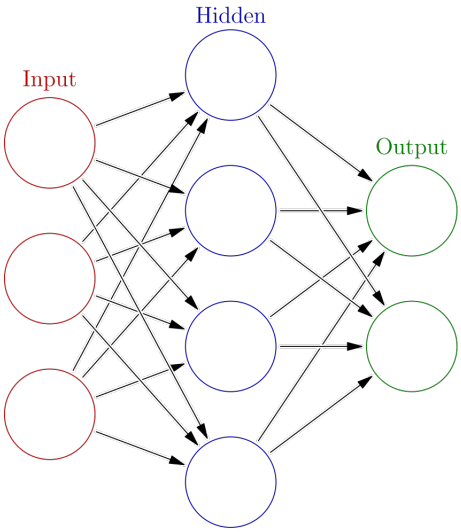

Figure 4.13: (Commons, 2013) Diagram of a fully-connected feed-forward neural network with three input variables, a single hidden layer that contains four nodes and two output nodes

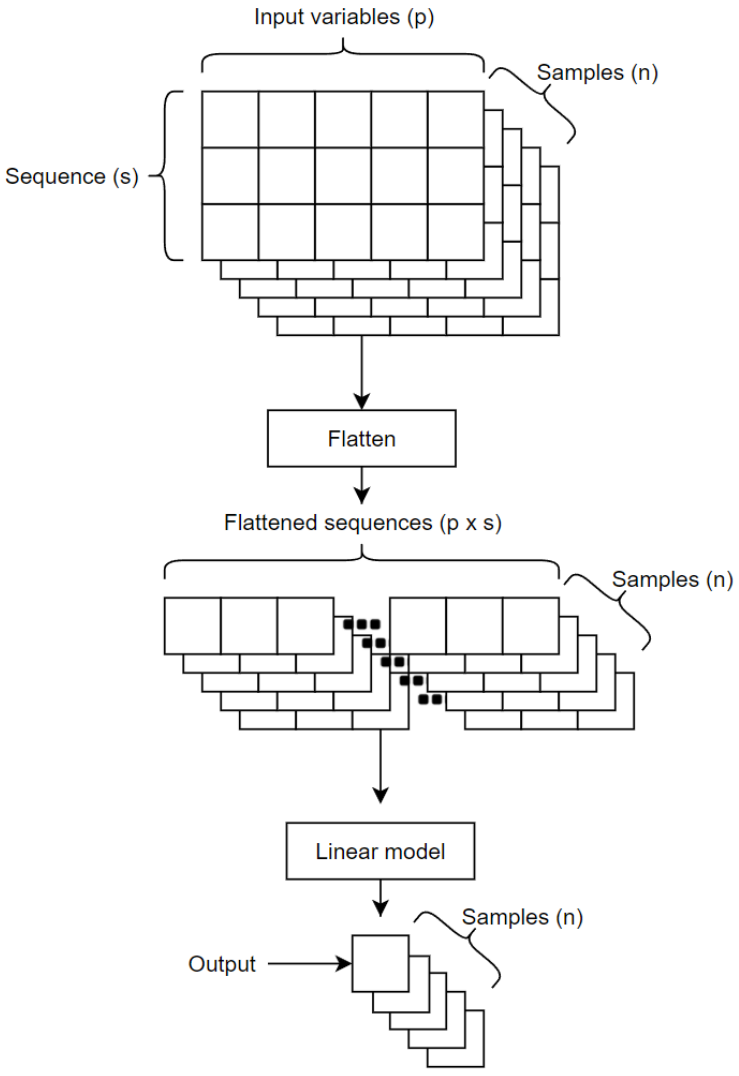

Figure 4.14: Diagram showing how each step in the linear model transforms the data. The input is a tensor with dimensions (n x s x p) where the sequence is the current and previous observations of the input variable

Linear model The first type of layer presented here is a fully-connected neural network with no hidden layers and a linear activation. This model is linear, and since it also has a squared error loss function, the only difference between this linear model and those more conventionally used is how it is optimized. The optimization is described in section 4.3.2. The main purpose of the linear model in this paper is to include it as a baseline of performance.

In order to make this model compatible with the same input as the other models, a flattening layer is added between the input and the linear model. This collapses the sequence into a vector of length equal to the product of the sequence length and the number of variables. See 4.14 for a diagram showing the shape of the data from input to output through this model.

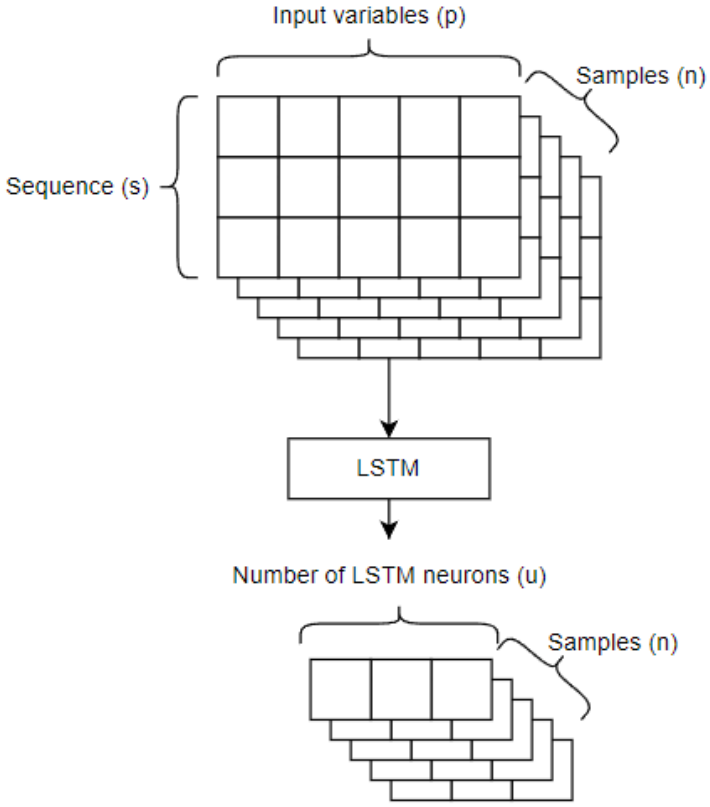

Figure 4.15: Diagram showing how each step in the LSTM transforms the data. The input is a tensor with dimensions (n x s x p) where the sequence is the current and previous observations of the input variable

Long Short-Term Memory (LSTM) is an artificial recurrent neural network (RNN) architecture. In contrast to the standard feedforward neural networks, the RNNs have feedback connections which mean that they not only get input from the layer before them but also pass information back into themselves so that the next time the node processes an input, it also has information from its previous state, see figure 4.16. The plain RNN will only pass forward a so-called 'hidden state',h , also known as the output state, while an LSTM also passes forward a 'cell state',c , whose value is regulated by three gates, σ_g. Gates are sigmoid functions, which are mathematical functions that have a characteristic "S" shape and return a value between 0 and 1 (Han & Moraga, 1995). The output of these functions can be thought of as representing an open, if the value is close to 1, or

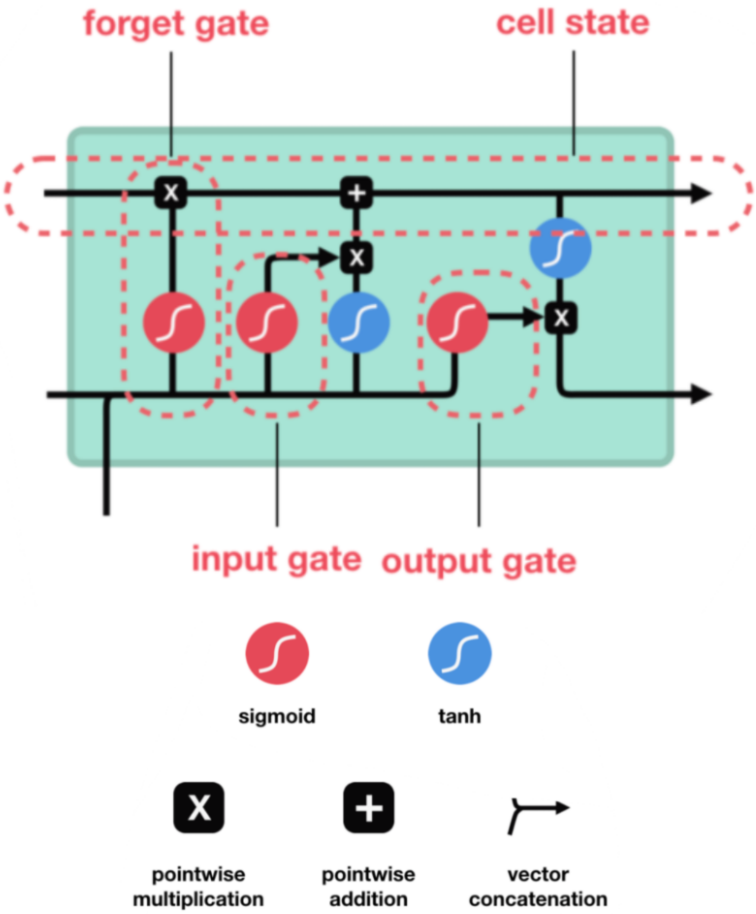

Figure 4.16: Image from (Phi, 2018). Diagram of the inner workings of an LSTM. Highlighted and labeled are each of the gates that control the flow of information to and from the cell state, which is shown as an arrow in the direction it travels through the cell. Below the diagram

closed, if the value is close to 0, gate. They act as gates because they are applied with the Hadamard product, an element-wise product, it can scale the values by any number between 0 and 1, which can be thought of as the open and closed states. Thus the gate is considered closed since nothing passes through. The three gates of the LSTM are the input gate, i, the output gate, o, and the forget gate, f. The input gate regulates what may be added to the cell, the output gate regulates when the value of cell state may passed to the hidden state, and the forget gate regulates how much of the previous state is preserved. The exact equations for each forward pass of an LSTM are:

$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \sigma_h(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
h_t &= o_t \circ \sigma_h(c_t)
\end{aligned}
$$

where the initial values are $c_0 = 0$ and $h_0 = 0$ and the operator $\circ$ denotes the Hadamard product, t is the index of the time step, x_t, input vector to the LSTM unit at time t, h_t is the hidden state vector, $\tilde{c}_t$ is the cell input activation vector, c_t is the cell state vector, W,U, and b are the weight matrices learned by the model during training. The function σ_h is the hyperbolic tangent function(Hochreiter & Schmidhuber, 1997).

Temporal convolution A convolutional neural network (CNN) is similar to a recurrent neural network in that the same weights are applied multiple times across different inputs. CNNs become more relevant when the data are tensors rather than vectors. A tensor is a generalization of vectors that extends to more dimensions. A vector is a 1-way tensor, and a matrix is a 2-way tensor. CNNs are useful when working with tensor data because there is often a shared structure or attribute across one or more of the tensor dimensions.

CNNs share weights across different inputs. In the case of temporal convolutions, the weights are only shared across time. Figure 4.17 shows a single sample with a sequence length of two and three input variables per time step in the sequence. The figure also shows the output from a two-neuron temporal convolution, and the colours demonstrate the connectivity. The green arrows represent the weights applied by the first neuron, and the blue arrows the weights applied by the second neuron. Then figure 4.18 shows how this looks when a temporal convolution with two convolutional neurons is applied to multiple samples, each with three sequences. The implementation of this temporal convolution is the Conv1D model from Tensorflow (Martín Abadi et al., 2015).

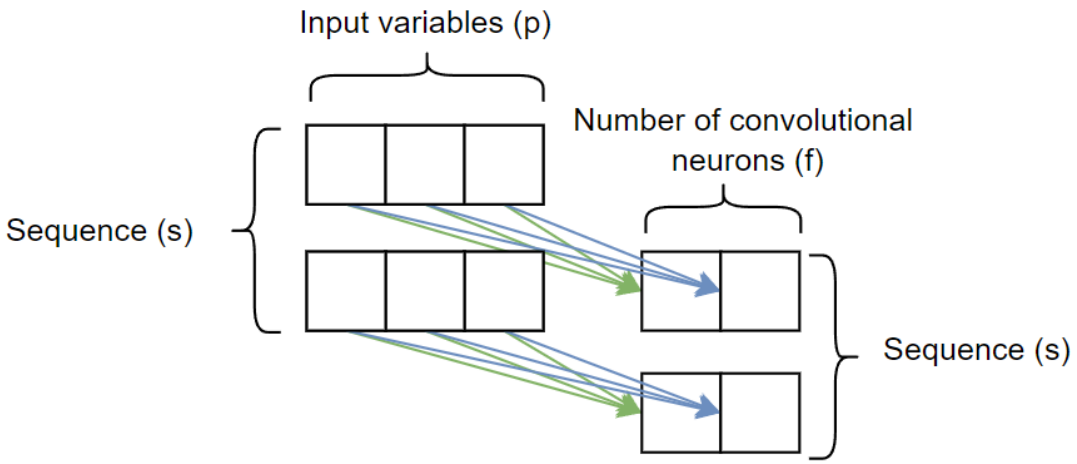

Figure 4.17: A 1D temporal convolution across a single sample with a sequence of length 2 and 3 input features. The number of convolutional neurons is 2. The weights of each neuron are represented with the colored arrows, showing that the same weights are applied to every sequence for each neuron.

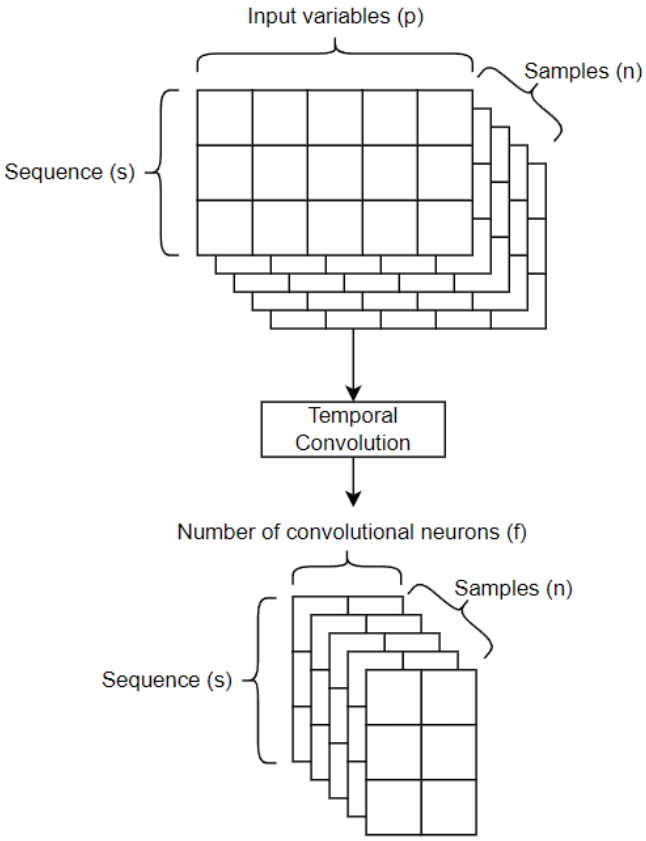

Figure 4.18: Diagram showing how each step in the temporal convolution transforms the data. The input is a tensor with dimensions (n x s x p) where the sequence is the current and previous observations of the input variable

Combinations of model

One of the main benefits of ANNs is their modularity, as discussed in section 4.3.1. That means that they can easily be combined into a single model. Although the models were all trained with one hyperparameter search, there are effectively three different model structures depending on the number of LSTM neurons and the number of convolutional neurons. When either of them equals zero, that layer is not included. A flattening layer is included before the linear model to ensure that the sequences are collapsed into a single vector. If the number of LSTM units is 0 and the number of convolutional filters is also 0, then what is left is only a linear model. If the LSTM is 0 and the filters are non-zero, then you have a Conv1D layer followed by a flattening layer and a linear model. Conversely, if there are 0 filters and non-zero LSTM, the LSTM is followed by a linear model. However, if both have non-zero filters and non-zero LSTM units, the convolutional layer is first. All four variations of this model are shown in figure 4.19.

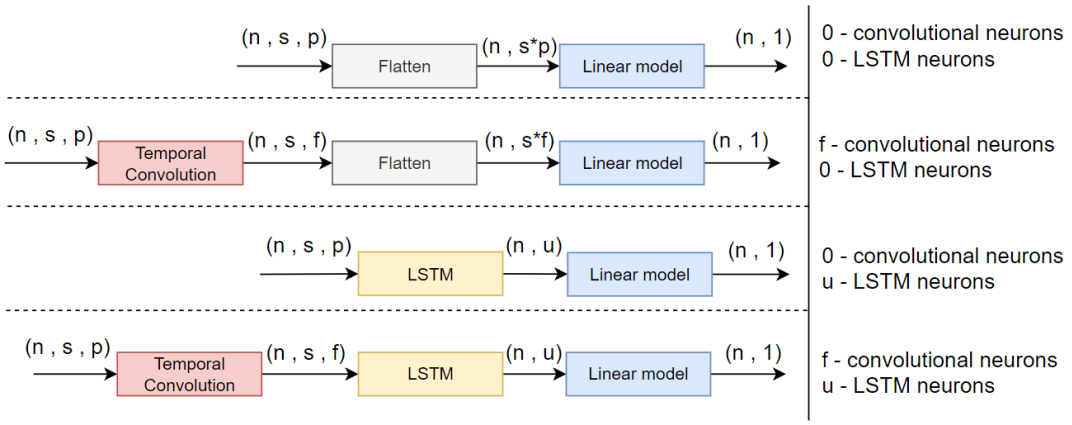

Figure 4.19: The different variations of the neural network components described in section 4.3.1

4.3.2 Model training

Model training is a process that involves the minimization of a loss function using the coefficients of a model. The loss function is often a type of distance function that is zero when the inputs are the same and large when they are different. For these models the following weighted mean squared error loss function was used:

$$WMSE = \frac{1}{n} \frac{\sum_{i=0}^{n} (w_i * ({y_i}' - y_i)^2)}{\sum_{i=0}^{n} w_i}$$

where n is the number of samples, i is the sample index, w_i is the sample weight, y_i and y_i' are the correct and predicted flows respectively for sample i. The weighting scheme was used as a hyperparameter to evaluate if it could be used to improve the performance of the models. Weight scheme a in tables 4.1 and 4.2 has equal weights and is thus identical to the commonly used loss function mean squared error. Weight scheme b is different in that the higher the value of the correct flow, the higher its weight during training. This approach is sometimes used the loss function is different from the function used to evaluate the performance for a given task; see section4.3.3 below for more details on the choice of metric.

Table 4.1: flow thresholds for Gelgjutangi and weights corresponding to the two weighing schemes tested for the models

		Weight scheme	
Flowrate [l/s]	Proportion	a	b
$0 \leq y < 100$	5.41	1	0.5
$100 \leq y < 200$	55.35	1	1
$200 \leq y < 300$	30.67	1	1.5
$300 \leq y < 400$	4.26	1	2
$400 \leq y < 500$	1.65	1	4
$500 \leq y < 600$	0.89	1	6
$600 \leq y < 700$	0.56	1	8
$700 \leq y < 800$	0.41	1	10
$800 \leq y < 900$	0.25	1	12
$900 \leq y < 1000$	0.17	1	15
$1000 <$	0.37	1	20

Table 4.2: flow thresholds for Boðagrandi and weights corresponding to the two weighing schemes tested for the models

		Weight scheme	
Flowrate [l/s]	Proportion	a	b
$0 \leq y < 15$	0.03	1	0.5
$15 \leq y < 30$	6.05	1	1
$30 \leq y < 45$	32.35	1	1.5
$45 \leq y < 60$	30.82	1	2
$60 \leq y < 75$	20.17	1	4
$75 \leq y < 90$	4.82	1	6
$90 \leq y < 105$	1.87	1	8
$105 \leq y < 120$	1.43	1	10
$120 \leq y < 135$	1.22	1	12
$135 \leq y < 150$	0.63	1	15
$150 <$	0.61	1	20

The models were trained using mini-batching and stochastic optimization. This optimization method works by using a subset of the training data, the size of which is called the batch size, to estimate the loss, after which the gradient of the loss function with respect to each coefficient of the model is computed. For neural networks the particular method used to update the weights is called backward propagation of errors, or just backpropagation. After each batch the updates the weights of each coefficient in the opposite direction of the gradient, to minimize the loss. The size of this update is called the step size and is a product of the size of the gradient and the learning rate, which is an important hyper-parameter for neural networks.

The optimizer used is a variation of stochastic gradient descent called Adam, which is an adaptive optimizer based on adaptive estimates of lower-order moments (Kingma & Ba, 2014). Due to limited computational resources, it was impractical to perform a thorough hyper-parameter search, and among the hyper-parameters that were kept fixed was the learning rate. Two methods were used to compensate for not including the learning rate in the hyper-parameter optimization: early stopping and learning rate decay. Early stopping is a method that monitors the loss of a separate validation data set not used for training. A parameter of the early stopping, called patience, determines how many epochs may pass without an improvement to the validation loss before the training is stopped. Learning rate decay works similarly and has a patience parameter, but if this patience parameter is lower than that of the other patience parameter, the learning rate is reduced by some predetermined factor after it is triggered.

Every model was set to train for up to 100 epochs, the early stopping patience was set to 20 epochs, and the learning rate decay patience was set to 5 epochs, and the learning rate decay factor was set to 0.3, so if the learning rate is originally set to 0.1, then after five epochs without improvement it is set to 0.03, then 0.009, etc. The minimum number of epochs between these decays was set to 3. That means that after the first decay if three epochs pass without improvement, it gets triggered again. This repeats every three epochs until either the early stopping is triggered or the 100 epochs were complete.

All input features were transformed into their standard score, also known as z-score. This means that they are translated and scaled such that their mean is equal to zero and their standard deviation is equal to 1. This is often applied in machine learning to help with model training (Suthaharan, 2016). The equation to compute the z-score is $z = \frac{x-\mu}{\sigma}$, where μ is the mean of x, and σ is the standard deviation.

4.3.3 Evaluation methodology

Although the simulation task and forecast task were formulated such that the models are identical, the data they are using is not, so the two cannot be compared directly. The model selection will be performed separately for each data source, but the rain gauge and radar-based models will be compared together, and the NWP based model will be evaluated on its own. Section 4.3.3 explains what parameter is to pick the best model in section 4.3.3. Section 4.3.3 describes the model selection strategy and the hyper-parameters tested.

Choice of metrics

The choice of the metric by which models are selected can greatly affect the final results. It's important to make a distinction between three different concepts:

- Objective function
- Evaluation metric
- Real world benefit

The objective function is picked for computational properties such as being differentiable with respect to the model's coefficients. The evaluation metric is used to select between models. The distinction between the latter two metrics is made because the real world benefit is often hard to calculate or measure and so the evaluation metric can be used as an approximation. Ideally all three are mutually optimal but this is often hard to achieve or prove.

The main application drainage flow forecast considered in this project is early warning of potential overflow events. According to staff at the Veitur utility company in Reykjavík a warning at least 24 hours in advance is usually enough to prepare to make staffing and equipment preparations for such events. Since forecasts are updated every 6 hours the prediction distances used for evaluation will be those 24 to 30 hours into the future in order to include all observations of the target variable in the data set in the training and evaluation.

The models in this project are regression models but the task is one of classification. To convert the predictions into the desired binary classification form, a threshold is used such that if a model predicts the flow to be above the threshold, it's considered a positive prediction and otherwise a negative prediction. Binary classification models are usually evaluated based on the ratio of some combination of values from the confusion matrix shown in table 4.3

		Predicted		Total
		p	n	
Observed	p'	TP	FN	P'
	n'	FP	TN	N'
Total		P	N	T

Table 4.3: A confusion matrix for binary classification

The most straightforward classification metric is accuracy, which is the ratio of correct predictions to total predictions. The number of true negatives is not a strong indicator of performance for this application due to its relative abundance and low importance in the data. The critical success index shown in equation 4.1 is the accuracy when true negatives

are removed from the equation.

$$CSS = \frac{TP}{TP + FN + FP} \tag{4.1}$$

This metric is still biased towards common events since they will have more chance occurrences of true positives. A modified CSS known as the Gilbert skill score (GSS) in equation 4.2 takes into consideration the number of chance occurrences of true positives for a less biased score.

$$GSS = \frac{TP - TP_{random}}{TP + FN + FP - TP_{random}} \tag{4.2}$$

Where

$$TP_{random} = \frac{(TP + FN)(TP + FP)}{T} \tag{4.3}$$

The last missing criteria for the evaluation metric is the threshold selected to represent a positive observation. The higher the threshold the more important the event but also the less common the event, thus making evaluation of the model performance harder. Additionally the threshold will be different for each station since both the normal operating flow-rate and the maximum capacity of each pumping station will be different. Thus the threshold flow-rate is picked as a number within the following constraints:

- It is greater than the flow during all historical dry-weather conditions
- It is low enough that it occurs at least 100 times in all of the 5 cross-validation sets

The thresholds selected are 500 and 90 litres per second for Gelgjugangi and Boðagrandi respectively.

Table 4.4: Number of samples above the desired threshold for both pumping stations

Fold	No. of samples above threshold	
	Gelgjutangi (>500 l/s)	Boðagrandi (>90 l/s)
1	189	204
2	183	291
3	186	212
4	146	150
5	118	100

In order to turn the continuous predictions from the regression models into classification predictions, the following approach is used. First, to compute the correct classification answer c_t from a continuous target variable, y, the equation.

$$c_t = \text{class}(y, \text{t}) = \begin{cases} n & y < t \\ p & y > t \end{cases}$$

is used with the thresholds $t = 500$ and $t = 90$ for Gelgjutangi and Boðagrandi respectively. For predicted classes c'_t computed from continuous predictions, y', we also write:

$$c'_t = \text{class}(y', \text{t}) = \begin{cases} n' & y < t \\ p' & y > t \end{cases}$$

TP, FN, FP and TN can be computed from table 4.3 as described above. However, since the models are optimized initially to minimize a mean squared error function, it may be useful to compute the predicted classes $c'_{\hat{t}}$ with a threshold $\hat{t} \neq t$. In order to estimate $\hat{t}$ we simply try all predicted values for the training data set $\hat{y}_{train}$ and see which one minimizes the GSS. To reduce computational load at the risk of a slightly lower GSS, we round all predicted values to whole numbers and then use all unique numbers as thresholds, although the predictions are not rounded when computing the GSS or in any other evaluation only when selecting an optimal threshold. Then finally, GSS is computed using the threshold $\hat{t}$ that minimizes the GSS in the training set.

Model selection

The objective of model selection can be thought of as optimizing the model performance with respect to model parameters that can't be optimized by the same means as the model coefficients, and these coefficients are hyper-parameters. A grid search is when all possible permutations of the different hyper-parameters are tested separately. This often reveals trends in model performance when the performance is plotted against the selected parameters.

In machine learning, a hyperparameter is a parameter whose value cannot be optimized during training which, in contrast to coefficients (also often called parameters), but is often estimated with searching algorithms. A common practice in machine learning is to use a grid searching algorithm that tries every possible combination of hyperparameters from a set of hyperparameters, and the best one is selected at the end. The hyperparameters selected for this experiment were chosen with two primary considerations: to get the best performance possible with limited computational resources and assess the effect some of these hyperparameters have on the model results. The hyper-parameters used in the model selection are found in table 4.5

Hyper-parameter	**Values**
Data source	(Rain gauge, Radar, NWP)
No. of convolutional neurons	(0,1,5)
No. of LSTM neurons	(0,1,5)
Length of sequences	(1,12,24,48)
L2 Regularization	(0,0.1)
Weight scheme	(a,b)

Table 4.5: Overview of all hyper-parameters used in the model selection. The results with the NWP are evaluated separately from the rain gauge and radar models.

Cross-validation For both experiments, data from 2015 to the end of 2019 was used for training and validation. The data was split into 5 training and validation sets using a variation of K-fold cross-validation implemented in scikit-learn knowns as TimeSeriesSplit (Buitinck et al., 2013). The split is used explicitly for time series data to ensure that training data never contains data that occurs after the validation data and each consecutive validation set has higher indices than all those before it. Each validation set is about 1/6th of the total data. The first training data set is also about 1/6th of the data and then each consecutive training data set contains all previous training and validation data. Each model was initiated using the same random seed to ensure repeatability.

5 Results and discussion

this chapter aims to present the results from the evaluations and experiments described in chapter 4 and answer the research questions presented at the beginning. First, the data evaluation results are presented. Then the results of the model selection described in 4.3 are presented, showing both the best performing models as well as an overview of model performance grouped by different hyper-parameters. Lastly, the performance of the best models is evaluated on otherwise untouched data to get a realistic estimate of performance.

The goal of these results is to provide answers the the reasearch questions posed in the introduction, which as the following:

1. How accurately can urban flow be simulated and forecasted in Reykjavík, Iceland, with currently available data and simple machine learning methods.

2. How can simple data-driven simulations and forecasts be used to improve operations of the wastewater system in Reykjavík with currently available data?

3. What are the main improvements to the data or methods that would be required for better performance and additional applications of the simulations and forecasts?

5.1 Data evaluation results

This section will look at the results of the data evaluation described in section 4.2. The first data evaluated will be the radar data, which was processed from raw radar reflectivity values into rainfall estimates. The second data source evaluated is the numerical weather predictions for rainfall, which was provided by the IMO and used in section 4.3 as input for a drainage flow forecast. The next data source is the drainage flow from the pumping stations, which was used as the target variable for both the simulation and the forecasting experiments in section 4.3.

5.1.1 Radar

When radar data is used directly to estimate rainfall, there are a few variables that can affect the quality of the rainfall estimates. Among these are the altitude and the coefficients in the reflectivity to rainfall conversion. It is thought that Without proper adjustment, radar reflectivity can give very inaccurate estimates of rainfall. One of the prerequisites of radar rain gauge adjustments is a sufficiently dense network of high quality rain gauges, which Reykjavík does not have. Until better radar estimated rainfall is available, it is still useful know how accurate the rainfall estimates from the radar are as well as how the data might be improved.

In table 5.1, the results from section 4.2.2 are presented. The table uses two metrics, mean squared error (MSE) and the Pearson product-moment correlation coefficient (correlation), to evaluate the radar estimated rainfall. For this experiment, the estimated rainfall from the radar comes from the pCAPPI pixel closest to each rain gauge. A low MSE is better, and a high correlation is better. Highlighted in each row is the best result for each sensor. Each row represents a particular combination of altitude and period and then for each of those, there is one row for the default Marshall-Palmer relation for the Z-R relationship, $a = 200, b = 1.6$ and one where a and b are the combination that minimizes the mean square error.

The best results for every sensor for both metrics were from the row with the 1.25km CAPPI data for the second period and using the MSE estimated Z-R coefficients. If we only consider the results from the default Z-R coefficients, the best results for both metrics

and all sensors are also found for the latter period at the lower altitude, with the exception of the correlation for sensor 1481, which was better at 2km for that period.

Since the main difference between those periods was the radar strategy, which for the latter period included a 0.9-degree beam between the 1.3-degree beams and 0.5-degree beams. One possible explanation is that this 0.9-degree beam was low enough to get interference from the ground but high enough to be included in the interpolation of the pCAPPI for the earlier part of the data set.

Table 5.1: Radar data evaluation results. The best scores in each column are marked with **bold face** and the best scores using the Marshall-Palmer relation for the Z-R coefficients ($a = 200, b = 1.6$) is marked with an underline. Each cell contains the score of the radar estimated rainfall from a pCAPPI data set created from the raw radar scans at the altitude specified in the leftmost column. The rainfall estimate is computed with the Z-R coefficients specified in the third column, and the reflectivity values come from the pCAPPI pixel closest to the sensor. The pCAPPI value is evaluated over the whole period (Jan15-Dec19) as well as individually for the two periods that had different radar strategies.

		metric	**MSE**			**Correlation**		
Altitude	period	Z-R coefficients	1473	1475	1481	1473	1475	1481
1.25km	Jan15-Dec19	a=100, b=2.5	.0537	.0899	.1116	.5825	.5958	.6338
1.25km	Jan15-Dec19	a=200, b=1.6	.0616	.0891	.1196	.5366	.5937	.5927
1.25km	Jan15-Mar18	a=117, b=2.62	.0558	.0984	.1201	.5716	.5722	.6187
1.25km	Jan15-Mar18	a=200, b=1.6	.0673	.1024	.1247	.5254	.5673	.5862
1.25km	Mar18-Dec19	a= 79,b=2.25	**.0501**	**.0608**	**.0951**	**.6096**	**.7209**	**.6740**
1.25km	Mar18-Dec19	a=200, b=1.6	.0507	.0665	.1111	.5767	.6993	.6091
2km	Jan15-Dec19	a=100, b=2.5	.0582	.0809	.1245	.5485	.6351	.5700
2km	Jan15-Dec19	a=200, b=1.6	.0615	.0896	.1425	.4921	.6067	.4981
2km	Jan15-Mar18	a=117, b=2.62	.0608	.0884	.1349	.5410	.6119	.5338
2km	Jan15-Mar18	a=200, b=1.6	.0650	.0971	.1552	.4753	.5882	.4397
2km	Mar18-Dec19	a= 79,b=2.25	.0508	.0699	.1035	.5803	.6761	.6455
2km	Mar18-Dec19	a=200, b=1.6	.0545	.0766	.1206	.5621	.6456	.6172

5.1.2 Numerical weather prediction

Numerical weather predictions (NWP) are the only available rainfall forecast for this project. The NWP is evaluated with the same metrics as the radar data. It is possible to evaluate the performance of the NWP with respect to the distance of the forecast. The correlation and MSE of the NWP can be seen on figures 5.1 and 5.2. In both plots, the performance of a persistence model is also shown for each rain gauge. Interestingly the performance of the persistence model outperforms the NWP for roughly 3 hours. Comparing the correlation of the NWP with the radar estimated rainfall reveals that the two are roughly similar, which should not be expected since one is an observation and the other a forecast. This highlights just how bad the performance of the radar estimated rainfall is when not adjusted with rain gauge observations.

The MSE plots in figure 5.2 reveal an unfortunate result for the NWP, showing at least in one case (gauge 1473) of three that a persistence model performs about as well as the NWP. However, it is known that NWP models often have considerable bias both in location and in forecast distance and in hydrological research, an adjustment is often performed on the NWP data in order to correct for this. A simple way to adjust the data is by fitting a linear regression model with each forecast distance and each gauge individually. E.g. all the estimates for gauge 1475 at 24 hours are used as input and all the observations for gauge 1475 for the corresponding times are used as outputs. Then the adjusted prediction

can be compared again with the rain gauge observation. There is no training and test set here, but each distance for each meter has more than 5500 samples in this data set, so while this does not give conclusive results for the real world expected performance, it shows how important the adjustment can be. The MSE comparison between the unadjusted and adjusted NWP predictions is seen in figure 5.3

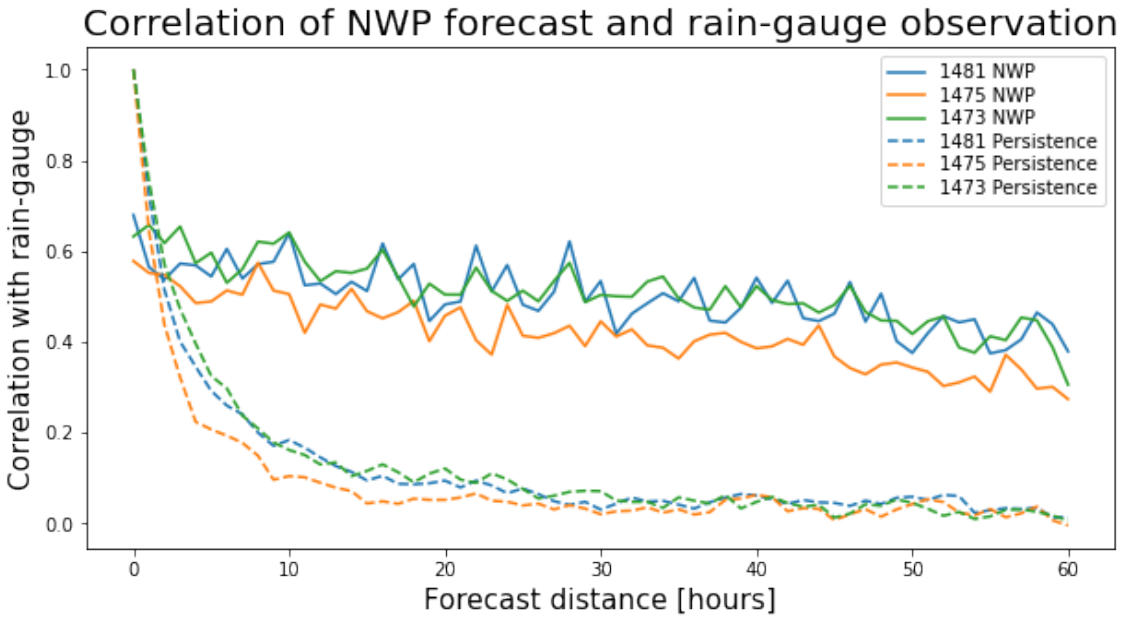

Figure 5.1: Correlation of NWP rainfall forecast for predicting the later observed rainfall from rain gauges. Also shown is the Correlation of a persistence forecast

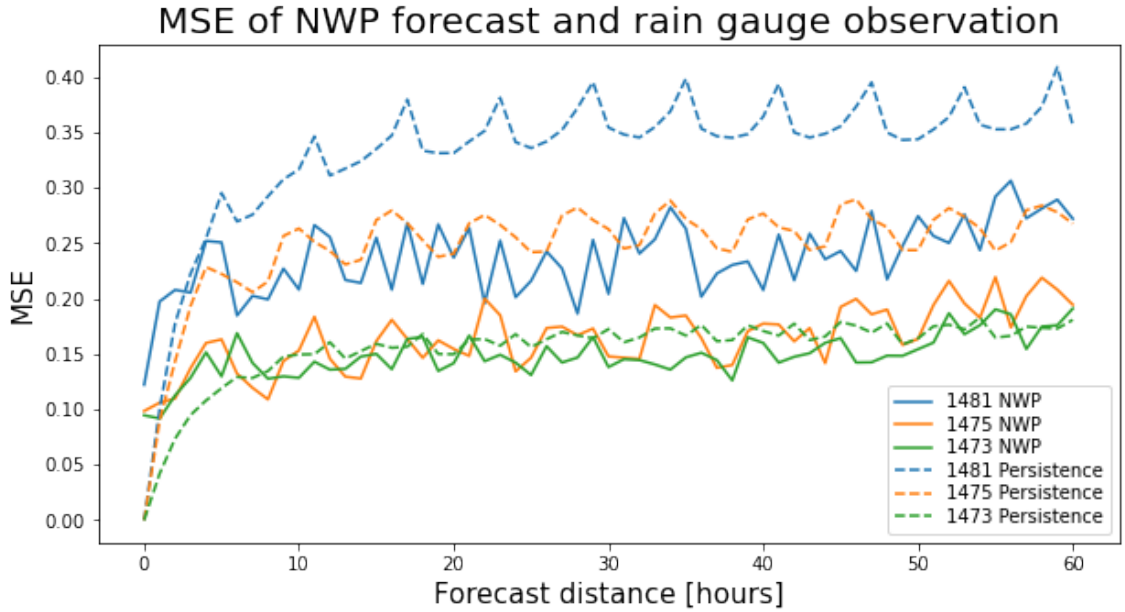

Figure 5.2: Mean squared error of NWP rainfall forecast for predicting the later observed rainfall from rain gauges. Also shown is the MSE of a persistence forecast

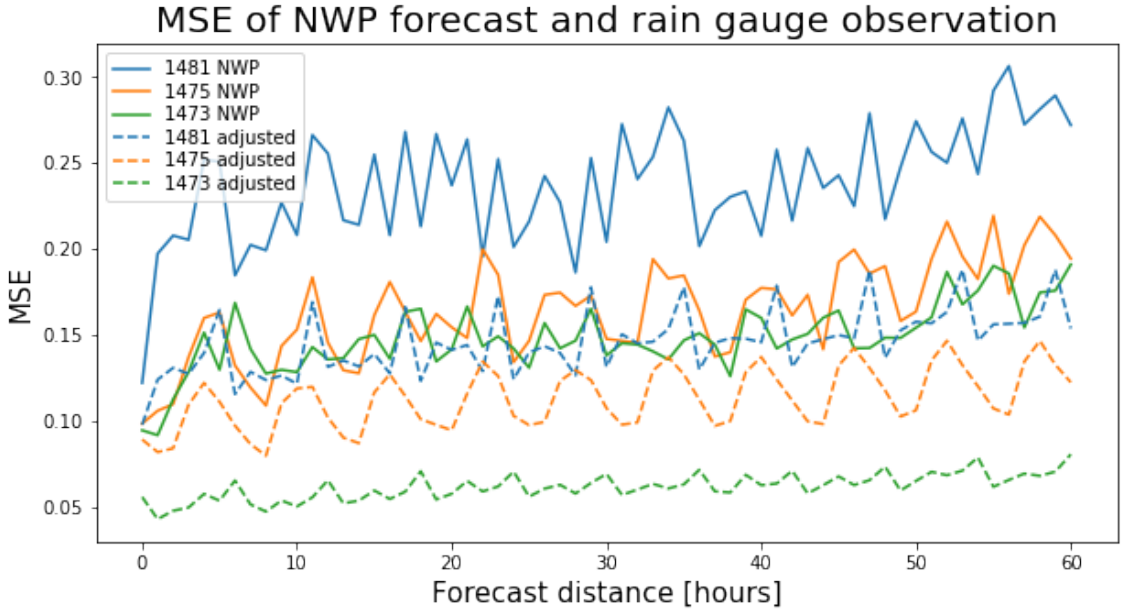

Figure 5.3: Mean squared error of NWP rainfall forecast for predicting the later observed rainfall from rain gauges. Also shown is the MSE of an adjusted NWP forecast

5.1.3 Drainage flow data

The drainage flow data was used as the target variable in the experiments described in section 4.3. For this evaluation, it's important to understand that for urban drainage applications, it's the inflow rather than the outflow of the pumping stations that is of interest. The data used in the experiments are from the outflow. The assumption that the inflow and outflow are the same breaks down for two reasons discussed in sections 4.2.3 and 3.3. One of the causes of this discrepancy is caused by the change in volume within the pumping station between the start and end of every hour. Because the historical data used in this paper is hourly outflow average, if the volume within the station changes during that hour, the inflow and outflow aren't equal. The resolution of the saved data from the water level and flow meters was increased from hourly to a value every minute from. This was done so that the water level at the start of the hour could be more accurately estimated.

Table 5.2: Statistics about the difference between the input and output at Boðagrandi and Gelgjutangi pumping stations. Overestimate refers here to the output being greater than the input.

Pumping station	Boðagrandi	Gelgjutangi
Start of period	2021-08-13 01:00:00	2021-08-13 01:00:00
End of period	2021-12-28 23:00:00	2021-12-28 23:00:00
Mean flow	61.84 [l/s]	193.44 [l/s]
Max flow	193.74 [l/s]	1145.41 [l/s]
Min flow	13.16 [l/s]	37.96 [l/s]
Mean absolute difference	1.88 [l/s]	12.12 [l/s]
Max overestimate	9.28 [l/s]	125.43 [l/s]
Max underestimate	8.79 [l/s]	148.26 [l/s]

The first part of the approach described in section 4.2.3 requires the area of the cross-section of the pumping station's wet well in order to estimate the volume of water inside

the station from the height. This approach assumes that the reconstruction of the inflow using the correct estimated area would be the smoothest of all reconstructions. For this, a metric for smoothness was defined as the sum of all changes in the reconstructed flow. See 4.2.3 for the equations used. Figures 5.4, and 5.5 show the smoothness metric for both stations as functions of the area used to make the reconstruction. Both plots show very clear maxima, and from these, we estimate the areas to be $300m^2$ and $30m^2$ for Gelgjutangi and Boðagrandi, respectively.

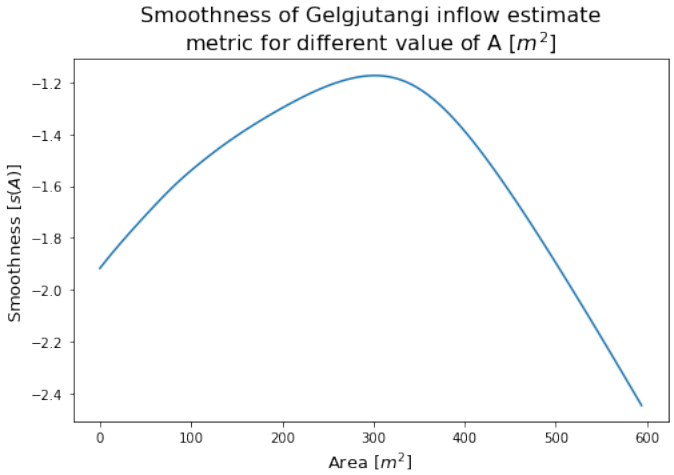

Figure 5.4: The smoothness of the estimated inflow for Gegjutangi as defined in 4.2 as a function of the area shown on the x-axis. The smoothness is assumed to be maximized when the area is equal to the true area of the chamber within the pumping station.

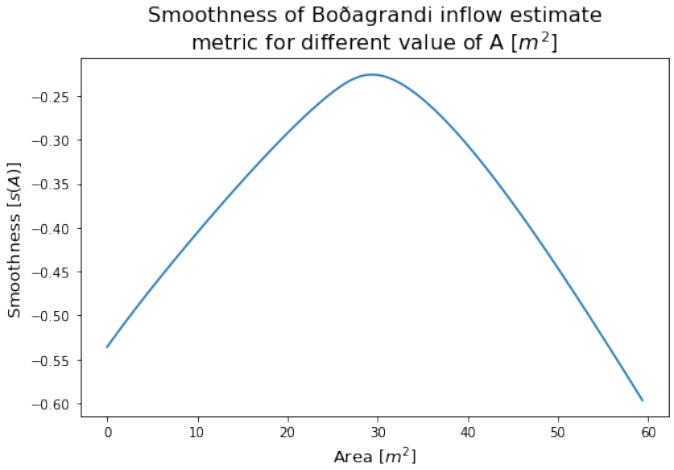

Figure 5.5: The smoothness of the estimated inflow for Boðagrandi as defined in 4.2 as a function of the area shown on the x-axis. The smoothness is assumed to be maximized when the area is equal to the true area of the chamber within the pumping station.

Next, the estimated areas are used to reconstruct the inflow for each hour as was described in 4.2.3. Table 5.2 shows some simple statistics about the flow in these stations during this time as well as some statistics about the difference between the inflow and outflow. Figures 5.6 and 5.7 show the estimated average hourly inflow and observed outflow for the 24 hours with the greatest total absolute error.

These results demonstrate how large this discrepancy can be, and in this analysis, the other sources of discrepancy were not considered since they cannot be evaluated. It has

not been explored how significantly this would affect drainage flow models like those made in this paper, but it's clear that any evaluation of the models that use the outflow data will be affected.

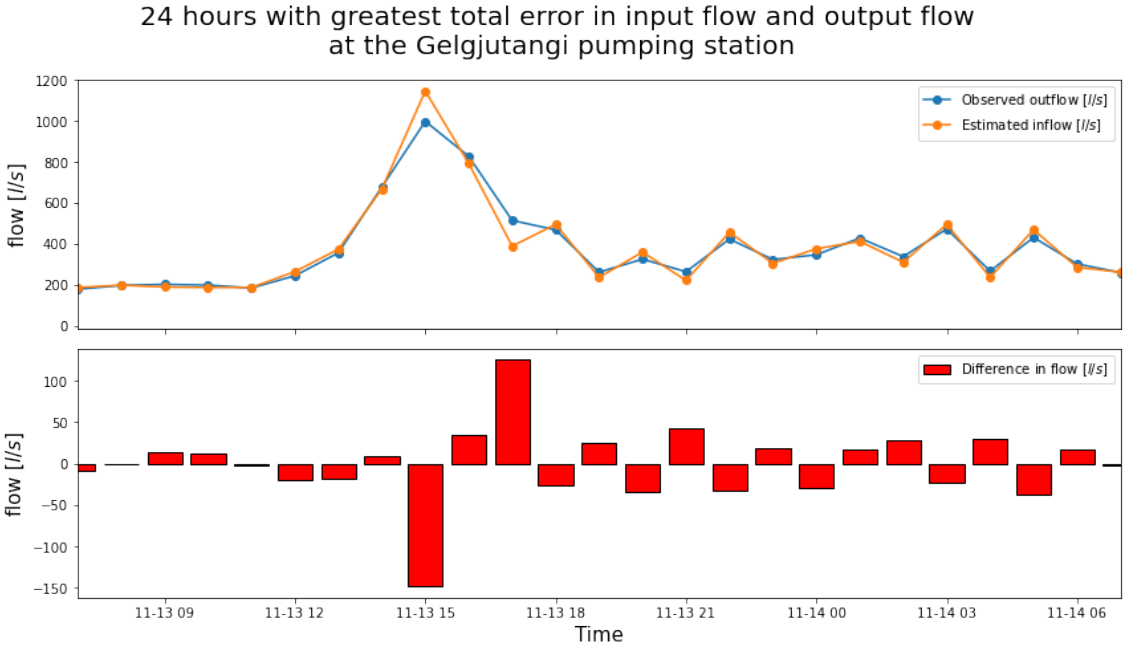

Figure 5.6: (top) the 24 hours with the greatest total error in the input flow and outflow at the Gelgjutangi pumping station. (Bottom) The difference between the estimated inflow and the observed outflow during this period

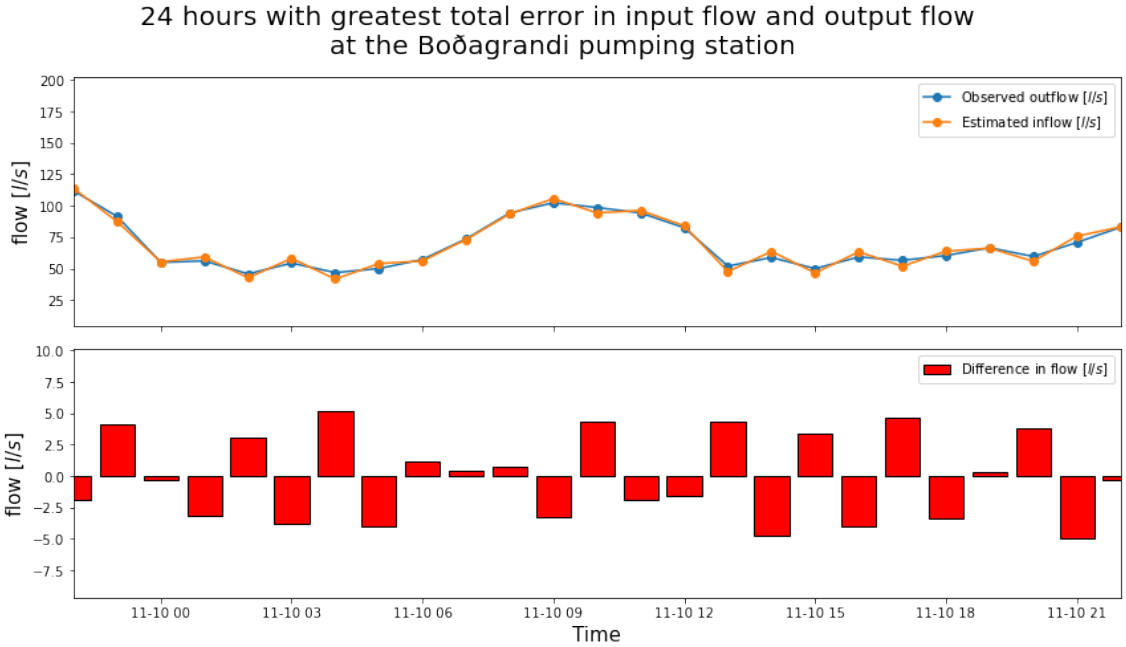

Figure 5.7: (top) the 24 hours with the greatest total error in the input flow and outflow at the Boðagrandi pumping station. (Bottom) The difference between the estimated inflow and the observed outflow during this period

5.2 Experiment results

This section will look at the results of the rainfall-runoff simulation experiments described in chapter 4.3. First, the drainage flow simulation results will be presented and discussed, beginning with the results of the model selection process and then looking at the model evaluations on new data. After that, the drainage flow forecast results will be presented, looking at the model selection and evaluation results.

Some of the hyperparameters included in the model selection in these following sections were specifically selected to evaluate the effects of certain factors relevant to the simulation and forecast performance that are not specific to the models themselves and some aspects of the model design were also intended to reveal these factors. The main factors of interest in this regard are: which data source is used as input and how many previous time steps of data are included. Another unanswered question that the experiments hope to answer relates to the quality of unadjusted radar estimated rainfall. It is known that radar produces a less accurate rainfall estimate when not adjusted with rain gauge observation. However, radar does have the benefit of more spatial information and information about rainfall in regions that may be too far from the rain gauge for it to provide a reliable estimate of rainfall there.

5.2.1 Drainage flow simulation results

This section begins by looking at the results of the model selection, which involves selecting the best model from many models with different hyperparameters. After that, this section looks at the model evaluation, which evaluates the performance of the best models from the model selection using new data to ensure an accurate assessment of the performance that can be expected of these models in the real world.

The drainage flow simulation models are rainfall-runoff models but also have to predict the flow due to other wastewater like sewage and, because of how Iceland's district heating system works, hot water used for district heating. These models use either rain gauge data or radar data as the rainfall input, but all use the outside temperature and the time of day in one-hot encoded form.

Model selection

Table 5.3: Hyper-parameters of the best performing models on both rain gauge and radar data. The individual GSS for both Gelgjutangi and Boðagrandi are included but the average GSS was used to select the best models

Data source	gauge	radar
Number of filters	5	5
Observations	48	24
Regularization	0.0	0.1
Number of LSTM units	0	0
Weight scheme	a	a
GSS average	0.373	0.235
GSS (Gelgjutangi)	0.424	0.259
GSS (Boðagrandi)	0.322	0.211

The results in the top right plot in figure 5.8 shows that the best results with radar data were notably worse than a large portion of the results with rain gauge data indicating that the added spatial information is not enough to counteract the worse accuracy, at least for these catchments that are relatively close to the rain gauge. This indicates that the less accurate rainfall estimate greatly outweighs the benefits of the added spatial information from the radar data.

There seems to be a pattern in the best performing models looking at the middle left plot in figure 5.8. The plot shows that the greater the length of the sequence, the better the performance of the best models. This indicates that past information about the system's state is relevant for determining the current flow.

The most apparent effect of any hyper parameter appears to be the number of LSTM neurons in the bottom right corner of figure 5.8. It appears that almost all but a few models with greater than 0.3 performance were models that did not include an LSTM layer. In addition to that, the non-LSTM layer appears to contain two clusters, one where all the models have a GSS above 0.3 and one where they have a GSS less than 0.25. Comparing those clusters to the data source cluster in the top left plot on figure 5.8, we can deduce that the subset of the model that used the rain gauge input and had no LSTM units accounts for all but two of the models with a greater than 0.3 GSS.

None of the results on the right side of figure 5.8 appears to show any interesting results. The hyperparameters of the best models for both the radar and rain gauge models, shown on table reffig:bestHyperSim, both include weight scheme a and five convolutional neurons.

Model evaluation

The objective of model evaluation is to take the best model from the model selection, retrain them and test them on previously unseen data. This is because, even though the cross-validation used in model selection is considered a good way to select between different models, it tends to produce optimistic performance estimates. The best models were selected using the Gilbert skill score (GSS), see 4.3.3 for elaboration on this metric, but tables 5.4 and 5.5 also include four more metrics: mean square error (MSE), which was the loss function used during the training of the models, root mean squared error (RMSE), a common metric that shows the average deviation of the predicted values, Pearson product-moment correlation (Correlation), a measure of the strength of a linear association between two variables (-1 for perfect negative linear relationship, 0 for no linear relationship, 1 for perfect positive linear relationship), Nash–Sutcliffe model efficiency coefficient (NSE), a common metric in hydrology used to assess the predictive skill of rainfall-runoff models.

The results of tables 5.4 and 5.5 both show similar differences in performance between the training and test data as well as between Gelgjutangi and Boðagrandi. According to every single metric, the best performing model that uses rain gauge data outperforms the best performing model that uses radar data, supporting the results of the hyper-parameter tuning about the performance difference between the data sources.

It should be noted before a comparison is made with drainage flow models in other papers that many other papers use the last known value of the drainage flow as input into the model. While this difference is very likely to improve the performance of the prediction, it would likely not have produced the same insights about the data and other modelling aspects. To demonstrate that including the previous value allows for much better results, see table 5.6, which shows the same evaluation metrics applied to a persistence model, which is a model that predicts the current output always to be equal to the last observed output. Lastly, table 5.7 shows the performance of a linear regression model that used the predictions from the best rain gauge model and the last available observation as input to predict the drainage flow. We see again that all metrics are improved over both the persistence and the best rain gauge models, except the GSS for Boðagrandi. This model may also be thought of as an ensemble of the best rain gauge and persistence models.

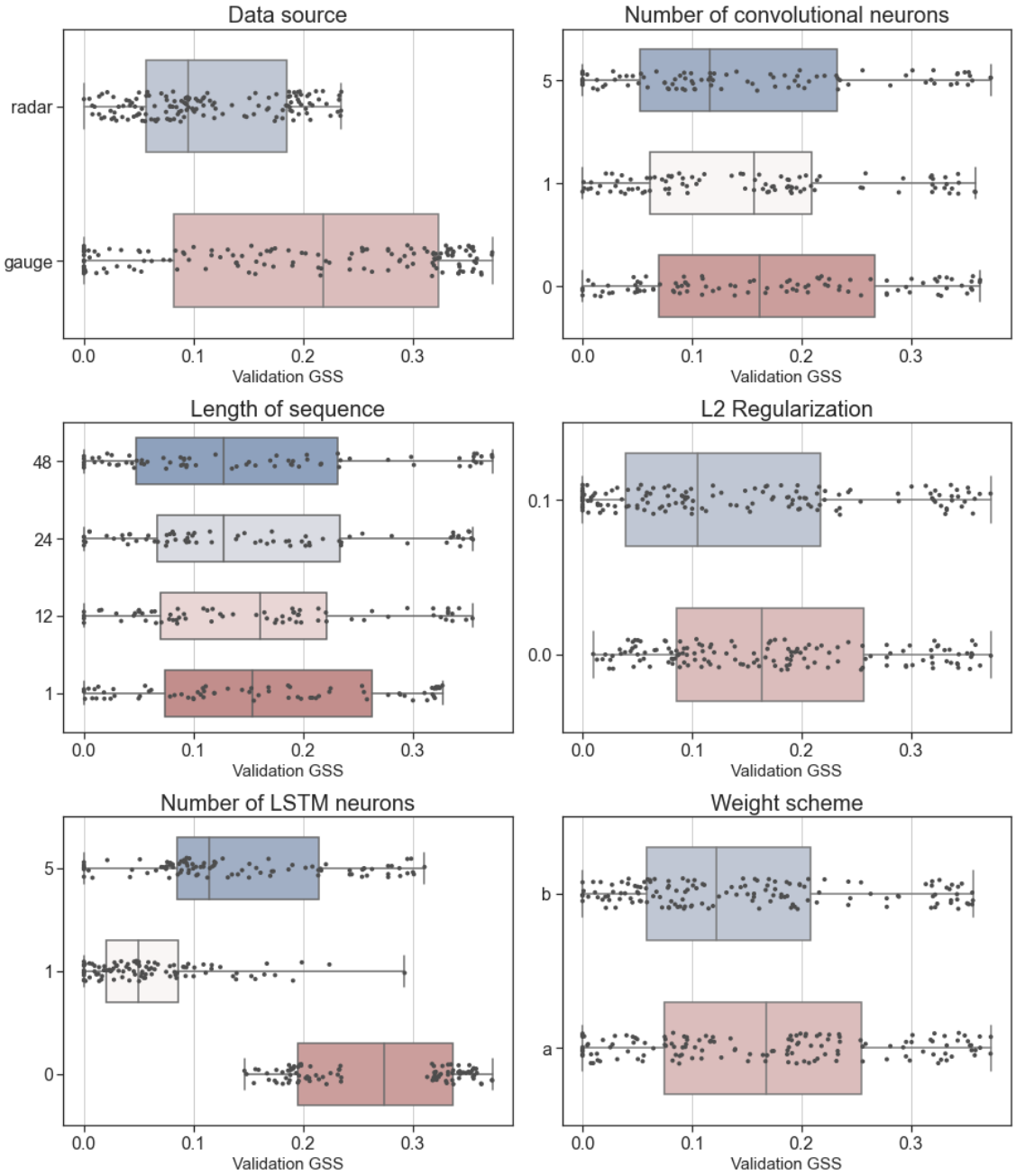

Figure 5.8: Each of the six plots shows the results of every single model trained with rain gauge and radar data for the task of simulating drainage flow. The x-axis is the GSS (see 4.3.3 for a description of this metric). Within each plot are 2-4 horizontal box and whisker plots with jitter along the y-axis to better show the density of the samples used to generate the box and whiskers. Each black dot represents a single model with a permutation of hyperparameters that included that hyperparameter in the value shown on the label on the left of that particular box and whiskers plot. For example, in the top-right plot, the upper box includes all models that used radar data and the lower box all models that used rain gauge data

Table 5.4: Training and testing results for both Gelgjutangi and Boðagrandi for the best performing model on rain gauge data

	Test		Train	
	GEL	BOD	GEL	BOD
MSE	4131.468	321.550	5042.482	256.401
RMSE	64.276	17.932	71.010	16.013
Correlation	0.783	0.588	0.793	0.701
NSE	0.611	0.316	0.629	0.489
GSS	0.351	0.222	0.444	0.245

Table 5.5: Training and testing results for both Gelgjutangi and Boðagrandi for the best performing model on radar data

	Test		Train	
	GEL	BOD	GEL	BOD
MSE	6098.810	351.138	7021.803	278.148
RMSE	78.095	18.739	83.796	16.678
Correlation	0.654	0.520	0.693	0.670
NSE	0.427	0.250	0.480	0.446
GSS	0.263	0.153	0.307	0.194

Table 5.6: Training and testing results for both Gelgjutangi and Boðagrandi for the persistence model

	Test		Train	
	GEL	BOD	GEL	BOD
MSE	3394.595	182.180	4295.448	138.064
RMSE	58.263	13.497	65.540	11.750
Correlation	0.840	0.806	0.842	0.862
NSE	0.681	0.613	0.684	0.725
GSS	0.461	0.410	0.519	0.516

Table 5.7: Training and testing results for both Gelgjutangi and Boðagrandi for a linear regression model that uses predictions from the best rain gauge model and the last observation as input to predict drainage flow

	Test		Train	
	GEL	BOD	GEL	BOD
MSE	2740.937	159.592	3364.708	119.723
RMSE	52.354	12.633	58.006	10.942
Correlation	0.862	0.814	0.867	0.873
NSE	0.742	0.661	0.752	0.762
GSS	0.485	0.399	0.539	0.529

One important aspect of model evaluation often forgotten is to visualize the data. This sometimes reveals faulty models and can point to issues that statistical metrics would not

otherwise reveal. Figures 5.9 and 5.10 show the observed and simulated flows from the best performing rain gauge model on the validation data. Figures 5.11 and 5.12 show the same for the best performing radar model. Each of these figures shows first the whole of 2020, aside from some windows where part of the input data was missing. These plots, specifically the plots that show Boðagrandi reveal a big change in the behaviour of the flow data starting right at the end of March. This change can be seen as a sudden increase in variance, visible by the thicker blue margin around the orange lines in figures 5.10. 5.12.

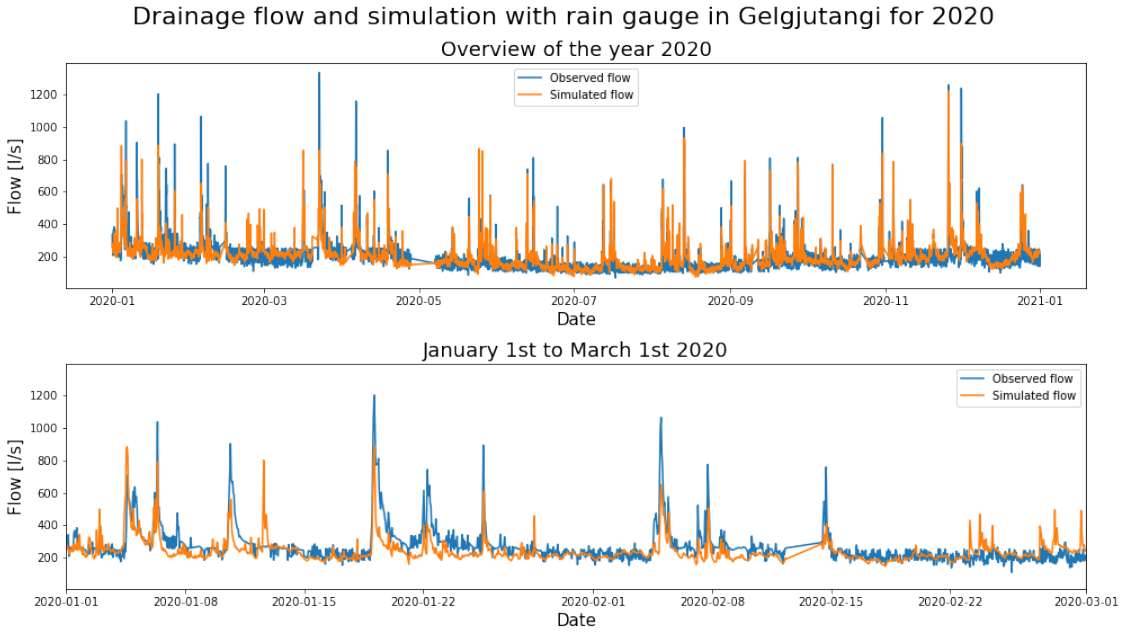

Figure 5.9: The observed and simulated flows from the best performing rain gauge model on the validation data for Gelgjutangi

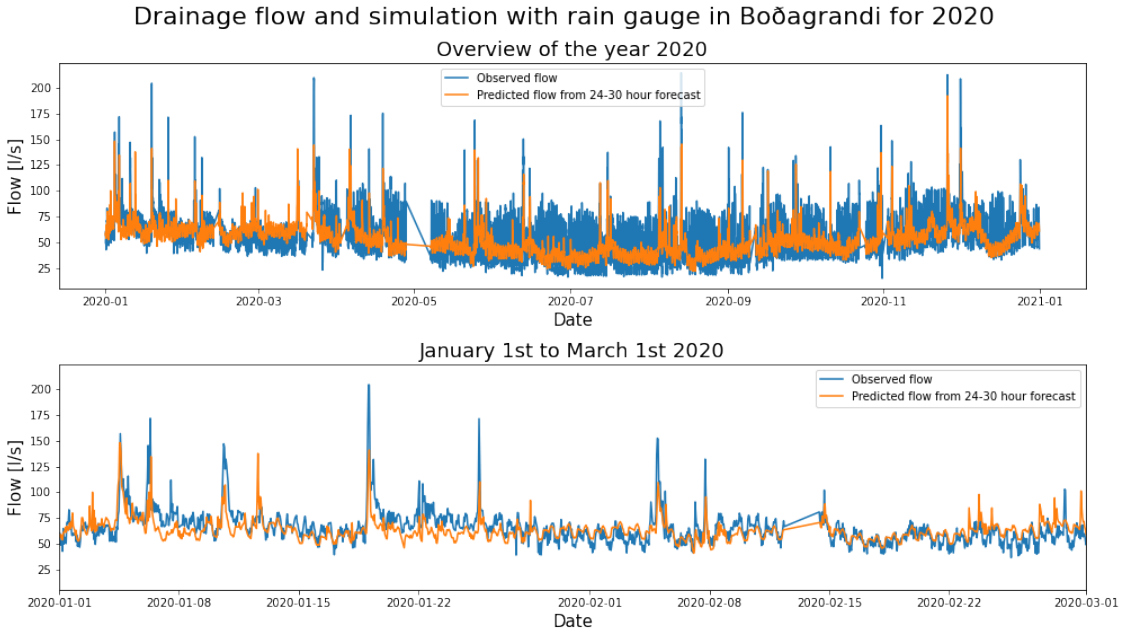

Figure 5.10: The observed and simulated flows from the best performing rain gauge model on the validation data for Boðagrandi

After consulting with experts in wastewater systems at Veitur, no probable cause for this change was found. Several ideas, including the installation of new pumps, construction in the area where water was being pumped from the pits or new residential buildings in the area. The installation of new pumps did not coincide with the start of the changed behaviour. No other theory was explored in more depth, so the cause for this change remains unknown.

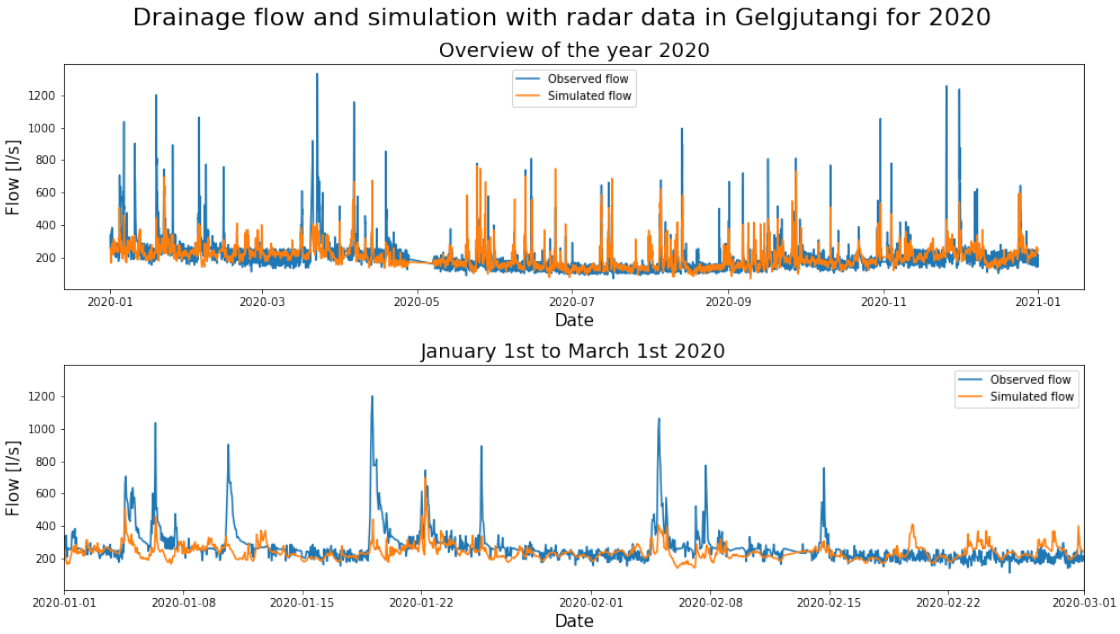

Figure 5.11: The observed and simulated flows from the best performing radar model on the validation data for Gelgjutangi

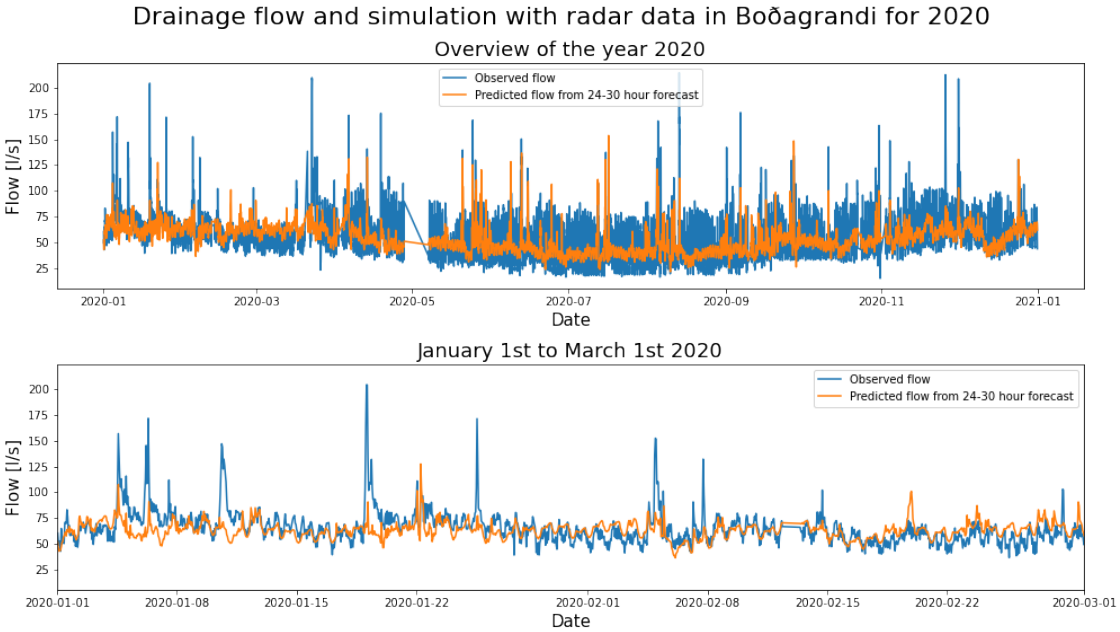

Figure 5.12: The observed and simulated flows from the best performing radar model on the validation data for Boðagrandi

5.2.2 Drainage flow forecasting results

This section mirrors the last section, beginning with an overview of model selection results and then looking at the model evaluation results. As noted in section 4.3.3 the performance of the models in the model selection was evaluated on the performance of predictions between 24 and 30 hours, the minimum time needed for most preparations concerning rainfall.

Model selection

The bottom left plot in 5.13 reveals a similar pattern as was seen in the simulation results. Almost all the best performing models do not include an LSTM layer. Looking then at the middle left plot in 5.13 we can infer that almost all the rightmost points belong to the non-LSTM models and we again see the best performing models improve in performance as the length of the sequence is greater. Similar results are seen for the parameters in the right column, where there are no notable results or patterns.

Table 5.8: Hyper-parameters of the best performing model. The individual GSS for both Gelgjutangi and Boðagrandi are included, but the average GSS was used to select the best models

Data source	NWP
Number of filters	1
Observations	48
Regularization	0.1
Number of LSTM units	0
Weight scheme	a
Validation GSS average	0.245
GSS (Gelgjutangi)	0.266
GSS (Boðagrandi)	0.224

Model evaluation

The drainage flow forecast model was evaluated on the 2020 data using the 24-30 hour forecasts. The same statistical metrics as in the simulation results are used in the evaluation in table 5.9. The same statistics for the persistence model can also be seen in table 5.10. Here we see a difference between the simulation results and the forecast results in that the forecast greatly outperforms the persistence model, at least when looking at the 24-30 hour forecast.

Since only the 24 to 30-hour forecast was used in the model selection, it has not been seen how well the forecast model performs for other prediction distances, and so for this evaluation, the best model from the model selection is used, and the performance for each metric is plotted in figure 5.16. Here we see that the persistence model performs better than the drainage flow model for up to around 3 hours, but after that, the performance of the forecast greatly outperforms the persistence model in all metrics. We also see how the drainage flow forecast metrics slightly decline over time, similarly to the NWP accuracy evaluated in 5.1.2.

Table 5.9: Performance of the best drainage flow forecast model on the test data set, which includes data from the year 2020.

	Test		Train	
	GEL	BOD	GEL	BOD
MSE	6199.454	352.819	7267.361	181.472
RMSE	78.737	18.783	85.249	13.471
Correlation	0.679	0.541	0.691	0.717
NSE	0.461	0.257	0.476	0.514
GSS	0.244	0.157	0.308	0.249

Table 5.10: Performance of a persistence model for a baseline in performance

	Test		Train	
	GEL	BOD	GEL	BOD
MSE	17093.591	595.516	18651.983	347.204
RMSE	130.742	24.403	136.572	18.633
Correlation	0.265	0.377	0.309	0.531
NSE	-0.486	-0.254	-0.345	0.070
GSS	0.049	0.046	0.078	0.073

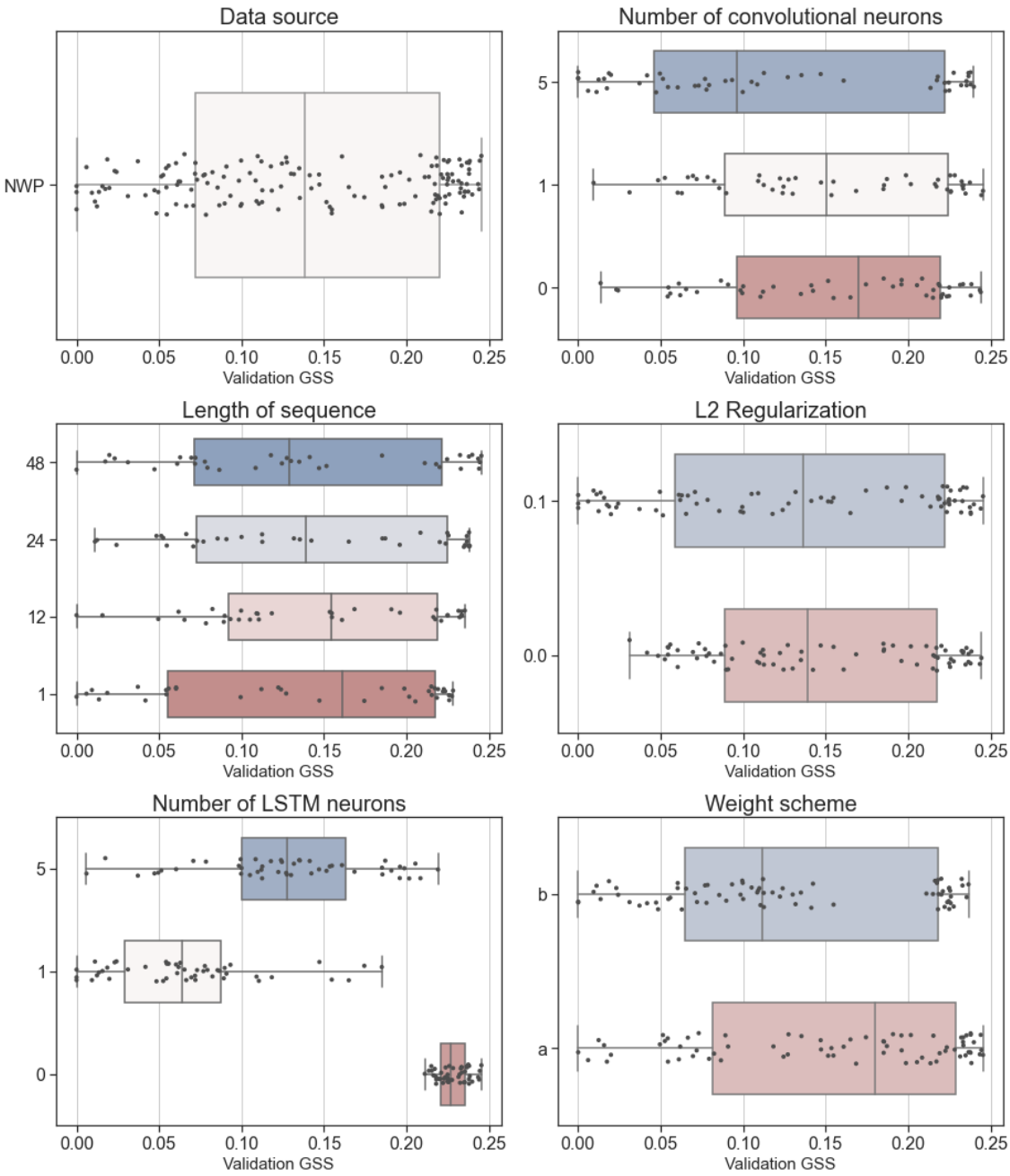

Figure 5.13: Each of the six plots shows the results of every single model trained with numerical weather prediction data for the task of forecasting drainage flow. The x-axis is the GSS for forecasts between 24 and 30 hours into the future (see 4.3.3 for a description of this metric). Within each plot are 2-4 horizontal box and whisker plots with jitter along the y-axis to better show the density of the samples used to generate the box and whiskers. Each black dot represents a single model with a permutation of hyperparameters that included that hyperparameter in the value shown on the label on the left of that particular box and whiskers plot.

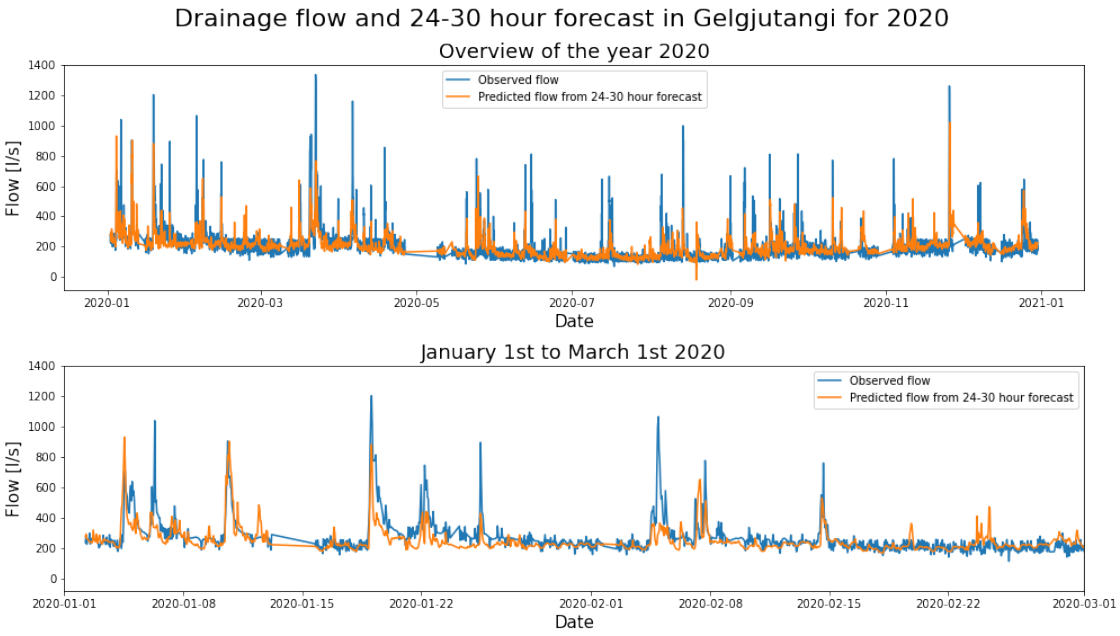

Figure 5.14: The observed and forecasted flows from the best performing rain gauge model on the validation data for Gelgjutangi

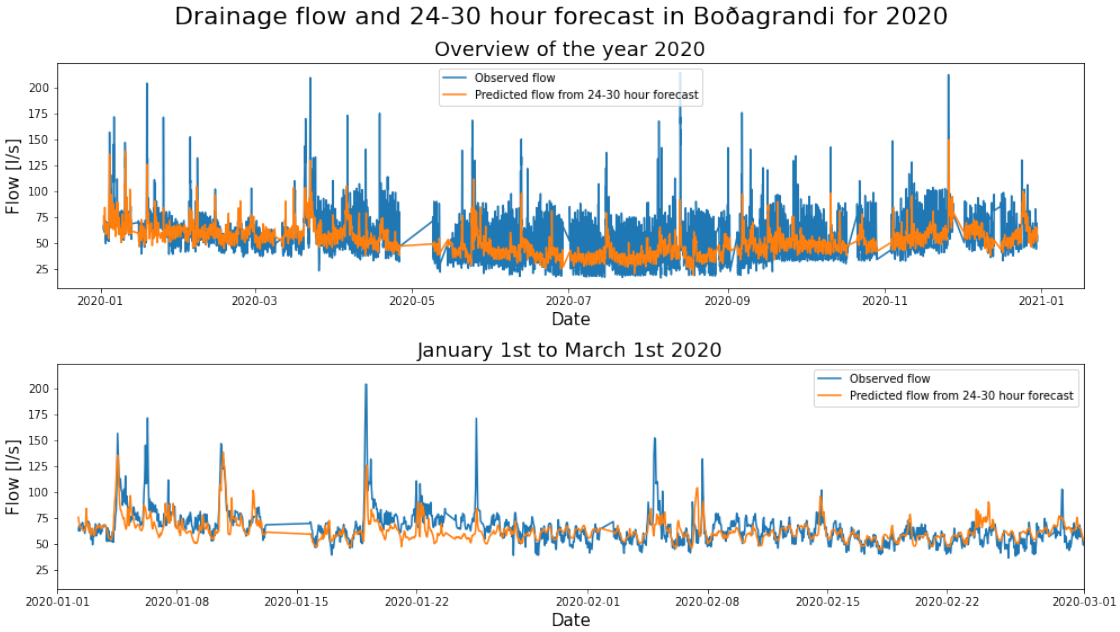

Figure 5.15: The observed and forecasted flows from the best performing rain gauge model on the validation data for Boðagrandi

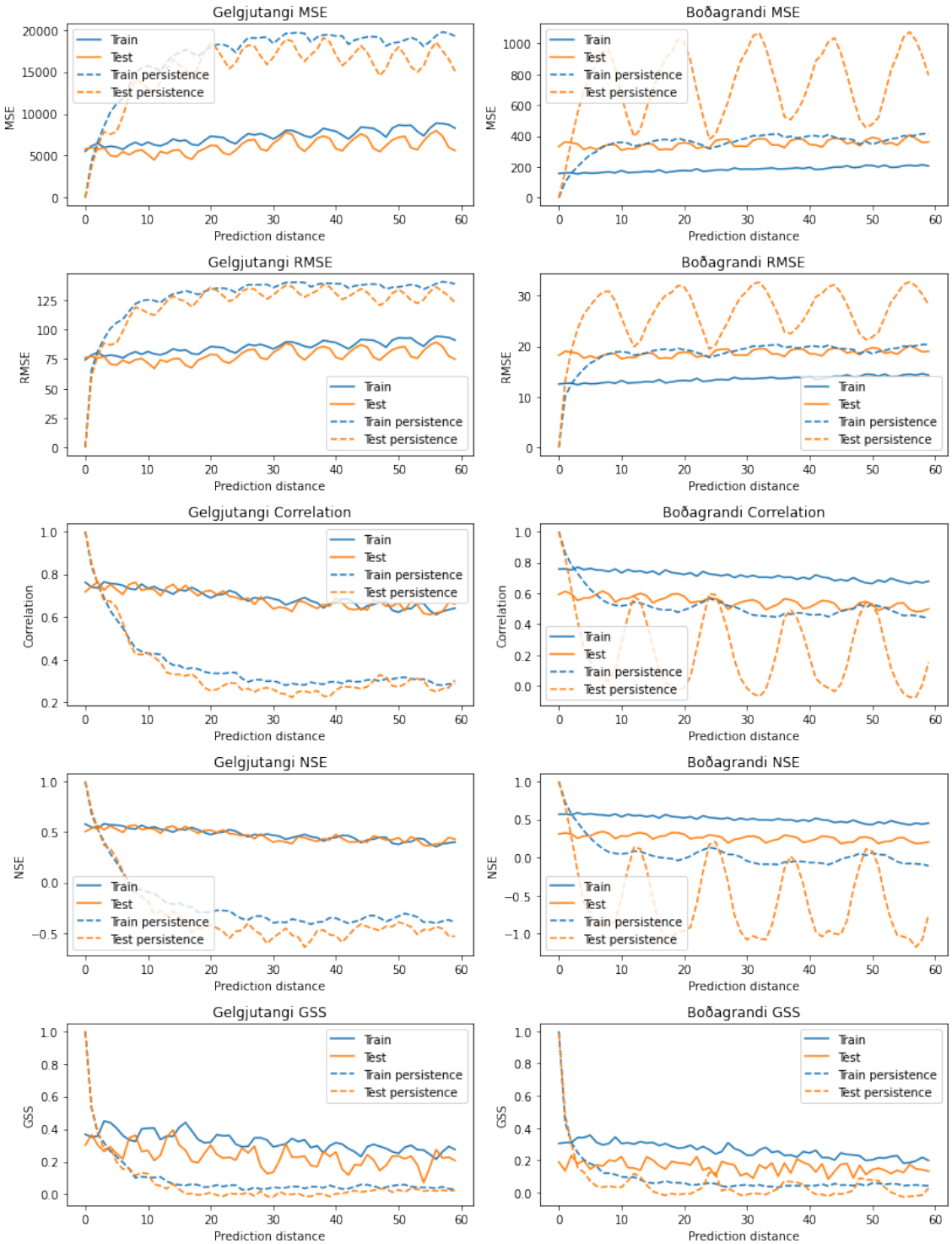

Figure 5.16: A comparison between the drainage flow forecast and a persistence model for Gelgjutangi (left) and Boðagrandi (Right) over the whole forecast horizon of 60 hours

6 Conclusions

The objective of this thesis was to design, implement and evaluate a simple data-driven drainage flow simulation and forecasting model for use in the Reykjavík wastewater and drainage system as well as to assess the data available for this task. The case study provided a qualitative assessment of the Reykjavík wastewater system and identified the most relevant available data sources. The experiments conducted showed how accurately drainage flow could be predicted using every rainfall data and rainfall forecast available in Reykjavík. The key applications for these simulations and forecasts were established, and several critical obstacles were discovered, analyzed and possible solutions identified.

As the name suggests, the most important variables in rainfall-runoff modelling, especially in the case of data-driven models, are the rainfall data and the runoff data i.e. combined wastewater flow. Since rain gauges are considered the most reliable data source, they were used to evaluate the accuracy of the other two rainfall data sources, radar and NWP. The evaluation of the radar data compared rainfall estimated at different altitudes and the effect of different radar strategies. It was found that the estimates of rainfall produced at a lower altitude are better but also more sensitive to interference from the ground. Rainfall estimates from the NWP were only slightly worse than that of the radar, confirming the findings of other research that highlighted the importance of radar-rain gauge merging.

Two different problems were identified in the drainage flow data, one of which is caused by a buffer between the inflow and outflow from the pumping stations. This discrepancy was investigated and found to be the cause of both underestimation and underestimation of the drainage flow. This can affect the training and evaluation of any simulation and forecasting model that uses the data.

Three neural network layers were tested in four combinations: a temporal convolutional layer, a long short-term memory layer, and a fully-connected linear layer. The temporal convolutional layer followed by a fully-connected linear layer performed the best of all models on every data source. The models were trained on a regression task but were evaluated by their ability to correctly classify instances where the flow surpasses a threshold of interest.

It was found that giving the models a longer sequence of input data improved the performance quite consistently, and it was found that models that use data from a single rain gauge outperform those that use radar estimated rainfall. The performance of the drainage flow forecast was similar or slightly better than the best model simulating drainage flow with radar data, thus highlighting the need for better methods for estimating rainfall from radar data.

The following is a list of obstacles that need to be overcome and recommendations to improve the performance of the drainage simulations and forecasts:

- The resolution of the drainage flow sensors should be increased significantly to allow for future applications like real-time control and the reconstruction of pumping station inflow. It could also open up more opportunities for offline applications. The exact resolution should be determined with experimentation and further study.
- The measured water level within the pumping stations should be logged at the start of every flow measurement interval to improve reconstructed inflow.
- Veitur needs to invest in better control mechanisms in the drainage system to utilize the drainage flow forecast for preemptive action. These could include increasing buffers and adding variable speed pumps.

- A radar nowcast is likely the only type of precipitation forecast suitable for real-time control applications.
- A good network of rain meters, as well as use radar-rain gauge merging, are needed to improve radar estimated rainfall.

A system that can operate using informed decisions can transition from reactive operations and damage mitigation to proactive operations and optimization. The project provides an overview of the challenges that need to be overcome so that Veitur can begin to modernize wastewater management in Reykjavík.

Bibliography

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938. https://doi.org/https://doi.org/10.1016/j.heliyon.2018.e00938

Abou Rjeily, Y., Abbas, O., Sadek, M., Shahrour, I., & Hage Chehade, F. (2017). Flood forecasting within urban drainage systems using narx neural network. *Water Science and Technology*, *76*(9), 2401–2412.

AghaKouchak, A., & Habib, E. (2010). Modeling radar rainfall estimation uncertainties: Random error model. *Journal of Hydrologic Engineering*, *15*(4), 265–274.

Beven, K. J. (2011). *Rainfall-runoff modelling: The primer*. John Wiley & Sons.

Britannica, E. o. E. (2020). Drainage basin. https://www.britannica.com/science/drainage-basin

Bruen, M., & Yang, J. (2006). Combined hydraulic and black-box models for flood forecasting in urban drainage systems. *Journal of Hydrologic Engineering*, *11*(6), 589–596.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: Experiences from the scikit-learn project. *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108–122.

Cole, S. J., & Moore, R. J. (2008). Hydrological modelling using raingauge- and radar-based estimators of areal rainfall. *Journal of Hydrology*, *358*(3), 159–181. https://doi.org/https://doi.org/10.1016/j.jhydrol.2008.05.025

Commons, W. (2013). *File:colored neural network.svg*. https://en.wikipedia.org/wiki/File:Colored_neural_network.svg

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

Dingman, S. L. (2015). *Physical hydrology*. Waveland press.

Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks—a review. *Pattern recognition*, *35*(10), 2279–2301.

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.-E., Quirmbach, M., Vaes, G., & Vieux, B. (2004). Towards a roadmap for use of radar rainfall data in urban drainage [Urban Hydrology]. *Journal of Hydrology*, *299*(3), 186–202. https://doi.org/https://doi.org/10.1016/j.jhydrol.2004.08.004

Fecarotta, O., Carravetta, A., Morani, M. C., & Padulano, R. (2018). Optimal pump scheduling for urban drainage under variable flow conditions. *Resources*, *7*, 73. https://doi.org/10.3390/resources7040073

Germann, U., & Joss, J. (2004). Operational measurement of precipitation in mountainous terrain. *Weather radar* (pp. 52–77). Springer.

Goodson, G., & Kopnitsky, A. (n.d.). Nimbostratus clouds: Precipitation layer [Accessed: 2021-12-09].

Goudenhoofdt, E., & Delobbe, L. (2009). Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrology and Earth System Sciences*, *13*(2), 195–203. https://doi.org/10.5194/hess-13-195-2009

Gunnlaugsson, E. (2004). Geothermal district heating in reykjavík, iceland.

Han, J., & Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. *International workshop on artificial neural networks*, 195–201.

Heistermann, M., Jacobi, S., & Pfaff, T. (2013). Technical note: An open source library for processing weather radar data (wradlib). *Hydrology and Earth System Sciences*, *17*(2), 863–871. https://doi.org/10.5194/hess-17-863-2013

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A*, *79*(8), 2554–8. https://doi.org/10.1073/pnas.79.8.2554

Íslands, V. (n.d.). *Harmonie - numerical weather prediction model*. https://en.vedur.is/weather/articles/nr/3232

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

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

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, … Xiaoqiang Zheng. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Software available from tensorflow.org]. https://www.tensorflow.org/

Ochoa-Rodriguez, S., Wang, L. .-P., Willems, P., & Onof, C. (2019). A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resources Research*, *55*(8), 6356–6391. https://doi.org/10.1029/2018wr023332

Pedersen, J. (2021). *Using numerical weather prediction and in-sewer sensor data for real-time monitoring and forecasting in urban drainage-wastewater systems* (Doctoral dissertation). Technical University of Denmark.

Phi, M. (2018). Illustrated guide to lstm's and gru's: A step by step explanation [Accessed: 2021-12-04].

She, L., & You, X.-y. (2019). A dynamic flow forecast model for urban drainage using the coupled artificial neural network. *Water Resources Management*, *33*(9), 3143–3153.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.

Sitterson, J., Knightes, C., Parmar, R., Wolfe, K., Avant, B., & Muche, M. (2018). An overview of rainfall-runoff model types.

Stránský, D., Bareš, V., & Fatka, P. (2007). The effect of rainfall measurement uncertainties on rainfall-runoff processes modeling. *Water science and technology : a journal of the International Association on Water Pollution Research*, *55*, 103–11. https://doi.org/10.2166/wst.2007.100

Suthaharan, S. (2016). Machine learning models and algorithms for big data classification. *Integr. Ser. Inf. Syst*, *36*, 1–12.

Tealab, A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, *3*(2), 334–340. https://doi.org/https://doi.org/10.1016/j.fcij.2018.10.003

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

Uijlenhoet, R., Van Der Wielen, S. H., & Berne, A. (2006). *Uncertainties in rainfall retrievals from ground-based weather radar: Overview, case study, and simulation experiment* (No. *4*). https://doi.org/10.5194/hessd-3-2385-2006

Vallabhaneni, S., Vieux, B. E., Donovan, S., & Moisio, S. (n.d.). Interpretation of radar and rain gauge measurements for sewer system modeling. *Global solutions for urban drainage* (pp. 1–10). https://doi.org/10.1061/40644(2002)231

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1989). Phoneme recognition using time-delay neural networks. *IEEE transactions on acoustics, speech, and signal processing*, *37*(3), 328–339.

Zell, A. (1994). *Simulation neuronaler netze* (Vol. 1). Addison-Wesley Bonn.

Technical
University of
Denmark

Richard Petersens Plads, Building 324
2800 Kgs. Lyngby
Tlf. 4525 1700

https://www.compute.dtu.dk/