UPPSALA
UNIVERSITET

UPTEC W 21012
Examensarbete 30 hp
Maj 2021

Using linear regression and neural network to forecast sewer flow from X-band radar data

Fredrik Wigertz

ABSTRACT

Using linear regression and neural network to forecast sewer flow from X-band radar data

Fredrik Wigertz

The climate adaptation of our cities and the optimization of our technical systems with regards to weather sets high demands on the availability and the processing of weather data. The possibility to forecast disturbances of influent flow rate to wastewater treatment plants allow control systems counteract these disturbances before they have a harmful effect on the treatment processes. These forecasts can be made by different models A neural network models complex patterns between different data sets through a multi-layered structure containing a large amount of transformation functions.

The aim of this project was to examine how the complex neural network performed compared with a simpler linear regression model when forecasting wastewater flow using high resolution X-band rain radar data. The study also investigated to what extent X-band rain radar data contributes to the performance of the model. The performance was evaluated at rain flow periods only.

Wastewater flow data were provided by Avedøre wastewater treatment plant in Copenhagen operated by BIOFOS. The X-band rain radar data was provided by HOFOR. The neural network was developed by Informetics on the TensorFlow platform.

This project concluded that the neural network and the linear regression model performed equally well at predicting when a rain flow period began. The neural network was more accurate at predicting the flow rate while the linear regression was better at approximating the accumulated flow over an entire rain flow period. Using additional rain data up to 30 km within the radar station location in comparison with using data only from within the catchment indicated a 20 to 30-minutes improvement of possible lead time. A conceivable lead time when forecasting the sewer flow to Avedøre wastewater treatment plant was estimated to be around 4 hours.

Keywords: Neural network, Linear regression, Flow forecasting, Wastewater, Wastewater treatment plant, Rainfall-runoff modelling, X-band radar.

D*epartment of Information Technology, Uppsala University (UU) Box 337 SE-75105 Uppsala, Sweden, ISSN 1401-5765*

REFERAT

Användning av linjär regression och neurala nätverk för att förutsäga avloppsflöde utifrån X-band radardata
Fredrik Wigertz

Det föreligger höga krav på tillgänglighet och bearbetning av väderdata för att kunna optimera tekniska system i förhållande till väder och klimat. Att kunna förutsäga ändrat inkommande flöde till avloppsreningsverk möjliggör för kontrollsystem att kunna motverka negativa konsekvenser på reningsprocesserna på grund av det ändrade flödet. X-band radardata kan användas för att prognoser av flöden med hjälp av olika modeller. Ett neuralt nätverk, reproducerar komplexa mönster mellan olika dataset genom en struktur med flera lager och en mängd överföringsfunktioner.

Målsättningen med det här projektet var att utvärdera hur ett komplext neuralt nätverk presterar jämfört med en enklare regressionsmodell i att förutsäga avloppsflöde med hjälp av högupplöst X-band radardata. I projektet undersöktes också hur tillgång av olika radardata kunde bidra till modellens prestanda. Modellerna utvärderades endast under regnflödesperioder.

Data över avloppsflödet som användes i projektet kom från Avedøre avloppsreningsverk i Köpenhamn. Reningsverket drivs av BIOFOS. Radardata kom från HOFOR. Det neurala nätverket som användes har utvecklats av Informetics på plattformen Tensorflow.

Slutsatser som kunde dras i projektet var att det neurala nätverket och den linjär regressionsmodellen var lika bra på att förutsäga när en regnflödesperiod startade. Det neurala nätverket kunde förutsäga det momentana flödet bättre än regressionsmodellen, medan det omvända gällde för att uppskatta den totala flödesvolymen under en hel regnflödesperiod. Genom att använda ytterligare regndata, upp till 30 kilometer från radarstationen, jämfört med att endast använda data från avrinningsområdet kunde en 20–30 minuters förbättring av den möjliga prognostiden påvisas. En tänkbar prognostiden för att förutsäga avloppsflödet till Avedøre avloppsreningsverk visades ligga omkring 4 timmar.

Nyckelord: Neurala nätverk, Linjär regression, Flödesprognosering, Avloppsvatten, Reningsverk, Avrinningsmodellering, X-bandradar.

Institutionen för informationsteknologi, Uppsala universitet (UU) Box 337-75105 Uppsala, ISSN 1401-5765

PREFACE

This master thesis of 30 ECTS finishes the Master Programme in Environmental and Water Engineering at Uppsala University (UU) and the Swedish University of Agricultural Sciences (SLU). The supervisor was Nicholas South, water resource consultant at Tyréns. The subject reviewer was Bengt Carlsson, Professor at Department of Information Technology at Uppsala University. The examiner was Gabriele Messori, Associate Professor at Department of Earth Sciences at Uppsala University.

First, I would like to thank my supervisor Nicholas South for providing this opportunity and for your valuable feedback and guidance during my master thesis. I would also like to thank my subject reviewer Professor Bengt Carlsson for aiding me with valuable insights and research.

Much appreciation to Informetics. There I would like to thank: Peter Rasch, for initiating this thesis and making sure that I received all the necessary data. Lasse Boerresen for vital tutorial in using the model in Python and for influencing me to think like a programmer. Charlotte Plum for assisting me with retrieving the X-band radar data and answering my questions about the radar very thoroughly.

Many thanks to Carsten Thirsing at BIOFOS and Margit Lund Christensen at HOFOR for your contribution to this project in the form of data and valuable feedback.

Fredrik Wigertz

Knivsta, April 2021

POPULÄRVETENSKAPLIG SAMMANFATTNING

Dagens avloppsreningsverk använder många olika metoder för att rena vatten, ofta en kombination av fysiska, kemiska och biologiska processer. Reningsverken blir allt mer avancerade och kräver mer automatiserad kontroll för att säkerställa fungerande och komplexa processer. Att kunna förutsäga förändringar i inflödet av avloppsvatten är viktigt för att möjliggöra för kontrollsystemen att vara beredda och kunna motverka störningar vid kraftigt förhöjda inflöden av vatten exempelvis i samband med skyfall. Det föreligger därför höga krav på tillgänglighet och bearbetning av väderdata för att kunna optimera de tekniska systemen i förhållande till väder och klimat.

Avloppsvatten består av spillvatten från hushåll och industrier, av dräneringsvatten och av dagvatten från nederbörd. Mängden vatten som når ett reningsverk i form av spillvatten varierar över dygnet, veckan och året beroende på variationer i den mänsklig aktiviteten som genererar inflödet. Dagvatteninflödet kan skapa dramatiska förändringar i inflödet beroende på nederbörden och det är därför viktigt ur både miljömässigt och ekonomiskt hänseende att kunna förutsäga och motverka dessa kraftiga svängningar.

Under de senaste årtiondena har man, allt eftersom datakapaciteten ökat, försökt att skapa så kallade neurala nätverk som i sin uppbyggnad liknar hur nervcellerna är förbundna och fungerar i hjärnan. Målsättningen med det här examensarbetet var att utvärdera hur väl ett komplext nätverk kunde förutsäga vattenflöden till ett avloppsreningsverk jämfört med en enklare statistisk sambandsmodell.

För att utvärdera modellerna användes data från Avedøre avloppsreningsverk i Köpenhamn. Flödesdata i reningsverket jämfördes med högupplöst radardata (X-band) som visade nederbörden. I arbetet identifierades perioder där avloppsvattenflödet till reningsverket var starkt påverkat av nederbörd, så kallade regnflöden. Dessa perioder kan skapa problem för reningsprocesserna och de utgjorde därför utgångspunkt för utvärderingen av hur väl modellerna kunde förutsäga flöde som är påverkat av regn.

Slutsatser som kunde dras i examensarbetet var att det neurala nätverket och den enklare statistiska sambandsmodellen var lika bra på att förutsäga när en regnflödesperiod startade. Det neurala nätverket var bättre på att förutsäga det ögonblickliga flödet än den enklare modellen, medan det omvända gällde för att uppskatta det totala flödet under en hel regnflödesperiod. Genom att använda ytterligare regndata, upp till 30 kilometer från radarstationen, jämfört med att endast använda data från avrinningsområdet kunde en 20–30 minuters förbättring av den möjliga prognostiden påvisas. Den maximala prognostiden för att förutsäga avloppsflödet till Avedøre avloppsreningsverk visades ligga mellan 4–5 timmar.

ABREVATIONS AND DEFINITIONS

Abbreviations:

WWTP: Wastewater Treatment Plant

NN: Neural Network

LRM: Linear Regression Model

MAE: Mean Absolute Error

reLu: rectified Linear unit (Activation function)

pdf: probability density function

Definitions:

Input signal/Input data/Input: The input data to a model that the prediction of the output signal is based upon.

Output signal/Output data/Output: The output data that the model is trying to predict.

Lead time: How long ahead in time of the input signal that the prediction is made. Also called prediction horizon.

Dry flow period: Periods where the wastewater flow is not influenced by rain.

Rain flow period: Periods where rain has infiltrated the sewer system, thus adding to the wastewater flow.

Evaluation period: A subset of rain flow periods deemed suitable for the performance of the models to be evaluated over.

Flow shift: The shift between a dry flow period and a rain flow period.

TABLE OF CONTENT

1. INTRODUCTION

Modern wastewater treatment plants (WWTP) use a variety of different processes of both physical, chemical, and biological nature to reach today's environmental demands on effluent water (Svenskt Vatten 2019). As WWTPs become more advanced the need for operational control increases to ensure that the complex treatment processes will remain functional and resources efficient. Forecasting the disturbances of the influent flow, primarily caused by precipitation, will allow for a feedforward control system (Bennet 1979) and action can be taken to counteract these disturbances before they have a harmful effect on the treatment processes. The benefits of forecasting are therefore both environmental and economical. Additionally, failing to clean wastewater may cause the spread of infectious diseases (Svenskt Vatten 2013).

Avedøre wastewater treatment plant is found in Hvidovre municipality in the southern part of the capital area of Denmark. Avedøre WWTP faces the sea at Køge Bugt, a bay area in the strait of Øresund. Avedøre WWTP is operated by BIOFOS, Denmark's largest wastewater organization. BIOFOS provides treatment services for 1.2 million inhabitants through three different WWTPs, Avedøre, Lynetten and Damhusåsen (BIOFOS 2021), BIOFOS is owned by 15 municipalities located within the capital area.

The treatment processes at Avedøre WWTP are controlled by the control software STAR Utility Solutions^{TM} developed by Veolia Water Technologies (Krüger, 2021). The control has for instance helped increasing the hydraulic capacity of the WWTP without increasing the process volume. This has reduced the risk of sludge escape (Veolia-1, 2016) and improved the quality of the effluent water. The control is adjusting the processes in relation to the incoming flow (Veolia-2, 2016). If the flow has substantially increased by the influence of rain Avedøre WWTP will shift the treatment processes to handle the rain flow that has other characteristics than the normal flow also referred to as dry flow. The sooner a shift from dry flow to rain flow can be predicted before its realisation, the more time will be given for the treatment processes to adjust.

A limiting factor when setting up a forecast is the data availability and the measuring techniques (Beven 2001). The influent flow to the WWTP has predominantly two sources of driving variables that need to be measured for forecasts. The first one being the wastewater resulting from human activities and therefore behaving periodically. The second one being the precipitation which disturbs the system randomly. The short response time in an urban catchment makes high temporal and spatial resolution of the precipitation data important to give accurate forecasts (Einfalt et al. 2004). This is especially important when measuring peak flows from intensive short lived and local weather (Thorndahl et al. 2017). Errors in precipitation data accounts for a large part of the uncertainty when modelling the rainfall-runoff relationship.

In early 2017 an X-band radar station was installed in western Copenhagen by HOFOR. HOFOR supplies the capital region of Denmark with water and operates parts of the sewer system. An X-band radar is a high-resolution radar excellent at measuring local weather occurrences with a 30-kilometre measurement range and a 50-kilometre

observational range. High-resolution radar measurements lead to a large amount of data. To store and process that amount of data large and effective computer resources are needed.

In urban hydrology, one application of short-term weather-related forecasts is the ability to predict hydrologic changes due to weather (Thorndahl et al. 2017). First, the relationship between the weather and the responding hydrologic feature needs to be modelled to allow for the prediction to be made. One modelling approach that has been extensively used in the past is to represent the natural processes numerically. Another approach data-driven modelling such as machine learning and that model does not require knowledge of the processes but instead relies on finding relationships in the available data.

One common model in machine learning is the linear regression model that was developed in the field of statistics. The model is best used when the relationship between the variables at large is linear. A neural network is another machine learning model that is used for many different applications (Guttag, 2017) in our data driven society. It resembles the human neural network and can efficiently find complex non-linear patterns in massive amount of data.

In this project a neural network software developed by Informetics will be used to forecast wastewater flow from primarily rain data measured by X-band radar. In 2020, Faust and Nelsson used the same Neural Network model with X-band radar to forecast the wastewater flow in Lund in their master thesis. They concluded that it was possible to accurately forecast 1 hour ahead of time for that relatively small catchment when using exclusively rain data from within the catchmentThey suggested further investigations to determine if rain data from outside the catchment could improve how far ahead the forecasts can be made.

1.1. PROJECT AIM

This project aimed to examine how the choice of machine learning model and rain data influence the forecast performance. A neural network was compared to a linear regression model when forecasting sewer flow of different lead times and using different extents of X-band radar data as input signal. Since the control strategy of Avedøre WWTP is mostly interested in when dry flow shifts to rain flow the performance of the models in this project was primarily evaluated by their ability to predict when this shift occurs. Their ability to accurately predict the flow rate was also of interest. The further into the future that the models can make well performing forecasts the better.

More specifically the aim was to answer the following questions:

- How does a neural network perform compared to a multivariate linear regression model when forecasting sewer flow with X-band radar data?
- What extent of X-band radar data improves the ability to make accurate sewer flow forecasts at longer lead times?

2. THEORY

The theory section is divided into three parts. The first part is concerned with how wastewater is produced and transported both generally and with application to Avedøre WWTP. The second part presents background theory about the X-band radar. The third part introduces machine learning and more specifically linear regression models and neural networks.

2.1. WASTEWATER SOURCES AND TRANSPORTATION

2.1.1 Wastewater sources

Wastewater, stormwater, drainage water, and leakage water as defined by Swedish Water and Wastewater association (Swedish: Svenskt Vatten) all contribute to the inflow of WWTPs (Svenskt Vatten 2013). Wastewater is the contaminated water that is primarily intended to be processed at the WWTP. Wastewater is produced in households, by services and industries. Wastewater production is periodical and varies depending on time of day, day of the week, and on the season. Depending on the proportions of different types of wastewater sources the flow rate over time can vary a great deal because industries and services are not usually active on weekends and holidays. Other water sources enter the sewer system, either led intentional or leaked unintentional, from surface runoff or groundwater. These sources are regenerated by rainfall.

2.1.2. Rainfall – runoff processes

The relation between rainfall and runoff are studied by hydrologists and are of great concern for proper water resource management. The fraction of a rain fall that will contribute to a point downstream (such as a WWTP) within a certain time frame as the runoff production and the distribution of the rainfall runoff over time are regarded as runoff routing (Beven, 2001). Runoff production and routing are dependent on the characteristics of the catchment area, the rain fall and the climate. These factors may differ a lot between different locations. The hydrological processes in a catchment are spatially heterogeneous, they are in part occurring underground, the driving variables are hard to measure, and the processes are affected by non-linearities and a constantly changing environment (Kirchner 2009). Therefore, it is important to substantially simplify and generalize the driving processes to make a feasible model over the rainfall runoff relationship. In urban catchments, the impervious surfaces and relatively small areas give a short response time between rainfall and rise in flow compared with a natural catchment area (Thorndahl et al. 2017).

The short response time of an urban catchment combined with a limited capacity of the water infrastructure to handle large flows can cause urban flooding during high intensity rain (Dahlström, 2006). High intensity rain fall are most likely to occur during the summer in the Nordic climate (SMHI, 2020), because of a larger temperature difference between the surface and air. They can also emerge locally under a short period making it harder for the intensity to be measured accurately with a low-resolution radar (Dahlström, 2006). The Swedish meteorological and hydrological institute (SMHI)

defines a heavy rain fall to give more than 2 mm rain over a 10 minute-period or more than 10 mm rain over a 1 hour-period (SMHI, 2015).

Bengt Dahlström (2006) constructed a formula, based on rain data from 47 locations in Sweden, that calculates rainfall intensity based on the duration and the return time. A 6-hour rain yielding 7.8 mm of rain is likely to occur once a month and a 6-hour rain yielding 17.8 mm is likely to occur once a year. In comparison, during the summer of 2014, the largest rainfall event in Swedish measured history occurred in Malmö with 110 mm rain from 6 hour of rainfall (VA SYD 2017). This is approximately 30 mm higher than the total rainfall from a 6-hour rain that is likely to occur once every 100 years (Dahlström, 2006).

2.1.3. Sewer systems

In the case of a WWTP inflow, the rainwater must infiltrate the sewer system to be able to reach the WWTP. There are three types of sewer systems: combined systems, duplicate systems, and separate systems (Svenskt Vatten 2013). The amount of rain that will end up in the sewer depends on the type and quality of the sewer system.

- A combined system carries water from all types of sources. This system will be affected by rainfall and the WWTP needs to be ready to handle large fluctuations in flow.
- In a duplicate system wastewater and stormwater are led in separate pipes. This diverts the stormwater flow to the recipient (river, sea or lake) instead of the WWTP.
- The separate system leads the stormwater by other means than a pipe. Instead, it may for example be transported through a ditch and apprehended in a local treatment system.

The sewer system may have a basin or an overflow system that creates an unnatural and nonlinear relationship between rainfall and inflow to the WWTP (Svenskt Vatten 2013). Extraneous water leak into the sewer system through cracks and joints, make a rainfall event more likely to increase the flow in all types of sewer system but with different magnitudes. On average the inflow to Swedish WWTP is twice as high as the registered water consumption indicating leakage into the sewer or usage of a combined system.

2.1.4. Catchment of Avedøre WWTP

The urban catchment of Avedøre WWTP covers 184 km^2 out of which 125 km^2 is connected to the sewer system (BIOFOS 2016). The catchment of Avedøre WWTP had in 2015 a population of 254 399 inhabitants and in 2025 the population is expected to have increased to 278 000. The municipalities serviced by Avedøre WWTP are Albertslund, Ballerup, Brøndby, Glostrup, Herlev, Hvidovre, Høje-Taastrup, Ishøj, Rødovre and Vallensbæk. The water management features within the catchment are presented in Figure 1. The dark blue line shows the catchment border, and the black line shows the border of the 10 municipalities. The wastewater is transported by the main pipes to Avedøre WWTP.

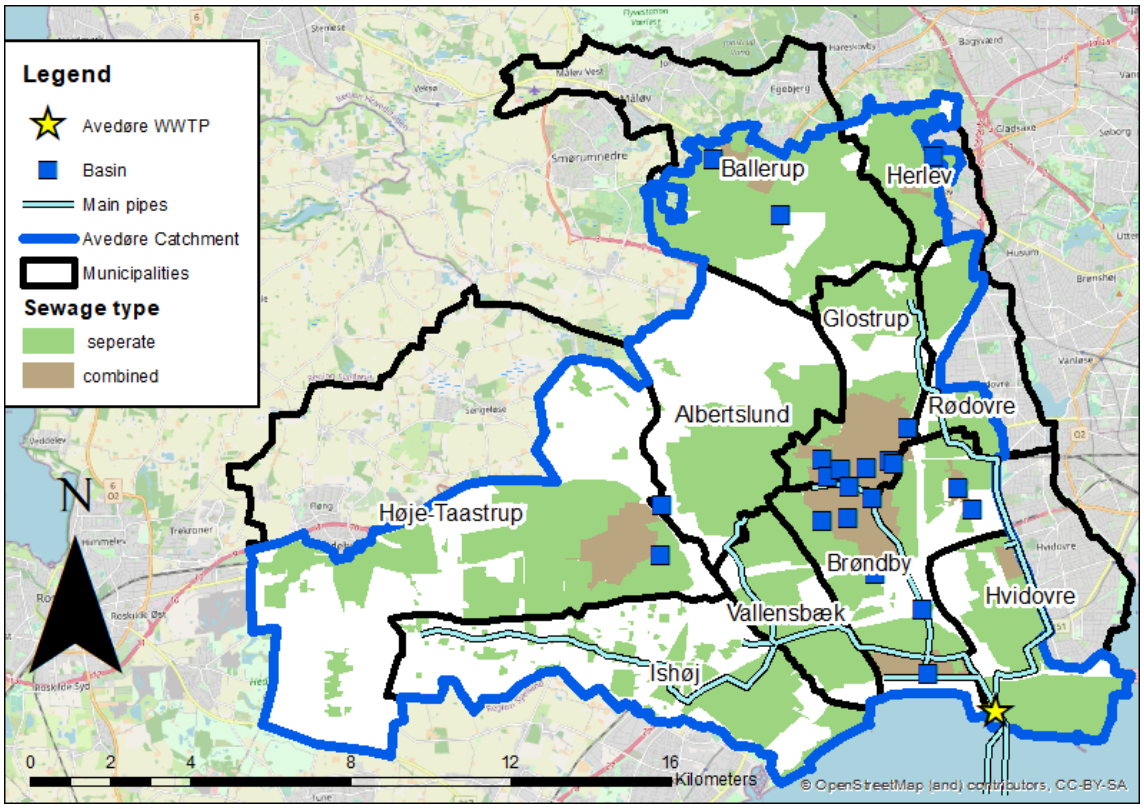


Figure 1: Map of Avedøre catchment showing the location of Avedøre WWTP, the sewer system by type, municipality borders and the catchment border.

In Figure 1 the separate system is shown in green, and the combined system is shown in brown. The basins that relieves the sewer system when storm water accumulates because of rainfall are shown as blue squares. Given the distribution of basins, it becomes quite clear that the combined sewer system is substantially more affected by storm water than the separate sewer system.

Water consumption data from six of the municipalities in 2014 indicate that households consume roughly 60-80% of the water and the remaining 20-40% is mainly consumed by industries and institutions. Water consumption data provides an estimate of the wastewater sources.

2.2. X-BAND RADAR

The short response time in an urban catchment makes high temporal and spatial resolution of the precipitation data important to enable accurate forecasts (Einfalt et al. 2004). This is especially important when measuring peak flows from intensive short lived and local weather (Thorndahl et al. 2017). Errors in precipitation data accounts for a large part of the uncertainty when modelling the rainfall-runoff relationship.

Precipitation measurements are carried out by rain gauges or by weather radar. These two techniques are fundamentally different when measuring precipitation and thus have different sources of error (Shellart et al. 2012). A rain gauge measures precipitation intensities directly in one spot while radar measures the reflectivity of rain particles continuously over a larger area.

The implementation of radar precipitation measurement has improved the monitoring of rainfall temporal and spatial variation (Beven 2001). A radar measures precipitation indirectly and has antennas that rotate while they send beams of electromagnetic pulses that are reflected by particles along their projected routes. It is assumed that the return signal is highly dependent of the precipitation intensity. The main variable that is measured is the reflectivity (Z) and it must be converted to rainfall rate (R) by a Z-R relationship (Einfalt et al. 2004). Depending on the type of rain different relationships are used. More advance radar systems also incorporate doppler and polarimetric measurements allowing better management of errors (van de Beek et al. 2010).

2.2.1. Radar data

Radar data are stored in a three-dimensional polar coordinate system with the radar in centre (South et al. 2019). The radar beam width is often 1 degree in azimuth leading to a variation in width from 100 metres to 1 kilometre depending on the distance from the radar (Einfalt et al. 2004), therefore the spatial accuracy will be higher closer the radar. The radar does not measure rain at the ground level and usually it measures on different elevation angles (Schellart et al. 2012). Data storage is also an issue. The X-band radar data used in South et al. (2019) produced 60 megabytes of data every minute. The total data over a 72-day period added up to 6.48 terabytes.

2.2.2. Sources of error and correction

The reflectivity is affected by the size of the rain droplets. Droplet sizes varies between different types of rainfalls making it important to correct the reflectivity – rainfall relationship with measurement of droplet size distribution (van de Beek et al. 2012). By relating the rain gauge measurement to radar a correction factor can be used to the earlier defined Z-R relationship (Achleitner et al. 2008). Correcting radar data with rain gauges has proven to be useful, but this relationship is based on a simplified assumption that radar and rain gauges are homogenous in time and space.

Since radar measures reflectivity with radar beams that are angled upwards into the air it is not certain that the measured rain will fall directly to the ground (Beven 2001), especially when winds are strong. With increasing distance from the radar, the angle of the radar beam will cause the measurement to occur on a higher altitude. This might cause the beam to overshoot the rain (Scheller et al. 2012).

Attenuation is the dampening of the radar beam through absorption of particles (Shellart et al. 2012). This problem can turn up with high intensity rain because of the increased number of particles. Any measurement beyond an event that causes attenuation will be underestimated or even blocked. For high frequency radar, such as X-band radar, attenuation becomes a significant problem (Einfalt et al. 2014). Radar will also measure disturbances such a clutter and background noises (Langfeld et al. 2014). Trees, houses and hills are examples of static clutter, and birds, insects, and other radar beams examples of dynamic clutter. Clutter often gives high reflectivity values. Background noises include atmospheric noises and noises from within the radar electrical circuits. There are many correction algorithms that try to filter out errors caused by attenuation, cluttering and background noises. More advanced techniques make use of data from

polarimetric and doppler radar data that are insensitive to attenuation and gives greater insights to clutter and attenuation detection (van de Beek et al. 2012; Thorndahl et al. 2017). They can reduce error further but not completely remove them.

The X-band radar that measured the rain intensity used in this project was installed by HOFOR early 2017 on top of a fire station. The X-band radar model is a WR-2100, produced by FURUNO (Furno 2021). WR-2100 is a compact dual polarimetric X-band doppler weather radar. WR-2100 is one of the smallest weather radars and it is aimed to measure local clouds within a 30 km radius, but it observes data up to 50 km radius. The radar data are interpolated to a cartesian grid system of desired spatial resolution and with a lowest possible resolution of 100x100m. The data from the HOFOR X-band radar are interpolated into a spatial resolution of 500x500 meters and temporal resolution of 1 minute. The measurements from the radar ended in June 2020 when the radar station was taken out of use.

2.3. MACHINE LEARNING

The exponential increase of computer power accelerates our journey towards a data driven society where massive amounts of observed data are stored. Machine learning algorithms can find patterns and relationships between the data to optimise the utility of technology and services (Guttag 2017). Whereas traditional programming finds the desired output signal by using a fixed model with sample data as its input signal, machine learning uses sampled data as both the input and output signal to find the model without knowing the details of the system. If the physical relationship between the input and output is delayed, such as the relationship between rainfall and runoff, this allows for the output to be forecasted ahead of time from the observed input data.

In this project two machine learning models are compared. The first model is the Linear Regression Model (LRM) which was developed in the field of statistics and assumes a linear relationship between the observed data. The second model is the neural network (NN) which is mathematically inspired by the neural network in our brains (Kartalopolous 1996).

2.3.1. Training and Validation

The goal with machine learning is to train a model such that it can predict from input signals, that it has not previously been trained upon, with the least deviation from the true output. Within the scope of this project the input signal are primarily rain data and the output signal are flow data. The available data are usually divided into a larger training set and a smaller validation set to ensure that the model is not exclusively fitted to the trained data.

The way in which a model can change its output is by varying the weights that transforms input data into output data. To find the ideal set of weights one must first define a criterion of what constitutes a better model. This criterion is called the loss function. There are many types of loss functions, but the general idea is to minimize the error between the output signal and the prediction (Carlson & Lindholm 2019). Informetics software uses a negative log likelihood loss function (L) shown in equation

(1) (Larochelle 2013). The prediction is not a single value, but a normal distribution that has a mean and a standard deviation. ذ means the likelihood of the prediction ó containing the output signal y. *u* is the input signal.

$$, \text{ó} \; \text{Ù} \; \text{سي طي} \; \text{ì} \; \text{ó} \; \text{ذ} \qquad (1)$$

To maximize the likelihood, we want to minimize the negative likelihood in equation (1) (Starmer 2017). The set of weights that minimizes the negative log likelihood loss function provides the best solution for the given model and results in the lowest loss value. Changing the weights will result in a different mean and standard deviation for the same set of input signals.

Informetics software provides both a loss value for the training and for the validation. Training the model on the complete data set one time is called an epoch. Each epoch should result in a lower loss value otherwise no additional learning has taken place. When the loss has reached a plateau, a minimum has been reached (Sanderson 2017). If the model is complex, there might exist multiple local minimums that the training can converge towards as shown in Figure 2. Finding the global minimum is not guaranteed but through training with different learning settings a new optimal solution might be reached.

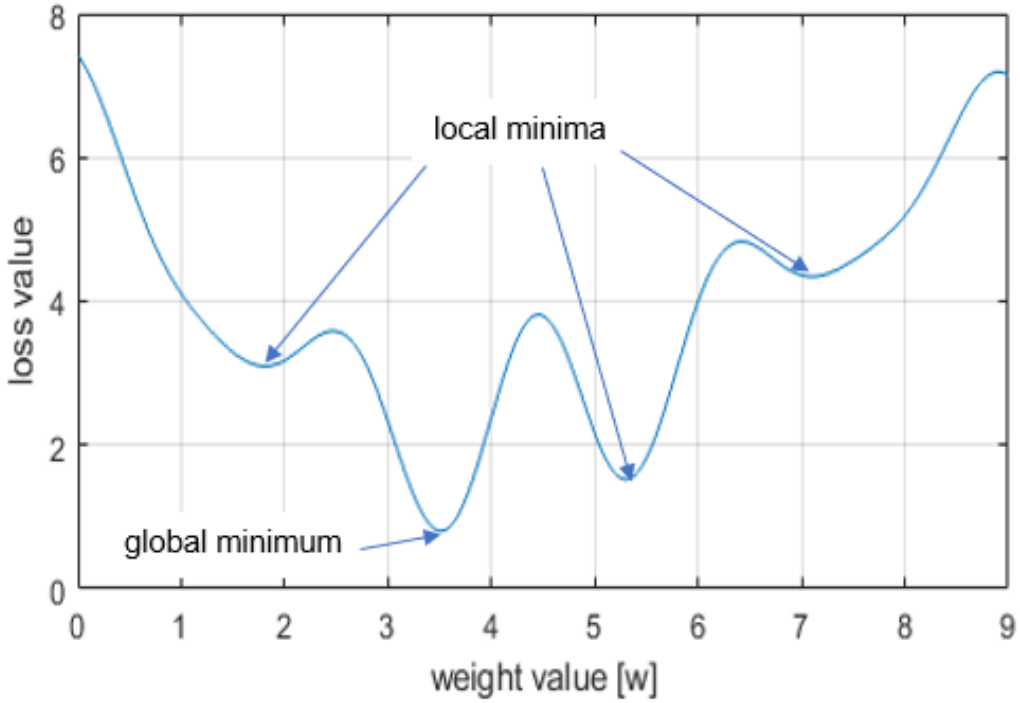


Figure 2: Graph of a complex loss function where the loss value varies by altering a single weight value.

2.3.2. Neural network

The architecture of a Neural Network can vary greatly from one to another and the Neural Network presented in this project is just one amongst many. The fundamental building block of the neural network consists of the node that corresponds to the neuron in the brain. The nodes are distributed in different layers (see Figure 3). The first layer from the left is the input layer where each input data set (u) corresponds to one node.

The last layer is the output layer where the red node corresponds to the prediction. The two layers in the middle in Figure 3 are called the hidden layers and every node in one of the layers is connected to all nodes in the previous layer for this Neural Network. Changing the number of layers and the number of nodes within a layer creates a more or less complex neural network.

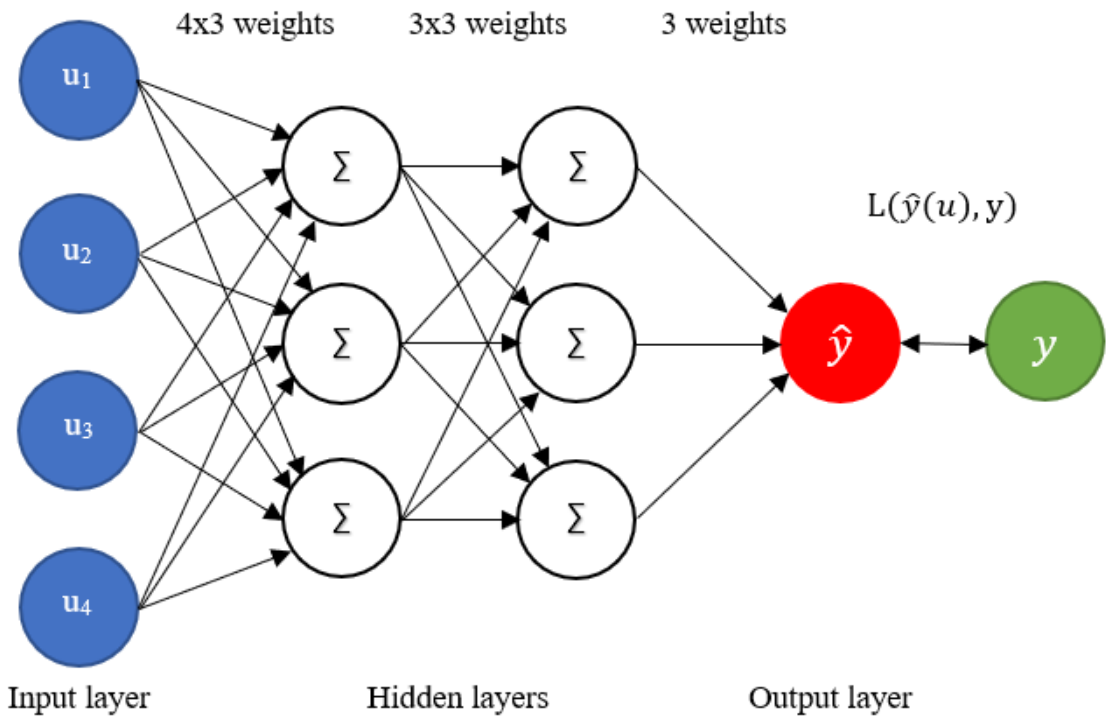


Figure 3: A fully-connected neural network consisting of 4 input signals, 2 hidden layers with 3 nodes each and 1 prediction node (output layer). Each node is connected to all the nodes in the previous layer where each connection has a weight (multiplier) attached to it. Training occurs when the loss function (L) is minimized.

In the brain the electrical potential difference created by chemical processes in the synapses determines if a neuron is activated or not (Kartalopoulos 1996). Similarly, the nodes in a neural network have an activation function (not shown in the Figure 3) attached to them that determines what the sum of the data from the nodes in the previous layer must be to activate the node (Starmer 2021) and pass through the data. The activation function used in Informetics software is a rectified linear unit (reLu). It transforms all negative values to zero and all positive values are kept the same (Starmer 2020). It is a simple and effective way of making the training process go faster than with other activation functions.

Each connection also has a weight attached that scales the data coming from each previous node. These are the weights that are being trained to optimize the model with regards to the loss function. The algorithm that is used in Informatics software is called the back-propagation algorithm and basically it starts with optimizing the weights attached to the last layer and then the second-to-last layer and so on (Kartalopoulos 1996).

Additional to the model itself there are hyperparameters that determine the properties of the training process. The hyperparameters that can be altered in the Informetics software are listed below with an explanation of their purpose. To find the optimal training of a model the impact of these hyperparameters needs to be evaluated.

- **Number and size of hidden layers:** The structure of the Neural Network. More hidden layers and more nodes within each layer means more weights to be optimized.
- **Number of Epochs**: Number of times the complete data set will be trained upon. More epochs leads to a more time consuming training session.
- **Learning rate**: Decides how much the weights may change when training on one batch of data points. Figure 4 shows the loss function progression with a fast learning rate (left) compared to a slow learning rate (right).
- **Learning rate decay**: Reduces the learning rate as the training progresses to avoid overshooting the loss function minimum. If learning rate decay is set to 0 the learning rate will remain constant.
- **Learning rate decay steps**: Determines how often the learning rate decays.
- **Hidden dropout rate**: Randomly discards nodes during training. This is done to reduce the risk of overfitting the model.
- **Batch size**: Amount of data points that are trained in one instance. Larger batch size means fewer times during an epoch that the weights are optimized by the back-propagation algorithm.

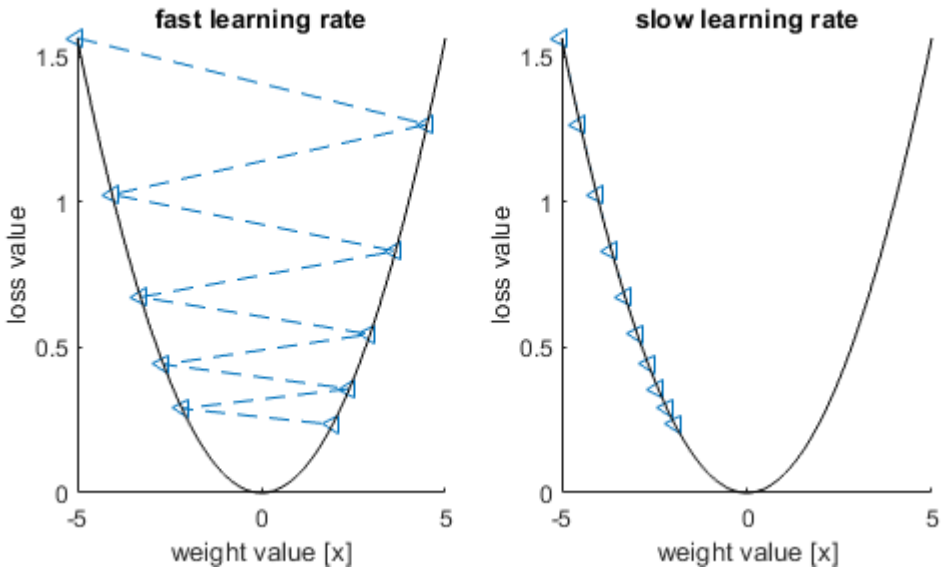


Figure 4: Optimization of a loss function with fast learning rate (left) and a slow learning rate (right). Each triangle indicates one epoch.

2.3.3. Linear regression model

The simplest mathematical model to describe a relationship between two variables is to infer a linear relationship (Chetwynd & Diggle 2011). In comparison with the neural network the linear regression model has only one input layer and one output layer without any activation functions in-between. Instead, there is one weight to be

optimized for each connection between a node in the input layer and a node in the output layer.

Models that use previously measured data of the output as an input signal are called autoregressive (Carlsson & Lindholm 2019). The ARX model stands for AutoRegressive model with an eXternal input. This is the linear regression model that best resembles the model used in this project, using both rain data and old flow data to predict the flow in the future.

An example of an ARX-model is described in equation (2). The *a:*s are the weights that model the relationship between the previous output signal *y* and the prediction $\hat{y}$. The *b*:s are the weights that model the relationship between the two input signals (or external data) u_1 and u_2 and the prediction $\hat{y}$. *t* is equal to discrete time and *k* is the lead time.

$$\hat{y}(t+k) = a_1 y(t) + a_2 y(t-1) + b_1 u_1(t) + b_2 u_1(t-1) + b_3 u_2(t) + b_4 u_2(t-1) \quad (2)$$

2.3.4. Lead times

Since the models are supposed to produce a forecast of a future flow, they also need to optimize their weights for that relationship. In Informetics software the lead times are defined as time shifts. This means that the output signal (flow) is shifted with the intended lead time compared with the input signal and is trained in the same way as if the flow was the current flow.

To save computational time, models of different lead times can be trained at the same time. This is done by simply adding additional nodes to the output layer. The consequence of this is shown in Figure 5. The neural network (right) that has hidden layers will share the model structure (same weights) until the last hidden layer. For the linear regression model (left) this will have no effect since the node in the input layer is transformed directly onto the nodes in the output layer.

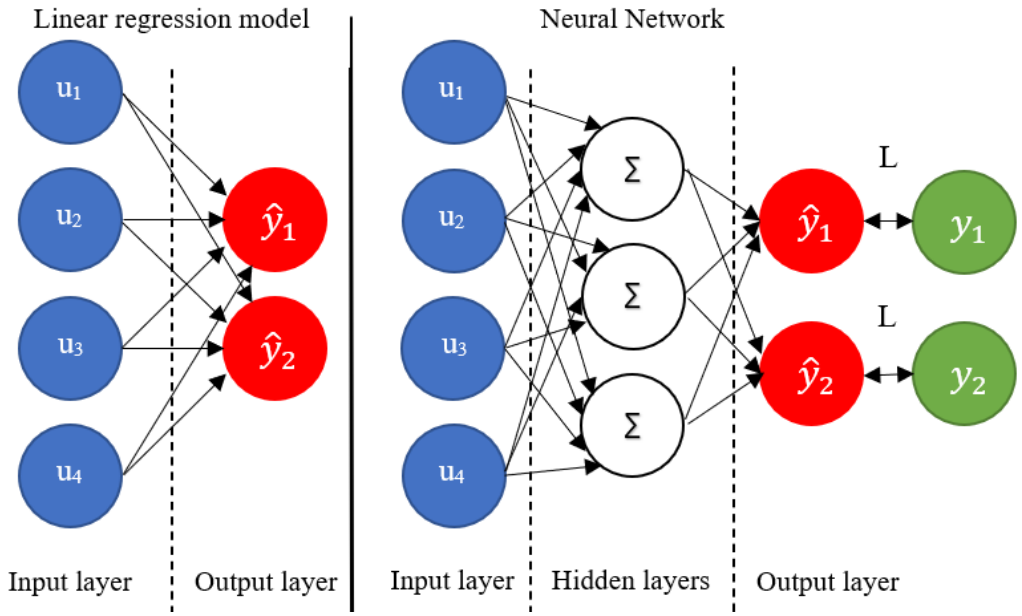


Figure 5: Linear regression model (left) with multiple forecasts of different lead times.Neural network (right) with multiple forecasts of different lead times. The number of weights between the input layer and hidden layer stays the same.

3. METHOD

The method section is divided into three parts. The first part presents the data and how the delimitation of the different datasets was conducted. The second part presents how the learning processes of the different machine learning models were conducted. The third part presents the evaluation method that the comparison of the models was based upon.

3.1. DATA AND INPUT

3.1.1. Flow data

The sewer flow data were provided by BIOFOS and covers the period of January 1, 2017 to June 30, 2020. The registered data points represent the pumping rate at the inlet to Avedøre WWTP, with a one-minute sampling time. The pumps are one by one activated by the water level of an adjacent basin and are either turned on full effect or turned off. Consequently, the flow time series will have stepwise shifts in flow. This also means that the time series will have a lot of periods of different length with zero flow.

Figure 6 presents two histograms of the probability density function (pdf) of flow rates from two different condition. The left histogram relates to flow rates from rainy condition and the right histogram relates to flow rates from relative dry conditions. The first peak of the histograms ranges from about 30 m^3/min to 45 m^3/min and corresponds to one active pump in the inlet of the WWTP. The second peak ranges from about 60 m^3/min to 85 m^3/min and corresponds to two active pumps. Because of the flow rate variability for each pump the peaks representing three or more active pumps becomes less distinct. The mean flow during the entire period, including both rain flow and dry flow periods, was 50 m^3/min and the mean flow during solely dry flow period was 38 m^3/min.

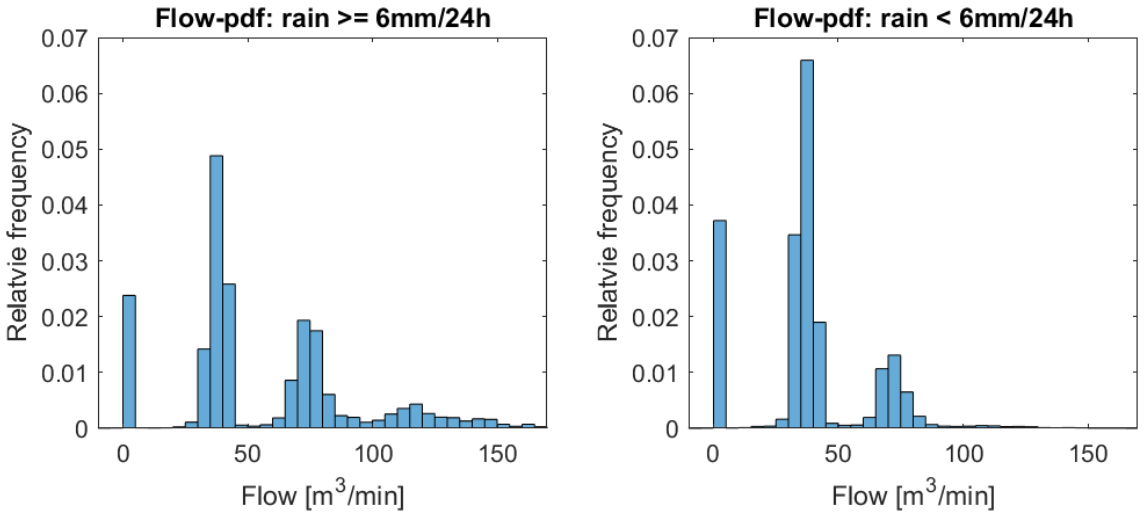


Figure 6: Histograms showing flow rate probability density function (pdf) by prevalence of rain. Left histogram represents rainy periods and right histogram represent dry periods.

Even though the right histogram, representing dry flow, shows considerably fewer flow rates corresponding to three pumps or more, these higher flow rates are present and make it harder to distinguish between rain flow and dry flow periods. A comparison of how the flow time series present itself during a dry flow period and a rain flow period is shown in Figure 7. Even during intense rain periods, the flow can for shorter periods shift to zero (left plot). The opposite occurs during the dry flow period (right plot) where there is no flow during longer periods and then suddenly it shifts to over 100 m^3/min to compensate for a shorter period with no active pumps. Additionally, the flow data consists of 2.2 % of datapoints with unknown flow. The longest consecutive period with no flow data is 14.6 days and starts at 19 December, 2018.

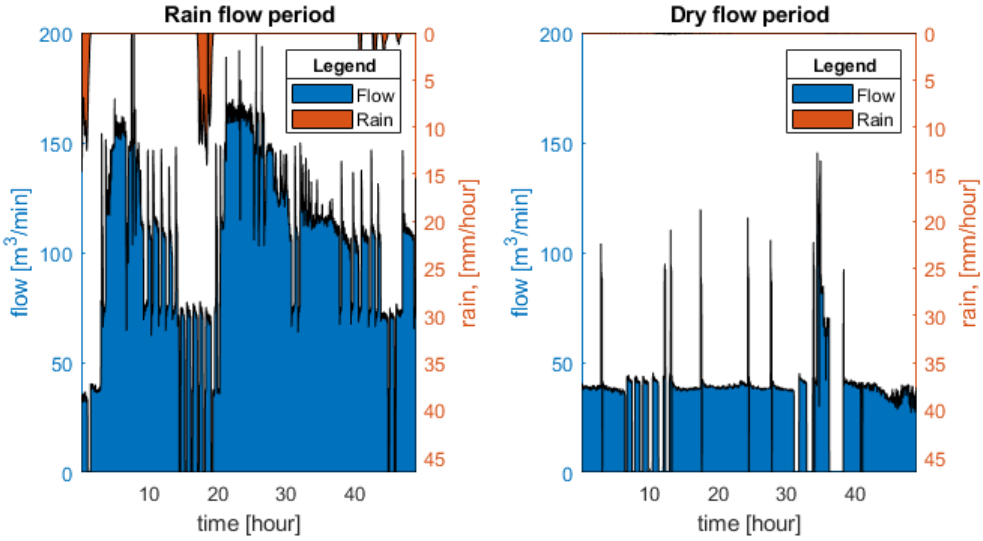


Figure 7: Examples of flow variation for a rain flow period (left plot) and for a dry flow period (right plot).

In the project the flow data were used both as an input signal and the output signal of which the prediction was trying to recreate.

3.1.2. Rain data

The rain data used in the project were measured by the X-band radar owned by HOFOR and is presented in Section 2.2.3. The radar data were downloaded through the VeVa api, made by Dryp. The temporal resolution of the data is 1-min and the spatial resolution was 500x500 meters. The catchment of Avedøre WWTP was 580 km^2 corresponding to approximately 2320 radar data points. Even though it is fully possible to use every datapoint as an individual input signal for the models, the project used different delimitations of radar data aggregates.

The strategy of evaluating how the extent of rain data impacts the performance of the forecast was done in two steps.

1. The first step investigated if the models could make use of the spatial distribution of rain data. This was done by comparing catchment rain data aggregated into one single rain data file with rain data aggregated into 10 rain data files from the 10 municipalities within the catchment. This allows to put different weights on rain data from different areas, which could be needed given the large catchment as well as the different sewer systems. The corresponding areas of the rain data files from within the catchment can be seen in Figure 8. The green dots represent every single data point. Given the different sizes of the municipalities the different files will be based on varied amount of data points.

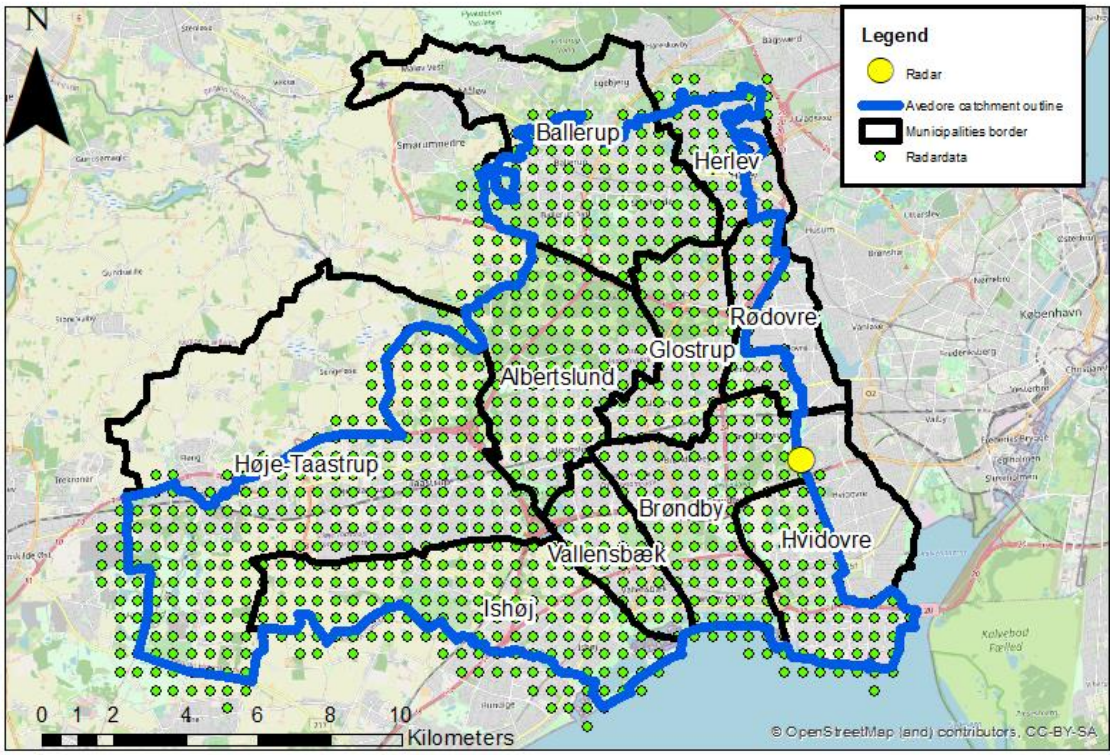


Figure 8: Radar data delimitation inside catchment. A green dot corresponds to a single radar data point.

2. The second step investigates if rain radar data from outside the catchment improve the ability to make better predictions based on the best model from the previous step. This is done by comparing the model based on radar data from the catchment with

additional radar data within 20 kilometres and 30 kilometres from the radar station. The delimitation for 0-20 kilometres rain radar data and 20-30 kilometres radar was done in 8 parts each. Figure 9 shows the delimitation done for the radar data from outside the catchment. Since the catchment already covers large areas west of the radar station these rain files will be made up of considerably smaller amount of data points than the eastern areas.

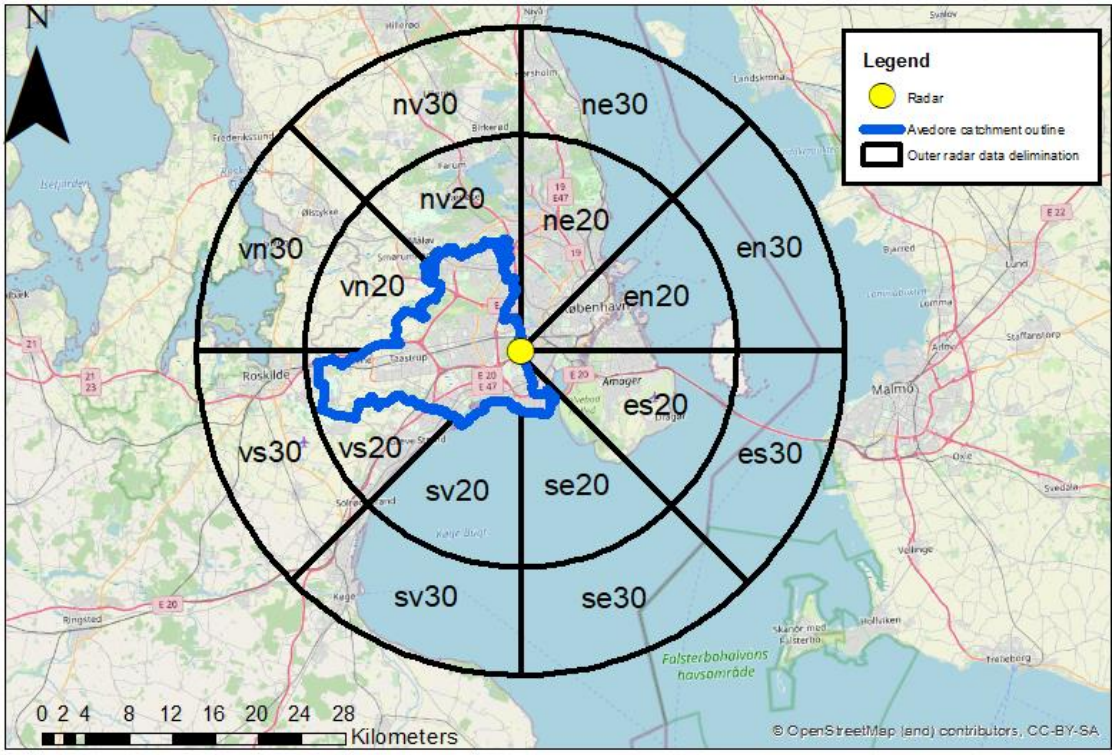


Figure 9: Radar data delimitation outside catchment. The first letter corresponds to the main point of the compass the second letter corresponds to the secondary point. The number indicates how far away in kilometres the data stretches from the radar station.

A table that summarises the rain data files can be found in appendix Section 8.1. Table 7 in Section 8.1 shows that among the different rain data files there are great variation in registered precipitation. The radar detected least rain towards the south west and most rain towards the north. The span of the total measured precipitation during the time periods for the rain files ranges from 2.91E+02 (sv30) to 5.35E+03 mm (Rødovre).

The different sets of rain data that will be compared as input signal for the models are referred to as:

- Full catchment rain data (FC): A single rain data file consisting of a mean made from 926 data points delimited by the Avedøre WWTP catchment border.
- Municipalities rain data (MUN): 10 rain data files consisting of means made from 31 to 225 data points delimited by the municipalities within Avedøre WWTP catchment,

- 20 kilometres rain data (MUN_20): Based upon MUN and adding 8 more rain data files consisting of means each made from 352 to 715 data points delimited by the outer areas 20 kilometres from the catchment.
- 30 kilometres rain data (MUN_30): Based upon MUN_20 and addings 8 more rain data files consisting of means each made from 653 to 859 data points delimited by the outer areas between 20 to 30 kilometres from the catchment.

As mentioned in Section 2.2.2. there are a few common errors associated with high resolution radar. With increasing distance from the radar station, the risk and occurrence of inaccurate data, due to for example attenuation or cluttering, also increases.

Figure 10 shows how the radar seems to fail in the registration of data, thus leaving gaps in the data. The errors that exist within the flow data and the rain data can all be deemed as being random noise that models will have to contend with.

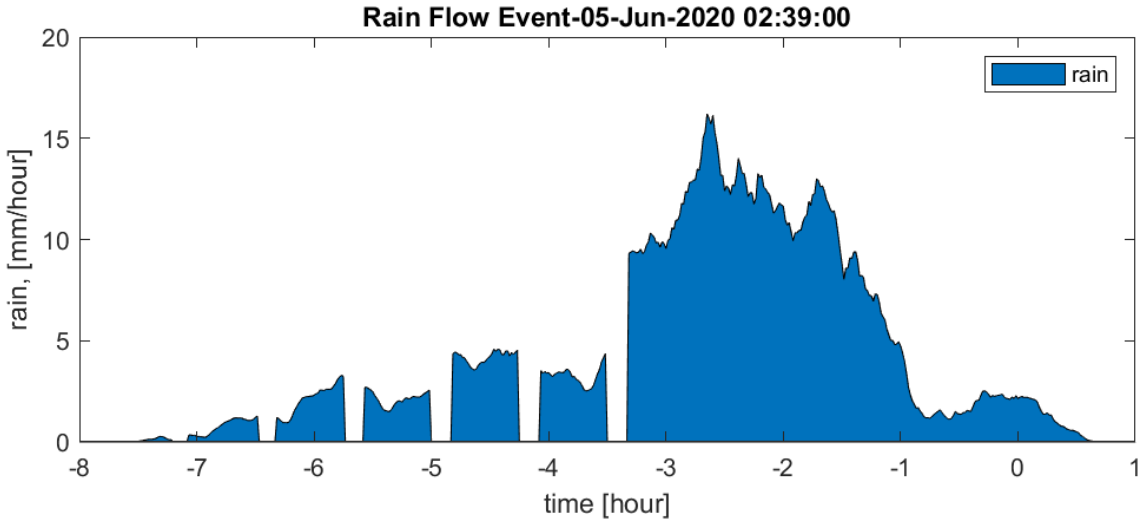


Figure 10: Example of a rain time series that is clearly impacted by errors.

3.1.3. Additional input signals

Beyond using flow data and the rain data as input signals to predict the flow in the future, additional input was created. These inputs were three different periodic functions indicating the time of year, the time of week and the time of day supporting trends recurring due to human activity. Since the response between rain and flow is not instant, rolling means of the data were added with different durations. These durations were set to 1, 2, 4, 8, 24 hours.

The lead times in the project were set to 0.5, 1, 3, 5 and 7 hours. The interesting aspect using these lead times when comparing the models is to pinpoint which model and dataset yields the least delay when determining the shift of the flow rate due to infiltration of rain. The shorter lead times (0.5 and 1 hour) will provide references on what the expectation of the longer forecast should be.

3.2. MODEL TRAINING

The model training process consisted of optimizing each model type and rain data combination with regards to the hyperparameters described in Section 2.3. First the full catchment and municipalities rain data were to be trained with both the LRM and the

NN. This led to four different sets of hyperparameters to be determined to optimize the training of each model data combination. Based on the optimal model and dataset from the first step two additional models were to be trained based on rain data files including areas within 20 kilometres and 30 kilometres from the radar station. In total 6 models were to be trained.

The data used in each model were divided into a training data set and a validation data set. The validation data set consisted of data from the first 10 days of each month. The training data set consisted of data from the remaining days of each month. The reason for the division is to distribute training and validation over the full period such that the daily, weekly, and yearly variation of the wastewater production are equally present for both data sets. It is also assumed that there is not a substantial variation within one single month.

3.2.1. Hyperparameter optimization

The evaluation of which hyperparameter setting that yielded the optimal training was based on three factors: validation loss, training loss and stability. Training loss indicates how well the model predicts the training data set and validation loss indicates how well the model predicts the validation data set. Since the training is performed only on the training data set the training loss is usually lower than the validation loss. If training loss is much better, then it indicates that the model has been overfitted to the training data. If the validation is better, it indicates that the model is too generalised and the validation data set being easier to predict. Stability of training considers how the loss improves over the course of the training . As the training progresses it is important that the loss becomes lower otherwise no training has occurred. If the loss oscillates it indicates that the model cannot find a low point. Examples of both stable and unstable training can be seen in Figure 11. Even though the training is unstable, the loss value might have ended on a low point. Evaluation of which hyperparameter is optimal is therefore a qualitative process, keeping both the stability of the training and the lowest possible loss value in mind.

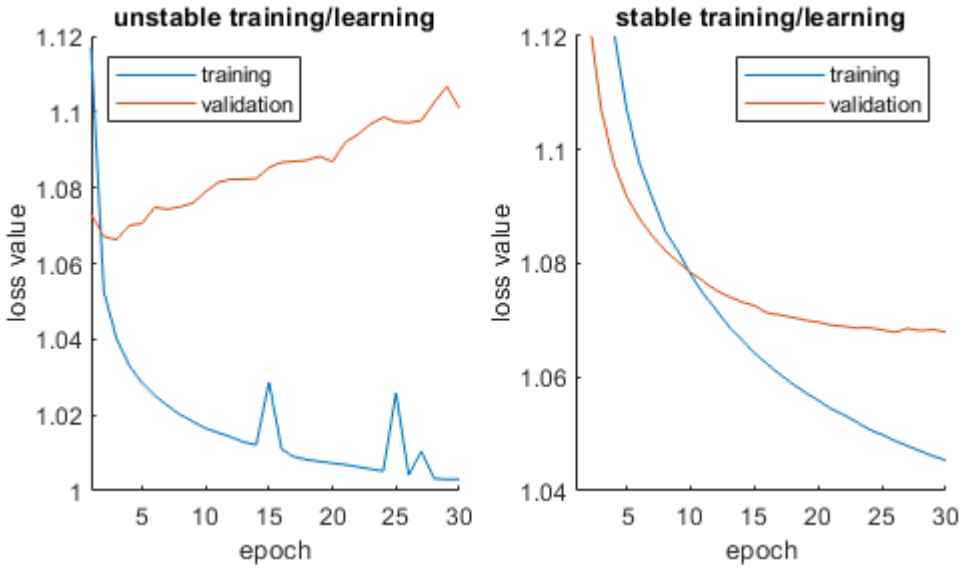


Figure 11: Training progress showing unstable training (left) and stable training (right).

Since the LRM does not have the hidden layer structure like the Neural Network there was no need to optimize the number of layers, number of nodes and the hidden dropout rate for that model. Still there were lots of possible hyperparameter combinations to be considered. To reduce the time spent on testing different hyperparameter sets, knowledge learned from training the first model was used for later training thus reducing the possible values for the hyperparameters which were tested. How the training was affected by varying the value of a single hyperparameter was done for all hyperparameters at least once. Also, the interdependency of the hyperparameters learning rate, learning rate decay, and learning rate decay steps was investigated. The number of epochs used in training were selected to ensure that the training and validation loss had reached its lowest value.

The hyperparameters that were tested and the maximum range of variables that were tried are summarised in Table 1. Summary of the conclusions from testing each hyperparameter are presented in appendix Section 8.2. Section 8.2 also includes the final hyperparameter settings used for training each model.

Table 1: Presents the maximal value range that was tested for each hyperparameter.

Hyperparameter	**Range**
Layers	0 - 2
Nodes	8 - 256
Epochs	10 - 60
Learning rate	0.1 - 1E-5
Learning rate decay	0 - 2
Learning rate decay steps	100 - 1E+6
Hidden dropout rate	0 - 1
Batch size	16 - 128

3.2.2. Resulting model

The largest model used in this project was a neural network with a single hidden layer consisting of 64 nodes and using 26 rain data files (MUN30) for input signals. This model is presented in Figure 12 where the rain data files are denoted P, flow data are denoted Q, the periodic function indicating the time is denoted T, the prediction is denoted $\hat{Q}_k$ and the nodes are denoted N. There are 5 predictions that uses 5 different lead times (k = 0.5h, 1h, 3h, 5h, 7h). Each input data sources uses the instantaneous value at time t as an input signal together with 5 different rolling means of past data. The rolling means are represented with an added dash above the abbreviations of the input signals.

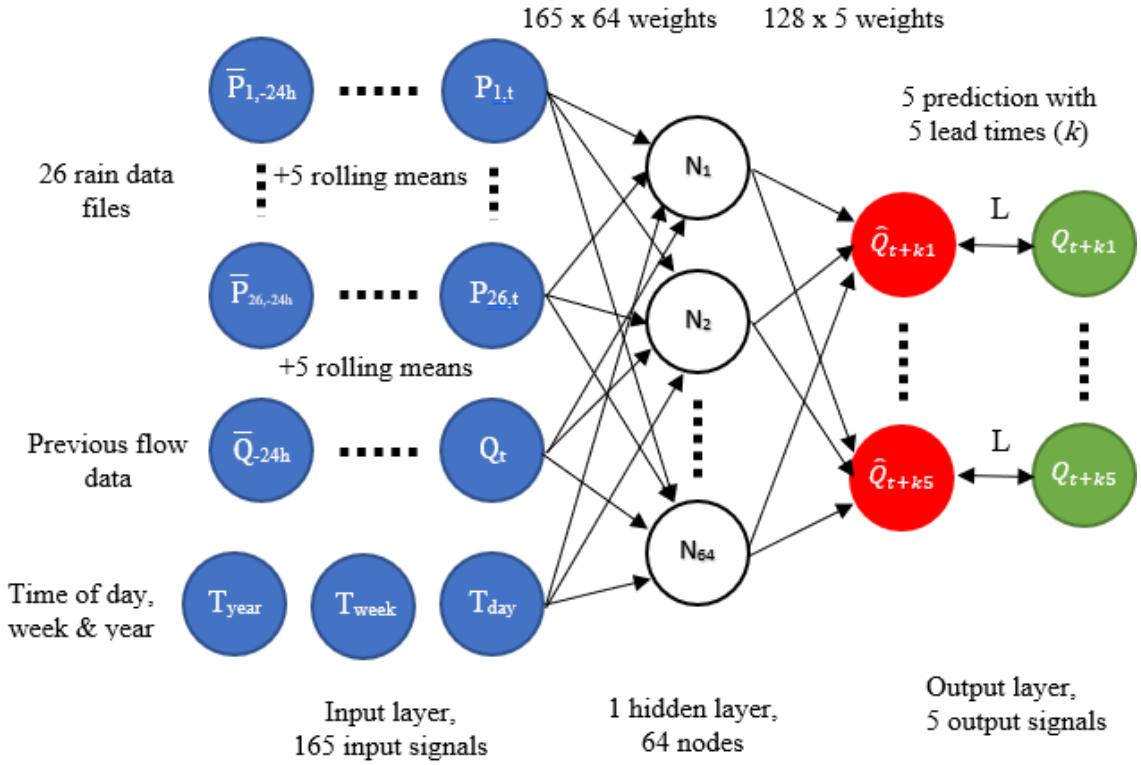


Figure 12: Largest model structure used in this project. There are 165 input nodes (blue) and for each input data set 5 additional rolling mean input signals are created from historic data. The hidden layer nodes are shown in white. The predictions are shown in red, and the output signals are shown in green. The weights (arrows) are modified to optimize the loss function L.

In total there are 165 different input signals leading to 10 560 weights between the input layer and the hidden layer. Also, each input signal is duplicated, and the duplicate is used to identify where in the time series there are non-existing values. The prediction, as mentioned in section 2.3, contains both a mean and a standard deviation that results in a duplication of nodes in the output layer. The actual number of weights that is trained for each model is presented in appendix section 8.2.

3.3. EVALUATION

The minimization of the loss function on which the training of the models is based upon is a measure of how well the forecast match the real flow as presented in Section 2.3.1. Since the intent of the forecast is to give a forewarning to Avedøre WWTP on when the sewer flow will switch from a dry flow to a rain flow the main performance to evaluate is timing. As mentioned in the introduction the control of the processes in response to the different flows is not continuous but rather a discrete on/off-function. This makes the precision of the forecast secondary to timing given that the flow rate is high enough to indicate a rain flow. The performance parameters that were used in this project are flow shift timing, relative volume, and MAE. These parameters are defined bellow.

- Flow shift timing is simply a comparison between the timing of the flow shift by the measured flow and by the forecast. Its purpose is to indicate how well the forecast can predict the timing of when a rain produces a substantial increase in flow and thus letting the WWTP know when to adjust the treatment processes.

- Relative volume is the comparison of the accumulated flow over the evaluation period between the forecast and the real flow. The forecast volume will be measured from when the shift of the forecast is deemed to have occurred. The relative volume shows how well the model approximates the flow.
- MAE is the comparison between the measured flow and the forecast at every timestep. This evaluates how precisely the forecast predicts the flow time series.

3.3.1. Selecting rain periods for evaluation

The first issue is to identify the rain periods of the flow time series. Given the characterization of the flow in Section 3.1.1. this is not a straightforward process since dry flow periods may provide periods of high flows and rain flow periods might have shorter breaks with no flow. Low intensity rain that are prolonged and not concentrated within a single period might result in a less distinctive switch of flow periods. The method used to identify rain flow periods are therefore initially set up to find all periods that may be representative of when a switch from dry flow to rain flow occurs. Following that a qualitatively analysis of these periods is done graphically. In the analysis, the periods with flows that show a clear shift from dry flow to rain flow in response to rainfall are selected and provided with a time stamp for that shift.

The quantitative identification process of the rain periods evaluates the rain and flow timeseries based on a set of threshold values. The variables that must reach these threshold values to start the possible rain period are:

- Instant flow rate ($\tilde{U}$),
- Mean flow of the upcoming 15 minutes ($\tilde{\theta}_{15}$)
- Mean flow of the upcoming 60 minutes ($\tilde{\theta}_{60}$)
- The accumulated rain during the last 6 hours. ($\tilde{U}_{6}$)

Three different flow rates are chosen to avoid the risk of choose a point of a sudden increase that decreases soon afterwards, given the nature of the flow time series. The rain duration of 6 hours is chosen with regards to the response time of the catchment.

When all threshold values were triggered the timing of the flow shift was set at that moment. Finding the end point was done by searching for the time-point when the mean of the rain flow period had returned to bellow 80m^3/min, given that the flow mean at some point was above that. Otherwise, the end point trigger was set right between the highest flow mean of that period and the threshold value that started the period.

The value of the threshold used in this project is presented with explanation in the following list.

- The value for all flow thresholds was set to 60 m^3 per minute. Making sure that at least two pumps or more were active during the period.
- The threshold value for the accumulated rain during the last 6 hours was set to 10 mm, which approximately corresponds to a rain which is likely to occur a slightly more seldom than once a month according to Dahlström (2006).

The rain periods that were identified through the quantitative process then were, one by one presented graphically. The aim was to select rain flow periods that showed a clear shift from dry flow to rain flow. Only the timespan around the start was analysed since it was that point that the performance evaluation was going to be based upon. Example of a rain flow period that shows a clear flow shift can be seen in Figure 13. First there is a compact and intense rain fall that 4 hours later at time zero provides a sharp increase in the flow. Example of a rain flow period that was deemed to be a bad representation of a shift from dry flow to rain flow is presented in Figure 14. The flow rate is already high before the rain fall and therefore the point of the shift cannot be decided.

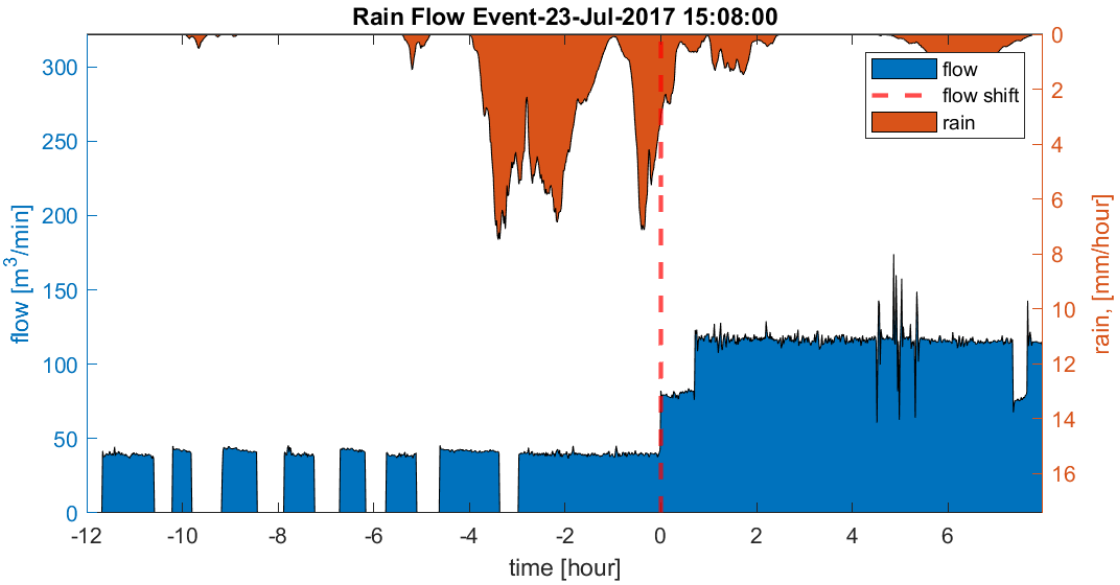


Figure 13: Example of a distinct shift from dry flow to rain flow.

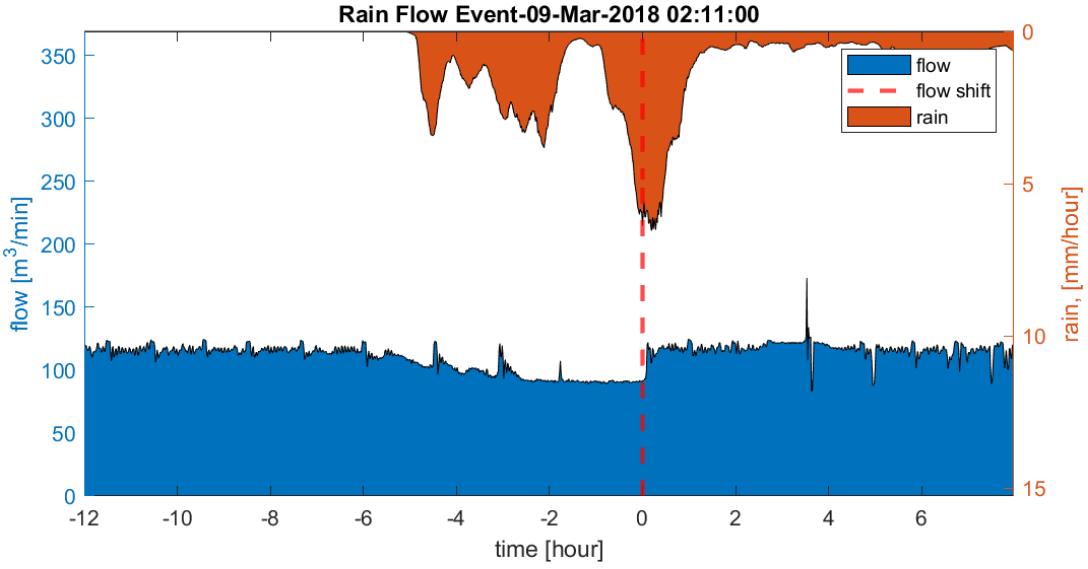


Figure 14: Example of a bad representation of a shift from dry flow to rain flow. The flow is already high, and this potential evaluation period must be disregarded.

A summary of all the rain periods that were selected to evaluate the performance of the model are presented in appendix Section 8.3.

3.3.2. Evaluation method

When the training of a model was completed, a csv-file containing time series for all forecasts at each lead time was created. These time series then wereevaluated with regards to the real flow within the evaluation periods. Both the forecast and the flow time series have a considerable amount of random noise which makes the comparison of the time series hard to assess.

Therefore, the time series were smoothed by moving average of all the data points ten minutes ahead and ten minutes behind. This removed almost all zero value flow rates during the studied periods. Before smoothing 3.7 % of the flow data points were zeros and after smoothing 0.3 % of the flow data points remained zero. An example of how the time series changed with smoothing can be seen in Figure 15 where each plot shows rain, flow and two forecasts one with 0.5-hour lead time and the other with 5-hour lead time - left plot before smoothing and right after. The rain time series are not shown in the figure and were not smoothed.

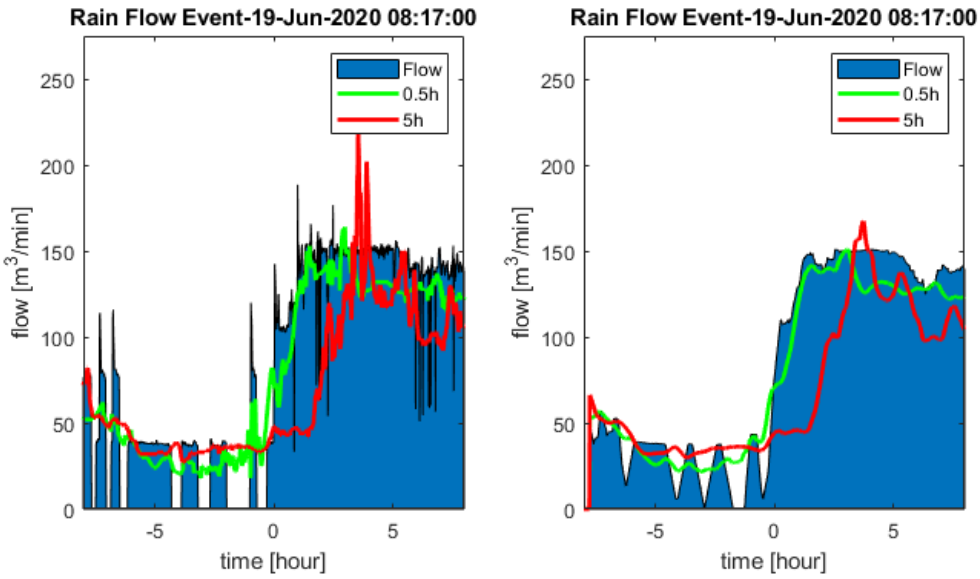


Figure 15: Flow and forecast time series before (left) and after smoothing (right).

For the evaluation of the forecast, starting points indicating a shift into a rain flow period needed to be determined. To stay consistent with the threshold values chosen to define the real rain flow period, the same value was chosen to define the shift of a forecast. Even though the time series where smoothed there was still oscillation which created multiple points of possible shifts within reasonable time frames from the rain weather. Therefore, every forecast was presented graphically, and the most feasible point of shift was selected. This process is shown in Figure 16 where the best guess is shown by the green vertical dashed line. The threshold value of 60m^3/min is shown by the green horizontal dashed line.

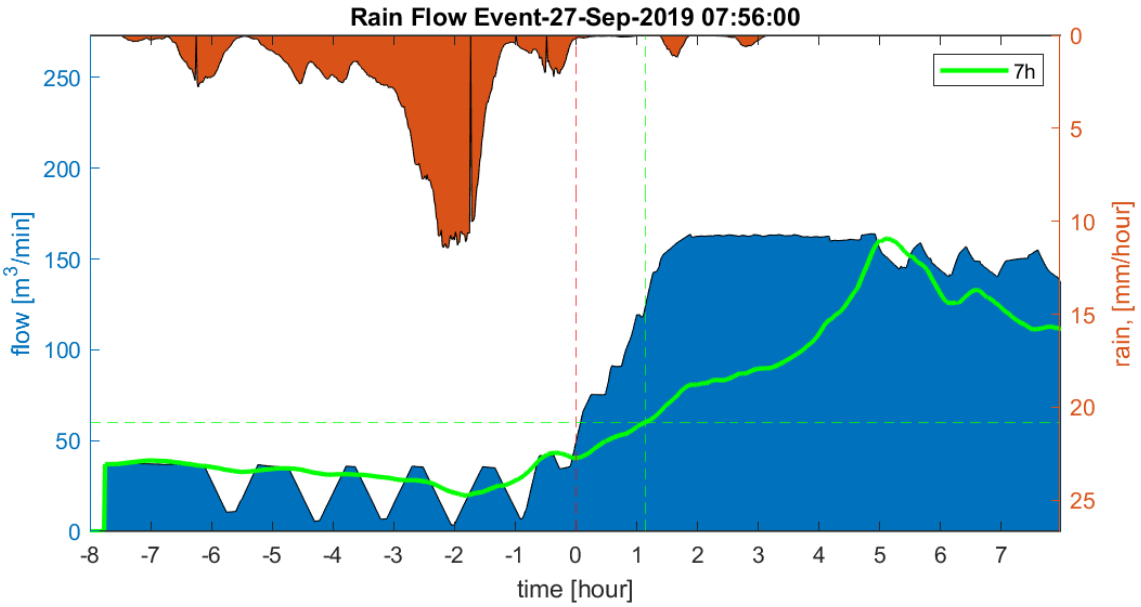


Figure 16: Determining timing of the flow shift for the 7h forecast at the evaluation period starting 27 September, (red vertical line). The timing is shown by the vertical green line and the threshold value defining the flow shift is shown by the horizontal green line. The rain is shown in orange and the wastewater flow is shown in blue.

If the forecast already was in a rain flow period by the definition given, this forecast was not evaluated within that evaluation period.

Flow shift timing was calculated by subtracting the timing of the flow shift for the real flow with the timing of the shift for the forecasted flow.The ideal value is zero and negative value means that the forecast gave indication of a flow shift ahead of the real shift. A negative value is considered to be better than a positive value since a delay would mean shorter time to adjust the treatment processes.

The relative volume was calculated by dividing the accumulated volume of the forecast with the corresponding volume of the real flow. The accumulation of the forecast began at its defined start point. The ideal value is one and corresponds to when the forecasted volume is equal to the real volume.

The mean absolute error (MAE) was calculated with equation 3 (Bowerman et.al 2005). The sum of all the absolute differences between the output signal y(t) and the prediction ŷ(t), at each instance t, is divided with the number n timesteps of the evaluation period. The ideal value is zero meaning that the forecasted flow and the real flow is equal at each time-step.

$$MAE = \frac{\sum_{t=1}^{n} |y(t) - \hat{y}(t)|}{n} \qquad (3)$$

4. RESULTS

The result section is divided into two parts. The first part presents the performance comparison between the linear regression model and the neural network together withthe comparison between the full catchment rain data set and the municipalities rain data set. The second part presents the performance comparison for the rain data sets with rain data from outside the catchment.

The results from the hyperparameter tuning are summarised in appendix section 8.2. In short, the hyperparameters that were tuned for each model were epochs, hidden layers, nodes and learning rate. Learning rate decay was discarded due to not significant improvements and the training strategy focused on a slow learning rate together with more epochs. Meaning that the training was relatively long and slow to ensure stable training. Hidden layer dropout rate and batch size were set to fixed values. The LRM was trained on 20 epochs while the NN was trained with 50-60 epochs because the NN needed longer time to converge to the lowest value of the given training set up.

The comparison between the different models was supported by a Wilcoxon rank sum test to examine if the differences shown in the median is significantly. A 5% significant level was used in this project. The Wilcoxon rank sum result is shown in Appendix 8.4.1 to 8.4.5 where each subsection presents the results for one evaluation parameter. If the p-value is above 5% then there is not enough evidence to assume that the difference is significant between two models.

4.1. Part 1: Comparison of LRM and NN

The result from the first part that focuses on the performance comparison between the linear regression model and the Neural Network using two types of delimitation of rain data are presented by each evaluation variable.

4.1.1. Flow shift timing

The results of how the forecasts in part 1 performed regarding the flow shift timing is summarised in Table 2. The different lead times are presented row wise, and the different models are presented column wise. The performance is given by the median timing error over the 31 evaluation periods together with the inter quartile range showing an indication of the spread. The optimal value is 0 or slightly negative given that the real flow has some lag in response. The table shows thatshorter lead times are better at timing the flow shift and between the 3 hours and 5 hours lead there is a substantial deterioration where the timing of the flow shift is predicted to occur later than the actual shift.

Table 2: Summary from part 1 of the flow shift timing result from the 31 evaluation periods in the form of median and inter quartile range (IQR). MUN = municipalities rain data, FC = full catchment rain data. LRM = linear regression model, NN = neural network

	MUN - LRM		MUN - NN		FC - LRM		FC - NN	
Lead time[h]	Median [h]	IQR [h]	Median [h]	IQR [h]	Median [h]	IQR [h]	Median [h]	IQR [h]
0.5	0.21	0.87	0.18	0.51	-0.14	1.02	0.22	0.50
1	0.30	0.98	0.38	0.62	-0.13	1.01	0.32	0.60
3	-0.16	0.91	-0.16	0.68	-0.53	1.14	-0.40	0.86
5	0.60	1.09	0.74	0.93	0.68	1.09	0.63	1.11
7	2.56	1.08	2.28	1.03	2.63	1.07	2.44	1.13

To get a better visualization of the differences between the different models the results summarised in Table 2 are presented in Figure 17 as box plots. The data is grouped by lead times of 3 hours, 5 hours and 7 hours and the different box plots within a group corresponds to the different models. The figure shows no substantial differences between the different models. There is slightly more spread amongst the models based on the full catchment rain file and slightly more spread for the LRM when predicting the 3-hour lead time compared to the NN. The Wilcoxon rank sum test with a 5% significant level (Appendix 8.4.1) showed no significant differences in the medians across the different model for each lead time.

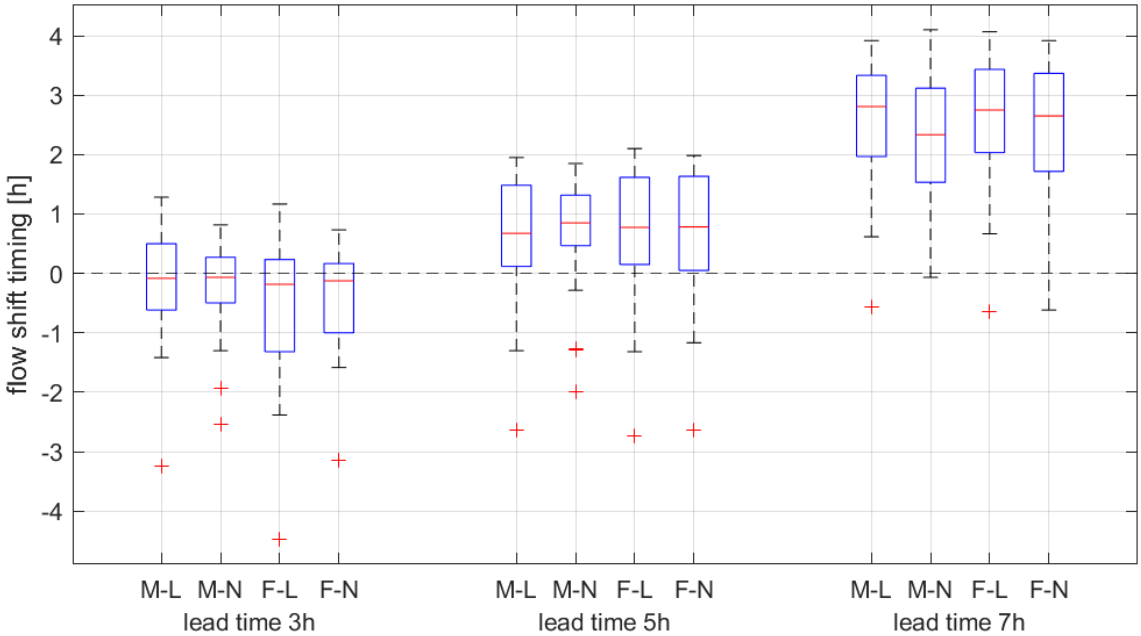


Figure 17: Box plot of the flow shift timing of different lead times (outer groups) and different models (inner groups). Whiskers are maximum 1.5 of the IQR. Outliers are shown as red dots. M = municipalities rain data, F = full catchment rain data. L = linear regression model, N = neural network

4.1.2. Relative Volume

The result from the relative volume evaluation is first summarized in Table 3 and then visualized as box plots in Figure 18. Both the table and the box plots show a clear difference between the LRM and the NN. This is reinforced in the Wilcoxon rank sum

test in Appendix 8.4.2 by showing significant differences in the medians among all lead times when comparing LRM and NN-based models. The LRM was able to almost perfectly represent the total volume that passed through during the evaluation period. The NN slightly underestimated the volume. All models showed low spread. The differences between the dataset showed a slight underestimation by the rain data delimited by municipalities. The significance of the difference between the two data set was only shown when using the Neural Network.

Table 3: Summary from part 1 of relative volume result from the 31 evaluation periods in the form of median and inter quartile range (IQR). MUN = municipalities rain data, FC = full catchment rain data. LRM = linear regression model, NN = neural network

	MUN - LRM		**MUN - NN**		**FC - LRM**		**FC - NN**	
Lead time[h]	Median	IQR	Median	IQR	Median	IQR	Median	IQR
0.5	1.00	0.04	0.94	0.03	1.00	0.04	0.97	0.04
1	1.00	0.04	0.93	0.03	1.01	0.04	0.96	0.04
3	1.00	0.06	0.91	0.05	1.04	0.09	0.96	0.07
5	0.99	0.07	0.90	0.07	1.04	0.10	0.95	0.08
7	0.95	0.07	0.87	0.07	1.00	0.08	0.91	0.08

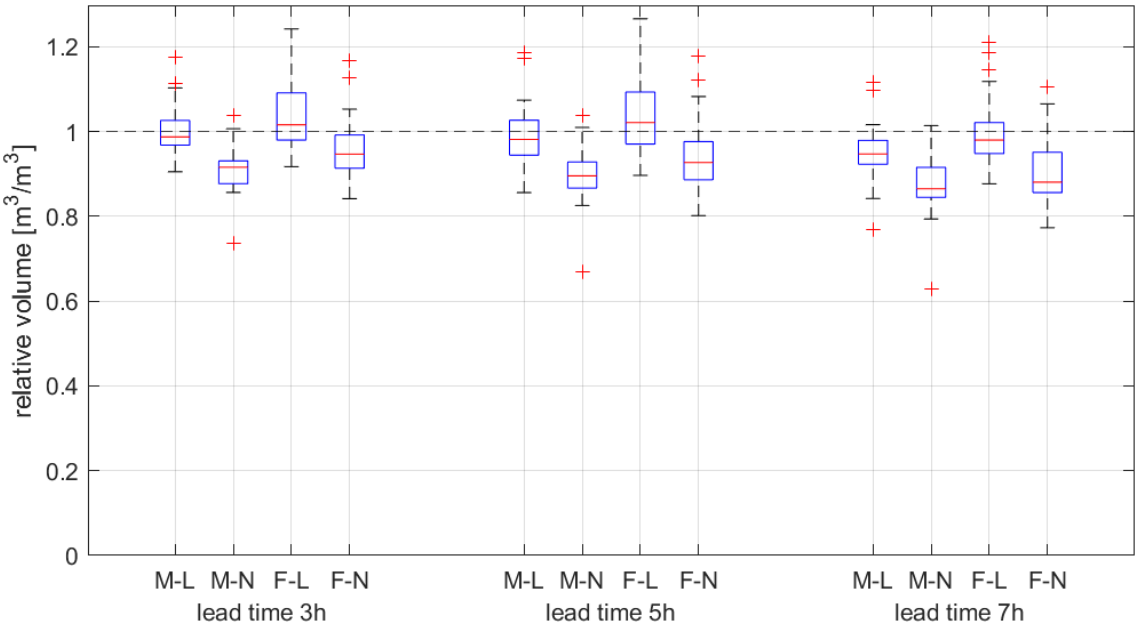


Figure 18: Box plot of the relative volume for different lead times (outer groups) and different models (inner groups). Whiskers are maximum 1.5 of the IQR. Outliers are shown as red dots. M = municipalities rain data, F = full catchment rain data. L = linear regression model, N = neural network

4.1.3. Mean absolute error (MAE)

The results from the MAE are summarised in Table 4 and visualized in Figure 19. A lower MAE value means that the model more precisely represented the flow. The Neural Network model seems to perform somewhat better, especially with longer lead

times. The Wilcoxon rank sum test (Appendix 8.4.3) showed that the medians of the MAE-results was significantly different between the LRM and the NN.

The differences between the two rain data sets are low. Visually, there is a slight improvement for the full catchment rain file, but this improvement was not significant. The MAE values of 3 and 5 hours lead times are quite similar followed bya larger jump in the 7-hour lead time. The mean flow rate of most evaluation periods is 80 m^3/min (because of the evaluation period definition). The lowest median MAE-value of 12.30 m^3/min is 15.4 % of the average mean flow rate of the evaluation periods. The highest median MAE-value of 28.49 m^3/min is 35.6 % of the average mean flow rate of the evaluation periods.

Table 4: Summary from part 1 of the mean absolute error result from the 31 evaluation periods in the form of median and inter quartile range (IQR). MUN = municipalities rain data, FC = full catchment rain data. LRM = linear regression model, NN = neural network

	MUN - LRM		**MUN - NN**		**FC - LRM**		**FC - NN**	
Lead time	Median	IQR	Median	IQR	Median	IQR	Median	IQR
h	m^3/min	m^3/min	m^3/min	m^3/min	m^3/min	m^3/min	m^3/min	m^3/min
0.5	15.00	3.68	13.63	3.15	14.75	3.40	12.30	3.04
1	15.14	4.04	14.59	3.39	14.77	3.73	13.50	3.79
3	21.20	5.44	17.81	4.63	20.28	4.60	17.42	5.27
5	23.48	6.81	19.39	4.91	22.89	6.01	19.95	5.65
7	28.77	9.18	24.29	6.70	28.49	8.01	25.79	8.08

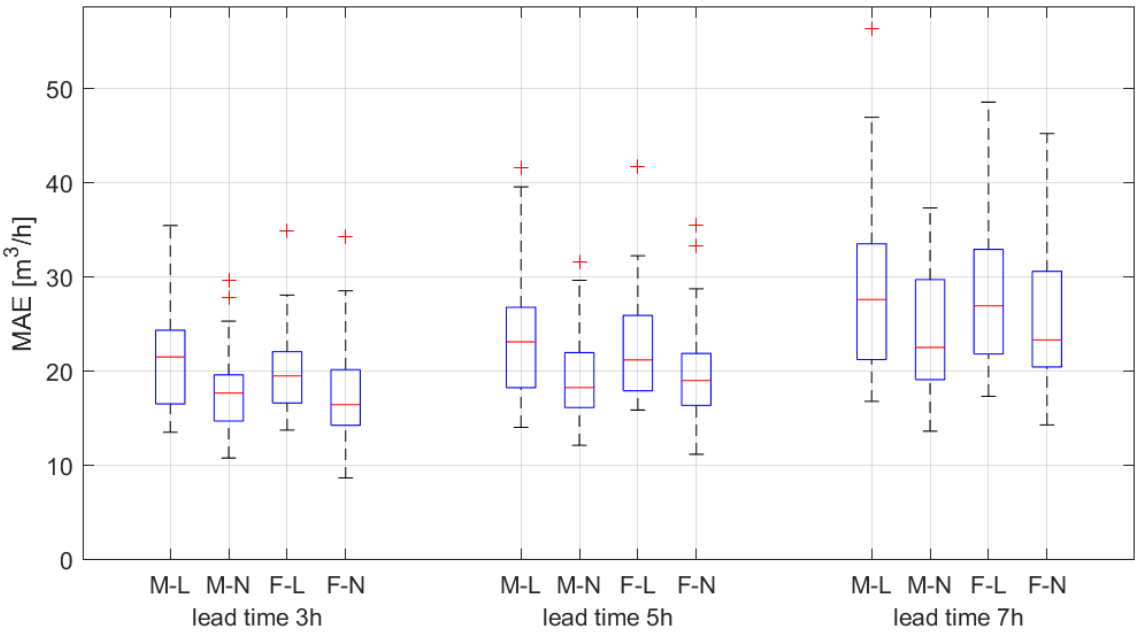


Figure 19: Box plot of the mean absolute error of the different lead times (outer groups) and different models (inner groups). Whiskers are maximum 1.5 of the IQR. Outliers are shown as red dots. M = municipalities rain data, F = full catchment rain data. L = linear regression model, N = neural network

4.1.4. Overall comparison

To represent the results from the large number of evaluation periods two new time serieswhere created. The first oneis the average of the 10 time series from the evaluations periods with the lowest rain intensity The second includes the average of the 10 periods with the highest rain intensity. The rain intensity is the maximal intensity found within the 6 hours leading up to the start point of the evaluation period (= the flow-shift). In Figures 20 and 21 these two types of periods are shown side by side.

In Figure 20, the LRM and NN are compared using the full catchment with a 5 h lead time. Their average is presented with the interval given by one standard deviation from the mean. The high intensity period on the right shows an extreme overshoot of the LRM together with a large variation. The NN on the other hand is somewhat underestimating the flow in the beginning of the period. Both models predict the retention of the flow quite well. For the low intensity period on the other hand there is a slight overestimation of the flow after the initial top. Both the LRM and the NN seems to answer more readily to high intensity rain than to low intensity, this is shown in the figures by the shifts in the low intensity period occurring later.

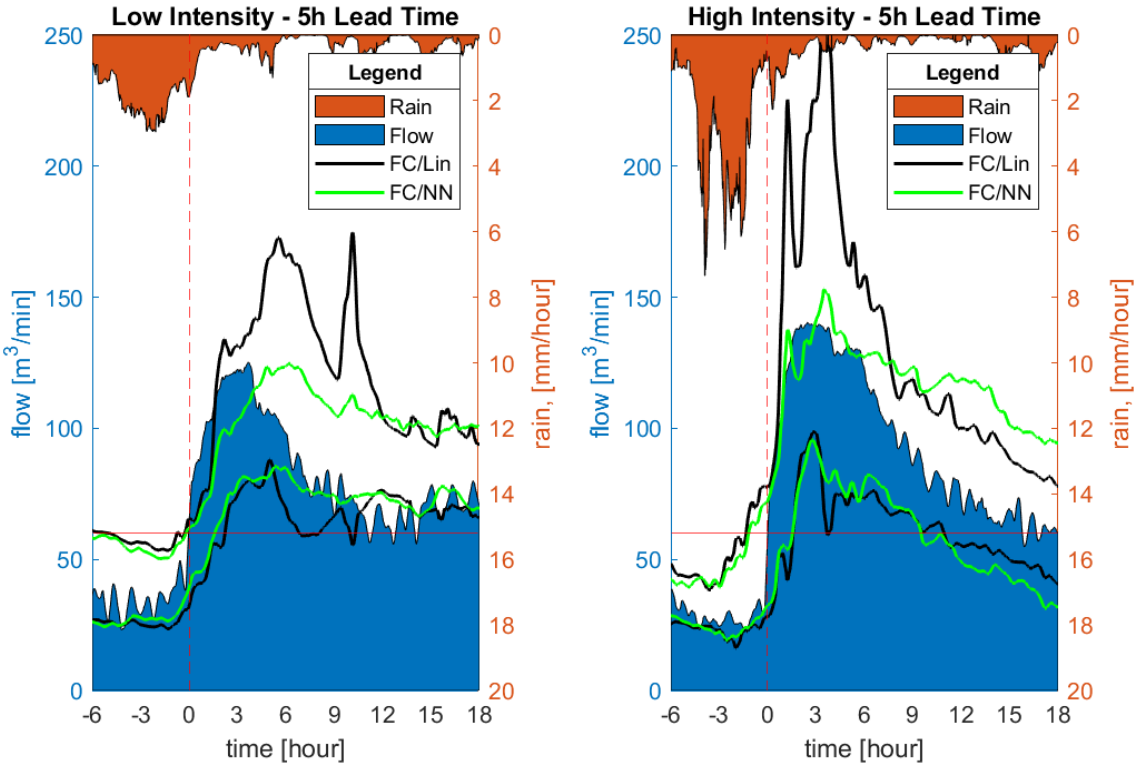


Figure 20: Evaluation period average where periods with low intensity rain is shown on the left plot and periods with high intensity rain is shown on the right. The models which are compered are linear regression model (Lin in figure) and neural network (NN) both using full catchment data. The lead time is 5 hours.

Figure 21 compares the LRM with the NN using the rain data delimited by municipalities. The most distinct change from Figure 20 is that the LRM overshoots have been dampened a bit. There also seems to be a bit less variation for the full catchment rain file. In comparison between LRM and NN the same trends are show. That is less overshoot and less variation by the NN.

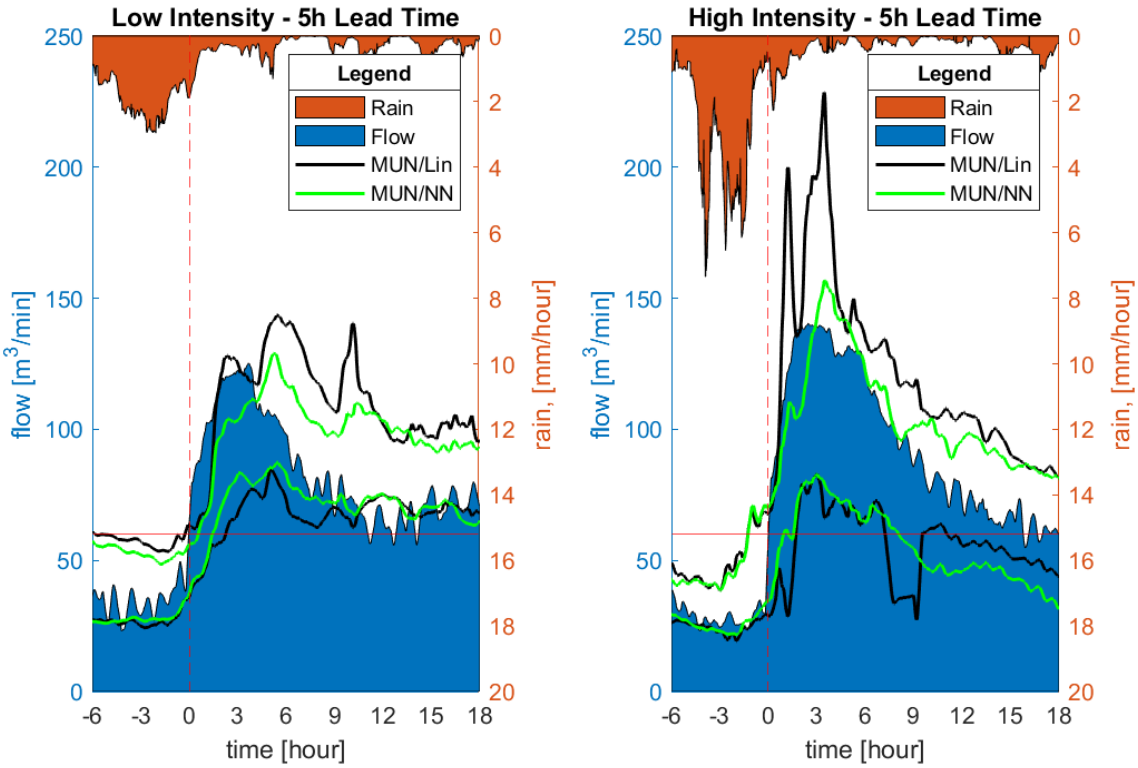


Figure 21: Evaluation period average where periods with low intensity rain is shown on the left plot and periods with high intensity rain is shown on the right. The models which are compered are linear regression model (Lin in figure) with neural network (NN) both using municipalities (MUN) rain data. The lead time is 5 hours.

4.2. PART 2: EXTENDED RAIN RADAR DATA

The results from the second part investigating if increased range of rain radar data could improve the performance when using longer lead times. Only NN was used for this investigation.

4.2.1. Flow shift timing

Table 5 summarises the results regarding flow shift timing for part 2. The models left to right use increasing amount of rain data. The table shows decreasing flow shift lags with increased data for the lead times of 5 and 7 hours. This is also true for 3 hours lead time where increasing negative value is shown with added rain data. The median flow shift timing has improved by about 20-30 minutes when comparing the rain data kept within the catchment with the farthest use of rain data of 30 kilometres. This is also shown in Figure 22 which reinforces the closing of the flow shift timing with the intended timing. Regarding the Wilcoxon rank sum test (Appendix 8.4.4) the only significant difference in median is shown between the original data set and the 30km data set at a 5-hour lead time. The differences between the models using the other two lead times lay slightly above the significance threshold of 5%.

Table 5: Summary from part 2 of the flow shift timing result from the 31 evaluation periods in the form of median and inter quartile range (IQR). MUN_20/30 = municipalities rain data with extended range (20 km and 30km), NN = neural network

	MUN - NN		MUN_20 - NN		MUN_30 - NN	
Lead time[h]	Median [h]	IQR [h]	Median [h]	IQR [h]	Median [h]	IQR [h]
3	-0.16	0.68	-0.33	0.71	-0.44	0.65
5	0.74	0.93	0.51	0.80	0.34	0.86
7	2.28	1.03	2.18	0.84	1.79	0.91

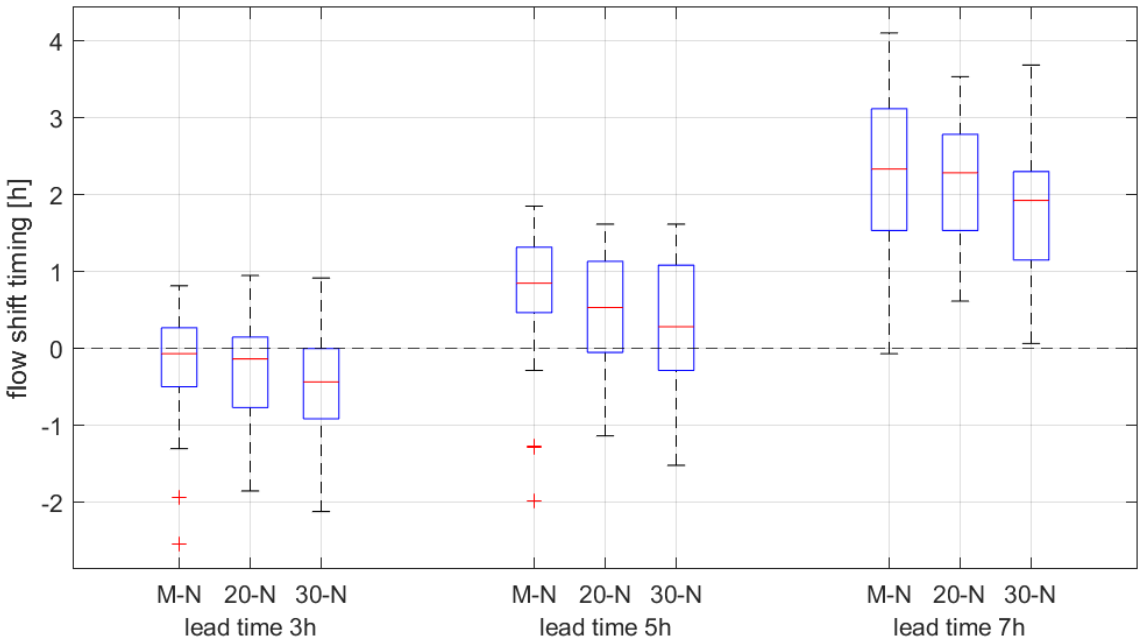


Figure 22: Box plot of the flow shift timing of different lead times (outer groups) and different models (inner groups). Whiskers are maximum 1.5 of the IQR. Outliers are shown as red dots. M/20/30 = municipalities rain data with extended range (20 km and 30km), N = neural network.

4.2.2. Mean absolute error (MAE)

The results from the MAE in part 2 are summarised in Table 6 and visualized in Figure 23. When comparing the 20 km rain data with the original municipality rain data set there seems to be no improvement at all, perhaps evena slight deterioration. But the 30 km rain data performs better for every lead time for both median and IQR. But this improvement is not proven to be significant when regarding the Wilcoxon rank sum test in Appendix 8.3.5. The median MAE-values range from 16.7 m^3/min to 24.6 m^3/min and are therefore 21% - 31% of the average mean flow rate of the evaluation periods.

Table 6: Summary from part 2 of the mean absolute error result from the 31 evaluation periods in the form of median and inter quartile range (IQR). MUN_20/30 = municipalities rain data with extended range (20 km and 30km), NN = neural network

	MUN - NN		MUN_20 - NN		MUN_30 - NN	
Lead time[h]	Median [m^3/min]	IQR [m^3/min]	Median [m^3/min]	IQR [m^3/min]	Median [m^3/min]	IQR [m^3/min]
3	17.81	4.63	17.26	4.78	16.66	4.56
5	19.39	4.91	19.54	6.06	17.51	4.55
7	24.29	6.70	24.63	7.49	22.40	6.20

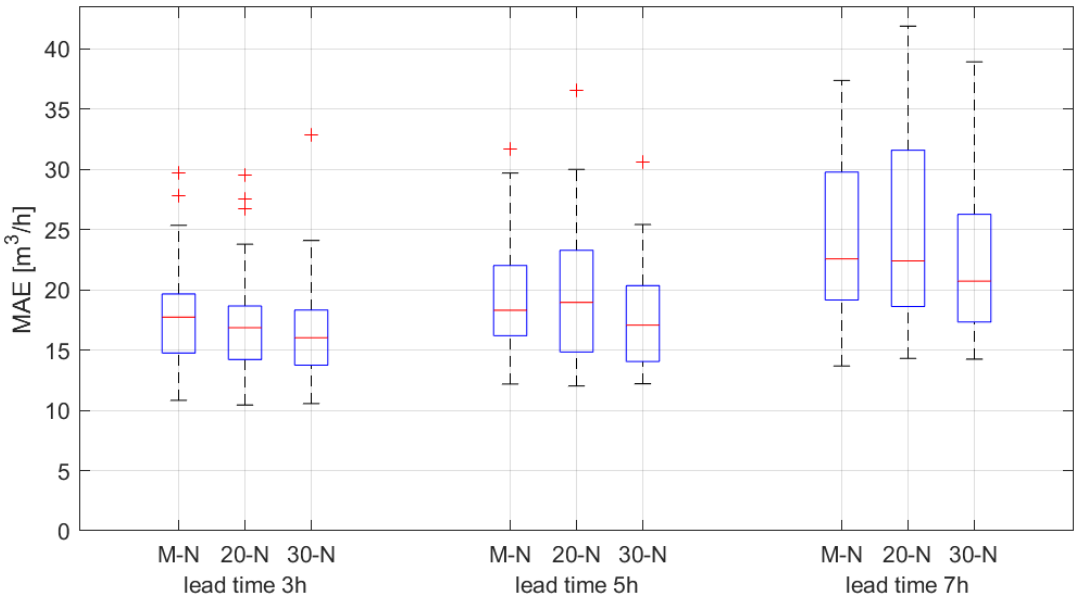


Figure 23: Box plot of the mean absolute error of the different lead times (outer groups) and different models (inner groups). Whiskers are maximum 1.5 of the IQR. Outliers are shown as red dots. Outliers are shown as red dots. M/20/30 = municipalities rain data with extended range (20 km and 30km), N = neural network.

4.2.4. Overall comparison

Looking at the evaluation period average in Figure 24 there are no large differences between the models when looking at the 5-hour lead time. There are slight gains for the flow shift timing with the 30km rain data that also got slightly higher flows. The timing is also clearly better for the high intensity period.

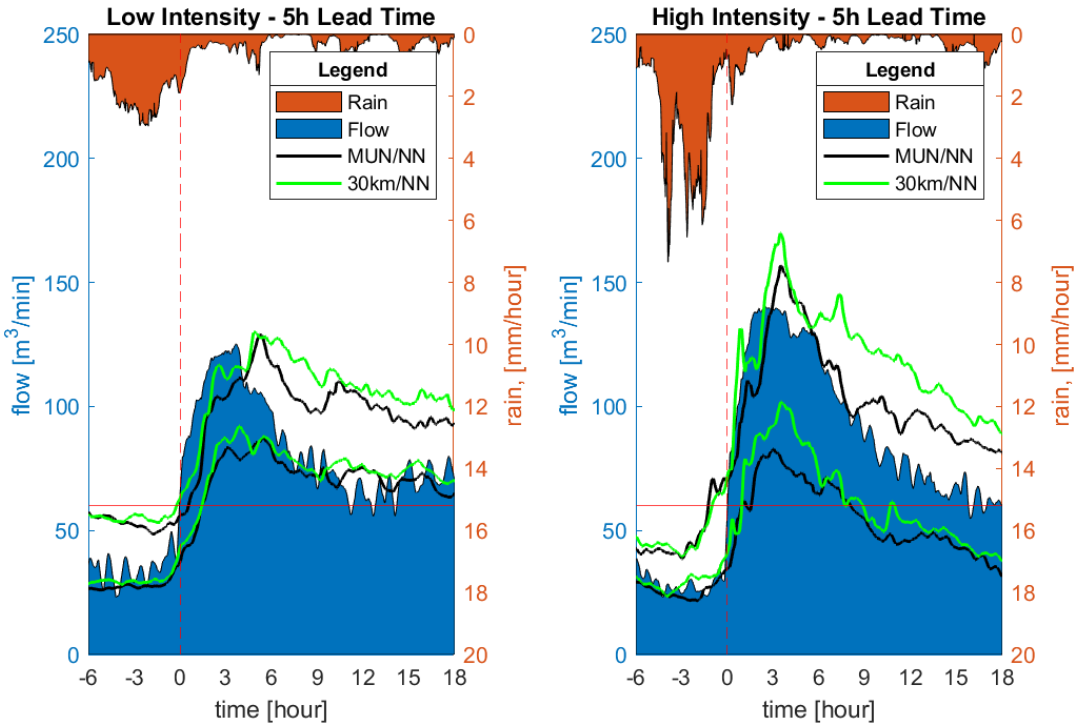


Figure 24: Evaluation period average where periods with low intensity rain is shown on the left plot and periods with high intensity rain is shown on the right. The models which are compered are neural network (NN) with municipalities rain data and neural network with 30 km radius rain data. The lead time is 5 hours.

5. DISCUSSION

The aims of this project were to compare the performance of a neural network with a linear regression model and to investigate what extent of X-band rain radar data improves the performance. The performance evaluation was case specific and evaluated by the ability of providing a forecast of the sewer flow into Avedøre WWTP with regards to the treatment process control system. The control system was a binary switch, either being set up to handle dry flow or to handle rain flow. Therefore, the most important information that the forecast should provide was the timing of when a flow switch occurs. Also, forecasts with longer lead time give Avedøre more opportunity to adapt its processes before the real flow switch occurs. In this project, the performance means how well the different models could forecast the 31 rain flow events with respect to three performance parameters. The first parameter was the timing of the shift from dry flow to rain flow. The flow shift threshold was set to 60 m^3/min. The second parameter was the approximation of the flow by calculating the relative volume of the forecasted flow compared with the real flow during that period. The third parameter were the precision of the forecast compared to the flow shown as the mean absolute error (MAE).

5.1 PART 1: Comparison of LRM and NN

The flow shift timing results from part one, which compared the NN with the LRM while using two different delimitations of rain data, showed that the median delay of the forecasts with 5 hours and 7 hours lead time were in the range of 0.60-0.74 hour and 2.28-2.63 hours. Given these ranges of delay the greatest possible lead time could be narrowed to 4.26-4.72 hours. Since the measured flow is somewhat delayed this range can be regarded as a slight overestimate. Comparing the models at lead times 3-5 hours from Figure 17 there were no significant difference in the median of the flow shift timing results. This show that both the NN and LRM were able to predict a flow response of sufficient magnitude equally well. It also showed that adding information about the spatial distribution within the catchment did not improve the prediction of the flow shift considering all the 31 evaluation periods. There might, however, exist individual periods where this added information of spatial distribution could be useful.

Findings from the performance regarding relative volume in Table 2 and Figure 18 showed that the LRM was good at approximating the flow when regarding the full evaluation period corrected by flow shift timing for all lead times. The NN on the other hand underestimated the volume and this became worse with longer lead times. The difference between the rain data sets were not as large as between the NN and the LRM but there was a trend that the municipalities rain data set performed worse than the full catchment rain data set.

The MAE performance results found in Table 3 and Figure 19 showed instead an opposite trend where the NN had a significant lower median value and lower IQR for all lead times compared with the LRM. This shows that the NN can more precisely predict

the real flow than the LRM. The difference in MAE performance between rain data set was not significant. When using the LRM, the municipalities rain set were slightly worse for all lead times compared to LRM/Full catchment combination. The NN combined with municipalities on the other hand were slightly better for the two longest lead times. The conclusion which may be drawn from these results is that the delimitation of the rain data by municipalities did not improve the forecast but if multiple data sets are to be used, the Neural Network seems to perform better. This could be because of the capability of the Neural Network to sort the input signals into multiple relationships that could be activate for different scenarios. But to prove this point the weights within the model should be examined to evaluate how the hidden layer was utilised.

These finding was also clearly shown in Figures 20 and 21 where the plot for the LRM forecast greatly overshoots the real flow at the early stage where as the NN is much more conservative and perhaps somewhat underestimating the initial rise in flow. The figures only show the first 18 hours from when a flow shift occurs and do not really show the point of when the LRM starts to underestimate the flow which it must do given the initial overshoot to get a relative volume close to 1. Since the precision is worse but the approximation is better for the LRM this means that the positive error cancels the negative error. The NN on the other hand trades the approximation of volume with being reasonably close to the real flow over the full period.

Also shown in Figures 20 and 21 was that forecasts where better at determining the flow shift at periods with high intensity rain. This is promising given that it is the high intensity rains that causes the most damaging rain flows.

5.2. PART 2: EXTENDED RAIN RADAR DATA

Part 2 of this project investigated if additional rain data from outside the catchment improved the forecast. Especially if forecast with longer lead times which failed at timing the shift in the previous part can shorten their delay. The additional two models were both Neural Networks and used the municipalities rain data set with added rain data up to 30km of the X-band radar station.

A slight improvement of the flow shift timing could be shown when the rain data range was increased (table 5 and Figure 22). With 5 hours lead time the improvement in the median using the 30km rain data set were 0.40 hours (24 min) and this improvement was significant when regarding the Wilcoxon rank sum test. With a 7 hour lead time the corresponding improvement in the median were 0.49 hours (29 min), but this was not significant. These findings indicate that it may be worthwhile using additional data that could train the model on secondary spatial relationship. The maximum range of rain data that could help to improve the forecast cannot be determined from the present results and need to be investigated further. Too far a distance might make the forecast more uncertain since the realisation of rain far away contributing to the sewer flow will be less likely.

The results of the MAE performance show almost identical results between the 20km rain data set and the original rain data set. Instead, there are a distinct improvement when regarding the 30km rain data set compared to the original. There is no clear explanation to why this is. It might be impacted by the hyperparameter optimization of the respective model or it might be just chance since the training of the models is not deterministic. But it at least shows that these additional information does not deteriorate the precision of the forecasts. When looking at Figure 24 the differences between the models are quite small. The 30 kilometres rain data set seems to have gotten a slight increase in the overall flow estimation compared with the original rain data set. The figure also suggest that the main improvement of flow shift timing occurred for the low intensity rain periods.

5.3. ERROR SOURCES

5.3.1. Flow data error

The flow data were used both as the output and input signal, making the model autoregressive. When the model trains using the flow data it assumes that the historic data is correct and thus tries to optimize the weights to fit the data overall. Since the flow data did not represent the natural flow of the sewers but instead were measured by the pumping rate at the inlet to Avedøre WWTP this resulted in the occurrence of high flow rates at periods which were regarded as dry flow periods and flow rates equal to zero at rain flow periods. This was enabled by the foregoing basin which could collect water before pumping it into the WWTP. The forecasts which try to recreate the flow will have both reduced precision in flow rate and an additional delay in response time. If the flow rate would instead have been measured at the basin inlet the uncertainties would have greatly decreased by eliminating the randomness which the pumping control strategy imposed. But non-linear processes upstream would still have an impact by providing random noise to the flow time-series.

5.3.2. Rain data error

The rain data measured by X-band radar also had errors tied to it, just as the theory about X-band radar suggests. The main error which was shown in this project was the occasional attenuation of the signal which caused gaps in the time series data. Also, the decreasing amount of data points further the distance from the radar station suggests that the error was higher for the outer data sets. But it can also be a consequence of more clutter and very low rates of precipitation is being registered close to the X-band radar. The direction of the rain measurement relative to the radar also gave different amount of data indicating different error sources in different directions.

If the error was just a dampening factor on the data which would remain roughly the same size within the same data set this would not be a problem. This is because every single data is given initially an individual weight by the model and can therefore correct the impact of that dampening. If the error varies across time the instances where the error occurs will cause the forecast to undervalue the impact of the rain and therefore risk missing the flow shift.

The high-resolution data were aggregated into rain data files that corresponds to larger areas. The comparison was done between two sets of aggregated data of different resolution. Even though it might point towards that some resolution is better than the other it does not show if the high resolution is necessary for this purpose. But since the quality of the aggregates is based on each data point there could still be a reason to have a local and accurate radar that can measure a high intensity rain fall.

5.4. APPLICATION TO AVEDØRE WWTP

The forecasts in this project were made with regards to the practical application of informing the wastewater treatment plant of Avedøre of when the sewer flow is affected by rain such that the efficiency of the treatment processes is deteriorated or even harmed. Given that the control strategy is binary the most important variable of the forecast to consider is when the sewer flow rate threshold is likely to be passed. The results in this project suggests that a flow shift could be forecasted with a lead time in the range of 3-5 hours. The uncertainty increases of course with longer lead times making it wise to utilise multiple lead times to correct the information as the flow realises itself. Including a larger range of rain data seems to increase the ability to make longer forecasts.

The implementation of a neural network-based forecast into the operational function of a wastewater treatment plant requires the input data sources to be on-line and quickly transferred to the neural network so it can compute a forecast as soon as possible. The short computational time is one of the strengths of a neural network and this helps to reduce the delay between measurement and forecast.

After the forecast has been made there is a second problem of how to interpret the information into a decision in the treatment control. The interpretation can be made by an operator or directly by a computer. It is important that this interpretation is made with respect to what the forecast can show and what the intended goals with the treatment control are. The flow rate that the neural network predicts is made with a certain amount of uncertainty and must be considered when taking further decisions. The neural network used in this project also provided the forecast with a standard deviation of the flow rate and gives some information of how certain the model is for each prediction value, but this was not considered in this project.

In this project only a sub selection of rain flow periods was used to evaluate the forecast performance. But to fully analyse the utility of the forecast made by a neural network it is necessary to formulate a distinct definition how the forecast will be interpreted continuously, and this definition will also be a factor of the performance. By evaluating the performance based on the definition it is interesting to see the number of times that the real flow is of a magnitude that would necessitate a shift in the operational mode but that the forecast did not detect or was substantially missed.

Another aspect that was not analysed in this thesis was the occurrence of false negatives that means the times when the forecast predicts a future rain flow that is not realised. The cost of a false negatives is the added energy and resource consumption made by an

operational shift in the treatment processes. To many false negatives might also reduce the trust in the forecast by the operator and might lead to the dismissal of the correct prediction. The optimal interpretation of the forecast should minimize the number of false negatives while also maximizing the number of flow-shifts accurately predicted ahead of time.

Phenomena such as changing climate, water usage and the occasional extreme weather situation will impact the nature of the rain and wastewater flow. If a neural network or another machine learning model would be implemented the initial training on historical data would perhaps not age well enough with time. Then it would be necessary to occasionally retrain the model on the newer data or implement a self-learning mechanism which updates itself.

5.5. FUTURE STUDIES

One aspect that was not investigated in this project were the number of false alarms which may have been activated if the forecast would have indicated a flow shift when it did not occur. Given the nature of the flow data this would of course have been somewhat hard to detect within the scope of this project since it is possible that the basin at the WWTP would have acted as a buffer and what would have been an intense flow over a shorter time becomes instead a small increase over a longer time period. The results in this project were not able to show that utilising spatially delimitated rain data improved the forecast in any substantial manner. Since this project did not investigate the rain fall events individually it was hard to determine if there were rain fall events that may have varied greatly over the catchment. Therefore, it would be interesting to see a study which evaluates delimitation of X-band rain radar data of different resolution with regards to rain falls of high intensity that vary across the catchment.

The need for qualitative assessments in this project with regards to model training and performance evaluation reduced the ability to test more model structures and evaluate more rain flow periods. To utilise the efficiency of the machine learning models it would be optimal if the performance parameters would be directly implemented into the model. Since the loss function only provides a performance estimate over the full time series adding additional parameters might secure the optimal training for the intended goal. If this process is properly automatized this would allow for more training sessions to take place, thus increasing the likelihood that the optimal model is found.

6. CONCLUSION

The results in this project showed no substantial differences between the Neural Network and the Linear Regression model at determining the time of the flow shift, which corresponds to the response time of the catchment. When using a lead time of 5 hours the median flow shift timing of both models was delayed by about three quarters of an hour. This indicates that the longest possible lead time for forecasting Avedøre WWTP is slightly above 4 hours. The neural network was better at predicting the flow at one single instance whereas the linear regression model was better at approximating the flow over the full period.

Using X-band rain data from areas up to 30 km from the radar station improved the median time shift difference in the range of 20-30 minutes indicating that increasing the rain data range might be useful when forecasting the sewer flow with longer lead times. Determining the maximum range of rain data which helps to improve the performance needs further investigation.

The delimitation of the rain data within the catchment area showed no improvements to the forecast compared to using all rain data as a single aggregate when considering all 31 rain flow events. But the results showed that the Neural Network were better at creating forecasts made from multiple rain data files compared with the Linear Regression Modell.

An additional finding of interest was that the flow shift of high intensity rain periods was better timed than low intensity rain periods. This increases the benefits of forecasting since the most problematic flows were forecasted with higher precision.

7. REFERENCES

Published

Achleitner, S., Fach, S., Einfalt, T. and Rauch, W., 2008. Nowcasting of rainfall and of combined sewage flow in urban drainage systems. *Water Science & Technology*, 59(6), pp.1145-1151. 10.2166/wst.2009.098

Bennet, R.J., 1979. *Spatial Time Series*. London: Pion limited

Beven, K.J., 2001. *Rainfall-runoff modelling.* Chichester: John Wiley & Sons LTD

Bowerman, B.L., O'Connell, R.T.and Koehler A.B., 2005. *Forecasting, Time Series and Regression.* 4th ed. Oxford Ohio: Brooks/Cole

Carlsson, B., Åmand, L., Hallin, S., 2019. *Tillämpad reglerteknik och mikrobiologi i kommunala reningsverk*. Uppsala: Svenskt Vatten AB.

Dahlström, B., 2006. *Regnintensitet i Sverige ï en klimatologisk analys*. Norrköping: Svenskt Vatten AB Available at: http://vav.griffel.net/filer/VA-Forsk_2006-26.pdf [Accessed 10 Mars 2021].

Diggle, P.J. and Chetwynd, A.G., 2011. *Statistics and Scientific method*. Oxford: Oxford University Press.

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.E., Qurimbach, M., Vaes, G. and Vieux, B., 2004. Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology*, 299(3-4), pp. 186-202. https://doi.org/10.1016/j.jhydrol.2004.08.004

Faust, F. and Nelsson, P., 2020. *Using X-band Radar with a Neural Network to Forecast Combined Sewer Flow*. M.Sc.Eng. Lund University. Available at: http://lup.lub.lu.se/student-papers/record/9019570 [Accessed 8 Mars 2021].

Kartalopolous, S.V., 1996. *Understanding Neural Networks and Fussy Logic*. Piscataway: IEEE Press.

Kirchner, J.W, 2009. Catchments as Simple Dynamical Systems: Catchment Characterization, Rainfall-Runoff Modeling, and Doing Hydrology Backward. *Water Resources Research*, 45(2). 10.1029/2008WR006912

Lengfeld, K., Clemens, M., Münster, H. and Ament, F., 2014. Performance of high-resolution X-band weather radar networks – the PATTERN example. *Atmospheric Measurement Techniques*, 7, pp. 4151-4166. 10.5194/amtd-7-8233-2014

Schellart, A.N.A., Shepherd, W.J. and Saul, A.J., 2012. Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale. *Advances in Water Resources*, 45, pp. 65-75. https://doi.org/10.1016/j.advwatres.2011.10.012

South, N., Hashemi, H., Olsson, L., Hosseini, S.H., Aspegren, H., Larsson, R., Berndtsson, R., Das, R., Marmbrandt, A., Olsson, J. and Persson, A., 2019. *Väderradarteknik inom VA-området ï test av metodik*, Bromma: Svenskt Vatten Utveckling.

Svenskt Vatten, 2013. *Avloppsteknik 1: allmänt*. Stockholm: Svenskt Vatten AB.

Thorndahl, S., Einfalt, T., Willems, P., Ellerbæk-Nielsen, J., ten Veldhuis, M.C., Arnbjerg-Nielsen, K., Rasmussen, M.R. and Molnar, P., 2017. Weather radar rainfall data in urban hydrology. *Hydrology and Earth System Sciences*, 21(3), pp. 1359-1380. 10.5194/hess-21-1359-2017

van de Beek, C.Z., Leijnse, H., Stricker, J. N. M., Uijlenhoet, R. and Russchenberg, H. W. J., 2009. Performance of high-resolution X-band radar for rainfall measurement in The Netherlands. *Hydrology and Earth System Sciences*, 14(2). 10.5194/hessd-6-6035-2009

VA SYD, 2017. *Åtgärdsplan för Malmös avloppsledningsnät,* Malmö: VA SYD. Available at: https://www.vasyd.se/Nyheter/Avlopp/atgardsplan-malmo [Accessed 8 Mars 2021].

Veolia-1, 2016. *Increase hydraulic capacity ï reduced sludge escape & footprints*. Available at: http://technomaps.veoliawatertechnologies.com/processes/lib/municipal/3780-2_Veolia_STAR_Utility-solutions_WWM.pdf [Accessed 10 Mars 2021].

Veolia-2, 2016. *Improved operation and quality ï reduced operating costs*. Accessed at: 3779-1_Veolia_STAR_Utility-solutions_WWM.pdf (veoliawatertechnologies.com) [Accessed 10 Mars 2021].

Unpublished

Carlsson, B. and Lindholm, C., 2019. *Linear Regression*, Department of Information Technology, Uppsala University, unpublished. Accessed through personal communication with the authors.

Greenhill, B., 2016. *Oplandsanalyse BIOFOS,* Copenhagen: BIOFOS. Accessed through personal communication with representative of BIOFOS.

Online

BIOFOS, 2021. *BIOFOS ï Om os.* Available at: https://biofos.dk/om-os [Accessed 9 Mars 2021].

Furno, 2021. *Furno väderradar modell WR-2100.* Available at: http://www.furuno.se/produkter/vaederradar/ [Accessed 9 Mars 2021].

Guttag, J., 2017. *Introduction to Machine Learning.* [video] Available at: https://www.youtube.com/watch?v=h0e2HAPTGF4&t=2213s [Accessed 9 Mars 2021].

Krüger, 2021. *Avedøre, Copenhagen, Biofos.* Available at: https://www.kruger.dk/teknologier-og-service/digital-services-aquavista/referencer-online-styring/avedoere-copenhagen [Accessed 10 Mars 2021].

Larochelle, H., 2013. *Neural networks [2.2] : Training neural networks - loss function.* [video] Available at: https://www.youtube.com/watch?v=PpFTODTztsU&t=129s [Accessed 9 Mars 2021].

Sanderson, G., 2017. *Gradient descent, how neural networks learn: Deep learning, chapter 2.* [video] Available at: https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=3 [Accessed 9 Mars 2021].

SMHI, 2015. *Nederbördsintensitet*. Available at: https://www.smhi.se/kunskapsbanken/meteorologi/nederbordsintensitet-1.19163 [Accessed 10 Mars 2021].

SMHI, 2020. *Regn.* Available at: https://www.smhi.se/kunskapsbanken/meteorologi/regn-1.648 [Accessed 10 Mars 2021].

Starmer, J., 2017. *StatQuest: Maximum Likelihood, clearly explained!!!*. [video] Available at: https://www.youtube.com/watch?v=XepXtl9YKwc [Accessed 9 Mars 2021].

Starmer, J., 2020. *Neural Networks Pt. 3: ReLU In Action!!!*. [video] Available at: https://www.youtube.com/watch?v=68BZ5f7P94E&t=94s [Accessed 9 Mars 2021].

Starmer, J., 2021. *Neural Networks Pt. 4: Multiple Inputs and Outputs*. [video] Available at: https://www.youtube.com/watch?v=83LYR-1IcjA&t=317s [Accessed 9 Mars 2021].

8. APPENDIX

8.1. RAIN DATA SUMMARY

Table 7: Summary of each rain data file used as input signals in this project.

Data files	radar data points	sum rain mm	minutes with rain	<0.5 mm/h	0.5 - 4 mm/h	>4 mm/h
name	**nr**	**mm**	**min**	**%**	**%**	**%**
Full Catchment	926	3.57E+03	1.28E+06	92.6	6.7	0.7
Albertslund	111	3.82E+03	2.57E+05	66.3	28.7	5.0
Ballerup	106	3.80E+03	9.01E+05	90.4	8.3	1.3
Brøndby	94	4.10E+03	2.83E+05	68.1	26.8	5.0
Glostrup	65	4.58E+03	3.45E+05	72.9	22.2	4.8
Herlev	70	4.24E+03	1.10E+06	92.1	6.8	1.2
Hvidovre	31	4.60E+03	2.41E+05	61.1	32.1	6.8
Høje-Taastrup	225	2.33E+03	2.17E+05	71.0	26.3	2.8
Ishøj	135	3.21E+03	2.55E+05	68.7	27.4	3.9
Rødvre	37	5.35E+03	5.28E+05	77.0	19.4	3.5
Vallensbæk	52	2.13E+03	6.42E+05	91.7	7.3	1.0
NV20	572	2.13E+03	6.42E+05	91.7	7.3	1.0
VN20	462	2.25E+03	2.12E+05	73.7	23.0	3.3
VS20	352	1.96E+03	1.85E+05	72.2	24.8	3.0
SV20	631	9.79E+02	1.59E+05	82.0	16.7	1.3
SE20	622	1.69E+03	7.20E+05	93.7	5.7	0.6
ES20	715	1.14E+03	5.72E+05	94.1	5.5	0.4
EN20	659	1.17E+03	8.52E+05	95.9	3.8	0.3
NE20	635	1.41E+03	7.51E+05	94.6	5.0	0.4
NV30	846	7.04E+02	9.57E+04	79.2	19.0	1.8
VN30	811	6.91E+02	8.76E+04	77.7	20.2	2.1
VS30	846	3.70E+02	4.96E+04	79.6	18.3	2.1
SV30	859	2.91E+02	4.81E+04	84.1	13.9	1.9
SE30	811	6.20E+02	9.18E+04	81.9	16.2	1.8
ES30	795	4.11E+02	8.89E+04	86.2	12.9	0.8
EN30	653	4.11E+02	7.25E+04	83.9	15.1	1.0
NE30	737	4.36E+02	8.37E+04	84.6	14.4	1.0

8.2. HYPERPARAMETER TUNING SUMMARY

Summary of the hyperparamter tuning process.

- **Layers & Nodes:** Generally increasing nodes could improve the loss value. But especially when training MUN20/MUN30 models the training loss became far better while the validation loss became worse. This indicates that the models were overfitted to the training data set and could not generalize enough to predict on the validation data. Adding an additional layer did not seem to improve the loss much but both 1 and 2 layers were used in the final training of the different models. The conclusion was to test nodes and layers for each model.
- **Number of Epochs**: The epochs were chosen with margin to ensure that the loss values reached an equilibrium. Linear regression model required far less epochs than the neural network.
- **Learning rate**: The linear regression model could be trained with far higher learning rate than the neural network. High learning rates could give low loss values but with an unstable training progression. Low learning rates led to a higher loss but a stable training progression, even with large number of epochs the loss equilibrium tended to remain somewhat higher. Learning rate became the hyperparameter that was most intensively tested for each model.
- **Learning rate decay**: Adding learning rate decay did not seem to improve the loss value any substantially. What could happen were that the training progress became a bit more unstable (jumpy). The conclusion was to turn learning rate decay off and instead start the training with a low enough learning rate.
- **Learning rate decay steps**: Same reasoning as learning rate decay.
- **Hidden layer dropout rate**: No substantial difference was detected. For the neural network a value of 0.4 was set for all models.
- **Batch size**: Larger batch size decreased the time to run each epoch. Larger values gave slightly worse loss value. No substantial difference for batch size 16 - 64. The conclusion was that all models were to have the same batch size of 32.

Table 8: Summery of the hyperparameter used to train each model and their resulting loss value and structure.

	MUN LRM	MUN NN	FC LRM	FC NN	MUN20 NN	MUN30 NN
Hidden Layers	-	1	-	2	1	1
(HL)Nodes	-	128	-	64	64	64
Epochs	20	50	20	60	60	60
Learning rate	1E-4	5E-5	1E-4	5E-5	5E-6	5E-6
Learning rate decay	0	0	0	0	0	0
Learning rate decay steps	-	-	-	-	-	-

Hidden dropout rate	-	0.4	-	0.4	0.4	0.4
Batch size	32	32	32	32	32	32
Training loss	1.011	0.9687	1.016	0.9683	1.031	1.024
Validation loss	1.029	0.9994	1.029	0.9777	1.07	1.077
Input signals	144	144	36	36	240	336
Output signals	10	10	10	10	6	6
Total amount of weights	1 450	19 850	370	7 178	15 814	21 958
Training time	5m 59s	25m 45s	5m 50s	29m 47s	22m 35s	23m 9s

8.3. EVALUATION PERIODS SUMMARY

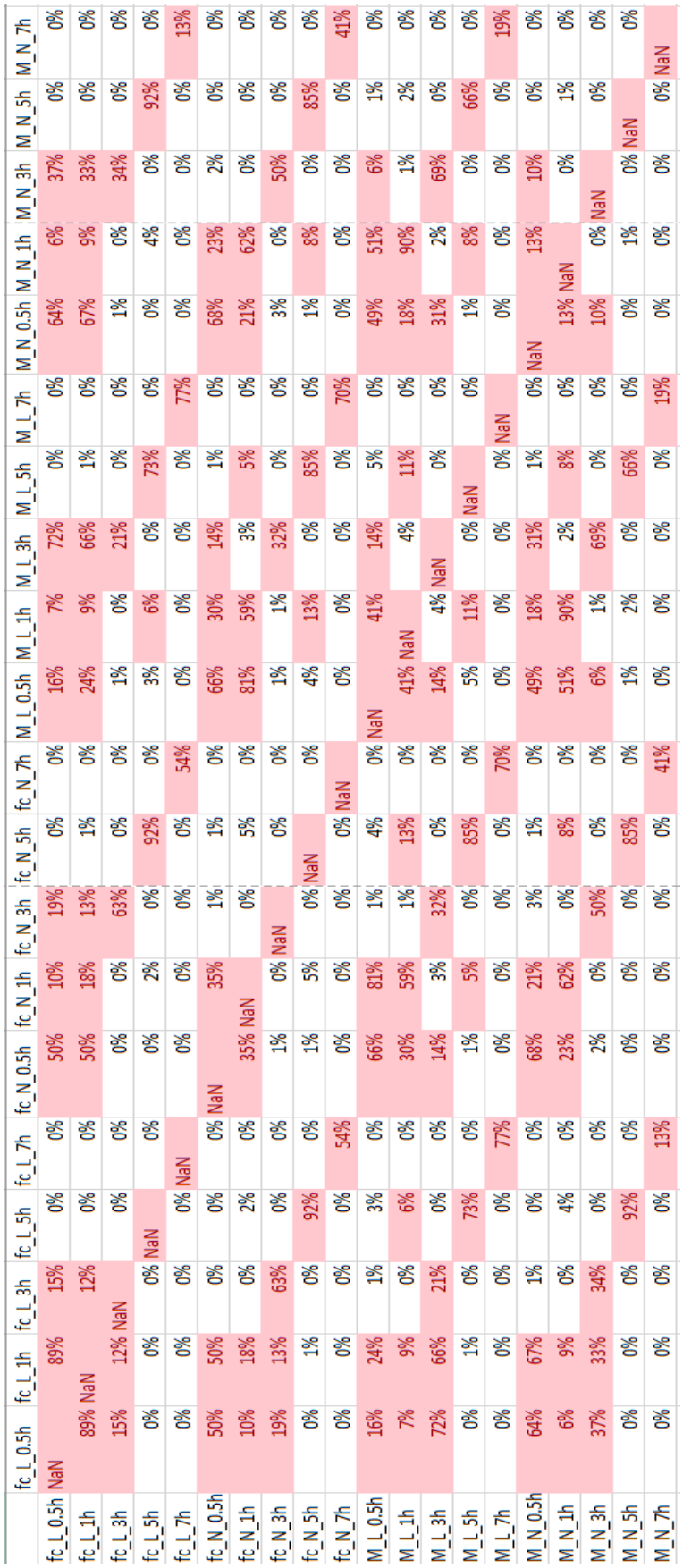
Table 9: Summary of the 31 evaluation periods. T = duration in hours. P-6h_sum = accumulated rain 6h before the period start. P-6h_max = max rain intensity 6h before the period start. Q-6h_mean = mean flow rate 6h before the period start. Q_mean = mean flow rate of the evaluation period. Q_max = max flow rate of the evaluation period.

Evaluation periods			**T**	**P-6h sum**	**P-6h max**	**Q-6h mean**	**Q mean**	**Q max**
nr	Start	Stop	h	mm	mm/h	m^3/min	m^3/min	m^3/min
1	2017-03-18 05:40	2017-03-19 06:21	24.7	10.4	5.4	42.9	80.0	156.5
2	2017-03-20 19:42	2017-03-24 06:01	82.3	10.0	5.7	57.1	80.0	184.0
3	2017-04-15 08:41	2017-04-17 18:25	57.8	14.2	6.0	42.5	80.0	201.6
4	2017-04-29 03:58	2017-05-04 14:24	130.5	20.7	6.9	38.7	80.0	225.5
5	2017-06-07 00:44	2017-06-10 08:44	80.0	10.0	5.0	44.6	71.1	188.9
6	2017-07-21 01:11	2017-07-22 18:31	41.4	10.0	4.4	37.2	69.5	195.3
7	2017-07-23 15:08	2017-07-25 17:33	50.4	14.6	7.5	32.4	81.2	193.1
8	2017-07-30 07:13	2017-08-01 15:47	56.6	12.0	11.8	26.1	80.0	175.7
9	2017-08-03 11:04	2017-08-09 01:21	134.3	10.0	4.6	29.9	80.3	225.9
10	2017-11-02 02:08	2017-11-03 23:12	45.1	10.0	9.0	53.7	80.5	179.9
11	2017-11-22 02:56	2017-12-02 12:56	250.0	10.0	6.3	50.1	80.3	216.1
12	2018-01-28 01:20	2018-01-29 23:54	46.6	10.0	4.1	47.8	82.5	181.3
13	2018-04-25 09:07	2018-04-27 20:51	59.8	22.6	14.7	31.2	79.6	194.4
14	2018-10-23 05:00	2018-10-24 02:41	21.7	16.6	4.8	31.7	79.9	181.4
15	2018-11-11 08:22	2018-11-11 19:51	11.5	11.9	4.3	25.5	80.0	176.2
16	2018-11-12 16:11	2018-11-13 12:08	20.0	12.5	7.3	40.7	80.0	177.0
17	2019-05-17 10:25	2019-05-18 08:15	21.9	14.3	11.1	24.3	80.0	203.3

18	2019-05-21 03:57	2019-05-22 02:58	23.0	16.6	16.9	35.6	80.0	206.8
19	2019-05-26 19:51	2019-05-27 06:53	11.1	14.5	10.9	42.7	80.0	145.3
20	2019-06-12 05:51	2019-06-16 11:56	102.1	21.1	23.5	32.6	80.0	223.3
21	2019-07-08 08:30	2019-07-09 11:38	27.2	16.3	8.5	22.4	80.0	209.4
22	2019-07-31 08:29	2019-07-31 18:47	10.3	13.8	4.9	31.3	80.0	142.3
23	2019-08-13 08:08	2019-08-14 13:27	29.3	13.6	16.4	23.0	67.0	198.2
24	2019-08-17 09:39	2019-08-17 17:57	8.3	13.2	6.2	22.4	80.0	230.1
25	2019-09-01 01:11	2019-09-01 21:54	20.7	17.9	29.4	33.8	80.0	206.1
26	2019-09-10 07:41	2019-09-14 04:12	92.5	14.0	10.0	32.3	80.0	204.7
27	2019-09-27 07:56	2019-09-28 19:07	35.2	18.6	11.4	23.8	80.0	204.7
28	2019-10-08 09:16	2019-10-09 04:50	19.6	14.6	9.0	25.0	80.0	210.5
29	2020-05-23 01:29	2020-05-23 14:06	12.6	10.4	5.8	36.8	80.0	206.9
30	2020-06-05 02:39	2020-06-06 22:13	43.6	33.2	16.2	33.4	80.1	216.8
31	2020-06-19 08:17	2020-06-22 10:21	74.1	25.2	28.3	24.9	80.0	188.5

8.4. Wilcoxcon rank sum test

8.4.1 Part 1-Flow shift timing

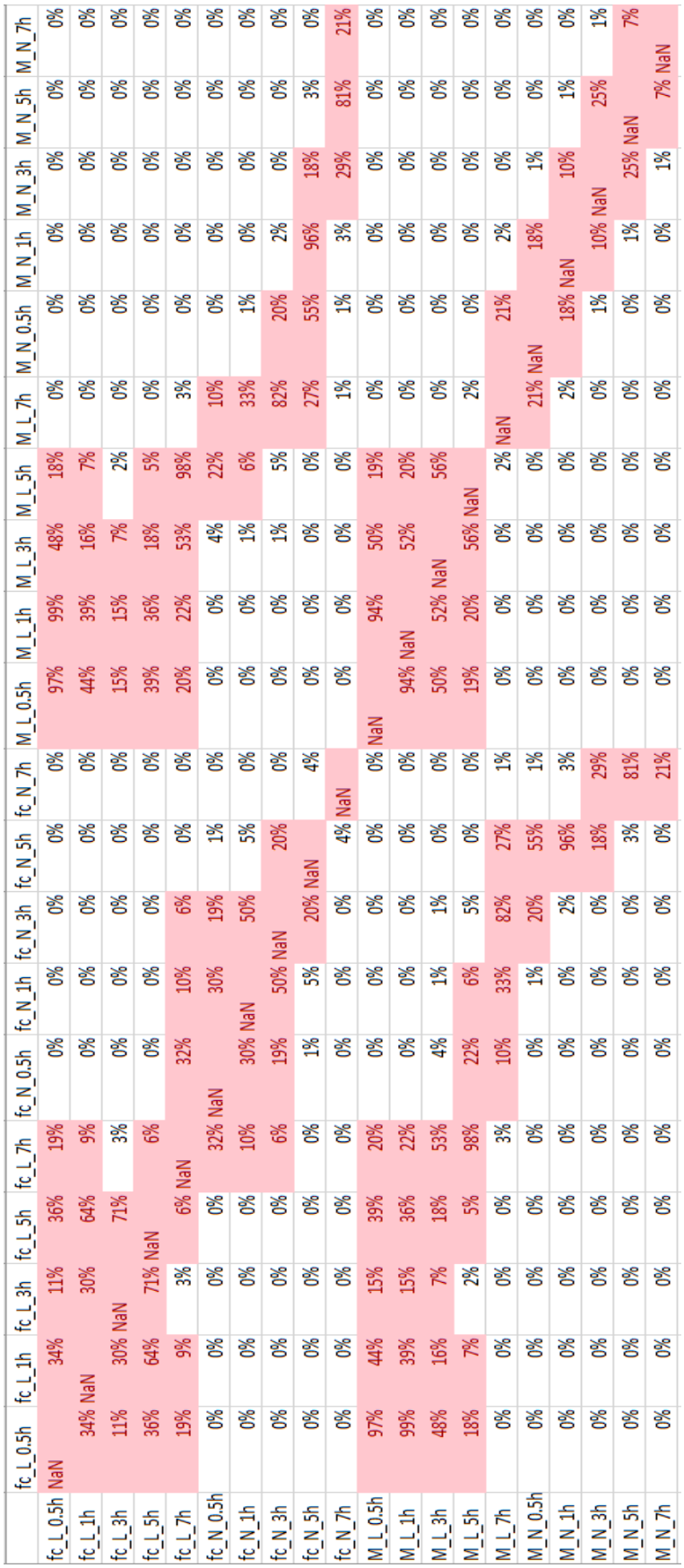


	fc_L_0.5h	fc_L_1h	fc_L_3h	fc_L_5h	fc_L_7h	fc_N_0.5h	fc_N_1h	fc_N_3h	fc_N_5h	fc_N_7h	M_L_0.5h	M_L_1h	M_L_3h	M_L_5h	M_L_7h	M_N_0.5h	M_N_1h	M_N_3h	M_N_5h	M_N_7h
fc_L_0.5h	NaN	89%	15%	0%	0%	50%	10%	19%	0%	0%	16%	7%	72%	0%	0%	64%	6%	37%	0%	0%
fc_L_1h	89%	NaN	12%	0%	0%	50%	18%	13%	1%	0%	24%	9%	66%	1%	0%	67%	9%	33%	0%	0%
fc_L_3h	15%	12%	NaN	0%	0%	0%	0%	63%	0%	0%	1%	0%	21%	0%	0%	1%	0%	34%	0%	0%
fc_L_5h	0%	0%	0%	NaN	0%	0%	2%	0%	92%	0%	3%	6%	0%	73%	0%	0%	4%	0%	92%	0%
fc_L_7h	0%	0%	0%	0%	NaN	0%	0%	0%	0%	0%	0%	0%	0%	0%	77%	0%	0%	0%	0%	13%
fc_N_0.5h	50%	50%	0%	0%	0%	NaN	35%	0%	1%	0%	66%	30%	14%	1%	0%	68%	23%	2%	0%	0%
fc_N_1h	10%	18%	0%	2%	0%	35%	NaN	0%	5%	0%	81%	59%	3%	5%	0%	21%	62%	0%	0%	0%
fc_N_3h	19%	13%	63%	0%	0%	1%	0%	NaN	0%	0%	1%	1%	32%	0%	0%	3%	0%	50%	0%	0%
fc_N_5h	0%	1%	0%	92%	0%	1%	5%	0%	NaN	0%	4%	13%	0%	85%	0%	1%	8%	0%	85%	0%
fc_N_7h	0%	0%	0%	0%	54%	0%	0%	0%	0%	NaN	0%	0%	0%	0%	70%	0%	0%	0%	0%	41%
M_L_0.5h	16%	24%	1%	3%	0%	66%	81%	1%	4%	0%	NaN	41%	14%	5%	0%	49%	51%	6%	1%	0%
M_L_1h	7%	9%	0%	6%	0%	30%	59%	1%	13%	0%	41%	NaN	4%	11%	0%	18%	90%	2%	2%	0%
M_L_3h	72%	66%	21%	0%	0%	14%	3%	32%	0%	0%	14%	4%	NaN	0%	0%	31%	2%	69%	0%	0%
M_L_5h	0%	1%	0%	73%	0%	1%	5%	0%	85%	0%	5%	11%	0%	NaN	0%	1%	8%	0%	66%	0%
M_L_7h	0%	0%	0%	0%	77%	0%	0%	0%	0%	70%	0%	0%	0%	0%	NaN	0%	0%	0%	0%	19%
M_N_0.5h	64%	67%	1%	0%	0%	68%	21%	3%	1%	0%	49%	18%	31%	1%	0%	NaN	13%	10%	0%	0%
M_N_1h	6%	9%	0%	4%	0%	23%	62%	0%	8%	0%	51%	90%	2%	8%	0%	13%	NaN	0%	1%	0%
M_N_3h	37%	33%	34%	0%	0%	2%	0%	50%	0%	0%	6%	1%	69%	0%	0%	10%	0%	NaN	0%	0%
M_N_5h	0%	0%	0%	92%	0%	0%	0%	0%	85%	0%	1%	2%	0%	66%	0%	0%	1%	0%	NaN	0%
M_N_7h	0%	0%	0%	0%	13%	0%	0%	0%	0%	41%	0%	0%	0%	0%	19%	0%	0%	0%	0%	NaN

Table 10: Shows the wilcoxcon rank sum test result between the flow shift timing results of all the used models and lead times from part 1. Each model-lead time combination exists in both the row and the column and each cell shows the Wilcoxon value between the result of the model found in the related row and the model found in the related column. A value above 5% (0.05) indicates that there are no substantial differences between the median of the respective results.

Model abbreviations: FC = full catchment (dataset)
M = municipalities (dataset)
L = Linear Regression Modell
N = Neural Network, 0.5-7 h = lead time

8.4.2 Part 1-Relative volume

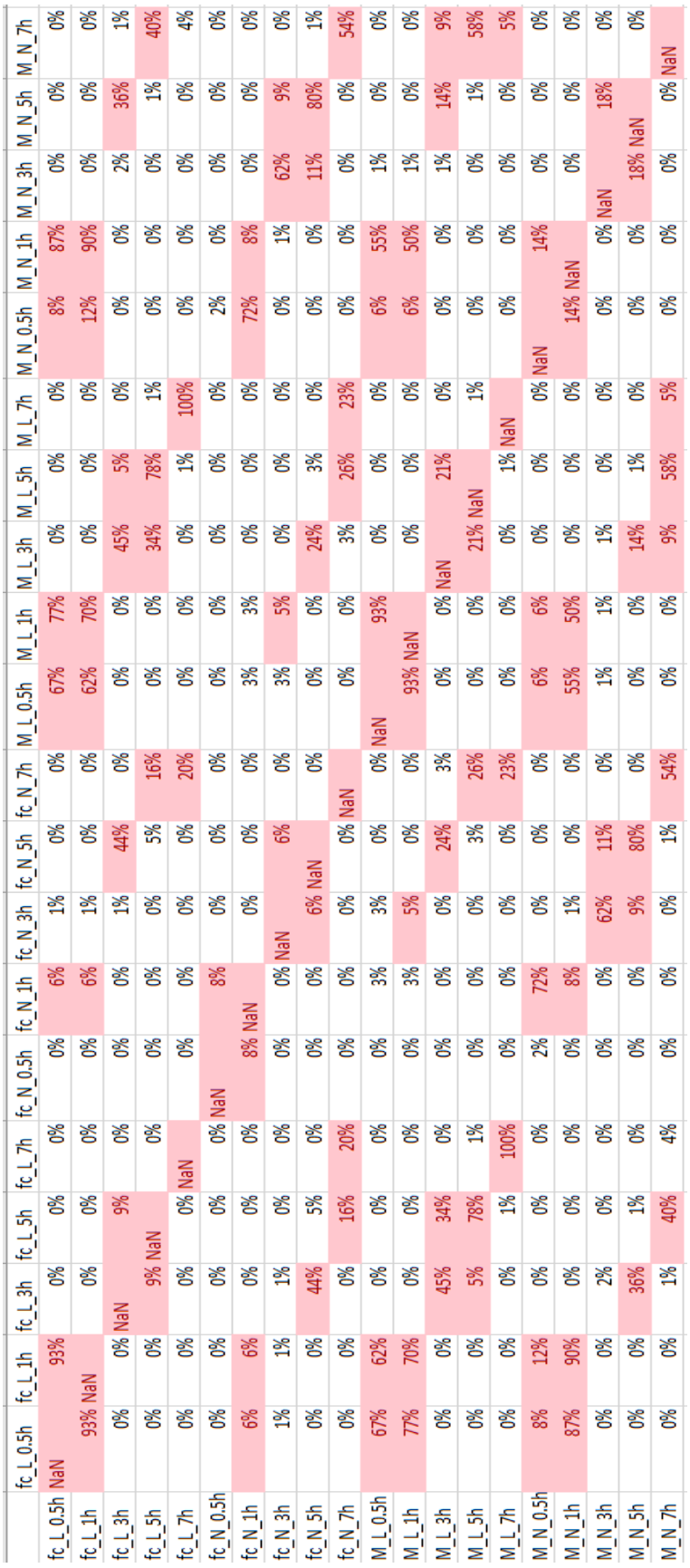


	fc_L_0.5h	fc_L_1h	fc_L_3h	fc_L_5h	fc_L_7h	fc_N_0.5h	fc_N_1h	fc_N_3h	fc_N_5h	fc_N_7h	M_L_0.5h	M_L_1h	M_L_3h	M_L_5h	M_L_7h	M_N_0.5h	M_N_1h	M_N_3h	M_N_5h	M_N_7h
fc_L_0.5h	NaN	34%	11%	36%	19%	0%	0%	0%	0%	0%	97%	99%	48%	18%	0%	0%	0%	0%	0%	0%
fc_L_1h	34%	NaN	30%	64%	9%	0%	0%	0%	0%	0%	44%	39%	16%	7%	0%	0%	0%	0%	0%	0%
fc_L_3h	11%	30%	NaN	71%	3%	0%	0%	0%	0%	0%	15%	15%	7%	2%	0%	0%	0%	0%	0%	0%
fc_L_5h	36%	64%	71%	NaN	6%	0%	0%	0%	0%	0%	39%	36%	18%	5%	0%	0%	0%	0%	0%	0%
fc_L_7h	19%	9%	3%	6%	NaN	32%	10%	6%	0%	0%	20%	22%	53%	98%	3%	0%	0%	0%	0%	0%
fc_N_0.5h	0%	0%	0%	0%	32%	NaN	30%	19%	1%	0%	0%	0%	4%	22%	10%	0%	0%	0%	0%	0%
fc_N_1h	0%	0%	0%	0%	10%	30%	NaN	50%	5%	0%	0%	0%	1%	6%	33%	1%	0%	0%	0%	0%
fc_N_3h	0%	0%	0%	0%	6%	19%	50%	NaN	20%	0%	0%	0%	1%	5%	82%	20%	2%	0%	0%	0%
fc_N_5h	0%	0%	0%	0%	0%	1%	5%	20%	NaN	4%	0%	0%	0%	0%	27%	55%	96%	18%	3%	0%
fc_N_7h	0%	0%	0%	0%	0%	0%	0%	0%	4%	NaN	0%	0%	0%	0%	1%	1%	3%	29%	81%	21%
M_L_0.5h	97%	44%	15%	39%	20%	0%	0%	0%	0%	0%	NaN	94%	50%	19%	0%	0%	0%	0%	0%	0%
M_L_1h	99%	39%	15%	36%	22%	0%	0%	0%	0%	0%	94%	NaN	52%	20%	0%	0%	0%	0%	0%	0%
M_L_3h	48%	16%	7%	18%	53%	4%	1%	1%	0%	0%	50%	52%	NaN	56%	0%	0%	0%	0%	0%	0%
M_L_5h	18%	7%	2%	5%	98%	22%	6%	5%	0%	0%	19%	20%	56%	NaN	2%	0%	0%	0%	0%	0%
M_L_7h	0%	0%	0%	0%	3%	10%	33%	82%	27%	1%	0%	0%	0%	2%	NaN	21%	2%	0%	0%	0%
M_N_0.5h	0%	0%	0%	0%	0%	0%	1%	20%	55%	1%	0%	0%	0%	0%	21%	NaN	18%	1%	0%	0%
M_N_1h	0%	0%	0%	0%	0%	0%	0%	2%	96%	3%	0%	0%	0%	0%	2%	18%	NaN	10%	1%	0%
M_N_3h	0%	0%	0%	0%	0%	0%	0%	0%	18%	29%	0%	0%	0%	0%	0%	1%	10%	NaN	25%	1%
M_N_5h	0%	0%	0%	0%	0%	0%	0%	0%	3%	81%	0%	0%	0%	0%	0%	0%	1%	25%	NaN	7%
M_N_7h	0%	0%	0%	0%	0%	0%	0%	0%	0%	21%	0%	0%	0%	0%	0%	0%	0%	1%	7%	NaN

Table 11: Shows the wilcoxcon rank sum test result between the relative volume results of all the used models and lead times from part 1. Each model-lead time combination exists in both the row and the column and each cell shows the Wilcoxon value between the result of the model found in the related row and the model found in the related column. A value above 5% (0.05) indicates that there are no substantial differences between the median of the respective results.

Model abbreviations: FC = full catchment (dataset)
M = municipalities (dataset)
L = Linear Regression Modell
N = Neural Network, 0.5-7 h = lead time

8.4.3 Part 1-Mean absolute error (MAE)



	fc_L_0.5h	fc_L_1h	fc_L_3h	fc_L_5h	fc_L_7h	fc_N_0.5h	fc_N_1h	fc_N_3h	fc_N_5h	fc_N_7h	M_L_0.5h	M_L_1h	M_L_3h	M_L_5h	M_L_7h	M_N_0.5h	M_N_1h	M_N_3h	M_N_5h	M_N_7h
fc_L_0.5h	NaN	93%	0%	0%	0%	0%	6%	1%	0%	0%	67%	77%	0%	0%	0%	8%	87%	0%	0%	0%
fc_L_1h	93%	NaN	0%	0%	0%	0%	6%	1%	0%	0%	62%	70%	0%	0%	0%	12%	90%	0%	0%	0%
fc_L_3h	0%	0%	NaN	9%	0%	0%	0%	1%	44%	0%	0%	0%	45%	5%	0%	0%	0%	2%	36%	1%
fc_L_5h	0%	0%	9%	NaN	0%	0%	0%	0%	5%	16%	0%	0%	34%	78%	1%	0%	0%	0%	1%	40%
fc_L_7h	0%	0%	0%	0%	NaN	0%	0%	0%	0%	20%	0%	0%	0%	1%	100%	0%	0%	0%	0%	4%
fc_N_0.5h	0%	0%	0%	0%	0%	NaN	8%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%
fc_N_1h	6%	6%	0%	0%	0%	8%	NaN	0%	0%	0%	3%	3%	0%	0%	0%	72%	8%	0%	0%	0%
fc_N_3h	1%	1%	1%	0%	0%	0%	0%	NaN	6%	0%	3%	5%	0%	0%	0%	0%	1%	62%	9%	0%
fc_N_5h	0%	0%	44%	5%	0%	0%	0%	6%	NaN	0%	0%	0%	24%	3%	0%	0%	0%	11%	80%	1%
fc_N_7h	0%	0%	0%	16%	20%	0%	0%	0%	0%	NaN	0%	0%	3%	26%	23%	0%	0%	0%	0%	54%
M_L_0.5h	67%	62%	0%	0%	0%	0%	3%	3%	0%	0%	NaN	93%	0%	0%	0%	6%	55%	1%	0%	0%
M_L_1h	77%	70%	0%	0%	0%	0%	3%	5%	0%	0%	93%	NaN	0%	0%	0%	6%	50%	1%	0%	0%
M_L_3h	0%	0%	45%	5%	0%	0%	0%	0%	24%	3%	0%	0%	NaN	21%	0%	0%	0%	1%	14%	9%
M_L_5h	0%	0%	5%	78%	1%	0%	0%	0%	3%	26%	0%	0%	21%	NaN	1%	0%	0%	0%	1%	58%
M_L_7h	0%	0%	0%	1%	100%	0%	0%	0%	0%	23%	0%	0%	0%	1%	NaN	0%	0%	0%	0%	5%
M_N_0.5h	8%	12%	0%	0%	0%	2%	72%	0%	0%	0%	6%	6%	0%	0%	0%	NaN	14%	0%	0%	0%
M_N_1h	87%	90%	0%	0%	0%	0%	8%	1%	0%	0%	55%	50%	0%	0%	0%	14%	NaN	0%	0%	0%
M_N_3h	0%	0%	2%	0%	0%	0%	0%	62%	11%	0%	1%	1%	1%	0%	0%	0%	0%	NaN	18%	0%
M_N_5h	0%	0%	36%	1%	0%	0%	0%	9%	80%	0%	0%	0%	14%	1%	0%	0%	0%	18%	NaN	0%
M_N_7h	0%	0%	1%	40%	4%	0%	0%	0%	1%	54%	0%	0%	9%	58%	5%	0%	0%	0%	0%	NaN

Table 12: Shows the wilcoxcon rank sum test result between the mean absolute error results of all the used models and lead times from part 1. Each model-lead time combination exists in both the row and the column and each cell shows the Wilcoxon value between the result of the model found in the related row and the model found in the related column. A value above 5% (0.05) indicates that there are no substantial differences between the median of the respective results.

Model abbreviations: FC = full catchment (dataset)
M = municipalities (dataset)
L = Linear Regression Modell
N = Neural Network, 0.5-7 h = lead time

8.4.4 Part 2-Flow shift timing

Table 13: Shows the wilcoxcon rank sum test result between the flow shift timing results of all the used models and lead times from part 2. Each model-lead time combination exists in both the row and the column and each cell shows the Wilcoxon value between the result of the model found in the related row and the model found in the related column. A value above 5% (0.05) indicates that there are no substantial differences between the median of the respective results.

Model abbreviations: m0k = municipalities (dataset), m20k = municipalities + 20km data, m30k = municipalities + 30km data, n = Neural Network, 3-7 h = lead time

	m0k_n_3h	m0k_n_5h	m0k_n_7h	m20k_n_3h	m20k_n_5h	m20k_n_7h	m30k_n_3h	m30k_n_5h	m30k_n_7h
m0k_n_3h	NaN	0%	0%	42%	0%	0%	5%	1%	0%
m0k_n_5h	0%	NaN	0%	0%	11%	0%	0%	2%	0%
m0k_n_7h	0%	0%	NaN	0%	0%	63%	0%	0%	6%
m20k_n_3h	42%	0%	0%	NaN	0%	0%	32%	0%	0%
m20k_n_5h	0%	11%	0%	0%	NaN	0%	0%	42%	0%
m20k_n_7h	0%	0%	63%	0%	0%	NaN	0%	0%	8%
m30k_n_3h	5%	0%	0%	32%	0%	0%	NaN	0%	0%
m30k_n_5h	1%	2%	0%	0%	42%	0%	0%	NaN	0%
m30k_n_7h	0%	0%	6%	0%	0%	8%	0%	0%	NaN

8.4.5 Part 2-Mean absolute error (MAE)

Table 14: Shows the wilcoxcon rank sum test result between the mean absolute error results of all the used models and lead times from part 2. Each model-lead time combination exists in both the row and the column and each cell shows the Wilcoxon value between the result of the model found in the related row and the model found in the related column. A value above 5% (0.05) indicates that there are no substantial differences between the median of the respective results.

Model abbreviations: m0k = municipalities (dataset), m20k = municipalities + 20km data, m30k = municipalities + 30km data, n = Neural Network, 3-7 h = lead time

	m0k_n_3h	m0k_n_5h	m0k_n_7h	m20k_n_3h	m20k_n_5h	m20k_n_7h	m30k_n_3h	m30k_n_5h	m30k_n_7h
m0k_n_3h	NaN	18%	0%	54%	32%	0%	27%	70%	0%
m0k_n_5h	18%	NaN	0%	8%	80%	0%	2%	12%	4%
m0k_n_7h	0%	0%	NaN	0%	0%	94%	0%	0%	23%
m20k_n_3h	54%	8%	0%	NaN	13%	0%	58%	81%	0%
m20k_n_5h	32%	80%	0%	13%	NaN	1%	6%	17%	5%
m20k_n_7h	0%	0%	94%	0%	1%	NaN	0%	0%	23%
m30k_n_3h	27%	2%	0%	58%	6%	0%	NaN	42%	0%
m30k_n_5h	70%	12%	0%	81%	17%	0%	42%	NaN	0%
m30k_n_7h	0%	4%	23%	0%	5%	23%	0%	0%	NaN