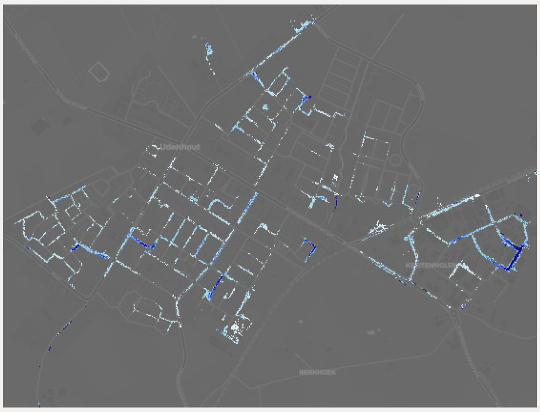
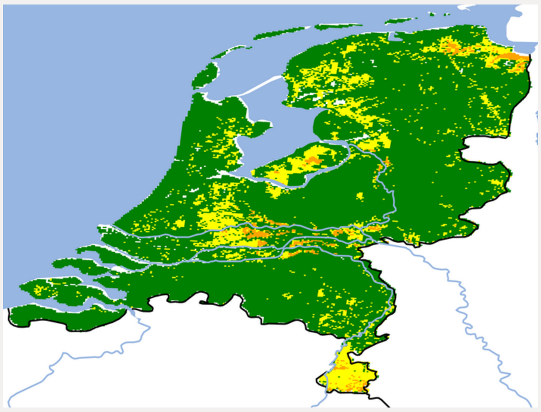
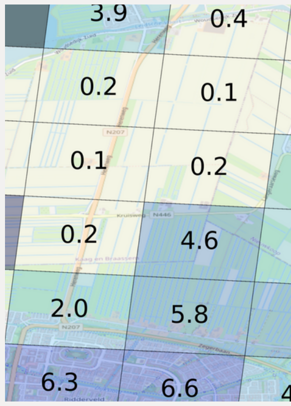
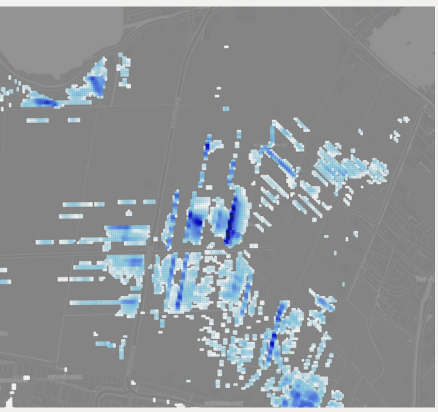
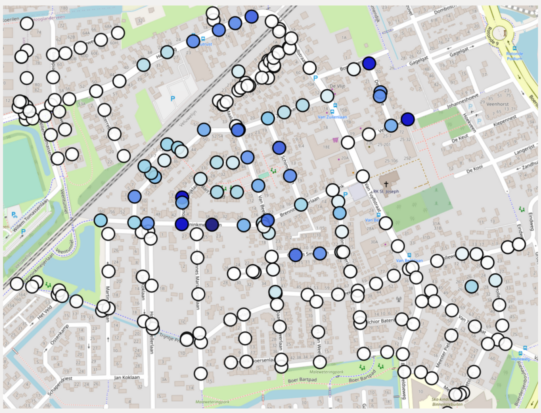
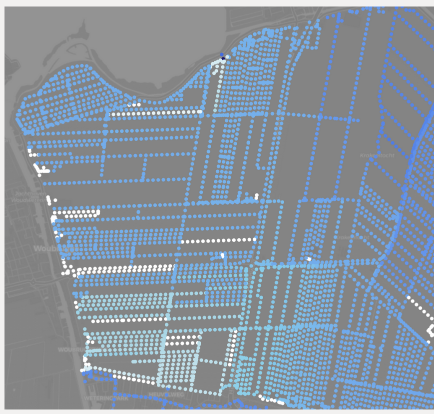
SURROGATE MODELS: A SOLUTION FOR REAL-TIME INUNDATION FORECASTING?

Master Thesis
Civil Engineering & Management

Laura Janssen
July 2023

HydroLogic
UNIVERSITY OF TWENTE.

Surrogate models: a solution for real-time inundation forecasting?

Surrogate modelling for three case studies in the Netherlands

Master Thesis

Presented for the degree of MSc Civil Engineering & Management

05-07-2023

Final version

Author

L. (Laura) Janssen

Graduation Committee:

Dr. Ir. M.J. Booij	University of Twente
Dr. Ir. T.M. Duong	University of Twente
Ir. B. Schnitzler	HydroLogic
L. Diender MSc.	HydroLogic

Preface

In front of you lies my Master thesis ‘Surrogate models: a solution for real-time inundation forecasting?’. This thesis is the final phase of the Master Water Engineering & Management at the University of Twente. In the past months, I had the opportunity to carry out my Master thesis at HydroLogic.

Throughout my research, I have had the support from many people around me, for which I am very grateful. First of all, I would like to thank my supervisors from HydroLogic: Bram Schnitzler and Ludo Diender. I would like to thank you for your valuable feedback and support throughout my master’s thesis. Your willingness to answer all my questions has been very helpful. I appreciate the time and effort you invested in assisting me throughout this journey. Besides, I would like to thank my supervisors Martijn Booij and Trang Duong from the University of Twente for their helpful meetings and feedback.

Lastly, I would like to thank Anne Vrouwe & Jos de Vries (Municipality of Amersfoort), Jan Janssens-Baan & Jaap Jansen (Municipality of Tilburg), and René van der Zwan & Jan Jelle Reitsma (Hoogheemraadschap van Rijnland). Your expert knowledge and experience on flood modelling within an operational setting were valuable inputs for my research.

I hope you enjoy reading my Master thesis.

Laura Janssen,

July 2023

Summary

The Netherlands is facing an increased risk of pluvial flooding, mainly due to the expected rise in the frequency and intensity of extreme rainfall events caused by climate change. Recent events have highlighted the importance of accurate flood forecasting to minimize damage. Hydrological inundation models are crucial in flood mitigation as they support water managers in an operational setting by predicting the size and timing of a flooding event, such that they can make informed decisions and implement appropriate measures (e.g. adjusting weir levels, alerting relevant parties). To ensure the operational usability of hydrological inundation models, they must meet certain criteria. Detailed hydrological inundation models provide accurate results but have long computational times, making them impractical for the operational context. Surrogate models have a shorter computational time by approximating the detailed hydrological inundation model. Surrogate models may offer a solution, but their quality and added value in an operational setting compared to existing hydrological models are yet to be determined.

Waarschuwing voor Wateroverlast (W2O) is an example of a conceptual bucket model that is currently used in the operational setting. It combines a probabilistic rainfall forecast with geographical characteristics (e.g. land use, soil type) and the available soil storage to calculate the probability of flooding. This conceptual hydrological inundation model covers the entire area of the Netherlands and has a computational time of only one minute. It could therefore be, next to surrogate models, a useful alternative to predict inundation in an operational setting.

In this study, the added value of surrogate models in an operational setting compared to the W2O model is researched. Three cases studies are selected to evaluate the hydrological inundation models' compliance with the end-users' requirements. These case studies include: the municipality of Amersfoort who would like to have an accurate forecast of the locations where inundation is expected within the city, the municipality of Tilburg who would like to have an accurate forecast of urban surface inundation, and Hoogheemraadschap van Rijnland who would like to know probability of flooding when the interaction between open waters (surface waters) and inundation on surface level is included.

Semi-structured interviews with the end-users took place to specify the modelling goals, perspectives for action, relevant output variables, and the required accuracy. For all three case studies, the end-users are interested in the size and location of the flooding on a 2D map to alert relevant authorities (e.g. contractors). The waterboard is, next the location and size of the flooding, also interested in the water levels in the study area's waterways, enabling proactive response to flood forecasts by changing weir levels and installing temporary pumping stations when necessary.

For each case study, a surrogate model is created based on available data, models, and the end-user's preference. For the municipality of Amersfoort, a Machine Learning (ML) model is used that predicts flood volume timeseries for manholes in the sewage system based on rainfall timeseries as input. This ML model meets most of the end-user's requirements. However, the output variable (flood volume per manhole), is relatively difficult to interpret and is therefore not suitable for non-experts. The ML model created for the municipality of Tilburg predicts the maximum inundation depth on surface level using the rainfall timeseries as input. This ML model meets all requirements by the end-user, except of the required accuracy of the model. The critical success index, a performance indicator that measures the model's ability to correctly predict flooding or not, is only 48% and therefore insufficient. Other studies in literature have shown that ML models are able to accurately reproduce the output of hydrological inundation models. This type of surrogate model has thus a high potential, but further research is necessary. For Hoogheemraadschap van Rijnland, a surrogate model is created based on a detailed hydrological inundation model by applying simplifications to the model schematisation. This model meets all end-user's requirements but is at its limits regarding the maximum computational time (30 minutes).

Also the quality of the W2O model is assessed using the requirements from the end-users. For both municipalities, the W2O model meets most of the requirements. Further research is necessary to determine if the W2O model also meets the accuracy requirement. For the waterboard, the W2O model suffices in the information provision on the 2D grid, but since the W2O model does not calculate the water level in waterways, it only partially meets the requirements for the required output variable.

Overall, it can be concluded that, for municipalities, the W2O model provides sufficient information in an operational setting. It is advised to further research the accuracy of the W2O model to ensure that the W2O model also meets this requirement. The ML of the second case study has high potential in being an additional value to the W2O model. It is able to rapidly predict inundation depths on a high spatial resolution, but the accuracy of this ML model needs to be improved. Since generating the training data for the ML model is a computationally and memory expensive process, it is not advised to apply this type of surrogate model on larger model domains. Instead, ML models could be used as an addition to the W2O model, where the W2O model is used to get a general impression of the predicted inundation and the ML model is used additionally for the most vulnerable areas. For waterboards, the W2O model suffices in the information provision on the 2D grid, but additional models are needed to provide also the information regarding the water levels in waterways. The surrogate model created for the third case study could be used for this. Due to the long computational time, it is not advised to apply this surrogate to a larger model domain. Instead, this surrogate model could be used for the most vulnerable areas. Using the W2O model to get a general impression of the predicted inundation and combining this with a surrogate model for the most vulnerable areas, results in all information needed to make informed decisions on flood mitigation.

Contents

Preface 2

Summary 3

1. Introduction 5
   - 1.1 Problem context 5
   - 1.2 State of the art 6
   - 1.3 Research gap 7
   - 1.4 Research objective & research questions 7
   - 1.5 Selected case studies 8
   - 1.6 Outline 9
2. Model requirements 11
   - 2.1 Introduction 11
   - 2.2 Methodology 11
   - 2.3 Results 11
   - 2.4 Discussion 14
   - 2.5 Conclusion 15
3. Quality assessment surrogate model 16
   - 3.1 Introduction 16
   - 3.2 General research methodology 16
   - 3.3 Case study 1 – Municipality of Amersfoort 18
   - 3.4 Case study 2 – Municipality of Tilburg 24
   - 3.5 Case study 3 – Hoogheemraadschap van Rijnland 33
   - 3.6 Conclusion 45
4. Quality assessment W2O model 47
   - 4.1 Introduction 47
   - 4.2 Methodology 47
   - 4.3 Results 50
   - 4.4 Discussion 53
   - 4.5 Conclusion 54
5. Discussion 57
   - 5.1 Limitations 57
   - 5.2 Generalisation 58
   - 5.3 Practical relevance & model suitability 60
   - 5.4 Theoretical contribution 60
6. Conclusion & Recommendations 62
   - 6.1 Conclusion 62
   - 6.2 Recommendations 64

References 65

Appendices 68
- A Interview questions 68
- B Minutes of the interviews 69
- C Implications timestep rainfall forecast 88
- D Alternative machine learning model set-ups 89
- E Final model structure MLP model Case Study 2 93
- F Enlarged figures case study 2 94
- G Observed rainfall 09-09-2021 95

1. Introduction

1.1 Problem context

Climate change is expected to lead to an increase in the frequency and intensity of extreme rainfall events in the Netherlands (KNMI, 2021). When the rainwater cannot be absorbed effectively by the soil and the capacity of the drainage system is exceeded, this can cause pluvial flooding (Acosta-Coll et al., 2018; Hofmann & Schüttrumpf, 2020). As an example, in July 2021, southern parts of the province of Limburg, The Netherlands, experienced an extraordinary pluvial and fluvial flooding event with more than 160 mm of rainfall within a span of 48 hours, leading to flooded roads, houses and businesses (Deltares, 2022a). This event highlights importance of timely and accurately forecasting flood events as this can minimize the damage causes by the flooding. This enables the implementation of appropriate measures before and during the flooding, such as the placement of sandbags/barriers or the change of weir levels (de Moel et al., 2014; Wang et al., 2019). Pluvial flooding typically results from fast-moving storms with short lead times (Borga et al., 2010; Schanze, 2018), making their prediction challenging for flood forecasters.

Flood Early Warning systems (FEWS) play a key role in flood mitigation. Based on input variables such as rainfall forecasts, FEWS provide flood forecasts for a specific lead time and notify the relevant authorities when necessary (Verkade & Werner, 2011; Wang et al., 2019). Hydrological inundation models are a part of FEWS and predict the flooding based on the rainfall forecast. To use FEWS and thus hydrological inundation models in an operational setting, the hydrological inundation model must satisfy certain criteria.

Detailed hydrological inundation models provide accurate results but have long computational times (Wang et al., 2019), making them impractical for the operational setting. Surrogate models reduce the computational load by approximating the original detailed hydrological inundation model (Razavi et al., 2012). Alternatively, conceptual hydrological models could be useful in an operational setting. Conceptual hydrological models are physically based models but do not solve the numerical equations, resulting in a reduced computational load. Waarschuwing voor Wateroverlast (W2O) is an example of a conceptual bucket model that is currently used in the operational setting (Schnitzler, 2022). It combines a probabilistic rainfall forecast with geographical characteristics (e.g. elevation, soil type) and the available soil storage to calculate the probability of flooding. This conceptual hydrological inundation model covers the entire area of the Netherlands and has a computational time of ± one minute. The W2O model could therefore also be, next to surrogate models, a useful alternative to predict inundation in an operational setting. A schematization of the different model types is shown in Figure 1.

The quality and added value of surrogate models compared to the W2O model in an operational setting are dependent on the requirements of the end-user. This study focusses on defining these requirements, and subsequently evaluate the quality of three different surrogate models and the W2O model by assessing their compliance with these requirements.

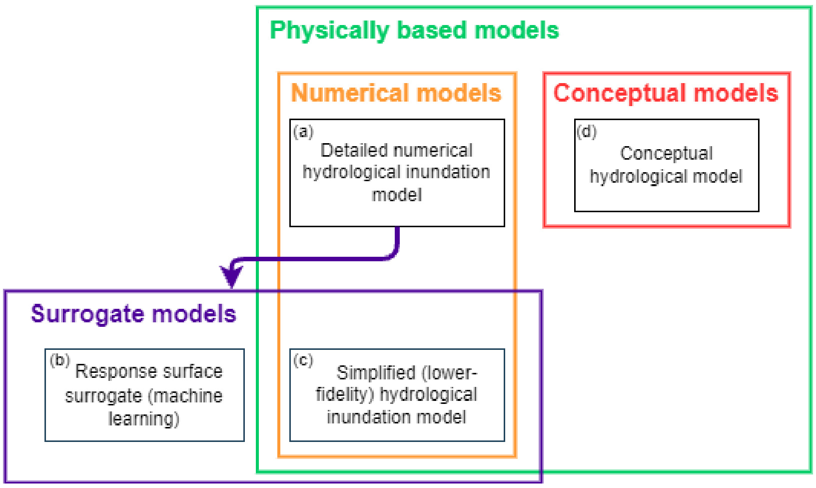


Figure 1 – Schematisation of different types of hydrological inundation models and their terminology. Terminology is specific for this thesis. Surrogate models (b & c) are derived from detailed numerical hydrological inundation models (a). A conceptual hydrological model (d) is a physically based model that does not solve the numerical flow equations.

1.2 State of the art

Waarschuwing voor Wateroverlast (W2O, translated: warning for flooding) is a conceptual hydrological inundation model (Schnitzler, 2022). The model combines a probabilistic rainfall forecast with a 2-hour and 48-hour lead time with the current available soil storage, and considers the infiltration based on the soil type and land use type. Using this, the probability of flooding in the next 2 and 48 hours can be calculated. The probability of flooding is defined as the likelihood (%) of more than 0mm, 10mm or 30mm water on the surface level within the next 2 and 48 hours. A more extensive explanation of the W2O model can be found in Section 4.2.

The W2O model covers the entire area of the Netherlands and calculates the probability of flooding for each 1 km x 1 km grid cell. The model covers a large spatial area and has a short computational time (± one minute) to ensure that it can be used in an operational setting. Since the W2O model is a conceptual hydrological model (assigned (d) in Figure 1), the numerical flow equations (i.e. Navier-Stokes equations) are not included in the model. Instead, a more simple bucket principle is applied. Not all hydrological and hydraulic processes are thus included in great detail. However, for specific applications, some of these hydrological processes could be considered relevant. For urban areas, these include the influence of sewage systems (Kilsdonk et al., 2022) as the overflowing of manholes in the sewage systems is a recurring cause of flooding in urban areas. Besides, urban areas have a dense and variable topography and infrastructure (Y. Wang et al., 2018). In case of pluvial flooding, this leads to small-scale and complex flow processes (Henonin et al., 2013). A high spatial resolution is required to accurately model floods in urban areas. Lastly, in certain rural areas of the Netherlands, the waterways play a crucial role in case of heavy rainfall. Therefore, waterboards consider it important that these surface waters, and the interaction between the surface waters and overland flow, are included in a hydrological inundation model.

As not (yet) all hydrological processes are included in the W2O model, alternative hydrological inundation models might better satisfy the criteria from the end-user. Detailed hydrological inundation models (assigned (a) in Figure 1) are able to accurately model all hydrological processes (Wang et al., 2019). However, due to the complex nature of floodings, solving the numerical equations in a hydrological inundation model results in a long computation time (Teng et al., 2017). As pluvial flooding can occur within a couple of hours, hydrological inundation models require a short computational time. Surrogate modelling could be a solution for this.

Surrogate models reduce the computational load by approximating the original detailed hydrological inundation model (Razavi et al., 2012). Two types can be distinguished: lower fidelity models (assigned (c) in Figure 1) and response surface surrogates (assigned (b) in Figure 1). Lower fidelity models are simplified and less detailed versions of a detailed hydrological inundation model where the physical processes of the model are preserved. Several examples of lower fidelity models can be found in literature. Bomers et al. (2019) created a lower fidelity hydraulic 1D2D model by lowering the dimensions, increasing the computational time step, and simplifying the shallow water equations of a detailed hydrological inundation model. Razavi et al. (2012) reviewed multiple lower fidelity surrogate models found in literature, where surrogate models are created using various techniques.

Next to lower fidelity models, response surface surrogates are another type of surrogate models. Response surface surrogates approximate the relationship between input and output variables by using data-driven techniques. Machine learning algorithms are an example of response surface surrogates. These algorithms are capable of recreating the training data. After training the machine learning model, the model can predict the outcome based on the given input. In the context of predicting inundation, simulations of hydrological inundation models serve as the training data as measurements are often not available. To predict pluvial flooding, neural networks have shown to outperform other methods as neural networks are capable of learning complex non-linear patterns (Bentivoglio et al., 2022; Kabir et al., 2020). Several examples of neural networks employed for pluvial flooding prediction can be found in literature. One example is the Long-Short Term Memory (LSTM) model developed by Hop (2023), which predicts inundation depths at specific time steps. Another example is the LSTM model created by Kilsdonk (2022), which predicts the flood volume for each manhole within the sewage system at different timesteps. Lastly, the Multi-Layer Perceptron (MLP) model introduced by Berkhahn (2019) predicts maximum inundations depths for two different urban areas using rainfall timeseries as input. The accuracy and runtime time of these neural networks is promising for the usage of machine learning models in an operational setting.

1.3 Research gap

To model the specific hydrological applications as stated in Section 1.2, alternative methods might better meet the requirements of waterboards and municipalities as not all hydrological processes are included in the W2O model. These alternative methods need to be usable in an operational setting, meaning that the model results should be accurate enough to support the policy decision making process, while minimizing the computational time. Three research gaps can be identified in this study.

First of all, it is currently unknown what the requirements are of a hydrological inundation model to be usable in an operational setting. This is dependent on how the end-user is using the model in their decision-making process. Requirements include for example the maximum computational time and the required accuracy of the model results. Detailed hydrological inundation models are able to cover the applications as asked for by the waterboards and municipalities. However, due to their long computational time, these models are not usable in an operational setting. To ensure the short runtime, a surrogate version of the detailed hydrological inundation models is needed. It is unknown to what extent these surrogate models are able to meet all the requirements from the end-users, which is the second research gap. When the surrogate model does not fulfil all requirements of the end-user, the W2O model could be a valuable alternative, depending on the requirements and modelling goal. Consequently, in these cases, the already existing W2O model may prove to be the preferred tool for hydrological inundation modelling. The third research gap is that it is currently unknown whether surrogate models have an added value compared to the already available W2O model within an operational setting.

1.4 Research objective & research questions

As elaborated above, three research gaps are currently present regarding the usage of surrogate models in an operational setting. This includes the end-users' requirements of a method, to what extent a surrogate model is able to meet these requirements, and if these models have an added value in an operational setting compared to the W2O model. Therefore, the aim of this study is to research to which extent surrogate models have an added value (compared to the W2O model) in supporting the end-user in an operational setting. A surrogate model has an added value when it better meets the end-users' requirements compared to the W2O model. Three case studies serve as an example: they are described in Section 1.5.

To reach the research objective, three research questions are defined:

1. What are the end-users' requirements of a hydrological inundation model in an operational setting?

Waterboards and municipalities want to research the added value of surrogate models compared to the W2O model as not (yet) all hydrological processes are currently covered by the W2O model. They would like to have a hydrological inundation model that provides sufficient and accurate information that supports their decision-making process in case of heavy rainfall. It is important to specify what the end-users' requirements are to assess the overall quality of both the surrogate models as well as the W2O model. These requirements include for example the required accuracy of the model results and the computational time. To find out what the end-users' requirements are of a hydrological inundation model, semi-structured interviews will take place. The questions for the interviews are drafted based on a literature study. Using the interviews, the modelling goal, perspectives for action in an operational setting, the relevant output variables, and required accuracy are specified.

2. To what extent are surrogate models for the case studies as described by 2a, 2b, and 2c able to meet the requirements that follow from research question 1?
    a. Machine learning techniques applied on sewage models
    b. Machine learning techniques applied on hydrological inundation models
    c. Simplified hydrological 1D2D inundation modelling

Three case studies are selected based on the municipalities' and waterboard's interest in the project, see Section 1.5. Surrogate models are created for each case study to see if these models are able to meet the requirements that follow from research question 1. For each case study, a different type of surrogate model is used. The type of surrogate model is based on the available data and models, and the preference of the waterboard or municipality.

3. What is the added value of a surrogate model compared to the W2O model in an operational setting?

Research question 2 has led to a surrogate model for each case study including a description of its quality with respect to the detailed hydrological inundation model and the requirements of the end-user. The created surrogate models will meet the requirements of the end-user to a certain extent. If the surrogate model is not able to fully satisfy the requirements of the end-user, it could be that the surrogate model does not have an added value compared to the already available W2O model. To define the added value of the surrogate model with respect to the W2O model, the two models are qualitatively compared to each other and the requirements as defined by the end-user. A schematization of the methodology is presented in Figure 2.

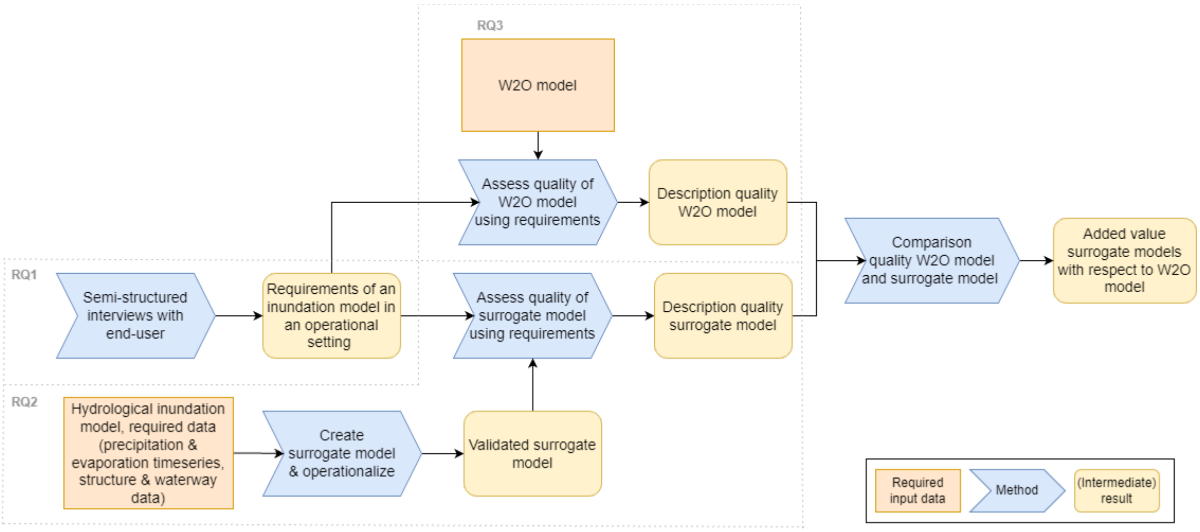


Figure 2 - Flowchart of the general research methodology

1.5 Selected case studies

For the purpose of this thesis, three case studies serve as an example. The following case studies are selected:

- The municipality of Amersfoort would like to have a more accurate forecast of the size and location of inundation in an urban area.
- The municipality of Tilburg would like to have a more accurate forecast of the size and location of inundation in an urban area.
- Hoogheemraadschap van Rijnland would like to know probability of flooding when the interaction between open waters (surface waters) and inundation on surface level is included.

Due to limited available time, it was not possible to create a surrogate model that covers the complete study area of the two municipalities and waterboard. Instead, a smaller study area is selected based on the already available data and models. These smaller areas accommodate the construction time of both the hydrological model and the dataset. The selected study areas are shown in Figure 3, where the figures in the middle show the municipalities and waterboard, and the figures on the right represent the selected study areas for this thesis. These selected study areas serve as a proof of concept: if it turns out that the created surrogate models meet the requirements from the end-users, then surrogate models could be created that cover the entire study area.

1.5.1 Case study Amersfoort: Hooglanderveen

For the case study of the municipality of Amersfoort, the area Hooglanderveen is selected as study area for this thesis. Hooglanderveen is a residential area within Amersfoort, the Netherlands. It is located in the northeast of Amersfoort, see Figure 3. The size of the area is about 1 km^2, mostly consisting of urban area. In total, around 5000 inhabitants live in Hooglanderveen. The neighbourhood has a combined sewer system and has frequently experienced pluvial flooding in the past. This study area has been selected since a Long Short-Term Memory (LSTM) neural network based on a hydrological inundation model was already available (Kilsdonk, 2021). This LSTM model predicts flood volumes (timeseries) for each manhole in the sewage system based on rainfall time series as input. More detailed information on this model can be found in Section 3.3.1.

1.5.2 Case study Tilburg: Udenhout

The village Udenhout is selected for the case study of the municipality of Tilburg. Udenhout is a residential area within the municipality of Tilburg, the Netherlands. It is a small village with 8440 inhabitants located northeast from Tilburg. In total, the village Udenhout covers around 6 km^2, mostly consisting of urban area. Udenhout is selected as study area since this area has experienced pluvial flooding in the past, which makes it an interesting research area. Besides, a calibrated and validated numerical hydrological inundation model is already available for this area via Royal HaskoningDHV. More detailed information on this model can be found in Section 3.4.1. Generating the dataset to train and validate the machine learning model is done using the input and output of the hydrological inundation model provided by Royal HaskoningDHV.

1.5.3 Case study Hoogheemraadschap van Rijnland: polder Vierambacht

For Hoogheemraadschap van Rijnland, the polder Vierambacht is selected as study area for this thesis. The polder is located north of the city of Alpen aan den Rijn in the Netherlands, and has an area of around 17.9 km^2. The area is low-lying: the average elevation of the surface level is 4.5 meters below sea level. Most of the polder consists of rural landscape with meadows for agricultural purposes. As can be seen in Figure 3 (bottom right), there are many waterways and ditches in the area. On the north side of the polder, the boezem waters called Leidsche Vaart, the Braassemermeer and the Paddegat are located. On the west side, the boezem water Woudwetering forms the border of the polder. On the southern side, a dike forms the border while on the east side quay walls are located. The polder is a relatively closed system where most water exits the study area via one of the two pumping stations. The polder is well-known and understood by experts since multiple studies have been conducted in the past for this study area as part of the PolderLab project (Hoogheemraadschap van Rijnland, 2022).

1.6 Outline

In the sections below, the research questions as stated in Section 1.4 are answered. In Section 2, the requirements from the end-user for each case study are defined. Section 3 starts with a general introduction and general research methodology. The case study specific methodology, results and discussion are described per case study in Section 3.3, Section 3.4 and Section 3.5 for Case study 1,2, and 3 respectively. Section 3 finalizes with a general conclusion, which can be found in Section 3.6. In Section 4, the quality of the W2O model is assessed using the requirements from the first research question. Finally, in Section 5 the discussion can be found and in Section 6 the conclusions and recommendations are described.

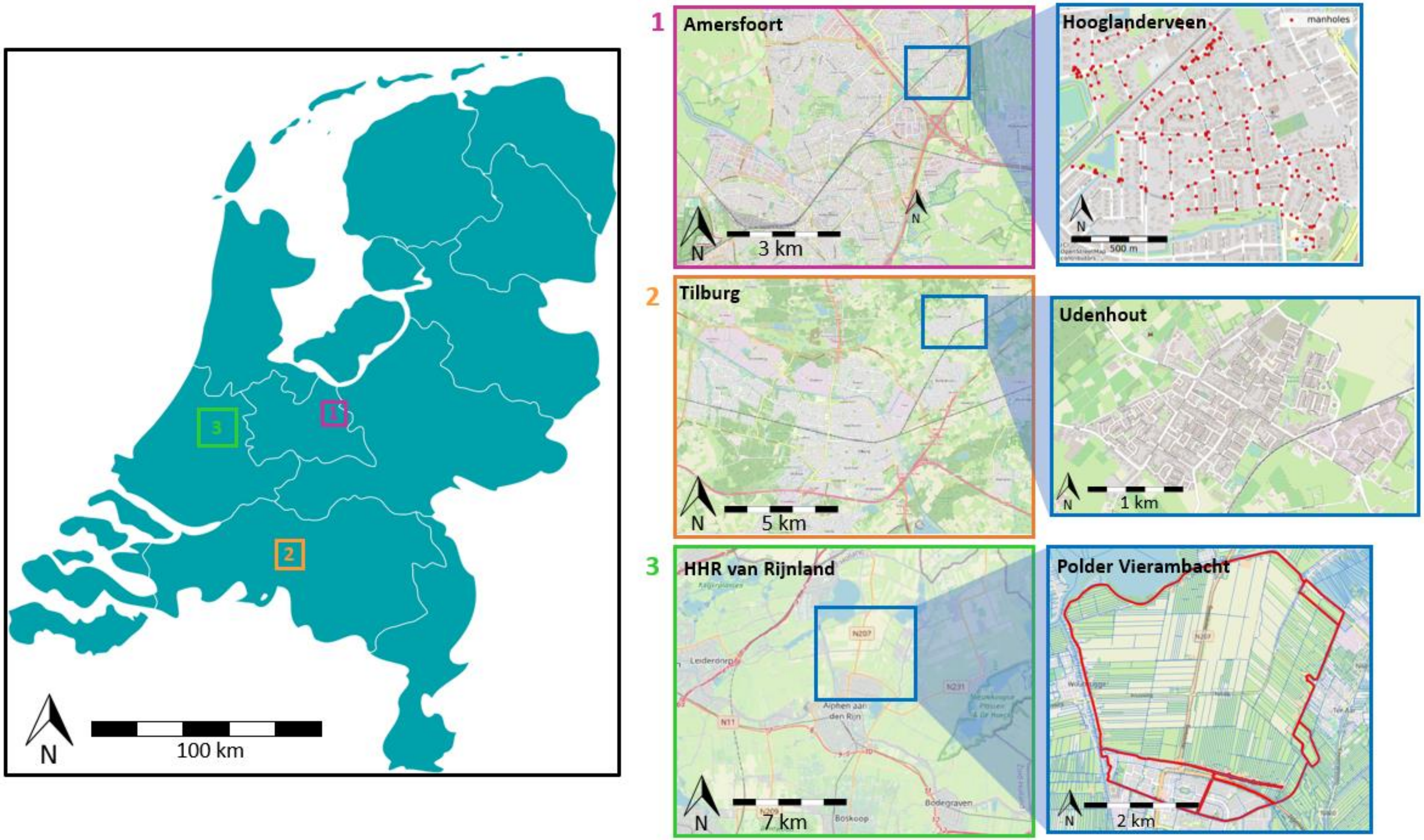


Figure 3 - Selected study areas. Left: the location of the study areas within the Netherlands. Middle: the area of both municipalities and the waterboard with the selected study area in the blue rectangle. Right: The selected study areas for this thesis that serve as a proof of concept.

2. Model requirements

2.1 Introduction

To assess the quality of the surrogate models and the W2O model, the end-users' requirements must be defined. The first research question addresses this. In this chapter, the methodology, results, discussion, and conclusion of the first research question are described. Section 2.2 describes the used methodology to set-up the questions for the interviews. Section 2.3 summarizes the outcome of the interviews. In Section 2.4, the results are discussed and to conclude in Section 2.5, the requirements are summarized to answer the first research question.

2.2 Methodology

Waterboards and municipalities want to research the added value of surrogate models compared to the W2O model as the W2O model does not (yet) cover all hydrological processes and inundation applications. They would like to have a hydrological inundation model that better supports their decision-making process in case of heavy rainfall. It is important to specify what the end-users' requirements are to ensure that these surrogate models meet their expectations as good as possible.

Through literature study, first a general overview of relevant criteria is set up to assess the usability of a model in an operational setting. These criteria include the temporal and spatial resolution of a hydrological model, the maximum computational time such that the model can be used for real-time forecasting, and required output variables. Using this overview with criteria, the questions for the semi-structured interview were drafted. Next, semi-structured interviews with the end-users were held to specify the modelling goal and the requirements of the inundation model. This information is case study specific and depends on the operational setting within the municipality or waterboard. Also, a discussion was held about how the model is used in an operational setting and what the corresponding relevant output variables, required resolution, and performance indicators are. Finally, the requirements for each case study are defined. This will be used in research question 2 and 3 to assess the model's quality and added value.

2.3 Results

First of all, questions for the interviews were drafted based on a literature study. Jakeman et al (2006) described multiple points that should be considered when assessing if a model fits its purpose. First of all, the model's purpose and scope should be clearly defined, including the specific questions and issues it will address, the interest groups involved (including clients and end-users) of the model, the desired outputs, the forcing variables, the expected accuracy, and the temporal and spatial boundaries (e.g. scope, scale and resolution). Furthermore, a fixed time frame should be defined for completing the model, as well as the available resources and effort for both setting up and operating the model.

Using this information, the questions for the semi-structured interviews were drafted. The main goal of the interviews is to define what the requirements of a model are to be usable in an operational setting according to the end-users of the models. The interviews focus on the operational setting including the perspectives for action and the corresponding lead times, the modelling goal, the required output variables and the accuracy of the output. The complete list with questions for the interviews can be found in the Appendix A.

2.3.1 Case study 1 – Municipality of Amersfoort

On 13/02/2023 an interview with Anne Vrouwe (consultant on water & climate adaptation) and Jos de Vries (sewage system expert) from the Municipality of Amersfoort took place.

For the municipality of Amersfoort, the cause of the inundation plays an important role. In Amersfoort, the sewage system is determinative for inundation on surface level. However, overflowing of the sewage system can have multiple causes. One of them is when heavy rainfall exceeds the capacity of the system, but also other factors play an important role (deferred maintenance, clogged manholes). It is important to realise that these factors are (probably) not considered in a hydrological inundation model. These models only consider overflowing due to the incoming water exceeding the maximum capacity. Using these models as a forecast could thus result in missed alerts or underestimations of the size of the flooding.

For the municipality of Amersfoort, the main goal of the model is to give more insights and information beforehand on the locations where inundation might occur. The municipality does not have any measures they can take beforehand. However, by knowing the expected inundated locations, they can alert relevant contractors and citizens. Also contractors cannot take measures beforehand, but it is useful if they are on stand-by such that they take appropriate actions as soon as necessary. Warning citizens about the possible future event is mainly useful for the citizens themselves as they can take measures to minimize the damage of their own property. One can think of installing flood barriers in front of their doors, see Figure 4, or move valuables to higher locations. These kinds of measures are the citizens' own responsibility, but the municipality is willing to help by informing them about the forecast. To warn the citizens of Amersfoort, the municipality would like to communicate the output of the model directly to its citizens. This implies that the output should be understandable for (almost) every inhabitant within the area. With the information, inhabitants can decide for themselves which measures they want to take. Historic events have shown that the measures taken by citizens themselves are the ones that contribute the most to minimizing the damage. Alerting the contractors and informing the inhabitants is the best a municipality can offer on short notice when inundation is expected.

Figure 4 - Flood barrier in front of a door: a measure to minimize damage due to flooding (FloodSafe Projects, 2021)

In the future, the municipality of Amersfoort might want to use a hydrological inundation model for reservoir management. Currently, the municipality has limited designated locations where they can temporarily store water to ease the flow to the urban water system. However, in the future they might want to use parking garages or roof tops to temporarily store water in case of heavy rainfall. The inundation model can then be used to decide whether or not to use these storages. The rainfall forecast can help with the storage management, such that is emptied again on time. For these types of applications, a detailed and accurate model is required as these measures are costly.

Based on the interview with the municipality of Amersfoort, a couple of requirements can be defined. Within Amersfoort, the different neighbourhoods can each handle a different rainfall intensity. How much a neighbourhood can handle, depends on its characteristics like the dimensions of the sewage system, size of paved area etc. Since these differences between the neighbourhoods are quite large, it is important that the model distinguishes in this. The threshold value that determines if an alert is given, should consider this as well.

Regarding the spatial resolution of the model, the municipality prefers a larger grid size as long as there are large uncertainties in the rainfall forecast. A grid size that is too small could lead to a false sense of security. Due to historic events and stress tests, the municipality knows the vulnerable locations that are prone to flooding. They can keep an eye on these locations during the rainfall event. According to the municipality, a grid size of about 1 km x 1km meets the needs. When the rainfall forecast is more accurate, a smaller grid size would be an option.

Warning the contractors and informing the inhabitants is a relatively simple measure that does not cost a lot of time and money to implement. The sooner the alert is send out, the more time the citizens have to prepare. An early warning is thus preferred, but not necessarily required for the municipality. A warning up to a couple of minutes before the flooding event still has additional value compared to no warning at all. The risk in warning relevant authorities and citizens is the false-alarm ratio: when an alert is send out while the event did not take place, people might not take the alert seriously in case of a future event.

Regarding the output of the model, mainly the expected inundation depth on surface level (timeseries) are of interest for the municipality. Also the locations of the manholes that will overflow are of interest. Preferably, this output is visualized on a 2D map that distinguishes between the neighbourhoods. The output variable should be relatively easy to interpret, also when communicating it to the inhabitants. When the hydrological inundation model output is interpreted by an expert from the municipality with knowledge about statistics, a probabilistic forecast is wishful as this communicates the uncertainty in the rainfall forecast.

2.3.2 Case study 2 – Municipality of Tilburg

On 15/02/2023 an interview with Jan Janssens-Baan (consultant on urban water management) and Jaap Jansen (interim policy advisor) from the Municipality of Tilburg took place.

When heavy rainfall is expected, the consultants of the municipality inform the on-call service (Dutch: consignatiedienst), the managers of the sewage system, and the management team of the municipality. The on-call services act when a calamity occurs: they perform assistance by for example calling the emergency services and putting up fences. The consults of the municipality alert the on-call service and the managers of the sewage system such that they are on stand-by in case they need to act. There are no measures that they can pro-actively take before the start of the rainfall event: they only respond to occurring events. In Tilburg, there is a combined sewage system for both sewage and urban surface runoff. In the period before the event, when there is no rainfall yet, these pipes are almost empty. Pumping has no additional value in that case. Besides warning the relevant authorities and possibly the inhabitants, there is not much the municipality can do before the rainfall event takes place.

A hydrological inundation model can support in this process by identifying the locations where inundation due to heavy rainfall is expected. Also the maximum flood depth of the forecasted event is of interest. In this way, relevant authorities are better informed beforehand. However, there are only passive measures available that minimize the damage: they can only facilitate the surroundings (e.g. putting up fences to close off inundated roads). The municipality would like to have a model that directly alerts the citizens and contractors, without the consultants as a step in between. The relevant authorities are warned sooner in that way. The reliability of the forecast is very important to the municipality of Tilburg. If there are a lot of misses and false hits, people will not take the warning seriously during a future event.

It is known that there are large differences between the neighbourhoods in how much rainfall the area can handle. However, it is unknown what rainfall intensity each area can handle specifically. It is important that the differences between the neighbourhoods are also considered in the model. The grid size of the model should thus be the size of a neighbourhood at maximum. However, when determining the grid size, also the spatial resolution of the rainfall forecast should be considered.

In the future, the municipality of Tilburg would like to use the model outcomes to alert inhabitants. They can take measures themselves as well to minimize the risk at their own property. The municipality knows that there are multiple people interested in these alerts. However, the municipality first want to use the model results for internal processes.

Since the municipality does not have a lot of measures that they can implement before the event takes place, they prefer to have a more accurate forecast with shorter lead times (compared to longer lead times with larger uncertainties). They would like to receive as much information as possible on the characteristics of the event (size, shape, duration), even when the event is expected to happen within tens of minutes. The required lead time is thus short: only a couple of minutes will already suffice. Since the consultants who analyse the models have knowledge about statistics, they are able to interpret the model output in a probabilistic format. The municipality sees the additional value of a probabilistic forecast as it is possible to communicate the uncertainty that comes with the rainfall forecast.

2.3.3 Case study 3 – Hoogheemraadschap van Rijnland

On 21/02/2023, an interview with René van der Zwan (consultant water management) and Jan Jelle Reitsma (consultant water quantity) took place.

Hoogheemraadschap van Rijnland has multiple models and tools in place that support them in water management. These include weather forecasts and Delft-FEWS. The latter controls their pumping stations and automatic weirs to ensure that the water levels are within the range as prescribed. The waterboard has multiple measures that they can take to reduce water levels. First of all, by pumping water to the boezem before the event, they can lower the water levels in the polder waterways which creates more storage for the rainfall later on. Besides, the hoogheemraadschap can temporally install additional pumping stations such that the pumping capacity is increased during the event. Lastly, they can inform and communicate with other relevant authorities about their approach, such as safety regions in case of extreme events.

If the waterboard is able to accurately forecast the event, they are able to anticipate based on this forecast. They can start with pumping as soon as the rainfall event is forecasted. Preferably, the model should provide lead times between 48 and 24 hours as this time is needed to significantly increase storage capacity in waterways due to pumping. However, the more time available, the better, as every hour of pumping will increase the storage capacity a bit more.

A hydrological inundation model that visualizes the expected inundation on a 2D map would be a useful addition for Hoogheemraadschap van Rijnland as their currently used models do not provide this: Delft-FEWS operates on polder level so no spatial variations within a polder are shown in the model output. In case of a predicted flooding, this spatial resolution does not allow for the identification of the location of the flooding. Therefore, a smaller grid size is preferred where the location of the flooding can be identified.

Waterboards base their decisions on the measured and expected water levels. Hoogheemraadschap van Rijnland can control the water levels in their study area and is also responsible for this. Therefore, they are mainly interested in the maximum water level that will be reached due to heavy rainfall, and if this is higher or lower than the level where damage occurs. The management of water levels is the order size of centimetres. It is therefore important that the predicted water level does not deviate more than 5 to 10 cm (maximum absolute error) from the observed value. If the capacity of the waterways is exceeded and the water overflows onto the land, Hoogheemraadschap van Rijnland would like to know where this occurs (location) and what the inundation depth on surface level will be.

A balance needs to be found between the required accuracy and the computation time. In case of an extreme rainfall event, the consultants from Hoogheemraadschap van Rijnland would like to receive an update every 30 minutes. This implies that the computational time should be within 30 minutes, including the computation for all ensemble members. When the model is extended for a larger study area of the waterboard, it is acceptable for the waterboard to run models simultaneously on multiple devices. If it is not possible to model the complete study area due to capacity issues, the waterboard would like to focus on the most vulnerable areas.

2.4 Discussion

The interviews show that there is not one model that is a perfect fit for every case study. Where the municipalities want to use the model to get more insight in the situation and warn the inhabitants and relevant authorities (e.g. contractors), the waterboard also wants to use the model to pro-actively take measures to minimize the flooding. This difference is reflected in the required lead time of the model: both municipalities prefer a longer lead time but still benefit from lead times shorter than 1 hour. On the other hand, Hoogheemraadschap van Rijnland requires a lead time between 48 and 24 hours because of their perspectives for action.

One common requirement is that the model results should show the location of the flooding on a 2D map. Besides the location of the flooding, also the flood depth, and the shape and duration of the flooding event are of interest. For both municipalities, the most relevant output variable is the expected inundation depth on surface level, while the waterboard is also interested in the maximum water level in the waterways for their perspectives for action. Both municipalities explain that there are large differences between the neighbourhoods in their area. Since each neighbourhood responds differently to a rainfall event, it is important that a model distinguishes in the neighbourhood's characteristics. The grid size of the model should thus not be larger than the area of a neighbourhood.

All three parties emphasize that the uncertainty in the rainfall forecast is an important issue. As long as there are large uncertainties in the rainfall forecast, a small model grid size does not have added value and could lead to a false sense of security. Besides, the uncertainty in the rainfall forecast could lead to a high false-alarm ratio. When alerting relevant authorities, the high false-alarm ratio might cause that alarms are not taken seriously. In case of the waterboard, the high false-alarm ratio has an even bigger impact whereby the water may be unnecessarily pumped into the boezem system which could cause problems in case of dry periods. Probabilistic rainfall forecasting has an added value according to both municipalities since it is a way to communicate the uncertainty in the forecast. It is thus wishful to include probabilistic forecasting in the hydrological inundation model to show the consequences of the uncertainty in the rainfall forecast.

2.5 Conclusion

For all three case studies, it holds that the end-user is interested in a hydrological inundation model that predicts the location of a flooding including the corresponding flood depth on a 2D map. They use this information to alert relevant authorities. Hoogheemraadschap van Rijnland has additional perspectives for action where they can pro-actively respond to a flood forecast (i.e. installing temporal pumping stations and pumping water into the boezem system). As a result, the waterboard is also interested in the water levels in water ways, as this information is needed in the decision-making process to these implement measures. In Table 1, the requirements per case study are summarized. In Section 3, three different surrogate models are created and validated. The quality of those models is assessed based on the descriptions below.

Table 1 - Requirements per case study

	#1 Municipality of Amersfoort	**#2 Municipality of Tilburg**	**#3 Hoogheemraadschap van Rijnland**
Modelling goal	Identify locations where inundation is expected including predicted flood depth	Identify locations where inundation is expected including predicted flood depth	Predict water levels in water ways & visualize expected flood depth on 2D map
Perspective for action	Alert contractors & citizens	Alert on-call service, managers of pumping stations & MT municipality	Pump water to the boezem system, install temporal pumping stations & alert relevant authorities
Output variable	Flood depth on 2D map: timeseries and maximum*	Flood depth on 2D map: timeseries and maximum*	Water level in water ways & flood depth on 2D map
Required lead time	Couple of minutes	Couple of minutes	24 – 48 hours
Grid size	Dependent on rainfall forecast, at maximum the size of a neighbourhood	Dependent on rainfall forecast, at maximum the size of a neighbourhood	5 – 100 meters*
Accuracy	Low false-alarm ratio (high critical success index*)	Low false-alarm ratio (high critical success index*)	Maximum absolute error of 10 cm in water ways
Maximum computational time	Order of minutes	Order of minutes	30 minutes

* This requirement is defined by the author (instead of by the end-user). The definition is based on expert judgement in combination with the results from the interviews.

3. Quality assessment surrogate model

3.1 Introduction

To answer the second research question, surrogate models are created for each case study to see if these models are able to meet the requirements that followed from the first research question. This chapter describes the creation and validation of the surrogate models including the quality assessment with respect to the requirements of the end-user. For each case study, a different type of surrogate model is used. The type of surrogate model is based on the available data and models, and the preference of the waterboard or municipality. An overview of this is shown in Table 2.

In Section 3.2, the general research methodology that is applied to all case studies is described. In Section 3.3, the methodology used for the first case study including the results and discussion can be found. Similarly, the methodology, results, and discussion for the second and third case study can be found Section 3.4 and Section 3.5 respectively. In Section 3.6, the overall conclusion is drawn where the quality of the surrogate models for each case study is assessed based on the requirements following from the first research question.

Table 2 - Type of surrogate model per case study

Case study	End-user	Problem context	Method surrogate model
1	Municipality of Amersfoort	The probability of flooding in urban areas	ML model based on hydrological inundation model
2	Municipality of Tilburg	The probability of flooding in urban areas	ML model based on hydrological inundation model
3	Hoogheemraadschap van Rijnland	The probability of flooding and water level in waterways including the interaction between open waters and inundation on surface level	Lower fidelity 1D2D inundation model

3.2 General research methodology

For each case study, the first step is to create and validate a surrogate model. The surrogate model is based on either a dataset (input and output variables) generated by a hydrological inundation model (case study 1 & 2) or the hydrological inundation model itself (case study 3). The specific method used to create, calibrate, and validate the different surrogate models can be found in Section 3.3.1 for case study 1, in Section 3.4.1 for case study 2, and in Section 3.5.1 for case study 3. The next step is to adapt the surrogate models to ensure that they can be used in an operational setting. This implies that the input of the surrogate model will be a probabilistic rainfall forecast including timeseries. The advantage of using a probabilistic rainfall forecast, compared to using a deterministic forecast, is that a probabilistic rainfall forecast is able to communicate the uncertainty in the forecast by using different ensemble members. For the purpose of this thesis, the probabilistic rainfall forecast KNMI Harmonie MOS (KNMI, 2023) is used, but other probabilistic rainfall forecasts will work as well. The KNMI Harmonie MOS forecast is a probabilistic rainfall forecast containing 50 ensemble members. It forecasts rainfall intensity timeseries [mm/hour] with lead times up to 48 hours, and the available output has a time interval of 1 hour. The output is in gridded form (latitude & longitude), and it covers the entire area of the Netherlands.

When using a probabilistic rainfall forecast as input for the surrogate model, the output will be determined for each of the ensemble members. In an operational setting, the surrogate model will be used in the form of a dashboard and therefore the output of the surrogate models should be post-processed such that it can be visualized on these dashboards (i.e. in only one figure). For all calculation points on the 2D map, the surrogate model output contains 50 ensemble members and possibly also timeseries. The post-processing of the ensemble members is done in a similar way as the W2O model. To process the 50 ensemble members, two different options are applied. Either a certain percentile of ensemble members per calculation point is calculated for each timestep, or the percentage of ensemble members that forecast a value (e.g. inundation depth) larger than a certain threshold value is computed. The percentile and threshold value can be easily adapted to the wishes of the end-user. Finally, if timeseries are available in the surrogate model's output (i.e. case study 1 & 3), the maximum of the surrogate model's output is taken over the considered time period as this reveals the most relevant information for the end-user, see Section 2.5.

To show how the surrogate model would be used in an operational setting, a heavy historic rainfall event is used as an example. On the 14th of July 2021, large parts of the Netherlands experienced heavy rainfall. Therefore, this historic event is used for all case studies as an example to show how the surrogate model output will be presented to the end-user. The available rainfall forecast for this event can be found in the results section of each specific case study.

Finally, the quality of the surrogate model is assessed using the requirements that followed from the first research question. For all criteria, the surrogate model is compared to the requirement. A description including a colour (see Table 3) that describes to what extent the requirement is met, is given. To determine if the surrogate model meets the requirement for the modelling goal and relevant output variables, a qualitative description is given. The grid size, computational time, and lead time of the surrogate model following from the model set-up will also be judged based on a qualitative description. For the computational time, the runtime of each model is determined using the same device (12th Gen Intel(R) Core(TM) i7-1270P 2.20 GHz processor and 32.0GB RAM). To determine if the surrogate models meet the standard for the required accuracy, three different performance indicators are used. For each case study, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the Critical Success Index (CSI) are calculated. This is based on the validation dataset where the surrogate model predictions are compared with those of the detailed hydrological inundation model. The detailed hydrological model inundation thus serves as a benchmark and is considered as the truth.

Table 3 - Legend with colours used in quality assessment of the surrogate models

Fully meets the requirement
Partially meets the requirement
Requirement not met

The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to assess the difference between the output of the surrogate model and the detailed hydrological inundation model (benchmark). The lower the value of the MAE and RMSE, the better the performance of the surrogate model. The RMSE emphasizes large errors between the surrogate model and the benchmark model. Equation 1 and 2 show how the MAE and RMSE can be calculated.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{predicted,i} - y_{benchmark\ model,i}\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{predicted,i} - y_{benchmark\ model,i}\right)^2} \tag{2}$$

Where $y_{predicted,i}$ is the output value in grid cell/node i according to the surrogate model, $y_{benchmark\ model,i}$ is the output value in grid cell/node i according to the benchmark model, and *n* is the total number of grid cells/nodes considered.

The critical success index (CSI) and accuracy are performance indicator that measures the model's ability to correctly predict flooding or not in each cell. If some cells/nodes are not flooded in most of the runs, this would typically lead to a high accuracy as the number of correct negatives is high. The CSI tackles this problem by not considering the number of true negatives. Therefore, the CSI is used as a performance indicator next to the generic accuracy equation. The CSI is calculated using equation 3, and the accuracy is calculated using equation 4:

$$CSI = \frac{\#True\ positives}{\#True\ positives + \#False\ negatives + \#False\ positives} \tag{3}$$

$$Accuracy = \frac{\#True\ positives + \#\ True\ negatives}{\#True\ positives + \#\ True\ negatives + \#False\ negatives + \#False\ positives} \tag{4}$$

Where a cell is considered positive if it is flooded and considered negative when it remains dry. To determine if a cell is flooded, an inundation threshold must be defined. Below this threshold, the cell is considered dry. For the inundation depth on surface level (case study 2 & 3), a threshold value of 0.05m is used, similarly to the work of Hop (2023) and Zanchetta & Coulibaly (2020). For the flood volume of a manhole (case study 1), a threshold value of $0.1m^3$ is used, similarly to the work of Kilsdonk (2021).

3.3 Case study 1 – Municipality of Amersfoort

3.3.1 Methodology

From the interviews with the municipality of Amersfoort, it can be concluded that they are interested in a model that forecasts the location and depth of inundation on a 2D map. Since the sewage system plays an important role in flood management in Amersfoort, it is important that the sewage system is considered in the model. In 2021, Kilsdonk successfully created a Long Short-Term Memory (LSTM) machine learning model that can be used as a flood early warning system for sewage systems (Kilsdonk, 2021). Kilsdonk has trained and validated the LSTM model for Hooglanderveen, a neighbourhood within Amersfoort. For the purpose of this thesis, this LSTM model will be used as a 'proof of concept' to answer the research questions.

3.3.1.1 Hydrological inundation model

The surrogate model is based on a calibrated and validated hydrological inundation model containing the sewage system. The hydrological inundation model was build using the software Infoworks Integrated Catchment Modelling (Infoworks ICM) (Kilsdonk, 2021). This numerical hydrological inundation model is a 1D2D model where the sewage system is included as a 1D component and the surface level system (streets, ditches, surface water channels) is included as the 2D component. The model uses the Shallow water equations to solve the 1D flow in the sewage system (Henonin et al., 2013). All relevant structures and properties of the sewage system are included in the model (e.g. pumps, manholes, overflows, pipes) (Kilsdonk, 2021). To determine the flow path from the surface level into the sewage system, the shortest flow path in combination with the area of the surface level are used. No topographic gradients are included in this hydrological inundation model. The hydrological inundation model uses rainfall timeseries as input, and has flood volumes per manhole in the sewage system (timeseries) as output. The flood volume is defined as the amount of water [m^3] that cannot be stored in the manhole due to exceeding capacity, and is therefore flooded onto the surface level. The flood volume per manhole is calculated for each timestep. Negative flood volumes indicate available storage inside a manhole.

3.3.1.2 Characteristics of the dataset

The validated hydrological inundation model is used as a benchmark to assess the accuracy of the LSTM model as there are no measurements (flood volumes or inundation depths) available. The LSTM model uses, similarly to the hydrological inundation model, rainfall timeseries [mm/hour] as input. The model is trained and validated by Kilsdonk (2021) for synthetic rainfall timeseries with a 5-minute timestep and 289 timesteps in total (=24 hours and 5 minutes). Table 4 and Figure 5 show the characteristics of the datasets used for training and validating the LSTM neural network. As can be seen in the figure, a wide variety of rainfall intensities, patterns, and durations are used to train and validate the model. Kilsdonk (2021) used 7 different rainfall patterns as input for the hydrological inundation model and LSTM model. Patterns contain either 1 rainfall peak, 2 rainfall peaks or are more uniformly distributed over time. The height of the peaks is based on a certain fraction of the cumulative rainfall, varying between 12.5% (more evenly spread-out event) to 87.5% (high peak). Besides the rainfall patterns, also the rainfall intensity and rainfall duration are varied. Rainfall intensities vary between 30 and 88.34 mm/hour and rainfall durations are either 4, 8 or 12 hours. In total, 126 rainfall events are synthetically created by Kilsdonk (2021) using this method.

Table 4 - Characteristics of the training and validation dataset for the LSTM model for case study 1

	Training dataset	**Validation dataset**
Number of events	101	25
Maximum rainfall intensity [mm/hour]	88.34	49.84
Maximum rainfall in 24 hours [mm]	105.00	103.16

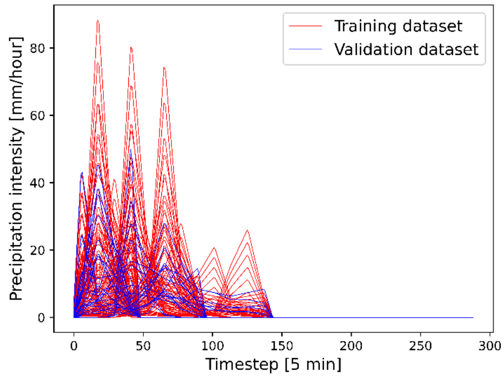


Figure 5 - Rainfall timeseries used for training and validating the LSTM model for case study #1

3.3.1.3 Training and validation machine learning model

Kilsdonk (2021) used a Bayesian Optimisation algorithm to find the optimal hyper parameters for the neural network. Bayesian optimisation is a global optimisation method for noisy black-box functions where a probabilistic model of the function mapping from the hyper parameter values to the objective function is created (Snoek et al., 2012). The method finds the optimum combination of hyper parameters by iteratively evaluating promising parameter combinations and updating the parameters based on their performance. Bayesian optimisation is chosen as method for the hyper parameter optimisation, since this method is more efficient in finding the best hyper parameter combination compared to other methods (Kilsdonk, 2021).

The final architecture resulting from the hyper parameter optimisation is presented in Table 5. On the validation dataset with synthetic rainfall events, the LSTM model has an MAE of 0.062 m^3, a Nash-Sutcliffe Efficiency of 0.84, and R^2 equals 0.99 (Kilsdonk, 2021). The LSTM neural network is also validated on four different historic events, which results in a MAE of 0.19 m^3, a NSE of 0.61, and a R^2 of 0.99. The performance indicators are also shown in Table 6. These lower values for the performance indicators on the historic events are due to the different rainfall patterns of the historic events compared to the synthetically created events (the historic dataset has higher peaks confined in a smaller timespan and more fluctuations/noise). The validation shows that the LSTM model has the tendency to underestimate high flood volumes, while the timing of the peak is predicted correctly (Kilsdonk, 2021).

Table 5 - Neural network architecture for case study 1, obtained from Kilsdonk (2021)

LSTM layer 1 #units	289
Dropout after layer 1	0 – 0.2
Learning rate	0.01
Dense layer #units	230 (nodes)
Batch size	10

Table 6 - Performance indicators following from the validation done by Kilsdonk (2021)

	Synthetic validation dataset	Historic validation dataset
MAE	0.0621 m^3	0.19 m^3
NSE	0.84	0.61
R^2	0.99	0.99

3.3.1.4 Quality assessment & practical implementation

For the purpose of this thesis, the LSTM neural network created and validated by Kilsdonk (2021) is adapted such that it is able to handle probabilistic rainfall forecasts as input (see Section 3.2). The used rainfall forecast (KNMI Harmonie MOS (KNMI, 2023)) has an hourly time interval while the LSTM model is trained with a 5-minute timestep. Besides, the KNMI Harmonie MOS rainfall forecast has a lead time up to 48 hours while the LSTM model requires an input of only 289 timesteps (=24 hours and 5 minutes). The rainfall forecast by the KNMI should thus be converted as it is required that it is in the same shape and format as the training data of the LSTM model. To do so, each rainfall intensity is repeated 12 times (12 times * 5 minute timestep = one hour). The consequence of this conversion is discussed in Section 3.3.2. Besides, only the first 289 timesteps of the forecast are used as this matches the length of the training data. The remaining part of the forecast is thus not considered resulting in a lead time of 24 hours for the LSTM model.

As explained in Section 3.2, the surrogate model is adapted such that it can be used in an operational setting where a probabilistic rainfall forecast is used as input. The output of the surrogate model is post-processed such that the results can be shown in a figure. This operationalized version of the surrogate model is used to assess the quality of the surrogate model using the requirements that followed from the first research question. For this, the modelling goal, output variables, accuracy, and computational time of the surrogate model are qualitatively compared to the requirements. See also Section 3.2 for a more elaborate description.

3.3.2 Results

The LSTM model created by Kilsdonk (2021) has been validated for 26 synthetic rainfall events and for four different historic events. However, for the purpose of this thesis, we zoom in on one specific event (11-08-2020) to show the consequences of the pre-processing of the rainfall timeseries and the post-processing of the machine learning model output. This event is selected as it shows the largest fluctuations in rainfall intensities within an hour (compared to the other three historic events), meaning that the pre-processing of the rainfall timeseries has the largest implications for the LSTM model output.

On August 11, 2020, the citizens of Hooglanderveen reported a flooding incident. The observed rainfall data is presented in Figure 6. The graph illustrates that the rainfall reached an intensity of more than 100mm/hour during 15:55 and 16:00 that day. When using a model in an operational setting, the rainfall forecast will be used instead of observed rainfall intensities. Since some rainfall forecasts (e.g. the KNMI Harmonie MOS forecast (KNMI, 2023) as used in this thesis) make use of a timestep of one hour instead of 5 minutes, the hourly average is taken from the observed rainfall to show the implications of the larger timestep. This yields the same amount of total rainfall within an hour, without the variation in the rainfall intensity. The hourly averaged rainfall is presented in red in Figure 6.

The hydrological inundation model serves as a benchmark to define the location and flood volumes for this event. The observed rainfall timeseries are used as input for the hydrological inundation model. Figure 7 shows the output of the hydrological inundation model where first the output (flood volume) is averaged over the hour, and then the maximum flood volume over time is computed for each manhole. As can be seen in Figure 7, flood volumes up to 90m^3 have been reached for certain manholes.

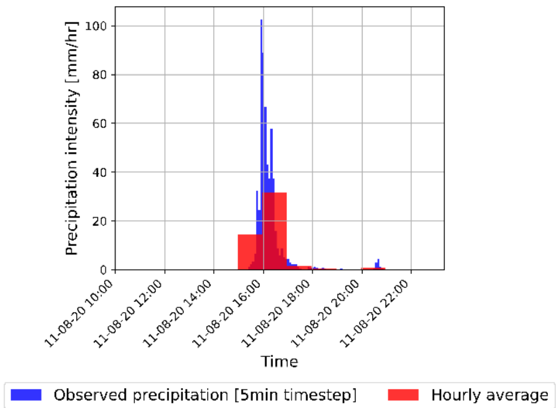


Figure 6 - Observed rainfall on 11-08-2020 (5 minute timestep) and the hourly average.

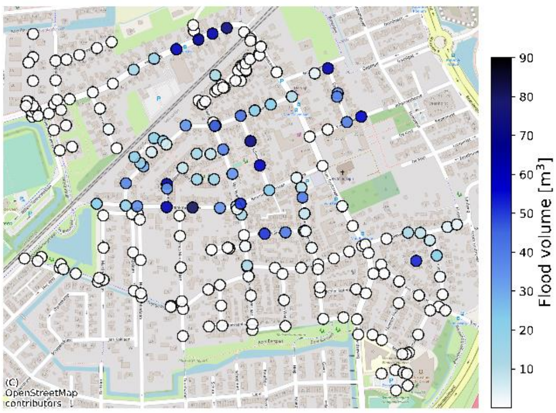


Figure 7 – Maximum flood volumes over time according to the hydrological model output (Infoworks ICM) when using observed rainfall timeseries (hourly average) as input. White nodes indicate manholes that are not flooded.

Figure 8 illustrates the output of the ML model when using the observed rainfall timeseries as input. Figure 9 shows the difference in flood volumes between the ML model output (Figure 8) and the hydrological inundation model output (Figure 7) when using the hourly average of the rainfall time series. Again, the maximum flood volume over time is computed for each manhole. The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Critical Success Index (CSI), and accuracy are calculated with respect to the output of the hydrological inundation model (benchmark). For the CSI and accuracy, a threshold value of 0.1 m^3 is used, similar to the work by Kilsdonk (2021), indicating that a flood volume lower than 0.1 m^3 represents no flooding. The results are presented in Table 7.

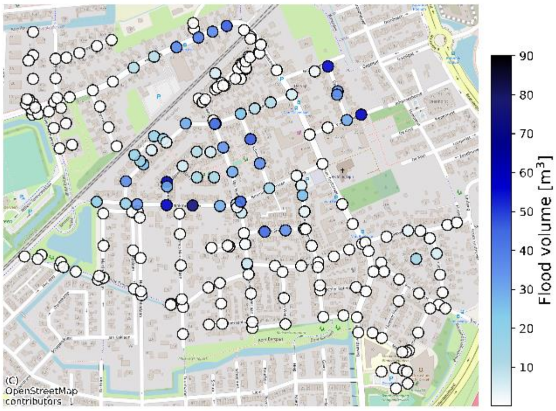


Figure 8 – Maximum flood volume over time according to ML model prediction when using observed rainfall timeseries (hourly average) as input. White nodes indicate manholes that are not flooded.

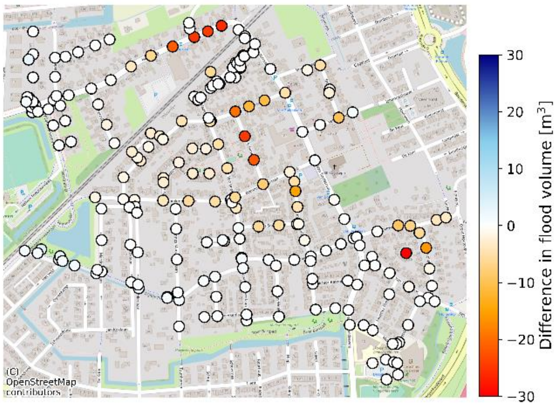


Figure 9 – Difference between ML model output (Figure 8) and hydrological model output (Figure 7). Negative flood volumes thus indicate an underestimation by the ML model.

Table 7 shows the performance of the ML model when using the hourly averaged rainfall timeseries with respect to the hydrological model (Infoworks ICM). Figure 9 and Table 7 show that the locations of the flooded manholes are predicted quite accurately. On average, the machine learning model is also able to accurately predict the flood volumes. However, there are some manholes where the ML model prediction is underestimating the flood volume as can be seen in Figure 9. With 230 manholes in a study area of 1 km^2, and around 50% of the total area being buildings (Klimaateffectatlas, 2021), a flood volume of 30 m^3 per manhole represents an inundation depth on surface level of around 14 mm. Similarly, a flood volume of 5.74 m^3 (which is the value of the RMSE), represents an inundation depth on surface level of around 3 mm. Translating the flood volumes to inundation depth on surface level shows that the errors in the ML model prediction are relatively small.

Table 7 - Performance indicators for the ML model when using the observed rainfall timeseries as input

	Accuracy ML model
RMSE [m^3]	5.74
MAE [m^3]	2.35
CSI [%]	81.52%
Accuracy [%]	92.61%

3.3.2.1 Practical implementation

As explained in Section 3.2, a probabilistic rainfall forecast will be used when the surrogate model is used in an operational setting. The rainfall event on the 14th of July 2021, which induced flooding in large parts of the Netherlands, serves as example. The KNMI Harmonie MOS rainfall forecast (KNMI, 2023) available at 12:00 that day for Hooglanderveen is shown in Figure 10. Note that the available rainfall forecast has a lead time up to 48 hours, but only the first 24 hours of this forecast are used as input for the ML model due to the required input format (see Section 3.3.1). This event serves as an example to show the implications of using a probabilistic rainfall forecast as input for the machine learning model.

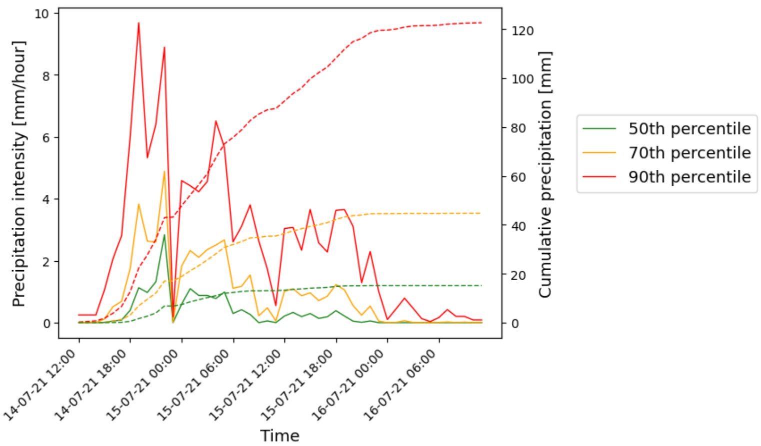


Figure 10 - Rainfall forecast available on 14-07-2021 at 12:00 for Hooglanderveen. Dashed lines indicate the cumulative rainfall over the 48 hours.

Figure 11 and Figure 12 show the output of the machine learning model when using the probabilistic rainfall forecast from 14-07-2021 12:00 as input. Figure 11 shows the 90th percentile of ensemble members. As explained above, a flood volume of 30 m^3 represents an inundation depth on surface level of around 13 mm. Figure 12 shows the probability that the flood volume exceeds a threshold of 0 m^3 within the next 24 hours, where the probability is defined as the percentage of ensembles that predict a flood volume higher than this threshold. These figures show what the end-user would see when the model is used in an operational setting.

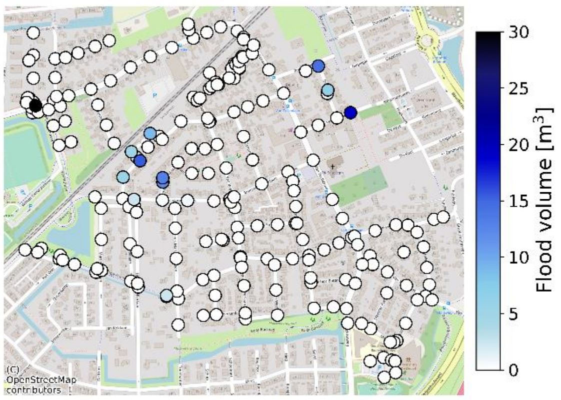


Figure 11 – Result ML model 90th percentile when using rainfall forecast as input. White nodes indicate manholes where no flooding is predicted.

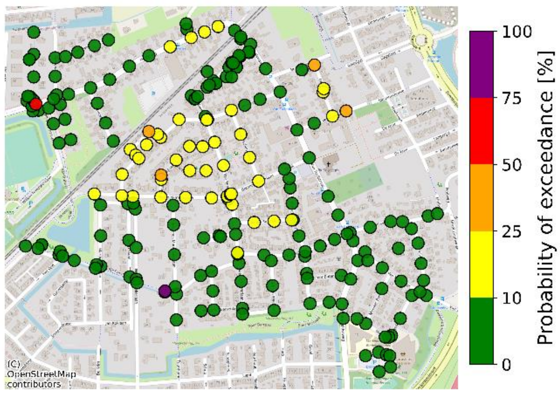


Figure 12 - Result ML model when using rainfall forecast as input. The probability of exceedance is defined as the percentage of ensemble members [%] that predict a flood volume higher than the threshold value of 0 m^3.

As there are no observations available for the historical event on the 14th of July, is it not possible to determine the accuracy of the machine learning model for this specific event. Besides, the rainfall forecast also adds an additional layer of uncertainty as the rainfall forecast itself could also be inaccurate. However, the validation of the ML model done by Kilsdonk (2021) has shown that the surrogate model is able to accurately reproduce the detailed hydrological inundation model, see also Section 3.3.1. Assuming that the rainfall forecast is correct, the machine learning model is thus able to accurately forecast the flood volume for each manhole in the sewage system.

3.3.3 Discussion

Model performance

The study done by Kilsdonk (2021) has shown that the LSTM model is able to rapidly and accurately reproduce the output of the hydrological model. The CSI score is used to assess if the surrogate model meets the accuracy requirement of the end-user (see Section 2.5). According to literature, a CSI score of at least 80% needs to be met in order to be suitable in operational flood management (Erechtchoukova et al., 2016; Zanchetta & Coulibaly, 2020). The surrogate model used in this case study has a CSI score of 81% and thus meets the accuracy requirement. Comparing the overall model performance with other studies is not possible, as there are no other studies found in literature that use an LSTM model to predict flood volumes in manholes.

An important limitation of the hydrological inundation model and thus also the LSTM model, is that it only considers overflowing of manholes due to heavy rainfall. However, as explained by the municipality in Section 2.3.1, there are multiple other causes (e.g. clogged manholes due to leaves) that could also lead to flooding. These are not considered and therefore the hydrological inundation model as well as the LSTM model might underestimate the event size, or even miss complete events. Besides, the ML model tends to underestimate the flood volumes in general, as can be seen in Figure 9 and is also described by Kilsdonk (2021). This implies that, if the rainfall forecast is correct, and the ML model predicts flooding, the actual event will likely have higher flood volumes than the prediction by the ML model.

Model output parameter

The output of the hydrological inundation model and the LSTM machine learning model is flood volume per manhole [m^3] for each timestep. This output variable is relatively difficult to interpret, especially for non-experts. As indicated by the end-user in Section 2.3.1, it is important that the output of the model is understandable for non-experts. If this model will be used in an operational setting, it is advised to reconsider the output variable in further research. One can for example think of a post-processing method that translates the flood volume to inundation depth on surface level by using an elevation map.

Limitations of rainfall data

A wide variety of synthetically created rainfall events is used to train and validate the LSTM model. However, the used rainfall dataset to train and validate the LSTM model has two main limitations. First of all, the rainfall event with the shortest duration has a duration of 4 hours. Peak rainfall events can have a much shorter duration (i.e. only a couple of minutes) (Hop, 2023; Schnitzler, 2022), but the LSTM model is not specifically trained for these short durations. In principle, machine learning models should not be used outside the scope they are trained for (Lange & Sippel, 2020; Zanchetta & Coulibaly, 2020). Furthermore, the synthetically created rainfall events do not have many fluctuations/noise over time. As this is the case for real rainfall events, the LSTM model performs less well on historic/real rainfall events.

For the purpose of this study, the KNMI Harmonie MOS forecast (KNMI, 2023) is used as input for the machine learning model. This forecast has an hourly timestep and a lead time up to 48 hours. A limitation of this rainfall forecast is the timestep of one hour. The municipality indicated that the shape/course of the rainfall event is important. With a large timestep, not all variations in the rainfall intensity are captured. This influences the model results: flood volumes can be underestimated when using an hourly timestep (see Appendix C). The ML model is able to handle rainfall timeseries with a 5-minute timestep, but the rainfall forecast by the KNMI falls short on this. Other rainfall forecasts with a shorter timestep could be used to overcome this shortcoming.

3.4 Case study 2 – Municipality of Tilburg

3.4.1 Methodology

From the interviews with the municipality of Tilburg, it can be concluded that they are interested in a model that forecasts the location and depth of a flooding on a 2D map. As the study area is an area with a dense and variable topography and infrastructure, a high spatial resolution is required to accurately model the flooding (Henonin et al., 2013). As neural networks have shown to be successful in accurately and rapidly predicting inundation depth (Berkhahn et al., 2019; Hop, 2023), it is decided to train a machine learning model for the study area based on an already available hydrological inundation model. In this section, the method used to train and validate the machine learning model is explained. Besides, the ML model is adapted such that it can be used in an operational setting (e.g. use probabilistic rainfall forecast as input and visualize the output).

3.4.1.1 Characteristics of the hydrological inundation model

A calibrated and validated numerical hydrological inundation model is used to generate the dataset for training and validating the machine learning model. This hydrological inundation model is created in the software Infoworks ICM (Arcadis, 2017; Breijn, 2012). Similarly to the hydrological inundation model in the first case study (Section 3.3.1.1), the hydrological inundation model is a 1D2D numerical hydrological model where the sewage system is included as a 1D component and the 2D surface level system is included as 2D component. The hydrological inundation model uses the Shallow water equations to solve the 1D flow in the sewage system (Henonin et al., 2013). In the hydrological inundation model, a rainfall-runoff component is included that determines the runoff based on the rainfall input data (Breijn, 2012). The topography of the area, the available storage on surface level, and the infiltration capacities for different types of paved area are all included in this hydrological inundation model.

Different compared to the hydrological inundation model used for the first case study, this hydrological inundation model also calculates the inundation depth on surface level as model output (Breijn, 2012). In this hydrological inundation model, the 2D surface level is not only used to determine the flow path from the surface level into the sewage system, but also the other way around. The 2D model reproduces the urban surface topography, allowing simulation of the flow spreading across the area (Henonin et al., 2013). The interaction between the sewage system and surface level flow are thus possible due to this.

In 2017, Arcadis has validated the hydrological inundation model (Arcadis, 2017). The hydrological simulations of historic events are compared to the (limited) available measurements and reports from inhabitants. The report by Arcadis (2017) concludes that the hydrological inundation model output is in line with the measurements. The hydrological inundation model is currently managed by Royal HaskoningDHV on behalf of the municipality of Tilburg. This external company also generated the dataset to train and validate the machine learning model.

3.4.1.2 Characteristics of the dataset

The dataset generated by the hydrological inundation model used to train and validate the machine learning model, has certain characteristics. Due to the large computational time of the hydrological inundation model, the size of the dataset is limited to 88 rainfall events. The impact of the size of the dataset will be discussed in Section 3.4.3. The input of the hydrological inundation model are rainfall timeseries, with most of them (~90%) being synthetically created timeseries (e.g. timeseries used for stress testing). Some of the events contain timeseries with an hourly timestep, others with a 15-minute or 5-minute timestep. The events also vary in length: the shortest rainfall events are only 1 hour long while the longest events have a length of more than 20 hours. The characteristics of the rainfall dataset are shown in Figure 13, Figure 14, and Figure 15.

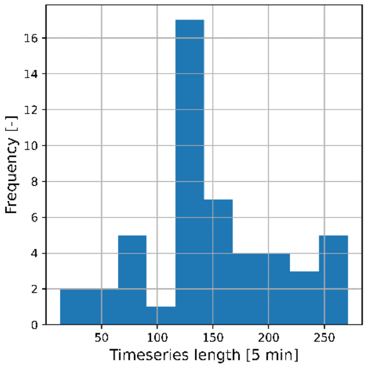


Figure 13 - Length of rainfall timeseries in the dataset

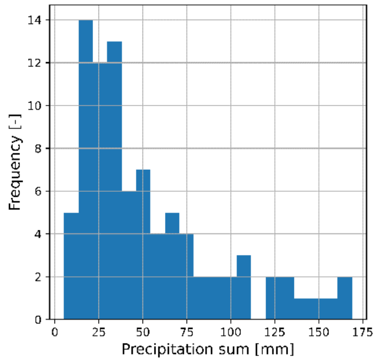


Figure 14 - Sum of rainfall timeseries in the dataset

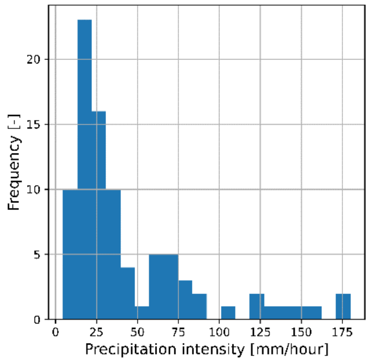


Figure 15 - Maximum rainfall intensity in timeseries

The output of the hydrological inundation model is a shapefile with a 2D grid containing the maximum water depth over time in that grid cell. The maximum water depth over time is chosen as output variable (instead of timeseries) as the dataset is quite small: using timeseries would result in an extra dimension for the machine learning model to understand.

For the machine learning model, a rectangular 2D grid will be used to plot the water depth as output. The hydrological inundation model uses a 2D irregular grid with different types of polygons, where each grid cell is of a different size (varying from $1m^2$ to $5m^2$). Since storing and post-processing an irregular grid are time-consuming and memory expensive processes, it is decided to convert the hydrological inundation model output to a rectangular grid. Besides, the number of grid cells is so large (>280,000), that it is costly to store this amount of data. Lastly, the resolution is unnecessary small for the usage in an operational setting as indicated by the end-user, see Section 2.3.2. Therefore, it is decided to first rasterize the output shapefile from the numerical hydrological model to a rectangular raster. For the purpose of this study, a grid size 5 m x 5 m is chosen but the spatial resolution could be decreased when this type of surrogate model would be used for a larger study area.

3.4.1.3 Data pre-processing

As the complete dataset needs to be in the same format (e.g. length and timestep), the rainfall timeseries are pre-processed by converting them to a 5 minute timestep. This is done by repeating each value multiple times until the timestep of 5 minutes is obtained (e.g. repeat 12 times to convert hourly timestep). Besides, the rainfall timeseries are extended to 288 timesteps by adding zeros to the end of the dataset until this length is reached. A length of 288 timesteps (=24 hours) is chosen as this is equal to the length of the longest rainfall event in the dataset, see Figure 13.

Both the rainfall timeseries as well as the water depths are normalized by dividing it by the maximum value of the dataset. There are grid cells that are never flooded in any of the model runs. When using the machine learning model in an operational setting, only grid cells that have the potential to be flooded, are of interest. When a cell is never flooded in the training dataset, this would imply extrapolation of the inundation depth. As a machine learning model should not be used for extrapolation purposes (Berkhahn et al., 2019; Zanchetta & Coulibaly, 2020), grid cells that are never flooded, are removed from the dataset. Before removing any grid cells, there are 282,338 grid cells in total. After removing the grid cells that are never flooded, there are only 28,114 grid cells left. This thus reduces the number of grid cells with > 90%, implying that the network will have to train less variables, making the training faster and less memory intensive.

Finally, the dataset is split in a training and validation dataset. It is chosen to use 90% of the total dataset for training the model, and 10% of the dataset for validating the model. This is in accordance with the study by Rajaee (2019), who reviewed 67 journal papers about machine learning methods. After splitting, the dataset for training contains 80 events and the validation dataset contains 8 events. The validation dataset is chosen such that the maximum value for both the rainfall intensity and rainfall sum, see Table 8, are within the range of the training dataset, as a machine learning model is likely to fail outside its training range (Lange & Sippel, 2020). Besides, the validation dataset is chosen such that it contains a variety of maximum rainfall intensities, timeseries lengths and rainfall sums. The maximum rainfall intensity present in each dataset and the maximum cumulative rainfall over the 24 hours are stated in Table 8.

Table 8 - Characteristics training and validation dataset

	Training dataset	Validation dataset
Number of events	80	8
Maximum rainfall intensity [mm/hour]	180	136.2
Maximum rainfall in 24 hours [mm]	168.98	93.60

3.4.1.4 Training and validation Machine Learning model

As explained in Section 1.2, neural networks have shown to outperform other machine learning algorithms to predict pluvial flooding (Bentivoglio et al., 2022; Kabir et al., 2020). Frequently used neural network include Long-Short Term Memory (LSTM) models (Hop, 2023; Kilsdonk, 2021) and Multi-Layer Perceptron (MLP) models (Berkhahn et al., 2019). Both LSTM models and MLP models make use of an input layer, one or more hidden layers, and an output layer (Zanchetta & Coulibaly, 2020). Each layer contains a certain number of neurons with a corresponding weight, which are connected to other layers via links. Each neuron receives input from other neurons and uses an activation function to process this. During the training of the model, the learning rate determines the step size per epoch (= iteration) at which the weights are adjusted. The main difference between a MLP model and a LSTM model, is the way they process the sequential data (Bentivoglio et al., 2022). A LSTM model is a type of recurrent neural network, meaning that the model structure can deal with long term dependencies by keeping track of previous states (Hop, 2023). A MLP model is a type of feedforward neural network, meaning that the data is processed in a sequential manner and predictions are only based on the current state (Berkhahn et al., 2019). To see which neural network type and set-up is most suitable for this case study, several set-ups have been tested.

First, the set-up of the Long-Short Term Memory (LSTM) model by Hop (2023) was used. The model set-up by Hop (2023) has shown to be successful in predicting timeseries of inundation depth on a 2D grid (10 m x 10 m resolution) with rainfall timeseries as input. The model set-up by Hop (2023) is adjusted such that it is able to handle the dataset for this case study (i.e. changing the spatial and temporal resolution of the machine learning model output). Also the model set-up of the LSTM model by Kilsdonk (2021) is tested for the dataset of this study. This model set-up is also adjusted such that it can handle the dataset for this case study.

LSTM models are specifically useful when predicting timeseries, as this model structure can deal with long term dependencies (Hop, 2023; Kilsdonk, 2021). However, as explained above, the provided dataset does not contain timeseries in the hydrological inundation model output. As a consequence, there are no long-term dependencies, and the LSTM model structure might not be the best fit for this specific case study. Therefore, also another machine learning algorithm is tested: a Multi-Layer Perceptron (MLP) model set-up. This algorithm is not specifically developed for long term dependencies (e.g. timeseries) might therefore better fit the purpose. First, the model set-up found by Berkhahn et al. (2019) was used. This MLP model is able to accurately predict maximum water levels during a flash flood event based on rainfall timeseries as input. In the study by Berkhahn et al. (2019), the model was successfully tested for spatially uniformly distributed synthetic rain events for two urban catchments. This model set-up is applied on the dataset for this case study.

Besides the model structure by Berkhahn et al. (2019), also two other MLP model structures are tested. These structures are found by doing hyper parameter optimisation. By varying the parameters as explained in the first paragraph of this section, and using an objective function, the optimal set of parameters is found leading to an optimal neural network architecture (Hop, 2023). Random search and Bayesian optimisation are hyper parameter optimisation methods that have shown to outperform other methods (Bergstra & Bengio, 2012). Therefore, first a Random search is used where the parameter combinations are randomly selected from the manually specified subset of hyper parameters space. Besides, also Bayesian optimisation is used to find the optimum model architecture. Bayesian optimisation is a global optimisation method for noisy black-box functions (Hop, 2023). Bayesian optimisation creates a probabilistic model of the function mapping from the hyper parameter values to the objective function (Snoek et al., 2012). The method finds the optimum combination of hyper parameters by iteratively evaluating promising parameter combinations and updating the parameters based on their performance. For both hyper parameter optimisations, a total of 500 iterations is done to find the optimum hyper parameter combination. As objective function, the Mean Squared Error (MSE) is used to find the best model architecture, similar to the work by Hop (2023) and Kilsdonk (2021).

For both hyper parameter optimisation procedures, a range of hyper parameters must be given. The range for the number of layers is set to a minimum of 1 and a maximum of 3 layers, next to the input and output layer. The range of nodes per layer is set between 30 and 1200, based on the study done by Kilsdonk (2021). The activation function for each layer is either 'relu', 'sigmoid', or 'tanh'. After each layer, a dropout is added with a value in the range of 0 to 0.5. Values in this range are common for the dropout rate (TensorFlow, 2022). Lastly, the learning rate of the model is varied between a minimum of $1*10^{-4}$ and 0.01, similarly as Kilsdonk (2021). An overview of the optimized hyperparameters and ranges is shown in Table 9.The hyper parameter optimisation using Random search and Bayesian optimisation both lead to an optimal model architecture for each of the optimisation methods.

Table 9 - Range of hyper parameters for hyper parameter optimisation

	Range of hyper parameters of parameter optimisation
Number of layers	1 – 3
Number of nodes per layer	30 – 1200
Activation function	Relu, sigmoid, tanh
Dropout	0 – 0.5
Learning rate	$1*10^{-4}$ – 0.01

3.4.1.5 Model validation

The performance of all model architectures as mentioned above is determined using the validation dataset. The accuracy is determined using the performance indicators as described in Section 3.2, where the neural network predictions are compared to the output of the hydrological inundation model. The model architecture with the best performance is used in the remaining part of this study. As explained in Section 3.2, the model set-up is adapted such that it is able to handle probabilistic rainfall forecasts as input. Finally, the quality of the operationalized ML model is assessed based on the requirements that followed from research question 1. For this, the modelling goal, output variables, accuracy, and computational time of the surrogate model are qualitatively compared to the requirements. See also Section 3.2 for a more elaborate description.

3.4.2 Results

As explained in Section 3.4.1.4 different types of machine learning model architectures are tested and compared to each other to see which model set-up results in the best performing model. The results of the comparison are described in Section 3.4.2.1. The best performing machine learning model is validated using the performance indicators (Section 3.4.2.2) and adjusted such that it can be used in an operational setting (Section 3.4.2.3).

3.4.2.1 Neural network architectures

After pre-processing the training and validation data, different types of neural networks are tested and compared to each other. First of all, the LSTM model created by Hop (2023) was used. It turns out that this LSTM model is not able to reproduce the training dataset: the model will always predict the mean, independently of the input variables. The application of the LSTM model by Hop (2023) is described in more detail in Appendix D1. When applying the LSTM model by Kilsdonk (2021), which was used in a second attempt, again the same water depth is predicted independent of the rainfall timeseries that is used as input. A more elaborate explanation of the results for the LSTM model by Kilsdonk (2021) can be found in Appendix D2. As explained in Section 3.4.1.4, LSTM models are specifically useful when predicting timeseries. However, the dataset used in this study does not contain timeseries in the hydrological model output (only the maximum), which is probably the reason why LSTM models are not suitable for this specific case study.

Next, three different MultiLayer Perceptron (MLP) architectures are used: the model set-up used in the study by Berkhahn (2019), a relatively simple set-up found using random search for the hyper parameter optimisation, and a more complex set-up found using Bayesian optimisation for the hyper parameter optimisation.

The model set-up found by Berkhahn (2019) performs reasonably well on the training data, but gives inaccurate predictions on the validation dataset. The results of this model set-up are presented in Appendix D3. A possible cause for the model functioning well on the training dataset but not on the validation dataset, is overfitting. Overfitting means that the machine learning model has found a perfect fit between training samples and their targets, which results in a mapping without any generalization power (TensorFlow, 2022). When making predictions on previously unseen data, the model will then not be able to accurately predict the target.

There are multiple solutions to handle overfitting (TensorFlow, 2022). The most effective way is to add more data to the training dataset. This can be done by either adding more model runs from the hydrological inundation model, or by using data augmentation techniques. Another technique to prevent overfitting is to use a simpler set-up for the machine learning model (fewer layers and nodes) such that the machine learning model simply does not have the capacity to learn the noise in the dataset. Therefore a simple MLP model is used next, where the hyperparameters are found using random search. However, it turns out that this led to a model architecture that is too simple, resulting in a machine learning model that is not able to understand any of the relationships between the input and output of the dataset, also not in the training dataset. The details on the hyper parameter optimisation using random search and the training of the model can be found in the Appendix D4.

As a third MLP model architecture, a more complex MLP model is used where the hyper parameters are found using Bayesian optimisation. After running the Bayesian optimisation for 500 iterations, the best hyperparameters for the model are found. This resulted in a MLP model with 1 layer with 855 nodes, the activation function 'relu', a dropout rate of 0.1 and a learning rate of 0.0037. The results are also presented in Table 10, and the python code for this model can be found in Appendix E. This model has a similar complexity compared to the model set-up by Berkhahn et al. (2019).

Table 10 - Range of hyper parameters for Bayesian optimisation and the final model architecture

	Range for Bayesian optimisation of hyper parameters	**Final model architecture after parameter optimisation**
Number of layers	1 – 3	1
Number of nodes per layer	30 – 1200	855
Activation function	Relu, sigmoid, tanh	relu
Dropout	0 – 0.5	0.1
Learning rate	$1*10^{-4}$ – 0.01	0.0037

After finding the optimal combination of hyper parameters, the model is trained and validated using the training and validation dataset. The machine learning model is able to accurately reproduce the trainings data as can be seen in Figure 16a. When looking at the scatter plot of the validation dataset in Figure 16b, a couple of observations stand out. First of all, the scatter plot shows that the machine learning model is able to predict water depths near the water depth according to the hydrological inundation model, with only a small underestimation. For larger water depths, the size of this underestimation increases. The last thing that stands out when looking at Figure 16b, are the dots that represent the dry grid cells (grid cells that are not flooded i.e. where the water depth equals 0m). For some grid cells, the machine learning model predicts a dry cell while the cell was actually flooded.

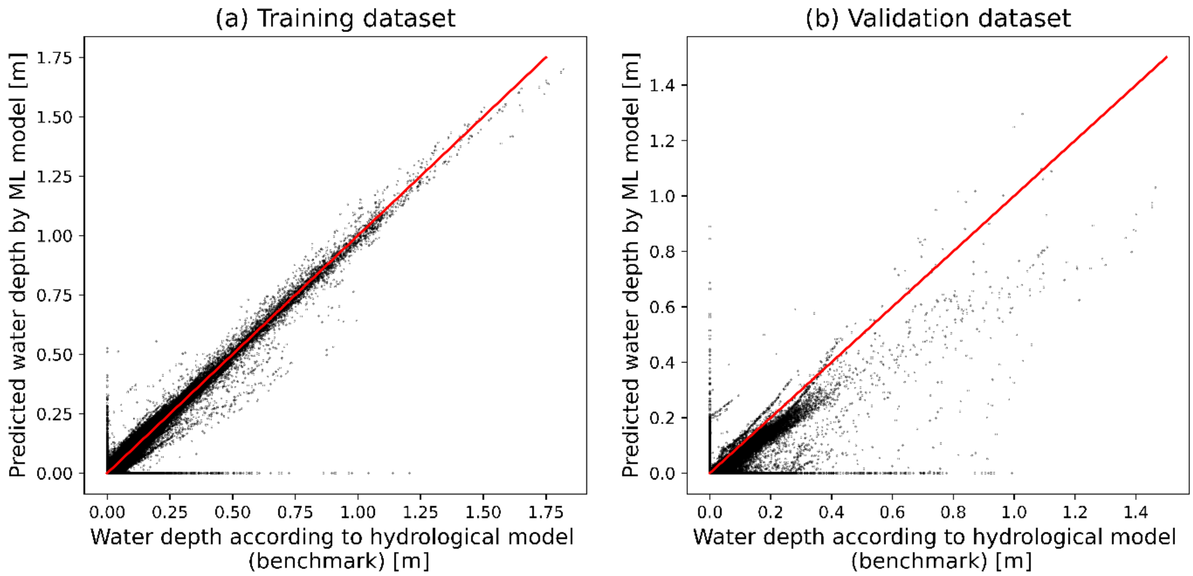


Figure 16 - Scatter plot for water depth according to hydrological model versus predicted water depth by the ML model. Left figure (a) is on the training dataset while the right figure (b) is on the validation dataset. Each black dot represents a grid cell where the water depth is calculated. The red line represents the perfect prediction: the predicted water depth by the ML model is equal to the water depth according to the hydrological model.

Since this MLP model is able to accurately reproduce the training dataset but performs less well on the validation dataset, it can be concluded that the ML model is again overfitted on the training dataset. This is confirmed by Figure 17, where the Mean Squared Error of the validation dataset (in orange) is increasing after around 20 epochs. This shows that the model is probably overfitted on the training dataset. It can thus be concluded that the relationships in the dataset are too complex to be understood by a simple model, while a more extensive model with more capacity is unable to generalize the relationships between the input and output data. The latter results in a model that is overfitted on the training dataset. Due to limited available time, the size of the training dataset is not increased. Instead, it is decided to continue with the best performing model: the MLP model with the hyper parameters as stated in Table 10. The consequences of the overfitted model are explained in Section 3.4.2.2.

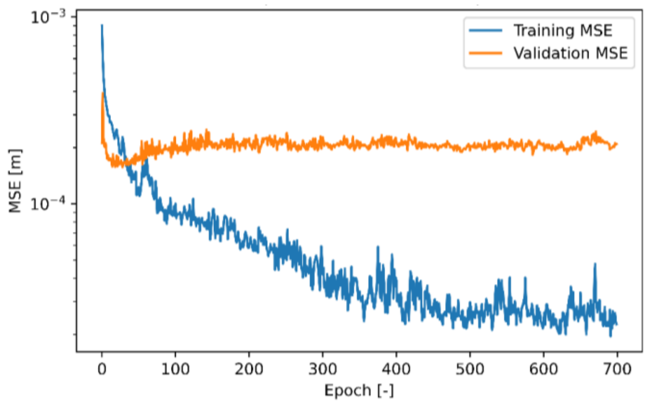


Figure 17 - Mean Squared Error (MSE) of the machine learning model over the epochs

3.4.2.2 Model performance indicators

Table 12 shows the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), the Critical Success Index (CSI), and accuracy of the machine learning model on the validation dataset. For the CSI, the threshold of flooding is set to 5 cm, similar to the work of Zanchetta & Coulibaly (2020), and Hop (2023). The low CSI score of only 48% shows what can also be seen in Table 11: there are a large number grid cells that were predicted as dry by the ML model, but were flooded according to the hydrological model (called misses). Due to the large number of correct negatives, the total accuracy of the ML model is high (94.7%) Also the values for the RMSE and MAE are small. The ML model is thus able to accurately predict the water depth in most grid cells (with a small underestimation, see Figure 16).

Table 11 - Number of true positives, true negatives, false positive and false negative grid cells in validation

	Actual dry cell	**Actual wet cell**
Predicted dry cell	202020 correct negatives	10038 misses
Predicted wet cell	1840 false alarms	11014 hits

Table 12 - Results accuracy assessment on training and validation dataset

	Training dataset	**Validation dataset**
RMSE [m]	0.0077	0.033
MAE [m]	0.0016	0.0087
CSI [%]	92%	48%
Accuracy [%]	99.1%	94.7%

In Figure 18, the Mean Absolute Error (MAE) is shown for every grid cell over the 8 validation runs. The figure shows that for most grid cells, the predicted water depth on average deviates only a couple of centimetres from the water depth according to the benchmark model. However, there are also some of grid cells (coloured red to blue in Figure 18) where the predicted water depth on average deviates more than 0.1m from the water depth according to the benchmark model. For these cells, the water depth is large according to the hydrological model output. The ratio between mean absolute error and the water depth according to the hydrological model output is thus relatively small. This is confirmed by Figure 19, which shows the relative MAE where the MAE is divided by the average water depth according to the hydrological model in that grid cell. When looking at the relative MAE, the error is about spatially uniform, and no clear patterns can be seen. For most grid cells, the error is 2 times smaller than the water depth in that cell. On average, the machine learning model output deviates 65% from the hydrological model output.

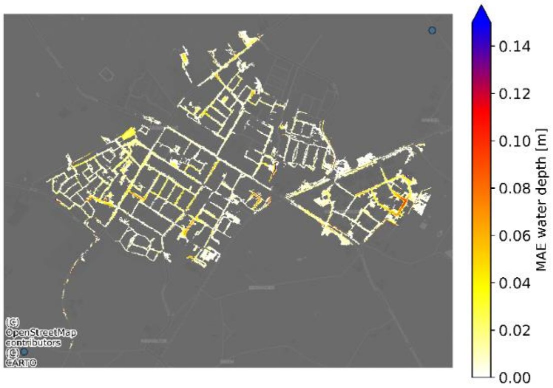


Figure 18 - Mean Absolute Error (MAE) over the 8 validation runs. The predicted water depth according to the ML model is compared to the hydrological inundation model output.

Figure 19 - Relative MAE over the 8 validation runs. The MAE in each cell (see Figure 18) is divided by mean water depth according to the hydrological inundation model output in that cell. E.g. a value of 1 indicates that the MAE is as large as the average water depth in that cell.

3.4.2.3 Practical implementation

As explained in Section 3.2, a probabilistic rainfall forecast will be used as input for the machine learning model when the model is used in an operational setting. The rainfall event on the 14th of July 2021, which induced flooding in large parts of the Netherlands, serves as example. The KNMI Harmonie MOS rainfall forecast (KNMI, 2023) available at 12:00 that day for Udenhout is shown in Figure 20. Note that the available rainfall forecast has a lead time up to 48 hours, but only the first 24 hours of this forecast are used as input for the ML model as this is the required input format (see Section 3.4.1.3). This event serves as an example to show the implications of using a probabilistic rainfall forecast as input for the machine learning model.

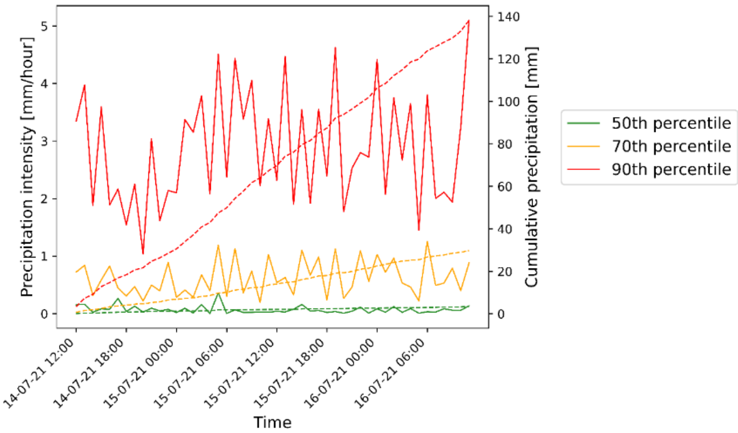


Figure 20 - Rainfall forecast available on 14-07-2021 at 12:00 for Udenhout. Dashed lines indicate the cumulative rainfall over the 48 hours.

Figure 21 and Figure 22 show the results of the machine learning model when using the probabilistic rainfall forecast as input. An enlarged version of these figures can be found in Appendix F. Due to the lack of available observational data, it is not possible to determine the accuracy of the machine learning model for this specific historic event. The accuracy of the surrogate model is one of the end-user's criteria in the quality assessment. For this, the results of the validation with respect to the hydrological inundation model (see Table 12) are used instead.

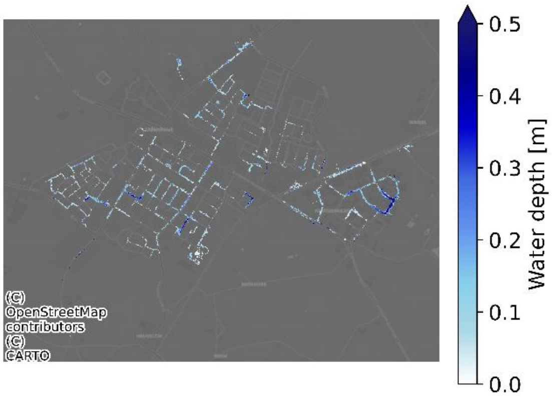


Figure 21 – Maximum water depth according to 90th percentile of ensemble members

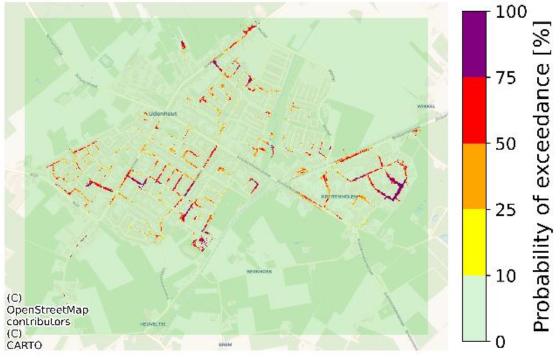


Figure 22 – Probability that the water depth exceeds a threshold of 0m based on the probabilistic rainfall forecast. Probability is defined as the percentage of ensemble members that forecast a value larger than the threshold.

3.4.3 Discussion

Model performance

The MLP machine learning model created for this case study is able to rapidly (order of milliseconds) predict the flood depth and location of flooded grid cells. The MAE of the model on the validation dataset shows that, on average, the model is able to predict the water depths near the water depth according to the hydrological inundation model. When analysing the RMSE, which emphasizes large errors, it can be seen that the model performs less well. This is also shown in Figure 16, which shows that the model has difficulties with predicting large water depths. The CSI score is the performance indicator used to assess if the surrogate model meets the accuracy requirement (see Section 2.5). According to literature, a CSI score of at least 80% needs to be met in order to be suitable in operational flood management (Erechtchoukova et al., 2016; Zanchetta & Coulibaly, 2020). The CSI of only 48% for this surrogate model is thus insufficient. The model output should be handled with care as the validation dataset has proven that the model can be inaccurate. The relatively low accuracy of the surrogate model is not in line with the accuracy of machine learning models found in other studies (Berkhahn et al., 2019; Hop, 2023; Zanchetta & Coulibaly, 2020). The paragraph Model overfitting below will elaborate further on this.

Dataset requirements

The dataset used for training and validating the surrogate model is relatively small. It was therefore decided to use only maximum water depths as output, instead of timeseries. This implies that there is thus no information on when these maximum water depths will take place. It is advised to communicate the rainfall forecast with the output of the surrogate model to get a sense of the course of the event. Despite the small size of the training and validation dataset, the variety in the characteristics of the dataset is good. As explained in Section 3.4.1.2, a wide variety of maximum rainfall intensities, event durations, event courses, and total rainfall intensities is used. So even though most of the rainfall events are synthetically created, they do represent a large number of possible events. This implies that the machine learning model is thus trained for many different possible rainfall events and is widely applicable.

The machine learning model is trained on a dataset where the input data (rainfall timeseries) contains 288 timesteps with a 5-minute time interval. Since it is required that the input data always has the same shape and format, it is thus needed to pre-process the KNMI Harmonie MOS rainfall forecast as also explained in Section 3.4.2.3. When a rainfall forecast with a timestep of 1 hour is used, like the KNMI Harmonie MOS rainfall forecast, the full potential of the machine learning model is thus unused. Other rainfall forecasts with different timesteps could be used to overcome this shortcoming.

Model overfitting

As explained in Section 3.4.2, the model is overfitted on the training dataset implying that the model is not able to generalize the relationships between the input and output data. As a consequence, the model will not perform well on unseen data like the validation dataset. To prevent overfitting, more data should be generated to train the model such that the model has more data to learn the complex relationships between the input and output data. Another suggestion for future research is to reduce the number of input and/or output variables. This could be done in two different ways. First of all, the hydrological inundation model output could be rasterized to a larger grid size (e.g. 10 m x 10 m instead of 5 m x 5 m). In this way, the number of variables to train is reduced, making the training less intensive. Another option would be to apply classification, where the inundation depths or precipitation timeseries are allocated to a certain fixed number of classes (e.g. inundation depth between 0-5cm, between 5-10cm etc., or classes for different rainfall intensities). As this reduces the complexity in the dataset, a more simple model set-up (e.g. as used in Appendix D4) could be sufficient. Also other types of neural networks could be used to predict inundation (e.g. conventional neural networks, which can deal with raster images (Hop, 2023)). For future research, it is advised to investigate these possibilities as neural networks have proven to be successful in other studies (Berkhahn et al., 2019; Erechtchoukova et al., 2016; Hop, 2023; Zanchetta & Coulibaly, 2020). It is expected that generating more data or simplifying the input-output relationship will lead to a more accurate ML model. Cloud computing techniques could be used to rapidly generate more data (Hop, 2023).

3.5 Case study 3 – Hoogheemraadschap van Rijnland

3.5.1 Methodology

For Polder Vierambacht, a hydrological inundation model was not yet available. Since the detailed hydrological inundation model will serve as a benchmark to validate the surrogate model, the first step is to create this detailed hydrological inundation model. In Section 3.5.1.1 till Section 3.5.1.4, the methodology for creating, calibrating, and validating the hydrological inundation model is explained. Section 3.5.1.5 describes how a surrogate model is created based on this detailed hydrological inundation model. A 1D-2D hydrological inundation model with a coupled rainfall runoff component is created for this study. This is done using a Python package called D-HyDAMO (Deltares, 2022b) where a hydrological inundation model can be automatically created using a Python script. This script also allows for easily changing model parameters, which is useful for calibration purposes and the creation of surrogate models.

A 1D2D coupled model is used where the waterways are modelled as 1D links using data on the cross sections. Water enters the waterways via the Rainfall Runoff (RR) component: this translates the rainfall input to infiltration into the soil and runoff to the waterways. Water can exit the system via one of the two pumping stations. Lastly, a 2D grid is added which calculates the flow of water over land. A short explanation on the different components is given below.

3.5.1.1 Rainfall Runoff (RR) model

The modelling of rainfall-runoff processes is done using the rainfall runoff component in the hydrological inundation model (RR model). This model simulates the hydrological processes in both rural and urban areas during dry and wet conditions. Within the study area, there are in total 37 catchments based on the drainage level regions. Each catchment is linked to three different 'buckets': paved areas, unpaved areas, and open waters. Also the different soil types (BRO, 2020) and seepage (OWASIS, 2023) in the study area are considered. At the RR links, water is transferred from the RR model to the 1D component (i.e. the waterways) of the model. Depending on the ground water level and the water level in the water ways, water can interchange in both directions. Table 13 shows an overview of the used data for the RR model including the source.

Table 13 – Overview required data for RR model

Data	Description	Date	Reference
Digital elevation map	Raster with elevation in every cell (5 m x 5 m). Used to determine flow path	2020	(AHN, 2020)
Land use	Land use raster (1 m x 1 m) to determine paved area, unpaved area & open water.	2020	(STOWA, 2020)
Soil type	Raster (5 m x 5 m) with different soil types in the study area. Used to calculate infiltration.	2020	(BRO, 2020)
Seepage	Seepage, constant and assumed to be spatially uniform over study area	2023	(OWASIS, 2023)
Rainfall	Timeseries with observed rainfall, assumed to be spatially uniform over study area	2022	(HydroNET, 2023)
Evaporation	Timeseries potential evaporation, assumed to be spatially uniform over study area	2022	(HydroNET, 2023)

3.5.1.2 1D model

The 1D component in the hydrological inundation model represents the channel flow. Profile measurements are available for the primary waterways, and the database of the waterboard (Dutch: legger data) is used to assume cross sections for the remaining waterways. There are also several structures included in the 1D model, namely: culverts, weirs, and two pumping stations. Target water levels (Dutch: streefpeilen) are used as operating levels for the weirs and pumping stations (see Figure 23). Two boundary conditions are added to the model: downstream the two pumping stations, the water level is fixed at the target water level. This implies that the full discharge capacity of the pumping stations is assumed to always be available. No upstream boundary conditions were added, as water can only enter the system via rainfall/seepage (there are no additional inlets). In case of very dry periods, water could be pumped from the boezem into the polder via one of the two pumping stations. In the model, 1D calculation nodes are placed every 20 meters and 1 meter upstream and downstream of every structure. An overview of the used data for the 1D model can be found in Table 14. Figure 23 shows a schematisation of the waterways and locations of the structures.

3.5.1.3 2D model

The hydrological inundation model also consists of a 2D component, which is used to simulate water flow over land. For this, a structured rectangular grid is used with a 20 meter x 20 meter resolution as a starting point for the benchmark model. It is decided to not refine the grid around the waterways since there are many waterways, which would thus result in large parts of the grid needed to be refined. Instead, an overall higher resolution grid is used. For each grid cell, the roughness is determined based on the land use type, and using a conversion table in the D-HyDAMO toolbox (Deltares, 2022b). At every calculation point of the 1D model, the 1D and 2D model exchange water via 1D2D links. An overview of the used data for the 2D model can be found in Table 14.

Table 14 - Overview required data for 1D and 2D hydrological inundation model

Data	Description	Date	Reference
Digital elevation map	Raster with elevation in every cell (5 m x 5 m). Used to determine flow path on 2D grid	2020	(AHN, 2020)
Land use	Land use raster (1 m x 1 m) to determine roughness on 2D grid	2020	(STOWA, 2020)
Waterway data	Locations and dimensions of waterways in the study area	2022	Rijnland
Structure data	Locations and dimensions of structures: culverts, weirs, and pumps	2022	Rijnland

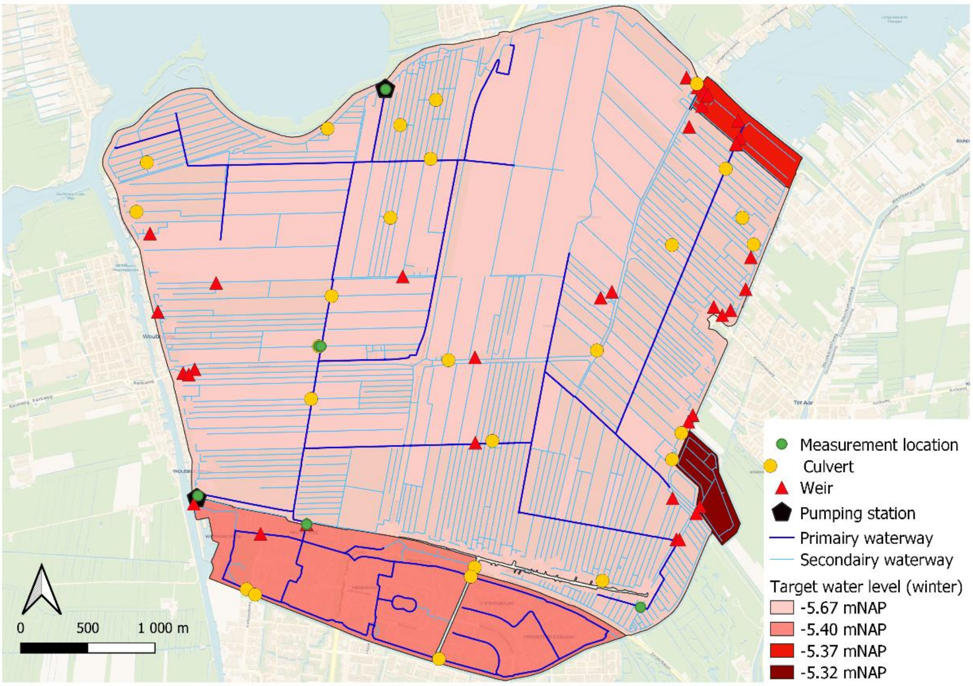


Figure 23 – Structures, waterways, measurement locations and target water level areas in Polder Vierambacht.

3.5.1.4 Model calibration & validation

After setting up the detailed hydrological inundation model, the hydrological inundation model is calibrated and validated. The goal of the calibration is to determine the values for uncertain model parameters such that the model output matches the observations. For the model calibration, two different events are used: the week of February 21st 2022, and the week of June 24th 2022. The rainfall timeseries for these events can be seen in Figure 24 and Figure 25. For these events, measurements of water levels are available at multiple locations (5 in total, see Figure 23) across the study area. At one of the two pumping stations (the southern pumping station), also discharge measurements are available. Based on an exploratory sensitivity analysis and expert judgement, four parameters are selected for the calibration (see Table 15). These parameter values are spatially uniform and there are thus no local deviations within the study area. The values of the parameters are varied within the range as specified in Table 15. By comparing the model output to the available measurements and calculating the MAE (see Section 3.2, equation 1) and the RMSE (see Section 3.2, equation 2), the combination of calibrated parameters that result in the lowest MAE, is found. Finally, the hydrological inundation model is validated using a third event: November 15th 2022, which is shown in Figure 26. This event is used to determine the accuracy of the benchmark hydrological inundation model.

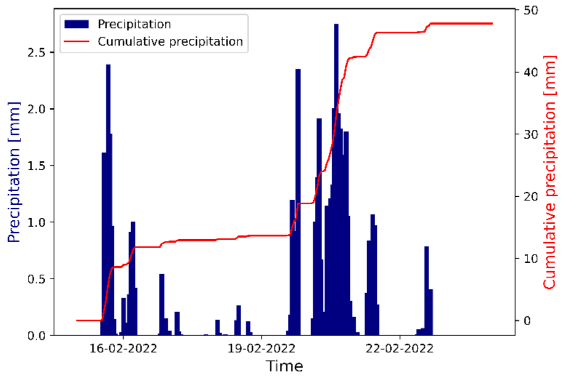


Figure 24 - Rainfall timeseries for event #1 used for hydrological inundation model calibration

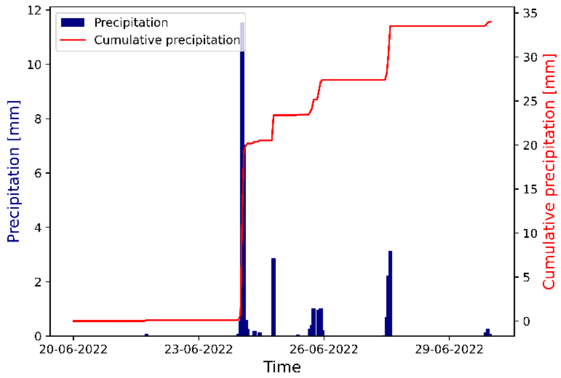


Figure 25 - Rainfall timeseries for event #2 used for hydrological inundation model calibration

Table 15 - Parameters used in calibration. The minimum and maximum values indicate the range used in the calibration.

Parameter	Description	Minimum value	Maximum value	Unit	Reference
1D channel roughness	Bed roughness in 1D waterways (Strickler)	10	30	$m^{1/3}s^{-1}$	(Hop, 2023)
Channel width	Width of the channels where no cross section was available in the provided dataset	0.5	2	m	Satellite imagery
Drainage resistance	The ability to resist water flow underground for each soil layer	0.5	2	mm	(Hop, 2023)
Drainage layer depth	The depth of the soil layers with respect to surface level	0	2	m	Expert judgement

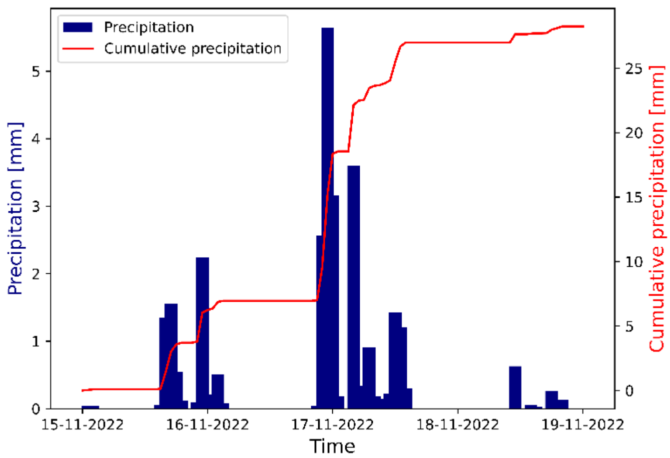


Figure 26 - Rainfall timeseries used for validation of the hydrological inundation model

3.5.1.5 Creation surrogate model

The detailed hydrological inundation model as described above is not suitable for usage in an operational setting due to the long computational time (> 100 minutes). Several adaptations are done to decrease the computational time. First of all, a restart-file is used instead of a long warm-up period. Using a restart-file allows for a shorter calculation time frame, thus less time steps resulting in a shorter computational time. The restart-file describes the initial conditions of both the 1D, 2D and RR model. By continuously updating the restart-file with observations, the initial conditions of the hydrological inundation model will match reality (Clark et al., 2008). This method is called data assimilation. Since the initial conditions of the hydrological model will (almost) match reality due to these techniques (Barthélémy et al., 2018; Clark et al., 2008), the error of the hydrological inundation model is increased only minimally. On average, the error increases 1 cm for the water levels in the waterways and 1 mm for the inundation depth on the 2D grid. Since the error of the hydrological inundation model including a restart-file is thus not significantly increased, the version of the hydrological inundation model where a restart-file is included, is considered as the benchmark for the remaining part of this thesis. The computational time of the detailed hydrological inundation model where a restart-file is included, equals 10.6 minutes to simulate 48 hours.

Next, five different simplifications are applied to further decrease the computational time:

- Increase distance between the 1D calculation points. In the benchmark model, the distance between the points equals 20 m. In the creation of the surrogate models, this distance is varied between 20 m and 200 m, see Table 16.
- Decrease spatial resolution of the 2D grid. The benchmark model has a rectangular grid with cells of 20 m x 20 m. For the surrogate models, the grid size is varied between 20 and 200 meters, see Table 16.
- Replace secondary waterways by retentions. As can be seen in Figure 23, the study area contains primary waterways and secondary waterways. The primary waterways are the main branches and are responsible for discharging the water to the pumping stations. The secondary waterways are mainly there to store the water until it flows into the primary waterways. By removing the secondary waterways from the hydrological inundation model, the number of branches and thus the number of 1D calculation nodes is decreased. The difference in storage due to the removal of the secondary branches is compensated by adding retentions to the surrogate model with a similar storage as the secondary branches in that area.
- Increase user timestep. The user timestep specifies the interval with which the meteorological forcings are updated (Deltares, 2023). This is not the same as the calculation timestep of the model, which is dynamic and dependent on the maximum Courant number.
- Increase maximum Courant number. The Courant number is calculated using equation 5.

$$C = \frac{U\,\Delta t}{\Delta x} \leq C_{max} \tag{5}$$

Where C is the Courant number [-], U is the flow velocity [m/s], Δt is the computational timestep, Δx is the grid length [m], and C_{max} is the maximum Courant number. When $C_{max} \leq 1$, this means that the distance that any information travels during the timestep length within the mesh, is equal or lower than the distance between mesh elements (Caminha, 2023). For the benchmark model, C_{max} is set to 0.7, which is the default value in D-HYDRO (Deltares, 2023). In literature, values up to 400 have been used for C_{max} (Aricò et al., 2016). However, it is important to check if the hydrological model does not become unstable when using a C_{max} larger than 1 (Aricò et al., 2016; Caminha, 2023). The size of the computation timestep Δt is automatically computed by the computational kernel of the model. A higher value for C_{max} allows for a larger computational timestep, and thus a shorter computational time overall.

An overview of the parameters and their ranges can be found in Table 16. A type of simplification that was considered but decided to not apply, is the simplification of the 2D shallow water equations. The diffusive wave equations could be used instead of the shallow water equations if flow separation and turbulence eddies can be neglected (Bomers et al., 2019). The study by Bomers et al. (2019) has shown the use of the diffusive wave equations resulted in a significant reduction of the computational time compared to using the shallow water equations in the hydrological modelling software HEC-RAS. However, in the software used for this thesis (D-HYDRO), the numerical equations to compute the flow characteristics cannot be changed easily. Therefore, it is decided to not apply this type of simplification in this study and use the five other types of simplifications instead.

Table 16 - Overview simplifications to create surrogate model and the range of the variations

	Range
Distance between 1D nodes [m]	20 – 200
Spatial resolution squared 2D grid [m]	20 – 200
Including secondary branches?	Yes/No
Courant number [-]	0.7 - 30
Timestep (user) [min]	10 – 30

By using different combinations of the simplifications, surrogate models are created. To find out the accuracy of the surrogate models with respect to the benchmark model, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Critical Success Index (CSI) are calculated. See Section 3.2 for the equations. Changing the distance between the 1D and/or 2D calculation points lead to a different number of calculation points and different coordinates for the calculation points. To compare the calculation points of the surrogate model with the calculation points of the benchmark model, the nearest calculation point is taken. Any additional calculation points in the benchmark model are then ignored.

Also the computational time of the surrogate models is determined. The computational time is partly dependent on the number of flooded 2D grid cells. The calculation points on the 2D grid only increase the computational time when these grid cells are actually flooded. To obtain a realistic value for the computational time in case of a flooding, it is thus important that an event is chosen that leads to a flooding. As there are no historic events known that have led to a flooding, a synthetic rainfall event is used. The event is shown in Figure 27.

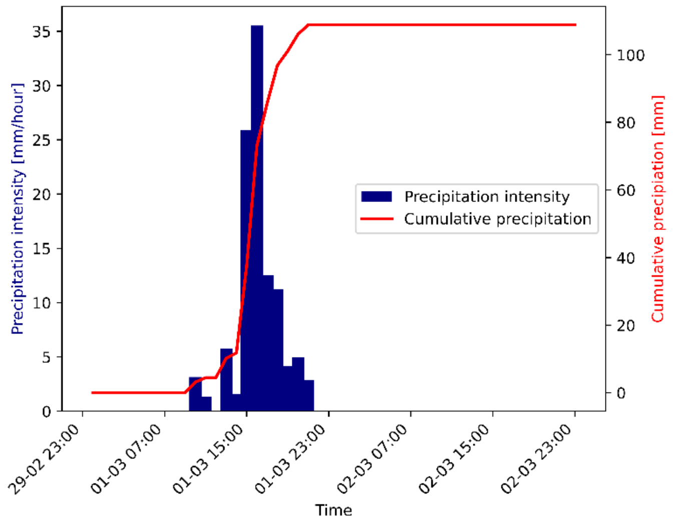


Figure 27 - Rainfall event used for testing the surrogate models

The characteristics of the surrogate models including their accuracy is then compared with their computational time to find a combination that meets the end-user's requirements (Section 2.5). As explained in Section 3.2, the best-performing surrogate model is adapted such that it can handle probabilistic rainfall forecasts as input. Finally, the quality of the surrogate model is assessed based on the requirements that followed from research question 1. For this, the modelling goal, output variables, accuracy, and computational time of the surrogate model are qualitatively compared to the requirements.

3.5.2 Results

As explained in Section 3.5.1, first a D-HYDRO hydrological inundation model is created using the D-HyDAMO toolbox (Deltares, 2022b) in Python. Next, the hydrological inundation model is calibrated and validated. The results of the calibration and validation are described in Section 3.5.2.1. In Section 3.2.3.2, the results of the different surrogate models are described and their accuracy and runtime is compared to the benchmark model. Finally, as explained in Section 3.2, the best-performing surrogate model is adapted such it can be used in an operational setting where a probabilistic rainfall forecast is used as input (Section 3.5.2.3).

3.5.2.1 Calibration & validation detailed hydrological inundation model

In this section, the results of the calibration and validation of the benchmark model are discussed. Four model parameters are calibrated. The resulting values are shown in Table 17. Besides these four model parameters, a couple of other adjustments have been made (manually) to the raw data to ensure that the hydrological model better represents reality. First of all, it was noticed that the water levels of the waterways in the urban area were significantly higher (>12cm, see the blue coloured graph in Figure 28) than the observed water levels after a rainfall event. A quick analysis showed that large amounts of water were entering the waterways from the paved areas in the catchments. Since the urban area has a separated sewage system, all rainwater is (via the sewage system), directly spilled into the open water instead of going to a waste water treatment plant. The amount of water running of from the paved area is directly related to the size/area of the paved area. The amount of paved area is determined using the land use raster (Deltares, 2023). A comparison between google satellite imaginary, Klimaateffectatlas (Klimaateffectatlas, 2021) and the land use raster shows that the amount of paved area is significantly overestimated (>15% for some parts). For example, the backyards of some houses are indicated as 'paved' in the land use raster while google satellite images show that this is grass and thus unpaved. Using Klimaateffectatlas, which gives an estimation of the amount of paved area per neighbourhood, the paved area is decreased up to 15% for some catchments. The model results after the correction for the paved area are shown in Figure 28 in green.

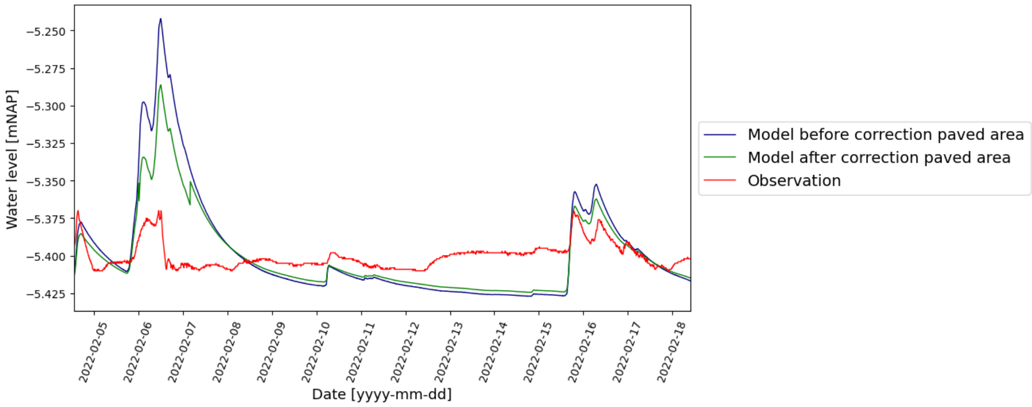


Figure 28 - Water level at a measurement point in the urban area. The overestimation of the water levels is (partly) due to the overestimation of paved area in the urban area.

Another adjustment that has been made, are the operating levels of the two pumping stations. When analysing the water level observations, it can be seen that the pumping stations aim for a slightly lower water level than the target water level that was mentioned in the raw data (Dutch: legger data). Therefore, some adjustments have been made to ensure a better match. However, after these adjustments, there was still a small error in the operation of the pumping stations. Since no Real Time Control component (D-RTC) is included in the D-HYDRO model, it is not possible to include different pumping capacities based on the current water level. In reality, the pumping capacity does depend on the water level: the observations show that the pumping stations operate with different capacities depending on the water level. To simulate this effect, two additional pumping stations are added at the same location as the original ones. When the water level is slightly higher than the target level, only one of the two pumping stations is pumping. Only when the water level increases significantly, both pumping stations are in operation. Even after this adjustment, the modelled water levels still do not fully match the observations. In the observations, it can be noticed that the waterboard decided to lower the water levels a couple of days before the rainfall events by turning on the pumping station This can for example be recognized by the drop in the water level around the 18th of February in Figure 29. These decisions can be seen in the observational data, but are not included in the hydrological model as it is a deviation from the standard operation procedure. The hydrological model thus overestimates the water levels due to this.

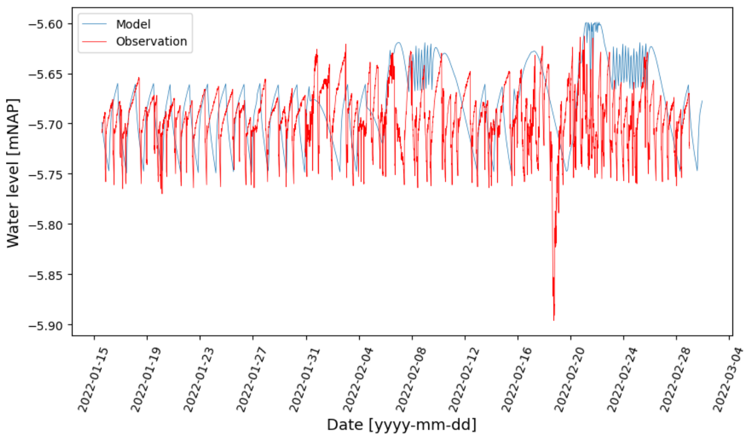


Figure 29 - Water level at one of the observations points after calibrating the hydrological model

Table 17 - Results calibration of the hydrological model

Parameter	Description	Minimum value	Maximum value	Final value	Unit
1D channel roughness	Bed roughness in 1D waterways (Strickler)	10	30	12	$m^{1/3}s^{-1}$
Channel width	Width of the channels where no cross section was available in the provided dataset	0.5	2	1.5	m
Drainage resistance	The ability to resist water flow underground for each soil layer	0.5	2	1.8	mm
Drainage layer depth	The depth of the soil layers with respect to surface level	0	2	0.2	m

Table 18 shows the values for the MAE and RMSE for the three events after the calibration. The error is still relatively high after the calibration. However, since an extensive calibration does not fit within the scope of this thesis, it is decided to accept this error for the purpose of this thesis. For the creation of the surrogate model, the calibrated hydrological inundation model is set as the benchmark and is considered as the truth. It is thus important to realize that the benchmark itself already has a relatively large error with respect to observations. The error of the surrogate model is thus additional to this error.

Table 18 - Performance indicators of the calibrated hydrological inundation model for the calibration and validation events

	Date	MAE	RMSE
Calibration event #1	February 20th 2022	0.042m	0.052m
Calibration event #2	June 24th 2022	0.071m	0.087m
Validation event	November 17th 2022	0.053m	0.064m

3.5.2.2 Creation surrogate models

Table 20 shows the combinations of simplifications that lead to the different surrogate models. First, the number of calculation points is decreased by decreasing the spatial resolution of both the 1D and 2D model. This results in a decrease in computational time, but also an increase in the error as can be seen in Figure 30. By removing the secondary branches and replacing them by retention areas, the computational time also decreases but the error increases significantly.

In a third attempt, the maximum Courant number is increased. When comparing the results of benchmark model with the results of surrogate model #12, it can be seen in Figure 30 that increasing the maximum Courant number significantly decreases the computational time without losing accuracy. However, when increasing the maximum Courant number even further, the computational time does not decrease anymore, see results of surrogate model #14. Increasing the maximum Courant number is thus a very efficient simplification up to a certain extent. To lower the computational time even further, it is needed to decrease the number of calculation points.

After testing multiple combinations, it turns out that surrogate model #10 has the shortest computational time with 0.6 minutes (= 36 seconds). As explained in Section 2.3.3, the end-user would like to have a maximum computational time of 30 minutes for all probabilistic rainfall forecasts. The KNMI Harmonie MOS rainfall forecast has 50 ensemble members, meaning that the computational time of one model run should be equal to or smaller than 0.6 minutes. This is the case for surrogate model #10. As also explained in Section 2.3.3, the end-user would like to have a MAE of 0.1m as maximum in the waterways. Since the benchmark model already has a MAE of 0.053m on the validation dataset, the surrogate model can have a MAE of 0.047m in a worst case scenario (when the errors are in the same direction: i.e. both overestimation or underestimation). Table 19 shows an overview of the accuracy of surrogate model #10 with respect to the benchmark model. As can be seen, the accuracy of the surrogate model meets the requirements of the end-user since the MAE of the water level in the waterways is smaller than 0.047m.

Table 19 - Accuracy surrogate model #10 with respect to the benchmark model

Performance indicator	Accuracy Surrogate model #10
MAE timeseries water level in waterways [m]	0.012
RMSE maximum water level in waterways [m]	0.017
MAE timeseries water depth on 2D grid (all nodes) [m]	0.003
MAE timeseries water depth on 2D grid (no True Negatives) [m]	0.022
RMSE maximum water depth on 2D grid (all nodes) [m]	0.005
RMSE maximum water depth on 2D grid (no True Negatives) [m]	0.033
CSI maximum water depth on 2D grid [%]	58%

In Figure 33, the output of the surrogate model can be seen. Besides the water depth on the 2D grid, also the water levels in the waterways are part of the output of the surrogate model. Figure 34 shows the difference between maximum water depth according to the benchmark model and the surrogate model. As can be seen, the surrogate model tends to underestimate the water depth on the 2D grid for most grid cells with a couple of centimetres. The opposite holds for the water levels in the waterways: here the surrogate model overestimates the water level as can be seen in Figure 35. The difference in under- and overestimation is probably caused by a combination of the difference in spatial resolutions and the methodology used to calculate the differences in water depth and water level (see Section 3.5.1). Figure 36 shows the water level at one of the measurement locations over time. This figure confirms, together with the graphs from the other measurement locations, that the timing and shape of the peak are predicted correctly.

Table 20 - Overview characteristics surrogate models. Yellow coloured cells indicate changes with respect to the benchmark model.

RUN	Benchmark	Sug #1	Sug #2	Sug #3	Sug #4	Sug #5	Sug #6	Sug #7	Sug #8	Sug #9	Sug #10	Sug #11	Sug #12	Sug #13	Sug #14
Distance between 1D nodes [m]	20	50	20	50	50	100	100	200	50	50	50	50	20	20	20
Spatial resolution squared 2D grid [m]	20	20	50	50	50	100	100	200	50	50	50	50	20	50	20
Secondary branches?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Courant constant	0.7	0.7	0.7	0.7	1	1	1	1	0.7	5	10	10	10	10	30
Timestep (user) [min]	10	10	10	10	10	10	20	30	10	20	20	10	10	10	10
Computation time [min]	10.62	8.28	5.34	3.54	2.64	1.78	1.62	1.26	2.22	0.66	0.60	0.71	2.04	1.05	2.07
MAE water level 1D timeseries [m]	0	0.004	0.012	0.012	0.012	0.016	0.015	0.026	0.035	0.012	0.012	0.012	0.0002	0.012	0.0002

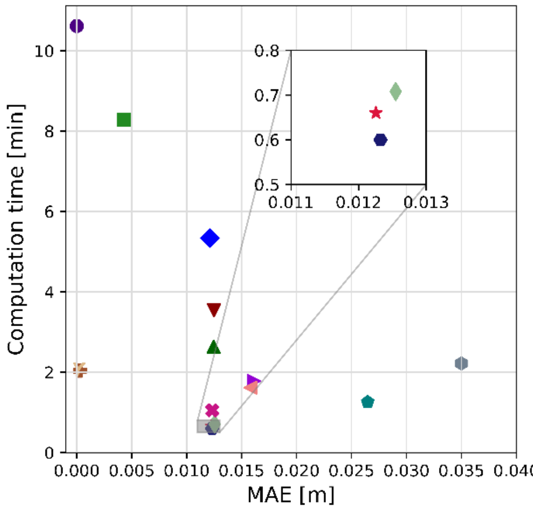


Figure 30 - Mean Absolute Error (MAE) of water level timeseries in 1D calculation points with respect to the benchmark model.

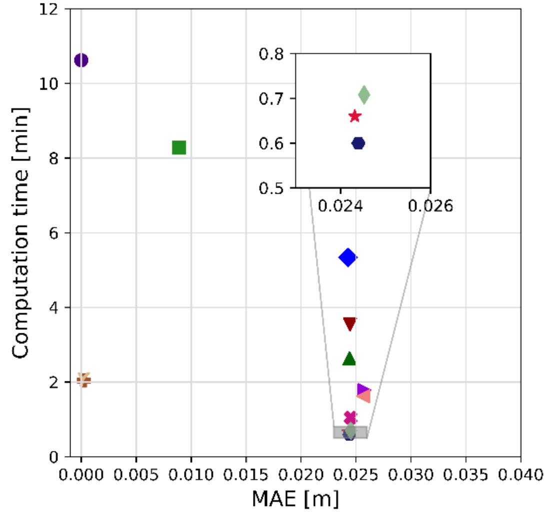


Figure 31 - Mean Absolute Error (MAE) of maximum water level over time in 1D calculation points with respect to the benchmark model.

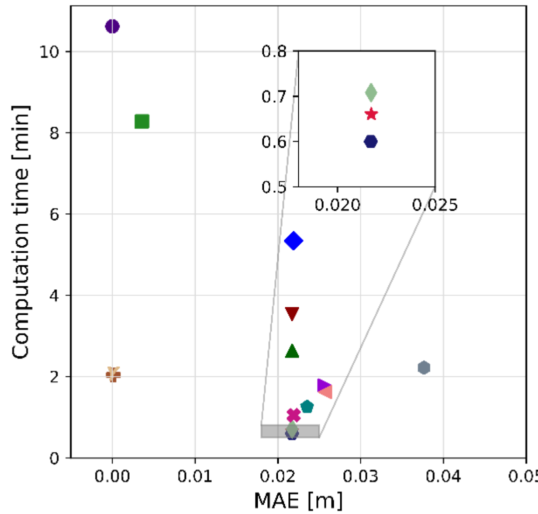
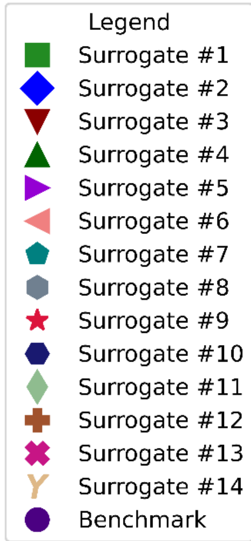


Figure 32 - Mean Absolute Error (MAE) of water depth timeseries on 2D grid cells with respect to the benchmark model. True Negative cells are not considered.



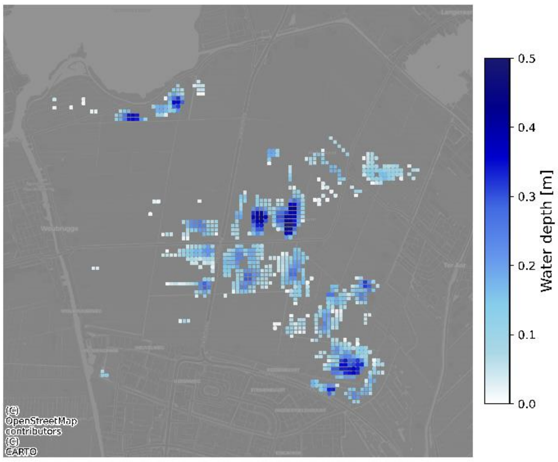


Figure 33 - Output surrogate model #10: maximum water depth over time on 2D grid. Figure 27 shows the rainfall timeseries used as input.

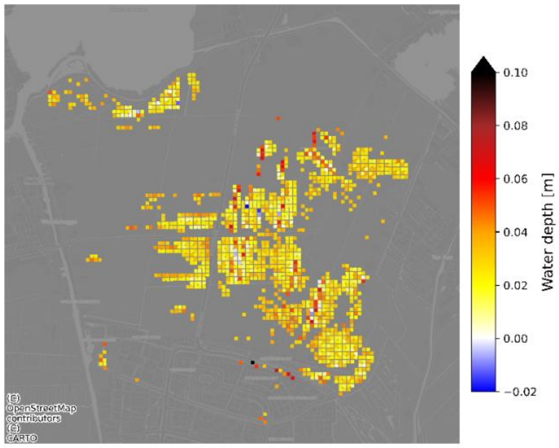


Figure 34 - Difference plot on maximum water depth over time on 2D grid. Difference = maximum water depth benchmark model - maximum water depth surrogate model

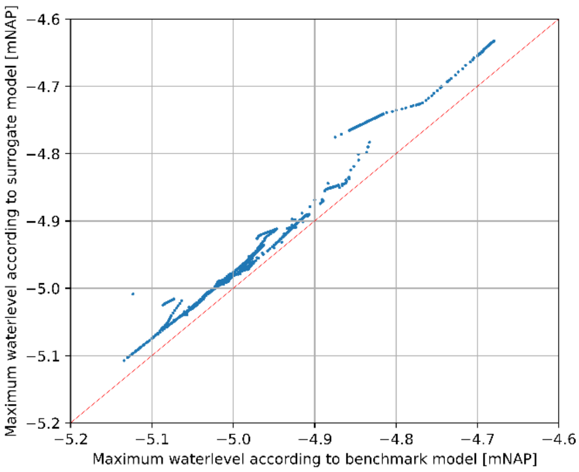


Figure 35 - Scatter plot for maximum water levels in 1D waterways surrogate model with respect to the benchmark hydrological model.

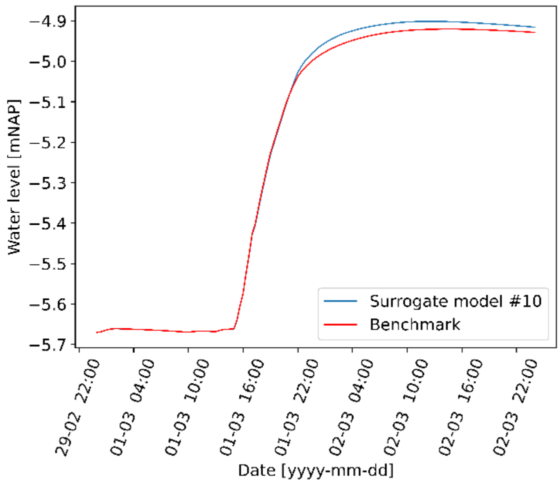


Figure 36 – Water level at one of the observation locations for the surrogate model and benchmark model.

3.5.2.3 Practical implementation

As explained in Section 3.2, a probabilistic rainfall forecast will be used as input for the surrogate model when the model is used in an operational setting. The rainfall event on the 14th of July 2021, which induced flooding in large parts of the Netherlands, serves as example. The KNMI Harmonie MOS rainfall forecast (KNMI, 2023) available at 12:00 that day for Polder Vierambacht is shown in Figure 37. This event serves as an example to show the implications of using a probabilistic rainfall forecast as input for the surrogate model.

Figure 38, Figure 39, and Figure 40, the output of the surrogate model can be seen when using the probabilistic rainfall forecast as input. For all output variables, also a csv-file is available that contains the exact numbers. Especially for the water levels in the waterways, this is an easier method to communicate to the end-user compared to plotting them as in Figure 40, as reading the exact water level increase from this figure can be challenging. For reference: computing this calculation for all 50 ensemble members took 27 minutes and 55 seconds in total.

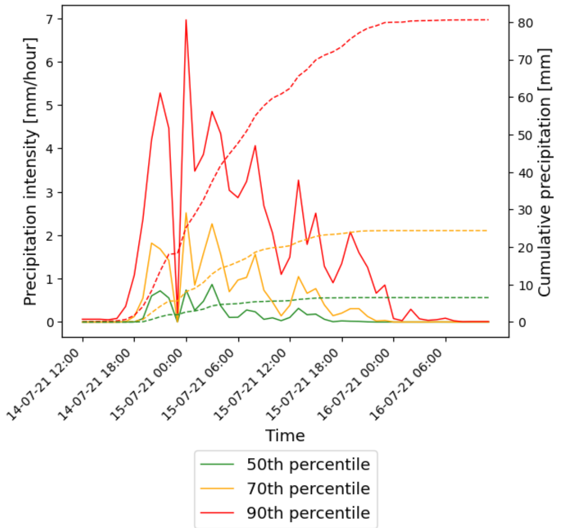


Figure 37 – Rainfall forecast KNMI Harmonie MOS (probabilistic forecast) that was available for Polder Vierambacht on the 14th of July 2021 at 12:00. Dashed lines indicate the cumulative rainfall over the 48 hours.

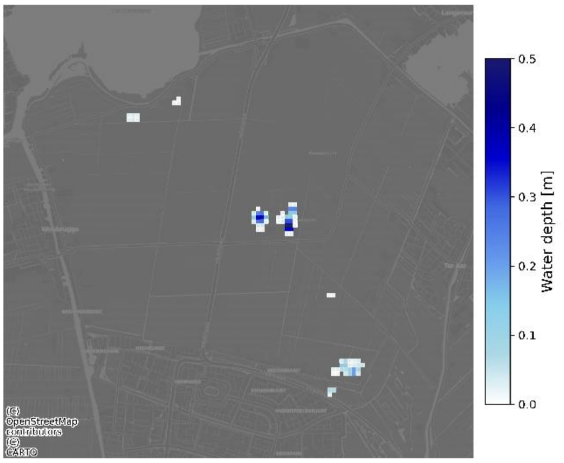


Figure 38 - Output of the surrogate model: maximum water depth over time on the 2D grid for the 90th percentile of ensemble members.

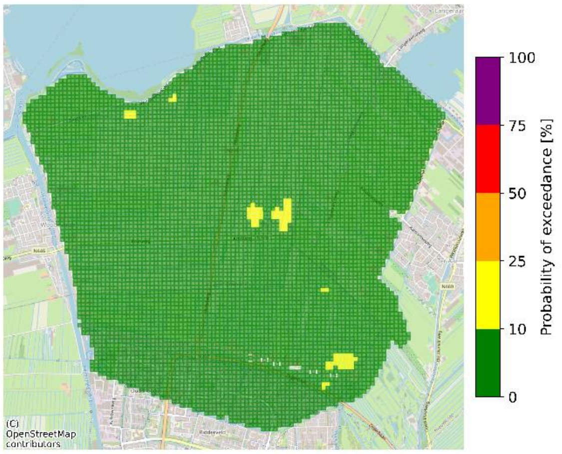


Figure 39 - Output of the surrogate model: probability that the water depth exceeds a threshold of 0 m based on the probabilistic rainfall forecast. Probability is defined as the percentage of ensemble members that forecast a value larger than the threshold.

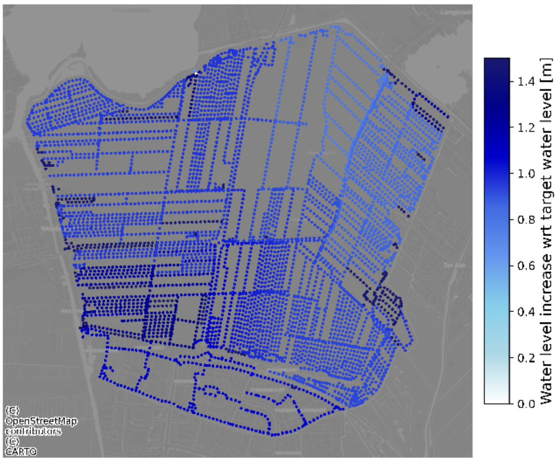


Figure 40 - Output of the surrogate model: maximum water level increase (with respect to the target water level) on the 1D calculation points (1D waterways) for the 90th percentile of ensemble members.

3.5.3 Discussion

Model performance

The surrogate model created in this study has a MAE of around 0.02 m for the 2D grid, and a MAE of 0.01 m for the waterways in the 1D grid. To calculate the deviation of the surrogate model with respect to the benchmark model for the 2D grid, only flood depths larger than 0.1m are considered as including lower flood depths results in an unrealistically high percentage. The surrogate model deviates on average 24% from the benchmark model for the 2D grid. For the maximum water levels in the 1D waterways, the surrogate model deviates on average only 0.8% from the benchmark model. This is in line with the performance of surrogate models found in literature: the lower fidelity surrogate model created by Bomers et al. (2019) deviates on average 1% from its benchmark hydrological inundation model.

Lack of complete and coherent data

Data about the exact operation of the weirs is not available and the operation during heavy rainfall events is often done manually by local water managers. In the model, the target water levels for the associated fixed drainage level area are used as operating levels for the weirs. In reality, the operation of the weirs and pumping stations is based on the actual measured water levels. The calibration shows that these types of errors make it difficult to obtain good results for the calibration and validation. The calibrated and validated hydrological inundation model that serves as a benchmark has a relatively large MAE of 0.05m. The error of the surrogate model with respect to the benchmark model is relatively small (around 0.01m) and therefore much accuracy can be gained by improving the accuracy of the benchmark model. Due to a lack of data (e.g. discharge only measured at one of the two pumping stations), it is difficult to identify the cause of the relatively large error after calibration. There is also no data available on the inundation on surface level, making it impossible to calibrate the model on the inundation depth. Instead, the hydrological inundation model is only validated on the water level in the waterways. An extensive calibration procedure on the benchmark model is not within the scope of this thesis, but could be done if the waterboard is interested in using this model in an operational setting. Besides, a D-RTC component could be implemented in the hydrological inundation model to be able to better simulate the operation of the pumping stations and weirs in the study area.

Maximum computational time

Even though the surrogate model meets the requirements regarding the computational time (< 30 minutes), there is not much room for a finer grid or covering a larger study area. Surrogate model #10 is of all 15 combinations the only model that meets the requirements regarding the maximum computational time. If the waterboard is interested in covering a larger study area, multiple devices will be needed to do the computations. For further research, it is advised to look into the possibilities of parallel computing as this increases the computational power. Figure 31 shows that the step from a 20 m x 20 m grid to a 50 m x 50 m grid results in a significant increase in the MAE. When using parallel computing techniques, it could be possible that surrogate model #12 meets the requirement of the computational time. In that case, one does not have to compromise on the accuracy.

The computational times as stated in Table 20 are based on an extreme rainfall event where numerous grid cells are flooded. A quick analysis has shown that the runtime is significantly shorter (up to 35%) when no grid cells are flooded compared to the ensemble member with the highest number of grid cells flooded (14% of grid cells flooded). During normal operation, without high rainfall intensities forecasted, the runtime will thus be shorter than 30 minutes.

3.6 Conclusion

Using the requirements that followed from the first research question (Section 2.5), the quality of the surrogate models for each case study are assessed. The sub-sections below summarize the requirements of each case study and qualitatively describe to what extent each surrogate model is able to meet the requirement.

3.6.1 Case study 1 – Municipality of Amersfoort

According to the municipality of Amersfoort, the main goal of a hydrological inundation model in an operational setting is to give more insights and information beforehand on the locations where inundation might occur. The LSTM model used in this case study has shown that it is able to accurately predict both the location of the flooding and the flood volume. A CSI score of 81% is sufficient for operational flood management (Erechtchoukova et al., 2016; Zanchetta & Coulibaly, 2020). It can be concluded that the LSTM model is able to accurately and rapidly (order of milliseconds) reproduce the hydrological inundation model. The output meets the requirements of the municipality: it shows the location of the flooding on a 2D map. For further research, it is advised to reconsider the flood volume as output variable. Instead, inundation depth on surface level would be a better fit, especially when communicating the model output to non-experts.

An overview of the quality assessment can be found in Table 21. It can be concluded that the model in general meets the requirements from the end-user. However, a post-processing method is advised to make the model output more understandable for non-experts.

Table 21 - Overview quality assessment of surrogate model for case study #1

	Requirement	**Surrogate model case study #1**
Modelling goal	Identify locations where inundation is expected including predicted flood depth	Locations are predicted correctly, but the output variable does not describe the flood depth.
Perspective for action	Alert contractors & citizens	Partly possible, output variable is difficult to interpret by non-experts.
Output variable	Flood depth on 2D map: timeseries and maximum	Flood volume per manhole [m^3]. With probabilistic forecast: certain percentile [m^3] or probability of exceedance [%]. Timeseries are known but not visualized.
Required lead time	None	Dependent on rainfall forecast, up to 24 hours.
Grid size	Dependent on rainfall forecast, max 1 km x 1 km	Per manhole
Accuracy	Low false-alarm ratio (high critical success index)	CSI = 81.52%
Maximum computational time	Order of minutes	Run time < 1 minute

3.6.2 Case study 2 – Municipality of Tilburg

The surrogate model created for the second case study has a high potential in meeting the requirements by the end-user. As described in Section 3.4.2, the surrogate model is overfitted on the training dataset and is therefore not able to perform well on unseen data. This is confirmed by the CSI score of only 48%, which is lower than 80% and thus insufficient (Erechtchoukova et al., 2016; Zanchetta & Coulibaly, 2020). An overview of the requirements and the performance of the surrogate model is given in Table 22. Despite the inaccuracy, the surrogate model performs well on the other criteria from the end-user. The computational time of less than a minute, the grid size, and the output variable all meet the requirements. For future research, it is advised to enlarge the training dataset or reduce the number of output variables such that the overfitting can be reduced. It is advised to also generate timeseries for the output variable (water depth) as this is the municipality's requirement. Cloud computing methods could be used to rapidly generate the required training data (Hop, 2023). This will lead to a surrogate model that performs well on all requirements by the end-user.

Table 22 – Overview quality assessment of surrogate model for case study #2

	Requirement	Surrogate model case study #2
Modelling goal	Identify locations where inundation is expected including predicted flood depth	Locations can be identified on a 5 m x 5 m scale
Perspective for action	Alert on-call service, managers of pumping stations & MT municipality	Assuming the prediction is correct, then the model output provide sufficient information for the perspectives for action.
Output variable	Flood depth on 2D map: timeseries and maximum	Flood depth [m], but only the maximum. Timeseries are unknown. With probabilistic forecast: certain percentile [m] or probability of exceedance [%].
Required lead time	None	Dependent on rainfall forecast, up to 24 hours
Grid size	Dependent on rainfall forecast, the size of a neighbourhood at maximum	Rectangular cells of 5 m x 5m
Accuracy	Low false-alarm ratio (high critical success index)	CSI = 48%
Maximum computational time	Order of minutes	Run time < 1 minute

3.6.3 Case study 3 – Hoogheemraadschap van Rijnland

The surrogate model created for the third case study fully meets the requirements as specified by the end-user. An overview of the requirements and the performance of the surrogate model is given in Table 23. The surrogate model is able to accurately reproduce the detailed hydrological inundation model that serves as a benchmark. The error of the surrogate model with respect to the benchmark is small (around 0.01 m for water level in waterways). However, the benchmark model has a relatively large error with respect to the observations (0.05 m). This makes the total error significantly larger, but still within the requirements of the end-user. For further research, it is advised to carry out a more extensive calibration for the benchmark model such that the total error is decreased.

The surrogate model is at its limit regarding the maximum computational time (<30 minutes). It is advised to look into the possibilities of parallel computing, as this increases the computational power, allowing for a smaller error by using a finer grid. In conclusion, the surrogate model meets all requirements. However, the requirement of the maximum computational time is just met, resulting in no room for additional requirements or increasing the model domain with the current set-up of the surrogate model.

Table 23 - Overview quality assessment of surrogate model for case study #3

	Requirement	Surrogate model case study #3
Modelling goal	Predict water levels in water ways & visualize expected inundation depth on 2D map	Water levels can be predicted at each calculation point & inundation depth is plotted on 2D map
Perspective for action	Pump water to the boezem system, install temporal pumping stations & alert relevant authorities	Possible, the model results provide sufficient information for the perspectives for action.
Output variable	Water level in water ways & flood depth on 2D map	Both flood depth [m] as well as water level [m] timeseries are known. With probabilistic forecast: certain percentile [m] or probability of exceedance [%].
Required lead time	48 to 24 hours	Depends on rainfall forecast, up to 48 hours
Grid size	Range of 5 – 100 meters	50 meters for both 1D and 2D grid
Accuracy	Maximum absolute error of 10 cm in water ways	MAE of 0.026 m for surrogate with respect to the benchmark model.
Maximum computational time	30 minutes	Run time around 30 minutes for 50 ensemble members.

4. Quality assessment W2O model

4.1 Introduction

The previous chapter has led to a surrogate model for each of the case studies that meets requirements of the end-user to a certain extent. As explained in Section 1.2, the W2O model is an already available hydrological model created by HydroLogic. To see if the surrogate models have an additional value compared to the W2O model, the quality of the W2O model is also assessed based on the requirements following from the first research question. Section 4.2 describes the used methodology for this. In Section 4.3, the results of the quality assessment are described. Finally, the results are discussed (Section 4.4) and conclusions are drawn (Section 4.5).

4.2 Methodology

Waarschuwing voor Wateroverlast (W2O, translated: warning for flooding) is a conceptual hydrological model that predicts the inundation depth for each grid cell (Schnitzler, 2022). The algorithm behind the W2O model is continuously updated and improved by colleagues from HydroLogic. For this thesis, the most recent version (June 2023) is used to assess the quality of the W2O model. However, since the model is still work in progress, more research will be necessary to fully assess the quality of the W2O model.

4.2.1 Description and characteristics of the W2O model

The W2O model covers the entire area of the Netherlands and uses a 1 km x 1 km grid size. The concept of the W2O model is a bucket model where each grid cell is represented by three different buckets with certain characteristics. These buckets include: inundation on surface level, open waters, and storage in the soil and sewage system. Interaction between these buckets takes place in all directions. An overview of the model schematisation is presented in Figure 41. The currently available soil storage is obtained from OWASIS (2023), a service by HydroLogic. OWASIS uses rainfall data (radar) in combination with evaporation data and the national hydrological model (Dutch: Landelijk Hydrologisch Model) to calculate the current available soil storage on a 250 m x 250 m resolution for the entire area of the Netherlands.

The infiltration of water into the soil is based on the soil type (BRO, 2020). The W2O model combines this information with the topography and a probabilistic rainfall forecast to calculate the inundation depth for each grid cell. As explained in Section 3.2, the ensemble members are visualized by either showing a certain percentile of ensemble members or the percentage of ensemble members that predict an inundation depth of more than a certain threshold value. To show the output on a 2D map in a dashboard, the maximum and average over the considered time period are calculated. When clicking on a certain grid cell, the timeseries will be shown in a graph for that grid cell.

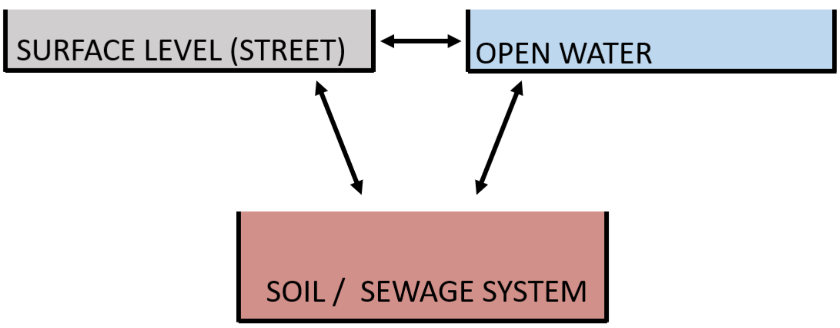


Figure 41 - Simplified model schematisation of the W2O model

Since the W2O model is a physically based conceptual model (and not a numerical model), it does not solve the numerical flow equations (e.g. Navier-Stokes equations). The principles and equations behind the algorithm are thus more simple compared to a numerical model (e.g. a D-HYDRO model). This results in the computation time being significantly shorter. The W2O model is able to do 50 computations within ± one minute, making it very suitable for probabilistic forecasting.

As the W2O model covers the entire area of the Netherlands, it is possible to get a general overview of the predicted inundation. The rainfall forecast that is used as input for the W2O model, is in gridded form and spatially distributed over the Netherlands, see for example Figure 42. As a result, also the rainfall forecast and predicted inundation near areas of interest are available. Knowing the predicted inundation for a larger area than just the area of interest can be useful information for the end-user. If it turns out that the rainfall forecast is spatially incorrect, the end-user would still be aware that peak rainfall intensities (and possibly inundation) are predicted for neighbouring areas. With this information, the end-user can also be on stand-by when inundation is forecasted in a neighbouring area, and act faster as soon as necessary. This is different compared to the surrogate models created in Section 3, where only the nearest rainfall forecast is used as input for the surrogate model and rainfall forecasts and inundation depths for neighbouring areas are not shown.

Regarding the input data of the W2O model, the algorithm is mainly based on maps containing soil types (BRO, 2020), elevation (version AHN4, (Actueel Hoogtebestand Nederland, 2020)), and available soil storage (OWASIS, 2023). Furthermore, there are a couple of parameters that can be calibrated (e.g. relationship between slope and runoff, available storage in open waters per square meter of open water in that grid cell). In the current version of the W2O model, these parameters are spatially uniform over the entire model domain. Due to the type of input data used in the W2O model, changes in the input variables and model parameters are relatively easy to make.

The W2O model uses a probabilistic rainfall forecast as input. Different types of probabilistic rainfall forecasts are available (Schnitzler, 2022). One of them is the KNMI Harmonie MOS (KNMI, 2023), which is also used for the surrogate models in Section 3. For the purpose of this thesis, this probabilistic rainfall forecast is also used as input for the W2O model.

4.2.2 Model validation

Due to limited available measurements of inundation depth on surface level, it is difficult to calibrate and validate the W2O model quantitatively. A qualitive validation can done by comparing the model results to soft data (e.g. news articles). The uncertainty and error in the rainfall forecast bring an additional layer of uncertainty to the output of the W2O model when comparing it to observations. Therefore, for validation purposes, an event is chosen where the rainfall forecast was (almost) in line with the observations. This is the case for the event on the 9th of September 2020: the 90th percentile of the rainfall forecast predicts rainfall intensities close to the observed intensities for most parts of the Netherlands (Schnitzler, 2022). On this day, the south-western part of the Netherlands experienced heavy rainfall that induced flooded streets, houses, and shops. The rainfall event has led to flooding in Middelburg, Terneuzen and Axel (PZC, 2022).

Figure 42 shows the available rainfall forecast for the 9th of September 2020 at 06:00. The figure shows the cumulative rainfall over the lead time of 48 hours for the 90th percentile of ensemble members. The rainfall forecast for that day is close to the observed rainfall (see Appendix G for the observed rainfall). For most parts of the Netherlands, the predicted cumulative rainfall (90th percentile) was of similar size as the observed cumulative rainfall. The middle part of the Netherlands (Utrecht, the east of Zuid-Holland, and the North of Noord-Brabant) is an exception: here the 90th percentile of the rainfall forecast overestimates the observed rainfall. As can be seen in Appendix G, the cumulative observed rainfall equals around 10 to 15 mm for this area (instead of 45 – 75 mm as forecasted). As for most parts of the Netherlands, the 90th percentile of ensemble members of the rainfall forecast is in line with the observations, this event is used to show the accuracy of the W2O model.

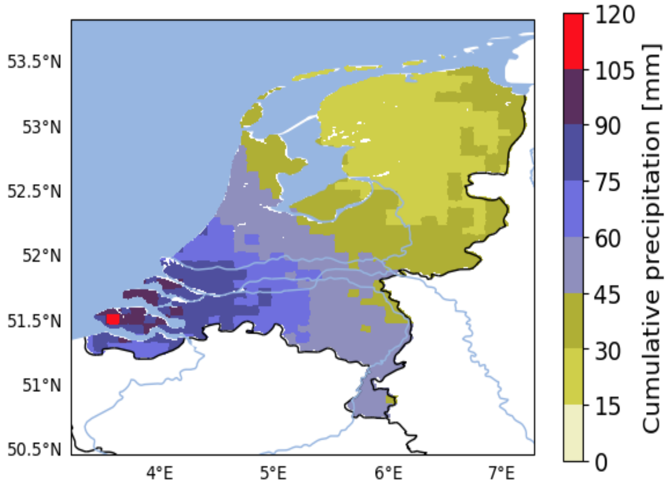


Figure 42 - Rainfall forecast KNMI Harmonie MOS available on 09-09-2022 at 06:00. Figure shows the cumulative rainfall over the 48 hours for the 90th percentile of ensemble members

4.2.3 Quality assessment

To assess the quality of the W2O model using the requirements as defined by the end-user in Section 2.5, a different rainfall event is chosen. For this purpose, the rainfall event on the 14th of July 2021 is chosen as the available rainfall forecast predicts heavy rainfall for the entire area of the Netherlands, including the study areas for the selected case studies. The rainfall forecast available on the 14th of July 2021 at 12:00 is shown in Figure 43. This figure shows the cumulative rainfall over 48 hours for the 90th percentile of ensemble members. The rainfall forecast available for the specific case studies can be found in Section 3.3.2.1, Section 3.4.2.3, and Section 3.5.2.3 for case study 1, 2 and 3 respectively.

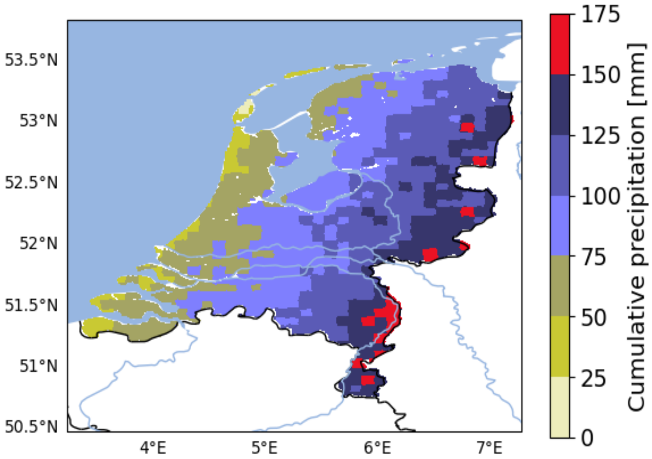


Figure 43 - Rainfall forecast KNMI Harmonie MOS available on 14-07-2021 at 12:00. Figure shows the cumulative rainfall over the 48 hours for the 90th percentile of ensemble members. Note the difference in scaling of the colour bar when comparing this figure to Figure 42.

This rainfall forecast is used as input for the W2O model. The output is given for the entire area of the Netherlands, but figures are also created where we zoom in on the area of the specific case studies. The quality of the W2O model is then assessed using the requirements that followed from the first research question. Similarly to the methodology for the surrogate models as described in Section 3.2, the quality of the W2O model is qualitatively compared. A description including a colour (see Table 3) that describes to what extent the requirement is met, is given for each requirement.

4.3 Results

4.3.1 Model validation

The probabilistic rainfall forecast KNMI Harmonie MOS (KNMI, 2023) that was available on the 9th of September 2022 at 06:00 (see Figure 42) is used as input for the W2O model. In Figure 44 and Figure 45, two of the output figures of the W2O model are shown. These figures represent the percentage of ensemble members that predict an inundation depth of more than 0 mm (Figure 44) or 10 mm (Figure 45). These figures show the maximum probability over time, but timeseries are also available for each grid cell.

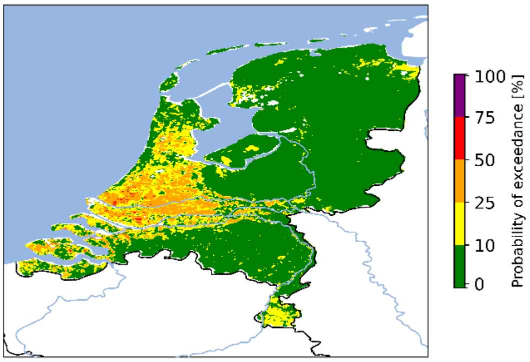


Figure 44 - Output of the W2O model. Probability is defined as the number of ensemble members that predict an inundation depth of more than 0 mm. The maximum probability over time is shown for each grid cell.

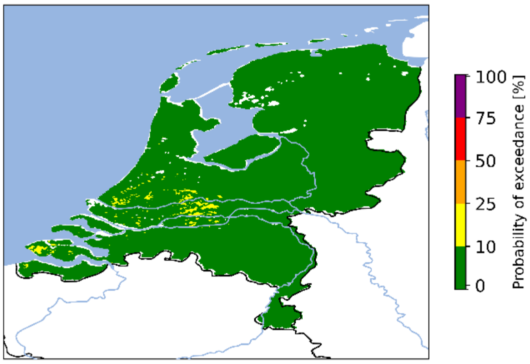


Figure 45 - Output of the W2O model. Probability is defined as the number of ensemble members that predict an inundation depth of more than 10 mm. The maximum probability over time is shown for each grid cell.

As can be seen in Figure 44, between 10% and 50% of the ensemble members predict at maximum more than 0 mm inundation on surface level for large parts of Zeeland, Zuid-Holland, Noord-Holland, and the south of Limburg. More than 10 mm inundation on surface level (Figure 45) is predicted by only a couple of ensemble members (10-25%) for a small part of Zeeland (near Middelburg) and the middle of the country. Figure 46 shows the inundation on surface level for the 90th percentile of ensemble members where the maximum over time is taken for each grid cell. This figure shows that the inundation around Middelburg is accurately predicted. An overestimation of the inundation depth can be seen in the middle of the Netherlands. This could be caused by the rainfall forecast that overestimated the rainfall intensities in this area (see Section 4.2.2).

Figure 46 - Output of the W2O model for 09-09-2022 at 06:00. Water on surface level [mm] for the 90th percentile of all ensemble members. The maximum over time is taken for each grid cell.

The W2O algorithm is continuously improved as described in Section 4.2.1. This, in combination with the limited available observations on inundation depths, makes it difficult to quantitatively validate the W2O model. The version of the W2O model presented in this thesis has shown that it is able to quite accurately predict the location of inundation for areas where the rainfall forecast is correct. However, future research is necessary to fully validate the W2O model including the improvements that are currently still work in progress.

4.3.2 Quality assessment

For the quality assessment of the W2O model, the rainfall forecast available for the 14th of July 2021 12:00 is used as input (see Figure 43). As explained in Section 4.2.3, this event is selected as heavy rainfall was forecasted for large parts of the Netherlands. The output of the W2O model is presented in Figure 47 and Figure 48. Figure 47 shows the percentage of ensemble members that predict an inundation depth of more than 0 mm. The maximum probability over time is shown in this figure, but timeseries are available for each grid cell. Figure 48 shows the inundation depth for the 90th percentile of ensemble members, where the maximum over time is shown in the figure.

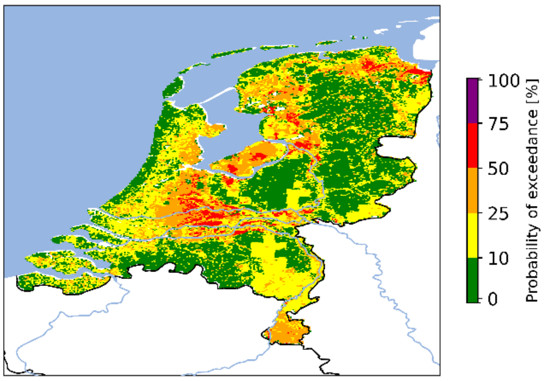


Figure 47 - Output of the W2O model for 14-07-2021 at 12:00. Probability is defined as the number of ensemble members that predict an inundation depth of more than 0 mm. The maximum probability over time is shown for each grid cell.

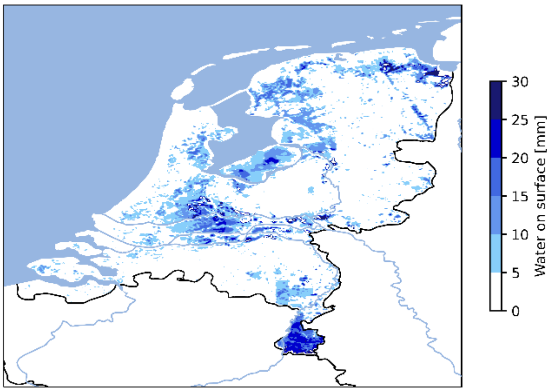


Figure 48 - Output of the W2O model for 14-07-2021 at 12:00. Water on surface level [mm] for the 90th percentile of all ensemble members. The maximum inundation depth over time is taken for each grid cell.

In Figure 49 and Figure 50, the same output of the W2O model is shown but zoomed in on the selected case studies. Based on the rainfall forecast, little to no inundation is predicted by the W2O model for Hooglanderveen (Figure 49a and Figure 50a) and Udenhout (Figure 49b and Figure 50b). For Polder Vierambacht (Figure 49c and Figure 50c), up to 11.4 mm inundation on surface level is forecasted for the 90th percentile of ensemble members. For the southern part of Polder Vierambacht, 10 till 42% of the ensemble members predict more than 0 mm inundation.

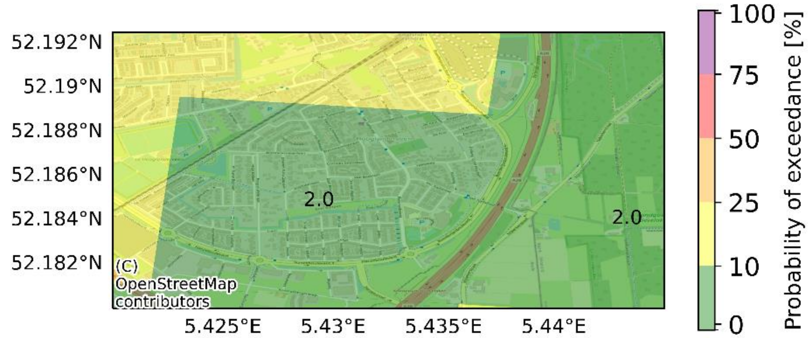


Figure 49a – Hooglanderveen. Probability of > 0 mm inundation on surface level, maximum over time.

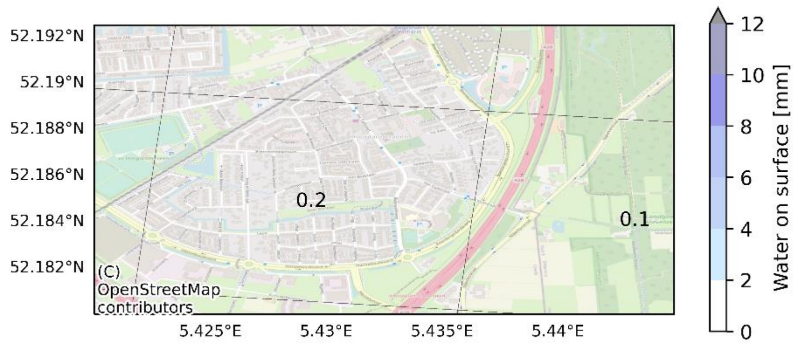


Figure 50a – Hooglanderveen. Maximum over time for the 90th percentile of ensemble members

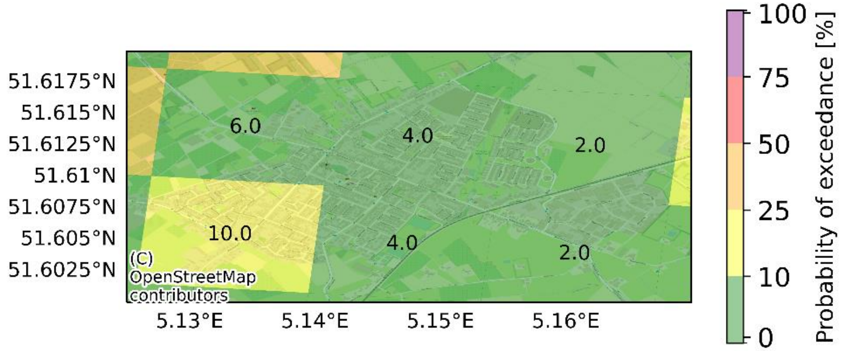


Figure 49b – Udenhout. Probability of > 0 mm inundation on surface level, maximum over time.

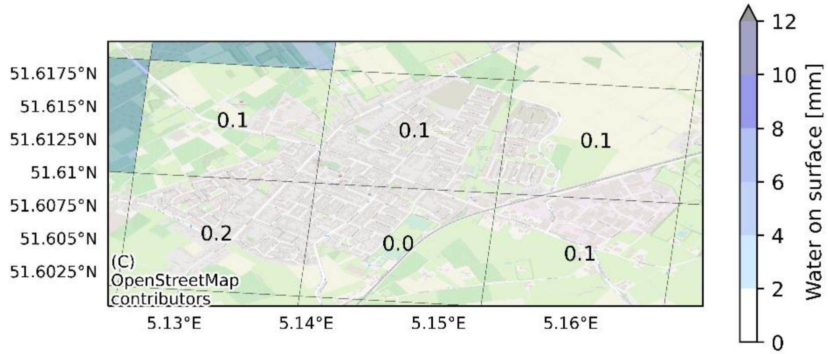


Figure 50b – Udenhout. Maximum over time for the 90th percentile of ensemble members

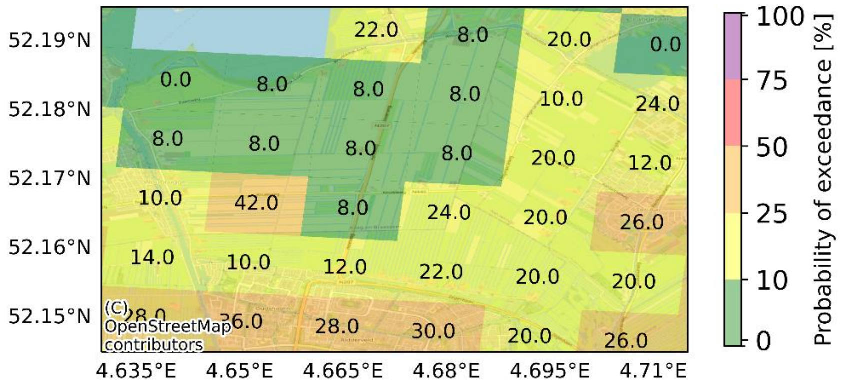


Figure 49c - Polder Vierambacht. Probability > 0 mm inundation on surface level, maximum over time.

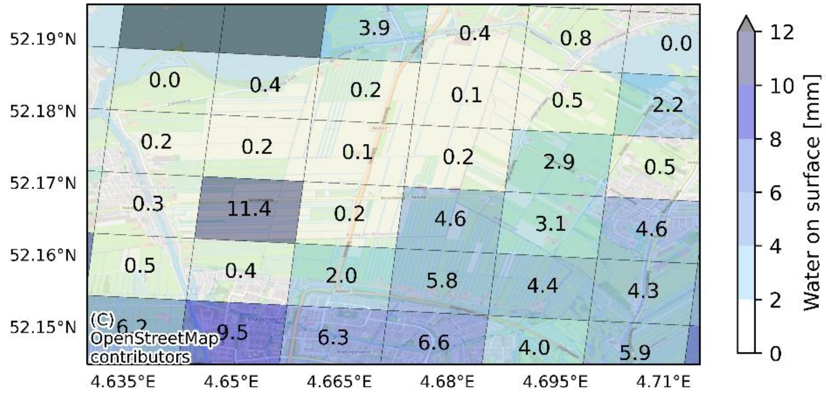


Figure 50c - Polder Vierambacht. Maximum over time for the 90th percentile of ensemble members

4.4 Discussion

4.4.1 Improvements of W2O model

As explained in Section 4.2.1, the W2O model is still work in progress and is continuously being improved. The version used in this thesis already contains significant improvements in the accuracy compared to previous versions (Schnitzler, 2022), but there are also still improvements possible. The latest version of the W2O model contains multiple parameters that can be calibrated. However, since this version was only recently made available, no extensive calibration procedure has been done yet. Besides, due to the lack of available quantitative data, it is difficult to numerically calibrate and validate the W2O model. Instead, only a qualitative comparison with the observations can be done to determine the accuracy. Possible future improvements of the W2O algorithm also include the improvement of the model schematisation and parameters. For example, the infiltration capacity is currently computed using the soil type and land use type only. However, in reality, the infiltration capacity is not constant over time and dependent on multiple other factors (e.g. available soil storage). Including an improved algorithm to calculate the infiltration will likely increase the accuracy of the model results. After implementing these revisions, it is advised to qualitatively calibrate and validate the W2O model to further determine the accuracy of the model output.

4.4.2 Model performance

As explained in Section 4.4.1, it is not possible to quantitatively validate the W2O model due to a lack of available data. This makes it impossible to calculate the CSI and MAE of the W2O model. Therefore, it is also not possible to determine if the W2O model meets the accuracy requirement of the end-user. Besides, as described in Section 2.2.3, the accuracy requirement of the waterboard is a maximum MAE of 0.1m for the water levels in the waterways. Since waterways are not directly included in the W2O model (only storage in open waters is included), water levels cannot be calculated using the W2O model. This makes it impossible to determine if the W2O model meets the accuracy requirement of the waterboard. Instead, only a qualitative description can be given for the accuracy of the inundation depth on surface level. The version of the W2O model used in this thesis has shown that is able to accurately predict the location of inundation when the rainfall forecast is accurate. Further research is necessary to determine if the improved version of the W2O model (see Section 4.4.1) meets the accuracy requirement of the end-user.

4.4.3 Comparison to surrogate models

Due to the different types of models (as explained in Section 1.2) and differences in model output variables, it is difficult to compare the model results of the W2O model with the model results of the surrogate models. Some differences between the W2O model output (Figure 49 and Figure 50) and the surrogate models' output (Section 3.3.2.1, Section 3.4.2.3 and Section 3.5.2.3) can be seen in both the location and size of the inundation. However, due to the large differences in spatial resolution, output variables, and also the nature of the models, it is not possible to determine what the exact differences are.

4.5 Conclusion

4.5.1 Case study 1 – Municipality of Amersfoort

Table 24 shows the quality assessment of the W2O model with respect to the requirements for the first case study. As can be seen in the table, the W2O model meets all requirements from the end-user except the accuracy. As explained in Section 4.4, further research is necessary to fully validate the accuracy of the W2O model. As the W2O model is still work in progress and the algorithm is continuously being improved, it is advised to validate future (improved) versions of the W2O model. For the municipality of Amersfoort, the W2O model provides sufficient information for their perspectives for action: the output variables, lead time and grid size all meet the requirements. Also the computational time of ± 1 minute is sufficient. Besides, the W2O model is already available for the entire study area, meaning that no time has to be invested to generalize the model for a larger model domain (which is the case for the surrogate model).

Table 24 - Overview quality assessment W2O model for case study #1

	Requirement	Surrogate model case study #1	W2O model
Modelling goal	Identify locations where inundation is expected including predicted flood depth	Locations are predicted correctly, but the output variable does not describe the flood depth.	Locations can be identified on 1 km x 1 km scale
Perspective for action	Alert contractors & citizens	Partly possible, output variable is difficult to interpret by non-experts.	Spatial resolution of 1 km x 1 km is sufficient to alert citizens and contractors.
Output variable	Flood depth on 2D map: timeseries and maximum	Flood volume per manhole [m^3]. With probabilistic forecast: certain percentile [m^3] or probability of exceedance [%]. Timeseries are known but not visualized.	Flood depth is indirectly shown on 2D map: a probability of more than *x* mm water on surface level within the next 48 hours is shown. Timeseries are visualized when clicking on certain grid cell.
Required lead time	None	Dependent on rainfall forecast, up to 24 hours.	Up to 48 hours
Grid size	Dependent on rainfall forecast, max 1x1km	Per manhole	1 km x 1 km
Accuracy	Low false-alarm ratio (high critical success index)	CSI = 81.52%	Further research is necessary to validate the W2O model
Maximum computational time	Order of minutes	Run time < 1 minute	± 1 minute

4.5.2 Case study 2 – Municipality of Tilburg

In Table 25, the results for the quality assessment of the W2O for the second case study can be seen. Similarly to the first case study, the W2O model meets all requirements from the end-user except the accuracy. On all other requirements, the W2O model suffices. It meets the requirements in terms of output variables, lead time, and grid size, and the computational time of ± 1 minute is also sufficient. Besides, the W2O model is already available for the entire area of the municipality of Tilburg (compared to only Udenhout for the surrogate model). Similarly to the first case study, it is advised to validate the improved version of the W2O model.

Table 25 - Overview quality assessment W2O model for case study #2

	Requirement	Surrogate model case study #2	W2O model
Modelling goal	Identify locations where inundation is expected including predicted flood depth	Locations can be identified on a 5 m x 5 m scale	Locations can be identified on 1 km x 1 km scale
Perspective for action	Alert on-call service, managers of pumping stations & MT municipality	Assuming the prediction is correct, then the model results provide sufficient information for the perspectives for action.	Spatial resolution of 1 km x 1 km is sufficient to alert relevant authorities.
Output variable	Flood depth on 2D map: timeseries and maximum	Flood depth [m], but only the maximum. Timeseries are unknown. With probabilistic forecast: certain percentile [m] or probability of exceedance [%].	Flood depth is indirectly shown on 2D map: a probability of more than *x* mm water on surface level within the next 48 hours is shown. Timeseries are visualized when clicking on certain grid cell.
Required lead time	None	Dependent on rainfall forecast, up to 24 hours	Up to 48 hours
Grid size	Dependent on rainfall forecast, max the size of a neighbourhood	Rectangular cells of 5 m x 5 m	1 km x 1 km
Accuracy	Low false-alarm ratio (high critical success index)	CSI = 48%	Further research is necessary to validate the W2O model
Maximum computational time	Order of minutes	Run time < 1 minute	± 1 minute

4.5.3 Case study 3 – Hoogheemraadschap van Rijnland

Table 26 shows the quality assessment of the W2O model with respect to the requirements from the end-user for the third case study. Besides the inundation depth on surface level, in which the W2O model suffices, the end-user is also interested in the water level in the waterways. This output variable is not part of the W2O model and will likely not be an output variable in the future. Since waterways are not directly included in the W2O model (only storage in open waters is included), it is not possible to identify which waterways are overflowing based on the W2O model output. Therefore, the W2O model does not fully provide in the information needs for the perspectives for action of the waterboard. Regarding the inundation depth on a 2D map, the W2O model does meet the requirements of the waterboard. The W2O model also meets the requirements regarding the computational time and lead time. Besides, the W2O model is already available for the entire area of Hoogheemraadschap van Rijnland, meaning that no additional time has to be invested to cover the total study area (which is the case for the surrogate model). Similarly to case study 1 and 2, further research is necessary to determine if the accuracy of the W2O model is sufficient.

Table 26 - Overview quality assessment W2O model for case study #3

	Requirement	Surrogate model case study #3	W2O model
Modelling goal	Predict water levels in water ways & visualize expected inundation depth on 2D map	Water levels can be predicted at each calculation point & inundation depth is plotted on 2D map	Inundation depth is visualized on 2D map, but water levels are not calculated.
Perspective for action	Pump water to the boezem system, install temporal pumping stations & alert relevant authorities	Possible, the model results provide sufficient information for the perspectives for action.	Actions could be taken based on the predicted inundation depth on surface level, but the waterboard currently uses water levels (not calculated by W2O) to decide what measures should be implemented.
Output variable	Water level in water ways & flood depth on 2D map	Both flood depth [m] as well as water level [m] timeseries are known. With probabilistic forecast: certain percentile [m] or probability of exceedance [%].	Flood depth on 2D map is calculated, but no information on water levels in waterways.
Required lead time	48 to 24 hours	Dependent on rainfall forecast, up to 48 hours	Up to 48 hours
Grid size	Range of 5 – 100 meters	50 meters for both 1D and 2D grid	1 km x 1 km. This spatial resolution makes it difficult to identify overflowing waterways.
Accuracy	Maximum absolute error of 10 cm in water ways	MAE of 0.026 for surrogate with respect to benchmark.	Not possible to assess (W2O model does not calculate water level in water ways).
Max computational time	30 minutes	Run time around 30 minutes for 50 ensemble members.	± 1 minute

5. Discussion

This section discusses the findings of the research. First by elaborating on the limitations (Section 5,1), then by discussing the generalisation of the surrogate models (Section 5.2), and finally by reflecting on the practical relevance (Section 5.3) and theoretical contribution (Section 5.4).

5.1 Limitations

5.1.1 Requirements end-user

For the first research question, semi-structured interviews took place to define the requirements of a hydrological inundation model in an operational setting according to the end-user. For this, three different case studies were selected that serve as an example. The remaining part of this thesis is based on the answers of these interviews. It could be that other municipalities or waterboards would have answered the interview questions differently. It is likely that the answers will not be completely different, as most tasks and responsibilities are also described by law. However, minor differences could be present from case to case. If, in the future, similar types of surrogate models are used for other municipalities or waterboards, it is important to consider this.

5.1.2 Rigidity surrogate models and W2O model

Changes in the study area require a change in the model set-up. In case of a physically based hydrological inundation model, this can easily be implemented. However, a machine learning model does not provide this flexibility as the physical processes within the model are not preserved (Razavi et al., 2012). These surrogate models are specifically trained for a study area with certain input and output variables including the correct format. Changes in the study area require that the machine learning model is trained again. Thus, a new dataset will have to be generated that includes the updated model set-up using the hydrological inundation model. When changes to the model's set-up are frequently needed, this can be a time-consuming process. For the third case study, where no machine learning model is present, changes to the model set-up can be implemented relatively easily. However, dependent on the type of change, the computational time of the surrogate model could increase. Due to the set-up of the W2O algorithm, model parameters and input data can relatively easy be adapted as also explained in Section 4.2.1. Regarding the rigidity, the W2O model thus has an advantage compared to the surrogate models created for each case study.

5.1.3 Quality rainfall forecast

In this study, the accuracy of three different surrogate models (and the W2O model) has been tested. When these models will be used in an operational setting, the rainfall forecast will be used as input for the model. The accuracy of the rainfall forecast adds an extra level of uncertainty to the model output. The surrogate models' accuracy as described in Section 3.6 is obtained when a perfect forecast would be used. As this is not realistic, the accuracy of the model output in an operational setting thus also largely depends on the accuracy of the rainfall forecast. As long as the rainfall forecast has a lower accuracy than the surrogate model, the accuracy of the surrogate model is not a limitation in an operational setting.

Validating the rainfall forecast is outside the scope of this thesis, but it is important to consider this when the models are used in an operational setting. The example used in Section 3.3.2 for the first case study, has shown that the quality of the rainfall forecast significantly influences the accuracy of the model output. Not only the peak rainfall intensity, but also the shape of the event highly influences the model results. Since the used rainfall forecast (KNMI Harmonie MOS (KNMI, 2023)) has a temporal resolution of 1 hour and a spatial resolution of 25 km^2, it could be that there are local variations that are not captured by the rainfall forecast due to the coarse resolution.

The machine learning models created in the first and second case study make use of rainfall timeseries with a 5-minute time interval as input. When using a rainfall forecast with an hourly timestep like the KNMI Harmonie forecast, the full potential of the models is thus not used. For future research, the STEPS nowcast from Zware Buien (Schnitzler, 2022) could also be used as rainfall forecast. This nowcast has a 5-minute timestep and a lead time of 2 hours. The algorithm is specifically developed to predict extreme rainfall events with short lead times. The LSTM model for the first case study is not trained on rainfall events with a duration shorter than 4 hours. It is unsure what the effect on the output is when rainfall events are (more than) 50% shorter. This should be researched before implementing the STEPS nowcast in an operational setting for the first case study. The surrogate model from the third case study uses a timestep of 20 minutes, and therefore the STEPS nowcast should be processed (for example by time-averaging) before using it as input for this surrogate model.

5.1.4 Visualisation

A recurring point of attention is the visualization of the surrogate model output. All end-users indicated that they are interested in timeseries of the water depth/water level on a 2D map. On the other hand, they also indicated that the uncertainty in the rainfall forecast is an important point of attention. To communicate the uncertainty in the rainfall forecast, a probabilistic forecast is used. Since the probabilistic rainfall forecast contains 50 ensemble members, it is impossible to visualize all dimensions (2D map, timeseries, ensemble members) in only one figure. Certain choices in the visualisation of the model output are therefore made.

For the first and third case study, the timeseries are known and end-users indicated that they are specifically interested in the timing of the event. However, the currently only the maximum water level/water depth is shown in the figures. A clickable dashboard containing the timeseries in each grid cell will thus definitely have an additional value. Creating dashboards is outside the scope of this thesis but it advised to consider this in further research as it will improve the quality of the surrogate models. For the W2O model, timeseries of the inundation depth per grid cell are also known. In the operational setting, the output of the W2O model is visualized on a dashboard where the timeseries is shown when the end-user clicks on a certain grid cell. The W2O model thus fully meets the requirements of the end-user on this criterium.

A consequence of the probabilistic forecast, is that output is generated for each ensemble member. Therefore, inundation depths and water levels cannot be directly shown on a 2D map, as 50 output values are obtained per grid cell/node. Instead, a certain percentile of ensemble members or a percentage of ensemble members that exceeds a certain threshold is shown on a 2D map. This value is more difficult to interpret, but it is the only possibility when using a probabilistic rainfall forecast.

5.1.5 Initial state of the system

An advantage of using a physically based hydrological inundation model directly, compared to a ML model, is that the initial conditions (e.g. water levels in the waterways) can easily be taken into account. A machine learning model is trained on data from a hydrological inundation model with certain initial conditions that are fixed over the model runs. However, especially for water levels in waterways, these values will differ per event. In an operational setting, these initial conditions might be of great importance. For example, it might be relevant if the soil is already fully saturated, or what the current water level in the waterways is. By using data assimilation techniques (Barthélémy et al., 2018; Clark et al., 2008) and combining them with restart files, the initial conditions of the physically based surrogate model can match the observations.

For the machine learning models in case study 1 and 2, the initial state of the system is not considered as the same initial conditions are applied to all model runs in the training and validation dataset. This implies that, when a new model run is started, the ML model does not use the current situation as a starting point. When large amounts of rainfall have fallen in the past and the storage in the sewage system is already (partly) filled, this is thus not considered. This could lead to an underestimation of the event size when the ML model prediction is done during a rainfall event. For future research, it is advised to investigate the effect of changing the initial conditions in the different model runs. These initial conditions will then be, next to the rainfall timeseries, input for the machine learning model.

In the W2O model, the initial state of the system is considered. As explained in Section 4.2.1, the available soil storage for each day is obtained via OWASIS (2023). OWASIS uses rainfall data (radar) in combination with evaporation data and the national hydrological model (Dutch: Landelijk Hydrologisch Model) to calculate the current available soil storage on a 250 m x 250 m resolution for the entire area of the Netherlands. By making use of radar data, the available soil storage in the W2O model is thus in agreement with reality. Also the available storage in the sewage system and open waters are part of the initial conditions. These values are updated using the rainfall data (radar) of the past 24 hours. The available storage in each of the buckets in the W2O model thus has an initial condition based on the current state.

5.2 Generalisation

The selected case studies serve as an example to show the added value of surrogate models with respect to the W2O model. As explained in Section 1.5, a smaller area is selected within the study area of the end-user that serves as a proof of concept. This section describes to what extent the created surrogate models can be applied to a larger area (e.g. the entire study area of the end-user) (Section 5.2.1), and to what extent the used methodology can be applied other study areas in the Netherlands or international (Section 5.2.2).

5.2.1 Extension to a larger model domain

Generalizing the surrogate models such that they cover a larger area is not necessarily a simple procedure. The first two case studies make use of a machine learning model that is trained on a dataset generated by a hydrological inundation model. If the model domain increases, the training data would be more expensive to generate. For example, the entire area of Amersfoort has more than 2.3×10^4 manholes (compared to 230 manholes for Hooglanderveen) (Kilsdonk, 2021). If an output is taken from each manhole for each timestep, the dataset will become extremely large: up to 30-60 GB per model run (Kilsdonk, 2021). The same holds for the second case study, where the selected area is around 6% of the total area of the municipality of Tilburg (7.6 km^2 for the selected study area versus 118 km^2 for the complete area of Tilburg). Since generalising the model domain will be memory expensive, it is advised to either create only surrogate models for vulnerable areas, or use a larger spatial and/or temporal resolution.

For the third case study, an upscaled version of the D-HYDRO model will likely not meet the requirements of the end-user as the computational time of 30 minutes will be exceeded. To solve this issue, one could think of using parallel computing techniques to speed up the calculation. Another possibility is to compute the calculations for less than all 50 ensemble members. One could think of merging ensemble members (e.g. if multiple of the rainfall forecast ensemble members predict no rainfall at all, this computation has to be performed only once), or compute only the calculations for certain percentiles of ensemble members based on for example the maximum rainfall intensity or cumulative rainfall. Computing the calculations for less ensembles will work for study areas that cover a slightly larger area than this case study ($\approx$18 km^2) but this will not work for the entire area of Hoogheemraadschap van Rijnland, as this area is more than 60 times larger ($\approx$1100 km^2) than polder Vierambacht. As also indicated by the waterboard, it would be an option to compute the calculations for vulnerable areas only.

For the third case study, creating and calibrating the D-HYDRO benchmark model is definitely the most time-consuming part as gathering the data, putting it into the right format for the D-HyDAMO toolbox, and correcting errors takes time. For this study, script are created to clean the data which can be used in future studies as well. This will likely reduce the time to set-up the hydrological inundation model in future studies. As soon as the benchmark model is created, calibrated, and validated using the D-HyDAMO toolbox, it is relatively simple to create the surrogate models: by changing the value of only one parameter in the Python script, the simplifications are implemented. For the first and second case study, the calibrated and validated hydrological models were already available. Generating the training and validation data for the machine learning model takes some computational time, but once the runs are prepared (in batch), most of this time is waiting time to finish the simulations. For the purpose of this thesis, Python scripts were written to put the data into the correct format and train the machine learning model. These scripts can be re-used when the models are generalized to a larger study area. When a calibrated and validated hydrological inundation model is already available, creating the surrogate model is thus relatively simple for all three case studies.

5.2.2 Application to other study areas

The three surrogate models as well as the W2O model could also be applied to other study areas in the Netherlands or international. It is expected that the two machine learning models created for case study 1 & 2 will perform similarly well if trained for other study areas, using the hydrological simulations for that specific area. Other studies have shown that machine learning models are able to accurately predict inundation depth based on rainfall data for many different types of areas (e.g. different topographical variations, other land-use types, other climates) (Berkhahn et al., 2019; Hop, 2023; Kabir et al., 2020; Y. Wang et al., 2018; Zanchetta & Coulibaly, 2020). Also the methodology used to create the surrogate model created for the third case study can be applied to other study areas. It is expected that a similar model performance is obtained for study areas with similar spatial characteristics and topography (i.e. land-use type, elevation). Further research should be conducted to evaluate the performance of this type of surrogate model in areas with a different topography.

An advantage of the W2O model, is that is already available for the entire area of the Netherlands. To use this model in an operational setting within the Netherlands, no generalization to a larger model domain is thus needed. Parts of the required input data (e.g. available soil storage) are, with the currently used sources, only available for the Netherlands. This makes it challenging to directly apply the W2O algorithm to areas outside the Netherlands. Further research is also necessary to determine how accurate the W2O model is for areas with a different topography (e.g. mountainous areas).

A high accuracy of a surrogate model or the W2O model does not necessarily imply that it meets all requirements by the end-user. As also explained in Section 5.1.1, other end-users might have different requirements for a hydrological inundation model to be usable in an operational setting. It is expected that other municipalities have similar requirements as the municipality of Tilburg and Amersfoort, and that other waterboards have similar requirements as Hoogheemraadschap van Rijnland. If this assumption is correct, then the quality assessment of the W2O model and surrogate models are also applicable to other case studies.

5.3 Practical relevance & model suitability

The surrogate models as well as the W2O model have many practical applications. All models have high potential in being useful in an operational setting as they are able to rapidly produce inundation forecasts. These models can support water managers in case of a predicted flooding by informing them about the size and location of the flooding. Especially the W2O model and the machine learning models used in case study 1 and 2 are orders of magnitude faster compared to traditional detailed hydrological modelling methods.

As discussed in Section 3.3.3, the surrogate model for the first case study meets all end-users' requirements except for the correct output variable. The study done by Kilsdonk (2021) has shown that the LSTM model is able to rapidly and accurately reproduce the output of the hydrological inundation model. The used methodology is thus suitable to predict flood volume timeseries per manhole based on rainfall forecasts. However, since this output variable is not of interest for the end-user, this type of surrogate model has limited added value in an operational setting.

The machine learning model created for the second case study is able to rapidly predict the maximum flood depth and location of flooded grid cells. The MAE and RMSE of the model on the validation dataset shows that, on average, the model is able to predict the water depths near the water depth according to the hydrological inundation model. The CSI score of this surrogate model (48%) does not meet the accuracy requirement of the end-user (>80%). Due to this, the used methodology in its current format is not suitable to predict inundation depths with a high CSI score. However, since other studies have shown that machine learning models can be able to accurately predict inundation depths with a high CSI score (Berkhahn et al., 2019; Bermúdez et al., 2018; Erechtchoukova et al., 2016; Hop, 2023; Shi et al., 2017), this type of surrogate model has a high potential in being an added value in an operational setting. For further research, it is advised to improve the machine learning model as it has high potential in being a perfect fit for the end-users within municipalities. As described in Section 3.4.3, improving the accuracy of the machine learning model will require additional testing for other model types, generating more data, and/or using a classification for input and output variables.

The surrogate model created for the third case study is able to accurately compute water levels in waterways and inundation depths on the 2D grid near the values of the benchmark model. The used methodology is thus suitable to accurately predict water levels and inundation depths. Due to the relatively large computational time (30 minutes), this model has added value in an operational setting for smaller areas (see also Section 5.2.1).

A qualitative validation of the W2O model has shown that this model is able to accurately and rapidly predict the location of inundation when the rainfall forecast is accurate. The W2O model is already available for the entire area of the Netherlands on a 1 km x 1 km spatial resolution. The computational time of only 1 minute, a lead time up to 48 hours, and the easily interpretable output variable are also strong points of the W2O model. Further research is necessary to fully validate the accuracy of the W2O model. If the accuracy of the W2O model meets the requirement of the end-users, the W2O model will have great additional value for the end-users of municipalities. For waterboards, the W2O model does fulfil the requirement of predicting inundation depth on a 2D map, but not the requirement of predicting water levels in waterways. Since the surrogate model created in the third case study does fulfil this requirement, the W2O model and the surrogate model combined result in the most additional value in an operational setting.

5.4 Theoretical contribution

In this research, the quality of three types of surrogate model and the W2O model is assessed using the requirements from the end-users in an operational setting. The problem investigation identified a research gap in the current body of literature, namely that it is unknown what the requirements are of a hydrological inundation model to be useable in an operational setting. In existing literature, model requirements are defined (Biondi et al., 2012; Harmel et al., 2014; Jakeman et al., 2006; Ritter & Muñoz-Carpena, 2013; Sargent, 2011; Teng et al., 2017), but were never specifically applied to assess the quality of hydrological inundation models in an operational setting for municipalities and waterboards in the Netherlands. To fill this gap, semi-structured interviews took place with end-users from three different case studies that served as an example. Based on these semi-structured interviews, the requirements of a hydrological inundation model in an operational setting were identified. This research contributes to the existing body of literature by defining requirements of a hydrological inundation model in an operational setting in the Netherlands. In other countries, the responsibilities and tasks of municipalities and waterboards might be different and it is therefore difficult to generalize the outcome of this research to other countries.

Another research gap found is that it is currently unknown to what extent the W2O model and surrogate models are able to meet the requirements from the end-user. In literature, multiple examples can be found of surrogate models that are able to rapidly and accurately reproduce detailed hydrological inundation models (Berkhahn et al., 2019; Bermúdez et al., 2018; Hop, 2023; Kilsdonk, 2021; Liu et al., 2015; Rajaee et al., 2019). However, the quality of these models with respect to the requirements from the end-user in an operational setting, was never described in literature. In this research, three types of surrogate models were created, validated, and operationalized. The quality of these models and the W2O model is assessed using the requirements following from the first research question. It is difficult to compare the quality of the surrogate models found in this study with the models found in other studies. This is because the output variables and flooding types (pluvial versus fluvial flooding) are different. Only a qualitative comparison can be done, which is described in Section 3. This is the first study where the quality of the surrogate models is also assessed using other requirements (e.g. spatial and temporal resolution, computational time). It is therefore not possible to compare the performance on the complete quality assessment with other studies.

6. Conclusion & Recommendations

6.1 Conclusion

The main objective of this study is to research to which extent surrogate models have an added value (compared to the W2O model) in supporting the end-user in an operational setting. Three different research questions were formulated to guide the research and achieve the research objective. In this section, the research questions will be answered.

6.1.1 Model requirements

First of all, the requirements of a hydrological inundation model in an operational setting are defined for the three selected case studies. By doing semi-structured interviews, the modelling goal, perspectives for action, relevant output variables, and required accuracy were specified. The municipality of Amersfoort and the municipality of Tilburg have similar requirements: they want to use the model to identify locations where inundation is expected, including the predicted inundation depth. They use this information to alert relevant parties like contractors and citizens. As output variables, they are interested in the timeseries of the inundation depth on a 2D map. The spatial resolution should be the size of a neighbourhood at maximum, and the maximum computational time should be in the order of minutes. The perspectives for action of a waterboard are different compared to municipalities: next to alerting relevant authorities, they can also pro-actively respond by pumping water into the boezem system or install a temporal pumping station. This perspective for action requires a different type of information provision. Hoogheemraadschap van Rijnland is mainly interested in the water levels in the waterways, next to the expected inundation depth on a 2D map including the locations where the waterways will overflow. The time to implement these measures is longer, and therefore the waterboard requires a lead time between 24 and 48 hours. Regarding the accuracy of the model, the municipalities would like to have a low false alarm ratio, while the waterboard is interested in minimizing the MAE in waterways. These differences in requirements lead to different types of surrogate models.

6.1.2 Quality assessment surrogate models

For the second research question, surrogate models are created to asses to what extent these models are able to meet the requirements following from the first research question. For the first case study, Amersfoort, a Machine Learning model by Kilsdonk (2020) is used that predicts flood volume timeseries for all manholes in the sewage system based on rainfall timeseries as input. This surrogate model meets most requirements by the end-user. However, the output variable, flood volume per manhole, is relatively difficult to interpret and is therefore not suitable for non-experts. The machine learning model created for Tilburg predicts the maximum inundation depth on surface level using the rainfall timeseries as input. This model meets all requirements, except for the accuracy of the model output (Critical Success Index of 48%) and the output variable that does not contain timeseries. Other studies in literature have proven that ML models are able to accurately reproduce the output of hydrological inundation models. This type of surrogate model has thus a high potential, but further research is necessary. For Hoogheemraadschap van Rijnland, a surrogate model is created based on a detailed hydrological inundation model by applying simplifications to the model schematisation. This surrogate model meets all requirements by the end-user.

6.1.3 Quality assessment W2O model

Thirdly, the quality of the W2O model is assessed using the requirements from the end-users. As the algorithm of the W2O model is continuously improved and still work in progress, it is difficult to assess the accuracy of the W2O model. The most recent version (June 2022) has shown that the accuracy is improving compared to previous versions, but further research is needed to fully validate the W2O model. For both municipalities, the W2O model meets all other requirements by the end-user. For Hoogheemraadschap van Rijnland, the W2O model only partially meets the requirements as it does not calculate the water levels in waterways. The W2O model provides sufficient information on the size and location of the inundation, but additional models are needed to also provide the water levels in the waterways and the locations where the waterways will overflow.

On all other criteria, the W2O model fully meets the requirements by the end-user for each case study. The computational time of ± 1 minute is perfect for all three studies operational settings. Besides, the spatial resolution of 1 km x 1 km is sufficient to get a first impression of the predicted inundation on a 2D map. This spatial resolution is also sufficient to alert relevant authorities, which is a perspective for action for the end-user of all three case studies. Also the lead time of 48 hours allows for enough time to implement measures and thus the perspectives for action. The inundation depth per grid cell (timeseries) is an output variable that is easy to interpret and also preferred by the end-users to show the size and location of the inundation on a 2D map. Another advantage of the W2O model is that it is already available for the entire area of the Netherlands. The set-up of the W2O algorithm allows for easy changes in the model parameters and input variables. This makes the W2O model flexible.

6.1.4 Main research objective

Using the answers on the research questions, the main research objective can be achieved. To which extent surrogate models have an added value (compared to the W2O model) in supporting the end-user in an operational setting, mainly depends on the case study. For municipalities, in principle, no additional surrogate models are necessary as the W2O model meets all requirements, provided that the model results are accurate. As the W2O model is still work in progress, further research is necessary to determine if the accuracy of the W2O model is sufficient. On all other criteria, the W2O model fully meets all requirements of both municipalities. The main advantages of the W2O model are the short computational time (± 1 minute), the lead time, and the output variables that are easy to interpret and provide the needed information on both the size and location of the predicted flooding. Besides, the W2O model is already available for the entire area of the Netherlands.

The surrogate model created for the first case study has limited added value, as the output variable (flood volume per manhole) is not the municipality's interest. Post-processing the model output is needed before this surrogate model will have added value in an operational setting.

The machine learning model of the second case study has a potential to be of great additional value. However, the ML model in its current format does not meet the requirements regarding accuracy. It is advised to further research the possibilities, as other studies have shown that machine learning models are able to accurately reproduce hydrological inundation models. Suggestions for future research include the generation of more training data, reducing the number of output variables (e.g. decrease spatial resolution or apply classification), and trying other types of machine learning algorithms. If the machine learning model does meet the accuracy requirement, then this type of surrogate model is a good addition to the W2O model for vulnerable areas (e.g. city centres). The W2O model can be used to get a general impression of the flooding event, while the machine learning model predicts the output on a much finer spatial resolution to ensure a high level of accuracy for the most vulnerable parts of the study area.

For waterboards, the W2O partially meets the requirements by the end-user. The model provides sufficient information on the expected inundation on surface level. Also the short computational time, the lead time, and output variable to show the inundation on a 2D map are strong points of the W2O model for waterboards. However, the W2O model does not fulfil the informational needs regarding the water levels in waterways and the locations of the waterways that will overflow. Additional models are needed to provide all information needed to make informed decisions on implementing measures. The surrogate model created for the third case study fulfils these needs. A limitation of this surrogate model is that it is difficult to apply it on a larger model domain, as the computational load is high. Instead, this type of surrogate model could be used as an addition to the W2O model for vulnerable areas. When using the W2O model to obtain a general impression on the predicted flooding, and using the surrogate model as an additional tool for the most vulnerable areas, they complement each other such that all requirements are met.

6.2 Recommendations

Multiple recommendations can be made regarding future research and the practical context. Throughout this thesis, these recommendations are already mentioned. In this section, the recommendations are listed and elaborated.

6.2.1 Accuracy ML model case study 2

With a critical success index of 48% and a relative mean absolute error of 65%, the accuracy of the machine learning model created for the second case study is relatively low. Since the machine learning model meets the requirements on most other criteria, it is advised to try to further increase the accuracy of the surrogate model. This could be done by either reducing the number of input or output variables, or by generating more data as explained in Section 3.4.3. Despite the fact that the model does not describe the output as timeseries, this type of surrogate model has high potential in having an added value in an operational setting if the accuracy of the model output is increased.

6.2.2 Accuracy W2O model

For both municipalities, the W2O model meets all requirements except the accuracy requirement. As the algorithm of the W2O model is continuously improved and still work in progress, it is difficult to assess the accuracy of the W2O model. Besides, to the lack of available data, it is difficult to numerically validate the W2O model. The most recent version (June 2022) has shown that the accuracy has improved compared to previous versions (Schnitzler, 2022), but further research is needed to fully validate the W2O model. When the accuracy of the W2O model is increased to a CSI of at least 80%, the W2O model meets all requirements of the municipalities' end-users and is thus a perfect fit.

6.2.3 Visualisations

Due to the large number of dimensions of the surrogate model output (2D grid, timeseries, ensemble members, water level/ water depth), it is decided to simplify the model output such that it can be visualized in one figure. For the first and third case study, the model output contains a timeseries component, while this is not visualized in the figures as the maximum over time is taken. Since all end-users indicated that they are interested in timeseries, it is advised to include this when the model is used in an operational setting, for example by creating a clickable dashboard like the W2O model.

6.2.4 Rainfall forecast quality

Currently, only the KNMI Harmonie MOS (KNMI, 2023) probabilistic rainfall forecast is used as input for the surrogate models in an operational setting. The accuracy of the rainfall forecast (indirectly) highly influences the quality of the surrogate model's output. The uncertainty in the rainfall forecast is thus of great importance when interpreting the output of the surrogate models. Part of this uncertainty is captured by communicating the uncertainty using a probabilistic forecast. As practical recommendation, it is advised to, next to the KNMI Harmonie forecast, also use other rainfall forecasts with different lead times and timesteps. For example, algorithms specifically developed to predict extreme rainfall events with short lead times could be used (Schnitzler, 2022).

References

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

Actueel Hoogtebestand Nederland. (2020). *Actueel Hoogtebestand Nederland*.

Arcadis. (2017). *Aanbevelingen meetplan Berkel-Enschot & Udenhout*.

Aricò, C., Filianoti, P., Sinagra, M., & Tucciarelli, T. (2016). The FLO Diffusive 1D-2D Model for Simulation of River Flooding. *Water*, *8*, 200. https://doi.org/10.3390/w8050200

Barthélémy, S., Ricci, S., Morel, T., Goutal, N., Le Pape, E., & Zaoui, F. (2018). On operational flood forecasting system involving 1D/2D coupled hydraulic model and data assimilation. *Journal of Hydrology*, *562*, 623–634. https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.05.007

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, *26*(16), 4345–4378. https://doi.org/10.5194/hess-26-4345-2022

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *The Journal of Machine Learning Research*, *13*, 281–305.

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

Bermúdez, M., Ntegeka, V., Wolfs, V., & Willems, P. (2018). Development and Comparison of Two Fast Surrogate Models for Urban Pluvial Flood Simulations. *Water Resources Management*, *32*(8), 2801–2815. https://doi.org/10.1007/s11269-018-1959-8

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., & Montanari, A. (2012). Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Physics and Chemistry of the Earth, Parts A/B/C*, *42–44*, 70–76. https://doi.org/https://doi.org/10.1016/j.pce.2011.07.037

Bomers, A., Schielen, R. M. J., & Hulscher, S. J. M. H. (2019). Application of a lower-fidelity surrogate hydraulic model for historic flood reconstruction. *Environmental Modelling & Software*, *117*, 223–236. https://doi.org/https://doi.org/10.1016/j.envsoft.2019.03.019

Borga, M., Anagnostou, E. N., Blöschl, G., & Creutin, J.-D. (2010). Flash floods: Observations and analysis of hydro-meteorological controls. *Journal of Hydrology*, *394*(1), 1–3. https://doi.org/https://doi.org/10.1016/j.jhydrol.2010.07.048

Breijn. (2012). *Basisrioleringsplan Kern Udenhout*.

BRO. (2020). BRO Bodemkaart. In *https://app.pdok.nl/viewer/#x=160000&y=455000&z=3&background=BRT-A%20standaard&layers=*.

Caminha. (2023, June 1). *The CFL Condition and How to Choose Your Timestep Size*. SimScale.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., & Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. https://doi.org/https://doi.org/10.1016/j.advwatres.2008.06.005

de Moel, H., van Vliet, M., & Aerts, J. C. J. H. (2014). Evaluating the effect of flood damage-reducing measures: a case study of the unembanked area of Rotterdam, the Netherlands. *Regional Environmental Change*, *14*(3), 895–908. https://doi.org/10.1007/s10113-013-0420-z

Deltares. (2022a). *Juli 2021: overstroming en wateroverlast in Zuid-Limburg*.

Deltares. (2022b). *D-HyDAMO and HYDROLIB github*. Https://Github.Com/Deltares/HYDROLIB/Tree/Main/Hydrolib/Dhydamo.

Deltares. (2023). *D-HYDRO Suite User Manual*.

Erechtchoukova, M. G., Khaiter, P. A., & Saffarpour, S. (2016). Short-Term Predictions of Hydrological Events on an Urbanized Watershed Using Supervised Classification. *Water Resources Management*, *30*(12), 4329–4343. https://doi.org/10.1007/s11269-016-1423-6

FloodSafe Projects. (2021). *Demountable flood barriers*. Https://Floodsafeprojects.Co.Uk/How-Do-i-Protect-My-House-from-Flooding/.

Harmel, R. D., Smith, P. K., Migliaccio, K. W., Chaubey, I., Douglas-Mankin, K. R., Benham, B., Shukla, S., Muñoz-Carpena, R., & Robson, B. J. (2014). Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. *Environmental Modelling & Software*, *57*, 40–51. https://doi.org/https://doi.org/10.1016/j.envsoft.2014.02.013

Henonin, J., Russo, B., Mark, O., & Gourbesville, P. (2013). Real-time urban flood forecasting and modelling – a state of the art. *Journal of Hydroinformatics*, *15*(3), 717–736. https://doi.org/10.2166/hydro.2013.132

Hofmann, J., & Schüttrumpf, H. (2020). Risk-Based and Hydrodynamic Pluvial Flood Forecasts in Real Time. In *Water* (Vol. 12, Issue 7). https://doi.org/10.3390/w12071895

Hoogheemraadschap van Rijnland. (2022). *Rijnland Digitaal op Koers*. https://rijnland.bestuurlijkeinformatie.nl/Document/View/bc9acb73-4714-48a0-9406-9877fc2b1e8a

Hop, F. (2023). *Rapid generation of probabilistic inundation forecasts by utilizing cloud computing and deap learning* [Master Thesis]. University of Twente.

HydroNET. (2023). *HydroNET, your water control room*. Https://Www.Hydronet.Nl/.

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). *Ten iterative steps in development and evaluation of environmental models*.

Kabir, S., Patidar, S., & Pender, G. (2020). A Machine Learning Approach for Forecasting and Visualizing Flood Inundation Information. *Water Management*. https://doi.org/10.1680/jwama.20.00002

Kilsdonk. (2021). *Predicting flooding due to extreme precipitation in an urban environment using machine learning algorithms* [University of Twente]. https://essay.utwente.nl/86086/

Kilsdonk, R. A. H., Bomers, A., & Wijnberg, K. M. (2022). Predicting Urban Flooding Due to Extreme Precipitation Using a Long Short-Term Memory Neural Network. In *Hydrology* (Vol. 9, Issue 6). https://doi.org/10.3390/hydrology9060105

Klimaateffectatlas. (2021). *Klimaateffectatlas, Grijs per buurt*. Https://Www.Klimaateffectatlas.Nl/Nl/.

KNMI. (2021). *Klimaatsignaal'21*.

KNMI. (2023). *KNMI Weather forecast Precipitation - radar*. Https://Dataplatform.Knmi.Nl/Group/Weather-Forecast.

Lange, H., & Sippel, S. (2020). *Machine Learning Applications in Hydrology* (pp. 233–257). https://doi.org/10.1007/978-3-030-26086-6_10

Liu, L., Liu, Y., Wang, X., Yu, D., Liu, K., Huang, H., & Hu, G. (2015). Developing an effective 2-D urban flood inundation model for city emergency management based on cellular automata. *Natural Hazards and Earth System Sciences*, *15*(3), 381–391. https://doi.org/10.5194/nhess-15-381-2015

OWASIS. (2023). *OWASIS, powered by HydroLogic*. Https://Www.Owasis.Nl/.

PZC. (2022, September 10). Flinke overlast door hoosbuien in Zeeland, Veiligheidsregio overspoeld door meldingen. *Https://Www.Pzc.Nl/Walcheren/Flinke-Overlast-Door-Hoosbuien-in-Zeeland-*

Veiligheidsregio-Overspoeld-Door-Meldingen~a5d4b73c/?Referrer=https%3A%2F%2Fwww.Google.Com%2F.

Rajaee, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, *572*, 336–351. https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.12.037

Razavi, S., Tolson, B., & Burn, D. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*, 7401. https://doi.org/10.1029/2011WR011527

Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, *480*, 33–45. https://doi.org/10.1016/J.JHYDROL.2012.12.004

Sargent, R. (2011). Verification and validation of simulation models. In *Engineering Management Review, IEEE* (Vol. 37). https://doi.org/10.1109/WSC.2010.5679166

Schanze, J. (2018). Pluvial flood risk management: an evolving and specific field. *Journal of Flood Risk Management*, *11*(3), 227–229. https://doi.org/https://doi.org/10.1111/jfr3.12487

Schnitzler, B. (2022). *Helpdesk Hydronet*. https://support-hydronet.atlassian.net/wiki/spaces/HLP/pages/1954512900/3.+Waarschuwing+wateroverlast+W2O

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W. (2017). Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. *CoRR*, *abs/1706.0*. http://arxiv.org/abs/1706.03458

Snoek, J., Larochelle, H., & Adams, R. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, *4*.

STOWA. (2020). *Landgebruik kaart*.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, *90*, 201–216. https://doi.org/10.1016/J.ENVSOFT.2017.01.006

TensorFlow. (2022, December 15). *Overfit and Underfit*. Https://Www.Tensorflow.Org/Tutorials/Keras/Overfit_and_underfit.

Verkade, J., & Werner, M. (2011). Estimating the benefits of single value and probability forecasting for flood warning. *IEEE Photonics Technology Letters - IEEE PHOTONIC TECHNOL LETT*, *15*. https://doi.org/10.5194/hessd-8-6639-2011

Wang, Kingsland, G., Poudel, D., & Fenech, A. (2019). Urban flood prediction under heavy precipitation. *Journal of Hydrology*, *577*, 123984. https://doi.org/10.1016/J.JHYDROL.2019.123984

Wang, X., Kingsland, G., Poudel, D., & Fenech, A. (2019). Urban flood prediction under heavy precipitation. *Journal of Hydrology*, *577*, 123984. https://doi.org/10.1016/J.JHYDROL.2019.123984

Wang, Y., Chen, A. S., Fu, G., Djordjević, S., Zhang, C., & Savić, D. A. (2018). An integrated framework for high-resolution urban flood modelling considering multiple information sources and urban features. *Environmental Modelling & Software*, *107*, 85–95. https://doi.org/https://doi.org/10.1016/j.envsoft.2018.06.010

Zanchetta, A. D. L., & Coulibaly, P. (2020). Recent Advances in Real-Time Pluvial Flash Flood Forecasting. In *Water* (Vol. 12, Issue 2). https://doi.org/10.3390/w12020570

Appendices

A Interview questions

Hoofdvraag: wat heeft de eindgebruiker nodig van een model/FEWS voor het voorspellen van wateroverlast door hevige regenval?

Main research question: what (information) does the end-user need from a model to be usable in an operational setting to predict inundation due to heavy rainfall?

➔ In kaart brengen van wensen, behoeftes, ideeën en verwachtingen

Gebruik/doel van het model

Gebruiken jullie het W2O model al? Zo ja, hoe wordt het W2O model ingezet in de operationele context? Zo nee, hoe bepalen jullie op dit moment wanneer welke maatregelen er worden genomen bij wateroverlast?

- Waarvoor gebruiken jullie het model? Dagelijks gebruik om een beeld te krijgen van de situatie? Of ook als tool om beslissingen mee te maken in geval van crisis (crisismanagement)?
- Welke voordelen zien jullie in het model?

Handelingsperspectieven met bijbehorende zichttijden

Stel: in de ideale wereld is alles mogelijk en alles kan gemodelleerd worden. Welke informatie hebben jullie van een model nodig zodat het voldoende handelingsperspectieven biedt?
-> idealiter versus minimaal nodig

- Stel er is een grote kans op wateroverlast. Welke stappen ondernemen jullie?
- En hoe worden deze stappen bepaald?

Besluiten tot handelen:
- kans op wateroverlast
- de bijbehorende potentiële schade
- de mogelijkheid tot het reduceren daarvan
- de kosten van het wel/niet handelen

Is er een overzicht van beschikbare maatregelen + de benodigde tijd voor implementatie? Achterliggende vraag: biedt het W2O model voldoende handelingsperspectieven?

Welke informatiebehoefte is er?
Welke informatie is er idealiter beschikbaar om te beslissen welke maatregelen worden genomen?

Relevante output & vereiste nauwkeurigheid

Welke output parameters zijn relevant voor de handelingsperspectieven? Tijdseries? Of maximale waarden? Inundatie diepte? Volumes? Of bijv ook stroomsnelheid?

Welke resolutie is er (minimaal) nodig voor de handelingsperspectieven? Zowel ruimtelijk als temporeel.

Wat is het meest relevant: timing, hoogte van de piek of de consistentie van de voorspelling?

Betrouwbaarheid/onzekerheid

Het gebruik van kansverwachtingen maakt het mogelijk om risicoafwegingen mee te nemen bij het nemen van beslissingen. Maar, het beslisproces moet wel zijn afgestemd op het ontvangen en verwerken van kansverwachtingen. Hoe is het beslisproces bij jullie ingericht?

Hoe groot mag de onzekerheid/afwijking zijn voor de handelingsperspectieven?

Het W2O model geeft een kans op wateroverlast, wat kunnen jullie daarmee?

Het is ook belangrijk dat de techniek achter de kansverwachting door gebruikers goed begrepen wordt. Heeft de kansverwachting toegevoegde waarde?

Laatste vraag: welke kansen/mogelijkheden zien jullie om het model verder te ontwikkelen?

B Minutes of the interviews

Minutes interview Municipality of Amersfoort

12-02-2023

Jos de Vries – adviseur leefomgeving. Rioleringsspecialist

Anne Vrouwe – adviseur leefomgeving. Water & klimaatadaptatie, bestaande stad en nieuwe ontwikkelingen

---

Laura: gebruiken jullie het W2O model op dit moment al actief bij jullie dagelijkse werkzaamheden?

Mevrouw Vrouwe: ik heb één keer een mail gehad, dat het heel hard zou gaan regenen in Hooglanderveen. En in de ScoreWater tijd hadden we de afspraak dat we het regelmatig zouden checken, dus dat heb ik toen ook wel gedaan. Maar verder niet actief in de dagelijkse werkzaamheden.

Laura: welke middelen gebruiken jullie dan op dit moment om te besluiten hoe jullie gaan handelen in geval van voorspelde wateroverlast?

Mevrouw Vrouwe: eigenlijk niks. We hebben meldingen achteraf.

Meneer de Vries: we hebben natuurlijk wel bepaalde buien waar je rekening mee houdt. Bij het ontwerpen van de riolering kijken we ook 70 jaar vooruit in de toekomst. Dat is namelijk de afschrijving van de riolering. We kijken dan naar buien die één keer in de 100 of 250 jaar voorkomen. We houden dus wel rekening met die stress-test.

Laura: maar dat is niet in de dagelijkse werkzaamheden?

Mevrouw Vrouwe: nee, maar dat is wel iets wat je meeneemt bij grootschalige renovaties. Als je met grotere gebieden (bijvoorbeeld de Indische buurt) aan de slag gaat, dan kijk je wel naar de effecten van hevige neerslag. Je kijkt dan naar de mogelijke maatregelen, bijvoorbeeld het effect van een wadi toevoegen. Maar we weten ook dat we met die wadi erbij nog steeds niet voldoen aan onze eis dat het water niet de panden instroomt bij T=250 jaar. In projecten en herontwikkelingen kijk je dus wel degelijk naar maatregelen. Maar in de bestaande stad, waar verder niks gebeurt, gaat het vooral op basis van waarnemingen. We hebben de stress-testen en we zien waar het water de panden in loopt. Dat zijn wel aandachtslocaties voor ons. Maar in veel gevallen gaan we pas aan de slag als er meerdere meldingen zijn.

Meneer de Vries: integraal inderdaad.

Laura: dus als er meerdere meldingen van wateroverlast binnen komen, dan is dat het moment waarop jullie tot handelen overgaan?

Meneer de Vries: ja, dan gaan wij eerst achterhalen wat de oorzaak is. Bijvoorbeeld afgelopen januari heeft het extreem veel geregend. Dan krijgen we dus veel meldingen. Mensen denken dan dat bijvoorbeeld het rioleringssysteem niet goed functioneert. Maar de oorzaak kan ook zijn dat een kelder zijn die niet waterdicht is. Het ligt dan dus niet aan de riolering zelf.

Mevrouw Vrouwe: zo was er een andere bewoner die last had van wateroverlast en dacht dat dit kwam door de hoge grondwaterstand. Maar uit de metingen bleek dat niet te kloppen. In werkelijkheid ging het om afstromend hemelwater vanuit een parkje.

Laura: dus de bewoner ervaart wateroverlast, maar kan niet altijd de oorzaak benoemen?

Mevrouw Vrouwe: de eerste stap bij meldingen van wateroverlast is dus de oorzaak achterhalen: wat is het eigenlijk? Waar komt het water vandaan? De meeste meldingen van wateroverlast op jaarbasis komen van riolering, verstopte kolken, vanuit beheer dat het niet voldoende functioneert. In de winter is het vooral door hoge grondwaterstanden.

Meneer de Vries: in de herfst zijn er bijvoorbeeld vallende bladeren. Dan is het net alsof het riool niet werkt of verstopt is, maar dat is dan niet zo. Dan kun je wel een model maken, maar als er blad op die kolk ligt, dan werkt die kolk gewoon niet. Dus dat is wel iets om rekening mee te houden: je moet weten waarom er sprake is van wateroverlast.

Laura: er zijn meerdere oorzaken die kunnen leiden tot de overstroming van een riool. Een zware regenbui is er een, maar niet de enige.

Meneer de Vries: dat klopt, er zijn heel veel dingen die meespelen.

Mevrouw Vrouwe: een recente calamiteit was onder de N-weg richting de A28. Daar was een oud riool dicht geschuimd. Er bleek echter nog een zijtakje aan te zitten en dat was even vergeten. Dat zorgde dus voor een groot probleem bij de eerste keer zware regen. Het is bijna nooit standalone dat de zware neerslag voor het probleem zorgt. Hoewel dat met de zomerse piekbuien wel vaker het geval is, dat het rioolsysteem echt vol zit.

Laura: dus het komt wel voor, maar lang niet alle keren dat het riool overstroomt, komt het alleen door veel regen?

Mevrouw Vrouwe: klopt.

Meneer de Vries: en het gaat steeds heviger regenen in kortere periodes. Daar zijn de riolen niet op berekend. Dat kan ook niet, want als het riool in 80 jaar vervangen moet worden, dan wil dat niet zeggen dat het hele systeem meteen op orde is.

Mevrouw Vrouwe: we hebben nu een opgave voor de binnenstad: het saneren van overstorten. In het coalitieakkoord is afgesproken de waterkwaliteit te verbeteren door overstort uit het riool terug te dringen. Om ervoor te zorgen dat er geen vervuild hemelwater meer wordt overgestort, zouden we de 24 overstorten kunnen weghalen maar dan hebben we een probleem. Het gaat natuurlijk over hele oude panden. De regenpijpen aan de voorkant zie je misschien nog, maar aan de achterkant kom je er sowieso niet bij. Daar zitten een heleboel uitdagingen.

Meneer de Vries: en wat er allemaal onder de grond gebeurt, dat is elke keer weer een verrassing.

Laura: mijn volgende vraag is dan wat de toegevoegde waarde van een model dan is voor jullie. Waar zien jullie de voordelen en kansen bij het voorspellen van wateroverlast?

Meneer de Vries: het is wel fijn om te weten waar de wateroverlast is. Daarbij neem je dan aan dat het onderhoud etc. allemaal in orde is. Het is voor ons wel heel belangrijk om te weten waar de overlast precies is. En het is natuurlijk wel fijn om van tevoren te weten dat er een zware regenbui verwacht wordt. Dat is dan wel korte termijn politiek natuurlijk.

Mevrouw Vrouwe: ja, want wij gaan daar niet meer op handelen.

Meneer de Vries: maar wij kunnen wel de bewoners waarschuwen dat er mogelijk een bui komt die voor overlast zal zorgen. We kunnen ook niks tegen die bewoners zeggen, en dan loopt het water sowieso die huizen binnen. Van tevoren waarschuwen is dan al een eerste stap.

Laura: stel er is binnen een aantal uur wateroverlast voorspelt. Dan is het natuurlijk niet mogelijk om dan nog ter plekke de capaciteit van het rioleringssysteem te vergroten. Maar zijn er, behalve bewoners waarschuwen, nog andere maatregelen of handelingsperspectieven die jullie wel hebben?

Meneer de Vries: we hebben nu toevallig een grondwatersanering, en dan willen ze graag gebruik maken van de riolering. Hier zitten gevaarlijke stoffen in, en dat mag je niet lozen op het open water. Deze schadelijke stoffen mogen alleen via het dwa-riool geloosd worden. Bij zware buien kan er sprake zijn van overbelasting van het rioolstelsel, waarbij deze schadelijke stoffen er dan wel uitgaan. Dit geldt ook voor de onttrekking van schoon grondwater bij bijvoorbeeld bouwprojecten. Het is dan dus nodig om vooraf ruimte in je rioleringssysteem te creëren. Want in het geval van hevige regenval kan geëist worden dat de grondwateronttrekking tijdelijk stilgelegd wordt als dit ernstige schade kan veroorzaken door wateroverlast.

Laura: het waarschuwen van bewoners is dus een maatregel die jullie kunnen nemen, maar verder is het altijd eerst achterhalen wat de oorzaak is voordat jullie tot actie overgaan?

Mevrouw Vrouwe richting Meneer de Vries: maar zou jij als gemeente bewoners individueel gaan waarschuwen wanneer er wateroverlast voorspeld is?

Meneer de Vries: als ik van tevoren weet dat er een zware bui is voorspeld en dat dit waarschijnlijk zal leiden tot wateroverlast en schade, dan zou ik ze wel van tevoren inlichten.

Mevrouw Vrouwe: maar stel er is een bui voorspelt van 150mm in 2 uur, dan weten we nu al dat dat overal in de stad voor overlast zal zorgen. Dat betekent dat we in de hele stad die waarschuwing moeten gaan uitgeven.

Meneer de Vries: ja dan wordt het misschien lastiger, maar minder hevige buien kunnen ook al voor wateroverlast zorgen. Veel wijken hebben bijvoorbeeld een groepsapp. Ik ga niet iedereen afbellen met welke maatregelen zij moeten nemen, maar ik wil best iemand informeren met de neerslagverwachting. En als je dat kunt communiceren van tevoren, daar zijn de bewoners dan toch altijd blij mee. Eigenlijk zijn dit maatregelen die je niet wil, want eigenlijk moet het rioolsysteem gewoon op orde zijn. En bij echt zware buien is er dan sprake van overmacht.

Laura: en in stedelijk gebied heeft water op straat eigenlijk altijd te maken met het rioleringssysteem?

Meneer de Vries: ja water op straat is omdat het niet voldoende afgevoerd kan worden, en dat is in Amersfoort toch vooral via de riolering. Als de capaciteit van de riolering is overschreden, dan houdt het op natuurlijk.

Mevrouw Vrouwe: tot nu toe hebben we in Amersfoort heel weinig wadi's en andere boven- en ondergrondse infiltratievoorzieningen. Amersfoort is voor een groot deel wel rioolafhankelijk. In andere gemeenten heb je veel meer oppervlaktewater waar het water naartoe loopt waardoor er veel minder problemen met de capaciteit van de riolering zijn. In Amersfoort is het riool wel bepalend voor de wateroverlast. Dat is altijd wel zo in stedelijk gebied, maar de mate verschilt wel per stad.

Laura: zijn er ook nog maatregelen die direct gekoppeld zijn aan die riolering? Bijvoorbeeld extra pompen inzetten of iets dergelijks?

Meneer de Vries: er zijn wel riolen die altijd vol staan met hemelwater ja, verdronken riolen. Maar als het water aan de buitenkant ook hoog staat, dan heeft pompen geen zin.

Mevrouw Vrouwe: wij pompen niet vooraf. Je hebt natuurlijk wel zomer- en winterpeil, maar verder is het heel beperkt.

Laura: jullie hebben niet vooraf een plan met mogelijke maatregelen of handelingsperspectieven die jullie kunnen inzetten als er zware buien zijn voorspeld?

Meneer de Vries: klopt, het is echt maatwerk.

Mevrouw Vrouwe: het enige wat daar op lijkt, is dat er in de herfst extra vaak geveegd wordt. Op de berg hebben we roostergoten. Als al het blad gevallen is, dan lopen we die roosters nog een keer na om te kijken of alles echt schoon is. Of als na een regenbui zitten de roosters helemaal vol zitten, dan halen we die ook weer leeg. Dat is meer seizoensafhankelijk en niet afhankelijk van de voorspelling.

Meneer de Vries: klopt, dat is gewoon structureel werk en dat staat vermeldt in het bestek. En dat voert een aannemer dan uit. Wat wel voor korte termijn zou zijn, is ervoor kiezen om bijvoorbeeld een tunnel onder water te zetten. Maar dat soort dingen doen wij op dit moment nog niet met het W2O model of met Zware Buien. Het zou in de toekomst wel een overweging kunnen zijn, als er echt geen ruimte is om je water kwijt te raken. Als het dan schade gaat veroorzaken, dan maar in de kelder. En in dat geval moet je dat wel ruim van tevoren weten, dat de kelder nu leeg moet want wij verwachten een bui.

Laura: en wat is dan ruim van tevoren in die context?

Meneer de Vries: soms misschien wel een week van tevoren. Maar kun je een week van tevoren een zware bui voorspellen? Waarschijnlijk niet.

Mevrouw Vrouwe: vaak gaat het dan ook niet om de hele kelder, maar een bepaald segment dat onder water wordt gezet. Dat is dan specifiek gereserveerd daarvoor.

Meneer de Vries: voor de toekomst zou dit wel een keuze kunnen zijn. Afhankelijk van de intensiteit van de bui, kan dan bewust een bepaald gedeelte onder water worden gezet als berging.

Mevrouw Vrouwe: In andere gemeenten gebeurt dat bijvoorbeeld in fietsenkelders. Daar mag en kan 15 of 20cm water in de fietsenkelder staan zonder dat dat schade toebrengt aan de fietsen. Maar daar moet fysiek een hendel omgezet worden voordat die kelder onder water wordt gezet.

Laura: er gaan natuurlijk allerlei afwegingen aan vooraf, voordat zo'n beslissing wordt gemaakt. Stel dat Amersfoort in de toekomst wel de mogelijkheid heeft om water te bergen in bijvoorbeeld kelders of garages. Welke informatie hebben jullie dan vanuit een model nodig om een beslissing te kunnen maken?

Mevrouw Vrouwe: het water vasthouden zou bijvoorbeeld ook kunnen op daken van gebouwen. Deze daken zijn ook gemaakt om een aantal centimeters sneeuw te dragen, dan kan een aantal millimeter water misschien ook wel. Daken kunnen dan ook worden ingezet om het water tijdelijk te bergen. Dat zit dan wel nog steeds op het vlak van particulieren. Al hebben wij als gemeente daar natuurlijk zeker belang bij. Maar in principe is een perceeleigenaar zelf verantwoordelijk voor het opvangen van regenwater.

Meneer de Vries: je kunt dat bergen op het dak dan combineren met de neerslagvoorspelling. Als er dan een zware bui wordt voorspeld over een aantal uur, dan kan de berging nu worden geleegd zodat er ruimte ontstaat voor die bui.

Laura: stel jullie hadden de mogelijkheid om water te bergen op bijvoorbeeld daken of in parkeergarages. Welke informatie hebben jullie dan nodig van een model om een beslissing te nemen?

Mevrouw Vrouwe: ja de vraag is of je op basis van een model waar de riolering niet direct in verwerkt zit, een knoop gaat doorhakken. Dan ga je eigenlijk alleen uit van het maaiveld model, terwijl het probleem eigenlijk verplaatst wordt door de rioolbuizen die onder de grond liggen. En in het geval van water bergen in bijvoorbeeld parkeergarages, dan vind ik dat ook nog lastig omdat die parkeergarages niet van ons als gemeentes zijn.

Meneer de Vries: dat zouden we natuurlijk kunnen verplichten. Dat we afspraken maken met de eigenaren dat ze een bepaalde hoeveelheid water moeten kunnen bergen. Op een andere manier berging realiseren in het binnenstedelijk gebied is eigenlijk niet mogelijk.

Laura: stel een model geeft x% kans op meer dan x millimeter water op het maaiveld. Is dat een output waar jullie iets mee kunnen? Of zijn er andere parameters, bijvoorbeeld stroomsnelheid, misschien ook relevant? Welke informatie hebben jullie nodig om te besluiten tot welke handelingen jullie overgaan?

Meneer de Vries: ja het is wel handig om te weten hoe snel het water stijgt en misschien ook weer zakt. Daar heb je natuurlijk ook het riool bij nodig. Maar die tijd is wel een belangrijke component.

Mevrouw Vrouwe: ja verwachte waterdiepte zit op dit moment niet direct in het W2O model. Alleen de kans op meer dan 10, 20 of 30 mm water op het maaiveld. Het zou een waardevolle toevoeging zijn om ook de waterdiepte zelf direct te kunnen zien. Het handmatig aanpassen van drempelhoogtes voor de alerting zou ook interessant kunnen zijn. Het verschilt namelijk per wijk welke neerslaghoeveelheid voor wateroverlast zou zorgen.

Laura: modellen die wateroverlast voorspellen, gebruiken vaak de neerslagvoorspelling als input. Omdat dat een voorspelling is, zit daar natuurlijk ook al een bepaalde onzekerheid. En als die neerslagvoorspelling onjuist blijkt te zijn en de bui een kilometer verderop valt, dan zijn de gevolgen misschien wel heel anders. Het is dus wel mogelijk om te water op straat op kleine schaal te voorspellen, maar omdat de neerslagvoorspelling onzeker is, kan dat ook een soort schijnzekerheid creëren.

Mevrouw Vrouwe: ja zolang die neerslagvoorspelling zo onzeker is, is het denk ik niet nodig om op kleine schaal de wateroverlast te voorspellen. Dan kunnen we beter de grote schaal aanhouden en accepteren dat dit is wat het is.

Meneer de Vries: de bewoner moet dan zelf inschatten in hoeverre dat voor hem of haar risicovol is of niet. Dat is dan aan de bewoner.

Mevrouw Vrouwe: ja stel dat er 50% kans is op 30mm water op het maaiveld, dan is het aan de bewoner zelf om dat te interpreteren en te besluiten welke acties hij of zij onderneemt.

Laura: maar vinden jullie dat dit een verantwoordelijkheid is die je bij bewoners kunt neerleggen? Het interpreteren van die statistische resultaten vraagt natuurlijk een bepaald denkniveau van, in dit geval, de bewoners.

Meneer de Vries: ik denk het wel. Niet in alle gevallen misschien, maar soms kiezen bewoners er bewust voor om bijvoorbeeld laag te bouwen. Die keuze kan dan gevolgen hebben.

Mevrouw Vrouwe: ik denk ook wel dat het beste is wat je kan bieden op een gegeven moment. Er zitten grenzen aan wat wij als gemeente kunnen qua handelingsperspectieven. Dat geldt bijvoorbeeld bij de riolering in de binnenstad. Zoals ik net aangaf, is er in het coalitieakkoord afgesproken om de waterkwaliteit te verbeteren door overstort vanuit het riool terug te dringen. Als wij alle overstorten saneren, dan kan er water op straat komen bij hevige regen. Maar als gemeente hebben wij dan wel voldaan aan het coalitieakkoord. Het enige wat wij als gemeente dan nog kan bieden is het waarschuwen van bewoners en het communiceren van de kans op wateroverlast. Je zou het natuurlijk liever structureel op een andere manier oplossen.

Laura: dus als ik het goed begrijp, dan zouden jullie de kans op wateroverlast zoals deze nu in het W2O model zit, op die manier communiceren richting bewoners? En jullie laten het dan aan de bewoner over welke actie hij of zij onderneemt.

Meneer de Vries: klopt. Het is dan aan de bewoners om in te schatten welke maatregelen ze nemen of niet. Het is dan fijn als bewoners van tevoren weten hoeveel kans er is op wateroverlast. Je wil natuurlijk graag dat er maatregelen zijn die je nog snel kan nemen zodra de zware bui wordt voorspeld, zodat je de schade kunt beperken. Maar dat is wel moeilijk natuurlijk, het voorspellen kan al niet met 100% zekerheid en daar komt dan nog de snelheid waarmee je de maatregel kan nemen bij.

Laura: maar misschien zijn er maatregelen die niet veel tijd kosten, en daardoor nog wel op tijd geïmplementeerd kunnen worden zodra de wateroverlast wordt voorspeld. Dan kan een voorspelling dus wel van toegevoegde waarde zijn.

Mevrouw Vrouwe: er is bijvoorbeeld een wijk in Amersfoort die over de watergangen heen zijn gebouwd met souterrains. Die liggen 20cm hoger dan de watergang, dus als het water in de watergang stijgt, dan staat het souterrain onder water. Tegenwoordig zijn deze ruimtes ingericht als slaapkamer of keuken. Als gemeente zeggen we nu: het is een souterrain boven de watergang, houdt daar rekening mee. Maar voor die bewoners is het misschien ook wel fijn om zo'n voorspelling te hebben.

Laura: ja dus ook hier zouden jullie de resultaten van het model direct naar de bewoners communiceren?

Mevrouw Vrouwe: ja wij kunnen daar zelf ook geen fatsoenlijke kennisslag overheen slaan denk ik.

Meneer de Vries: ja het is aan de bewoners welke maatregelen ze nemen. De maatregelen die particulieren zelf kunnen nemen, dat zijn ook de maatregelen die uiteindelijk de schade het meest beperken.

Mevrouw Vrouwe: ja als gemeente accepteren wij het dat een weg een keer onder water staat. Dat is voor ons acceptabel. Dat het water een woning instroomt, dat is voor ons niet acceptabel. En voor bewoners zelf natuurlijk ook niet. Maar op dat moment zelf kunnen wij niet heel veel meer doen dan waarschuwen. En bewoners zelf kunnen dan bijvoorbeeld schotten plaatsen bij hun deuren.

Meneer de Vries: en bijvoorbeeld bij putdeksels die omhoog komen bij hevige neerslag, die zetten we dan vast. Maar dat zijn structurele maatregelen. Als we weten dat er zware neerslag wordt voorspeld, dan gaan we niet van tevoren die deksels weghalen of iets dergelijks.

Laura: en nog even terugkomend op die resolutie. Bij het generieke W2O model is nu gekozen voor een schaal van 1x1km omdat ook de neerslagvoorspelling op die schaal komt. Is dit een schaal waar jullie iets aan hebben? Of hebben jullie behoefte aan een andere schaal, met in het achterhoofd dat er een bepaalde onzekerheid zit in je neerslagvoorspelling?

Mevrouw Vrouwe: ja als we de informatie willen doorgeven aan bewoners, dan is het niveau van Amersfoort als geheel al voldoende. Dan geven wij zelf als gemeente aan wat de aandachtsgebieden zijn, bijvoorbeeld laagstgelegen delen of kwetsbare plekken in het rioolsysteem. Ik denk dat je op het niveau van heel Amersfoort, dat dat al van toegevoegde waarde kan zijn.

Meneer de Vries: vroeger werd er anders gerekend. Door de hevige buien die we nu steeds vaker hebben, wordt de dimensionering van de rioolbuizen helemaal veranderd. Dat is een groot project met hoge kosten. De vraag is ook of het de investering waard is, met een overstroming die een keer in de honderd jaar plaatsvindt. Het uitgangspunt is dan dus dat we een keer in de honderd jaar schade hebben door wateroverlast. Dat is dan overmacht. Bij minder hevige buien mag het water dan wel op het maaiveld staan, maar dan mag het geen schade veroorzaken.

Mevrouw Vrouwe: maar op een schaal van heel Amersfoort weet je dan wat de kwetsbare plekken zijn, ook door de stress-testen die al worden gedaan. Op wijkniveau kan ook, bijvoorbeeld 1x1km of 2x2 km.

Meneer de Vries: ja en bij een herinrichting weet je dat dus ook, door die stress-testen. Dan ga je maatregelen nemen voor de toekomst, bijvoorbeeld met wadi's. Of als het riool vervangen wordt, dan neem je de leiding een diameter groter. Maar als een zware bui binnenkort wordt verwacht, dan kunnen we op dat moment zelf niet zoveel meer doen.

Laura: en ben ik bij mijn laatste vraag. Welke kansen of mogelijkheden zien jullie in het voorspellen van wateroverlast?

Mevrouw Vrouwe: ik zie denk ik veel kansen nog bij particulieren en grootgrondbezitters. Meer dan bij ons als gemeente. En het waterschap misschien, omdat je het oppervlaktewater natuurlijk kan kijken naar stuwen etc. Zij hebben meer maatregelen die ze vooraf kunnen nemen om de schade te beperken. Als gemeente zetten wij daar ons daar, tot nu toe, helemaal niet op in.

Meneer de Vries: ja en het bergen en vasthouden van water, bijvoorbeeld op daken of in garages. Dat de berging leeg is voordat de volgende bui eraan komt.

Mevrouw Vrouwe: Ja dat is wel een van onze grootste vragen. We hebben op jaarbasis voldoende neerslag, maar niet op de juiste momenten. Misschien dat zo'n model daar ook in kan faciliteren. Dat zou een stap in de toekomstbestendige richting zijn. Maar ik moet wel eerlijk toegeven dat dit iets is wat wij super interessant vinden voor in de toekomst, alleen we zijn nu nog niet zover dat we dan ook daadwerkelijk zouden betalen voor het product.

Minutes interview Municipality of Tilburg

15-02-2023

Jan Janssens-Baan – adviseur Stedelijk Water gemeente Tilburg, afdeling ruimtelijke uitvoering , team vakspecialisten

Jaap Jansen – interim beleidsmedewerker gemeente Tilburg

Laura : in hoeverre gebruiken jullie het W2O model of iets vergelijkbaars op dit moment al in de operationele context?

Meneer Janssens-Baan: onze afdeling gebruikt het dagelijks. Zowel W2O als Zware Buien. Daarnaast gebruiken we ook het KNMI harmonie model om de verschillen te zien. Voordat wij deze modellen gebruikten, maakten we vooral gebruik van buienradar, KNMI, meteoalarm.eu. Op dit moment hebben Zware Buien en W2O de zwaarste doorslag, van de andere systemen maken we nauwelijks nog gebruik.

Laura: En hoe zetten jullie deze modellen dan in?

Meneer Janssens-Baan: wij gebruiken Zware Buien om de neerslag te kunnen zien aankomen. En het W2O model moet nog wel verder doorontwikkeld worden om het voor ons bruikbaar te maken. Want bijvoorbeeld de ondergrondparameters die erin zitten, die voldoen eigenlijk niet voor het gemeentelijke gebied. We gebruiken dus vooral Zware Buien. We zijn met vier collega's op onze afdeling en we rouleren per week wie de toerbeurt heeft om te kijken wat er aan neerslag zou kunnen vallen in die week. En bij potentiële calamiteiten of zware neerslag dan is het de bedoeling dat wij de consignatiedienst, de beheerder van de gemalen en ons managementteam daarover inlichten. En we houden dan voor hen ook bij in die periode wat er precies gaande is, hoe het ontwikkelt en of het afgeschaald wordt.

Laura: en de acties die bijvoorbeeld de consignatiedienst onderneemt, die worden ook echt bepaald door de neerslagverwachting uit Zware Buien?

Meneer Janssens-Baan: de consignatiedienst loopt buiten voor wanneer er een calamiteit plaatsvindt. Bij een ongeval verricht de consignatiedienst de hand- en spandiensten. Bijvoorbeeld hekken neerzetten en hulpdiensten inschakelen. Wij lichten de consignatiedienst in over de potentieel zware neerslag. Zij zijn dan alvast alert en weten dat ze misschien moeten gaan handelen Zij weten dan dat ze misschien een aannemer of hulpdiensten moeten gaan bellen. Dat is onze rol als gemeente daarin.

Meneer Jansen: ja het is vooral voorbereiden op, alleen maar mensen alert maken. Verder kunnen we niets van tevoren.

Laura: stel dat er over een aantal uur een zware bui wordt voorspeld, dan melden jullie dat bij de consignatiedienst. En wanneer gaan zij dan over tot actie?

Meneer Janssens-Baan: zij gaan dat pas doen wanneer er daadwerkelijk iets is gebeurt. Er is eigenlijk geen handelingsperspectief voor de consignatiedienst. Er zijn geen dingen die zij vooraf al kunnen doen, behalve alert zijn. Stel dat ze bijvoorbeeld toch al ergens hekken zouden neerzetten en de bui valt net wat verderop, dan is het voor niks.

Meneer Jansen: ja en daar schrikken bewoners van, terwijl het misschien helemaal niet nodig blijkt.

Meneer Janssens-Baan: en de operationeel beheerder van de gemalen, die kan eigenlijk ook nog niks doen in die periode. De meeste gemalen die we hier hebben zijn allemaal gemengde riolering gemalen. Dus die zijn gemaakt voor vuilwater- en regen afvoer. Dus er komen vrij grote leidingen daar binnen om ook het regenwater te kunnen verwerken. Maar in de periode voor de regenbui, is dat regenwater er nog niet. Het enige wat de pompen dan dus kunnen doen is heel hard pompen voor een heel klein beetje vuil water. Dus de pompen kunnen niet handelen vooraf, zo zit het rioolsysteem niet in elkaar. Het enige wat wel kan, is alvast het alert maken van bijvoorbeeld aannemers zodat zij weten dat ze eventueel gebeld kunnen gaan worden. Hetzelfde geldt voor het management team, dat zij graag op de hoogte gehouden willen worden zodat ze ook richting het bestuur kunnen communiceren. Dan zijn het bestuur en de wethouders op de hoogte dat de burgerij misschien gaat bellen/mailen/twitteren over de wateroverlast en het feit dat wij dit inmiddels ook in de gaten houden.

Meneer Jansen: het rioolsysteem is niet ontworpen voor de extreme gevallen en kan dat soort calamiteiten niet opvangen. En het van tevoren leger pompen van het riool levert zo weinig extra capaciteit, dat voegt eigenlijk niks toe. Het zou wel kunnen dat er in de toekomst systemen komen waar wel meer waterberging in zit. Dat je dus van tevoren actief die berging zou kunnen leegpompen. Er worden wel studies gedaan naar Real-time-control (RTC)-sturing van het rioolsysteem. Uiteindelijk is het stelsel dat hier in de gemeente ligt gewoon ontworpen op afvoer. Er zit wel een stuk berging in de leidingen maar die wordt eigenlijk al geledigd door droog weer. Dan trekt het in een paar uur vanzelf weer leeg. Dus stel er zou net een bui zijn gevallen, dan zou je nog een beetje extra kunnen pompen, maar dat is heel minimaal. Het gaat veel meer om de gevolgen bovengronds van die neerslag. Dus hoe kunnen we alert zijn of klaar staan bij tunnelbakken en knelpuntlocaties die we kennen. Daar kunnen we dan onze maatregelen nemen zodra de neerslag ook daadwerkelijk is gevallen. Op dit moment is er geen maatregelenplan voor hevige regenval omdat er dan over het algemeen geen sprake is van een veiligheidsrisico.

Meneer Janssens-Baan: ja dwars door Tilburg loopt het spoor. Op vijf locaties wordt het spoor doorkruist door een tunnel of verdiepte brugconstructie. En dat zijn vijf locaties waarvan we weten dat daar een probleem zou kunnen ontstaan bij extreme neerslag. Voor hulpdiensten is dat dan een probleem. We hebben dat het afgelopen jaar helemaal uit laten zoeken: bij welke neerslag er wat gebeurt, en ook wat de handelingsperspectieven zijn en wat we nu al vooruitlopend kunnen doen. Bijvoorbeeld iets met de gemalen of het anders inrichten bovengronds. Uit dat onderzoek is gebleken dat we ons moeten richten op één onderdoorgang want de andere onderdoorgangen zijn 'niet te redden'. Wij weten nu wat de situatie is en we zijn nu met onze afdeling verkeer aan het kijken hoe we nu verdergaan. Bijvoorbeeld of we bij extreme neerslag (als tunnels onderwater lopen) met digitale bebording kunnen aangeven dat mensen niet hierin moeten rijden maar moeten omrijden. Daar zijn we nu mee bezig met de beleidsmedewerkers van verkeer. We zoeken vooral een korte termijn oplossing met bijvoorbeeld borden omdat een betaalbare lange termijn oplossing eigenlijk niet haalbaar is. Bij één locatie willen we op een gegeven moment wel iets meer mee gaan doen omdat dat een belangrijke onderdoorgang is. Daar willen we kijken of we daar ook fysiek iets mee kunnen doen voor de lange termijn, maar dat gaat veel geld kosten. Tot die tijd is bewegwijzering dan misschien een optie. Dat kunnen we dan bijvoorbeeld koppelen aan peilmeters in het rioolsysteem. Dat gelijk automatisch de bebording aangaat als die meten dat er water op straat staat. We zijn dat nu aan het uitzoeken.

Laura: en het voorspellen van wateroverlast via bijvoorbeeld W2O daar nog een rol in kunnen betekenen?

Meneer Janssens-Baan: jazeker. Dan zouden we bijvoorbeeld per pixel kunnen benoemen wat het gebied nog kan hebben voordat het onder loopt.

Laura: en de digitale borden aanzetten om het verkeer om te leiden is een maatregel die relatief weinig tijd en geld kost als die borden er eenmaal staan.

Meneer Janssens-Baan: klopt. Maar het is niet actief de problematiek oplossen, het is heel passief omgaan met de wateroverlast.

Laura: maar het is wel een manier om de schade te beperken voordat het probleem met bijvoorbeeld de tunnels structureel is aangepakt. We noemden net al kort dat het W2O model ook een rol zou kunnen spelen bij het nemen van deze besluiten/maatregelen. Welke informatie zou het W2O model of een ander model moeten geven zodat jullie de juiste keus kunnen maken? Welke informatiebehoefte is er bij dit soort overwegingen?

Meneer Janssens-Baan: het zou mooi zijn als er alert (mail, whatsapp, pushbericht of iets dergelijks) komt bij voorspelde wateroverlast. Het is veel accurater als dat alert dan direct binnenkomt bij de aangewezen personen (bijvoorbeeld consignatiedienst, MT, de beheerder). Die lopen allemaal met een telefoon rond en die kunnen er dan iets mee gaan doen. Als dat systeem rechtstreeks bij hen komt, dan zitten wij er als menselijke beoordelaar niet meer tussen. Daar zouden wij als team van adviseurs graag naar toe willen: dat het systeem zo goed functioneert, dat we het rechtstreeks durven neerleggen bij de mensen die er iets van zouden moeten vinden. Op dit moment zijn wij ze nog aan het sturen naar de vervolgstappen. Het zou mooi zijn als de adviseurs daar tussenuit kunnen want ook zij hebben niet heel veel meer handelingsperspectieven anders dan faciliteren aan de omgeving.

Meneer Jansen: ja het voordeel van de rechtstreekse alerting is ook dat je sneller bent. De menselijke factor en de verloren tijd door een onnodige tussenstap gaan er dan uit. Ik denk dat de behoefte ook is dat we weten bij welke maatstaf het nou daadwerkelijk nodig is om alert te zijn. Je wil niet te pas en te onpas een alert uitsturen, dan wordt een alert niet serieus genomen. De betrouwbaarheid van het alert is dus ook heel belangrijk. We weten dat het per wijk in Tilburg verschilt hoeveel de wijk aankan qua neerslag. Maar hoeveel elke wijk precies aankan, welk volume echt een knelpunt wordt, dat zouden we graag willen weten.

Meneer Janssens-Baan: en daarvoor is het dus nodig om in W2O per pixel te definiëren wat de opnamecapaciteit is. We weten uit ervaring dat 40mm/uur in de ene wijk wel voor wateroverlast zorgt, en in de andere wijk niet. Het is voor ons dan belangrijk dat we dat verschil ook terugzien in het model. Dat is er dus echt per wijk/pixel onderscheid wordt gemaakt. We hebben het over die betrouwbaarheid ook wel gehad aan het begin van dit traject samen met het KNMI. Als je er altijd zeker van wil zijn dat je alert terecht is, dan ga je ook meldingen missen. Dat is dan de balans, daar waren we toen naar aan het zoeken en aan het kijken. Ik heb nu voor mijzelf een drempelwaarde gevonden die voor mij goed werkt (40% kans op meer dan 30 mm in 24 uur). Dat is minder dan 50%, maar voor mij persoonlijk is dit een goede drempel om alert te worden. En daarna ga ik dan bedenken of ik er ook iets mee moet. Ik ontvang dan niet te pas en te onpas een melding. Voor mij is deze drempelwaarde zo goed, zeker die 40%. Die 30mm is wel nog iets waar we naar moeten kijken, misschien dat die drempel moet verschillen per wijk of woonkern.

Meneer Jansen: ja het zou mooi zijn als we het niveau kunnen bereiken van een aanpak per wijk. En anders moeten we op z'n minst weten wat dan de ondergrens en bovengrens is. Het blijft een continu proces van niet teveel alerten, maar je wil liever ook niet missen. En dat zit ook in de onzekerheden van het model: die verwachting zal steeds scherper/nauwkeuriger worden naarmate de daadwerkelijke bui dichterbij komt.

Meneer Janssens-Baan: voor ons is het denk ik ook anders dan voor bijvoorbeeld een waterschap, en zeker gepolderde waterschappen. Die zien zware regenval over een heel gebied komen waarbij de boezem en alle watergangen al vol staan. Dan weten ze dat ze snel moeten gaan pompen. Ze moeten dan een halve dag tot twee dagen van tevoren beginnen met pompen om al dat regenwater op te kunnen vangen. Dat is echt een praktisch handelingsperspectief. Zij moeten veel verder van tevoren weten of de bui gaat vallen of niet. Want stel ze kiezen ervoor om te gaan pompen en de bui valt uiteindelijk niet, dan staan de watergangen misschien wel een aantal weken leeg. Die zuiverheid van de voorspellingen is dus zeker heel erg van belang.

Laura: ja dit soort maatregelen kosten natuurlijk geld, en als de bui dan uiteindelijk ergens anders valt, dan zijn er nog steeds gevolgen. Alleen alert worden kost verder niks, behalve dat mensen de waarschuwing minder serieus nemen als het vaak gebeurt. Dat is dus ook een belangrijk verschil tussen waarvoor het model ingezet wordt. Is het alleen om inzicht te krijgen en alert te worden, of worden er besluiten en maatregelen genomen op basis van de voorspelling.

Meneer Janssens-Baan: ja, maar dat kan natuurlijk ook samen gaan. Uit de consortiumbijeenkomst eerder deze week bleek ook dat er behoefte was om een andere gebiedsindeling te hebben (waterschapsgebieden, stroomgebieden, polderdelen etc.). Dat vind ik ook met het stedelijk gebied. Tilburg heeft nu pixels van 20x20m, maar eigenlijk zijn het pixels die vallen in één groot vlak van 1x1km omdat de neerslagverwachting op die schaal is. Als die neerslagverwachting één pixel opschuift, dan kleuren een heleboel pixels in Tilburg in een keer groen. De gevolgen zijn hetzelfde, alleen de pixels zijn kleiner. Hoe kleiner je de pixels maakt, hoe groter de afwijking kan zijn als de brongegevens veranderen. Dat zorgt voor schijnzekerheid. Het is dus wel belangrijk dat die brondata ook op kleine schaal beschikbaar is om die diversiteit te creëren. Tilburg is nog relatief plat (8meter hoogteverschil over 8km), maar bijvoorbeeld in Limburg is er al veel hoogteverschil binnen een pixel. Daar maakt dit dus wel degelijk een groot verschil.

Laura: nog even terugkomend op waar we het net over hadden. Stel we kijken naar wateroverlast op wijkniveau. Welke informatie heeft dan bijvoorbeeld de buitendienst nodig om de juiste beslissingen te maken? Gaat dat vooral over de maximale waterdiepte? Of ook over de snelheid waarmee het water stijgt? Welke informatie is daar relevant?

Meneer Janssens-Baan: dat is eigenlijk voor ons als adviseurs ook al interessant, niet alleen voor de buitendienst. Stel er gaat 40mm neerslag vallen de komende 24uur. Als dat gestaag valt, dan kan het riool dat wel aan. Maar als het binnen 2 uur valt, dan is het al een heel ander verhaal. De snelheid waarmee de neerslag valt en de locatie waar het valt, dat is allebei heel erg van belang. Zoals we eerder ook al zeiden, verschilt het heel erg per wijk hoeveel neerslag de wijk aankan. En de snelheid waarmee het valt, hangt daar mee samen. En het volume van de neerslag natuurlijk. Die drie dingen samen, de locatie, de snelheid, en het volume zijn het meest belangrijk. En dan kunnen wij dat hangen aan een bepaalde capaciteit per wijk.

Laura: ja dat is duidelijk, dat gaat over de eigenschappen van de neerslag. Maar als we dan specifiek kijken naar de wateroverlast en water op straat. Stel dat de voorspelling is: 60% kans op meer dan 50mm water op het maaiveld. Is dat dan informatie waar jullie iets mee kunnen? Is dat voldoende, of hebben jullie een andere informatiebehoefte?

Meneer Janssens-Baan: bedoel je bijvoorbeeld hoe snel het water ook weer zakt? ja dat is wel echt al een stap verder. Ik denk dat dat best wel ingewikkeld zal zijn. Dan moet je het hele rioolmodel (inclusief afvoercapaciteit, alle gemalen die ingeregeld zijn) verwerken in het model. En het risico daarvan is, als er in de praktijk iets gaat veranderen qua regulering van het gemaal, dan moet dat ook weer aangepast worden in het model.

Meneer Jansen: ja het helpt als je weet hoe het systeem van Tilburg in elkaar zit. In de leidingen onder de grond zit berging. Tot een millimeter of 10 is er berging in die leidingen beschikbaar. De pompen zijn aan het werk, maar die houden dat niet in een uur bij. Die zijn 10 tot 12 uur aan het werk om de leidingen leeg te pompen als ze vol staan. Als de berging hier in Tilburg vol zit, en het blijft regenen, dan hebben we een probleem. Dan gaan de overstorten werken, maar er zijn niet voldoende overstorten om de stad volledig te kunnen ontlasten. Dan gaat het water dus op straat staan. Dan is de vraag of dat in een woonstraatje is waar bijna geen verkeer is, of bijvoorbeeld een hoofdverkeersader. Dat maakt voor ons uit. Dus het systeem werkt zo dat het water op straat gaat staan als de berging vol zit. Dan kunnen wij dat niet meer wegpompen. Vanaf dat moment moeten we opletten en hopen dat de bui weer afzwakt en het niet de hele stad doortrekt. Daarom zijn wij alert op die buien, we weten hoeveel mm neerslag het systeem aankan qua berging en nog een beetje pompcapaciteit. En als het systeem dan vol zit, dan moeten we echt opletten.

Meneer Janssens-Baan: en in woonwijken vinden wij een beetje water op straat niet erg. Het wordt een probleem als het tegen de dorpel aan komt te staan en het water naar binnen stroomt. Of bij niet-zelfredzame mensen zoals ziekenhuizen, bejaardenhuizen etc. maar ook energievoorzieningen zoals hoogspanningskasten. Die locaties hebben wij op een kaart staan, en daar moeten we extra goed op letten.

Meneer Jansen: en ik zeg wel dat we alert moeten zijn en ‘alle hens aan dek’, maar eigenlijk kunnen we niks. We zijn gewaarschuwd en we weten dat we misschien in actie moeten komen. Daarnaast informeren we de bestuurders over wat we weten. Maar daar zit bij ons de grens, er is verder geen handelingsperspectief.

Meneer Janssens-Baan: misschien dat er in de toekomst wel op een andere manier gecommuniceerd gaat worden. Nu sturen wij hier intern een mailtje met een plaatje erbij van de voorspelling. Op die manier krijgt het MT zonder technische of inhoudelijke achtergrond er ook een beetje een beeld bij. Dat mailtje gaat ook naar de gemaalbeheerder. Maar stel we gaan op een andere manier communiceren, dan is de vraag of we dat bericht ook direct naar burgers willen sturen die in een onveilige omgeving wonen. Dat heeft ook heel veel met een risicodialoog te maken. Wij hebben al een aantal risicodialogen gedaan, bijvoorbeeld met een industrieterrein. We hebben toen een aantal bedrijven kenbaar gemaakt wat wij hebben uitgerekend, wat het probleem is wanneer het ernstig regent, en welke maatregelen wij gaan nemen. Die maatregelen kosten heel veel geld en gaan wij doen op gemeentegrond. Maar na het nemen van de maatregelen, zou het kunnen dat er toch nog problemen ontstaan bij een bepaald theoretisch voorval. En wij kunnen dat als gemeente niet oplossen, dus de vraag is of die bedrijven daar zelf iets aan kunnen en willen doen. Bij dat gesprek kan een model goed helpen. En als je de mensen hebt geïnformeerd over wat je als gemeente wel en niet kan doen, en je hebt de maatregelen ook uitgevoerd, is dat dan het moment dat je die pushberichten ook naar de burgers doorzet? Zover zijn we nu nog niet, maar ooit.

Laura: ja vandaar dat ik net ook vroeg naar de informatiebehoefte vanuit zo’n model. Want als de resultaten worden doorgezet naar burgers, dan vraagt dat om een andere aanpak. Dan moet elke burger in staat zijn om die resultaten te kunnen interpreteren. Een statistische kans op wateroverlast is dan misschien niet de juiste output om te communiceren.

Meneer Jansen: maar in de praktijk gebeurt dit wel. Die mensen gebruiken nu hun weer app, en sommige mensen weten dat er paniek is in Tilburg bij zoveel millimeter. Er zijn mensen die in de zomer op de camping zitten en dan bij een bepaalde neerslagvoorspelling snel naar huis racen om nog snel een zandzak of iets dergelijks neer te leggen. Dat is dan op eigen initiatief, maar die mensen zijn er. En als gemeente kan je dan wel op een punt komen waarop je zegt: wij kunnen jouw probleem niet oplossen, dit valt buiten onze verantwoordelijkheid, maar we kunnen jou wel op een zo goed mogelijke manier informeren. Er moet dan wel uitleg bij over wat het bericht betekent. Er zijn mensen bij die hier echt behoefte aan hebben. Zodat ze dan zelf een keuze kunnen maken over de maatregelen die ze nemen.

Meneer Janssens-Baan: maar zover zijn we nog lang niet. Eerst moeten we zelf goed kijken naar de brongegevens en hoe we daar intern mee omgaan en intern de mensen hierin meenemen. De vorm waarin het nu gepresenteerd wordt is voor ons als adviseurs goed leesbaar. En voor de mensen waar we het naartoe sturen, bijvoorbeeld het MT, daarvoor zijn de modelresultaten wel heel erg cryptisch. Zij hebben net wat meer uitleg nodig. De consignatiedienst en de gemaalbeheerders kennen de locaties heel goed, dus hen zegt de modeloutput wel iets. Zij begrijpen dit en kunnen er iets mee. Het verschilt dus nogal per groep welke resultaten goed leesbaar en te interpreteren zijn. Voor de burgers zal het echt op een andere manier moeten.

Laura: we hadden het er net al even over dat de betrouwbaarheid van de weersverwachting heel belangrijk is. Een voorspelling heeft meerdere eigenschappen. Bijvoorbeeld de timing van de piek, maar ook de hoogte/intensiteit van de regenbui. Bij een voorspelling, welk van deze dingen moet dan het meest accuraat zijn? Gaat het erom dat de timing van de voorspelling precies klopt, of meer om de intensiteit van de bui zelf? Uiteindelijk gaat het altijd om de balans, maar welke informatie is voor jullie het belangrijkst in de operationele setting?

Meneer Janssens-Baan: ja het is natuurlijk altijd een samenloop. Zoals ik net al aangaf, als er 25mm valt in 24 uur, dan zijn er geen grote problemen. Maar als dat in een korte periode valt (één uur), dan zorgt dat wel voor problemen. Daarnaast is de zuiverheid van de locatie ook erg belangrijk.

Meneer Jansen: ja als ik het even probeer te vertalen: wil je weten dat er serieuze neerslag is over exact 3 uur, of wil je weten dat er ergens de komende 20 uur neerslag valt met deze intensiteit. Voor Tilburg is dat tweede belangrijker, omdat dat bepaalt of het ook daadwerkelijk voor problemen gaat zorgen. Wij hebben geen handelingsperspectief. Dus als we het een uur eerder weten, dan is dat mooi, maar verder maakt dat voor ons niet heel veel uit. Het gaat er voor ons vooral om dat we zo concreet mogelijk weten hoe de bui eruitziet. De eigenschappen van de bui zijn voor Tilburg dus wel echt het belangrijkst.

Laura: dat komt natuurlijk ook omdat jullie niet actief aan de voorkant heel veel kunnen doen in de periode voordat de bui valt.

Meneer Jansen: precies. Ik verwacht dat dit voor een waterschap precies andersom is. Zij hebben handelingsperspectieven en willen weten hoeveel tijd ze nog hebben om bepaalde maatregelen uit te voeren. Zij willen het dus vooral tijdig weten, en wij willen de exacte gebeurtenis het liefst weten.

Meneer Janssens-Baan: de 48-uurs voorspelling kijk ik wel altijd door en dat maakt ons alert. Vooral ook om te kijken of er wat verandert in de voorspelling. Als de bui ineens steeds zwaarder wordt, vooral ook in de zomerperiode, dan ben ik daar extra alert op.

Meneer Jansen: ja dat heeft ook met die menselijke factor te maken. Als wij er als tussenpersoon tussenuit gaan, dan maakt het ons eigenlijk niet uit hoe ver van tevoren die voorspelling duidelijk is. Zolang we maar net genoeg tijd hebben om dat alert eruit te sturen.

Laura: het W2O model geeft een kansverwachting voor de wateroverlast, waarbij de onzekerheid in de neerslagvoorspelling dus terug te zien is in de modelresultaten. Voor het interpreteren daarvan, is dus een beetje kennis over statistiek nodig. Is die probabilistische voorspelling voor jullie van toegevoegde waarde?

Meneer Janssens-Baan: ja voor mij zeker wel. Dat is ook de meerwaarde ten opzichte van bijvoorbeeld Buienradar of KNMI.

Meneer Jansen: ja door die kansverwachting geef je die onzekerheid in de neerslagvoorspelling een plek.

Laura: en als we het hebben over die handelingsperspectieven. Is er dan bij jullie nog een verschil tussen bijvoorbeeld 70% of 80% kans op wateroverlast?

Meneer Janssens-Baan: nee, dat maakt niet zo heel veel uit verder. Het is meer om een beeld te krijgen. Het is vooral belangrijk dat we op tijd alert genoeg worden. Ik heb dat dus voor mezelf op 40% gezet, dat vind ik voor mezelf een goede drempel om alert te worden. Die 40% is een goede balans voor mijzelf, maar die is niet bij alle collega's hetzelfde. Anderen hebben bijvoorbeeld 30% ingesteld, dat is ook met opzet gedaan. Om te kijken hoe het nou precies werkt en wat wij fijn vinden. We hebben daar intern nog geen evaluatie over gedaan.

Laura: en als laatste vraag: welke kansen of mogelijkheden zien jullie nog om het W2O model of een ander model verder door te ontwikkelen? Het direct communiceren richting andere partijen is daarvan een voorbeeld. Maar zien jullie nog andere kansen?

Meneer Janssens-Baan: ja en ook dat verschil per wijkniveau. En waar we het gisteren over hadden bij de consortiumbijeenkomst: neural networks. Daar ben ik ook wel heel benieuwd naar. Zodat we veel sneller en heel nauwkeurig kunnen rekenen. Ik was wel onder de indruk van die presentatie van Fedde Hop. Dat heeft zeker potentie.

Laura: Ja daar ben ik het helemaal mee eens. Ik ga in mijn studie ook kijken naar de mogelijkheden voor Tilburg. Een lastig ding bij neural networks is wel dat je een heleboel (nauwkeurige) data nodig hebt om het model te trainen. Dus dat is wel een vereiste.

Meneer Janssens-Baan: ja en natuurlijk het up-to-date houden van het model. Als het systeem verandert, of we gaan nog meer bouwen, dan heeft dat invloed op het model.

Meneer Jansen: ja als we gaan ingrijpen op de bestaande structuur, dan heeft dat natuurlijk gevolgen.

Meneer Janssens-Baan: ja zeker als er structurele veranderingen zijn. Bijvoorbeeld een nieuw riool bij een nieuwe woonwijk, of ander grondgebruik. Dat scheelt ook heel veel. Dat is wel belangrijk om dan mee te nemen. Ik ben dan ook echt benieuwd naar de resultaten van die neural networks. Dat het zo snel en zo nauwkeurig kan rekenen.

Laura: ja daar gaat mijn onderzoek dus onder andere over. Ik ga onderzoeken of zo'n neural network misschien kan bieden wat jullie zoeken.

Meneer Janssens-Baan: ja wij zijn ook benieuwd naar de resultaten.

Minutes interview Hoogheemraadschap van Rijnland

21-02-2023

Jan Jelle Reitsma– adviseur waterkwantiteit bij Hoogheemraadschap van Rijnland

René van der Zwan – adviseur watermanagement bij Hoogheemraadschap van Rijnland

---

Laura: gebruiken jullie het W2O model op dit moment al actief bij jullie dagelijkse werkzaamheden?

René: nee, eigenlijk niet. En als we het gebruiken, dan is het te weinig. We zouden het vaker moeten gebruiken maar meestal gebruiken we nu de waarschuwingen die we vanuit het KNMI krijgen of vanuit onze eigen weerprovider. Van onze eigen weerprovider DTN krijgen we dan een mailbericht en de dienstdoende peilbeheerder wordt opgeroepen door de meteoroloog via zijn semafoon. De peilbeheerder gaat dan bellen met DTN en dan krijgt hij nadere duiding van de situatie.

Jan Jelle: en even voor mijn beeldvorming, wat is nu de uitvoer van het W2O model?

Laura: het W2O model berekent een kans op wateroverlast. Dat is dan gedefinieerd als meer dan x% kans op meer dan 10, 20 of 30mm water op het maaiveld. De gebruiker bepaalt zelf bij welke kans hij of zij een alert per mail wil ontvangen.

Jan Jelle: en dat is dus per peilvak, met een bepaalde bodemberging en hoeveelheid oppervlaktewater?

Ludo Diender: nee, dat zit er op dit moment nog niet in. Alleen de bodemberging wordt nu meegenomen in het algoritme. We zijn nu aan het onderzoeken hoe we de interactie met het openwatersysteem kunnen meenemen in het W2O algoritme. Dat zijn de volgende stappen om toe te voegen aan het algoritme.

Laura: als ook die component van het openwater is toegevoegd aan het W2O algoritme, dan is de vraag of het model daarna wel voldoende handelingsperspectieven biedt. Of past een hydraulisch model misschien beter?

Jan Jelle: qua handelingsperspectieven kunnen wij twee dingen doen in een gebied. We kunnen van tevoren gaan voormalen om al extra berging te creëren. Daarnaast kunnen we op het moment zelf nog Tijdelijke Pomp Installaties (TPI) erbij plaatsen. Dat zijn de twee dingen die we in de meeste gevallen kunnen doen. In een aantal polders hebben we automatische stuwen waar we misschien nog iets mee zouden kunnen. In de meeste gevallen gaat het een polder met een gemaal, en dat is je knop.

René: anticiperen is natuurlijk altijd een hele goede maatregel. Als we eerder die bui zien aankomen en ook de zekerheid hebben dat de bui gaat vallen, dan kunnen we daar wat mee. In de zomer zijn de peilen vaak opgezet om zoetwater te bufferen. Als we dan gaan anticiperen door het zoetwater weg te pompen en de bui valt niet of ergens anders, dan creëren we ons eigen probleem voor de rest van de zomer, omdat dan de zoetwatervoorraad afgevoerd is en het de vraag is of het daarna nog lukt om die voorraad weer aan te vullen. Maar bij bijvoorbeeld de bui die in 2021 boven Limburg viel, konden we twee dagen van tevoren al zien dat er forse hoeveelheden neerslag zouden gaan vallen. Als die bui bij ons gevallen zou zijn, dan hadden we daar best op kunnen anticiperen. Al het water dat we van tevoren wegpompen, dat voorkom je uiteindelijk aan peilstijging aan de bovenkant. Het ergste wat kan gebeuren in een polder maar ook voor ons boezemsysteem is dat we naar een maalstop gaan. Als we het maalstoppeil bereiken, dan moeten we stoppen met malen. Dan moet het poldergemaal uit, of in een nog slechter geval het boezemgemaal (bij hoge waterstanden op Noordzeekanaal of afgesloten Hollandsche IJssel. Dan weten we dat de polders sowieso vollopen omdat er neerslag blijft vallen maar we het water niet meer kunnen wegpompen.

De maalstop voorkomt dat het water over de kades gaat en de kade gaat eroderen. In het maalstoppeil zit nog een waakhoogte van 25cm verwerkt voor o.a. windopzet. Als de kade doorbreekt, dan loopt de polder ongecontroleerd vol. Dat kan dan hele negatieve effecten hebben voor de rest van het systeem. Het stoppen met malen is nodig als de waterstanden te hoog worden in het hoofdsysteem. We hebben hier wel over extreme situaties. Ons watersysteem kan een bui weerstaan met een kans van voorkomen van eens in de honderd jaar, we hebben dit tot nu toe nog niet meegemaakt. De Limburgbui zat in de orde grootte van eens in de duizend jaar.

Jan Jelle: maar een hevige regenbui binnen een polder maken we wel regelmatig mee. Eens per een of twee jaar valt er wel ergens in ons gebied een extreme bui.

René: wij willen dan graag zo vroeg mogelijk weten of die bui eraan komt. Je zou dan met kansen kunnen werken natuurlijk. En als de bui valt, wat betekent dat dan voor de waterstanden in elk gebied. We willen dan graag weten in welk gebied water staat, dat is waar we geïnteresseerd in zijn.

Laura: natuurlijk wil je het liefst zo vroeg mogelijk weten dat een bui eraan komt. Maar welk tijdsbestek is voor jullie de grens om nog maatregelen uit te kunnen voeren?

Jan Jelle: als we het pas een paar uur van tevoren weten, dan heeft voormalen niet zo heel veel nut meer. Als de voorspelling heel extreem is en je weet dat het alleen in een specifiek gebied is, dan zouden we al noodpompen kunnen neerzetten. Dan zouden we dat vast kunnen gaan organiseren. Als je wilt gaan voormalen dan wil je dat eigenlijk wel 48 tot 24 uur van tevoren weten. Een dag voormalen helpt ook al wel wat, alles helpt een beetje natuurlijk.

René: het hangt ook van je uitgangssituatie af natuurlijk. Of het droog is of nat, dat maakt wel uit in hoeveel effect het heeft.

Jan Jelle: die uitgangssituatie zoals beschikbare bodemberging kun je meenemen in je berekening vanuit Owasis. Als we weten dat het kritisch gaat worden, dan kunnen we wat doen. Hoe eerder, hoe effectiever. Maar als je onze rol vergelijkt met de handelingsperspectieven van een gemeente, dan heeft de gemeente misschien niet zoveel knoppen waar ze aan kunnen draaien, maar ze kunnen wel mensen waarschuwen. Dat is dan bij ons wat minder relevant, wij hebben andere handelingsperspectieven.

René: en bij hele ernstige situaties, dan staan wij niet meer aan het roer. Wij zijn dan adviseur richting de veiligheidsregio's. Die zullen onze adviezen waarschijnlijk wel overnemen. Maar als het bijvoorbeeld gaat om mensen evacueren, dan is dat echt aan de veiligheidsregio's. Maar dat geldt ook voor gemeentes.

Jan Jelle: ik denk dat we sowieso niet snel zouden overgaan op evacueren. We zouden dan een paar miljoen mensen moeten evacueren, ik zou niet weten waar die naartoe moeten. En bij wateroverlast met 10cm water op straat, daar gaan we niet voor evacueren. Tenzij er echt een kade doorbreekt, dan misschien wel. Maar dat is dan weer niet te voorspellen met dit soort modellen.

Laura: nog even terugkomend op mijn eerste vraag. Jullie gebruiken het W2O model dus niet dagelijks. Zijn er andere modellen of tools die jullie wel actief gebruiken in de operationele context?

Jan Jelle: wij gebruiken het Beslis- Ondersteunend Systeem BOSBO, dat is dan meer beredeneert vanuit de boezem. Daar zitten de poldermodellen in, maar qua uitvoer wordt daar vooral gekeken naar de boezems. In die modellen wordt op basis van de neerslagvoorspelling berekend wat de voorspelde boezempeilen zijn (m.b.v. SOBEK en een RTC model). Het BOS bepaalt dan welke gemalen we moeten gaan inzetten op welk moment. Maar bij dit model wordt er niet specifiek naar de polders gekeken.

René: bij een aantal polders zijn nu we nu bezig met machine-learning modellen en daar wordt wel de verwachting van de waterstanden in de polders berekend.

Jan Jelle: Ja we hebben verder ook niet heel veel handelingsperspectieven. We kunnen voormalen op basis van de neerslagvoorspelling. Maar behalve malen bij neerslag kunnen we verder niet veel. Het BOS geeft als output of we gaan malen en op welke locaties. Dit zou je als ondersteuning kunnen gebruiken om te bepalen waar je gaat voormalen.

René: Maar op dit moment wordt BOS niet actief gebruikt voor de polders om te bepalen of en waar we gaan voormalen. Dat gebeurt op basis van de informatie van de weerprovider. Als de weerprovider een bui voorspelt die meer is dan we kunnen wegpompen, dan gaan we kijken hoe we dat trapsgewijs kunnen wegmalen. Als je alle poldergemalen tegelijk aanzet, dan gebeurt er niet zoveel in de boezem. Daar wil je ook ruimte creëren. Daarom gaan we eerst kijken welke polders het meest kwetsbaar/gevoelig voor wateroverlast. We handelen nu dus vooral op basis van de weersvoorspellingen.

Jan Jelle: voor de polders handelen we nu dus vooral op basis van de weersvoorspellingen. En bij een aantal polders gebruiken we dus machine-learning modellen waar ook het voormalen in verwerkt zit. Het BOS wordt vooral gebruikt om de boezemgemalen aan te sturen, het hoofdsysteem dus.

Laura: en welke informatie missen jullie dan in deze modellen? Hoe zou een model zoals het W2O model van toegevoegde waarde kunnen zijn?

René: in W2O zie je op de kaart geplot waar de wateroverlast optreedt. Dat zit hier niet in.

Jan Jelle: en in W2O zit vanuit Owasis bijvoorbeeld de bodemberging verwerkt. Daar zitten de metingen in, waardoor je een beter beeld hebt van de werkelijke toestand. In Delft-FEWS zit een inregeltijd en het model zelf is ook best grof: een hele polder is één bakje. In W2O zou zo'n grid kleiner kunnen.

Laura: ja dat klopt. De neerslagvoorspelling is op een grid van 1x1km, dus dat is ook de schaal waarop W2O nu is. En vanuit Owasis komen de gegevens op 250x250m. Het is wel gebleken dat de rekentijd aanzienlijk toeneemt als W2O op die schaal wordt gedraaid. Dat komt onder andere door de ensemble members die het model doorrekent.

René: ja je zou nog eens kritisch kunnen kijken naar die ensemble members. In plaats van 30 ensembles kun je ook een realistic, worst-case en best-case scenario doorrekenen. Dat scheelt in de rekentijd.

Laura: ja dat is een keuze die afhankelijk is van de wens van het waterschap of de gemeente. Liever 50 ensembles en een minder gedetailleerd conceptueel model, of slechts 3 ensembles en een geavanceerder hydraulisch model bijvoorbeeld.

Jan Jelle: ja dan is de vraag wat meer detail in je model uiteindelijk meer gaat opleveren aan informatie. Vergeleken met hoe het nu is, zal het sowieso beter zijn. Maar als je het W2O model gaat uitbreiden met bijvoorbeeld oppervlaktewater, hoe groot is dan nog de winst?

Laura: en die extra informatie en de vereiste nauwkeurigheid daarvan, hangt ook weer samen met de handelingsperspectieven. Als jullie besluiten te gaan voormalen om wat meer berging te creëren en de neerslagverwachting blijkt niet te kloppen, dan heeft dat wel gevolgen. Dan is het dus wel belangrijk dat de voorspelling nauwkeurig is.

Jan Jelle: ja, maar dat zou je ook kunnen koppelen aan die ensembles. Dat je bijvoorbeeld gaat voormalen zodra minimaal 20 van de 30 ensembles regen voorspellen. Voormalen kan verder niet heel veel kwaad als er wel sowieso regen valt. Dan wordt het wel weer aangevuld. Voormalen zorgt alleen voor problemen als het een hele droge periode is en er helemaal geen regen valt.

Laura: maar die piekbuien in de zomer zijn wel lastig te voorspellen. Die komen soms in één keer opzetten en het zou dan kunnen dat ze de bui net een paar kilometer verderop valt.

Jan Jelle: maar als de bui wat verderop valt, dan is dat ook niet direct een probleem. Al onze polders zijn namelijk aangesloten op de boezem. Als de bui valt in een naastgelegen polder, dan kunnen we dat uitmalen op de boezem en vanuit daar weer inlaten in deze polder. Dat is anders in hogere delen in Nederland waar je moeilijk water kunt krijgen. In al onze polders kunnen we in principe water krijgen omdat ze lager liggen dan de boezem. Het is eerder de vraag of we op dat moment voldoende zoetwater beschikbaar hebben. Dat is 90-95% van de tijd wel het geval maar in uitzonderlijke situaties moeten we zoetwater aanvoeren vanuit het oosten. Dan moeten we wel zuinig zijn op het water. Maar dat zou je ook op dat moment nog kunnen bepalen. Dan kun je iets kritischer zijn met voormalen in een KWA periode.

René: en specifiek voor de polder van jouw studie, daar is heel veel onderbemaling. De vraag is hoe die vertaald wordt in een model zoals het W2O model. Onderbemalingen zijn peilvakken door een particulier beheerd met een eigen pomp. Die hanteert vaak een lager peil dan het reguliere peil. Als waterschap hebben we de eis dat de pompen van de particulieren in verhouding niet groter mogen zijn dan onze eindpomp. Dat zorgt ervoor dat ze niet afwentelen op het hoofdvak. En als er een maalstop is in de polder, dan moeten die onderbemalingen ook stoppen met pompen want anders wentelen ze het weer af op het grotere vak.

Laura: en heeft deze onderbemaling veel invloed op het grotere systeem?

Jan Jelle: in een polder als deze wel, want er zijn heel veel onderbemalingen. Als die allemaal een hele grote pomp neerzetten en dat gaan pompen op het hoofdvak als het heel hard regent, dan verzuipt het hoofdvak. En als je dan daar als boer zit, dan heb je pech omdat jij afhankelijk bent van ons gemaal dat minder groot is dan de gemalen van de onderbemalingen.

Laura: als ik een model ga maken in D-HYDRO, en ik neem deze onderbemalingen daarin mee. Stel dat de particulieren dan niet de richtlijnen volgen die jullie hanteren, dan klopt het model en dus de voorspelling niet. Daar ben je dan dus wel van afhankelijk.

Jan Jelle: ja dat klopt. Dan zou je dit als een soort extra onzekerheid/ensemble mee kunnen nemen. Ik vraag me wel af in hoeverre een D-HYDRO model heel veel gaat toevoegen ten opzichte van een redelijke simpele bakjes benadering. Uiteindelijk gaat het toch om de hoeveelheid neerslag die valt en die je kwijt wil in je gebied. Voor onze handelingsperspectieven is het vooral relevant om te weten bij welk peil je een probleem hebt en er schade ontstaat. Wanneer zeg je: dit moet voorkomen worden met het voormalen? Je wil graag weten of je dat peil gaat bereiken. Zolang de openwater component niet in het W2O model zit, dan zal het D-HYDRO model zeker wat toevoegen. Als dat er niet inzit, dan overschat je heel erg het probleem.

Laura: en als er wel een openwater component in het W2O model zit, dan is het eerst belangrijk om te kijken op welke manier dat in het algoritme verwerkt zit. Of dat realistisch genoeg is om resultaten te krijgen die nauwkeurig genoeg zijn? Of is het zo versimpeld dat het nog steeds voor een significant verschil zorgt vergeleken met de daadwerkelijke situatie? Daarbij hoort dan de vraag voor jullie: hoe nauwkeurig moet de voorspelling voor jullie handelingsperspectieven zijn?

Ludo: ja we zijn nu dus bezig om die openwater component toe te voegen aan het W2O model. Het concept is uitgedacht en het zal een hele simpele benadering worden. Het gaat om een bakje voor het open water en voor de bodemberging. Als deze allebei vol zitten, dan zal het water op het maaiveld staan. Het 2D stromen van het water vanuit de watergang het land op en andersom, dat zit dus niet direct verwerkt in het W2O model. Die specifieke interactie zit niet in het W2O model en zal er ook niet in komen. Dat zit dan wel in het D-HYDRO model.

Jan Jelle: die RR koppeling wordt dan dus een belangrijke toevoeging in het D-HYDRO model. Dat 2D model kun je redelijk makkelijk geautomatiseerd opzetten. En in deze polder zit natuurlijk ook nog een klein stukje met stedelijk gebied. Dat kun je zo gedetailleerd modelleren als je zelf wil.

Laura: ik ga dus beginnen met een gedetailleerd D-HYDRO model dat gaat fungeren als een benchmark. Vervolgens ga ik aanpassingen doen aan dat gedetailleerde model zodat het model sneller gaat runnen en bruikbaar wordt voor de operationele setting. Door die aanpassingen zal wel de nauwkeurigheid van de resultaten afnemen. Hoe nauwkeurig moet de output van een model zijn zodat jullie daar wat mee kunnen voor de handelingsperspectieven?

Jan Jelle: als waterschap handelen wij uiteindelijk op basis van een peil. Voor ons gaat de voorspelling dus om het maximale peil dat je bereikt, en of dat hoger of lager is dan het peil waarbij we schade krijgen. Dat is voor ons de vraag. Daar zal dan een soort marge omheen zitten waarbij je 5 tot 10 cm onder dat peil moet blijven. Als het boven die marge komt, dan willen we sowieso voormalen.

Laura: oké duidelijk. Zoals ik net vertelde, ga ik het gedetailleerde D-HYDRO model gebruiken als een benchmark. Stel dat ik daar aanpassingen aan ga doen om de rekentijd te verkorten, maar de resultaten van het surrogaat model wijken dan bijvoorbeeld 15 of 20 cm af van de benchmark. Dan is dat voor jullie dus niet nauwkeurig genoeg?

Jan Jelle: ja we willen met die nauwkeurigheid wel binnen 10cm blijven. Ons hele peilbeheer is centimeter werk. Dus 5 tot 10cm is wel echt de grens daarin, dat wil je wel voor elkaar krijgen. In sommige gebieden hebben we wel extremere peilstijgingen, bijvoorbeeld in een droogmakerij.

René: we hadden het net al even over het stedelijk gebied in de polder waar jij mee aan de slag gaat. Het stedelijk gebied ligt een stuk hoger, dus het is belangrijk om dat wel mee te nemen in je model omdat het stedelijk gebied via een stuw afwatert op de rest van de polder.

Jan Jelle: en die stuw is ook geautomatiseerd dus daar kunnen we mee sturen. Daar hebben we ook bedieningsinformatie over. In de praktijk zullen we de stuw omlaag zetten als het hard regent, om zoveel mogelijk het stedelijk gebied te ontzien.

René: ja dat past in het schade-reductie verhaal. De schade bij wateroverlast in het stedelijk gebied is vele malen groter dan in de polder.

Jan Jelle: ja die geautomatiseerde stuw is hier dus wel een extra sturingsknop. We kunnen niet alleen het gemaal bedienen maar ook de stuw. Maar waarschijnlijk gaan we dat pas doen als het echt uit de hand loopt. Je kunt met die stuw dus ook 'voormalen'. Als je van tevoren de stuw alvast wat omlaag zet, dan ga je ook in het stedelijk gebied alvast wat lager zitten.

René: en in het stedelijk gebied is waarschijnlijk niet zoveel openwater, dus elke druppel leidt al snel tot een peilstijging.

Laura: ik ga dus aanpassingen doen aan het gedetailleerde D-HYDRO model om het bruikbaar te maken in de operationele context. Aan welke rekentijden moet ik dan denken om het bruikbaar te maken? Hoe vaak willen jullie een update van de voorspelling? De neerslagvoorspelling wordt elke 5 minuten geupdate, dus dat is niet een limiterende factor daarin.

Jan Jelle: als je alle neerslaginformatie wil gebruiken, dan heb je dus een model nodig dat snel rekent.

René: maar als het model elke paar minuten met een nieuwe voorspelling komt, dan betekent dat ook dat je elke paar minuten moet kijken. De vraag is ook hoe reëel dat is. Ik denk dat eens per halfuur of uur een mooi streven is. Het BOS hoofdmodel draait ook eens per uur, en die rekentijd is ongeveer 10 tot 15 minuten. En met de neerslag nowcast kijk je de komende 2 uur vooruit. Is het pragmatisch voor jou om eens per uur aan te houden? Dan kan je later nog optimaliseren en het frequenter te laten draaien.

Laura: ja uiteindelijk is het een balans tussen de rekentijd en de nauwkeurigheid van de resultaten. Ik kan het model heel erg versimpelen en heel veel aannames doen waardoor het heel snel rekent, maar dan zijn de resultaten misschien niet meer accuraat genoeg.

René: ja precies, duidelijk. De nowcast kijkt de komende twee uur vooruit, dus dan lijkt me 4 keer daarbinnen, dus elk halfuur een mooi streven. Dat is wel acceptabel.

Jan Jelle: eens. Dat betekent dus dat je rekentijd maximaal een half uur is en dat je binnen een half uur alle ensembles doorrekent. Dat zijn 30 ensemble members, of minder als je daar een keuze in maakt.

Laura: en nog even terugkomend op die 5 tot 10cm die de peilen maximaal mogen afwijken van de benchmark. Over welke locaties hebben we het dan? Ik neem aan dat de peilen in ieder geval worden gemeten bij de gemalen. Is dat ook de plek waar het surrogaat maximaal 5 tot 10cm mag afwijken van de benchmark?

Jan Jelle: nee ik denk juist ergens midden in de polder. Die locatie wijs je dan aan in je benchmark model en daar ga je het mee vergelijken. In deze polder heb je meerdere peilvakken dus je zou er dan ook nog een referentiepunt kunnen plaatsen in het stedelijk gebied. Als je dan gaat versimpelen, dan mag het surrogaat model niet teveel afwijken van wat je daar berekend hebt met het D-HYDRO model.

René: ja in de buurt van gemalen is juist de minst goede plek om waterstanden te meten. Die metingen worden gebruikt als procesinformatie voor het gemaal, maar die zijn niet representatief voor de rest van de polder. In de buurt van de gemalen worden veel lagere waterstanden gemeten dan in de rest van het gebied. Maar in deze polder zitten er wel meer meetpunten verspreid, ook niet in de buurt van gemalen. Dat is in kader van het polderlab, daardoor wordt op meer plekken de waterstand gemeten. Die kan je in ieder geval gebruiken voor de kalibratie van je model.

Jan Jelle: je zou in je gedetailleerde model kunnen kijken of er bij een maatgevende/extreme bui veel grote peilverschillen optreden binnen de polder. Als je dat ziet, dan wil je die verschillen ook graag terugzien in de versimpelde versie van je model. Als blijkt dat de polder als een grote bak water werkt zonder grote verschillen, dan is dat wat makkelijker om het model te vereenvoudigen.

Jan Jelle: die RR koppeling in je model is natuurlijk wel heel belangrijk in deze studie. Heb je al bedacht hoe je het stedelijk gebied wil meenemen in je benchmark model? Ik neem aan dat je geen rioolstelsel gaat meenemen, maar bijvoorbeeld wel de verhardingsgraad. Die afvoer van de verharding moet op een manier verwerkt zitten in het model. Je hebt natuurlijk ook de berging op straat. Mijn ervaring is dat je snel overschattingen krijgt waarbij je enorm snel afvoer krijgt richting je oppervlaktewater. In werkelijkheid is dat wat minder. Dat komt door je aannames die je doet over wat verhard is en wat niet. Als je veel stedelijk gebied hebt, dan is dat wel bepalend.

Laura: hebben jullie misschien al ideeën voor aanpassingen die ik aan het model kan doen om het sneller te laten runnen? We hadden het net al over bepaalde watergangen niet meenemen. Jullie kennen het gebied, dus misschien hebben jullie hier al over nagedacht?

René: het gebied is herverkaveld waardoor er nu best veel grote watergangen liggen. Die hebben dus best wel veel invloed en kun je niet zomaar weglaten.

Jan Jelle: je zou alleen de primaire watergangen mee kunnen nemen. Voor de overige watergangen neem je dan alleen de berging mee. Dat scheelt waarschijnlijk al behoorlijk. Je kunt er ook voor kiezen om de onderbemaling er wel of niet in te zetten. Als je aanneemt dat de onderbemalingen net zo reageren als de rest van het gebied, dan kun je de onderbemalingen negeren. Je kunt kijken of dat veel uitmaakt. Uiteindelijk gaan die onderbemalingen via het hoofdgemaal op het hoofdsysteem. Als het allemaal klopt, dan zouden al die onderbemalingen in verhouding moeten zijn met het hoofdgemaal. Dan zou het qua waterbergingsverdeling op dezelfde manier reageren.

René: je kan bijvoorbeeld de kleine slootjes weglaten. Maar je moet dat volume dan wel ergens als een bak ergens maken. Voor je berging is het namelijk wel van belang.

Jan Jelle: je kan bijvoorbeeld ergens in het midden van je model een bergingsknoop maken waarin je al het overige water stopt. Je kunt kijken of dat nog veel uitmaakt.

Laura: ik denk inderdaad wel die richting op. Om het model snel te laten runnen, zijn dat soort aanpassingen nodig.

Jan Jelle: je zou alle onderbemalingen ook nog als één onderbemaling kunnen beschouwen. Dan ga je ze aggregeren en dan neem je het oppervlakte van alle onderbemalingen als een gebied. Dan heb je dus geen koppeling meer met hoe het fysiek zit. Op die manier heb je nog wel de mogelijkheid om te kijken wat er gebeurt als blijkt dat de onderbemalingen in werkelijkheid een veel grotere capaciteit hebben. Dan kun je kijken of dat nog veel uitmaakt voor je uitkomsten.

René: in deze GIS-kaart kun je al wel zien dat er veel niet-primaire watergangen zijn. Als je al die watergangen niet meeneemt, dan zal je model waarschijnlijk niet zo representatief meer zijn. Je zou daar slimme aannames in kunnen doen.

Jan Jelle: je zou ook nog kunnen kijken naar je 1D2D-koppeling, ik weet niet voor hoeveel extra rekentijd dat zorgt. Waar zit de rekentijd vooral in? Waar valt het meest halen? Bijvoorbeeld kleine lijnelementen zullen je rekentijd wel beïnvloeden. Die kun je er vaak vrij makkelijk uithalen zonder dat dat veel invloed heeft op de uitkomsten.

Laura: stel ik ben uiteindelijk in staat om voor dit gebied een model te maken dat snel genoeg runt en waarbij ook de modeluitkomsten nauwkeurig genoeg zijn. Dat model is dan hopelijk bruikbaar voor jullie. Maar als jullie dit voor jullie hele beheersgebied willen doen, dan vraagt dat om een andere aanpak. Dan heb je bijvoorbeeld meerdere systemen nodig die tegelijk runnen. Hebben jullie daar over nagedacht?

René: ja dat parallel draaien zal dan nodig zijn.

Jan Jelle: je kunt dan ook de keus maken om alleen naar de kwetsbare polders te kijken. Een aantal polders zijn minder kritisch, dus die doen we dan niet. Daar wonen minder mensen en zal er minder schade zijn in geval van hevige regen. Die polders hebben ook minder bemalingscapaciteit waardoor we sowieso minder handelingsperspectieven hebben.

René: de Vierambachtspolder heeft akkerbouw, dus dat is wel wat kwetsbaarder. En stedelijk gebied natuurlijk. Dat wil je ten alle tijden drooghouden. Daar zitten ook wat spanningen met de akkerbouwers omdat zij het gevoel hebben dat ze 'de sigaar zijn' door het stedelijk gebied dat voorrang krijgt. Daarom willen we problemen graag zoveel mogelijk voorkomen door bijvoorbeeld voor te malen.

Laura: dat parallel rekenen is dus wel een oplossing voor jullie?

René: ja dat zal dan wel moeten. Als we elk halfuur een update willen, dan is dat de enige mogelijkheid. Als we in een crisissetting zitten, dan willen we graag regelmatig updates. Dan ontkom je er dus niet aan om het parallel te draaien.

Laura: of een conclusie zou dan ook kunnen zijn dat het generieke W2O model misschien een beter passendere oplossing is?

René: ja dat zou ook kunnen. Maar voor die paar kwetsbare polders willen we denk ik toch wel graag die detail informatie hebben.

Laura: en welke detailinformatie is dan voor jullie vooral van belang?

Jan Jelle: de peilen in de watergangen. Dat kunnen wij namelijk sturen en waar wij ook verantwoordelijk voor zijn. Voor ons is het minder van belang waar en hoeveel het precies op het maaiveld staat. Als het water op het maaiveld blijft staan, daar gaan wij niet over en daar kunnen we ook niks aan doen.

René: maar als de watergangen vol zitten en het water stroomt vanuit daar op het land, dan krijgen we wel de vraag waar het water op het land komt (en welk percentage van het land).

Jan Jelle: maar dat bereken je dan met je 2D component.

Laura: klopt. Daar ga ik de komende weken mee aan de slag. Waarschijnlijk heb ik nog wel meer vragen als ik eenmaal bezig ben met het operationaliseren. Dan zal ik contact met jullie opnemen.

C Implications timestep rainfall forecast

The rainfall forecast has an hourly timestep, which results in no variation of rainfall intensity within an hour. In reality, there will likely be a variation in the rainfall intensity within an hour. To simulate the effect of the hourly time interval of the rainfall forecast, the observed rainfall timeseries was averaged over the hour, as depicted in Figure 51 in red. This yields the same amount of total rainfall within an hour, without the variation in the rainfall intensity.

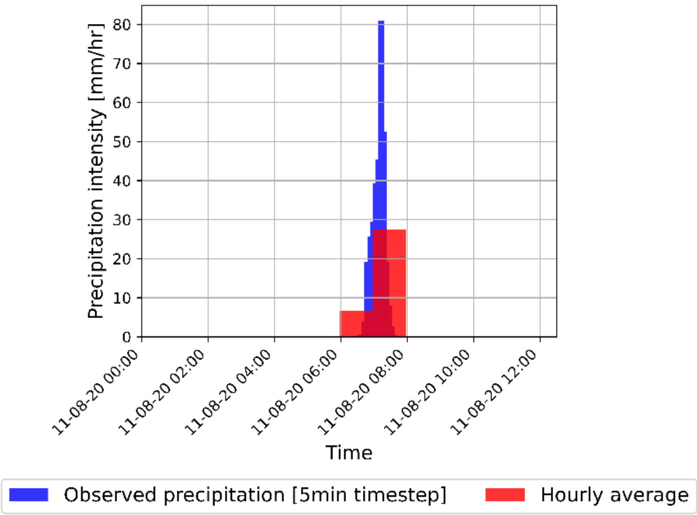


Figure 51 - Observed rainfall versus hourly averaged rainfall on 11-08-2020

Figure 52 shows the difference in flood volumes when using the hourly average of the rainfall time series (red in Figure 51) versus using the rainfall timeseries with a 5 minute timestep (blue in Figure 51). The maximum flood volume over time is computed for each manhole. As can be seen in Figure 52, when using the hourly average of the rainfall timeseries, the flood volumes are underestimated up to $16m^3$. Using rainfall timeseries with a timestep of 1 hour thus has large consequences for the accuracy of the ML model output.

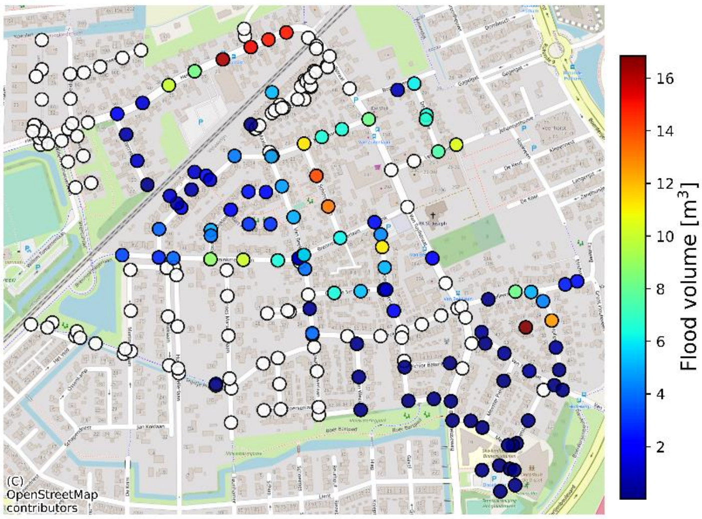


Figure 52 - Difference ML model output with hourly averaged observed rainfall versus observed timeseries with 5-minute timestep (difference = hourly average – 5 minute timestep). White dots indicate manholes that are not flooded in any of the runs.

D Alternative machine learning model set-ups

As explained in Section 3.4.1.4, different alternative ML model set-ups are tested for the second case study. The results of the final model set-up are presented in Section 3.4.2. This section shows the results of the other model set-ups that have been tested. These alternatives do not perform as good as the final model set-up and are therefore not considered in the main report. However, to give a complete overview, these results are presented here.

D1 Alternative 1 – LSTM model by Hop (2023)

First of all, the model set-up that is found in the study by Hop (2023) is applied. This LSTM model has proven to be successful in accurately reproducing a hydrological inundation model for a rural area in the Netherlands. The input of the LSTM model are rainfall timeseries and the output is the timeseries of the water depth for each grid cell in the model domain. An overview of the model architecture found by Hop (2023) is presented in Table 27. This model architecture is adapted such that it can be applied on the dataset for the case study in this thesis. This implies that the number of grid cells in the dense layer is adjusted. Besides, the dataset in this case study does not has water depth timeseries as output (only maximum) and therefore second LSTM layer should not return a sequence.

Table 27 - Model set-up of LSTM model found by Hop (2023)

LSTM layer 1	256 neurons
Dropout after layer 1	0.02
LSTM layer 2	256
Dropout after layer 2	0.2
LSTM layer activation function	Hyperbolic tangent
Learning rate	0.001
Dense layer	28114 (number of grid cells)
Dense layer activation function	relu

The performance indicators on the validation dataset when using the model set-up found by Hop (2023) are presented in Table 28. Figure 53 shows the scatter plots of the prediction of the machine learning model versus the hydrological model on the training dataset (Figure 53a) and on the validation dataset (Figure 53b). As can be seen in these plots, certain patterns (lines) can be recognized when looking at the scatters. These patterns are caused by the model: it always predicts the mean water depth in that grid cell. This is confirmed by looking at the model prediction for a certain (random) grid cell: independent of the rainfall timeseries, always the mean water depth is predicted, see Figure 54. This relatively bad prediction is also reflected in the model performance indicators, see Table 28.

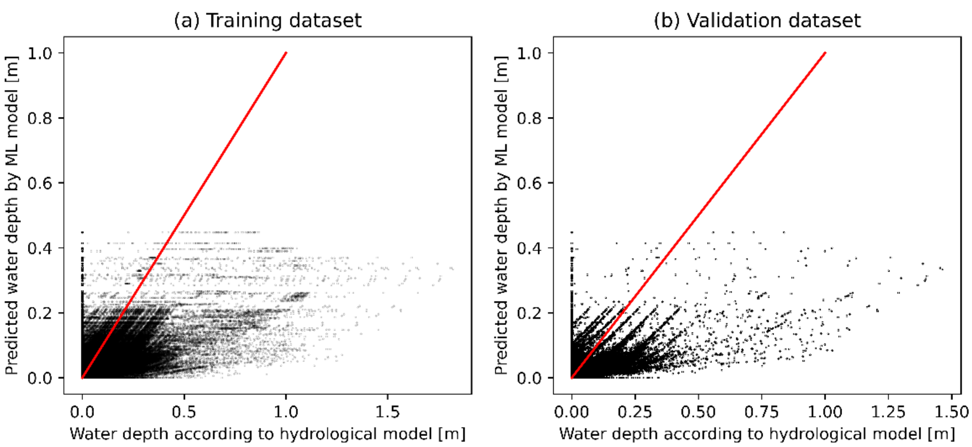


Figure 53 - Results of LSTM model with set-up found by Hop (2023)

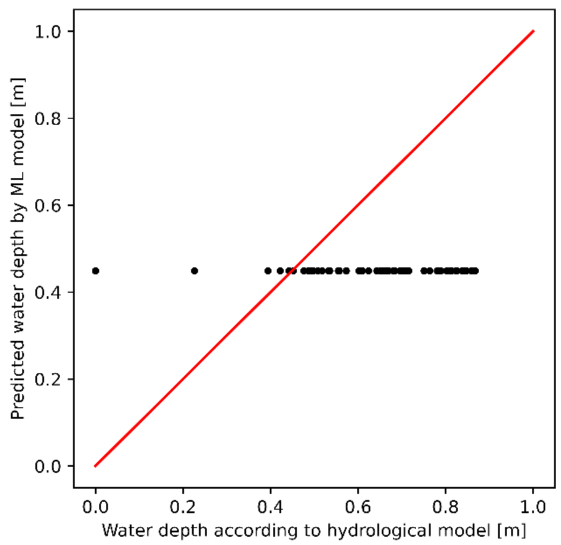


Figure 54 - Predicted water depth according to ML model and hydrological inundation model for a random grid cell. This scatter plot shows that the ML model always predicts the same value (the mean), independent of the input variable.

Table 28 - Performance of LSTM model (Hop, 2023) on validation dataset

	Score on validation dataset
MAE	0.02m
RMSE	0.05m
CSI	21%

D2 Alternative 2 – LSTM model by Kilsdonk (2021)

Secondly, the LSTM model set-up found by Kilsdonk (2021) is tested. The used model architecture is presented in Table 29. Similarly to the LSTM model by Hop (2023), the model architecture is adjusted such that it can be applied on the dataset for this case study. This implies that the number of neurons in the dense layer is adjusted to the number of grid cells of this case study, and the LSTM layer does not return a sequence.

Table 29 - Model set-up of LSTM model found by Kilsdonk (2021)

LSTM layer 1	230 neurons
Dropout after layer 1	0.2
LSTM layer activation function	Hyperbolic tangent
Learning rate	0.01
Dense layer	28114 (number of grid cells)

The performance indicators on the validation dataset when using the model set-up found by Kilsdonk (2021) are presented in Table 30. Figure 55 shows the scatter plots of the prediction of the machine learning model versus the hydrological model on the training dataset (Figure 55a) and on the validation dataset (Figure 55b). As can be seen in these plots, also here certain patterns (lines) can be recognized when looking at the scatters. Similarly to using the model set-up found by Hop (2023), the ML model always predicts the mean water depth.

Table 30 - Performance of LSTM model (Kilsdonk, 2021) on validation dataset

	Score on validation dataset
MAE	0.01m
RMSE	0.05m
CSI	16%

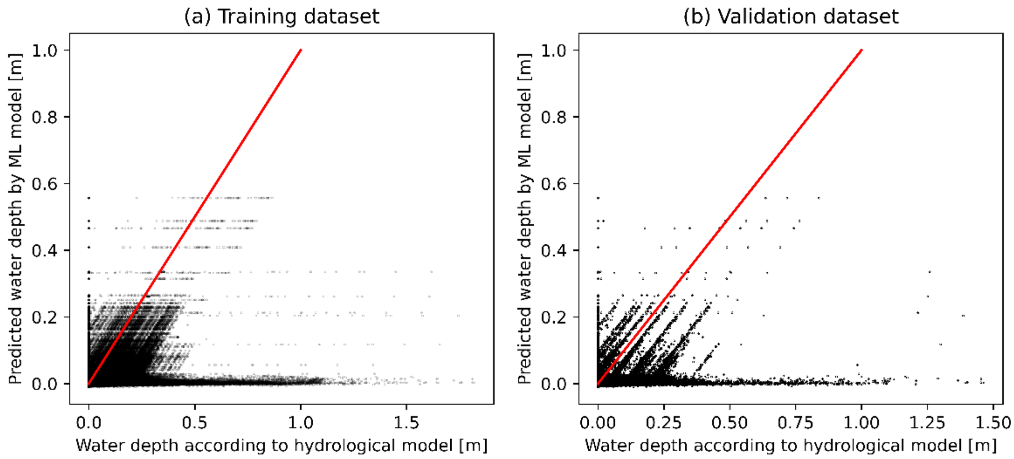


Figure 55 - Results of LSTM model with set-up found by Kilsdonk (2021)

D3 Alternative 3 – MLP model by Berkhahn et al. (2019)

Since it turned out that LSTM models are not a suitable Machine Learning model type for this specific problem (no timeseries), Multi-Layer Perceptron (MLP) models are tested as an alternative. The MLP model found by Berkhahn et al. (2019), has shown to successfully reproduce the output of a hydrological inundation model for two different case studies. Using rainfall timeseries as input, this model is able to accurately predict maximum water depths in the study by Berkhahn (2019). Since this model set-up has proven to be successful, it is decided to apply the same model architecture to the dataset of this case study. Figure 56 shows the scatter plots of the prediction of the machine learning model versus the hydrological model on the training dataset (Figure 56a) and on the validation dataset (Figure 56b). As can be seen in the plots, and as confirmed by the performance indicators in Table 31, this MLP model architecture is not able to accurately reproduce the output of the hydrological inundation model.

Table 31 - Performance of MLP model (Berkhahn, 2019) on validation dataset

	Score on validation dataset
MAE	0.02m
RMSE	0.03m
CSI	41%

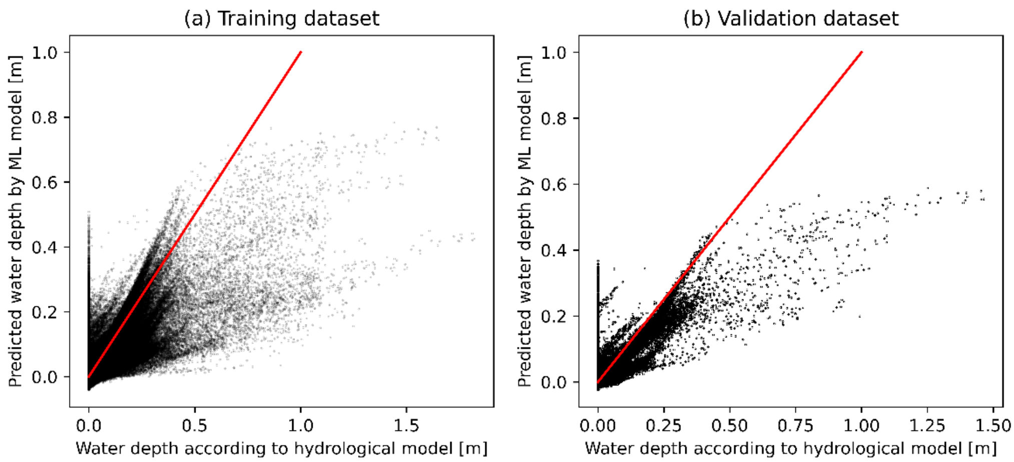


Figure 56 - Results MLP model with set-up found by Berkhahn (2019)

D4 Alternative 4 – MLP model found using Random Search

The fourth alternative model set-up is using a simple Multilayer Perceptron (MLP) model. The hyperparameters are found using a random search. The ranges used for the random search are presented in Table 32. This results in a best functioning model containing 2 hidden layers, both with 160 nodes. The activation function is 'relu' and the learning rate is set to 'adaptive' with an initial value of 0.0055. An overview of the final model set-up is also presented in Table 32.

Table 32 - Model architecture for MLP model Random Search

	Range for Random Search of hyper parameters	Final model architecture after parameter optimisation
Number of layers	1 – 3	2
Number of nodes per layer	10 – 500	160
Activation function	Relu, sigmoid, tanh	Relu
Learning rate	$1*10^{-4}$ – 0.01	0.0055

A scatter plot of the model results is shown in Figure 57 below. As can be seen in the figures, the model is not able to accurately predict the water depths, not even on the training dataset. This is also confirmed by the performance indicators on the validation dataset, see Table 33. Since the model is not even able to perform well on the training dataset (Figure 57a), this shows that the model set-up is too simple to understand the complex processes in the data and capture the relationship between the input and output.

Table 33 - Performance of MLP model (Random Search) on validation dataset

	Score on validation dataset
MAE	0.013m
RMSE	0.03m
CSI	47%

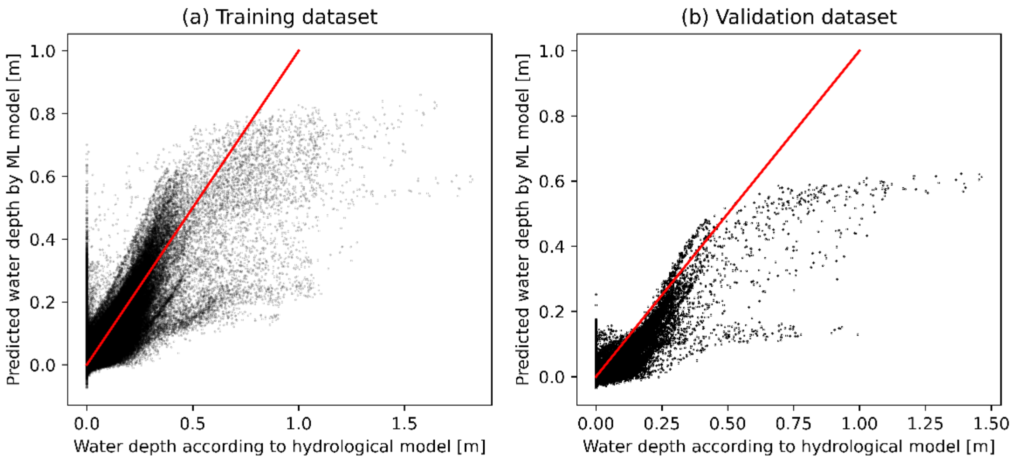


Figure 57 - Results MLP model with set-up found using random search

E Final model structure MLP model Case Study 2

```
n_epoch = 1000
n_batch  = 10
model = Sequential()
model.add(tf.keras.Input(shape = xdata_training.shape[1]))
model.add(Dense(855 , activation='relu'))
model.add(Dropout(0.1))
model.add(Dense(ydata_training.shape[1], activation='relu'))
opt = tf.keras.optimizers.Adam(learning_rate =0.0037)
model.compile(optimizer=opt,loss='mean_squared_error', metrics=['mse'])

fitting=model.fit(xdata_training, ydata_training, epochs=n_epoch,
batch_size=n_batch,validation_data=(xdata_validation,ydata_validation),verbose=2)
history = fitting.history
```

F Enlarged figures case study 2

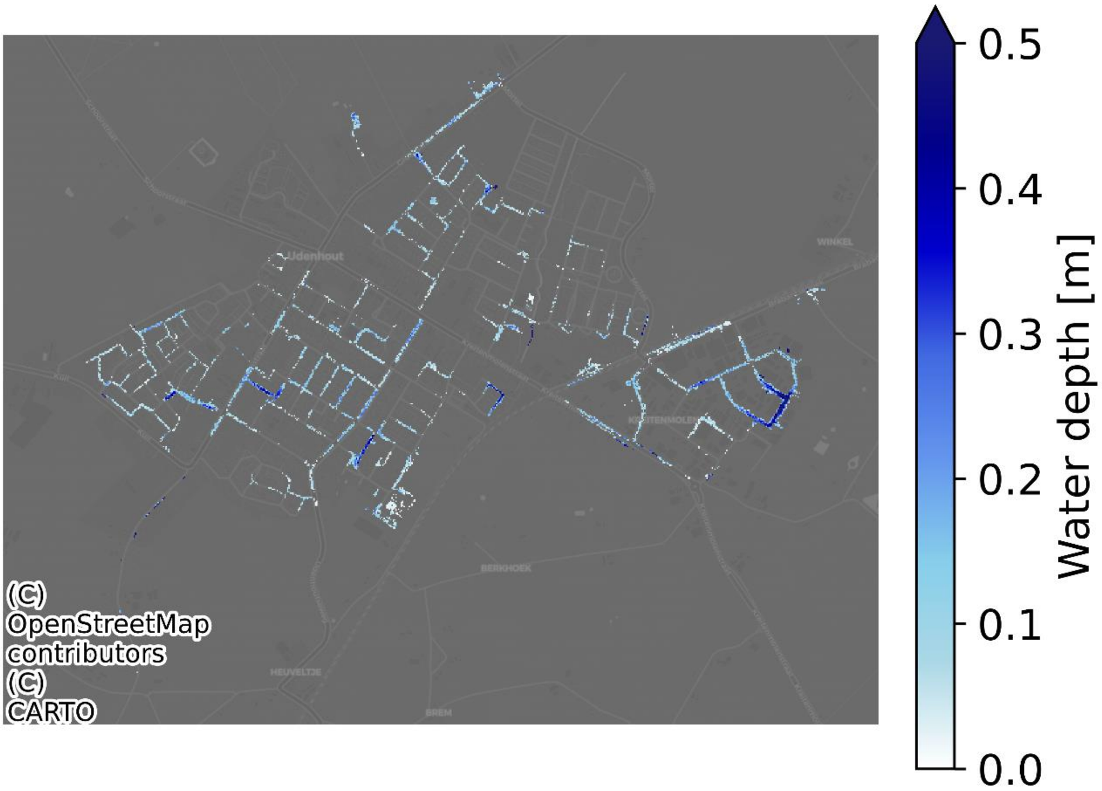


Figure 58 - Enlarged version of Figure 21

Figure 59 - Enlarged version of Figure 22

G Observed rainfall 09-09-2021

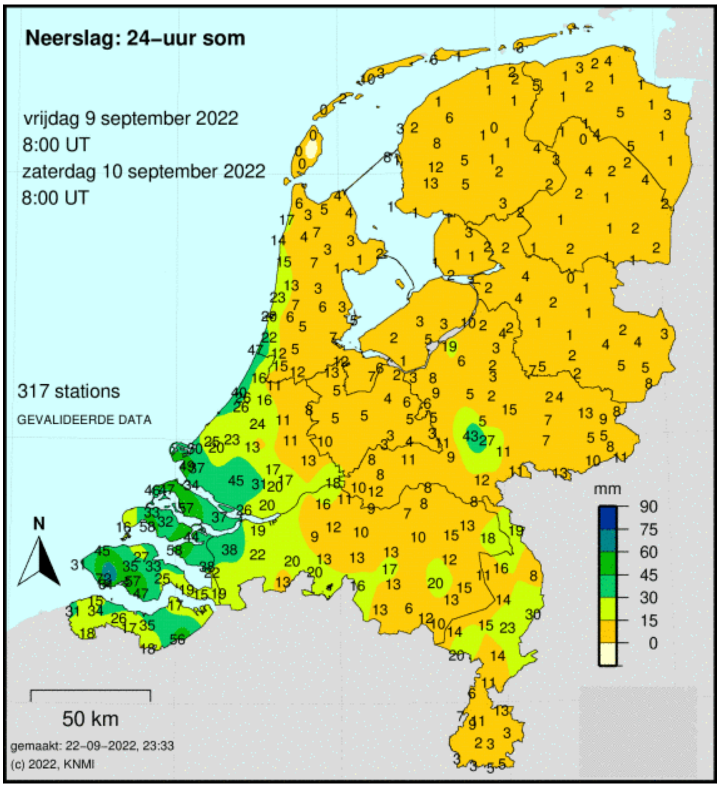


Figure 60 - Observed rainfall [mm] between 09-09-2022 at 08:00 and 10-09-2022 at 08:00 (cumulative). Figure is obtained from KNMI (2022).

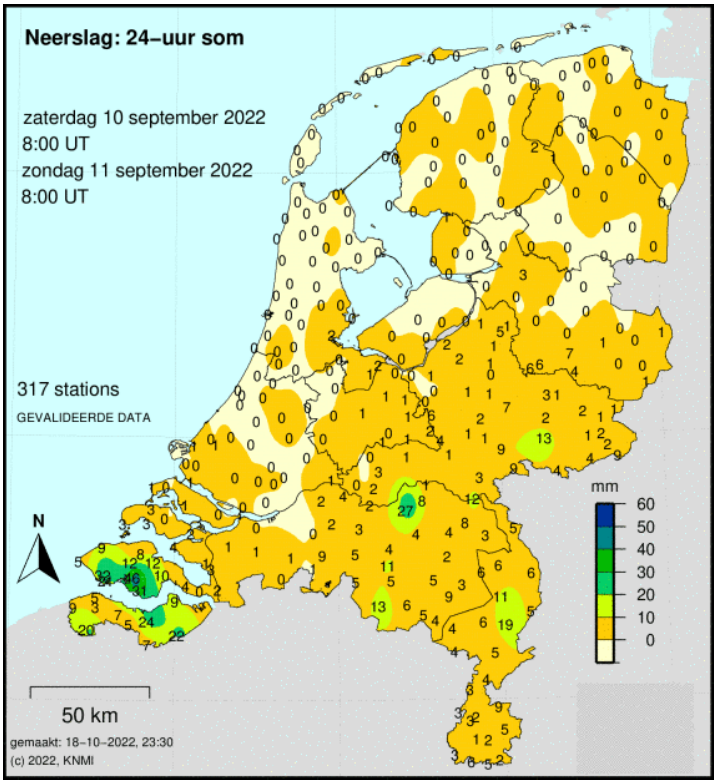


Figure 61 - Observed rainfall [mm] between 10-09-2022 at 08:00 and 11-09-2022 at 08:00 (cumulative). Figure is obtained from KNMI (2022).