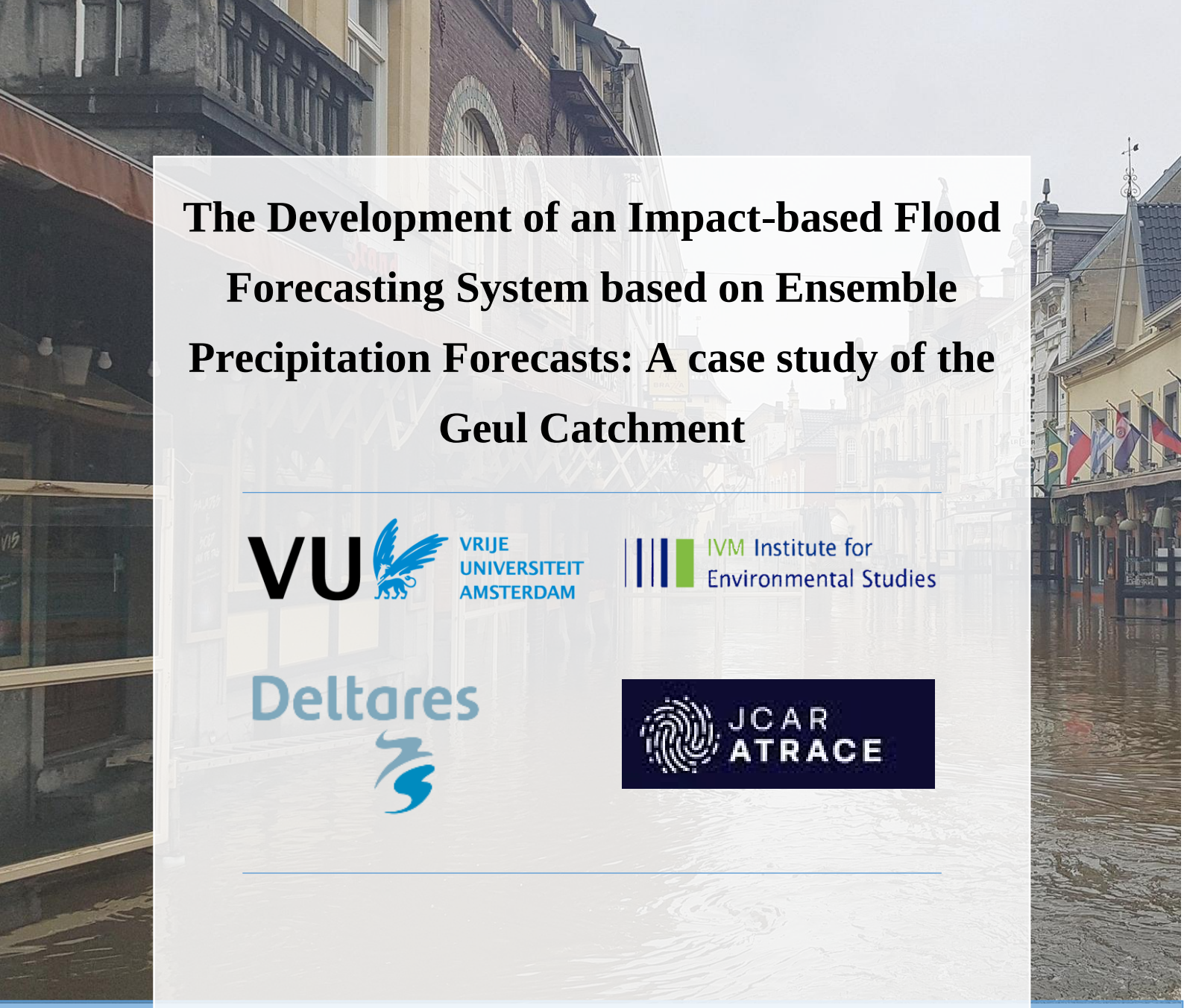

# The Development of an Impact-based Flood Forecasting System based on Ensemble Precipitation Forecasts: A case study of the Geul Catchment


**Author**:

Roy Pontman - 2818232

**Institutions**:

Institute for Environmental Studies of the Vrije Universiteit Amsterdam (IVM, VU)

&

JCAR ATRACE program at Deltares


**Date:**

June 30th, 2025

EARTH SCIENCES – MASTER HYDROLOGY

# The Development of an Impact-based Flood Forecasting System based on Ensemble Precipitation Forecasts: A case study of the Geul Catchment

Master thesis

To obtain the degree of Master of Science at the Vrije University of Amsterdam

Submitted in partial fulfilment of the requirements for the VU degree of Master of Science in Hydrology

To be presented on June 20th, 2025

This thesis is the result of a research project initiated by Deltares and part of the JCAR ATRACE program

# Colophon


| Authors: | Student number: | Email: |
| --- | --- | --- |
| *Roy Pontman (R.K.)* | *2818232* | *R.K.Pontman@student.vu.nl* |

Faculty:

*Earth Sciences Vrije Universiteit Amsterdam (VU)*

Study:

*MSc Hydrology*

Location:

*Amsterdam*

| Deltares Project supervisor: | Email: |
| --- | --- |
| *Dr. Ir. K. Slager* | *Kymo.Slager@deltares.nl* |
| IVM First research supervisor: | Email: |
| *Dr. T.S. Busker* | *Tim.Busker@vu.nl* |
| IVM Second research supervisor: | Email: |
| *Prof. Dr. J.C.H.J. Aerts* | Jeroen.Aerts@vu.nl |
| IVM Third research supervisor: | Email: |
| *R. Gouveia Loureiro Oliveira MSc* | *R.Gouveialoureirooliveira@vu.nl* |



Document:

*Research report*

Version:

*Final 1.0*

Publication date:

*30 June 2025*

# Preface

Dear reader,

With this thesis, I finish my Master's degree in Hydrology at the Vrije Universiteit of Amsterdam. Simultaneously, this marks the end of my student time in Amsterdam, where I first finished the University of Applied Sciences in Amsterdam whereafter I started with this master. My research involved the development of an Impact-based flood forecasting warning system in the Geul catchment using a modelling chain. I started this research with the goal to show the potential of combining different innovative components within the flood risk assessment. Hopefully, the conclusions from this study can add to the current state of the art in the usage of flood forecasting systems inside transboundary regional river catchments and help local water authorities in flood preparedness and mitigation.

This master thesis has taught me how to deal with ensemble weather predictions, hydrodynamical processes, and impact estimations. Additionally, I learned a lot about the needed components that are part of doing scientific research. On a personal level, I also learned a lot about how to deal better with my perfectionism and intrinsic motivation. Simultaneously, by connecting the different components in the chain, I improved my programming skills in python. This programming primarily involved modelling and data preparation, analysis and visualization but I also learned to create an interactive dashboard. One of the highlights in this period was the presentation of my research at the water authority Limburg where I was able to present this dashboard as well.

Fortunately, I did not do this research on my own and, therefore, I would like to thank some people for the support (in all forms) and guidance throughout this thesis period. First, this thesis did not originate without the prior open conversations with one of the supervisors Tim Busker. Moreover, Tim formed together with Rafaella Gouveia Loureiro Oliveira and Jeroen Aerts the supervision team from the VU. Without their expertise, their valuable time for questions and discussions, their feedback, and their support the thesis would be a much tougher period. They also helped me deepening the results of this study with their critical view for opportunities. I also want to thank Kymo Slager for the supervision at Deltares, his critical view, and his expertise of flood risk management. Additionally, I would like to thank Jens de Bruijn and Veerle Bril from the IVM, who helped me solving modelling issues and understanding the modelling framework.

Finally, I would like to thank my family and friends for their support during my thesis but also for the amazing time and distraction. Furthermore, I would like to thank Rianne Sijm and Paul Kuijn for their listening ears.

Hopefully you enjoy reading my thesis!

*Roy Pontman*

*Amsterdam, June 30$^{th}$, 2025*

# Abstract

Impact-based forecasting is recognized as crucial for effective flood risk management since it translates meteorological forecasts into actionable information about potential impacts on critical facilities, residents, and infrastructure. Despite its growing relevance, including impacts within existing flood early warning systems remains in its infancy with limited operational implementations and few performance evaluations.

This study has developed an Impact-based flood forecasting system for the Geul catchment in the GEB framework, using eight historic operational ECMWF ensemble and control precipitation forecasts to simulate the 14 July 2021 flood event. By using these different forecasts at different lead times as input, the SFINCS hydrodynamic model simulated flood maps for the eight lead times ranging from 88 hours till 5 hours before the event. Simultaneously, a study baseline is created by using the ECMWF's ERA5 reanalysis dataset in the framework. To improve the computation efficiency, this study has introduced an innovative approach where percentiles are calculated for each ensemble to describe the ensemble' uncertainty. Each ensemble percentile and control-based flood map is then evaluated against observations and empirical data, whereafter it is combined with vulnerability and exposure data to assess flood impact distributions, probabilities and hits of critical facilities.

Up to 88 hours before the event, the results demonstrate adequate performance for the $90^{th}$ and $95^{th}$ ensemble percentiles in simulating flood extents across the catchment (CSI of ~0.6), while the performance decreases for lower percentiles by an increasing lead time. Although the simulated flood map still show room for improvement due to missing hydraulic structures and survey-based limitations, the developed model chain proves capable for capturing flood impacts as the event approaches. This is further underpinned by the detailed Valkenburg evaluation, which showed stronger performances for the flood extent simulations (CSI of >0.7) and for predicted impacts on a building level across the different forecasts and lead times. By integrating ensemble precipitation forecasts within the framework, this study offered valuable insights for the representation of forecast uncertainty and forms a proof-of-concept for future impact-based early warning applications. Moreover, it addresses the potential of probabilistic flood impact forecasting, while it emphasizes the need for further refinements and stakeholder engagement to improve the operational relevance inside the Geul basin.

# Table of contents

# List of Figures

# List of Tables

# Glossary

| | |
|---|---|
| ***CDF*** | Cumulative Distribution Function |
| ***DEM*** | Digital Elevation Model |
| ***CRPS*** | Continuous Ranked Probability Score |
| ***CWatM*** | Community Water Model |
| ***DRR-cycle*** | Disaster Risk Reduction cycle |
| ***ECMWF*** | European Centre for Medium-Range Weather Forecasts |
| ***EM-DAT*** | Emergency Events Database |
| ***EPS*** | Ensemble Prediction System |
| ***ESA*** | European Space Agency |
| ***EWS*** | Early Warning Systems |
| ***FEWS*** | Flood Early Warning Systems |
| ***GAR*** | Global Natural Disaster Assessment Reports |
| ***IbF*** | Impact-based Forecasting |
| ***IBFFWS*** | Impact-Based Flood Forecast and Warning Services |
| ***IIASA*** | International Institute for Applied Systems Analysis |
| ***IPCC*** | Intergovernmental Panel on Climate Change |
| ***IVM*** | Institute for Environmental Studies |
| ***NWP*** | Numerical Weather Predictions |
| ***OSM*** | Open Street Map |
| ***PIT-diagrams*** | Probability Integral Transform diagrams |
| ***RWS*** | Rijkswaterstaat |
| ***SFINCS*** | Super-Fast Inundation of CoastS |
| ***UNDRR*** | The United Nations Office for Disaster Risk Reduction |
| ***WMO*** | World Meteorological Organization |

# 1. Introduction

Globally, natural disasters are increasing in frequency and severity, leading to increasing economic losses and population displacements (EM-DAT & CRED / UCLouvain, 2024; Merz et al., 2020). Climate change amplifies these natural disasters resulting in more severe impacts (Bolan et al., 2023; Kreibich et al., 2022). In 2024, the 1.5 °C global warming threshold was exceeded for the first time, pushing climate goals further out of reach (Copernicus, 2025). This development of climate change will put mankind more at risk, resulting in even more severe impacts in the future. Simultaneously, together with global population growth and related urbanization, the vulnerability for natural hazards further increases (Dodson et al., 2020; Norrman, 2023). While different regions in the world faces different sets of natural hazards, flooding remains the most devastating (Rentschler et al., 2022). Between 1994 and 2013, floods affected globally nearly 2.5 billion people and caused more than $40 billion in annual losses (WMO, 2024). The number of people living in flood-prone areas rose by 24% between 2000 and 2015, and without action flood impacts will worsen (Trogrlić et al., 2022; WMO, 2024). Moreover, flood risk is shaped by diverse meteorological and hydrological drivers that differ across time and place (Merz et al., 2020; Piadeh et al., 2022). Ignoring the interaction of flood drivers can lead to a significant underestimation of the flood risk (Kumbier et al., 2018), especially if global extreme precipitation become more intense and frequent (Tradowsky et al., 2023).

In response to previous catastrophic flood events and climate change projections, global attention has shifted towards improved preparedness to reduce (future-) flood risk. In the last decades, Early Warning Systems (EWS) (UNDRR, 2023) have become a crucial component of the Disaster Risk Reduction (DRR) cycle. They support timely preparedness and response by integrating flood hazard, exposure, and vulnerability information (IPCC, 2012; Trogrlić et al., 2022). Growing attention has been giving to such non-structural approaches due to their rapid implementation, limited spatial requirements, and cost-effectiveness (Berndtsson et al., 2019; Piadeh et al., 2022).

In this context, the UNDRR recently launched the ‘Early Warnings for All’ initiative, aiming to ensure that every country has access to EWS by the end of 2027. The main objective is to build more resilience through societies, with the goal to save lives and livelihoods (UNDRR, 2023). Flood Early Warning Systems (FEWS) have proven to be a viable measure in mitigating flood risk by combining the scientific understanding of the natural processes that generates flood hazards, past experiences with flood hazards, and monitoring of current flood conditions (Merz et al., 2020; Perera et al., 2019; Ringo et al., 2023). The combination of factors made it possible to forecast the likelihood of precipitation, discharges, and water levels at different lead times and levels of confidence (Merz et al., 2020). In the end, when warnings become more precise, they result in a diminished number of false alarms, which increases the trust residents have in the forecasts (Fernández-Nóvoa et al., 2024; Lindenlaub et al., 2024).

However, to have an effective FEWS, it must be people-centred and must integrate the following four pillars: (i) Knowledge of the risk faced, (ii) Available technical monitoring and warning services, (iii) Dissemination of meaningful warnings to those at risk, and (iv) Public awareness and preparedness to act. Failure of one of these elements can mean failure of the entire FEWS chain (Fernández-Nóvoa et al., 2024; UNDRR, 2008).

One of the issues in current FEWS is communicating the warning so that it will trigger responses (Islam et al., 2025). De Perez et al. (2022) address the need for more focus on the communication and response capability in FEWS initiatives, as some of the deadliest and

costliest flood disasters of this century have happened despite these events were forecasted with a FEWS (e.g. Davidson & Ni, 2024; Endendijk et al., 2023; Khalid et al., 2015; Latona et al., 2024; Thieken et al., 2023).

Moreover, another important aspect is that current FEWS focus on flood hazard forecasting instead of including risk or impact information, such as expected amount and distribution of physical damage, consequences to important assets and humans, disruption of community services or financial loss (Merz et al., 2020). The providence of information on the forecasted impacts is important as it supports the translation towards response and action (Busker et al., 2025).

Building on this, a key limitation of current FEWS is the lack of communication and the absence of impact-based information, which has a large societal importance as this impedes an effective response. To bridge this gap, the integration of Impact-based Forecasting (IbF) into existing FEWS has been proposed (WMO, 2015; Busker et al., 2025). As a result, an Impact-Based Flood Forecast and Warning Services (IBFFWS) enhance public response, as people are more likely to act when provided with detailed information about the impacts of a hazard (Lindenlaub et al., 2024). This shifts the focus from forecasting what the weather will be to what it will do, which will heighten the response rate and communication to all the involved stakeholders (Merz et al., 2020; Potter et al., 2025). However, the effectiveness of IBFFWS has never been objectively tested and modelled as it comes with significant operational challenges, such as data scarcity of impacted areas in order to address vulnerability or exposure (Merz et al., 2020).

Therefore, the primary objective of this study is the following:

*To develop and test an Impact-Based Flood Forecast and Warning Services (IBFFWS), wherein ensemble precipitation forecasts are incorporated into a SFINCS hydrodynamic model.*

Another objective of this study is to implement ensemble forecasting within IBFFWS to investigate the added value of these forecasts compared to ERA5 reanalysis data or deterministic forecasts from ECMWF. Ensemble forecasts need to be implemented as they treat the uncertainty from the source leading to more information on uncertainty, resulting in different outcomes (Boelee et al., 2018). These different outcomes can improve communication, as a range of possibilities are clarified. Therefore, ensemble forecasts are shown to have higher value for decision-making (Verkade & Werner, 2011). In the end, the ensemble forecasts will be used to predict impacts in a probabilistic way.

Based on the defined research gap and the main objectives, early warning systems can be enhanced by providing impact-based forecasts to support the trigger of effective actions. To increase the knowledge base around IBFFWS, this study will investigate the development of such an updated system in the Geul catchment. The following research question (MQ) has been formulated:

***How could Impact-based Forecasts (IbF) from the hydrodynamic SFINCS model have been used to trigger effective early-warnings and actions, incorporating associated uncertainty for the 2021 flood in the Geul basin?***

To develop an IbF for the Geul basin, the forecasts leading to the flood in July 2021 will be used to forecast flood maps. These forecasted flood extents will be compared with the available observed data and then translated from flood extents and depths to impacts, including an estimation of uncertainty. Additionally, to answer the research question, the study is guided by different sub-questions:

***RQ1: How well do historic operational ensemble and deterministic forecasts relate to the observed precipitation patterns and totals of the July 2021 flood event, compared to ERA5 reanalysis data?***

***RQ2: What is the quality of flood forecasts from SFINCS for the July 2021 event at multiple lead times and how do variations in forecast data affect the outcomes of the SFINCS model?***

***RQ3: How does the predicted flood impact vary across different rainfall forecasts and lead times, and what is the added value of using ensemble-based predictions for estimating local damages per function and area?***

This study provides a state-of-the-art modelling IBFFWS chain, which can provide more information to enhance the communication and flood preparedness in regions. The Geul catchment serves as a practical example and proof-of-concept to highlight the potential of the study setup in forecasting and visualizing flood impacts based on local impact dynamics. Moreover, it provides lessons that contribute to the improved management of flood risk reduction in a transboundary region, where the system offers a practical contribution to the added complexity of cross-border flood management.

Ultimately, this study focuses only on developing an IbF for the Geul catchment. This includes the implementation and preparation of different precipitation ensemble forecasts within the SFINCS hydrodynamic model, the translation from flood extent to impact, the evaluation of the flood maps with the observed flood extent and flood depth, and the quantification of uncertainties within the precipitation ensemble forecasts and within the impact forecasts.

Further analysis and usage of the impact forecasts regarding communication, warning levels and response towards stakeholders is not considered in this study. This study is limited to providing insights in the usage of ensemble forecasts in hydraulic modelling and the translation of flood extents into impacts.

Moreover, in this study the used models were not calibrated from scratch, as a functioning and validated setup was provided for this research. As a result, only minor adjustments in SFINCS were made to improve the accuracy of the outcomes and to ensure consistency with the implemented rainfall forecasts. These adjustments involved aligning the coordinate system and refining the channel network representation to better match the spatial characteristics. This means that no additional calibration of model parameters was conducted to focus on testing the model's response to different forecast scenarios.

In addition, although one of the main reasons of this research is that weather events become more extreme due to climate change, climate change is not considered in this thesis. This study only applies historical precipitation ensemble forecasts and does not include climate scenarios.

# 2. State of the art

This chapter provides a literature review of the state of the art in flood forecasting systems, hydrological- and hydrodynamic models, and in ensemble weather forecasting. By examining and elaborating the mentioned research gaps, this chapter forms the foundation for further investigation and supports the stated research question and objectives. Therefore, the research gaps in current early warning systems are examined, providing an overview of the state of the art in the field. Additionally, the state of the art in hydrological- and hydrodynamic models, as well as in ensemble weather forecasting, are analysed to enhance the understanding of the key components used in developing the model-chain for impact-based forecasting.

## 2.1. Key characteristics of existing Flood Early Warning Systems (FEWS)

Within FEWS a sequence of steps take place from the moment first signs of a flood are recognized (t0), to the moment that a flood occurs 'Low point in route cut by flood water' (ti) (Esm et al., 2010). This timeline is also called **the maximum potential warning time,** and an example is illustrated in Figure 1. FEWS come into play in the first part (t0 till tw) of the timeline (depicted in Figure 1). During this phase, measured or forecasted hydrometeorological data is collected and evaluated to determine the flood severity. Meteorological institutes often provide information about (forecasted or measured) precipitation events, which will be translated to discharge or water level thresholds by a FEWS (Esm et al., 2010). A FEWS often combines hydrological models for the land-phase (rainfall runoff processes) with hydraulic models to simulate the propagation of precipitation events towards high discharges in streams (Arduino et al., 2005). The phase between the threat recognition (t0) and 'Low point in route cut by flood water' (ti) in Figure 1 is called the **lead time** (Esm et al., 2010). During the lead time warnings are given, which result in response and actions.

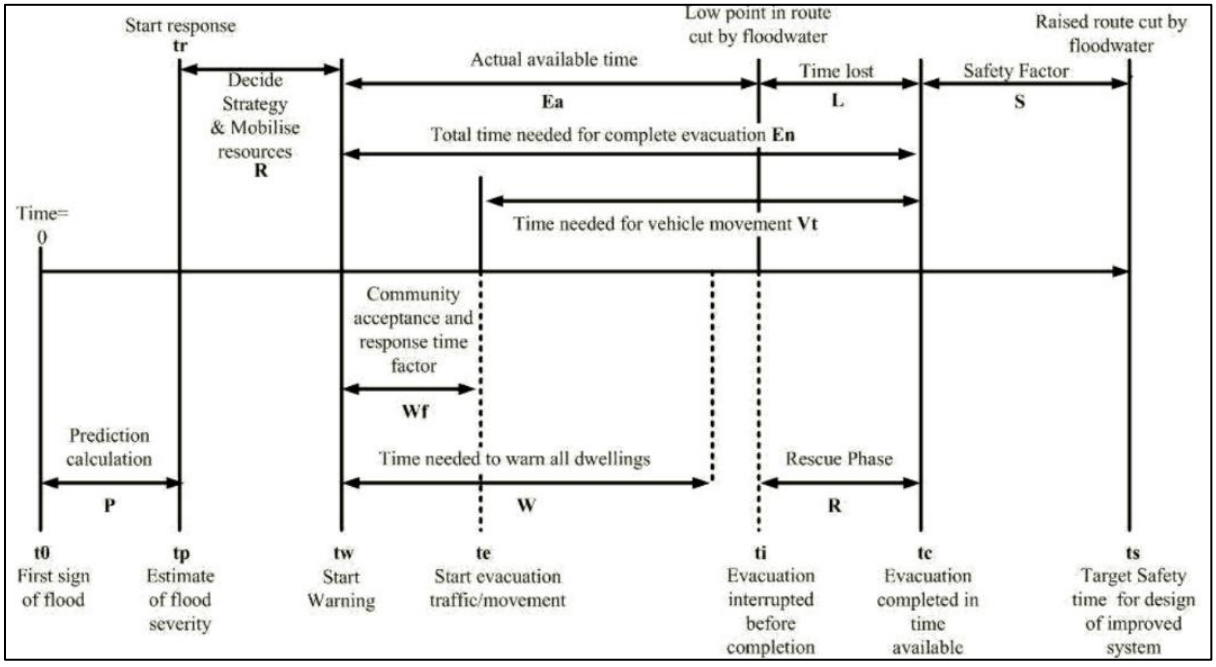


*Figure 1: Flood timeline. The sequence of consecutive steps from recognition to estimation of severity to decision-making to response to action (Esm et al., 2010).*

Therefore, current FEWS require information inputs such as real-time rainfall information, high-resolution numerical weather forecasts, and the operation of hydrological model systems (Collier, 2007). However, the content of the elements can vary within FEWS. In fact, the difference in content is dependent on the type of hazard. Typical triggers for flood events are short-duration rainfall, high-intensity rainfall, long-duration rainfall, rain on saturated soils, snowmelt, or a combination of rainfall and snowmelt (Merz et al., 2020). Hence, pluvial and flash floods do need more accurate short-term rainfall information due to the local rainfall peaks. These extreme peaks develop on different space-time scales compared to fluvial floods. As a result, dependent on the trigger of a flood hazard, the input information and the type of hydrological or hydrodynamic model can differ in order to capture the flood hazard on time (Merz et al., 2020).

Furthermore, dependent on the type of flood, the warning lead times vary with lead times being less than an hour for flash floods and extending up to several weeks for fluvial floods (Merz et al., 2020). In addition, these lead times depend on various catchment characteristics (e.g. catchment lag time, catchment size) (Jain et al., 2017; Piadeh et al., 2022).

### 2.1.1 Component I: Weather forecasting

In general, current FEWS can be divided into two parts: meteorological forecasting and hydrological forecasting (Das et al., 2022). Where weather is the state of the atmosphere over a shorter time span, meteorological forecasting models try to predict the future state of the atmosphere. Out of a number of meteorological factors, temperature, pressure, wind, humidity, and precipitation have the greatest influence on a location's weather (Kumar & Sharma, 2024). Many FEWS depend on meteorological inputs from observation stations or radar measurements. These measurements, however, provide forecasts for short lead times (1-2 days ahead). In order to provide early warnings for severe flood events, it is more useful to use medium-range weather forecasts (2-15 days ahead) to allow sufficient preparation time for civil protection authorities (Cloke & Pappenberger, 2009; Das et al., 2022).

To produce medium-ranged weather forecasts, meteorological variables from numerical weather predictions (NWP) are used. NWP is a method to predict likely future states of the atmosphere, by solving a set of differential equations based on the current atmospheric conditions (Cloke & Pappenberger, 2009; Teja et al., 2023). The forecasts with NWP can be generated in two ways: Deterministic, and Probabilistic. A deterministic forecast provides one specific outcome from many scenarios, typically with short lead times and without considering associated uncertainties (Teja et al., 2023).

Although deterministic forecasts are based on the best available initial atmospheric conditions, they may still under- or overestimate actual outcomes. As a result, the usage of these forecasts poses a challenge in communicating warnings or decision making in flood risk management. To overcome these limitations, the probabilistic or ensemble forecasts are developed to incorporate the uncertainties by simulating multiple equally probable future states of the atmosphere (Cloke & Pappenberger, 2009; Das et al., 2022; Pappenberger et al., 2019; Teja et al., 2023). By creating different forecast with slightly different initial conditions and parametrization compared to the best initial state, named as the control forecast, different forecasts are initialized with an Ensemble Prediction System (EPS). The multiple simulations run in parallel, where the varied initial conditions lead to perturbed forecasts. The divergence of the control forecast with the perturbed forecasts gives an estimate of the uncertainty and spread of the weather prediction on a particular day, making ensemble prediction systems valuable for risk-based decision-making (ECMWF, 2008).

Nonetheless, ensemble weather forecasts are complex and characterized by uncertainty primarily due to the chaotic nature of atmospheric dynamics. Even small perturbations in the initial conditions can lead to different atmospheric future states. This effect is known as the "butterfly effect". To represent this uncertainty in flood forecasting, multiple NWP ensembles can be forced through a single hydrological model, or a single NWP ensemble can be coupled with multiple hydrological models (Das et al., 2022).

In Europe, most meteorological weather forecasts are used from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Das et al., 2022). To use ensemble forecasts of ECMWF, the forecasts are initialized every six hours but released according to a fixed dissemination schedule (ECMWF, 2025a). This dissemination schedule delays user access and consequently reduces the effective lead time.

This study focuses on applying historical operational ensemble precipitation forecasts from ECMWF to evaluate their effectiveness in predicting the 2021 flood event. Unlike hindcasts, which are retrospective simulations, operational forecasts reflect real-time conditions and model configurations used at the time. These forecasts are used to force a single hydrological model with predicted rainfall inputs.

### 2.1.2 Component II: Hydrological forecasting

The second part of current FEWS is forcing the meteorological forecast data into a hydrological model. In general, hydrological models are utilized to describe parts and interactions of the hydrological cycle. As models are a representation of the reality, different hydrological models describe processes and interactions of the hydrological cycle in catchments differently (Pechlivanidis et al., 2013). Because of the simulation of natural processes, there is no single best model to use. Rather, there are many solutions, depending on the objective and needed complexity (Ogden, 2020). Additionally, while the hydrologic cycle is global in nature, large-scale meteorological conditions are processed on a local scale. This specific catchment response makes hydrology highly heterogeneous over space and time, which is also called the uniqueness of place in hydrologic modelling (Beven, 2000). Ultimately, the goal of hydrological models is to support and improve decision-making in water resource management and flood risk management in river catchments (Pechlivanidis et al., 2013).

### 2.1.3 Component III: Communication, Response & Early action

A common approach in medium-range flood forecasting, is by forcing a hydrological model with an ensemble of NWP to estimate the probable future hydrological conditions. In this case, the hydrological model predicts whether predefined hydrological exceedance thresholds are surpassed with a certain probability (Alfieri et al., 2019). When a threshold is surpassed, warnings can be given to the public or to crisis managers who make decisions whether to respond. Predefined meteorological and hydrological warning levels are often depicted in color codes as illustrated in Figure 2. These warning levels can be connected to predefined actions, outlined in crisis management plans (Busker et al., 2025).

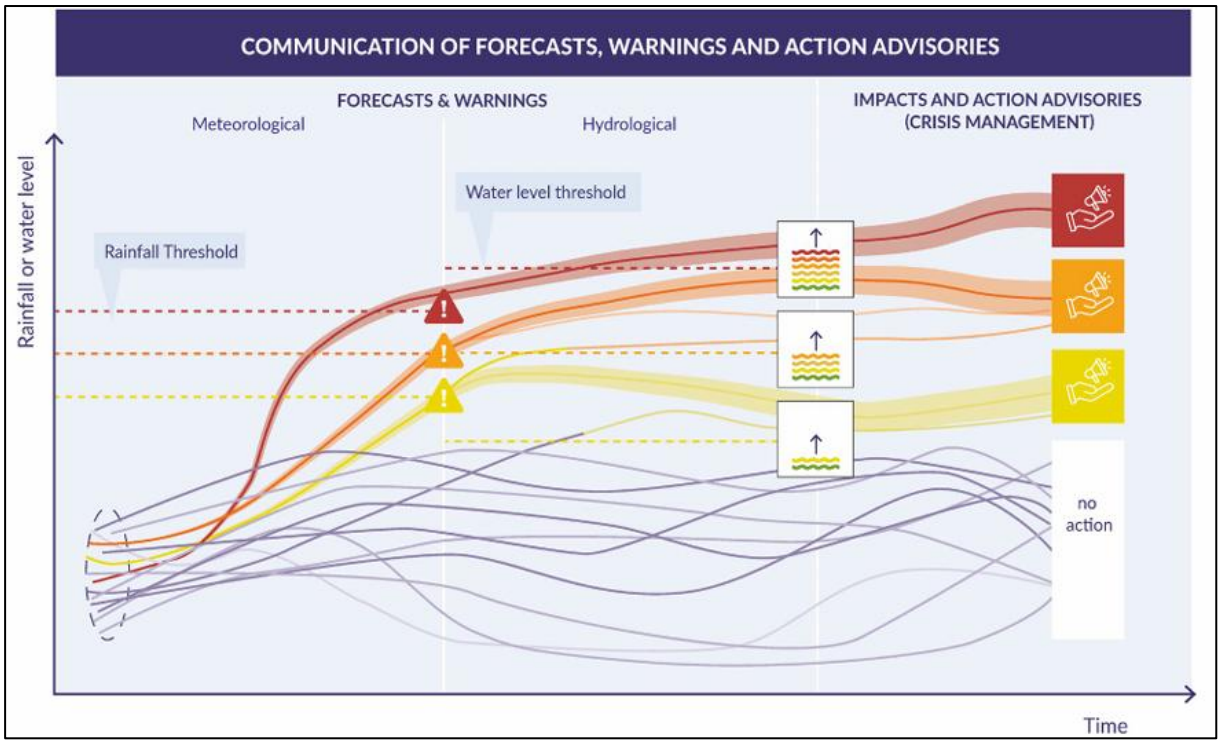


*Figure 2: Conceptual representation of a FEWS. Rainfall ensemble forecasts (left) are used in hydrological models to predict river water levels (middle). Warnings are issued if a share of the ensembles exceed the predefined thresholds, which result in triggering (sometimes predefined) early actions (right) (Obtained from Busker et al. (2025)).*

Subsequently, if authorities have decided to warn the public, ideally warnings should be communicated in a way that the response is correct and that people act on the warning (Trogrlić et al., 2022). For instance, a warning can include evacuation of people and property or implementing measures such as barriers.

Due to the usage of forecasts and models in different components, a key challenge in the development and implementation of a FEWS is that uncertainty can propagate further to the next component. By going through all FEWS' components, the cumulative uncertainty propagation can have effects on the systems output (Parker & Priest, 2012). For example, uncertainty in the collection and preparation of the rainfall ensemble forecasts could lead to 'missed' forecasts where rainfall thresholds are not surpassed, while observed water levels did surpass the water level thresholds. On the other hand, uncertainty can also lead to forecasts that did surpass the thresholds, without actual exceedance of thresholds within the observations (Busker et al., 2025; Cloke & Pappenberger, 2009; Merz et al., 2020). These so-called false alarms and misses have effects on the trust that people have in current FEWS. The propagation of uncertainty can therefore undermine the credibility of FEWS (Sawada et al., 2022).

## 2.2. Major challenges in current FEWS and in developing Impact-based Forecasting Warning Systems (IBFFWS)

Although current EWS have a proven substantial monetary benefit against the damages of natural hazards (Pappenberger et al., 2015), these systems are based on the ability of people to use the given information and take effective action (WMO, 2015). Moreover, current FEWS are used to predict the magnitude, location, and timing of potential damaging events (Merz et al., 2020).

The study of Busker et al. (2024) concludes that current FEWS can effectively trigger early action up to 2-3 days in advance across much of Europe, while the value of the forecasts largely disappears at a 5-day lead time. However, despite the focus on developing the hydro-meteorological monitoring and forecasting systems (Fernández-Nóvoa et al., 2024; Najafi et al., 2024; Pappenberger, 2019; Sawada et al., 2022), the decline of uncertainty is not a sole prerequisite for reducing the impacts of flood hazards. Consequently, an accurate and timely warning does not guarantee safety of life or prevention of major economic disruption (WMO, 2015).

For example, during the floods of July 2021 in Northwestern Europe a lot of people were not warned or aware of the flood severity, despite having FEWS in place. Although the risk was acknowledged and forecasted, the final pillar 'Public awareness and preparedness to act' for a good functioning FEWS was not sufficient (Busker et al., 2025; Endendijk et al., 2023).

Moreover, post-event analysis of the flood event has revealed that FEWS solely focusing on hazard metrics, such as maximum local rainfall depths or maximum water levels, resulted in misinformed actions, delayed responses, and at times, no action at all (Najafi et al., 2024). Here the communication of flood warnings to first responders and the public appears challenging (Busker et al., 2025; Merz et al., 2010; Parker & Priest, 2012).

Furthermore, the rapid assessment of impacts immediately after an event or the provision of impacts prior to an event tend to be developed without the involvement of relevant stakeholders: the information sharing and interaction between stakeholders during an event is often not well integrated (Merz et al., 2020; Potter et al., 2025). Moreover, for governments, economic sectors and the public to take appropriate action, they must know how hazards might impact their lives, livelihoods, and the economy (WMO, 2015).

In addition, interviewees in the study of Busker et al. (2025) stressed that estimates of flood areas are highly needed to improve the effectiveness of early actions. As the flood estimations can be combined with overlays of exposed buildings, land covers, and critical infrastructure, flood impacts can be mapped (Najafi et al., 2024). However, a key factor contributing to inaction is ineffective communication of the forecasts. Interviewees stressed that the information needs for emergency management are often not met. Moreover, interviewees emphasized the challenge of interpreting probabilistic forecasts, resulting in the preference for using deterministic forecasts to take action (Busker et al., 2025). Operational services such as safety regions, fire brigades, police, and ambulance address the need for impact-based communication, as they underscored that their knowledge about potential impacts of 150 mm or 200 mm rainfall events remains limited (Busker et al., 2025; Lindenlaub et al., 2024; Najafi et al., 2024). Therefore, a multi-disciplinary approach with all the stakeholders is needed to heighten the response rate, information sharing, and mitigate the impacts of natural hazards.

To forecast flood maps and associated flood impacts, current FEWS are often extended with the integration of additional components, for instance, depth-damage curves or probabilistic multivariable vulnerability models. A major challenge in this approach lies in producing timely

and accurate estimates of water levels and flood areas, which are essential to determine impacts (Najafi et al., 2024). This is achieved either by fast hydrodynamic simulation approaches or by using predefined flood maps (Merz et al., 2020; Najafi et al., 2024).

### 2.2.1 Including hydrodynamic models in FFEWS

As hydrological models are utilized for the simulation of the interaction processes in the hydrologic cycle, one key aspect in the behaviour of large-scale river basins that has been partially neglected in large-scale hydrological models is river hydrodynamics (Paiva et al., 2011). Often hydrological models use simplified flow routing models which do not include dynamical processes that occur in rivers such as backwater and floodplain storage effects. These effects play significant roles in fluvial systems such as in biogeochemical processes, feedback between land, atmosphere and water (Paiva et al., 2011). The simulation of flood inundation provides the basis for the representation of all these processes (Paiva et al., 2011). Using hydrodynamic flow routing (inclusion of dynamical processes) provide also model outputs such as river stages, flow velocities and slopes (Paiva et al., 2011).

Within this context, one-dimensional (1D) and two-dimensional (2D) hydrodynamic models are often used to simulate flood behavior. First, 1D models assume flow varies only in the longitudinal (streamwise) direction, neglecting lateral and vertical variations. Water levels are calculated using discharge-water level relationships or flood wave propagation equations. Some models can partially account for transverse (lateral and vertical) flow effects by including different cross sections. After predicting the water levels at fixed points in the river, 1D models can extrapolate the flood depths at the floodplains based on an elevation map (Hamdi et al., 2019).

In contrast, 2D models compute depth-averaged water surface elevations and fluxes over a two-dimensional grid. As a result, these models require more detailed topographic data, roughness data, and boundary condition data. Hence, 2D approaches better capture spatial variability in flood modelling, offering a more accurate estimate of flood risk (Hamdi et al., 2019).

Therefore, 2D Hydrodynamic models have been in common use for several decades worldwide, for the simulation of flood events for engineering, planning and risk assessment studies (Nicholas, 2003). As these models rely on numerically solving partial differential equations to spatially evaluate flood extents (Leijnse et al., 2021; Paiva et al., 2011), this is much more computationally intensive compared to hydrological models. Currently, different studies tried to improve the time expensive models by simplifying the partially differential equations or by using Machine Learning methods (Haces-Garcia et al., 2024; Leijnse et al., 2021).

In the context of the UNDRR mission to create FEWS in all countries by 2027, the focus of this study is to apply a 2D SFINCS hydrodynamic model together with ensemble forecasts to generate flood map forecasts. In fact, current FEWS use hydrological models to warn people based on hydrological thresholds. In contrast, this study uses ensemble weather forecasts in a 2D hydrodynamic model together with a hydrological model, which initializes the hydrological processes to forecast flood extents and depths.

### 2.2.2 Estimating impacts in FFEWS with depth-damage curves

To complement an IBFFWS, depth-damage curves can be utilized to translate simulated flood water levels into direct economic damages. However, having adequate depth-damage curves to describe the vulnerability on a local scale is a significant challenge that stands in the way of flood risk modelers (Pita et al., 2021).

Specifically, damage assessments are currently limited in evaluating the impacts from flooding due to the absence of a comprehensive global database of flood damage functions that can translate flood water levels into direct economic damage (Huizinga et al., 2017; Merz et al., 2010). Moreover, the evaluation and reliability of the damage assessments are complicated due to the scarcity of reported impact data at the local scale (Merz et al., 2020; Potter et al., 2025).

To overcome the lack of site-specific functions, modelers usually adopt functions from other locations, extrapolate from limited flood damage data, or conduct expert-judgement surveys (Huizinga et al., 2017; Pita et al., 2021). Additionally, the damage fractions in the curves are intended to span from zero (no damage) to one (maximum damage). Therefore, it is necessary for each location to normalize the curve based on the local maximum damage (Huizinga et al., 2017). Furthermore, the technical report of Huizinga et al. (2017) presented depth-damage functions for 214 countries with the following damage categories: Residential buildings, commerce, industry, transport, infrastructure, and agriculture.

Ultimately, this study will apply depth-damage curves in order to translate flood levels into direct damages. Since the development of IBFFWS is still in its infancy, the majority of current FEWS are focusing on extending their system with qualitative or quantitative impact assessments (Busker et al., 2025; Merz et al., 2020). These assessments are commonly based on direct damages to the residential and commercial sector (Huizinga et al., 2017; Merz et al., 2020). In contrast, impacts on critical infrastructure or indirect impacts are rarely considered (Merz et al., 2020).

However, to act effectively and to communicate as early as possible to the local public, it is important to know what the impacts would be for emergency services, critical infrastructure and facilities. Hence, the development of IBFFWS faces further challenges related to the definition of relevant impact information. Reasonably, the forecasted impact still must be aligned with the responsibilities, knowledge and requirements of the involved stakeholders (Busker et al., 2025; Merz et al., 2020). Subsequently, it is important to effectively share the propagation and representation of uncertainties within the IbF (Najafi et al., 2024).

Despite the challenges of extending current FEWS with IbF, the future thresholds (Figure 2) must be based on the expected impact on people and their assets (Busker et al., 2025). Therefore, there is a need for developing and testing impact-based forecast systems with different impact-based thresholds (WMO, 2015; Red Cross Red Crescent Movement & Climate Centre (RCCC), 2020; Merz et al., 2020). Multiple IbF initiatives are already launched by WMO (2021).

# 3. Study area

The study takes the transboundary Geul catchment as a case study for the implementation of an IBFFWS. The river is located between 50°39′ to 50°54′N latitude and 5°65′ to 6°6′E longitude and flows through Belgium, Germany and Limburg (Southern province of the Netherlands), as illustrated in Figure 3. First, the Geul catchment is described (section 3.1), then a description is given of the EWS in the Netherlands (section 3.2), and finally the flood event of 2021 is described (section 3.3).


*Figure 3: Topography of the Transboundary Geul Basin. Source: (JCAR ATRACE, 2025-a)*

## 3.1 The Geul catchment

The Geul River is an important tributary of the Meuse and drops about 250 meters over approximately 60 km, making it one of the few steeply sloping rivers in the Netherlands with an average slope of 6% (ENW, 2021; Tsiokanos et al., 2024). The catchment's elevation is illustrated in Figure 4. The river is prominently rain-fed and has an average discharge of approximately 3.2 $m^3/s$. Consequently, its discharge can vary enormously during flood and drought events (Tsiokanos et al., 2024). Due to the flash flood character of the 2021 event, the Geul area experienced higher impacts during and after the flood as compared to damages along the Meuse River (Endendijk et al., 2023; Slager et al., 2022). The upstream part of the Geul river is characterized by a relatively fast response to precipitation. The fast rainfall runoff originates from a low infiltration capacity caused by poorly permeable rocks in the subsurface. In addition, recent flood management measures (e.g. river widening and channel deepening) prevented flooding in the main branch of the Meuse (JCAR ATRACE, 2025-a; Slager et al., 2022).

The Geul catchment area is 340 $km^2$ and exists for 46% of grassland, 19% of arable land, 12% of buildings, 5% of roads, and 20% of forests (JCAR ATRACE, 2025-a). This catchment is characterized by small built-up areas in a rural area. Many small towns and villages are in the stream valley, where the river often flows through narrow village centers. As a result, the water level increases in the village centers at higher discharge compared to the water levels in the

rural area (ENW, 2021). Therefore, to reduce the flood risk, many rainwater buffers have been constructed throughout the catchment (JCAR ATRACE, 2025-a).

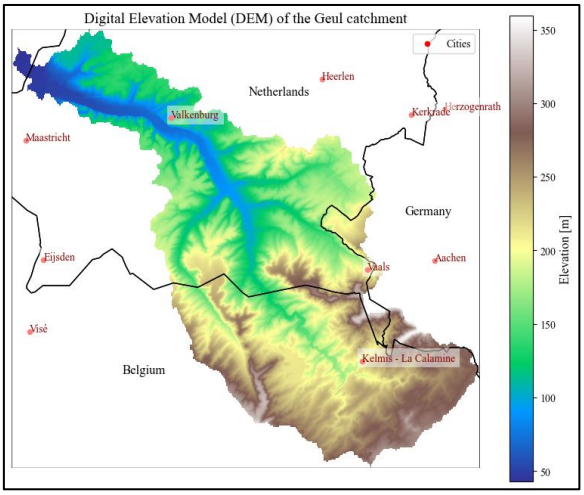


*Figure 4: Elevation map of the study catchment, including land boundaries and nearby cities.*

## 3.2 Early Warning Systems in the Netherlands

In the Netherlands there is an explicit distinction between major floods from the main waterways and local floods from regional waterways or local rainfall (Kok et al., 2017). Flood risk management for main waterways addresses the protection against main rivers such as the Rhine and the Meuse, the largest lakes such as the Ijssel Lake and the sea. When these main waterways flood, the consequences in terms of casualties and damages cannot be foreseen. On the other hand, the regional flood risk management focuses on the flood risk of smaller streams and canals, which are usually caused by local rainfall (K. De Bruijn et al., 2023).

Specifically, in the Netherlands, Rijkswaterstaat (RWS) is responsible for giving local authorities early warnings for high water levels or extreme events which are relatable to water management (Ministerie van Infrastructuur en Waterstaat, 2024b). RWS monitors the discharges and water levels of the main rivers in the country (Ministerie van Infrastructuur en Waterstaat, 2024a). Simultaneously, regional waterboards obtain precipitation forecasts of the KNMI through an automatic warning system when a critical precipitation limit is likely to be exceeded (KNMI, 2003). Both RWS and the local Waterboards have focused their EWS on rainfall and hydrological forecasts without information about impacts. This is confirmed by Merz et al. (2020) who described more in general that often only magnitudes of events are modelled in EWS and not impacts.

Although, it is acknowledged that a grey zone exists between the management of these two regimes. The distinction extends to the governance of protection standards and the geographical distribution of these responsibilities. For instance, the estimated exceedance probability of the July 2021 event did exceed the protection standards (causing flooding and damages) of the smaller local rivers but was within the protection range around the Meuse River (where protection infrastructure functioned well) (K. De Bruijn et al., 2023).

## 3.3 The Flood event of July 2021

In retrospect to the flood event in July 2021, extraordinary precipitation events between 12 and 15 July covered a large area of the Meuse and Mosel basins in Germany, Belgium and the Netherlands (ENW, 2021; K. De Bruijn et al., 2023). The meteorological conditions were characterized by a cutoff low-pressure system over Central Europe that supplied warm and very humid air masses to Central Europe from the Mediterranean in a rotating movement (Tradowsky et al., 2023). Simultaneously, with the low-pressure system slowly approaching from France towards Germany, the troposphere was increasingly unstably stratified (Tradowsky et al., 2023). Both meteorological drivers were forced upwards due to the presence of the western low mountain ranges (e.g. Eifel, Ardennes) causing damming effects. As a result, a cold pit originated where the air from a large area circulated in the direction of the pit. In the cold pit, the air went up due to the temperature gradient, cooled and rained out (ENW, 2021). This system caused heavy rainfall, where the most precipitation fell in the Ardennes Vestre River basin and in the southern Walloon part of the Geul river basin (Asselman et al., 2022). In addition, in the core of the weather system the rainfall peak amounts exceeded 250 mm (K. De Bruijn et al., 2023).

### 3.3.1 Meteorological conditions in the Geul catchment

In terms of rainfall intensity and atmospheric conditions, the weather system itself was not uncommon for the region during summer (Asselman et al., 2022). However, the stagnant position of the system over the area led to exceptionally high and prolonged precipitation, resulting in extreme cumulative rainfall totals (Asselman et al., 2022; ENW, 2021). The severity of the event in the Meuse catchment is illustrated by a Gumbel distribution of extreme precipitation values, with return periods exceeding 1:10,000 years in eastern sub-catchments of the Meuse (ENW, 2021). Between 13 and 15 July 2021, total precipitation in the Geul catchment ranged from 160 to 180 mm (Asselman et al., 2022; K. De Bruijn et al., 2023). On average, 128 mm fell within the 48 hours, corresponding to a return period of 1:500 years when orographic effects are considered (Asselman et al., 2022). Specifically, the (accumulated)

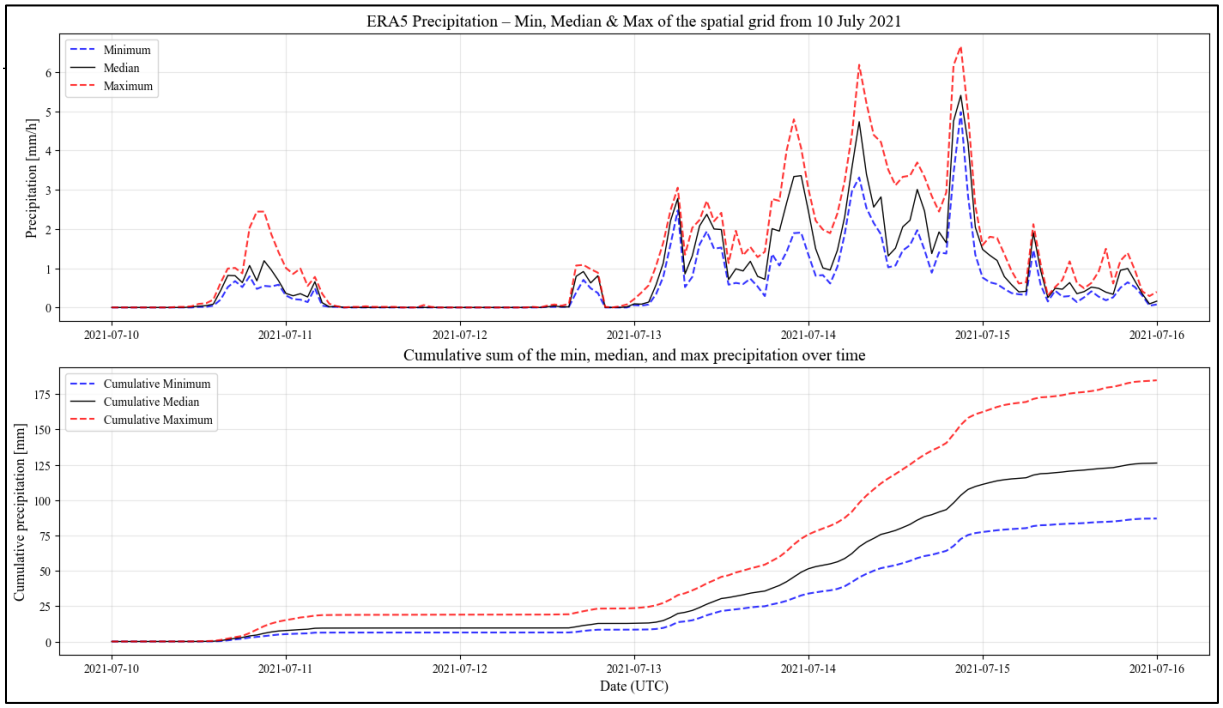


*Figure 5: The amount of rainfall in the Geul catchment from 10 July till 16 July 2021. The top figure visualizes the max, median, and minimum precipitation intensity [mm/h] that fell across the catchment per time step. The bottom figure depicts the maximum, median, and minimum cumulative sum of precipitation [mm] in the catchment.*

amount of rainfall that fell in the Geul basin is presented in Figure 5, based on ERA5 reanalysis data from ECMWF. Note the event had three high rainfall peaks between 14 and 15 July. In contrast, the average monthly July precipitation in Zuid-Limburg is around 80mm, highlighting the event's exceptional nature (Asselman et al., 2022).

In addition, until the forecast issued on 10 July, the extremeness of the event was not predicted. The first indications of rainfall appeared in the forecast on 11 July. From that point onward, predicted rainfall amounts and peak discharges were frequently revised upward. Consequently, the ENW evaluation report (2021) concluded that the real-time rainfall data from the KNMI radar was of insufficient quality, underestimating the cumulative amount of precipitation with approximately a factor three. Therefore, it is possible that probabilistic forecasting systems could have reflected these extremes (ENW, 2021). In addition, the evaluation report of Van Heeringen et al. (2022) demonstrates that only the deterministic forecasts ICON-EU from the German Weather Institute (DWD) estimated the extreme rainfall amounts for 11 July. The following days the ECMWF ensemble forecasts became more accurate, resulting in overall accurate ensemble forecasts. This means that the report concludes that there was at least a timely indication of very high precipitation initialized by ECMWF and the DWD. Although it must be acknowledged that for this specific event the ECMWF forecasts were behind to those of the DWD, where the DWD gave an indication much earlier (Van Heeringen et al., 2022). Ultimately, it appears that different weather forecast institutes predict different outcomes for the days before the flood, where the KNMI did not came close to the observed rainfall with their forecasts (ENW, 2021). The DWD provided the first indication of extremeness on 11 July at 00:00 UTC, after which it underestimated the event until 12 July. Lastly, the ensembles of the ECMWF underestimated the event till 11 July 12:00 UTC (Van Heeringen et al., 2022).

### 3.3.2. Extreme water levels and peak discharges in perspective of the Geul catchment

This record-breaking precipitation resulted in record high discharges in the Meuse and its tributaries. The peak discharge in the Meuse arrived in Eijsden at 22:00 on 15 July and took approximately 113 hours to travel to the most downstream point of the Meuse near Rotterdam. Simultaneously, as the peak of the Meuse reached the confluence with the Geul river, the discharge peak of the Geul also arrived at this junction (ENW, 2021). As a result, this temporal overlap restricted effective drainage, leading to elevated water levels in the Geul estuary. These water levels likely contributed to the downstream impacts of the event (Asselman et al., 2022).

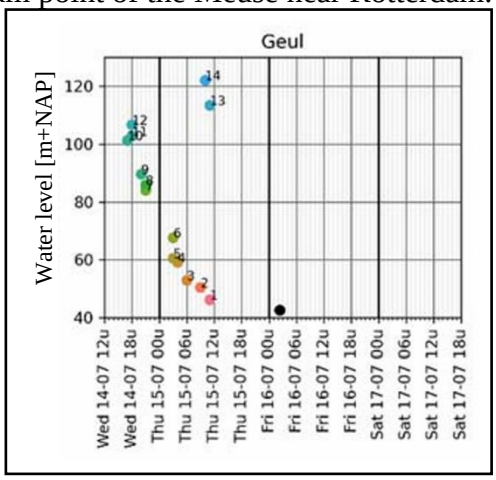


*Figure 6: Moment and height of the peak water level along the Geul River at different measuring stations (each a different number and colour). From upstream (number 14, blue) to downstream (number 1, Pink) and in black the moment and height of the water level at the measuring station in the Meuse near the outlet of the Geul. Obtained from ENW (2021)*

Furthermore, the timing of maximum water levels across the catchment varied, as illustrated in Figure 6. The figure demonstrates that peak water levels upstream in the Geul (measuring stations 13 and 14) occurred later than those downstream due to a secondary rainfall peak. The highest water levels in the Geul were recorded between the evening of 14 July and the morning of 15 July, where the first peak occurred around Mechelen at 18:00 on 14 July. This peak resulted in water levels higher than the T100 water level in the Geul.

As a result, the actual water levels during the event exceeded the water levels associated with the designed return period of T25 (ENW, 2021). This indicates that the event falls within the scope of accepted flood risk since the water levels were above the T25 return period. Ultimately, the tributaries of the Meuse such as the Geul, Geleenbeek, and the Roer had a water level exceedance probability of approximately 1:100 till 1:1000 years (ENW, 2021).

### 3.3.3 Evacuation and taken measures in the Geul catchment

As water levels in the Geul rapidly exceeded design thresholds, authorities were forced to respond under extreme and unforeseen circumstances, especially in the tributaries of the Meuse. In these catchments existing disaster protocols were not prepared for events of this severity, leaving villages and areas such as Valkenburg not evacuated. Consequently, many residents were caught by surprise during the night of 14 and 15 July 2021 when water levels rose very fast (ENW, 2021). In Figure 7 the timeline of the warnings, communication and response in the Geul catchment is demonstrated. Based on the timeline, the start of the flood occurred at 22:45 on 14 July 2021.

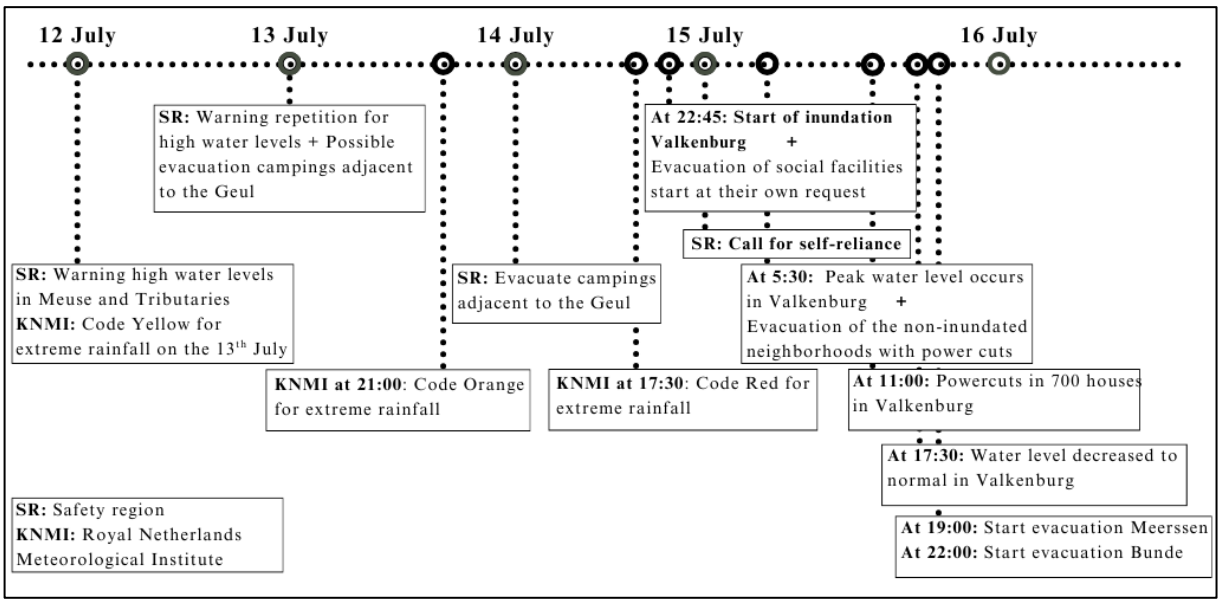


*Figure 7: Timeline of warnings, communication, and response in the Geul catchment, created from the timeline of ENW (2021).*

In general, the warning communication was disseminated through a combination of formal and informal channels, including municipal authorities, safety regions, NL-Alerts, social media, police loudspeakers, church bells, and the regional broadcaster 'L1'. In addition, many people received informal warnings via WhatsApp groups or through family or acquaintances living upstream (ENW, 2021).

Although official warnings were issued by the Safety region Zuid-Limburg at approximately 17:00 and 22:30 on 14 July. These messages emphasized the potential danger of the forecasted rainfall and rising water levels, advising to remain indoors. However, they lacked critical information such as expected flood depths, spatial flood extents, and impact zones. Within the scope of the existing protocols, evacuations were carried out in high-risk zones directly adjacent to the river throughout 14 July. As a result, several campsites and social facilities were evacuated in Valkenburg and Meerssen (ENW, 2021).

Despite these efforts, the majority of residents in the Geul valley were caught by surprise. The floodwaters had particularly impacts on communities in Valkenburg and Meersen, as well as downstream areas along the Juliana channel in the villages Bunde and Geulle. Moreover, two neighbourhoods in Valkenburg experienced power cuts and approximately 1000 people were

exposed to flooding and only evacuated after water had already flooded the area (ENW, 2021). Furthermore, respondents of the survey by Endendijk et al. (2023) indicated that the respondents living in the more upstream flooded area of the Geul River were also caught by surprise.

### 3.3.4 Impacts in the Geul catchment

While large-scale evacuations were limited and often delayed, the severity of the flood led to several physical impacts. The primary driver of the extensive damage was the exceptional precipitation that fell over the catchment (ENW, 2021). The rainfall in combination with the fast rainfall runoff processes, and the hilly terrain caused high discharges in the Geul river, which ultimately, resulted in the flooding of floodplains. Due to the presence of low protected villages in the stream valley, the impact of the event was relatively high for agriculture and residential buildings (Endendijk et al., 2023; ENW, 2021; JCAR ATRACE, 2025-a). In total, the extreme flood event of July 2021 in North-Western Europe caused 230 fatalities and around €40bn of damage (Lehmkuhl, et al., 2022). Fortunately, unlike in Germany and Belgium, the flood event did not result in direct casualties in the Netherlands. However, the observed flow velocities, water depths exceeding 1.5 meters, and the rapid water level rise in several parts of the Geul valley indicate that these extreme conditions could have led to more casualties (ENW, 2021; Slager et al., 2022).

Nevertheless, the event caused severe disruption throughout the catchment, even though the duration of the flood was relatively short (not longer than one day; Figure 7). The impacts were dependent on the timing of peak flow, increase in water depth and velocity (ENW, 2021). Moreover, as the limited early warning and the flash flood character affected the response ability, different stakeholders in the catchment suffered high losses due to peak seasonal usage during the summer period. Examples are damages to agricultural lands and campsites (ENW, 2021).

In addition, by early August, insurers had reported nearly 13,000 damage claims from Limburg, with the majority originating from the different tributaries. Ninety percent of the damage claims were related to private homes and vehicles. Moreover, approximately 600 businesses were affected by the flood, of which around 70% were located along the Geul. Ultimately, the total damage in the Netherlands is estimated in the order of € 350 – 600 million (ENW, 2021).

Looking at the impacts after the flood event, Endendijk et al. (2023) investigated that the disaster prevention was effective during the flood, as only households located in areas without dikes have been flooded. Those who received a warning employed more flood damage mitigation measures than households without a warning. These adaptation measures reduced the damage by approximately 20% to 50%.

### 3.3.5 Post-flood actions in the Geul catchment

After the flood event of July 2021, the safety regions Limburg-Noord (VRLN) and Zuid-Limburg (VRZL) have prepared together with Rijkswaterstaat and Waterboard Limburg a disaster response plan for anticipating high water levels in the main rivers and tributaries of the province. The goal of the plan is to have a close collaboration with the local authorities and emerging services, whereby the consequences of future high-water levels will be prevented as much as possible. The plan will support the operational deployment by defining preventive actions and responses based on the discharges of the rivers and scenarios (Veiligheidsregio Zuid-Limburg et al., 2024).

Moreover, in response to the summer floods of 2021, different research programs are initiated, including a Joint Cooperation program for Applied scientific Research on flood and drought

risk management in regional river basins (JCAR). The program's goal is to Accelerate Transboundary Regional Adaptation to Climate Extremes (ATRACE) by evaluating impacts and implications, aiming to improve preparedness and to support collaboration between partners in the transboundary basins of Belgium, Germany, Luxembourg and the Netherlands (JCAR ATRACE, 2025-b). This research will contribute to JCAR by focusing on the development of IbF based on precipitation ensemble forecasts.

Furthermore, authorities currently assess measures to mitigate impacts and increase preparedness for similar extreme events. It can be concluded that the exceptional magnitude and spatial extent of the July 2021 floods, along with their severe impact on neighboring countries, heightened the awareness that a similar hydrometeorological hazard occurring in a different region in the Netherlands could have caused much larger impacts (K. De Bruijn et al., 2023). In response to this awareness, the development of this IbF for the Geul catchment could hypothetically lead to a wider application of impact-based models in the Netherlands, what will contribute to improved responses.

# 4. Methodology

This chapter outlines the methodological framework created to investigate the development of an IBFFWS for the Geul catchment. The aim of this chapter is to provide an overview of the methods and models used for data collection, processing, interpretation, evaluation, and limitations and uncertainties. Figure 8 represents the overall study setup in a flow chart. In general, this study applies different models from the Geographical, Environmental, and Behavioural (GEB) model framework (J. A. De Bruijn et al., 2023).

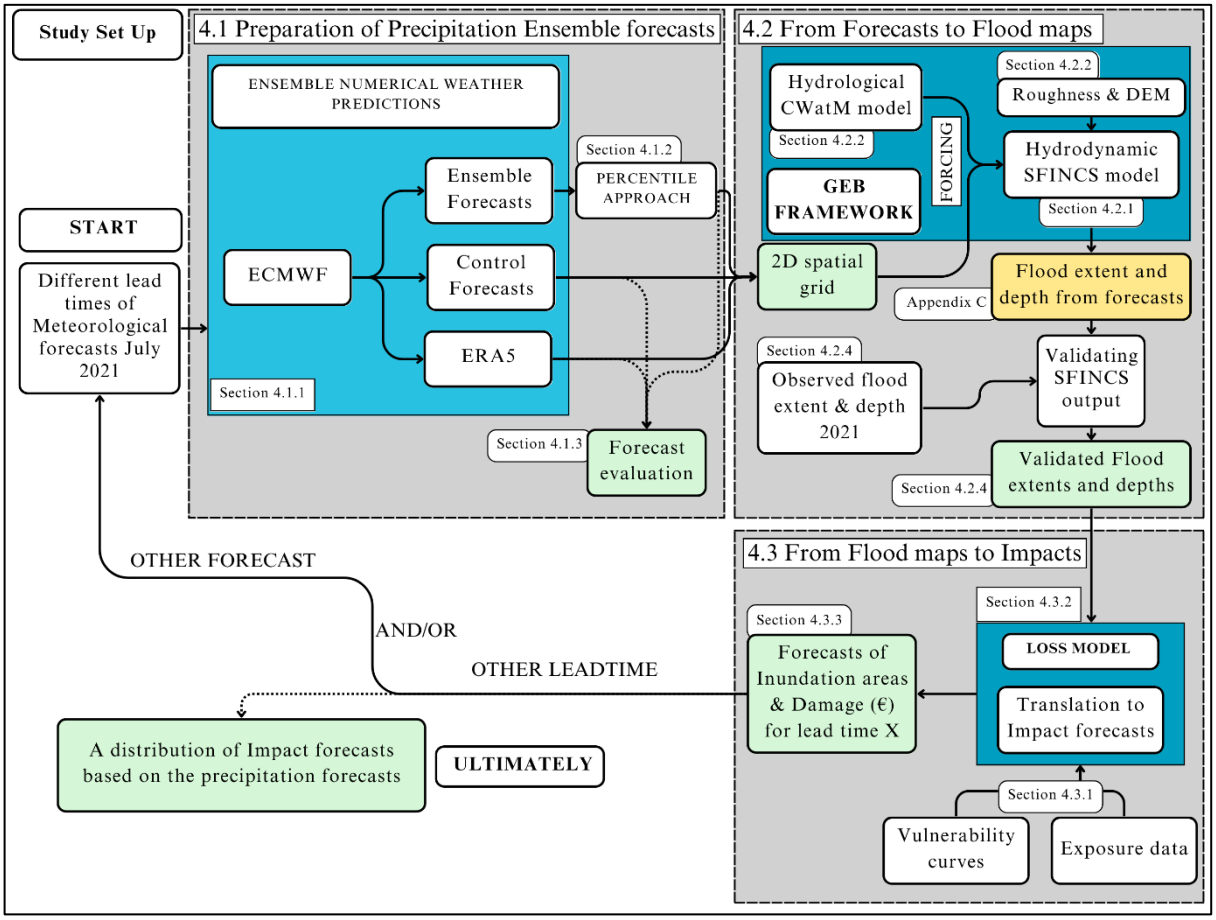


*Figure 8: A flow chart of the study, where the different colours depict different products used or outcomes, with blue representing the different models, yellow the calibration, and green representing the output. Specifically, the dark blue boxes are integrated in the GEB model.*

Furthermore, in Figure 8 each grey-striped box represents a step in the study: 'The preparation of precipitation ensemble forecasts' (Section 4.1), 'From forecasts to flood maps' (4.2), and 'From flood maps to impacts' (4.3). The blue boxes represent different models, where the dark blue boxes are models which are part of the GEB framework. Moreover, the green boxes represent outputs from this study and the yellow box represents the improvement of the SFINCS outcomes (Appendix C). Each step has different input data, which is specified in each section (4.1.1, 4.2.2, and 4.3.1).

As the study is the development of a model chain, the output of one box will be input data for the next step in the chain. As a result, the processed rainfall grids (4.1) together with the CWatM model will be used for forcing the SFINCS hydrodynamic model of the Geul catchment (4.2). Subsequently, the flood extents will be processed and evaluated before it will be used in a loss model (4.3). This chain is developed in the GEB framework (4.2). Ultimately, this framework iterates over different forecasts, varying by a decreasing lead time up to the flood event. Moreover, the framework has been tested under conditions that reflect operational forecasting practices, ensuring the relevance of the results for real-world applications. The elaboration of the methodological framework now starts with an explanation on how the precipitation forecasts are prepared to be used in the IbF model.

Lastly, SFINCS is initialized with initial soil moisture conditions and daily ERA5 precipitation data from the CWatM model. However, the initialisation of CWatM is not part of the scope of this study. The study of Burek et al. (2020) elaborates the coupling of CWatM within GEB, this study can be used for more information. Most important is that the modelling of hydrological conditions in this study is provided with CWatM and proven with the studies of Burek et al. (2020) and (J. A. De Bruijn et al., 2023). The conditions of CWatM are used in SFINCS but are not elaborated in this study.

## 4.1 Preparation of Precipitation Ensemble forecasts

The first step in the methodological framework involves the preparation of precipitation input data for all subsequent hydrodynamic and impact modelling. Moreover, this section describes the procedures for the data collection, cleaning, and preprocessing of the precipitation data. As one of the objectives of this study is to investigate the skill of ensemble forecasts compared to the ERA5 dataset from ECMWF, different forecasts are processed with varying lead times.

### 4.1.1 Forecast data

In this study different precipitation timeseries are used for forcing the CWatM hydrological model and SFINCS hydrodynamic model. First, the ERA5 dataset will be explained, whereafter the ensemble forecasts of the ECMWF are outlined.

**ECMWF ERA5 dataset**

To establish a reliable baseline for the July 2021 event, this study uses the ERA5 reanalysis dataset from ECMWF, hereinafter referred to as observational data. This dataset combines global observations with model data to provide hourly and daily gridded precipitation estimates (Hersbach et al., 2020; ECMWF & Copernicus, 2025). The daily precipitation grids are used to warm up the framework and to initialize hydrological conditions in CWatM, whereafter hourly values are used inside the SFINCS model. Therefore, precipitation data from 1 January to 31 July 2021 was extracted and clipped to the Geul catchment. The selected grid spans longitudinal steps of 0.1 degrees from 5.7-6.1 degrees in the x-direction and latitudes from 50.7-50.9 latitudes in the y-direction, resulting in 15 grid cells with an approximate horizontal resolution of 6 km and vertical resolution of 11km. Ultimately, the hourly ERA5 precipitation data is used to simulate the event from 11 to 16 July 2021 and serves as the baseline for the evaluation of model performance under observed conditions. Additionally, daily ERA5 data is used in the

warming-up (spinning) phase of the framework where the CWatM model sets the hydrological conditions of the catchment.

**ECMWF ensemble forecasts**

In addition to ERA5, this study uses ECMWF perturbed ensemble precipitation forecasts from the Ensemble Prediction System (EPS), version CY47R2 (ECMWF, 2025d) to force SFINCS. This version was selected to ensure that forecasts were issued before the event (historic operational forecasts), hereby avoiding the nature of hindcasts. For this reason, the evaluation of the forecast skill is improved as new model updates from the ECMWF are not included and without extra reanalysis. Forecasts were retrieved through the ECMWF operational archive with an API request where perturbed and control forecasts are downloaded. Together they represent an ensemble forecast with 50 perturbed forecasts and one control forecast. Each forecast spans 15 days, with hourly timesteps for the first 90 hours, three-hourly timesteps from 93 to 144 hours, and six-hourly intervals from 150 till 360 hours (ECMWF, 2025b). For the first 90 hours the forecasts are updated each six hours following a dissemination schedule (ECMWF, 2025a).

Furthermore, the spatial resolution of the forecasts is defined within the API request. This study employs the grid F640 Gaussian grid, which was used in the operational forecasts in the days leading up to the 2021 flood event. The F640 grid contains 640 lines between the equator and the poles, translating to a vertical resolution of approximately 0.14 degrees (~15.6 km), with a comparable horizontal resolution of approximate 8 km at 50 degrees latitude.

Utilizing both control and perturbed forecasts facilitates a more robust understanding of the hydrological response to forecasted precipitation. Since flooding in Valkenburg began on 14 July at 22:45 (section 3.2), eight forecasts were selected from 11 July to 14 July, including both forecast initialization times of 00:00 and 12:00. As a result, the study' lead times range from approximately 5 hours to 89 hours for both control and ensemble forecasts. Table 1 outlines these lead times and dissemination times, which vary depending on whether hourly or extended three-hourly forecast intervals are applied (ECMWF, 2025a).

The forecast data is downloaded with an API request through the 'ecmwfapi' package, specifying parameters for both control and perturbed forecasts as detailed in Appendix A. The request accesses the ECMWF's operational archive 'od', selecting the surface-level 'sfc', and specifies the total precipitation (cumulative) as the variable of interest along with other more technical and geographical parameters. Note that this study uses the operational archive instead of the hindcasts to use the forecasts which were available for users at that moment in time (see Discussion).

*Table 1: Overview of lead times with and without considering the dissemination schedule of ECMWF (2025a) and based on the determined start of flooding on 14 July 22:45 from section 3.2.*

| *FORECAST* | *INITIAL LEAD TIME [hours: minutes before]* | *FORECAST TIME TO DISSEMINATE TO USERS* | *LEAD TIME IN REALITY [hours: minutes before]* |
|---|---|---|---|
| *11 July 2021 00:00* | 94:45 | 06:27 | 88:18 |
| *11 July 2021 12:00* | 82:45 | 18:27 | 76:18 |
| *12 July 2021 00:00* | 70:45 | 06:27 | 64:18 |
| *12 July 2021 12:00* | 58:45 | 18:12 | 52:33 |
| *13 July 2021 00:00* | 46:45 | 06:12 | 40:33 |
| *13 July 2021 12:00* | 34:45 | 18:12 | 28:33 |
| *14 July 2021 00:00* | 22:45 | 06:12 | 16:33 |
| *14 July 2021 12:00* | 10:45 | 18:12 | 4:33 |
| ***Start of flooding*** | 14 July 2021 22:45 | - | 14 July 2021 22:45 |

### 4.1.2 Processing input data

Before the forecast data is used as input in the framework it is first processed using a standardized Python workflow. Each API request returns a GRIB-format file, where specifically for forecasts from 11 July and 12 July additional data requests were necessary. Since these forecasts did not cover 15 July with hourly timesteps entirely, the lacking hours are filled with 3-hourly timesteps to extend the coverage up to at least 00:00 on 16 July. Additionally, forecasts issued at 12:00 lacked precipitation for the first 12 hours of the day. To meet the input requirements of SFINCS zero precipitation grids with similar dimensions were inserted for the missing period to prevent introducing artificial rainfall. Regarding the ERA5 baseline simulation in SFINCS, the processing steps are not needed as the data already captures the entire month July (see Section 4.1.1).

In addition, the forecasts that required extension are merged to get consistent input data. As the original forecast files contain cumulative precipitation in meters over time, all the data is decumulated and converted to millimeters. When three-hourly timesteps or zero-precipitation grids are inserted, the forecasts are carefully aligned and converted to hourly intervals by interpolating to get a consistent timestep.

Furthermore, to efficiently incorporate the entire perturbed ensemble forecasts of 50 members, this study defines an ensemble percentile approach. For each grid cell, the $25^{th}$, $50^{th}$, $75^{th}$, $90^{th}$ and $95^{th}$ percentile are calculated across the ensemble. This new method describes the uncertainty and spread within the ensemble but significantly reduces the computation time in the next modelling steps. Without this step, all the 50 members would have been processed individually. An example of the spatial averaged ensemble members is presented in Figure 9.

Once each forecast meets the required temporal coverage, the spatial grids are clipped to the catchment boundaries before being used as input in the SFINCS hydrodynamic model.

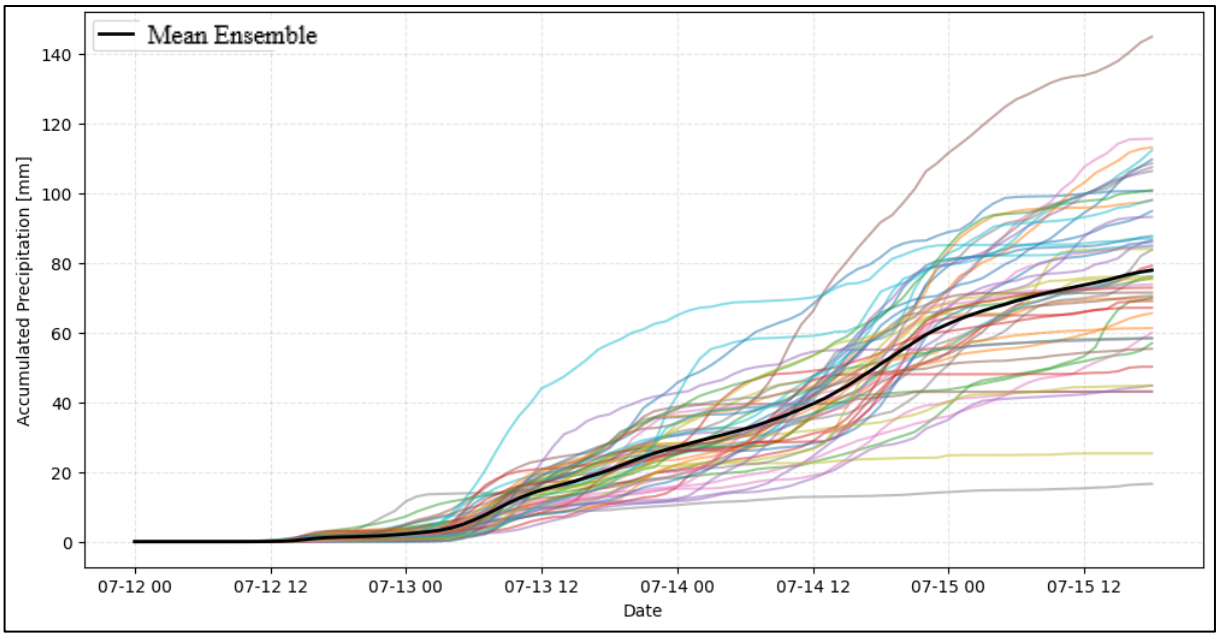


*Figure 9: An example of a spatial averaged ECMWF ensemble precipitation forecast with 50 members (in distinct colours), and the mean forecast created from the members (in black). This example is a perturbed forecast initialized on 12 July 2021 00:00.*

### 4.1.3 Evaluating the different forecasts

To evaluate the quality of the new percentile approach, a combination of metrics is applied to examine the overall performance of the forecasts, the accuracy, the reliability, and the spatial representation of the forecasts. Each metric targets a specific aspect of forecast quality to ensure a comprehensive comparison. Since the ERA5 dataset is used as observational data, this study investigates for every lead time the quality of each perturbed percentile and control forecast compared to the ERA5 dataset. Based on the outcomes of the forecast evaluation, the first sub-question RQ1 can be answered: "How well do historic operational ensemble and deterministic forecasts relate to the observed precipitation patterns and totals of the July 2021 flood event, compared to ERA5 reanalysis data?

As a first step in the evaluation, scatterplots were used to visually compare the ensemble percentile and control forecasts against the ERA5 observations (WCRP & WWRP, 2017). To ensure clarity, only the maximum forecasted value across the spatial grid were included per percentile and lead time in the analysis. For the control forecasts, all individual grid cells are included per lead time and presented in different colors. This visualization initializes the subsequent more statistical metrics (PIT, CRPS, MAE, Bias, and FSS) by highlighting the forecast-observation relationship.

***Evaluating the reliability of the Ensemble percentile forecasts:***

Following the visual comparison with scatterplots, the reliability of the ensemble percentile forecasts is evaluated using Probability Integral Transform (PIT) histograms. These histograms are essential for evaluating how well the spread of the percentiles corresponds to the observed data. This metric evaluates whether the percentile forecasts are under dispersed, over dispersed or exhibit systematic biases. Therefore, a PIT histogram reveals the relation between the observed frequency against the forecast probability (Crochemore et al., 2016; WCRP & WWRP, 2017; Yang et al., 2021).

To construct the PIT diagrams, PIT values were calculated for each lead time and percentile by comparing observations with the corresponding ensemble forecasts. Specifically, for each timestep, the rainfall intensity values were sorted per percentile, and the rank of the observed value within this sorted list was determined. Then, this rank was normalized by dividing it by the total number of ensemble members to obtain the PIT value. This process was applied for each grid cell and time step across all the lead times and percentile members. To assess the overall model performance, the PIT values were averaged across all grid cells, which resulted in the creation of the PIT histograms. When the members are close to the 1:1 line it means that it perfectly fits the ERA5, when it is below the line it is under dispersed and when the member is above the 1:1 line it is over dispersed and thus systematically biased (Crochemore et al., 2016; WCRP & WWRP, 2017; Yang et al., 2021).

***Evaluating the overall performance of the Ensemble percentile forecasts:***

Subsequently, the overall performance of the ensemble forecast is calculated with the continuous ranked probability score (CRPS). The CRPS quantifies the squared difference between the cumulative distribution of the forecast and the observed outcome, represented as a step function. A lower CRPS indicates higher forecast skill, as it reflects a closer alignment between the predicted and observed distributions (Crochemore et al., 2016). In this study, the CRPS score is calculated for the rainfall intensity per grid cell for each lead time based on the ensemble percentiles, following the equation derived by Hersbach (2000) (see Equation 4.1).

$$CRPS_{(i,j)} = \frac{1}{N}\sum_{k=1}^{N}\left|x_k^{(i,j)} - x_{obs}^{(i,j)}\right| - \frac{1}{2N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left|x_k^{(i,j)} - x_l^{(i,j)}\right| \quad (4.1)$$

(i, j): Grid cell dimensions

N: Total amount of percentiles

$x_k$: Forecasted value for percentile k

$x_l$: Forecasted value for percentile l

$x_{obs}$: Observed value in ERA5

Equation 4.1 is employed in a standardized python workflow using the ‘crps_for_ensemble’ function, which interprets ensemble forecasts as an empirical Cumulative Distribution Function (CDF). Figure 10 illustrates the CRPS calculation, where the grey area between the orange forecast CDF and the blue observation CDF represents the CRPS score. Before the CRPS score is calculated across lead times, the spatial resolution of the percentile ensemble is first aligned with the resolution of ERA5. Subsequently, CRPS values are computed for each grid cell and time step, and then averaged over time to generate a spatial CRPS map. These

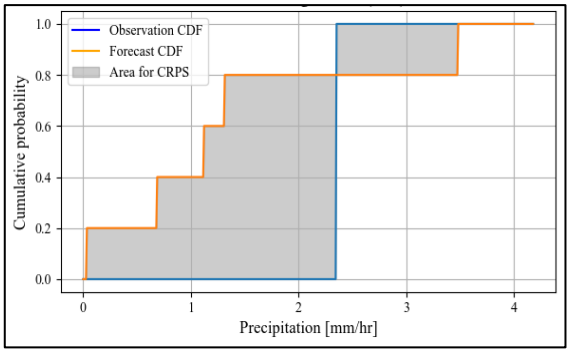


*Figure 10: An example of how the CRPS value is calculated based on the CDF curves of the ensemble forecast and the observation per time step and per grid cell.*

maps are then used to derive a lead time-based line plot, which summarizes the average CRPS over time and grid per lead time.

***Evaluating the accuracy of the Ensemble percentile & Control forecasts:***

Following the evaluation of ensemble spread, it is important to assess the accuracy of the forecast values compared to observations. To compare the different forecasts with the observations, the accuracy of the control and ensemble percentile forecasts are quantified by calculating the Mean Absolute Error (MAE) and Bias. These metrics are calculated for each grid cell, enabling the identification of regional variations in forecast performance. This allows for a fair and consistent comparison between the rainfall intensity of the deterministic control forecast and the percentiles of the ensemble forecast. The equations of these metrics are outlined in Equation 4.2 and 4.3, with the goal to give insights in the magnitude of the error per grid cell and if the forecast is underestimated or overestimated (Crochemore et al., 2016; WCRP & WWRP, 2017; Yang et al., 2021).

$$MAE = \frac{1}{N}\sum|F_i - O_i| \tag{4.2}$$

$$Bias = \frac{\frac{1}{N}\sum F_i}{\frac{1}{N}\sum O_i} \tag{4.3}$$

N: Number of measurements

$F_i$: Forecasted value

$O_i$: Observed value

Subsequently, for the ensemble percentile forecasts, the MAE and Bias are averaged across all the ensemble percentiles to produce a single spatial map for each lead time. In addition, the standard deviation of both metrics is derived to illustrate the variability in forecast accuracy among the different percentiles.

***Evaluating the spatial representation of the Ensemble percentile & Control forecasts:***

Moreover, traditional verification metrics such as MAE compare forecasts and observations individually at each location. This becomes a problem when an event such as a heavy rainfall is forecast slightly off the location where it occurred. At that moment, a double penalty arises as a forecast will be doubly penalised once for missing the feature in the correct spot, and once for the false alarm in the wrong place (Haiden & Lledó, 2023). Therefore, the spatial evaluation of the forecasts is quantified with the Fractions Skill Score (FSS) metric. This metric is widely used for threshold-based events such as heavy rainfall where it benefits in including both the spatial distribution and displacement errors. (WCRP & WWRP, 2017; Haiden & Lledó, 2023). In this study, the FSS is applied for both the control forecast and the ensemble percentiles ($25^{th}$, $50^{th}$, $75^{th}$, $90^{th}$ and $95^{th}$) by comparing the predicted and observed values per grid cell. The calculated scores, ranging from 0 (no skill) to 1 (perfect agreement), allow for an objective comparison of how well each forecast captures the spatial structure of the event over different lead times. This spatial evaluation complements other metrics and supports the assessment of which forecast member or percentile best represents the observed rainfall.

The calculation of the FSS is standardized in a structured python workflow. First, all datasets are synchronized in both spatial resolution and time to ensure consistent spatial and temporal alignment with the ERA5 data. Subsequently, the FSS is calculated for the deterministic control forecast by evaluating each grid cell over time. This is done by binarizing precipitation values using a different rainfall intensity threshold (e.g., 0.5 mm/h) and comparing the exceedance frequencies between the forecast and observed values (See equation 4.4). The result provides a map with FSS scores per grid cell (Necker et al., 2024).

$$FSS_{control} = 1 - \frac{(\sum_{i=1}^{I}(NP_{in,f} - NP_{i,o}))^2}{\sum_{i=1}^{I}(NP_{in,f})^2 + \sum_{i=1}^{I} NP_{i,o}^2} \tag{4.4}$$

I: Total number of grid cells across the study area

$NP_{in,f}$: The number of grid cells above the threshold from the deterministic forecast

$NP_{i,o}$: The number of grid cells above the threshold from the observations

For the ensemble forecasts, a probabilistic FSS (pFSS) is computed. In contrast to the FSS control, precipitation values for each percentile are binarized and aggregated to determine the frequency with which rainfall exceeds the thresholds. This aggregated frequency is then compared to observations to calculate the pFSS per grid cell, accounting for the spread within the ensemble (see equation 4.5) (Necker et al., 2024).

$$pFSS = 1 - \frac{\sum_{i=1}^{I}(\sum_{n=1}^{N} NP_{in,f} - NP_{i,o})^2}{\sum_{i=1}^{I}(\sum_{n=1}^{N} NP_{in,f})^2 + \sum_{i=1}^{I} NP_{i,o}^2} \tag{4.5}$$

N: Total number of ensembles

I: Total number of grid cells across the study area

$NP_{in,f}$: The number of ensemble members predicting the event above the threshold at each time step each grid cell

$NP_{i,o}$: The number of observations of event $i$ for grid cell $i$

Once both FSS control and pFSS values are stored as spatial data arrays for each lead time, this approach enables a robust comparison between spatial deterministic and probabilistic forecast performance. Additionally, to provide an overview, multiple thresholds (0.5, 1, 3, 4 mm/h) are applied, and summary statistics are derived from the FSS maps. For each threshold, the median, maximum, and minimum scores are extracted per lead time and visualized using boxplots.

Ultimately, a quantitative summary of the metrics is presented in a table to compare the different scores. These metrics are averaged over both the spatial and temporal dimensions of the forecast and are based on the rainfall intensity. Once the overview is created, it becomes possible to examine which lead time best represents the observed rainfall from ERA5.

## 4.2 From forecasts to flood maps in the GEB framework

The precipitation forecasts from 4.1 are applied within the GEB framework. GEB integrates, and couples different models, including water cycle models (CWatM, MODFLOW), Human behaviour models (DYNAMO, ADOPT), a vegetation dynamics model (plantFATE), and a hydrodynamic model (SFINCS). These models can function independently or in combination depending on the research design (J. A. De Bruijn et al., 2023).

This study applies the GEB framework to simulate hydrological processes for the Geul catchment for a 10-year period (warming up period) in CWatM, whereafter SFINCS is initialized to simulate the flood extent and depth. While the framework handles the model coupling, the core of the SFINCS setup remains unchanged. This allows the study to focus on testing the effects of forecast input on hydrodynamic processes.

Although the rainfall data were pre-processed, GEB required adjustments to be able to simulate ensemble forecasts without changing the initial conditions of the model. To efficiently run multiple ensemble members within SFINCS, a custom function is developed (named 'Multiverse') within the GEB framework to automate the execution of the different forecast percentiles. This function stores the initial hydrologic conditions from CWatM and imports the processed forecast before SFINCS is initialized. After the conditions are stored, the function runs each forecast percentile as an individual model simulation. When a percentile run is completed, the function automatically restores the initial conditions, whereafter another percentile can be simulated with the same initial conditions. As a result, the function ensures that each forecast member starts with the same constant baseline, which leads in the end for a better comparison between percentiles. In addition, it prevents the influence from previous simulations on future results.

### 4.2.1 The Super-Fast INundation of CoastS (SFINCS) Model

Building on the framework overview, this study applies SFINCS (v2.1.3) to simulate flood extents and flood depths resulting from forecasted precipitation. The model is developed to efficiently simulate hydrodynamical processes such as compound flooding events at limited computational cost and good accuracy (Leijnse et al., 2021). To keep the efficiency, the model solves Simplified Shallow Water Equations (SSWE) or Local Inertial Equations (LIE). These equations are part of a reduced-physics model derived from the full momentum and continuity Saint-Vernant equations. The difference between the two equations is that SSWE incorporates the advection term to compute wave dynamics and fluxes over a spatial grid in more detail, while the LIE excludes this term for increased computational efficiency (Leijnse et al., 2021).

Moreover, SFINCS further includes spatially varying processes such as infiltration, precipitation, and surface roughness, which are essential for simulating compound flooding (e.g. pluvial, fluvial, tidal components). Although the model excludes atmospheric pressure gradients and the Coriolis effect, SFINCS remains the first reduced-physics model to include all these processes, which makes it very useful for flood risk assessments (Leijnse et al., 2021). The numerical equations used for simulating the hydrodynamical processes are presented in Appendix B.

On a different note, the model consists of a rectangular staggered 2D grid where bed levels, water levels and water depths are defined in the grid cell centres and fluxes are incorporated as velocity points. This water level is dependent on the calculation of the fluxes between the grid cells. An example of the structure of the model grid is outlined in Figure 11.

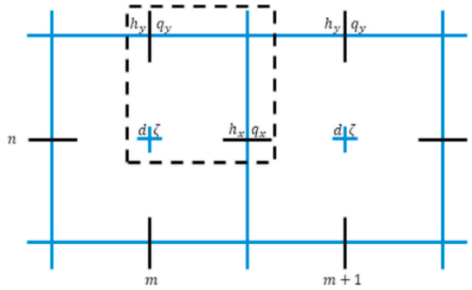


*Figure 11: SFINCS rectangular grid with bed level (d) and water level (ζ) in the cell centres (+) and water depth (h) and fluxes (q) in velocity points. Variables within the dashed box share the same grid indices, which are indicated with m and n for x- and y-direction. Obtained from Leijnse et al. (2021).*

```
-----------------------------------------
Processes
-----------------------------------------
Subgrid topography    : yes
Quadtree refinement   : no
Advection             : yes
Viscosity             : yes
Coriolis              : no
Wind                  : no
Atmospheric pressure  : no
Precipitation         : yes
SnapWave              : no
Wave paddles          : no
-----------------------------------------
```

*Figure 12: Included processes in the SFINCS model for this study.*

Furthermore, the advantage of SFINCS is that several types of forcing data can be applied, examples of forcing data are storm surge, precipitation or upstream discharge (Leijnse et al., 2021).

This study uses precipitation as forcing data since the Geul is a headwater catchment with no river or tidal inflow points. Therefore, different processes explained by Leijnse et al. (2021) are excluded in this study as depicted in Figure 12. Specifically, the processes more related to coastal flooding (e.g. wave paddles, snap wave, and wind) are excluded. Moreover, in the simulation the sub grid topography is active as shown in Figure 12. This process conducts sub grid-corrections, which makes it possible to run the model on a coarser resolution without losing small scale effects such as roughness variations. Van Ormondt et al. (2025) found that these corrections significantly improve the model accuracy, especially in complex areas where small-scale variations have a major impact on water movement. Since the Geul is a meandering tributary, sub grid corrections are crucial in this study. Additionally, the sub grid accelerates the computation time.

Ultimately, the model generates various hydrodynamic outputs throughout the simulation period. However, this study only uses the maximum flood map over the time to evaluate the maximum flood extent and depth. Hence, SFINCS is chosen because of its short computation time but still its proven high accuracy in coastal compound flooding hazards (e.g. Beveren, 2022; Deltares, 2025b; Eilander et al., 2023; Leijnse et al., 2021), and in estimating flood damages (e.g. Sebastian et al., 2021). This makes it possible to run the different rainfall forecast scenarios processed in Chapter 4.1. However, using SFINCS in headwater catchments such as the Geul has not yet been tested (Deltares, 2025b), therefore this study also tests the simulation of pluvial and fluvial floods in small tributaries.

### 4.2.2 Input data for SFINCS

Apart from the different processes in the model, the user's input such as initial conditions, boundary values, or parameter settings are specified in a text-based input file. This input serves to initialize the general mathematical framework for solving the equations elaborated in the previous section and to generate output based on the initiated settings (Leijnse et al., 2021). This study uses a coarse grid resolution of sixty meters, with a sub-grid correction of twelve

meters resulting in a 5-meter grid resolution. Moreover, the simulation period for the precipitation forecasts is dependent on the start time of the forecast but ranges from 12 hours till a maximum of four days. Besides, ERA5 data is used in a simulation period from 1 July till 16 July.

Beyond the warm-up period of the hydrological processes in CWatM and the processing of the rainfall forecasts, SFINCS also requires landcover data to determine spatial roughness, a DEM to determine the bed level, and a river network to complete the bathymetry.

***ESA landcover & Manning Roughness:***

First, land cover data is derived from the European Space Agency (ESA) World Cover 2021 dataset with a spatial resolution of 10 meters (Zanaga et al., 2022). The catchment contains mostly grassland, cropland, built-up, and tree cover as presented in Figure 13.

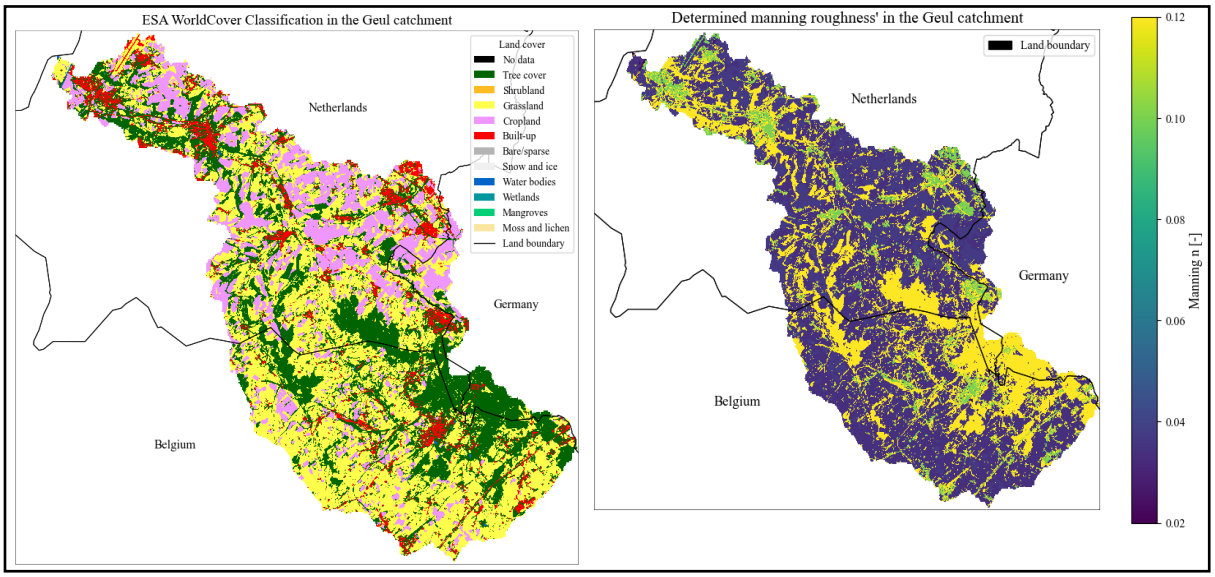


*Figure 13: With right the Landcover classification of the Geul catchment obtained from Zanaga et al. (2022) and left the determined manning roughness' in the Geul catchment based on the ESA world cover map in a 10 meter spatial resolution.*

Subsequently, the data from Figure 13 is used to parametrize the SFINCS model with Manning's roughness by using the relation between land use types and Manning roughness estimated by Deltares (2025a). As a result, every grid cell is coupled with a manning roughness, resulting in a spatial map with the Manning roughness based on the ESA World cover map with a similar resolution. The spatial map with manning roughness' is presented in Figure 13.

Incorporating spatial roughness into a hydrodynamic model improves the flow simulations as it allows the model to account for heterogeneity in surface types. Different surface types interact differently. This represents the different behavior of water over different land covers such as grass, paved areas or forests. As a result, the lagging effect of different land covers are included in the model making the flood extents, flow velocities, and water depths more reliable and accurate (Ye et al., 2018).

***Digital Elevation Model (DEM) & Mask file:***

The DEM is crucial in hydrodynamic modelling as changes in elevation determine flow gradients, which affect the flow velocities and discharges. Moreover, the source and resolution of the DEM significantly influences the simulation of flood extents and depths (Overhoff, 2024). In this study, the SFINCS model must be provided with a depth file (containing the vertical elevation of each grid cell) and a mask file (Leijnse et al., 2021). For this reason, the DEM presented in Figure 4 is used in the model. This DEM file is provided by Deltares with a 5x5 m resolution, where three national elevation model are combined: AHN4 (NL), Geoportail Wallonia (BE), and the Geoportal NRW (DE). In addition, infrastructure such as roads, train tracks, and bridges are included and depressions larger than 1 meter are already filled (Overhoff, 2024). Since the coarse grid resolution is set to 60 meters, the DEM is resampled using interpolation to match this resolution. Consequently, the resulting depth file reflects a 60-meter resolution, even though a sub grid correction at 12 meters is applied. However, it is important to note that the model still operates at a 5-meter resolution, but this is not visible in the exported depth output (Deltares, 2025b).

Moreover, the model requires a mask file which indicates the boundary cells, active cells, and inactive cells. Inactive cells can be cells located below or above certain elevation thresholds (e.g. cells of deeper water in lakes). As the Geul is a headwater catchment, only the grid cells inside the catchment are active. Additionally, no boundary cells are included in the model.

***River network:***

To complete the bathymetry, the DEM must be extended with the river bathymetry to have the depth of the rivers. Before the extension, the DEM is used in combination with the HYDRO-MT package to determine the subbasins of the river. Subsequently, the river bathymetry can be included in the DEM. However, since there is no river bathymetry available of the Geul river, the river dimensions are determined based on the bank full discharge of the observed 2-year return period (Sampson et al., 2015). The 2-year return period discharge (Q) is then utilized in the power law equations of Andreadis et al. (2013) to calculate the river width (w) and depth (d) (Equations 4.4 and 4.5).

$$w = 7.2Q^{0.50} \tag{4.6}$$

$$d = 0.63Q^{0.31} \tag{4.7}$$

Ultimately, for each subbasin the river width and depth are determined to complete the bathymetry. Subsequently, the model is improved with some minor adjustments to retrieve more reliable flood maps.

### 4.2.3 Evaluating SFINCS

Building on the previous sections where the input data and the model is explained, the model output is first improved in Appendix C. In this appendix, the taken steps are explained to retrieve more reliable outputs and to clarify the choices. Note that the preprocessing of the SFINCS outcomes is mainly conducted to improve the river network, to reduce the influence of certain water structures (like culverts) that are not represented in SFINCS, and to avoid a mismatch in the projection of the model. The last step of this methodological section involves the evaluation of the simulated flood maps from SFINCS. Both results are evaluated differently, where first the flood extent evaluation is described.

***Flood extent evaluation:***

The flood extent is evaluated with aerial measurements captured shortly after the event (Het Waterschapshuis, 2021). Based on these measurements, Slager et al. (2021) created an observed flood extent limited to the main branch of the Geul and Dutch part of the study area. As a result, the evaluation of the simulated extent is also limited to the observed extent as illustrated in Figure 14. In addition, as the original measured extent is not available, this study approximates the measured extent by applying a 700-meter buffer around the observed flood extent. This approach aims to recreate the map presented by Het Waterschapshuis (2021). The recreation of the measured extent allows that the evaluation is restricted to the area that was observed shortly after the event. This improves the reliability of the comparison between simulated and observed extents.

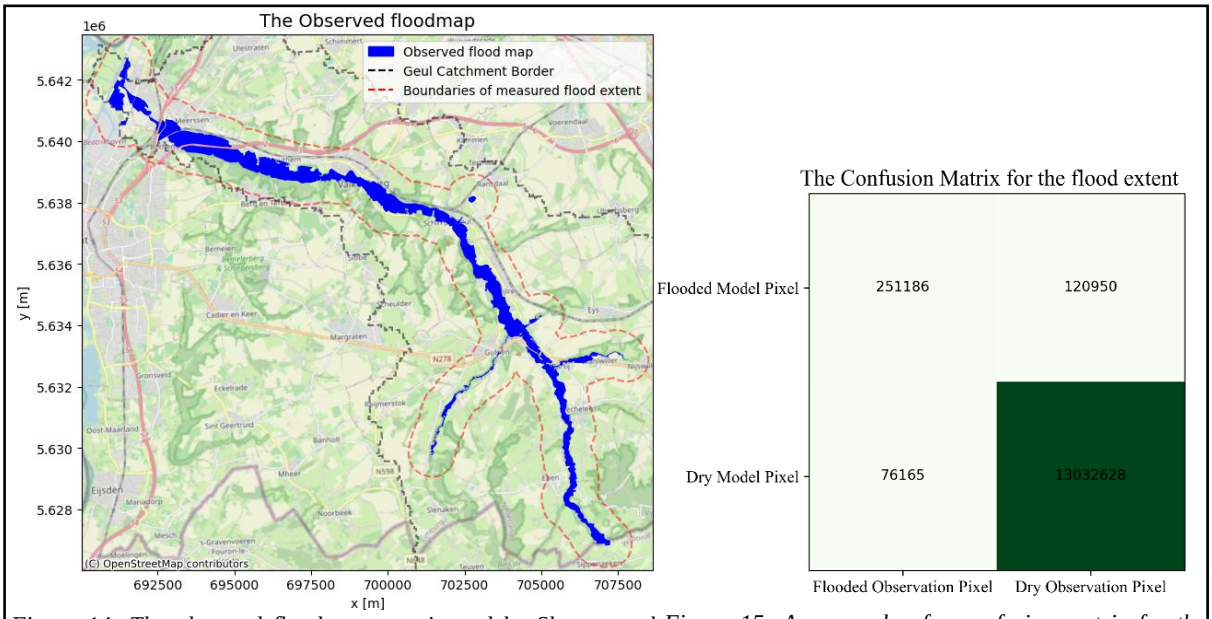


*Figure 14: The observed flood extent estimated by Slager et al. (2021) (in Blue) and the estimation of the measured extent (in red).*

*Figure 15: An example of a confusion matrix for the flood extent.*

Moreover, prior to the evaluation the river network is excluded from the flood extent with a buffer zone equal to half the river width. This step is important as the study focuses on evaluating the flood extent of the floodplains and surrounding areas, rather than the river channel itself. Finally, the simulated and observed flood maps are converted into a binary format, clipped on the catchment boundaries, and checked on characteristics. The conversion into a binary format result into maps where flooded areas are represented by ones and non-flooded areas by zeros. Another important step within the GEB framework is that a threshold of 15 cm is applied according to the study of Wing et al. (2017). This threshold means that a location is only considered flooded in SFINCS when the simulated water level exceeds this threshold.

If both maps match in characteristics in terms of resolution and shape, the extents are evaluated based on the studies of Bernhofen et al. (2018) and Wing et al. (2017). In these studies, both flood maps are laid on top of each other. Then for each grid cell, the overlapping determines whether both maps indicate flooding, in this case both simulated and observed grid cell are equal to one and the cell is classified as a **hit**. If the observed map shows flooding while the simulation does not, the cell is considered as a **miss**. Conversely, if the simulated grid cell indicates flooding but the observed cell does not, it is labelled as a **false alarm**. If both simulated and observed cell indicate no flood, the cell is labelled as a **correct negative**. These four classifications are presented in a confusion matrix in Figure 15.

Based on three classified categories hits, misses, and false alarms the performance metrics Hit rate (HR), False alarm rate (FAR), and Critical Success Index (CSI) are calculated as outlined in Equation 4.8, 4.9, and 4.10

$$HR = \frac{Hits}{Hits+Misses} \tag{4.8}$$

$$FAR = \frac{False\ Alarms}{Hits+False\ Alarms} \tag{4.9}$$

$$CSI = \frac{Hits}{Hits+Misses+False\ Alarms} \tag{4.10}$$

The hit rate is the ratio of correctly flooded pixels, ranging from 0 (no flooded pixels match) to 1 (all flooded pixels match). The metric examines the model's tendency toward underprediction. Next, the false alarm rate is the ratio of wrong predicted flooded pixels, ranging from 0 (no false alarms) to 1 (all false alarms). This metric gives an idea whether the model has the tendency to overpredict the flood extent. Lastly, the CSI balances the hit rate (underprediction) and the false alarm ratio (overprediction). This metric can range from 0 (no match between the simulated and observed flood) to 1 (perfect match between the simulated and observed flood) (Wing et al., 2017). According to Bernhofen et al. (2018) the model has a good performance when the CSI is above 0.7. Despite this, it is important to note that the CSI biases larger flood extents and in areas with smaller topographic gradients (Landwehr et al., 2024; Stephens et al., 2013). However, since this study compares the flood extent within the same flood, it is appropriate to use this metric (Bernhofen et al., 2018). Finally, the four classifications are visualized in a map to see the spatial distribution. Additionally, the three metrics are also presented in a line plot to visualize the score over the percentiles and lead times. An additional plot with the cumulative maximum over the forecast length and the spatial grid is inserted in this overview with the goal to visualize the total amount of rainfall per forecast type.

***Flood depth evaluation:***

To evaluate the model's flood depth performance, observed water depth data was collected empirically with surveys by Endendijk et al. (2023). In the surveys participants reported flood levels at their houses in an anonymized manner, limited to four- or six digital postal codes. This approach ensures the respondents privacy while enabling spatial comparison. The spatial comparison is enabled by combining the survey data with postal code 4 (PC4) & 6 (PC6) statistics of the CBS 2023 in the Geul catchment (Centraal Bureau voor de Statistiek, 2025). From the survey, the average depth against the exterior wall (in centimetres) is taken as observed flood level, an example of a spatial map for the postal four and six areas is presented in Figure 16.

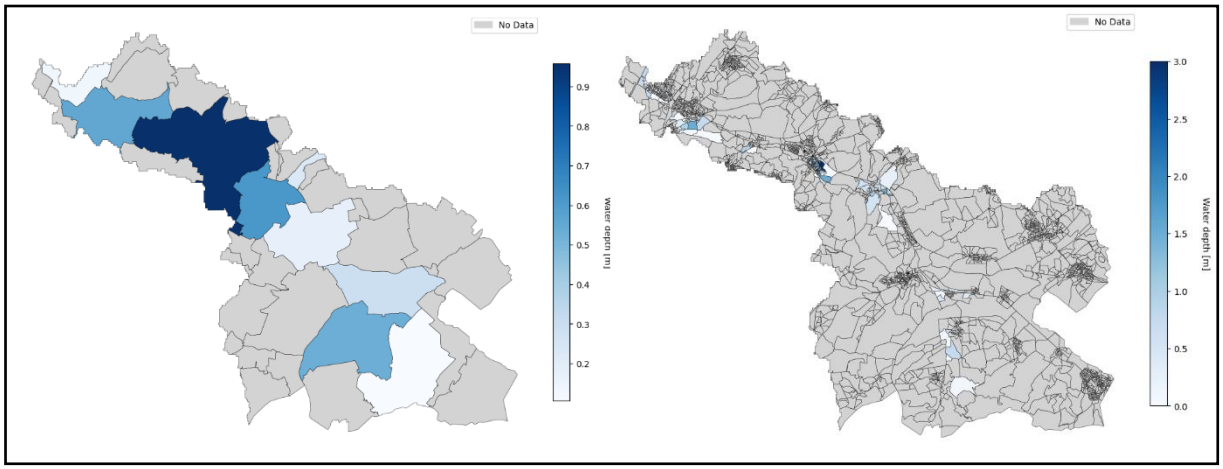


*Figure 16: The observed flood levels at the exterior wall of buildings (street level) scaled on 4-digit (left) & 6-digit (right) postal code areas (Centraal Bureau voor de Statistiek, 2025).*

Subsequently, the simulated flood map is clipped on the postal code areas, whereafter the average simulated flood depth is calculated per postal code area. As a result, for each area a single simulated value is matched against a corresponding observed value to visualize the performance over the catchment. For the spatial performance the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are used. Subsequently, to evaluate the overall performance across the catchment the Nash-Sutcliffe Efficiency (NSE) is also included. These metrics together provide a robust assessment of the model's accuracy and reliability in reproducing flood depths across the study area. First, the MAE indicates the average magnitude of error and is already outlined in Equation 4.2, then the RMSE emphasises larger differences between the simulated and observed values (outlined in Equation 4.11). Lastly, the NSE assesses how well the model simulates the depth relative to the mean of the observed data as outlined in Equation 4.12. The metrics NSE and RMSE are proven effective in evaluating flood depths against empirical data (e.g. Bermúdez et al., 2017; Khalaj et al., 2021; Manfreda & Samela, 2019), where the NSE ranges from < 0 (model simulates worse than taking the average from the observations) to 1 (where the model perfect fits the observations). NSE values exceeding 0.5 are considered acceptable and NSE values greater than 0.7 are considered very good (Nash & Sutcliffe, 1970).

$$RMSE = \sqrt{\frac{1}{N}\sum(F_i - O_i)^2} \tag{4.11}$$

$$NSE = 1 - \frac{\sum(O_i - F_i)^2}{\sum(O_i - \bar{O})^2} \tag{4.12}$$

$F_i$: Forecasted flood depth per postcode area

$O_i$: Observed flood depth per postcode area.

$\bar{O}$: Average observed flood depth over the study area

N: Total number of postal code areas

In addition, the number of attendees per postal code area from the survey results of Endendijk et al. (2023) is used to weight the RMSE and MAE statistic. The advantage of using weighted statistics lies in the ability to account for the varying number of responses per postal code area. This ensures that areas with more data have a greater influence, which improves the reliability

of the analysis by providing a more representative evaluation of model performance across the catchment.

Ultimately, the three metrics are calculated for each forecast, percentile, and lead time, included in a line plot. Moreover, the MAE and RMSE are also presented per postal code area for the ERA5 simulation to visualize the spatial differences across the study area. Based on the outcomes of the flood map evaluation, the second sub-question RQ2 can be answered: "What is the quality of flood forecasts from SFINCS for the July 2021 event at multiple lead times and how do variations in forecast data affect the outcomes of the SFINCS model?"

In the end, the evaluated flood depths and extents form an essential foundation for the next step in this methodological framework, which involves assessing the flood impacts in the affected areas.

## 4.3 From flood maps to impacts

In the final step of this methodological framework, the flood forecasts are translated into tangible impact forecasts. This process involves integrating the extent of flood depths with exposure and vulnerability data to estimate the consequences for people, assets, and infrastructure.

### 4.3.1 Input data

Following the evaluation of the flood maps, this section introduces the input data that is required for the loss model used for estimating the forecasted flood impacts. By overlaying the flood outputs with detailed exposure datasets, such as building footprints and landcover data the GEB framework quantified the forecasted damage with use of pre-defined vulnerability curves. This quantification is conducted in a loss model which is included inside the GEB framework. This study uses exposure data from the Open Street Map (OSM) dataset and from the ESA landcover map (presented in Figure 13). Moreover, to determine the forecasted damage at the exposed elements, different vulnerability curves are used.

***Exposure data:***

To represent buildings, roads, and railways in the Geul catchment, OSM is used, which is a free web-based map service. OSM is an object-based dataset with a satisfactory completeness for building locations and footprint geometries for most developed countries and urban areas (Cerri et al., 2021; Sieg & Thieken, 2022). Moreover, this dataset is often used to estimate flood damages to buildings or infrastructure (e.g. Cerri et al., 2021; Koks et al., 2019; Sieg et al., 2023). The roads are divided in four classifications: motorway, primary, secondary, and tertiary based on the study of Van Ginkel et al. (2021). For simulating the exposure of other landcover areas, this study uses the ESA World Cover 2021 database provided by Zanaga et al. (2022). Besides, this database is also used in a similar flood risk study of De Moel et al. (2011).

Moreover, to explore the number of hits for critical facilities in the region, OSM is used to filter specific tags. This study uses the following amenity tags to explore the number of critical facilities that are hit across the different forecasts and lead times. The amenities that are included are: hospitals, health clinics, doctors, fire stations, police stations, schools, kindergartens, universities, pharmacies, ambulance stations, nursing homes, childcares, townhalls, embassies, bus stations, and social facilities.

***Vulnerability curves:***

To connect the forecasted flood depth with the exposed elements in a catchment, vulnerability curves are applied in this study. Figure 17 illustrates the used vulnerability curves (derived from Bril et al. (2025)), where the relationship is presented between flood depth and the damage factor. This factor of the maximum damage is the damage fraction for a certain water depth and ranges from 0 (no damage) to 1 (maximum damage is reached) (Huizinga et al., 2017). As a result, the curve can show a higher damage ratio for a specific land use or asset type, meaning a greater proportion of the value is lost at a given flood depth. However, the total maximum damage value may still be lower for that function due to a lower economic value.

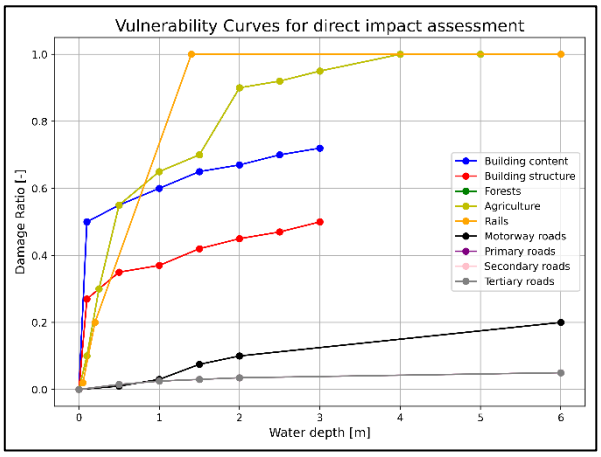


*Figure 17: The applied vulnerability curves in the study to connect exposed elements with the forecasted flood level.*

This study uses different vulnerability curves from different studies. For the building structures and contents, the study uses the curves empirically derived from the surveys of Endendijk et al. (2023), where they are based on the local experienced damages after the 2021 flood event (Endendijk et al., 2023). For roads, the European depth-damage relationships and maximum damage estimations are used for different road types (Van Ginkel et al. 2021). Next, for railways Austrian curves and maximum damages are used from Kellermann et al. (2016), due to the lack of these data for the Netherlands (Nirandjan et al., 2024). Lastly, for nature and agriculture, the study applied curves developed by De Moel et al. (2014). Specifically, for nature only clean-up costs are considered as the maximum damage (De Moel et al., 2014).

### 4.3.2 The loss model

The loss model uses the function ‘object scanner’ within the GEB framework to calculate the forecasted damage by combining object-specific exposure data, forecasted flood extents and depths, and depth-damage relations (Koks, 2022). The model supports different geometry types to account for the distinct functions and land covers present in the study area. Dependent on the geometry, the forecasted damage is determined. For polygons and lines, it calculates the affected fraction using the exact extraction of the coordinates. For point geometries, the flood depth is sampled at the object’s location. Ultimately, this approach allows for a spatially detailed damage estimation based on the actual impact of flood depth. On the other hand, the estimation remains dependent on the resolution of the input data.

### 4.3.3 Impact evaluation

The last step in the methodological framework involves the processing of the calculated impacts for each forecast and lead time. By combining all the impact data, this step enables a comprehensive assessment of how the impact evolves with forecast timing. Ultimately, this analysis answers the final sub-question of the study RQ3: "How does the predicted flood impact vary across different rainfall forecasts and lead times, and what is the added value of using ensemble-based predictions for estimating local damages per function and area?".

The results are presented in the next section, including a summary table with all the impacts per forecast per function, a trend analysis where for each function a box plot presents the total and median damage (per building) per lead time, and the total number of buildings hit with a range for ensemble forecasts. Subsequently, a hit intensity map is created to visualize which critical facilities are hit and how often they are hit across all the different lead times. Lastly, an interactive map is created to get spatial insights, and a focused evaluation of key facilities in Valkenburg to highlight local impacts.

***In-depth research centre of Valkenburg***

To evaluate the impacts in more detail, the municipality of Valkenburg is chosen to conduct an in-depth impact evaluation. This municipality experienced extensive flood damages to residential buildings, critical infrastructure, and public services during the 2021 floods (Asselman et al., 2022; ENW, 2021). Given its exposure and vulnerability, Valkenburg provide valuable insights into the overall accuracy of the developed chain. By zooming in on the centre of Valkenburg, the analysis aims to better understand how well the forecasting system captures localized flood dynamics, impact variability, and forecast uncertainty in an urban context. This local-level assessment serves as a critical step in evaluating the operational applicability of the IBFFWS for early warning and disaster response at municipal levels.

For this assessment the exposure data from OSM is used in order to determine the locations and geometries of the buildings and critical facilities. Besides, the vulnerability curve of the building structures (Figure 17) is used to determine impacts on buildings. This input data is used for the calculation of the flood and impact probability per building, where different exceedance thresholds are applied. Specifically, for the flood probability the first threshold is chosen to be 0.1 because the first damages occur at this threshold in the corresponding vulnerability curve (Bril et al., 2025; Sušnik et al., 2014). Subsequently, random water depths and impacts are chosen to visualize the difference in exceedance probabilities across different thresholds and lead times. Note that only ensemble forecasts are used in the probability calculations since they provide an uncertainty per lead time. In addition, for the impact probability is assumed that the buildings are residential buildings since this vulnerability curve is used. As a result, this impact is obviously not properly calculated for other functions such as critical facilities and stores that are present in the area.

# 5. Results

This chapter presents the key findings derived from the methodology outlined in the previous chapter. The results are structured to follow the green output steps (4.1.3, 4.2.3, and 4.3.3) of the research framework described in Figure 8, starting with the evaluation of the precipitation forecasts (5.1), followed by the evaluation of simulated flood maps (5.2). Subsequently, the evaluation of impact estimations is discussed (5.3). The outcomes of 5.2 and 5.3 are analyzed both at the scale of the catchment and on the municipality scale, where the municipality of Valkenburg is chosen as an example (5.4). Through a combination of performance indicators and visual representations, this chapter aims to provide insights into answering the main research question (MQ):

***How could Impact-based Forecasts (IbF) from the hydrodynamic SFINCS model have been used to trigger effective early-warnings and actions, incorporating associated uncertainty for the 2021 flood in the Geul basin?***

## 5.1 The rainfall forecasts issued by ECMWF for the July 2021 flood in the Geul

This section evaluates the rainfall forecasts that serve as key input for the hydrodynamic simulations. First, the analysis compares the different forecasts and visualizes the percentile approach (5.1.1). Afterwards, the performance of different forecasts is evaluated to assess statistical metrics for all the forecasts against the ERA5 dataset (5.1.2). Additional explanatory figures have been added in Appendix C. Ultimately, the findings of this paragraph will help answering the **RQ1** of this study:

***How well do historic operational ensemble and deterministic forecasts relate to the observed precipitation patterns and totals of the July 2021 flood event, compared to ERA5 reanalysis data?***

### 5.1.1 Overview and visual comparison of the issued forecasts

Figure 18 compares the maximum rainfall intensity across the Geul catchment for the processed ECMWF control and ensemble percentile forecasts against the ERA5 observations for the forecast initialization dates between 11 and 14 July. In general, the forecasts better align with ERA5 before the flood event (red line) compared to after the event. Especially the median ensemble percentiles (P50 & P75) tend to be closer to the ERA5 data, conversely to the more extreme percentile (P90 & P95) who present more extreme scenarios. Therefore, the percentile range (P25-P95) reflects the forecast uncertainty, but often overestimates the peak magnitudes, indicating an overestimation of extreme rainfall. However, none of the forecasts reproduce all three maximum observed rainfall peaks simultaneously and accurately since the forecasts often predict the peaks either earlier or later. Regarding the control forecast (green line), it exhibits distinct and variable behaviour across lead times, sometimes the forecast closely aligns with ERA5 and the median ensemble percentile (12 July 00 UTC forecast, aligns till the flood event).

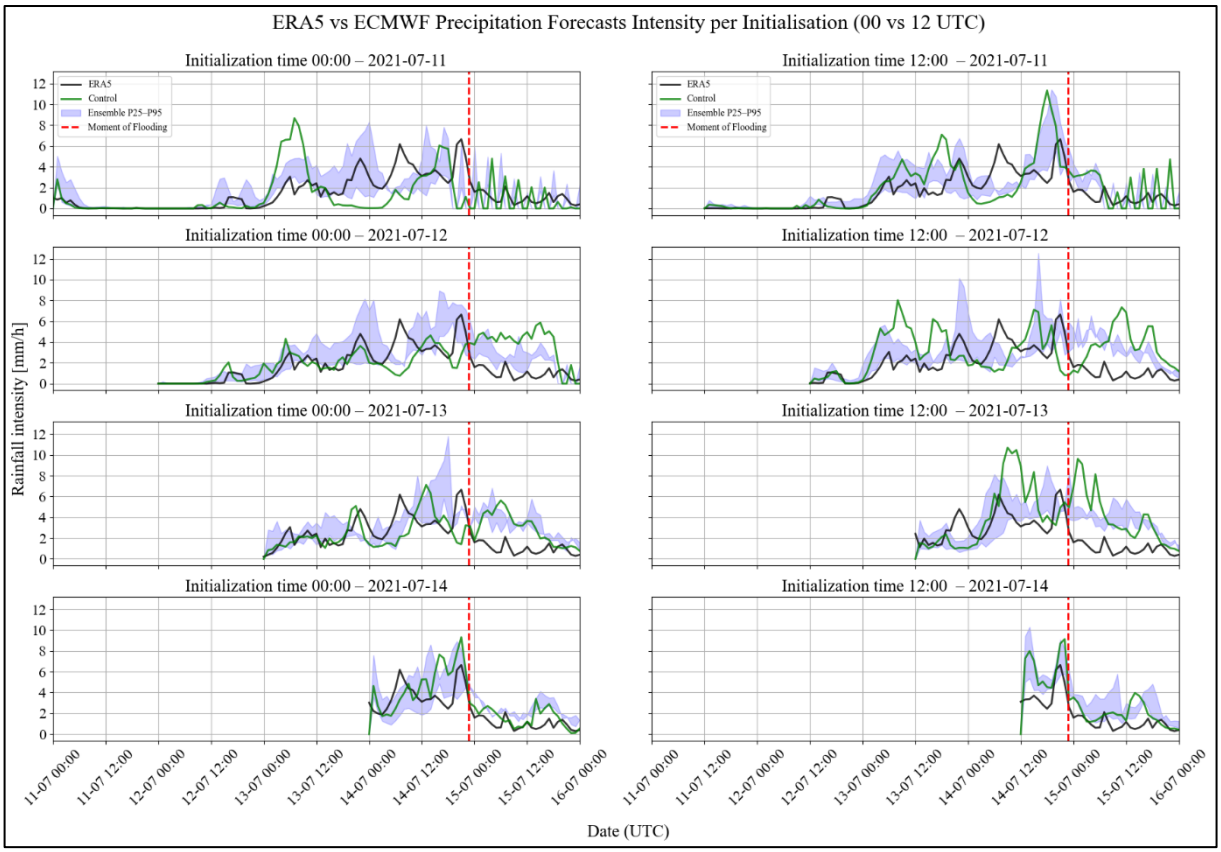


*Figure 18: A comparison of the maximum rainfall intensity across the Geul catchment between the processed control (green line), ensemble percentile forecasts (blue spread) of the ECMWF, and the ERA5 observations (black line) for various initialization dates between July 11 and 14 2021.*

Another initialization, the control forecasts behave differently and follows the more extreme percentiles (11 July 12 UTC forecast), where in some cases it even deviates completely from the other forecasts (11 July 00 UTC forecast on 13 July predicting a factor two more). These findings highlight that the control forecast represents a single deterministic realization without incorporated uncertainty, leading to significant deviations from both the ensemble forecasts and ERA5. Lastly, both deterministic and ensemble forecast tend to overestimate the ERA5 after the flood appeared, indicated by the drop of rainfall intensity of the ERA5 data.

Following the analysis of rainfall intensity, Figure C.1 (see Appendix C) presents the corresponding maximum cumulative rainfall totals for each forecast initialization across the Geul catchment. In contrast to the intensity plots, the cumulative forecasts of both control and ensemble percentiles show a strong alignment with ERA5 for the longest lead times. As lead time decreases, the ensemble spread narrows but overestimates the maximum total rainfall compared to ERA5. Moreover, the total rainfall decreases over the lead times due to the length of the forecast. As the forecast length decreases the amount of time where the rain could be forecasted decreases. In addition, some rainfall peaks have already occurred for the latest forecast initializations.

Building on the insights from Figure C.1, the spatial distribution of the cumulative rainfall further illustrates these patterns per grid cell as shown in Figure 19. The spatial maps of cumulative rainfall reveal clear patterns in both time and ensemble distribution. The rainfall increases across lead times until 14 July due to the shorter remaining forecast duration and the

rainfall peaks that fell earlier. Moreover, a horizontal increase is visible in each lead time, with higher precipitation amounts forecasted in the higher percentiles (e.g. P95) compared to lower percentiles (e.g. P25). In terms of spatial distribution, the ensemble and control forecasts predict more rainfall in the northern and eastern grid cells of the catchment, whereas ERA5 indicates more rainfall in the southeastern corner.

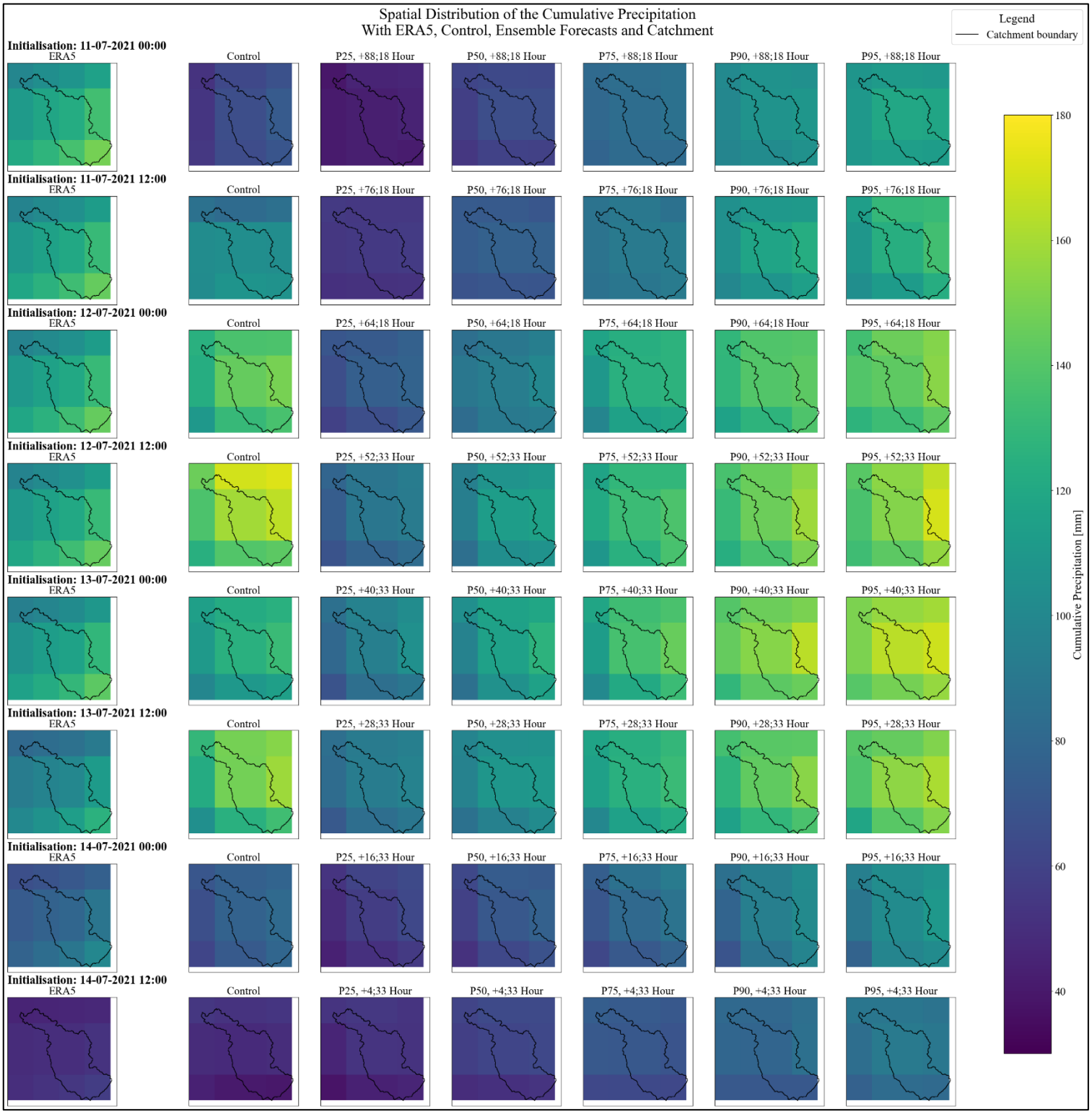


*Figure 19: Spatial distribution of cumulative rainfall (in mm) per grid cell over the Geul catchment for each lead time (in vertical direction), ERA5(most left), control (1 from left), and ensemble percentile forecasts (horizontal direction).*

This indicates a possible spatial mismatch between forecasted and observed distributions. In addition to the findings of Figure C.1, the more extreme percentiles are more aligned with ERA5 and control values until the 13 July forecast, after which the median percentiles (P50 & P75) align better. For the final forecast on 14 July 12:00, the lowest percentile (P25) corresponds most closely to ERA5 and the control forecast. However, again the control forecasts exhibit different behaviour across lead times as for some lead times it predicts more rainfall compared to the most extreme percentiles in Figure D.1 and Figure 19.

Following on these insights, the next section evaluates the forecast performance using various verification metrics to compare the forecasts and to quantify the overall performance, reliability, accuracy, and spatial representativeness.

### 5.1.2 Evaluation of forecast performance against ERA5 observations

To evaluate the performance of the control and ensemble percentile forecasts, the forecasts are compared against ERA5 observations by using a range of verification metrics. First, the relation between the control forecast and ERA5 for the cumulative sum of rainfall is presented in the left scatterplot in Figure 20. Each point in the left figure represents a grid cell and the closeness to the diagonal 1:1 line indicates a minimal deviation between the forecasted and observed value. The control forecasts initialized at the 4, 16-, 40-, 64-, and 76-hour lead times illustrate a good fit to the 1:1 line, suggesting accurate predictions. In contrast, the forecast for 88-hour lead time is located below the line, indicating a systematic underestimation. Conversely, the forecasts at 52 and 28 hours are more dispersed and lie above the line, which reflects a more consistent overestimation.

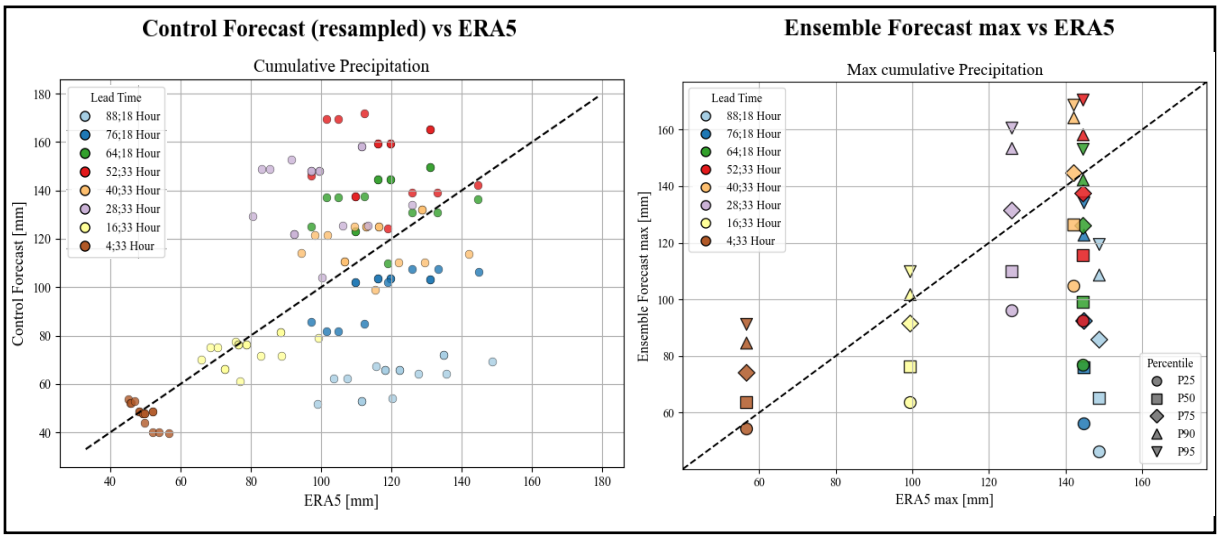


*Figure 20: Scatterplot comparison of cumulative rainfall forecasts against ERA5. The left panel presents the relation of the spatially distributed cumulative rainfall per grid cell per lead time between the control forecast and ERA5. The right panel only illustrates the relation of maximum cumulative rainfall across the catchment per percentile per lead time between the ensemble percentile forecast and the ERA5 observations.*

Moreover, on the right side of Figure 20 the maximum cumulative rainfall values across the catchment are presented for each ensemble percentile and lead time. This visualization complements the spatial rainfall patterns previously described, which illustrates how the ensemble captures the maximum values for each lead time. In line with Figure C.1, the highest percentiles (P90 & P95) are more aligned with ERA5 for the earliest lead times until the 13 July

initialization (40 hour lead time). From that point onward, the median percentiles (P75 & P50) present a better alignment.

Following the spatial and visual inspectation of the cumulative rainfall forecast accuracy, the statistical reliability of the ensemble percentile forecasts is assessed in Figure 21 using PIT diagrams. Across the different lead times, systematic deviations from the ideal reliable uniform distribution are presented in the data. This suggests limitations in forecast calibration, where for the longest lead times (88 to 76 hours) the curves lie below the diagonal 1:1 line till PIT value 0.6. As a result, crossing the 1:1 line around 0.6 indicates that observations tend to exceed most ensemble forecasts too often in the lower rainfall intesities, but are overestimated in the higher rainfall intesities. This suggests both underprediction and poor calibration since the observed values exceed the forecasted percentiles too frequent. At lead times (64 to 28 hours), the diagrams show steep jumps around Pit value 0 as a result of underdispersion. Tis suggests that the ensemble percentile spread is then too narrow, resulting in that the ERA5 observations fall outside the range of the forecast. For the lead times (16 & 4 hours) the PIT curves lie again below the diagonal line, which indicates underprediction. These patterns indicate that the ensemble percentiles are not calibrated with the ERA5 data and that they often underpredict the ERA5 observation as also illustrated in Figure 18.

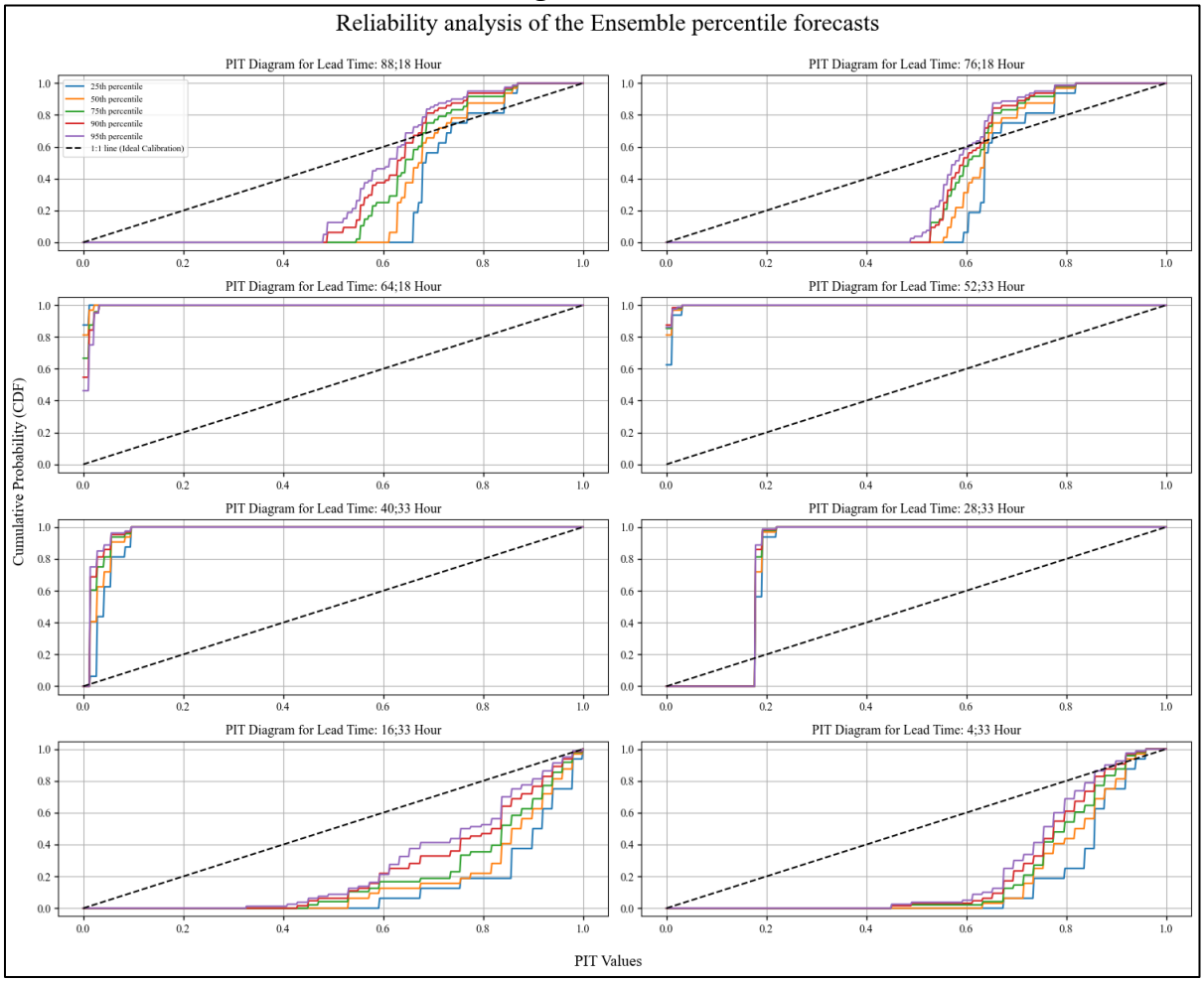


*Figure 21: The Probability Integral Transform (PIT) diagrams across the different forecast horizons, with lead times ranging from 88 hours to 4 hours. Each subplot represents the CDF of the PIT values for every percentile separately. These values are compared to the ideal calibration 1:1 line.*

Building on the reliability analysis, Table 2 provides a quantitative summary of the forecast verification metrics (CRPS, Bias, MAE, and FSS). In general, the ensemble percentile forecasts perform well for longer lead times (88 to 52 hours), with CRPS scores ranging between 0.52 and 0.60 mm/h. This indicates a moderate alignment between the ensemble percentile forecasts and ERA5 during these early lead times. However, the CRPS score increases from the 40 hour lead time onward, reaching a maximum of 1.061 mm/h at 4 hours. This indicates a significant drop in forecast skill as lead times shortens. An exception is the 16 hour lead time, which maintains a low CRPS (0.589 mm/h).

Regarding the bias and MAE, the ensemble and control forecasts generally shift from a small underestimation (negative bias) at longer lead times to slight overestimations (positive bias) from the 52 hour lead time onward. Additionally, from 52 hour lead time onward the standard deviation of the bias (σ) amplifies the overestimation for the entire ensemble as the bias further increases. In contrast, for the earlier lead times the standard deviation contradicts the underestimation since the deviation brings the entire ensemble closer to the ERA5 (0 mm/h bias). This suggests an increased overestimation in the ensemble forecast compared to the ERA5 closer to the event. Moreover, the control forecast show more fluctuations in the bias and have relative to the ensemble forecast a higher MAE. These verification metrics further support the findings of Figure 20 and 21 for the ensemble forecasts as the underprediction observed in the early and late lead times (88 & 76, 16 to 4 hours) is reflected in negative biases. In addition, the underdispersion for the median lead times (64 to 28 hours) reflects the increased MAE and CRPS.

*Table 2: Overview of verification metrics for the evaluation of the forecast performance against ERA5. These metrics describe the overall performance, accuracy and spatial representativeness of the rainfall intensity. Specifically, for the MAE and Bias is added the standard deviation to address the variability of the percentile members.*

| *VERIFICATION METRIC* | *CRPS [mm/h]* | *BIAS (σ) [mm/h]* | *MAE (σ) [mm/h]* | *FSS [-]* |
|---|---|---|---|---|
| *Ensemble percentile forecast – Lead time* | | | | |
| *88;18* | 0.595 | -0.327 (+0.218) | 0.801 (+0.057) | 0.608 |
| *76;18* | 0.537 | -0.255 (+0.204) | 0.753 (+0.064) | 0.506 |
| *64;18* | 0.520 | -0.068 (+0.277) | 0.824 (+0.121) | 0.446 |
| *52;33* | 0.581 | 0.041 (+0.256) | 0.905 (+0.134) | 0.423 |
| *40;33* | 0.797 | 0.165 (+0.332) | 1.167 (+0.132) | 0.361 |
| *28;33* | 0.950 | 0.059 (+0.296) | 1.274 (+0.116) | 0.353 |
| *16;33* | 0.589 | -0.011 (+0.321) | 0.908 (+0.092) | 0.471 |
| *4;33* | 1.061 | -0.212 (+0.274) | 1.322 (+0.154) | 0.443 |
| *Control forecast – Lead time* | | | | |
| *88;18* | - | -0.476 | 0.878 | 0.784 |
| *76;18* | - | -0.180 | 0.871 | 0.917 |
| *64;18* | - | 0.178 | 1.241 | 0.889 |
| *52;33* | - | 0.357 | 1.211 | 0.876 |
| *40;33* | - | 0.051 | 1.311 | 0.927 |
| *28;33* | - | 0.325 | 1.549 | 0.859 |
| *16;33* | - | -0.100 | 0.803 | 0.943 |
| *4;33* | - | -0.639 | 1.124 | 0.771 |

Furthermore, the FSS score of the ensemble forecasts illustrate a consistent decline from 0.608 at 88 hour lead time to 0.353 at 28 hours. Despite being deterministic, the control forecast achieve significantly higher FSS scores across all the lead times with a peak FSS score at 16 hours of 0.943. This indicates that the control forecast better represent the spatial distribution of rainfall intensity of ERA5, which mirrors the findings of Figure 19.

In addition to Table 2, Figure C.2 and Figure C.3 present the spatial distribution of the examined CRPS, Bias, MAE, and standard deviation (ensemble forecasts) scores. Regarding both forecasts, it is noteworthy that the CRPS, MAE and Bias scores are higher for the eastern grid cells and increases over the lead times. Again this is in line with the findings of Figure 19 where more rain was predicted in these cells. However this is not captured in the ERA5 data. Additionally, the control forecast reflects the trend a bit more across the catchment, resulting in fewer differences between grid cells. Moreover, the range of FSS scores over the different rainfall intensity thresholds are presented in Figure C.4 per lead time. The FSS boxplots reveal more clearly the difference of FSS scores over the different thresholds between the control and ensemble forecasts.

Together with all the metrics, the ensemble forecasts offer a valuable representation of uncertainty as they illustrate signs of underprediction, underdispersion, and limited spatial skill at shorter lead times compared to the ERA5 observations. Conversely, the control forecasts demonstrate greater consistency and higher spatial agreement with the ERA5 observations across all lead times. The findings of the forecast performance form an essential foundation for the next step, where resulting forecasts are implemented to simulate flood maps in the next step.

## 5.2 The forecast of flood maps in the Geul catchment

The next step in the analysis focuses on the evaluation of the forecasted flood maps in the Geul catchment. As a baseline, the flood map derived from the ERA5 precipitation is first assessed, whereafter the different flood maps of the forecasts and lead times are compared to this baseline map. This comparison will help answering the second sub research question (RQ2):

***What is the quality of flood forecasts from SFINCS for the July 2021 event at multiple lead times and how do variations in forecast data affect the outcomes of the SFINCS model?***

Figure 22 presents the flood map generated by SFINCS using the most extreme ensemble percentile (P95) at 88-hour lead time as an example. The map reveals that the most extensive flooding occurs in the downstream parts of the Geul River. Comparing the observed flood extent illustrated in Figure 14 with Figure 22, it appears that the model overestimates flooding near the catchment outflow. This deviation suggests an overrepresentation of flooding. Moreover, the depths presented in Figure 22 display water depth of approximate three meters. Due to the simulation of unrealistically high water depths single cells at the origins of the tributaries (as also identified in Appendix D) the legend in the figure depicts water depths up to 9 meters.

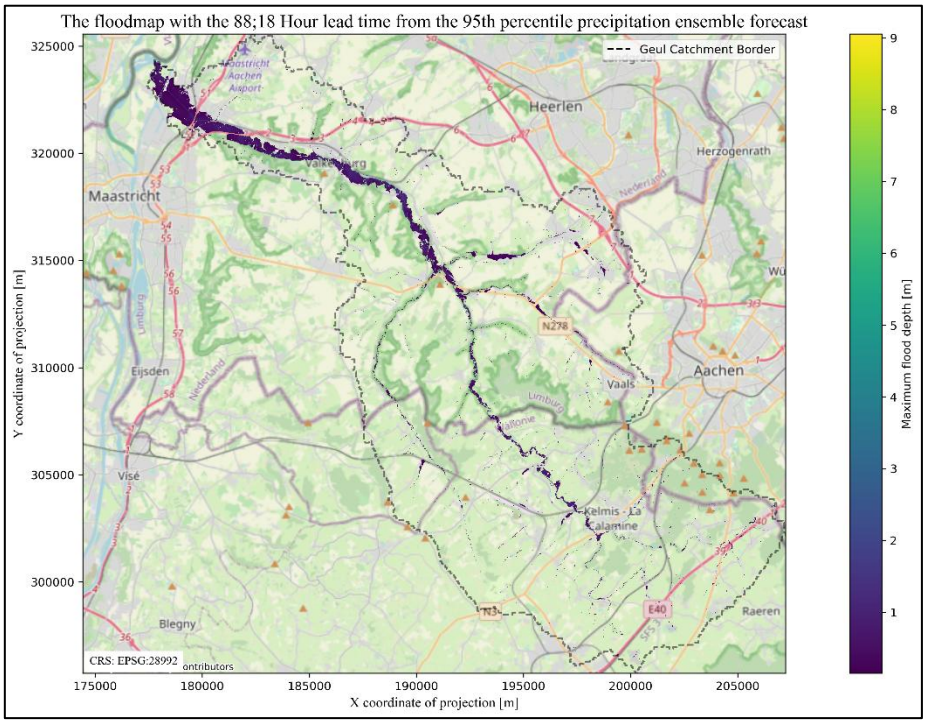


*Figure 22: Simulated flood map across the Geul catchment based on the 88 hours 95$^{th}$ percentile forecast input. The flood depths are spatially represented in meters and range till three meters, but single grid cells simulate extraordinary water depths as addressed in Appendix C, leading to the scaling of the legend as presented in the figure.*

Building on the flood map of the 95$^{th}$ percentile for the 88-hour lead time, Figure 23 displays the extent of the different forecasts. The forecasts and ERA5 show significantly higher extents compared to the observed flood extents, as this extent is only limited to the main Dutch sections of the Geul River. Moreover, all forecasts show larger flood extents compared to the ERA5 (green) extent, indicating that higher amounts of rainfall (presented in Figure 18 and C.1) do have effect on the flood extent.

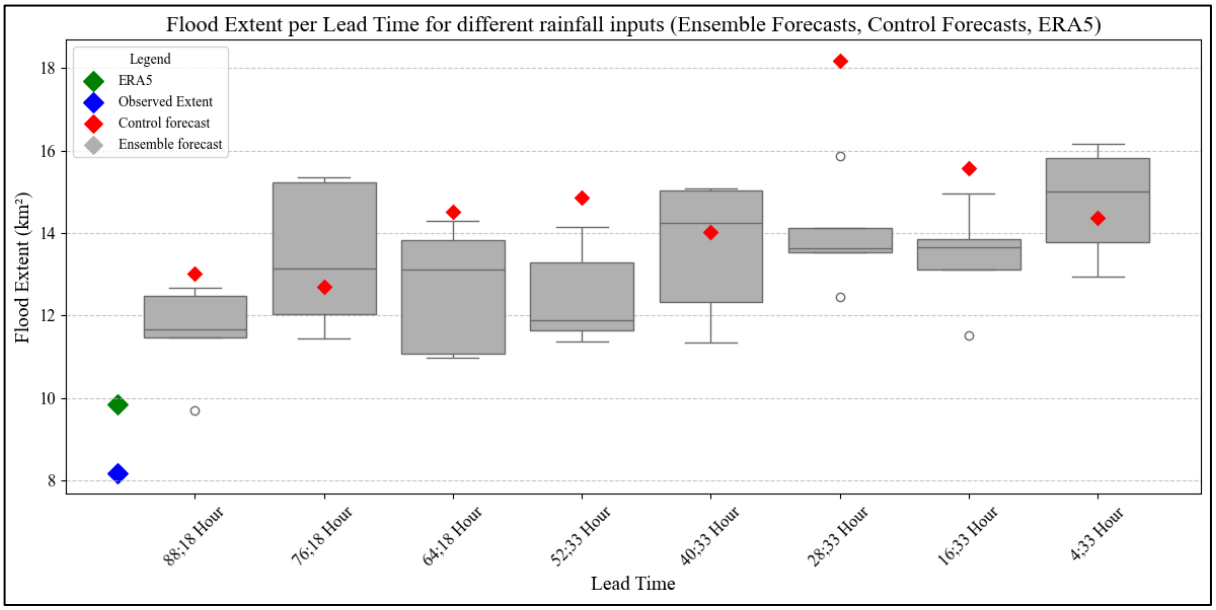


*Figure 23: The simulated flood extents in km² over the different lead times with the different forecasts as input. The blue shape represents the observed flood extent, the green shape the flood extent of ERA5, the red shape the flood extent of the control forecasts and the grey boxplots represent the five different percentiles of the ensemble percentile forecast.*

Moreover, Figure 23 presents that the lower percentiles are relatively close to the ERA5 (e.g. 25$^{th}$ percentile of the 88-hour lead time forecast). This behaviour is also noticed in the previous section 5.1 and again in Figure 23, the discrepancy between ERA5 and forecasted flood extents increases as lead times decreases. Regarding the ensemble extents, they show a minor increasing trend across the lead times, where the overall spread decreases. This indicates a decrease in uncertainty, presenting a more accurate simulated extent. Nonetheless, the final spread reflects an uncertainty of approximate three km$^2$ at the 4 h lead time. Additionally, the lower percentiles in the boxplots remain relatively constant over the lead times compared to the most extreme scenarios.

In terms of the simulated extents from the control forecasts (red) of Figure 23, the deterministic character is shown again. The extents often lay above the spread of the ensemble forecasts, which suggests a relative overestimation of the event compared to the other forecasts. Also, for these flood extents the 28-hour lead time remains an outlier.

Building on the map presented in Figure 22, the extent is evaluated by calculating the different evaluation metrics described in the methodological framework. Figure 24 illustrates the spatial flood extent evaluation for the 88-hour lead time and 95$^{th}$ percentile precipitation forecast. The comparison between simulated and observed flood extent (Figure 14) presents a high number of false alarms (0.3) in the downstream areas, which indicates an overprediction. On the other hand, in the upstream regions of the measured extent (in light blue) the model fails to capture a significant portion of the observed extent. These mismatches in false alarms and misses are also reflected in the intermediate CSI score of 0.54.

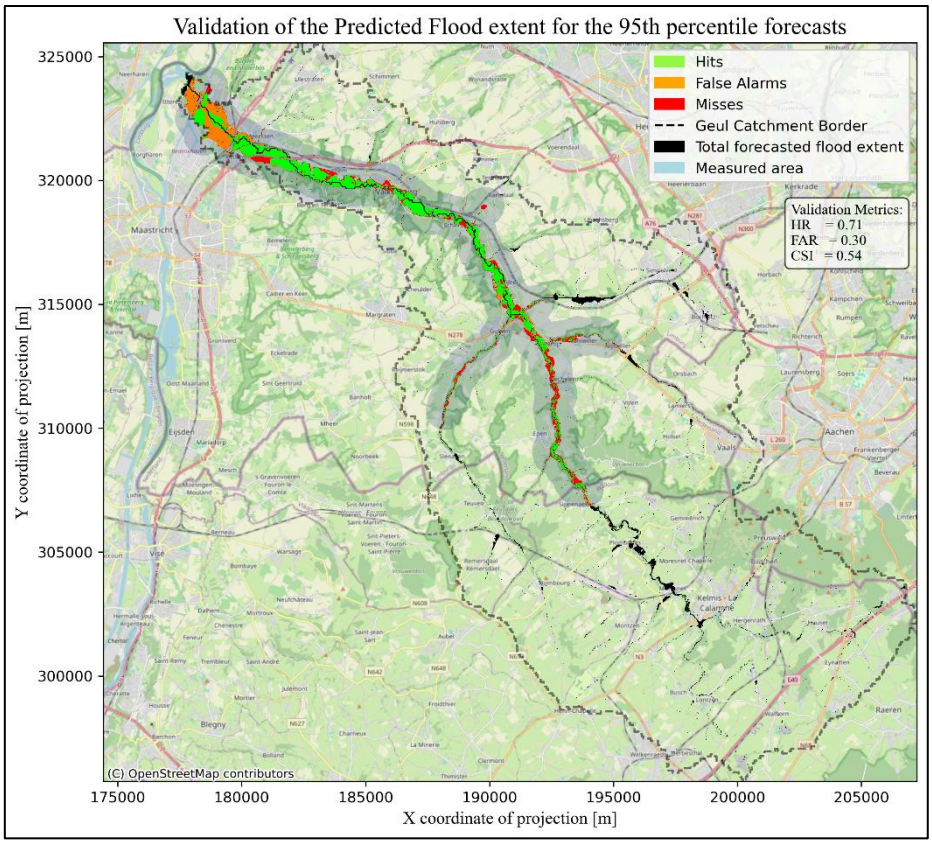


*Figure 24: Spatial evaluation of the simulated flood extent with the 88 hours 95$^{th}$ percentile forecast as input against the observed flood extent of Slager et al. (2021). The hits are represented in light green, the false alarms in orange, the misses in red, the measured area in light blue, and the total flood extent in black.*

Following the spatial flood evaluation, the different forecasted flood extents are evaluated over the lead times and displayed with the performance metrics in Figure 25. First, the hit rate shows a clear declining trend with increasing performance over decreasing lead times and decreasing performance over percentiles. This indicates that shorter lead times and higher percentiles result in better detection of the observed extent. The control forecast shows the highest hit rate at the 28-hour lead time, whereafter a general decline can be observed. In contrast, both the false alarm rate and CSI index remain relatively constant over the lead times but show similar behaviour as the hit rate performance. Simultaneously, lower percentiles score lower false alarms rates, while higher percentiles and the control forecast have higher false alarm rates. As a result, the higher hit rates and false alarm rates balances the CSI index values around 0.55-0.6. This indicates that despite higher detection rates the overall fit does not significantly improve, suggesting that the forecasts show similar mismatches as Figure 24. Notably, the control forecast displays an unusual peak at the 28h lead time across multiple metrics.

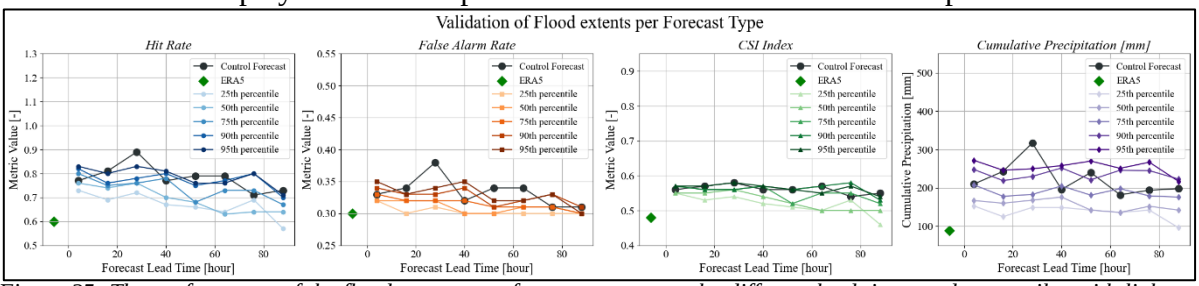


*Figure 25: The performance of the flood extents per forecast type over the different lead times and percentiles with lighter colours representing the lower percentiles, the darker colours the higher percentiles, anthracite representing the control forecast and the green shape represents the ERA5 scores.*

Moreover, the fourth subplot in Figure 25 presents the cumulative rainfall over the grid. Again, the lower percentiles are more aligned with the ERA5 compared to the higher percentiles and the control forecast. These forecasts overestimate the cumulative precipitation compared to ERA5. However, this has limited impact on the overall performance (CSI index).

The evaluation continues with the evaluation of the simulated flood depths for 88 h lead time and 95$^{th}$ percentile forecast by first focusing on the spatial evaluation based on postal code statistics PC4 and PC6. Figure 26 presents the PC4 and PC6 spatial evaluation, where for the PC4 postal code areas the MAE ranges between 0 and 0.4 meters and the RMSE from 0 to 0.16 meters. In contrast, for the PC6 postal code areas the MAE spans from 0 to 2.5 meters and the RMSE from 0 to 0.25 meters, reflecting a greater spatial detail but also more variability. The flood depth analysis reveals that the municipality of Valkenburg has the highest MAE and RMSE across the PC4 and PC6 postal code levels. However, the magnitude of the errors is significant higher in the PC6 postal code level due to the increased spatial detail of these postal code zones. Furthermore, only a limited number of postal code areas reported observed flooding, suggesting that much of the modelled flood extent lies outside the areas with recorded impacts. As a result, especially for the detailed postal code areas the evaluation can be better visualized on a municipality scale rather than at the full catchment scale regarding the PC6 statistics.

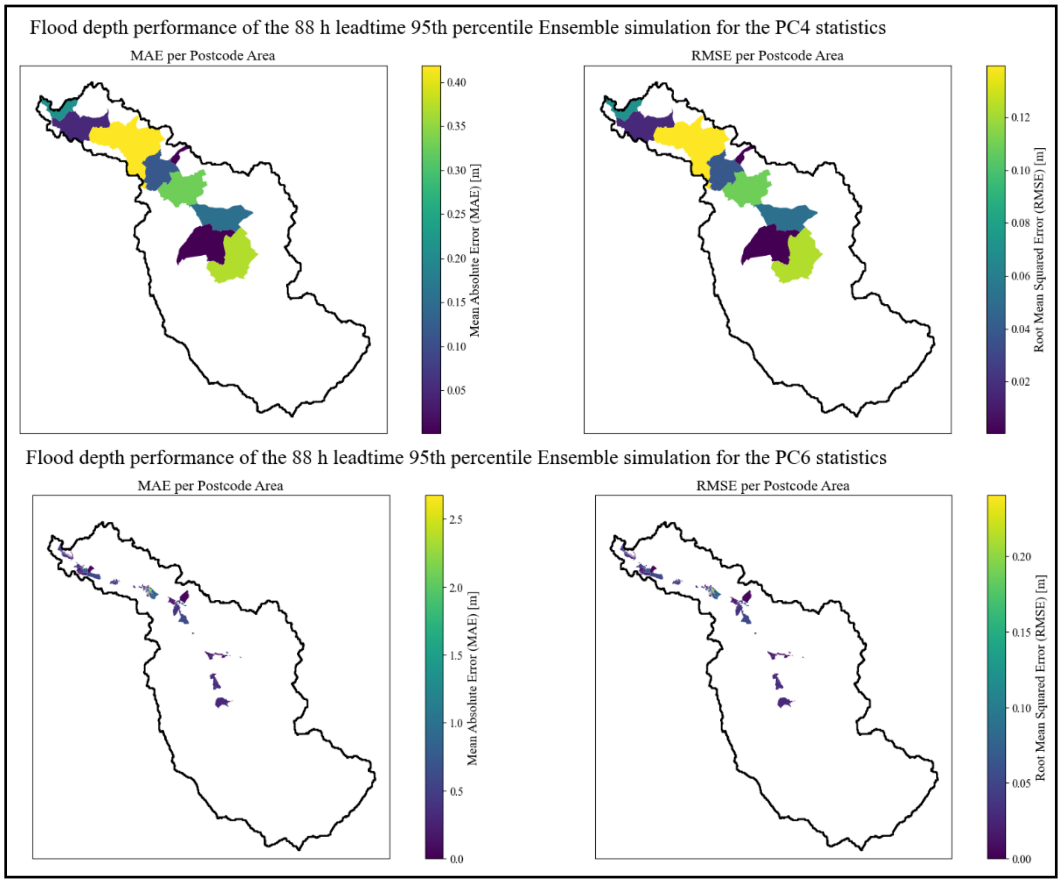


*Figure 26: The spatial evaluation of the simulated flood depths for 88 h lead time 95$^{th}$ percentile ensemble against the observations from the survey of Endendijk et al. (2023). This evaluation is based on the aggregation of the flood depths per postal code area detail level (PC4 and PC6)*

Building on the spatial evaluation of Figure 26, Figure 27 presents the evaluation metrics for the different forecasted simulated flood depths across the Geul catchment against the household survey data of Endendijk et al. (2023). The figure distinguishes the metrics per postal code detail level, forecast type, lead times, and weighted and unweighted statistics. In general, the control forecast shows a sharp negative discrepancy at lead time 28 hours in all metrics, where an exception is visible in the weighted PC4 MAE metric.

For the unweighted statistics, the PC4 evaluation shows a stable performance across the lead times with MAE values ranging from 0.17 to 0.28 meters, RMSE from 0.23 to 0.34 meters, and NSE scores between 0.3 and -0.4. Across the metrics, the lower percentiles (P25-P50) align more closely to the performance of ERA5. Especially at earlier lead times (88h & 76h) the performances matches the ERA5 performance, while the higher percentiles and the control forecast tend to overestimate the flood depths. In addition, the NSE values below 0.0 for the higher percentiles and the control forecast indicate that the model does not score better than taking the average out of the observational data. As a result, the NSE values suggest that the higher percentiles and control forecast does not describe the flood depths well and the lower percentiles and ERA5 do not score an acceptable NSE score (>0.5) either. Though, it is unexpected that the higher percentiles score worse over the different metrics regarding flood depth.

For the PC6 evaluation, the simulated flood depths perform less compared to the PC4 evaluation. The forecast errors increase significantly at longer lead times and lower percentiles, which can be observed in the degradation over the lead times for the MAE and RMSE metrics. Conversely, higher percentiles and the control forecast remain more constant over the lead times, where again the ERA5-based simulation performs similarly to the low percentile forecasts at early lead times. Across the PC6 evaluation metrics the MAE remains between 0.44 and 0.50 meters and the NSE values are slightly better compared to the PC4 postal code areas, as the higher percentile score better. However, these NSE values are still not acceptable.

Turning to the weighted statistics the trends are smoother, where postal code areas with more household responses carry greater influence. First, the NSE statistics remain unchanged since this metric is not affected by weighting. Over both the PC4 and PC6 the control forecast is now more aligned in the centre of the ensemble, except for the 28h lead time. For PC4, the weighted MAE increases slightly with lead time and lower percentile, where the percentile ranking has inverted. Higher percentiles (P75-P95) now show (as expected) lower errors, while lower percentiles show higher errors. This pattern indicates a shift in performance under weighting. In addition, this pattern also applies for the RMSE. Moreover, for the PC6 statistics this trend is amplified, where the error increases more sharply for low percentiles and early lead times. This highlights the sensitivity to poorly performing ensemble members. Overall, the results show the influence of weighting, spatial resolution, forecast type, and forecast lead time on model performance. In comparison to the ensemble forecast performance, the performance of the control forecast depicts a different behaviour with errors larger than the ensemble forecasts. Apart from the peak at the 28-hour lead time, the control forecast presents also a decrease in performance across the lead times.


*Figure 27: Evaluation metrics of forecasted flood depths across the Geul catchment using household survey data, shown for both PC4 and PC6 postcode levels. Each subplot displays a metric with the Mean Absolute Error (MAE) (blue at left), Root Mean Square Error (RMSE) (orange at centre), and Nash-Sutcliffe Efficiency (NSE) (green at right) for ensemble percentile forecasts (gradient of the colour), ERA5 (green shape) and control forecasts (anthracite) over various lead times. Metrics for both postal code areas are presented in both unweighted (top two rows) and weighted (bottom two rows) form to account for the number of survey responses per postcode.*

These findings underline the differences in reliability of forecasted flood extents and depths across different lead times and forecast types. To further examine the practical value of the different forecasts, the next section explores the translation of the forecasted flood maps into tangible impacts.

## 5.3 Forecasted impact in the Geul catchment

The last step in the analysis of the methodological framework focuses on the forecasted flood impacts. By translating the forecasted flood maps into affected areas, this section evaluates how variations in rainfall forecasts propagate through the modelling chain to influence impact estimations by answering the final research sub question RQ3:

***How does the predicted flood impact vary across different rainfall forecasts and lead times, and what is the added value of using ensemble-based predictions for estimating local damages per function and area?***

Table 3 provides a quantitative summary of the estimated flood impacts across different forecast type, lead time, and ensemble percentiles. This overview allows a comprehensive comparison of rainfall, flood extents, number of affected buildings, total damages across all the exposed assets, median damages per building, and the exposed critical facilities.

First, the table indicates that the variations in cumulative rainfall across lead times and percentiles is directly translated to proportional changes in flood extent, indicating an intermediate influence of total rainfall on the spatial extent of flooding. Additionally, for the lowest lead time the control forecast simulated an impact estimate identical to ERA5, although having a larger flood extent. Since the flood extent often differs with the ERA5 extent, the influence of the location of the flood extents is illustrated in the calculation of the damages. For example, the 40h control forecast simulated a flood extent of 14 $km^2$ and forecasted twice as much precipitation but simultaneously the estimated total damage is predicted lower compared to ERA5. In this case, a higher flood extent does not necessarily lead to higher total damages. This example highlights the added value of the buildings hit column since it presents the distribution of the extent. As the number of buildings hit increases it means that the flood extent is more distributed in urban areas.

Regarding the control forecasts, they constant forecasted higher cumulative precipitation amounts and flood extents as also presented in Figure 18, 23 and C.1. Moreover, again 28h lead time depict peak impacts with 317 mm of rainfall, 18.18 $km^2$ flood extent, 7855 buildings are affected, and a total damage of € $5.66*10^8$. These estimates significantly exceed the ERA5 baseline simulation. However, the control forecasts present fluctuations in its forecasts since the 16h lead time estimates lower impacts. Therefore, the control forecasts display more variation across lead times, as it provides a single deterministic estimate per lead time.

In contrast, the ensemble percentile forecasts present a more stable range of impact estimates. The more extreme percentiles (P90 & P95) constantly predict more precipitation, flood extents, damage estimates and number of affected buildings and critical facilities. As a result, at longer lead times (88h till 52h) these extreme percentiles are more aligned with ERA5, indicating a minor underprediction of impact estimates. However, at shorter lead times (40h till 4h) the more median percentiles (P50 & P75) correspond with ERA5.

Moreover, the number of affected critical facilities by simulated floods tend to be constant, except for higher percentiles and the early lead times where the hits increase up to sixteen facilities.

*Table 3: An overview with propagation of different rainfall forecasts to impact estimations in terms of total damages in euros, total number of buildings hit, median damage per building, and the number of critical facilities that are hit.*

| ***Forecast type*** | ***Lead time [h]*** | ***∑ Rainfall [mm]*** | ***Flood extent [ $km^2$]*** | ***Buildings Hit [N]*** | ***Total Damage [€ $* 10^8$]*** | ***Median damage per building [€]*** | ***Hit critical facilities [N]*** |
|---|---|---|---|---|---|---|---|
| *ERA5* | - | 89 | 9.86 | 5871 | 3.91 | 29693 | 11 |
| *Control forecast* | | | | | | | |
| | 88;18 | 198 | 13.01 | 5162 | 3.37 | 27738 | 13 |
| | 76;18 | 194 | 12.69 | 4938 | 3.18 | 26679 | 13 |
| | 64;18 | 182 | 14.52 | 6027 | 4.17 | 30445 | 13 |
| | 52;33 | 240 | 14.86 | 6156 | 4.24 | 30121 | 14 |
| | 40;33 | 196 | 14.02 | 5646 | 3.73 | 28786 | 13 |
| | 28;33 | 317 | 18.18 | 7855 | 5.66 | 33408 | 17 |
| | 16;33 | 243 | 15.56 | 6398 | 4.40 | 30762 | 15 |
| | 4;33 | 210 | 14.36 | 5871 | 3.91 | 29693 | 14 |
| *Lead time [h]* | *Ensemble percentile forecast* | | | | | | |
| *88;18* | $25^{th}$ | 96 | 9.71 | 3379 | 2.20 | 23726 | 12 |
| | $50^{th}$ | 142 | 11.48 | 4450 | 2.75 | 23931 | 13 |
| | $75^{th}$ | 176 | 11.67 | 4445 | 2.85 | 25889 | 13 |
| | $90^{th}$ | 223 | 12.48 | 4938 | 3.14 | 26678 | 13 |
| | $95^{th}$ | 217 | 12.67 | 4950 | 3.19 | 26888 | 13 |
| *76;18* | $25^{th}$ | 142 | 12.04 | 4522 | 2.86 | 24802 | 13 |
| | $50^{th}$ | 152 | 11.45 | 4416 | 2.81 | 25015 | 13 |
| | $75^{th}$ | 179 | 13.14 | 5184 | 3.36 | 27121 | 13 |
| | $90^{th}$ | 245 | 15.36 | 6129 | 4.20 | 30573 | 15 |
| | $95^{th}$ | 267 | 15.24 | 6093 | 4.15 | 30267 | 15 |
| *64;18* | $25^{th}$ | 136 | 11.08 | 4090 | 2.57 | 24108 | 13 |
| | $50^{th}$ | 136 | 10.98 | 4175 | 2.67 | 25325 | 13 |
| | $75^{th}$ | 198 | 13.13 | 5324 | 3.41 | 27395 | 13 |
| | $90^{th}$ | 246 | 14.30 | 5808 | 3.90 | 29655 | 14 |
| | $95^{th}$ | 251 | 13.82 | 5645 | 3.82 | 29418 | 14 |
| *52;33* | $25^{th}$ | 142 | 11.37 | 4315 | 2.77 | 25467 | 13 |
| | $50^{th}$ | 142 | 11.65 | 4335 | 2.81 | 26388 | 13 |
| | $75^{th}$ | 181 | 11.88 | 4540 | 2.95 | 27255 | 13 |
| | $90^{th}$ | 221 | 13.28 | 5300 | 3.50 | 28694 | 13 |
| | $95^{th}$ | 270 | 14.15 | 5755 | 3.79 | 29404 | 14 |
| *40;33* | $25^{th}$ | 149 | 11.36 | 4264 | 2.74 | 25619 | 13 |
| | $50^{th}$ | 176 | 12.33 | 4845 | 3.11 | 27089 | 13 |
| | $75^{th}$ | 205 | 14.24 | 5743 | 3.83 | 29310 | 14 |
| | $90^{th}$ | 252 | 15.04 | 6253 | 4.33 | 30921 | 14 |
| | $95^{th}$ | 258 | 15.09 | 6383 | 4.47 | 31288 | 15 |
| *28;33* | $25^{th}$ | 149 | 12.45 | 5012 | 3.35 | 28503 | 13 |
| | $50^{th}$ | 168 | 13.54 | 5514 | 3.79 | 30008 | 13 |
| | $75^{th}$ | 183 | 13.64 | 5597 | 3.79 | 29777 | 13 |
| | $90^{th}$ | 230 | 14.11 | 5763 | 3.92 | 29949 | 13 |
| | $95^{th}$ | 250 | 15.86 | 6569 | 4.61 | 31508 | 16 |
| *16;33* | $25^{th}$ | 125 | 11.53 | 4354 | 2.84 | 27056 | 13 |
| | $50^{th}$ | 160 | 13.12 | 5333 | 3.59 | 29114 | 13 |
| | $75^{th}$ | 178 | 13.66 | 5543 | 3.74 | 29205 | 13 |
| | $90^{th}$ | 219 | 13.86 | 5633 | 3.87 | 30036 | 13 |
| | $95^{th}$ | 246 | 14.95 | 6107 | 4.22 | 30714 | 15 |
| *4;33* | $25^{th}$ | 153 | 12.94 | 5260 | 3.51 | 28586 | 13 |
| | $50^{th}$ | 167 | 13.77 | 5627 | 3.77 | 29156 | 13 |
| | $75^{th}$ | 208 | 15.02 | 6160 | 4.22 | 30621 | 15 |
| | $90^{th}$ | 248 | 15.81 | 6580 | 4.52 | 30832 | 15 |
| | $95^{th}$ | 272 | 16.16 | 6816 | 4.66 | 31131 | 16 |

To continue the impact evaluation, the total damage across the different forecast types and lead times is also outlined in Figure 28. In the figure, an increasing trend can be observed over the lead times and over the percentiles, where the percentiles show a large spread. This spread contributes to the uncertainty description of the event as the percentiles deliver a collection of probable outcomes. This highlight the strong message that the control forecasts are different than the median of the ensemble, and that it should always be considered next to the ensemble members to address the complete uncertainty. Once again, the 28h lead time highlights that the control forecast has a large peak compared to the other lead times.

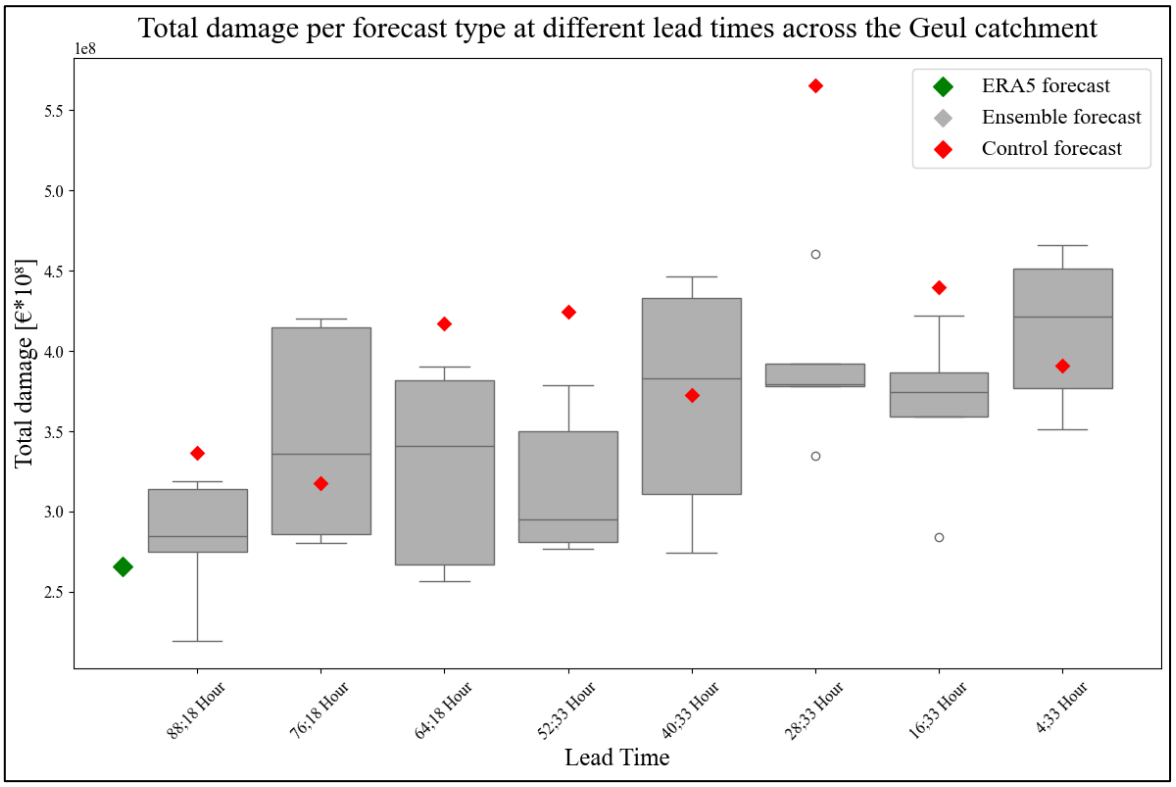


*Figure 28: The total damages across the Geul catchment displayed for each forecast type and lead time. The total damages include the damage to forests, agriculture, infrastructure and building structures, which are derived from the vulnerability curves. The control forecast is represented in Red, the ERA5 data in green, and the ensemble percentile forecasts with the grey boxplots.*

Building on the total damage visualization, the different parameters of Table 3 are presented in Appendix E. The figures E.1, E.2, and E.3 illustrate the total damage per function, the total number of buildings that are hit, and the median damage per impacted building. For the total damage per function in Figure E.1, a similar trend is presented as described for the total damages across the lead times. The figure indicates that the damages to the buildings are the largest part of the total damage since they have the highest impacts compared to the other functions. Another noteworthy finding is that all the three figures show a decreasing spread moving toward the event. Where again the 28h lead time of the control forecast shows a large peak compared to the other forecasts. Specifically, looking at Table 3 the estimates of the control forecast at that lead time lay far from the other forecasts, which in the end indicates that these estimates are outliers.

Moving to the critical facilities, the amenities in the catchment are merged with the flood maps to count the number of hits across all the forecast types and lead times. Figure 29 presents the hit intensity per critical facility, where most hits are located more downstream of the catchment. Additionally, schools are frequently hit whereafter several fire stations and police stations are hit. Interestingly, almost all the facilities are hit with every forecast, except for a clinic in Gulpen which is hit only once.

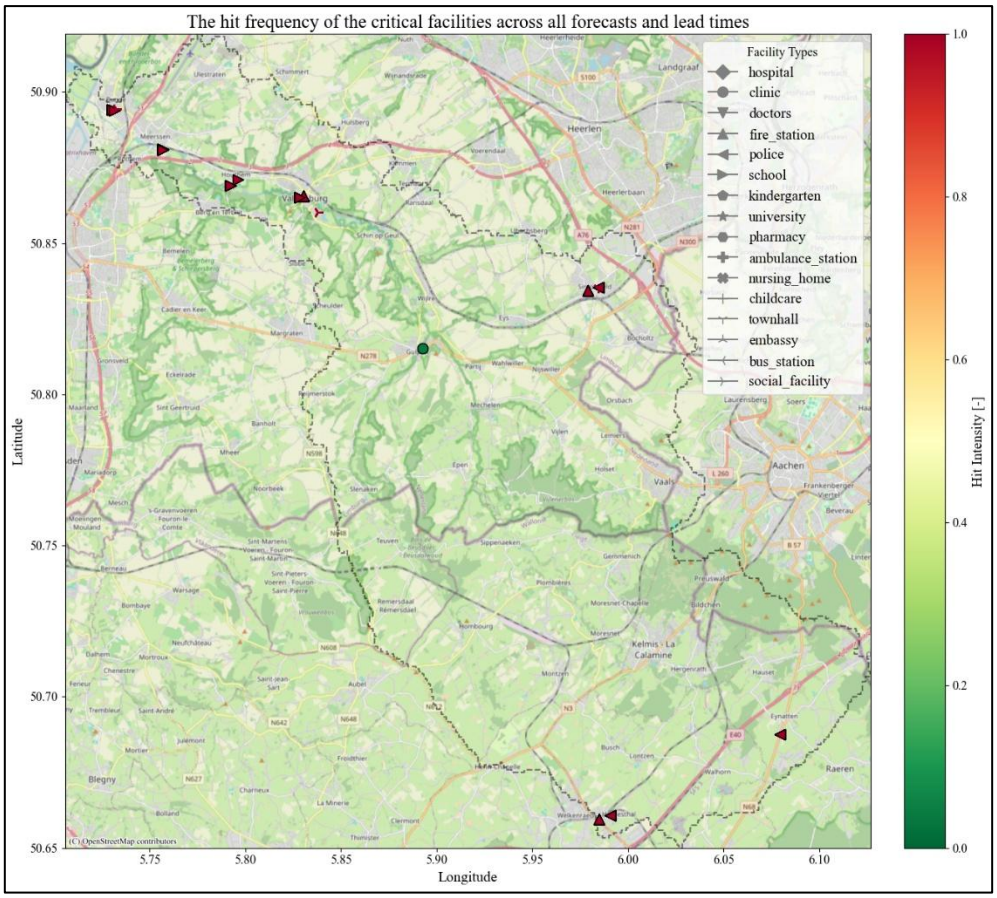


*Figure 29: The hit intensity of critical facilities across all the forecast types and lead times with the shape describing the critical amenity and the colour the ratio of hits across the entire collection of forecasts.*

These findings underline that the estimates of the impacts are aligned across the lead times resulting in a lower range of possible estimates reaching the lower percentiles. In addition, together with increasing impacts over the lead times the reliability of the estimates increases. To further examine the practical value of the modelling chain, the next section explores the impact chain for a municipality specifically. This is of an added value to explore the forecasts' performance at a more localized spatial scale to assess how well the system captures impacts at community or building levels. In the end, the results will be expanded with a finer spatial resolution, offering a deeper understanding of forecast utility in areas with high vulnerability and exposure.

### 5.3.1 Valkenburg as case study for detailed impact analysis

To complement the catchment-scale analysis, an additional in-depth evaluation is conducted for the municipality of Valkenburg. This section narrows the focus to the localized validation of the forecasted flood maps and more in-depth impact assessments. Specifically, it evaluates the damages to individual functions or buildings in the center of Valkenburg. Moreover, additional supporting figures are included in Appendix F.

First, Figure 30 presents the forecasted flood map for Valkenburg based on the 95$^{th}$ percentile of the ensemble forecast with an 88h lead time as an example. The figure illustrates extensive flooding in the central part of Valkenburg with water levels reaching up to 2 to 2.5 meters. In addition, significant flooding is simulated along the riverbanks near Houthem, where the flood depths in this area are relatively low compared to those in Valkenburg.

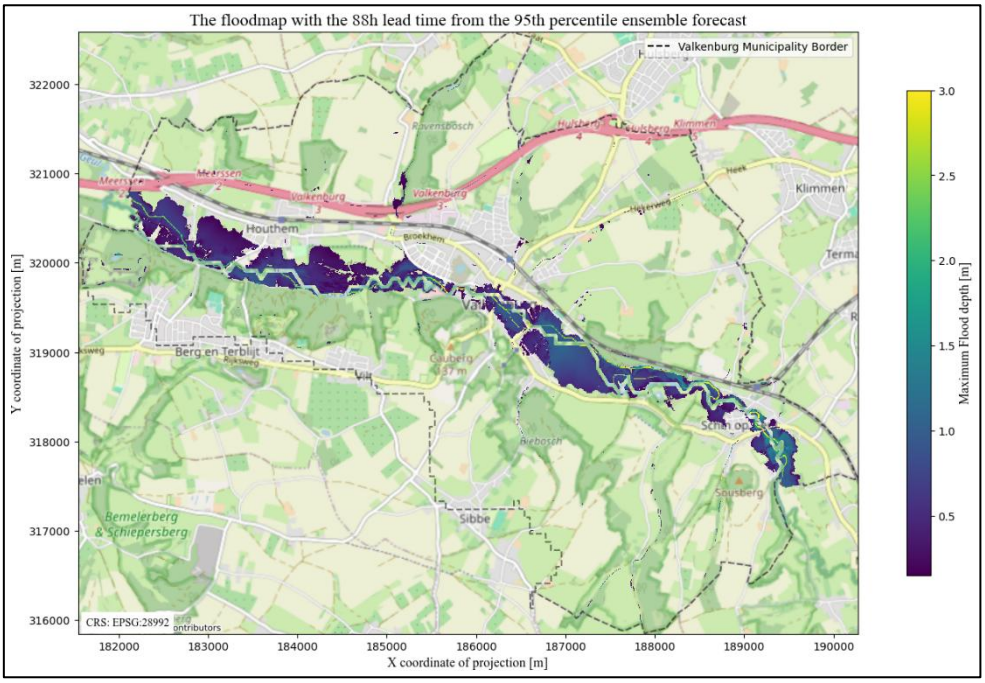


*Figure 30: Simulated flood map across the Valkenburg Municipality based on the 95$^{th}$ percentile forecast with a lead time of 88 hours. The flood depths are spatially represented in meters and the river geometry has been removed from the visualization to emphasize the flooding outside the riverbanks.*

Building on the flood map presented in Figure 30, Figure 31 visualizes the validation results of the flood extents across all the lead times and forecast types. Compared to the broader catchment-level evaluation presented in Figure 24, the performance metrics depict a significant improvement. Specifically, both the hit rates and CSI indices have increased with a peak performance (CSI > 0.8) at 16h lead time, while simultaneously the false alarms decreased. This indicates that the forecasts show an accurate performance (CSI > 0.7) in detecting the flood areas within the municipality of Valkenburg, which in the end highlight a stronger alignment between simulated and observed flooding patterns. In addition, Figure F.1 illustrate the spatial evaluation of the 88h lead time with the 95$^{th}$ percentile ensemble forecast as input. This figure supports the strong alignment where only minor misses are displayed on the edges of the extent, while no false alarms are detected.

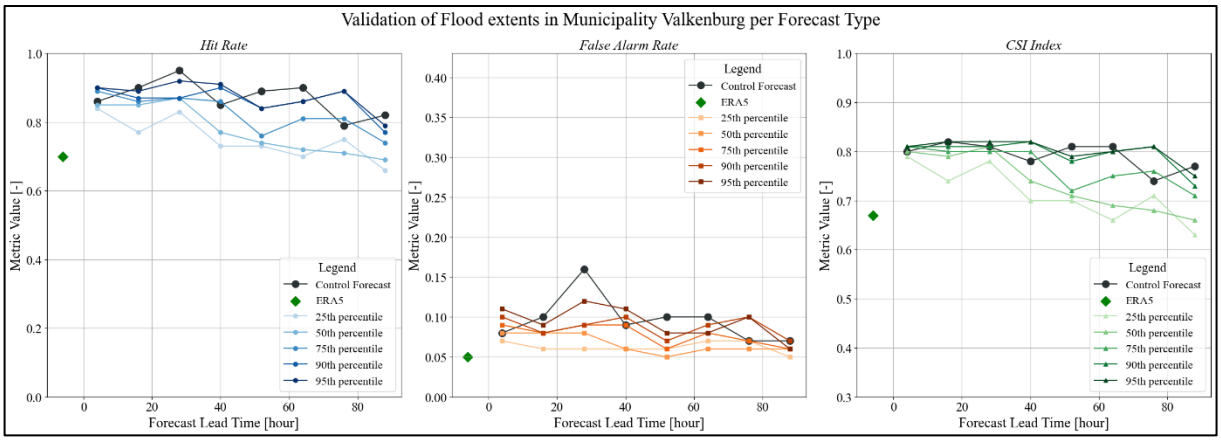


*Figure 31: The performance of the flood extents in Municipality Valkenburg per forecast type over the different lead times and percentiles with lighter colours representing the lower percentiles, the darker colours the higher percentiles, anthracite representing the control forecast and the green shape represents the ERA5 scores.*

Figure 31 further reveals that across all three-performance metrics, the forecast performance generally declines as lead time increases. Additionally, there is a clear performance discrepancy between the higher percentiles (P75, P90 & P95) and the lower percentiles (P25 & P50), where the lower percentiles constantly showing lower scores and a steeper decline over increasing lead times. Moreover, the control forecasts have a similar performance as the higher percentiles, although the 28h lead time still shows a peak in the hit rates and false alarms. However, this is again not reflected in the CSI index. Note that for the higher percentile ensemble forecasts and the control forecast the CSI index remains above 0.7, suggesting a strong alignment throughout the range of the forecasts. Regarding the ERA5 simulation, the CSI index depict a score of just below 0.7, which can be interpreted as an intermediate performance. Furthermore, the lower percentiles (P25 & P50) drop below the CSI threshold after the 40-hour lead times.

In contrast to the increased performance of the flood extent simulation, Figure 32 displays the reduced performance in simulating flood depths within the municipality of Valkenburg. In general, both the MAE and RMSE are higher compared to the catchment-scale flood depth analysis in Figure 27. The figure outlines that the MAE values range from 0.55 to 0.7 meters and the RMSE values between 0.8 and 0.95 meters. Like earlier findings, the higher percentiles show more stable performance across lead times, where in contrast the lower percentiles have greater variability and higher errors. Additionally, the ERA5 scores correspond more with the lower percentiles. Nonetheless, for the in-depth analysis the control forecasts exhibit a more moderated performance compared to Figure 27, which is more in line with the trend and magnitudes of the ensemble forecasts. The NSE values further reflect the reduced performance, where the values remain around 0.1 to -0.1 for the higher ensemble percentiles and control forecasts. For the lower percentiles and the ERA5 simulation, the NSE values further decreases up to -0.4. This indicates that in terms of the flood depth simulations, the model performs poorly against the survey data for both the catchment scale and this in-depth analysis for Valkenburg. In addition, Figure F.2 presents the spatial evaluation of the simulated flood depths shown in Figure 30. Note that the largest errors are concentrated in the center of Valkenburg, which suggests that the model constantly overestimates the flood depths compared to the survey responses.

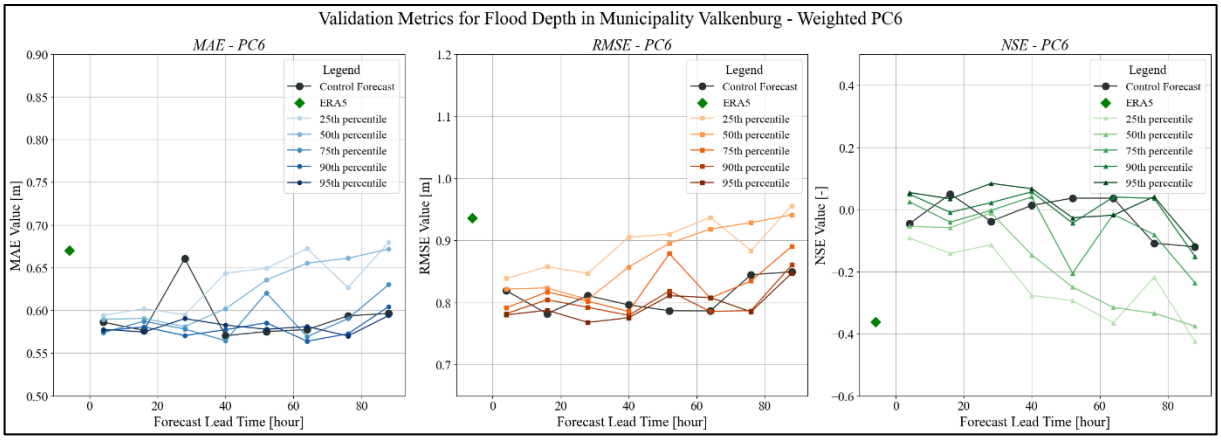


*Figure 32: Weighted evaluation metrics of forecasted flood depths across the municipality of Valkenburg using household survey data, shown for PC6 postcode levels. Each subplot displays a metric with the Mean Absolute Error (MAE) (blue at left), Root Mean Square Error (RMSE) (orange at centre), and Nash-Sutcliffe Efficiency (NSE) (green at right) for ensemble percentile forecasts (gradient of the colour), ERA5 (green shape) and control forecasts (anthracite) over various lead times. Metrics are presented in a weighted form to account for the number of survey responses per postcode.*

Building on the evaluation of the flood maps, the analysis now shifts focus to the forecasted impacts at a more detailed level. First, Figure F.3 displays a similar trend for the total damages across the lead times compared to Figure 28. Looking at the median ensemble, it appears that the municipality contributes approximately 25 to 30% to the total damages' prediction in the entire impact assessment in the Geul. Moreover, Figure 33 presents the hit intensity per critical facility across all forecast types and lead times in the municipality. The figure reveals that all the forecasts (hit intensity of 1) predicted impacts for several facilities located on the southern riverbank in Valkenburg.

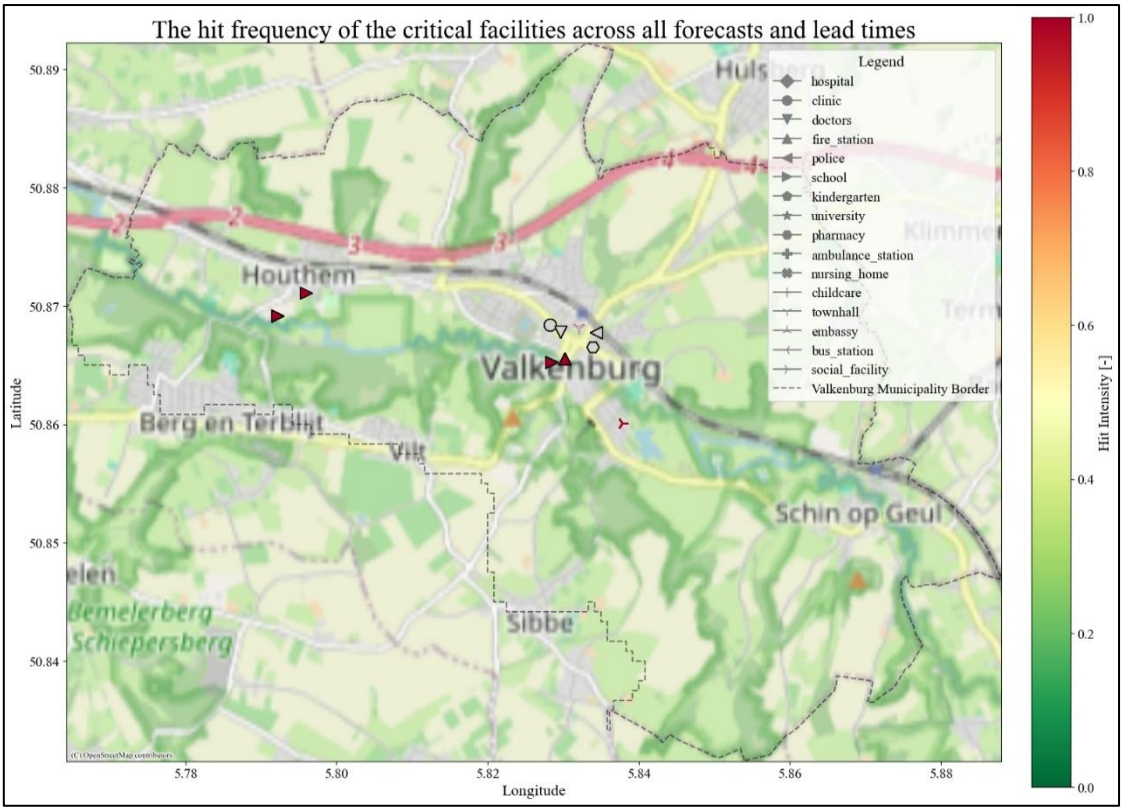


*Figure 33: The hit intensity of critical facilities across all the forecast types and lead times in the municipality of Valkenburg with the shape describing the critical amenity and the colour the ratio of hits across the entire collection of forecasts. Transparent shapes show the critical facilities that are not hit by any forecast.*

To enable targeted warnings for specific regions, stakeholders, and facilities, it is essential to understand flood impacts at the building level. Therefore, the three affected key facilities in Valkenburg (Figure 33) are the primary school, the fire station, and the social facility 'Oosterbeemd', which are selected for further analysis. Figure 34 outlines the forecasted water depths at the exterior walls of the buildings across the lead times and forecast types. For the fire station, a consistent pattern emerges across both the ensemble and control forecasts, where they show a minor increase over the lead times with a declining spread. Additionally, although the deterministic realization of the control forecast, the deviation with the ensemble spread is not significantly high. Besides, note that ERA5 also simulated a water depth which corresponds with the other forecast types. From the longest lead time (88 hours) onward a depth of approximately 2.5 meters is forecasted, which ultimately rises to 2.8-3 meters by the 4-hour lead time. Although the model seems to overestimate the flood depths, Figure 34 indicates the strength of combining the ensemble and control forecast.

Furthermore, a similar trend and forecast behaviour is observed for both the primary school and care facility. Apart from the lower water depth predictions, in particular the primary school shows a larger uncertainty in the ensemble forecasts with larger spreads in water depths up to the 40-hour lead time. In addition, the extraordinary behaviour of the control forecast at the 28-hour lead time is also present in the water depth for these two facilities. Besides, note that the ERA5 simulation does not predict flooding at the location of the primary school.

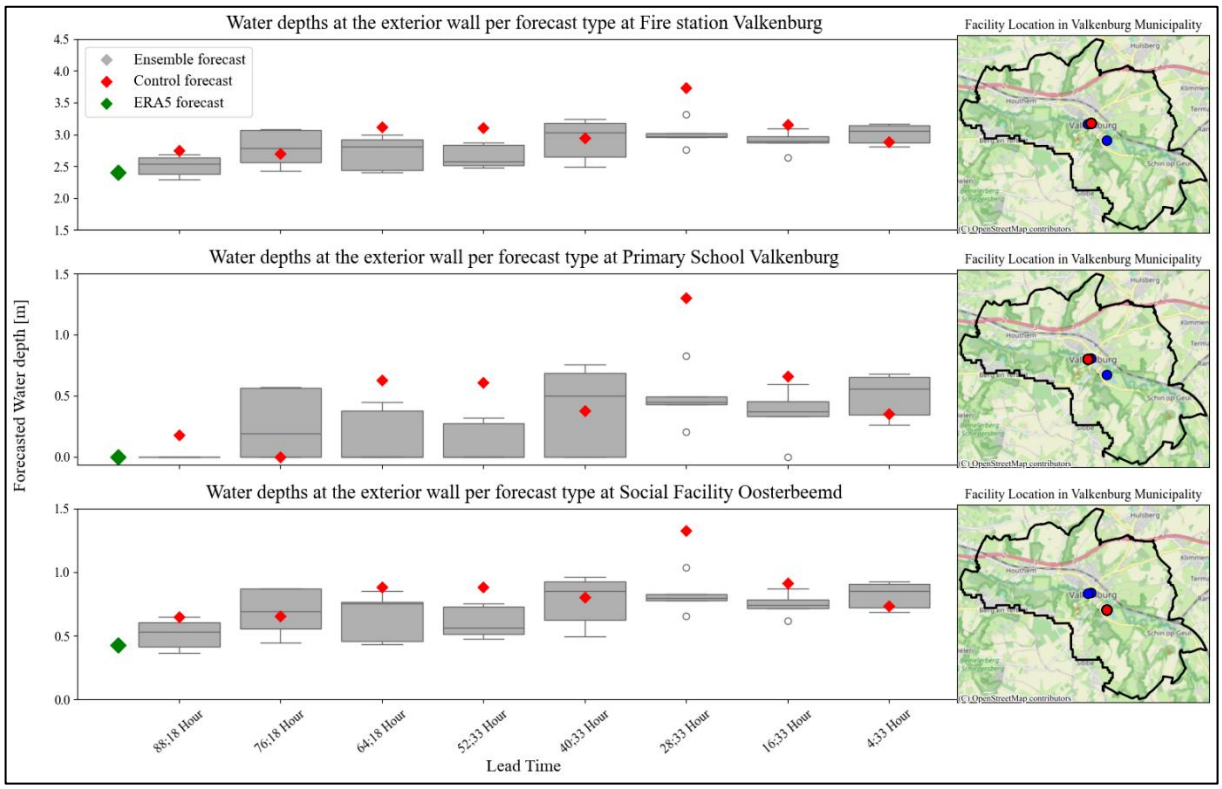


*Figure 34: The predicted water depths at three selected critical facilities in Valkenburg across all forecast types and lead times: fire station Valkenburg (top figure), primary school (centre figure), and social facility Oosterbeemd (bottom figure). The control forecast is represented in a red diamond, the ERA5 data in a green diamond, and the ensemble percentile forecasts with the grey boxplots. The facility' location is visualized in a map with a red circle presenting the location of the facility in the corresponding left graph and with blue for the other facilities presented in the figure.*

Building on the critical facility analysis, the flood probability per building in the centre of Valkenburg is displayed in Figure 35 for three flood depth thresholds 0.1 m, 0.95 m, and three meters. The figure reveals that for the 0.1 m and 0.95 m thresholds, the number of buildings with higher flood probability increases as lead time shortens. A substantial number of buildings already show a 100% exceedance of 0.1 meter, where first for the longest lead times only the buildings near the riverbank show this probability. However, as the lead time decreases more buildings get affected. Regarding the 0.95 m threshold, the exceedance probabilities vary more often across buildings, indicating that for some buildings it is uncertain that this depth will be exceeded. In contrast, the 3-meter threshold is often not exceeded with only a few buildings showing a 100% probability from the longest lead times onward. This indicates that severe flood depths of three meters are not predicted frequently during this event.

Subsequently, a similar trend is illustrated in Figure 36 where the impact probability is calculated for the ensemble forecasts using financial damage thresholds of €100,000, €500,000, and €1,000,000. As with the flood depth exceedance probability presented in Figure 35, the findings indicate a clear increase in the number of buildings with high probabilities near the riverbank as the lead time shortens. For the lower thresholds, many buildings have a high probability for surpassing the €100,000 threshold or in lesser amounts for the €500,000 threshold. The €1,000,000 threshold reflect a similar pattern of significant impacts as is indicated in the 3-meter flood depth threshold. These results suggest that ensemble forecasting provides valuable probabilistic insights into the financial flood impacts at a building level.

Ultimately, the final findings presented in Figures 34, 35 and 36 indicate that, despite model uncertainties, the ensemble forecasts constantly forecast flood depths at a building level well in advance of the event with a decreasing uncertainty closer to the event. This early prediction of water depths and impacts suggests that impact-based warnings could be issued on time for early action.

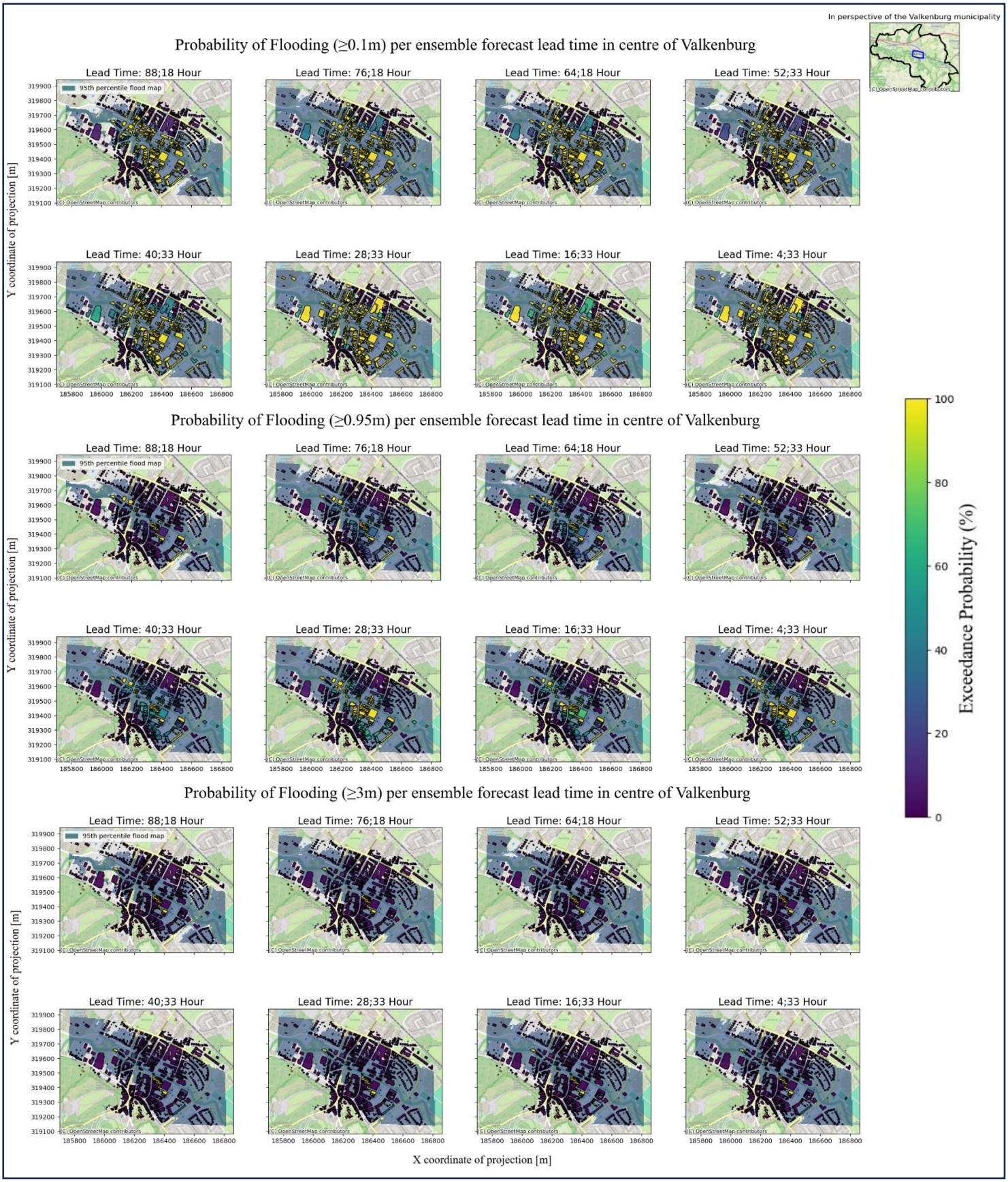


*Figure 35: The exceedance probability of flood depths for individual buildings in the centre of Valkenburg across three water depth thresholds: 0.1 m (top figure), 0.95 m (centre figure), and 3.0 m (bottom figure). In addition, the area for the in-depth evaluation (blue line) is presented in the top right figure in perspective with the municipality border (black line). Besides, to improve the understanding of the figure is the flood map (in ice blue) presented with the 95$^{th}$ ensemble percentile as an example for each lead time.*

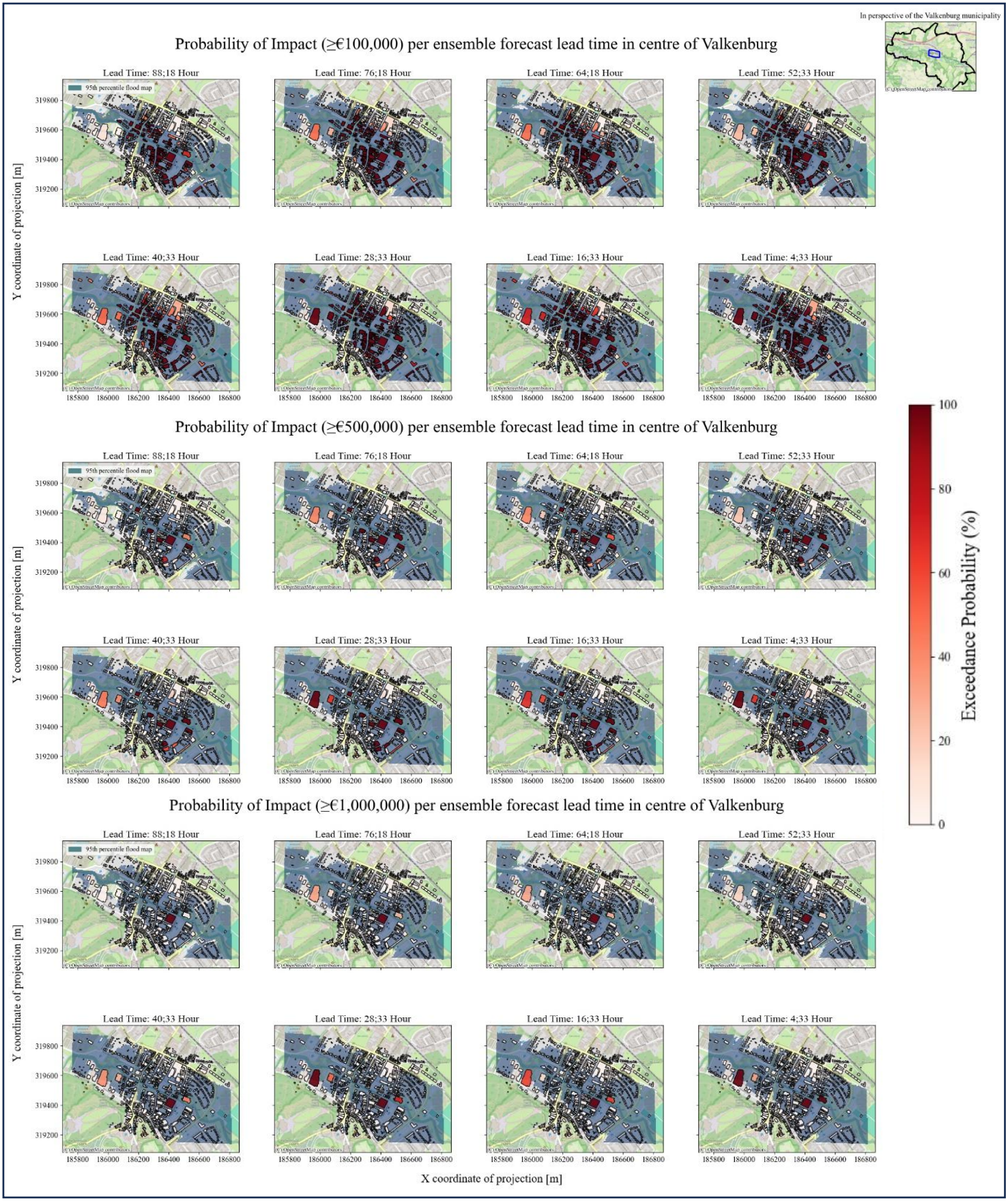


*Figure 36: The exceedance probability of financial impacts for individual buildings in the centre of Valkenburg across three impact thresholds: €100,000 (top figure), €500,000 (centre figure), and €1,000,000 (bottom figure). In addition, the area for the in-depth evaluation (blue line) is presented in the top right figure in perspective with the municipality border (black line). Besides, to improve the understanding of the figure is the flood map (in ice blue) presented with the 95$^{th}$ ensemble percentile as an example for each lead time.*

# 6. Discussion

This study has evaluated the performance of a flood impact-based forecasting model for the Geul catchment, using the GEB framework. The modelling framework consisted of the initialization of the hydrological CWatM model (out of scope), the hydrodynamic SFINCS model, and a loss model to translate the hydrological outputs to impacts. This chapter elaborates the limitations and the implications of the study to assess the results in a broader perspective. In addition, the chapter discusses in detail what is needed for future studies (academical recommendations) and what is needed to get the developed system operational (practical recommendations). First, the limitations are discussed to provide a more nuanced understanding of the findings for the entire modelling chain.

## 6.1 Limitations in the modelling framework

Addressing the limitations of this study is essential for the interpretation of the key findings. Therefore, the main uncertainties inside the input data, SFINCS model, and in the developed chain are elaborated.

***Limitations in the precipitation data***

First, this study has used different sources of rainfall data to force the GEB framework. Starting with ERA5 ECMWF reanalysis data, this data is used to create a reference baseline in the study. Although ERA5 is widely used in the fields of meteorology and hydrology and provide physical, spatial, and temporal consistent long time series at a global scale (Gomis-Cebolla et al., 2023), several studies have discussed important limitations that must be considered when reanalysis are used as a reference (e.g. Bližňák et al., 2022; Da Silva et al., 2024; Gomis-Cebolla et al., 2023; Wu et al., 2022). These studies conclude that ERA5 often tends to underestimate extreme precipitation intensities while in contrast it overestimates light and moderate rainfall events, particularly in summer conditions. In addition, it is important to note that ERA5 is a reanalysis dataset which is in fact a reconstruction. Therefore it cannot be used for operational decision-making (Hersbach et al., 2020).

Apart from the reanalysis data, this study also utilized historic operational ECMWF precipitation forecasts rather than hindcasts to simulate the 2021 flood event. However, the used forecasts deviated from the original operational forecasts since the applied forecasts uses full gaussian grids (e.g. F640) instead of the original octahedral reduced Gaussian grids (e.g. O640). While both grid types have similar resolutions (e.g. 640 lines), the octahedral grids ensure higher computational efficiency with a non-uniform longitudinal resolution leading to spatial irregular grids. These grids represent fewer points near the poles and more near the equator (Malardel et al., 2015). In contrast, full gaussian grids maintain regular spacing in both longitudinal and latitudinal direction (Malardel et al., 2015). For this reason, regular spacing is an advantage in terms of expecting spatial data on regular and constant coordinates for small study areas as the Geul catchment. Nonetheless, this approach required an interpolation from original grids towards regular by the ECMWF within the download of the forecasts. This may have shifted local rainfall on a small scale, leading to a relatively more distributed rainfall over the regular grid cells. Moreover, Van Heeringen et al., (2022) concluded that the extreme

precipitation was forecasted at different moments for the Geul catchment by different meteorological services. For example, ICON-EU (from the German Meteorological Service) predicted the extremes earlier compared to ECMWF. As a result, using a different meteorological service could lead to different impact estimates.

***Limitations in the SFINCS model***

To simulate flood maps for each forecast, SFINCS (v2.1.3) is applied due to its efficient character for simulating hydrodynamical processes with limited computational cost and good accuracy. However, originally the model is developed for the simulation of compound coastal flooding (Leijnse et al., 2021). Therefore, it is uncertain whether the pluvial components of the 2021 floods is fully captured in the modelling since the model is not evaluated or proven effective yet in small riverine catchments like the Geul. Although, the pluvial components are proven in coastal areas (e.g. Beveren, 2022; Sebastian et al., 2021) one notable limitation in the hydrodynamic modelling component of this study is the absence of detailed water infrastructure such as culverts, weirs in the SFINCS setup. Specifically, this simplification has likely contributed to the substantial number of false alarms observed in the estuary of the Geul of the simulated flood maps. Previous studies (e.g. Hailemariam et al., 2013; Fallowfield & Motta, 2024) have also identified the absence of water infrastructure as a factor leading to flow routing errors and water accumulation in flood simulations. In addition, analysing the observed extent it appears that the extent decreases after the Juliana channel. The reason for this is that during the 2021 flood the Geul estuary saw extremely high-water levels at peak times due to a combination of high discharges on the Geul and the Meuse rivers. De Jong and Asselman (2022) investigated in the estuary the effects of local barriers in the landscape such as bridge abutments. They concluded that the local water structures caused rising water levels up to several meters, allowing the water moving more northwards in the direction of Bunde. A vital role played the siphon under the Juliana channel where the water from the Geul flows into the Meuse. This structure appeared to have limited capacity which could not manage the discharge of the flood event. In addition, tubes of the siphon were clogged with silt and other natural materials which decreased the capacity even more (De Jong & Asselman, 2022).

Furthermore, the overestimated flood depths as shown in Figure 27 and 32 could be the result of the missing structures (Figure D.3). However, it is also important to acknowledge the limitations of the used survey data itself since the observations are derived from survey responses. These responses lack data of precise local observations due to its anonymous character, resulting in spatial uncertainty. Ultimately, this limits the accuracy of the validation in areas with high spatial variability in water depths such as urban areas.

***Limitations in the impact estimations and the developed modelling chain***

Finally, this study did not apply national used standards such as the Dutch Standard Damage Method (SSM). This is a national established model used to estimate flood damages across various assets and buildings by the government (Van Den Braak et al., 2020). As a result, the impact estimations rely on more generic vulnerability curves. Apart from the residential curves constructed from the surveys of Endendijk et al. (2023), these generic curves may not reflect

the most recent and location specific insights. Knowing the limitations of the survey and the depth-damage curves, the impact estimations could have been underestimated.

While the model chain presented in this study demonstrates an intermediate (Figure 25) to strong (Figure 31) performance for the Geul catchment flood extent, its transferability to other regions may be constrained due to different hydrological triggers (e.g. snow melt or rain on saturated soils). Moreover, it could be also constrained due to different climates or landscapes (e.g. fewer altitude differences) and its dependency on high quality spatial datasets. Changes in these factors could affect the performance of the chain. In addition, considering the applicability of the impact module, the transferability is limited due to the usage of location specific curves regarding the residential impacts and due to the lack of existing detailed vulnerability curves such as for critical facilities like fire stations (Merz et al., 2020; Nirandjan et al., 2024).

## 6.2 Implications of the key findings

The identified limitations highlight several implications for the interpretation of the results of the study. First, the underestimation of ERA5 in perspective of this study could have affected the accuracy of the baseline flood map and the evaluation of the forecast performance. Hence, the limitations of the reanalysis must be acknowledged when interpreting the flood extents and forecasts evaluations. It is important to nuance the difference between ERA5 and the different forecasts as this only illustrates what the different inputs are in SFINCS. The evaluation of the different forecasts against the ERA5 reanalysis data provided a quantitative overview of the input data used in the created modelling chain. The findings of the evaluation mainly present its potential in the joint use of deterministic and probabilistic forecasts. Besides, it addresses the added value of ensembles since they describe the uncertainty of weather forecasting. Several studies have already addressed the importance of using ensemble forecast inside risk modelling (e.g. Dawkins et al., 2023; Teja et al., 2023). This study indicates that the most extreme percentiles together with the control forecasts caused the best model performances where additional uncertainties are incorporated. This finding corresponds with several studies (e.g. Busker et al., 2025; Najafi et al., 2024).

Furthermore, the use of a full gaussian grid could have led to shifts in local rainfall intensities which in the end could have affected the flood extents downstream.

In the SFINCS model the missing structures caused the multiple false alarms more downstream in the estuary of the Geul. This affected the evaluation of the flood maps and including these structures could have a positive influence on the reliability of the forecasted flood extents. Additionally, Figure 31 already presented an increased fit (CSI > 0.7) of the modelled extent for the municipality of Valkenburg (where the downstream parts were not included). Moreover, apart from the limitations inside the used observations for the water depth evaluation, the inclusion of water structures could also have a positive influence on the results. For example, strange water depths are simulated in the ends of the tributaries due to the absence of water infrastructure such as culverts. (e.g. Figure C.3). As a result, the simulated flood depths within this study show high deviations compared to the observed flood depths (Figures 26,27, 32, and F.2). The errors within the 6-digit postal code areas are not negligible with on average a weighted MAE error of 0.5 m and a RMSE 0.72 m for the entire catchment (Figure 27). In

addition, in Valkenburg (Figure 32) the errors are on average 0.63 m and 0.85 m and therefore need reflection.

To put the flood map evaluation results in perspective, the CSI index and the RMSE values obtained in this study ranged from approximately 0.5 (for the entire catchment in Figure 25) to 0.8 (for Valkenburg in Figure 31) for the CSI index. Regarding the weighted 6-digit postal code RMSE values, the values ranged from 0.6 m (for the entire catchment in Figure 27) to 0.95 m (for Valkenburg in Figure 32). These values fall within the range of results reported in several other studies (e.g. Bentivoglio et al., 2023; Bernhofen et al., 2018; Woo et al., 2025). This comparison indicates that the developed forecasting framework performs within a realistic and acceptable range and shows its potential, while there is still room for improvements. Therefore, the limitations in particularly for the simulated flood depths and the extent in the estuary must be acknowledged. These two findings have the greatest uncertainty which could impact the entire performance of GEB.

The findings of this study demonstrate that the GEB framework can simulate reliable flood extents and estimate building-level impacts using ensemble forecasts. Beyond the presented case study, the modular structure of the GEB framework offers a strong potential for the transferability to other catchments (J. A. De Bruijn et al., 2023). Compared to established operational systems like Delft-FEWS (Flood Early Warning Systems), GloFAS (Global flood Awareness Systems), and Delft-FIAT (Flood Impact Assessment Tool), GEB offers a more modular and open-source alternative where hydrodynamic modelling is linked with exposure and vulnerability analysis. While FEWS is a global established real-time operational hydrological forecasting system (e.g. Chowdhury et al., 2022; Deltares, 2025c) and FIAT an established operational system for estimating flood impacts (Deltares, 2025d; Nederhoff et al., 2024). FIAT uses a similar approach in estimating the impacts, this shows that this methodology is already embedded in operational workflows under different climate zones and for different hydrological triggers. Unlike FEWS and FIAT, is GEB not yet embedded in such operational systems. However, this study highlights the potential of GEB for extending forecasting and impact tools in one framework what goes beyond existing separated systems. In addition, the validation of the GEB framework across different catchments must be emphasised before it can be used in operational services and before it is equivalent in status with the existing operational services.

## 6.3 Recommendations

This section provides suggestions for further investigations for the developed IBFFWS chain in the Geul catchment. First, general academical recommendations and further research regarding the methodology are given. Secondly, the practical recommendations are provided for the JCAR ATRACE program, IVM, Deltares, and the water authorities in the Geul catchment.

### 6.3.1 General recommendations and further research

Based on the identified limitations and implications, the following recommendations are proposed to guide future academic research. Future studies should test the GEB framework in a variety of catchments globally to examine its transferability across differences in catchment

size, hydrological triggers and climates. Simultaneously, the components of the framework should be benchmarked against established modelling systems like FEWS, FIAT or GloFAS. Moreover, research should further investigate the representation of hydraulic structures such as culverts and siphons in hydrodynamic models to improve flood map accuracy. Moreover, the facilities included in this study just show the potential of what can be included inside an impact-based assessment on a building level. This study forms a first concept of the possibilities where it is recommended to elaborate the vulnerability curves for critical facilities such as fire stations to improve the accuracy of the impact estimates for the critical facilities in catchments. Lastly, translating the impact-based forecasts into early action it is recommended to further elaborate this chain by connection real-time decision tools such as agent-based models to enhance the response and communication on the ground.

### 6.3.2 Recommendations for practical & operational implementations

From a practical perspective, this study highlights several key recommendations for future flood preparedness in the Geul catchment. First, to improve the developed chain it is recommended the integration of local hydraulic water structures inside the model to enhance the accuracy and reliability of the simulated flood maps. Additionally, it is recommended to refine the evaluation of the flood maps with calibrated flood maps from other sources such as other validated model simulations and remote sensing data to improve the accuracy of the flood maps. Moreover, the use of ensemble percentile forecasts is proved valuable in communicating uncertainty in both flood extents and impacts estimations. However, since the generation of flood maps for entire ensemble forecasts is not included in the study due to the required computational time. It is recommended to further investigate the description of the uncertainty across the model chain (with the right octahedral gaussian grid type) and how the outcomes differ compared to the outcomes of the percentile approach (with the full gaussian grid type). In the scope of this study, it has not been ruled out that the percentile approach deliver different outcomes compared to the entire ensemble. In addition, it will be an added value if the findings of this study can be compared with similar findings based on the usage of different meteorological services such as the DWD or KNMI. Subsequently, it is recommended to include more critical facilities inside the analysis since for example the energy sector and industries are not included right now. Furthermore, it is advised to evaluate the simulated impact estimates with the SSM standard described by Van Den Braak et al. (2020) to improve the reliability of the impact estimates. Finally, it is recommended to further clarify the needs in a multi-disciplinary stakeholder analysis to translate the impacts in effective measures on the ground. A new objective could thus be that these study findings lead to activation in the region with the different stakeholders engaging in a conversation about these results.

# 7. Conclusion

This study investigated the development of an impact-based flood forecasting system for the Geul catchment where data of the 2021 flood event has been used to evaluate the outcomes of the model chain. By answering the formulated sub question the chain is developed, whereafter an answer is given to the main research question.

***RQ1: How well do historic operational ensemble and deterministic forecasts relate to the observed precipitation patterns and totals of the July 2021 flood event, compared to ERA5 reanalysis data?***

The historic operational ECMWF forecasts do not reproduce the patterns of the ERA5 reanalysis data since the evaluation reveals meaningful differences in terms of spatial distribution and intensity. These differences do not indicate lower quality but rather highlight the differences in forecast behaviour. The control forecast, with a deterministic character, often deviates from both the ensemble spread and ERA5 prediction, though it could capture extreme values which are missed in the ensemble percentile spread. In contrast, ensemble forecasts provide valuable insights into the forecast uncertainty. As demonstrated in the study and supported in the discussion, the combination of deterministic control forecasts with ensemble forecasts reinforces the interpretation of the probabilities in rainfall. Moreover, it is important to recognize that ERA5 is a reanalysis product which is in fact a reconstruction. Although it functioned as a reference in the study, ERA5 cannot be used in real-time operational decision-making.

***RQ2: What is the quality of flood forecasts from SFINCS for the July 2021 event at multiple lead times and how do variations in forecast data affect the outcomes of the SFINCS model?***

The evaluation for the quality of the SFINCS flood forecasts for the July 2021 event shows that the chain demonstrates a moderate performance for flood extents at the catchment scale with the highest accuracy observed for the higher percentiles (P90 & P95) and the control forecasts (CSI ~0.6). At the local scale in the municipality of Valkenburg, the performance has further improved (CSI > 0.7). Across all the forecasts, the simulated flood extents often exceed the ERA5 extent, which is influenced by higher forecasted rainfall volumes. The ensemble forecasts presented a moderate increase in flood extent across shorter lead times with a decreasing spread. This indicated the reduction in uncertainty closer to the event. In addition, The higher percentiles (P75-P95) and control forecasts constantly outperformed the lower percentiles (P25-P50) in detecting observed extents.

In contrast, the quality of the flood depth performance remained limited across all lead times and forecast types. The model overestimated water depths in both postal code detail levels (PC4 & PC6). With the more detailed PC6 areas the model overestimated the water depth even more particularly in urban areas, leading to high MAE (~0.5m), RMSE (~0.72m), and NSE near or below zero across the catchment. These deviations are caused by the absence of hydraulic structures (e.g. culverts, weirs, and siphons), which also affected the simulated extents due to the large number of false alarms in the estuary of the river. Additionally, the used survey data also introduced uncertainty due to its limited spatial accuracy of the reported water levels.

Although the weighted performance metrics improved the overall performance for both postal code detail levels, the flood depth simulation remained of insufficient quality.

Ultimately, these results highlight the joint strength of combining ensemble forecasts with control forecasts since they both proved capable in predicting the flood extents, though the 28h lead time of the control forecast is identified as an outlier. The joint strength certainly applies for the in-depth evaluation in Valkenburg showing a good performance quality.

***RQ3: How does the predicted flood impact vary across different rainfall forecasts and lead times, and what is the added value of using ensemble-based predictions for estimating local damages per function and area?***

The predicted flood impacts show a clear variation across the several forecast types and lead times, here the ensemble-based predictions offering a significant added-value in estimating the spatially spread of distributed damages and probabilities. An increasing trend is observed as lead time decreases and percentile increases where the affected buildings is reflected in the spatial flood extents.

Moreover, the control forecast frequently deviates from both the ERA5 and ensemble spread, which underlines its deterministic nature and potential to produce outlier scenarios. However, it still needs to be considered since it may capture extremes that are not represented in the ensemble percentiles.

At the building level, the ensemble forecasts constantly predict flood depths several days before the event with declining uncertainty closer to the event. This provides insights into the probable damage ratios per asset. Nevertheless, the absence of specific vulnerability curves for critical facilities limits the depth of impact interpretation. Apart from the probability of flooding per facility or building, specific damage estimates for these facilities are therefore not made in this study.

In general, the impact findings show that ensemble-based impact forecasting enables more nuanced, probabilistic and location specific flood risk assessments which are in the end crucial for early action and early warning applications.

***How could Impact-based Forecasts (IbF) from the hydrodynamic SFINCS model have been used to trigger effective early-warnings and actions, incorporating associated uncertainty for the 2021 flood in the Geul basin?***

In conclusion, this thesis primarily shows that Impact-based forecasting (IbF) systems generated with the SFINCS model could have supported earlier and more informed warnings and actions for the 2021 flood in the Geul catchment. It is then necessary that the model is combined with ensemble precipitation forecasts. By integrating forecast uncertainty into flood extent and impact simulations at both catchment and building level, the developed chain offers actionable insights up to 88 hours in advance (>3 days). In particular, critical facilities and high risk urban areas like Valkenburg can be supported with timely actions and targeted response. This underlines the value of the developed chain where it forms an added value in assisting the decision-making in flood risk management. However, this developed chain is not transferable

yet to all types of flooding (hydrological triggers) and need to be validated in other catchments and benchmarked against existing frameworks. The greatest uncertainty in the chain lies in the simulation of flood depths and the absence of water structures since these caused the highest errors in the chain. These uncertainties must be refined to improve the reliability of the chain. Apart from the existing uncertainties inside the forecasted depths, this study provides a promising proof-of-concept for operationalizing IBFFWS. This study marks an important step toward more precautionary and location-specific flood risk management.

# References


Alfieri, L., Zsoter, E., Harrigan, S., Hirpa, F. A., Lavaysse, C., Prudhomme, C., & Salamon, P. (2019). Range-dependent thresholds for global flood early warning. Journal of Hydrology X, 4, 100034. https://doi.org/10.1016/j.hydroa.2019.100034

Andreadis, K. M., Schumann, G. J. P., & Pavelsky, T. (2013). A simple global river bankfull width and depth database. Water Resources Research, 49(10), 7164-7168. https://doi.org/10.1002/wrcr.20440

Arduino, G., Reggiani, P., & Todini, E. (2005). Recent advances in flood forecasting and flood risk assessment. Hydrology and Earth System Sciences, 9(4), 280–284. Https://doi.org/10.5194/ hess-9-280-2005

Asselman, N., Van Heeringen, K., De Jong, J., Geertsema, T. (2022) Juli 2021 overstroming en wateroverlast in Zuid-Limburg: Eerste bevindingen voor Valkenburg, Geulmonding, Roermonding en Eygelshoven. Deltares, https://www.waterschaplimburg.nl/publish/pages/7013/juli_2021_overstroming_en_wateroverlast_in_zuid-limburg.pdf (in Dutch)

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2023). Rapid spatio-temporal flood modelling via hydraulics-based graph neural networks. Hydrology And Earth System Sciences, 27(23), 4227–4246. https://doi.org/10.5194/hess-27-4227-2023

Bermúdez, M., Neal, J. C., Bates, P. D., Coxon, G., Freer, J. E., Cea, L., & Puertas, J. (2017). Quantifying local rainfall dynamics and uncertain boundary conditions into a nested regional-local flood modeling system. Water Resources Research, 53(4), 2770–2785. https://doi.org/10.1002/2016wr019903

Berndtsson, R., Becker, P., Persson, A., Aspegren, H., Haghighatafshar, S., Jönsson, K., Larsson, R., Mobini, S., Mottaghi, M., Nilsson, J., Nordström, J., Pilesjö, P., Scholz, M., Sternudd, C., Sörensen, J., & Tussupova, K. (2019). Drivers of changing urban flood risk: A framework for action. Journal of Environmental Management, 240, 47–56. https://doi.org/10.1016/j.jenvman.2019.03.094

Bernhofen, M. V., Whyman, C., Trigg, M. A., Sleigh, P. A., Smith, A. M., Sampson, C. C., Yamazaki, D., Ward, P. J., Rudari, R., Pappenberger, F., Dottori, F., Salamon, P., & Winsemius, H. C. (2018). A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique. Environmental Research Letters, 13(10), 104007. https://doi.org/10.1088/1748-9326/aae014

Beven, K.J (2000). Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences, 4(2), 203–213.

Beveren, J. V. (2022). *Global Changes of Flood Hazard in Cities* (Master's thesis).

Bližňák, V., Pokorná, L., & Rulfová, Z. (2022). Assessment of the capability of modern reanalyses to simulate precipitation in warm months using adjusted radar precipitation. Journal Of Hydrology Regional Studies, 42, 101121. https://doi.org/10.1016/j.ejrh.2022.101121

Boelee, L., Lumbroso, D. M., Samuels, P. G., & Cloke, H. L. (2018). Estimation of uncertainty in flood forecasts—A comparison of methods. Journal of Flood Risk Management, 12(S1). https://doi.org/10.1111/jfr3.12516

Bolan, S., Padhye, L. P., Jasemizad, T., Govarthanan, M., Karmegam, N., Wijesekara, H., Amarasiri, D., Hou, D., Zhou, P., Biswal, B. K., Balasubramanian, R., Wang, H., Siddique, K. H., Rinklebe, J., Kirkham, M., & Bolan, N. (2023). Impacts of climate change on the fate of contaminants through extreme weather events. The Science of the Total Environment, 909, 168388. https://doi.org/10.1016/j.scitotenv.2023.168388

Bril, V., De Bruijn, J. A., De Moel, H., Sadana, T., Busker, T., Botzen, W., Aerts, J.C.J.H. (2025). Assessing the effectiveness of nature-based solutions and building-level flood risk reduction measures [In Peer review]. Water Resources Research

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., & Wada, Y. (2020). Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management. Geoscientific Model Development, 13(7), 3267–3298. https://doi.org/10.5194/gmd-13-3267-2020

Busker, T. S. (2024). Anticipating weather and climate extremes. https://doi.org/10.5463/thesis.837

Busker, T. S., Castro, D. R., Vorogushyn, S., Kwadijk, J., Zoccatelli, D., Loureiro, R., Murdock, H. J., Pfister, L., Dewals, B., Slager, K., Thieken, A. H., Verkade, J., Willems, P., & Aerts, J. C. J. H. (2025, March 14). Comparing Flood Forecasting and Early Warning Systems in Transboundary River Basins. https://doi.org/10.5194/egusphere-2025-828

Bhuyan, K., Van Westen, C., Wang, J., & Meena, S. R. (2022). Mapping and characterising buildings for flood exposure analysis using open-source data and artificial intelligence. Natural Hazards, 119(2), 805–835. https://doi.org/10.1007/s11069-022-05612-4

Centraal Bureau voor de Statistiek. (2025). Kerncijfers per postcode. Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode

Cerri, M., Steinhausen, M., Kreibich, H., & Schröter, K. (2021). Are OpenStreetMap building data useful for flood vulnerability modelling? Natural Hazards And Earth System Sciences, 21(2), 643–662. https://doi.org/10.5194/nhess-21-643-2021

Chowdhury, M. E., Islam, A. S., Lemans, M., Hegnauer, M., Sajib, A. R., Pieu, N. M., Das, M. K., Shadia, N., Haque, A., Roy, B., Billah, M., Abdullah, F., Mamoon, W. B., Kaiser, S., Bala, S. K., Islam, G. T., Sarker, G. C., Rahman, S., & Bhuyan, A. (2022). An efficient flash flood forecasting system for the un-gaged Meghna basin using open-source platform Delft-FEWS. Environmental Modelling & Software, 161, 105614. https://doi.org/10.1016/j.envsoft.2022.105614

Cloke, H., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. Journal Of Hydrology, 375(3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Collier, C. G. (2007). Flash flood forecasting: What are the limits of predictability. Quarterly Journal of the Royal Meteorological Society, 133(622), 3–23. https://doi.org/10.1002/qj.29

Copernicus. (2025, January 10). Global Climate Highlights 2024. https://climate.copernicus.eu/global-climate-highlights-2024

Courant, R., Friedrichs, K., Lewy, H., 1928. Uber die partiellen Differenzengleichungen der mathematischen Physik. Math. Ann. 100, 32–74. doi:10.1007/BF01448839 (In German)

Crochemore, L., Ramos, M., & Pappenberger, F. (2016). Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. Hydrology And Earth System Sciences, 20(9), 3601–3618. https://doi.org/10.5194/hess-20-3601-2016

Das, J., Manikanta, V., Teja, K. N., & Umamahesh, N. V. (2022). Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications and future opportunities. Hydrological Sciences Journal, 67(3), 477–493. https://doi.org/10.1080/02626667.2021.2023157

Da Silva, Q. P., Moreira, D. S., De Freitas, H. C., & Domingues, L. M. (2024). Discrepancies between observation and ERA5 reanalysis in the Amazon deforestation context: A case study. Dynamics Of Atmospheres And Oceans, 106, 101442. https://doi.org/10.1016/j.dynatmoce.2024.101442

Davidson, H., & Ni, V. (2024, 8 November). China floods: death toll climbs as questions raised over preparedness. The Guardian. https://www.theguardian.com/world/2021/jul/22/china-floods-military-blasts-dam-release-water-as-floods-death-toll-climbs (Grey literature)

Dawkins, L. C., Bernie, D. J., Lowe, J. A., & Economou, T. (2023). Assessing climate risk using ensembles: A novel framework for applying and extending open-source climate risk assessment platforms. *Climate Risk Management*, *40*, 100510. https://doi.org/10.1016/j.crm.2023.100510

De Bruijn, J. A., Smilovic, M., Burek, P., Guillaumot, L., Wada, Y., & Aerts, J. C. J. H. (2023). GEB v0.1: a large-scale agent-based socio-hydrological model – simulating 10 million individual farming households in a fully distributed hydrological model. Geoscientific Model Development, 16(9), 2437–2454. https://doi.org/10.5194/gmd-16-2437-2023

De Bruijn, K., Van Den Hurk, B., Slager, K., Rongen, G., Hegnauer, M., & Van Heeringen, K. J. (2023). Storylines of the impacts in the Netherlands of alternative realizations of the Western Europe July 2021 floods. Journal of Coastal and Riverine Flood Risk, 2, p. 8. https://doi.org/10.59490/jcrfr.2023.0008

De Jong, J., & Asselman, N. (2022). Analyse Overstromingen Geulmonding: Watersysteemevaluatie Waterschap Limburg. In https://www.deltares.nl/expertise/publicaties/analyse-overstromingen-geulmonding-watersysteemevaluatie-waterschap-limburg (Nr. 11207700-000-ZWS-0013).

Deltares. (2025a). Hydromt_sfincs/hydromt_sfincs/data/LULc/ESA_worldcover_mapping.csv at main · Deltares/hydromt_sfincs. GitHub. https://github.com/Deltares/hydromt_sfincs/blob/main/hydromt_sfincs/data/lulc/esa_worldcover_mapping.csv

Deltares. (2025b). User manual - general — SFINCS 2.2.0_col_dEze documentation. https://sfincs.readthedocs.io/en/latest/input.html

Deltares. (2025c). Delft-FEWS applications. Delft-FEWS. https://oss.deltares.nl/web/delft-fews/

Deltares. (2025d). Delft-FIAT: Een tool voor overstromingseffecten | Deltares. https://www.deltares.nl/software-en-data/producten/delft-fiat-tool-voor-overstromingseffecten

De Moel, H., Aerts, J. C., & Koomen, E. (2011). Development of flood exposure in the Netherlands during the 20th and 21st century. Global Environmental Change, 21(2), 620-627. https://doi.org/10.1016/j.gloenvcha.2010.12.005

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

De Perez, E. C., Berse, K. B., Depante, L. a. C., Easton-Calabria, E., Evidente, E. P. R., Ezike, T., Heinrich, D., Jack, C., Lagmay, A. M. F. A., Lendelvo, S., Marunye, J., Maxwell, D. G., Murshed, S. B., Orach, C. G., Pinto, M., Poole, L. B., Rathod, K., Shampa, N., & Van Sant, C. (2022). Learning from the past in moving to the future: Invest in communication and response to weather early warnings to reduce death and damage. Climate Risk Management, 38, 100461. https://doi.org/10.1016/j.crm.2022.100461

Dodson, J. C., Dérer, P., Cafaro, P., & Götmark, F. (2020). Population growth and climate change: Addressing the overlooked threat multiplier. The Science of the Total Environment, 748, 141346. https://doi.org/10.1016/j.scitotenv.2020.141346

ECMWF. (2008). The ECMWF Ensemble Prediction System. https://www.ecmwf.int/sites/default/files/elibrary/2012/14557-ecmwf-ensemble-prediction-system.pdf

ECMWF. (2025a, 25 February). Dissemination schedule - Data and Charts - ECMWF Confluence Wiki. https://confluence.ecmwf.int/display/DAC/Dissemination+schedule

ECMWF & Copernicus. (2025, 24 may). ERA5 hourly data on single levels from 1940 to present. https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview

ECMWF. (2025b). Atmospheric Model high resolution 15-day forecast. https://www.ecmwf.int/en/forecasts/datasets/set-i

ECMWF. (2025c). Grid - keyword in MARS/Dissemination request - User Documentation - ECMWF Confluence Wiki. ECMW-Confluence WIKI. https://confluence.ecmwf.int/pages/viewpage.action?pageId=123799065#:~:text=MARS/Dissemination%20*%20grid%20=%20F160%20%2D%20full,latitude%20lines%20between%20the%20pole%20and%20equator.

ECMWF. (2025d). Integrated Forecasting System. https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model

Eilander, D., Couasnon, A., Leijnse, T., Ikeuchi, H., Yamazaki, D., Muis, S., Dullaart, J., Haag, A., Winsemius, H. C., & Ward, P. J. (2023). A globally applicable framework for compound flood hazard modeling. Natural Hazards And Earth System Sciences, 23(2), 823–846. https://doi.org/10.5194/nhess-23-823-2023

EM-DAT, CRED / UCLouvain (2024) – with major processing by Our World in Data. "Number of recorded natural disaster events – EM-DAT" [dataset]. EM-DAT, CRED / UCLouvain, "Natural disasters" [original data]. Retrieved January 13, 2025 from https://ourworldindata.org/grapher/number-of-natural-disaster-events

Endendijk, T., Botzen, W., De Moel, H., Aerts, J., Duijndam, S., Slager, K., Kolen, B., & Kok, M. (2023). Experience from the 2021 floods in the Netherlands: Household Survey Results on Impacts and Responses. Journal of Coastal and Riverine Flood Risk, 2, p. 9. https://doi.org/10.59490/jcrfr.2023.0009

ENW (2021) Hoogwater 2021: Feiten en Duiding. Expertise Netwerk Hoogwater (ENW), https://www.enwinfo.nl/publicaties/ (in Dutch)

Esm, S. O., Oam, P. C., & Davies, B. (2010). Timeline modelling of flood evacuation operations. Procedia Engineering, 3, 175–187. https://doi.org/10.1016/j.proeng.2010.07.017

Fallowfield, L., & Motta, D. (2024). The permanent flood risk of culverts and the impact of increasing debris blockage. Journal Of Flood Risk Management, 17(4). https://doi.org/10.1111/jfr3.13021

Fernández-Nóvoa, D., González-Cao, J., & García-Feal, O. (2024). Enhancing Flood Risk Management: A Comprehensive Review on Flood Early Warning Systems with Emphasis on Numerical Modeling. Water, 16(10), 1408. https://doi.org/10.3390/w16101408

Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., & Francés, F. (2023). Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). Atmospheric Research, 284, 106606. https://doi.org/10.1016/j.atmosres.2023.106606

Haces-Garcia, F., Ross, N., Glennie, C. L., Rifai, H. S., Hoskere, V., & Ekhtari, N. (2024). Rapid 2D hydrodynamic flood modelling using deep learning surrogates. Journal of Hydrology, 132561. https://doi.org/10.1016/j.jhydrol.2024.132561

Haiden, T., & Lledó, L. (2023, 24 January). Verifying high-resolution forecasts. ECMWF. https://www.ecmwf.int/en/about/media-centre/science-blog/2023/verifying-high-resolution-forecasts

Hailemariam, F. M., Brandimarte, L., & Dottori, F. (2013). Investigating the influence of minor hydraulic structures on modeling flood events in lowland areas. Hydrological Processes, 28(4), 1742–1755. https://doi.org/10.1002/hyp.9717

Hamdi, M., El Molla, D. A., & Gad, M. A. (2019). A comparison between 1D, 2D and semi 2D hydraulic models. *Al-Azhar University Civil Engineering Research Magazine (CERM)*, *41*(4), 295-305.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559–570

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal Of The Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Het Waterschapshuis. (2021). Hoogwater Limburg (juli 2021) op de luchtfoto. https://twiav.nl/nl/water/hoogwater2021limburg.htm?x=187754&y=318484&l=12&b=lichtgrijs

Huang, S., Gan, Y., Zhang, X., Chen, N., Wang, C., Gu, X., Ma, J., & Niyogi, D. (2023). Urbanization amplified asymmetrical changes of rainfall and exacerbated drought: Analysis over five urban agglomerations in the Yangtze River Basin, China. Earth S Future, 11(2). https://doi.org/10.1029/2022ef003117

Huizinga, J., De Moel, H., & Szewczyk, W. (2017). Global flood depth-damage functions: Methodology and the database with guidelines. RePEc: Research Papers in Economics. https://doi.org/10.2760/16510

IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp

Islam, M. M., Hasan, M., Mia, M. S., Masud, A. A., & Islam, A. R. M. T. (2025). Early Warning Systems in Climate Risk Management: Roles and Implementations in Eradicating Barriers and Overcoming Challenges. Natural Hazards Research. https://doi.org/10.1016/j.nhres.2025.01.007

Jain, S. K., Mani, P., Jain, S. K., Prakash, P., Singh, V. P., Tullos, D., Kumar, S., Agarwal, S. P., & Dimri, A. P. (2017). A Brief review of flood forecasting techniques and their applications. International Journal of River Basin Management, 16(3), 329–344. https://doi.org/10.1080/15715124.2017.1411920

JCAR ATRACE. (2025-a). Geul | JCAR ATRACE. https://www.jcar-atrace.eu/en/basins/geul#:~:text=The%20total%20length%20of%20the,Dutch%20part%20of%20the%20basin.

JCAR ATRACE. (2025-b). About us | JCAR ATRACE. https://www.jcar-atrace.eu/en/about-us

Kellermann, P., Schönberger, C., & Thieken, A. H. (2016). Large-scale application of the flood damage model RAilway Infrastructure Loss (RAIL). *Natural Hazards and Earth System Sciences*, *16*(11), 2357–2371. https://doi.org/10.5194/nhess-16-2357-2016

Khalaj, M. R., Noor, H., & Dastranj, A. (2021). Investigation and simulation of flood inundation hazard in urban areas in Iran. Geoenvironmental Disasters, 8(1). https://doi.org/10.1186/s40677-021-00191-1

Khalid, M. S., Mustaffa, C. S., Marzuki, N., Sakdan, M. F., Sipon, S., Ariffin, M. T., & Shafiai, S. (2015). Failure to react positively to flood early Warning Systems: Lessons Learned by flood Victims from Flash Flood Disasters: The Malaysia Experience. Engineering And Technology International Journal Of Social, Behavioural, Educational, Economic And Management Engineering, 9(5).

KNMI. (2003, December 12). KNMI - KNMI en waterschappen samen tegen wateroverlast. https://www.knmi.nl/over-het-knmi/nieuws/knmi-en-waterschappen-samen-tegen-wateroverlast

Kok, M., Jongejan, R., Nieuwjaar, M., and Tánczos, I. (2017). Fundamentals of flood protection. ENW Available at:https://www.enwinfo.nl/publish/pages/183541/grondslagenen-lowresspread3-v_3.pdf

Koks, E. E., Rozenberg, J., Zorn, C., Tariverdi, M., Vousdoukas, M., Fraser, S. A., Hall, J. W., & Hallegatte, S. (2019). A global multi-hazard risk analysis of road and railway infrastructure assets. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-10442-3

Koks. E.E. (2022). DamageScanner: Python tool for natural hazard damage assessments. Zenodo. http://doi.org/10.5281/zenodo.2551015

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H., Alvarez-Garreton, C., Aznar, B., Balkhi, L., Barendrecht, M. H., Biancamaria, S., Bos-Burgering, L., Bradley, C., Budiyono, Y., Buytaert, W., Di Baldassarre, G. (2022). The challenge of unprecedented floods and droughts in risk management. Nature, 608(7921), 80–86. https://doi.org/10.1038/s41586-022-04917-5

Kumar, U., & Sharma, N. (2024). A Scientific Analysis of Weather Forecasting Techniques and its Classification. Deleted Journal, 2024(4), 917–923. https://doi.org/10.53935/jomw.v2024i4.567

Kumbier, K., Carvalho, R. C., Vafeidis, A. T., & Woodroffe, C. D. (2018). Investigating compound flooding in an estuary using hydrodynamic modelling: a case study from the Shoalhaven River, Australia. Natural Hazards and Earth System Sciences, 18(2), 463–477. https://doi.org/10.5194/nhess-18-463-2018

Landwehr, T., Dasgupta, A., & Waske, B. (2024). Towards robust validation strategies for EO flood maps. Remote Sensing Of Environment, 315, 114439. https://doi.org/10.1016/j.rse.2024.114439

Latona, D., Pons, C., Lombardi, P., & Laing, A. (2024, 9 November). How slow warnings, poor defences left Spain exposed to deadly floods. Reuters. https://www.reuters.com/business/environment/spains-flood-disaster-was-its-worst-recent-history-heres-what-went-wrong-2024-11-09 (Grey literature)

Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., Völker, C., & Hollert, H. (2022). Assessment of the 2021 summer flood in Central Europe. Environmental Sciences Europe, 34(1), 107. https://doi.org/10.1186/s12302-022-00685-1

Leijnse, T., Van Ormondt, M., Nederhoff, K., & Van Dongeren, A. (2021). Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind- and wave-driven processes. Coastal Engineering, 163, 103796. https://doi.org/10.1016/j.coastaleng.2020.103796

Lindenlaub, S., Matthes, K. S., & Thieken, A. H. (2024). Improving the understandability of flood warning messages – explorative study on design preferences. International Journal of Disaster Risk Reduction, 111, 104685. https://doi.org/10.1016/j.ijdrr.2024.104685

Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kühnlein, C., Mozdzynski, G., Hamrud, M., & Smolarkiewicz, P. (2015). A new grid for the IFS. ECMWF Newsletter, Number 146–Winter 2015/16, 23–28. https://doi.org/10.21957/zwdu9u5i

Manfreda, S., & Samela, C. (2019). A digital elevation model based method for a rapid estimation of flood inundation depth. Journal Of Flood Risk Management, 12(S1). https://doi.org/10.1111/jfr3.12541

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article &quote; Assessment of economic flood damage" Natural Hazards And Earth System Sciences, 10(8), 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., Domeisen, D. I. V., Feser, F., Koszalka, I., Kreibich, H., Pantillon, F., Parolai, S., Pinto, J. G., Punge, H. J., Rivalta, E., Schröter, K., Strehlow, K., Weisse, R., & Wurpts, A. (2020). Impact Forecasting to Support Emergency Management of Natural Hazards. Reviews Of Geophysics, 58(4). https://doi.org/10.1029/2020rg000704

Ministerie van Infrastructuur en Waterstaat. (2024a, February 19). Waterbericht. https://www.rijkswaterstaat.nl/water/waterdata-en-waterberichtgeving/waterbericht

Ministerie van Infrastructuur en Waterstaat. (2024b, October 10). Waterbeheer in Nederland. Water | Rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/water/waterbeheer-in-nederland#:~:text=Taken%20waterbeheerders&text=RWS%20waarschuwt%20de%20verantwoordelijke%20overheden,en%20nevengeulen%20aan%20te%20leggen.

Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293.

Najafi, H., Shrestha, P. K., Rakovec, O., Apel, H., Vorogushyn, S., Kumar, R., Thober, S., Merz, B., & Samaniego, L. (2024). High-resolution impact-based early warning system for riverine flooding. Nature Communications, 15(1). https://doi.org/10.1038/s41467-024-48065-y

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal Of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., & Serafin, S. (2024). The fractions skill score for ensemble forecast verification. Quarterly Journal Of The Royal Meteorological Society, 150(764), 4457–4477. https://doi.org/10.1002/qj.4824

Nederhoff, K., Crosby, S. C., Van Arendonk, N. R., Grossman, E. E., Tehranirad, B., Leijnse, T., Klessens, W., & Barnard, P. L. (2024). Dynamic Modeling of Coastal Compound Flooding Hazards Due to Tides, Extratropical Storms, Waves, and Sea-Level Rise: A Case Study in the Salish Sea, Washington (USA). Water, 16(2), 346. https://doi.org/10.3390/w16020346

Nicholas, A.P., and C.A. Mitchell (2003). Numerical simulation of overbank processes in topographically complex floodplain environments, Hydrological Processes, 17(4), 727-746.

Nirandjan, S., Koks, E. E., Ye, M., Pant, R., van Ginkel, K. C. H., Aerts, J. C. J. H., and Ward, P. J.: Review article: Physical Vulnerability Database for Critical Infrastructure Multi-Hazard Risk Assessments – A systematic review and data collection, Nat. Hazards Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/nhess-2023-208, in review, 2024

Norrman, K. (2023). World population growth: a once and future global concern. World, 4(4), 684–697. https://doi.org/10.3390/world4040043

Ogden, F. L. (2020). Geohydrology: Hydrological Modeling. In Elsevier eBooks (pp. 457–476). https://doi.org/10.1016/b978-0-08-102908-4.00115-6

Overhoff, J. (2024). Evaluating nature-based solutions for flooding in the Geul catchment using openLISEM. https://www.jcar-atrace.eu/en/basins/geul

Paiva, R. C., Collischonn, W., & Tucci, C. E. (2011). Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. Journal of Hydrology, 406(3–4), 170–181. https://doi.org/10.1016/j.jhydrol.2011.06.007

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. Environmental Science & Policy, 51, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016

Pappenberger, F., Pagano, T. C., Brown, J. D., Alfieri, L., Lavers, D. A., Berthet, L., Bressand, F., Cloke, H. L., Cranston, M., Danhelka, J., Demargne, J., Demuth, N., De Saint-Aubin, C., Feikema, P. M., Fresch, M. A., Garçon, R., Gelfan, A., He, Y., Hu, Y.-., . . . Pozo, J. T. (2019). Hydrological ensemble prediction systems around the globe. In Springer eBooks (pp. 1187–1221). https://doi.org/10.1007/978-3-642-39925-1_47

Parker, D. J., & Priest, S. J. (2012). The fallibility of flood warning chains: Can Europe's flood warnings be effective? Water Resources Management, 26(10), 2927–2950. https://doi.org/10.1007/s11269-012-0057-6

Pechlivanidis, I., Jackson, B., McIntyre, N., & Wheater, H. (2013). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Global NEST Journal, 13(3), 193–214. https://doi.org/10.30955/gnj.000778

Perera, D., Seidou, O., Agnihotri, J., Wahid, A., & Rasmy, M. (2019). Flood Early Warning Systems: A Review Of Benefits, Challenges And Prospects. UNU-INWEH Report Series, (8). https: //doi.org/10.13140/RG.2.2.28339.78880

Piadeh, F., Behzadian, K., & Alani, A. M. (2022). A critical review of real-time modelling of flood forecasting in urban drainage systems. Journal of Hydrology, 607, 127476. https://doi.org/10.1016/j.jhydrol.2022.127476

Pita, G. L., Albornoz, B. S., & Zaracho, J. I. (2021). Flood depth-damage and fragility functions derived with structured expert judgment. Journal Of Hydrology, 603, 126982. https://doi.org/10.1016/j.jhydrol.2021.126982

Potter, S., Kox, T., Mills, B., Taylor, A., Robbins, J., Cerrudo, C., Wyatt, F., Harrison, S., Golding, B., Lang, W., Harris, A. J., Kaltenberger, R., Kienberger, S., Brooks, H., & Tupper, A. (2025). Research gaps and challenges for impact-based forecasts and warnings: Results of international workshops for High Impact Weather in 2022. International Journal Of Disaster Risk Reduction, 105234. https://doi.org/10.1016/j.ijdrr.2025.105234

Red Cross Red Crescent Movement & Climate Centre (RCCC). (2020). The future of forecasts: Impact-based forecasting for early action. Anticipation Hub. https://www.anticipation-hub.org/download/file-58

Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-30727-4

Ringo, J., Sabai, S., & Mahenge, A. (2023). Performance of early warning systems in mitigating flood effects. A review. Journal of African Earth Sciences, 210, 105134. https://doi.org/10.1016/j.jafrearsci.2023.105134

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. Water resources research, 51(9), 7358-7381. https://doi.org/10.1002/2015WR016954

Sawada, Y., Kanai, R., & Kotani, H. (2022). Impact of cry wolf effects on social preparedness and the efficiency of flood early warning systems. Hydrology and Earth System Sciences, 26(16), 4265–4278. https://doi.org/10.5194/hess-26-4265-2022

Sebastian, A., Bader, D. J., Nederhoff, C. M., Leijnse, T. W. B., Bricker, J. D., & Aarninkhof, S. G. J. (2021). Hindcast of pluvial, fluvial, and coastal flood damage in Houston, Texas during Hurricane Harvey (2017) using SFINCS. Natural Hazards, 109(3), 2343–2362. https://doi.org/10.1007/s11069-021-04922-3

Sieg, T., & Thieken, A. H. (2022). Improving flood impact estimations. Environmental Research Letters, 17(6), 064007. https://doi.org/10.1088/1748-9326/ac6d6c

Sieg, T., Kienzler, S., Rözer, V., Vogel, K., Rust, H. W., Bronstert, A., Kreibich, H., Merz, B., & Thieken, A. H. (2023). Toward an adequate level of detail in flood risk assessments. Journal of Flood Risk Management, 16(3). https://doi.org/10.1111/jfr3.12889

Slager, K., Becker, A., Bouaziz, L., & Kwadijk, J. (2022). Rapid Assessment study on the Geul River Basin: Screening of flood reduction measures. Deltares. https://www.deltares.nl/expertise/publicaties/rapid-assessment-study-on-the-geul-river-basin-screening-of-flood-reduction-measures

Slager, K., de Moel, H., & de Jong, J. (2021). *Maximum flood extents Limburg floods July 2021* (Version 1) [Data set]. 4TU.ResearchData. https://doi.org/10.4121/16817389.V1

Stephens, E., Schumann, G., & Bates, P. (2013). Problems with binary pattern measures for flood model evaluation. Hydrological Processes, 28(18), 4928–4937. https://doi.org/10.1002/hyp.9979

Sušnik, J., Strehl, C., Postmes, L. A., Vamvakeridou-Lyroudia, L. S., Mälzer, H., Savić, D. A., & Kapelan, Z. (2014). Assessing Financial Loss due to Pluvial Flooding and the Efficacy of Risk-Reduction Measures in the Residential Property Sector. Water Resources Management, 29(1), 161–179. https://doi.org/10.1007/s11269-014-0833-6

Teja, K. N., Manikanta, V., Das, J., & Umamahesh, N. (2023). Enhancing the predictability of flood forecasts by combining Numerical Weather Prediction ensembles with multiple hydrological models. Journal Of Hydrology, 625, 130176. https://doi.org/10.1016/j.jhydrol.2023.130176

Thieken, A. H., Bubeck, P., Heidenreich, A., Von Keyserlingk, J., Dillenardt, L., & Otto, A. (2023). Performance of the flood warning system in Germany in July 2021 – insights from affected residents. Natural Hazards And Earth System Sciences, 23(2), 973–990. https://doi.org/10.5194/nhess-23-973-2023

Tradowsky, J. S., Philip, S. Y., Kreienkamp, F., Kew, S. F., Lorenz, P., Arrighi, J., Bettmann, T., Caluwaerts, S., Chan, S. C., De Cruz, L., De Vries, H., Demuth, N., Ferrone, A., Fischer, E. M., Fowler, H. J., Goergen, K., Heinrich, D., Henrichs, Y., Kaspar, F., . . . Wanders, N. (2023). Attribution of the heavy rainfall events leading to severe flooding in Western Europe during July 2021. Climatic Change, 176(7). https://doi.org/10.1007/s10584-023-03502-7

Trogrlić, R. Š., Van Den Homberg, M., Budimir, M., McQuistan, C., Sneddon, A., & Golding, B. (2022). Early warning systems and their role in disaster risk reduction. In Springer eBooks (pp. 11–46). https://doi.org/10.1007/978-3-030-98989-7_2

Tsiokanos, A., Rutten, M., Van Der Ent, R. J., & Uijlenhoet, R. (2024). Flood drivers and trends: a case study of the Geul River catchment (the Netherlands) over the past half century. Hydrology and Earth System Sciences, 28(14), 3327–3345. https://doi.org/10.5194/hess-28-3327-2024

UNDRR. (2008, September 18). Global survey of early warning systems. https://www.undrr.org/publication/global-survey-early-warning-systems

UNDRR (2023). *GAR Special Report: Measuring Resilience for the Sustainable Development Goals*. Geneva, Switzerland: United Nations Office for Disaster Risk Reduction. https://www.undrr.org/gar/gar2023-special-report

Van Den Braak, W. E. W., De Bake, D., Ministry of Transport, Public Works and Water Management, & HKV in lijn. (2020). [Damage assessment in The Netherlands]. In Damage Assessment in The Netherlands. https://www.hkv.nl/wp-content/uploads/2020/08/DamageassessmentintheNetherlands_EvdB.pdf

Van Dongeren, A.R., Svendsen, I.A., 1997. Absorbing-generating boundary conditions for shallow water models. Waterw. Port, Coastal, Ocean Eng. 123, 303–313

Van Ginkel, K., Dottori, F., Alfieri, L., Feyen, L., & Koks, E. (2021). Flood risk assessment of the European road network. Natural Hazards and Earth System Sciences, 21(3), 1011–1027. https://doi.org/10.5194/nhess-21-1011-2021

Van Heeringen, K. J., Asselman, N. E. M., Overeem, A., Beersma, J. J., & Philip, S. (2022). Analyse overstroming Valkenburg : watersysteemevaluatie Waterschap Limburg | Deltares. https://www.deltares.nl/expertise/publicaties/analyse-overstroming-valkenburg-watersysteemevaluatie-waterschap-limburg

Van Ormondt, M., Leijnse, T., De Goede, R., Nederhoff, K., & Van Dongeren, A. (2025). Subgrid corrections for the linear inertial equations of a compound flood model – a case study using SFINCS 2.1.1 Dollerup release. Geoscientific Model Development, 18(3), 843–861. https://doi.org/10.5194/gmd-18-843-2025

Veiligheidsregio Zuid-Limburg, Veiligheidsregio Noord-Limburg, Rijkswaterstaat, & Waterschap Limburg. (2024, May 4). Rampbestrijdingsplan Hoogwater Limburg 2023-2026. Lokale Wet- En Regelgeving. https://lokaleregelgeving.overheid.nl/CVDR719417/1

Verkade, J. S., & Werner, M. G. F. (2011). Estimating the benefits of single value and probability forecasting for flood warning. Hydrology and Earth System Sciences, 15(12), 3751–3765. https://doi.org/10.5194/hess-15-3751-2011

WCRP & WWRP. (2017). Theme: Forecast Verification Methods Across Time and Space Scales. WWRP/WGNE Joint Working Group On Forecast Verification Research. https://www.cawcr.gov.au/projects/verification/#Methods_for_foreasts_of_continuous_variables

Wing, O. E., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous United States. Water Resources Research, 53(9), 7968-7986. https://doi.org/10.1002/2017WR020917

WMO. (2015). WMO Guidelines on Multi-hazard Impact-based forecast and warning Services (ISBN 978-92-63-11150-0). Geneva, Switzerland: WMO. https://etrp.wmo.int/pluginfile.php/16270/mod_resource/content/0/wmo_1150_en.pdf

WMO. (2021). WMO Guidelines on Multi-hazard Impact-based forecast and warning Services – Part II: Putting Multi-hazard IBFWS into practice (ISBN 978-92-63-11150-0). Geneva, Switzerland: WMO. https://www.anticipationhub.org/Documents/Manuals_and_Guidelines/WMO_Guidelines_on_Multi-hazard__part2.pdf

WMO. (2024, December 17). Floods. World Meteorological Organization. https://wmo.int/topics/floods#:~:text=Floods%20are%20the%20deadliest%20natural,increase%20in%20flood%20risks%20worldwide.&text=Between%201994%20and%202013%2C%20floods%20affected%20nearly%202.5%20billion%20worldwide.

Woo, H., Choi, H., Kim, M., & Noh, S. J. (2025). Physics-Guided Deep Learning for Spatiotemporal Evolution of Urban Pluvial Flooding. Water, 17(8), 1239. https://doi.org/10.3390/w17081239

Wu, G., Qin, S., Mao, Y., Ma, Z., & Shi, C. (2022). Validation of Precipitation Events in ERA5 to Gauge Observations during Warm Seasons over Eastern China. Journal Of Hydrometeorology, 23(5), 807–822. https://doi.org/10.1175/jhm-d-21-0195.1

Yang, Q., Wang, Q. J., & Hakala, K. (2021). Achieving effective calibration of precipitation forecasts over a continental scale. Journal Of Hydrology Regional Studies, 35, 100818. https://doi.org/10.1016/j.ejrh.2021.100818

Ye, A., Zhou, Z., You, J., Ma, F., & Duan, Q. (2018). Dynamic Manning's roughness coefficients for hydrological modelling in basins. Hydrology Research, 49(5), 1379–1395. https://doi.org/10.2166/nh.2018.175

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O., (2022). ESA WorldCover 10 m 2021 v200 (Version v200) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7254221

# Appendices

## Appendix A: API request to download ECMWF forecast data

*Table A: Outlined parameters for the API request to download rainfall data with the consistent parameter settings.*

| *API PARAMETERS* | *CONTROL FORECASTS* | *PERTURBED FORECASTS* |
|---|---|---|
| *CLASS* | 'od' | 'od' |
| *DATE* | 2021-07-11 till 2021-07-14 | 2021-07-11 till 2021-07-14 |
| *EXPVER* | 1 | 1 |
| *LEVTYPE* | 'sfc' | 'sfc' |
| *PARAM* | 228.128 | 228.128 |
| *STEP* | 0/to/90/BY/1 & 90/to/120/BY/3 | 0/to/90/BY/1 & 90/to/120/BY/3 |
| *STREAM* | 'enfo' | 'enfo' |
| *NUMBER* | - | '1/to/50' |
| *TIME* | ['00','12'] | ['00','12'] |
| *TYPE* | 'cf' | 'pf' |
| *GRID* | 'F640' | 'F640' |
| *AREA* | '52.5/3.5/49/8' | '52.5/3.5/49/8' |

## Appendix B: Numerical equations inside the SFINCS model

Section 4.2.2 focused on the structure of the model. Another key aspect is the methodology of how the model solves the hydrodynamic processes. First, the model calculates the volumetric flow rate between adjacent grid cells for the next time step for each dimension (x or y) (Equation 1).

$$q_x^{t+\Delta t} = \frac{q_x^t - \left(g h_x^t \frac{\Delta z}{\Delta x} + adv_x - \frac{\tau_{w,x}}{\rho_w}\right)\Delta t}{\left(1 + \frac{\left(g \Delta t n^2 q_x^t\right)^{\frac{7}{3}}}{h_x^t}\right)} \quad (1)$$

Equation 1 describes the volumetric flow rate for the x dimension in the next time step, where $q_x^t$ is the volumetric flow rate in the previous timestep t, $h_x^t$ defines the average water depth of the two adjacent cells in the previous timestep, $\Delta z$ is the difference in waterlevel between the adjacent cells, and the gravitational constant and the manning friction are given by g an n.

Moreover, two additional terms are incorporated within the momentum equation of Saint-Vernant in Equation 1. The first term is the wind drag term $\frac{\tau_{w,x}}{\rho_w}$, this term simulates locally generated wind set-up based on the wind shear stress $\tau_{w,x}$ and the water density $\rho_w$. In addition, the wind shear stress is computed as a function of a wind drag coefficient $C_d$, the air density $\rho_a$, the wind speed in the x-dimension $u_{w,x}$ and in the y-dimension $u_{w,y}$ (Equation 2). In the model the wind stress is reduced to zero in water depths less than 0.25 meter, to improve the numerical stability of the model.

$$\tau_{w,x} = C_d \rho_a u_{w,x} \sqrt{u_{w,x}^2 + u_{w,y}^2} \quad (2)$$

The second term that is included in the momentum equation is the advection term $adv_x$, which is divided into two parts (Equation 3). These two parts are related to the volumetric flow rates $q_x$ and $q_x q_y$ over the x- and y-dimensions and are also dependent on the grid size.

$$adv_x = adv_{x,1} + adv_{x,2} \quad (3)$$

The first part of the advection term $adv_{x,1}$ calculates the difference between the quotients of the squared flow rate of the current grid cell (q) and the water depth (h) for the adjacent grid cells in the x-dimension. The calculated difference is divided by the grid size in the x dimension. The equation is determined based on the state of flow rate and water depth in the current and previous time step (Equation 4a and 4b)

$$for\ q_{x,m,n}^t > 0:\ adv_{x,1} = \frac{\left(\frac{\left(q_{x,m,n}^t\right)^2}{h_{x,m,n}^t} - \frac{\left(q_{x,m-1,n}^t\right)^2}{h_{x,m-1,n}^t}\right)}{\Delta x} \quad (4a)$$

$$for\ q_{x,m,n}^{t-\Delta t} < 0:\ adv_{x,1} = \frac{\left(\frac{\left(q_{x,m+1,n}^t\right)^2}{h_{x,m+1,n}^t} - \frac{\left(q_{x,m,n}^t\right)^2}{h_{x,m,n}^t}\right)}{\Delta x} \quad (4b)$$

The second sub equation of the advection term $adv_{x,2}$ is similar to the Equations 4a and 4b but focuses on the difference over the y-dimension. In this equation, also the cells diagonal to the current cell are taken into account (m+1 & n+1). This part is not conditioned based on the current and previous timesteps (Equation 5).

$$adv_{x,2} = \frac{\left(\frac{\frac{1}{2}\left[q_{x,m,n}^t + q_{x,m,n+1}^t\right]\frac{1}{2}\left[q_{y,m+1,n}^t + q_{y,m,n}^t\right]}{\frac{1}{2}\left[h_{x,m,n}^t + h_{x,m,n+1}^t\right]} - \frac{\frac{1}{2}\left[q_{x,m,n}^t + q_{x,m,n-1}^t\right]\frac{1}{2}\left[q_{y,m+1,n-1}^t + q_{y,m,n-1}^t\right]}{\frac{1}{2}\left[h_{x,m,n}^t + h_{x,m,n-1}^t\right]}\right)}{\Delta y} \quad (5)$$

When the components are calculated Equation 3 can be solved and implemented in Equation 1 together with the determined wind drag term. Solving Equation 1 per timestep results in the determination of the fluxes through the 2D spatial grid.

After the water depth and discharges in x- and y-directions are computed, the water levels ζ inside the cells are updated based on the current water level and the change in flow rate (Equation 6).

$$\zeta_{m,n}^{t+\Delta t} = \zeta_{m,n}^t + \left(\frac{\left(q_{x,m-1,n}^{t+\Delta t} - q_{x,m,n}^{t+\Delta t}\right)}{\Delta x} + \frac{\left(q_{y,m,n-1}^{t+\Delta t} - q_{y,m,n}^{t+\Delta t}\right)}{\Delta y} + S_{m,n}\right) \Delta t \quad (6)$$

In Equation 6 a source term $S_{m,n}$ is incorporated to represent an additional flux such as an infiltration rate, precipitation rate, or a discharge from a user-defined point source. Additionally, Equation 6 is the 2D version of the continuity equation. Moreover, A flow depth limiter is incorporated in the model to distinguish between dry and wet grid cells. The limit can be altered dependent on the type of flood. As a result, after the limit of a grid cell is reached, the cell is marked as wet.

Furthermore, the timestep $\Delta t$ is determined dynamically based on the Courant-Fiedrich-Lewy (CFL) condition. This condition ensures that the computation of the equations does not move further than one spatial grid point per time step. This is important because the equations attempt to calculate values in next time steps that has not yet arrived at the current timestep, which can lead to numerical instability (Courant et al., 1928). By including the CFL-condition, numerical instability in the model is minimized (Equation 7).

$$\Delta t = \alpha \frac{\Delta x}{\sqrt{g h_{max}}} \tag{7}$$

The timestep in the model cannot be greater than the calculated timestep of Equation 7. When this is the case, it results in numerical instability what will cause inaccurate results.

Ultimately, the seven equations are applied for each time step and in each grid cell. As a result, the hydrodynamical processes are simulated.

***Boundary conditions:***

Next to the simulation of hydrodynamic processes, SFINCS defines open boundary conditions, which have a weakly reflective and absorbing character. This character allows outgoing fluxes to leave the computational grid without influencing incoming fluxes at the same boundary (Van Dongeren and Svendsen, 1997). Boundary fluxes are calculated at the velocity points of grid cells, where h is the total water depth at the boundary, $\bar{u}$ the mean velocity, and $\zeta_i$ is the water level in the first regular grid cell inside the spatial grid (Equation 8).

$$q = h\left(2\sqrt{\frac{g}{h}}\,(\zeta - \zeta_0) - \sqrt{\frac{g}{h}}(\zeta_i - \zeta_0) + \bar{u}\right) \tag{8}$$

# Appendix C: Additional plots for the issued rainfall forecast evaluation

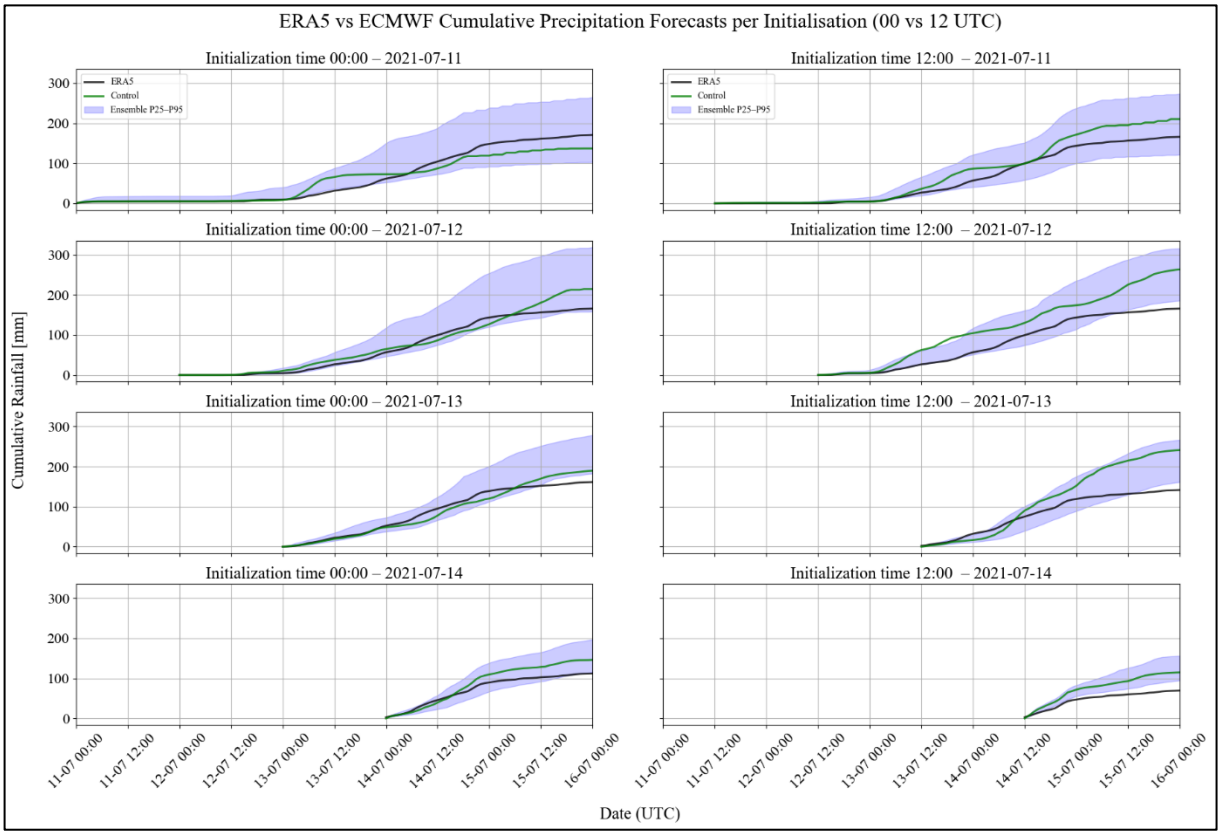


*Figure C.1: A comparison of the maximum cumulative rainfall over the entire Geul catchment between the processed control (green line) and ensemble percentile forecasts (blue spread) of the ECMWF and the ERA5 observations (black line) for various initialization dates between July 11 and 14 2021.*

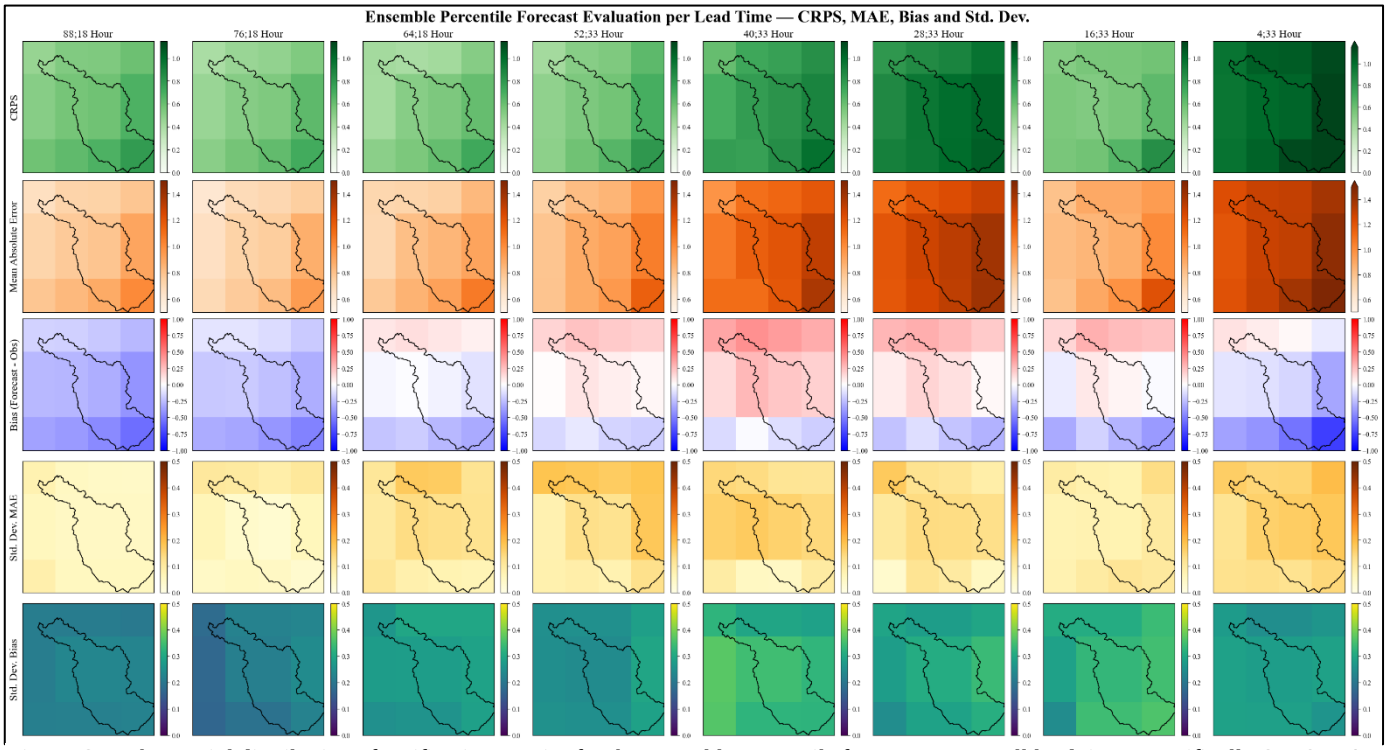


*Figure C.2: The spatial distribution of verification metrics for the ensemble percentile forecasts across all lead times, specifically CRPS, MAE, Bias, Standard deviations in mm per hour.*

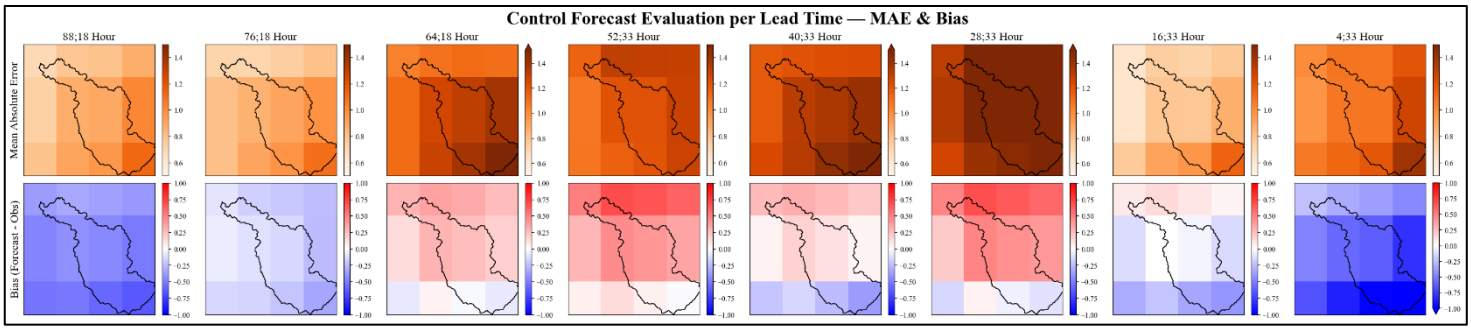


*Figure C.3: The spatial distribution of verification metrics for the control forecasts across all lead times, specifically MAE, Bias in mm per hour.*

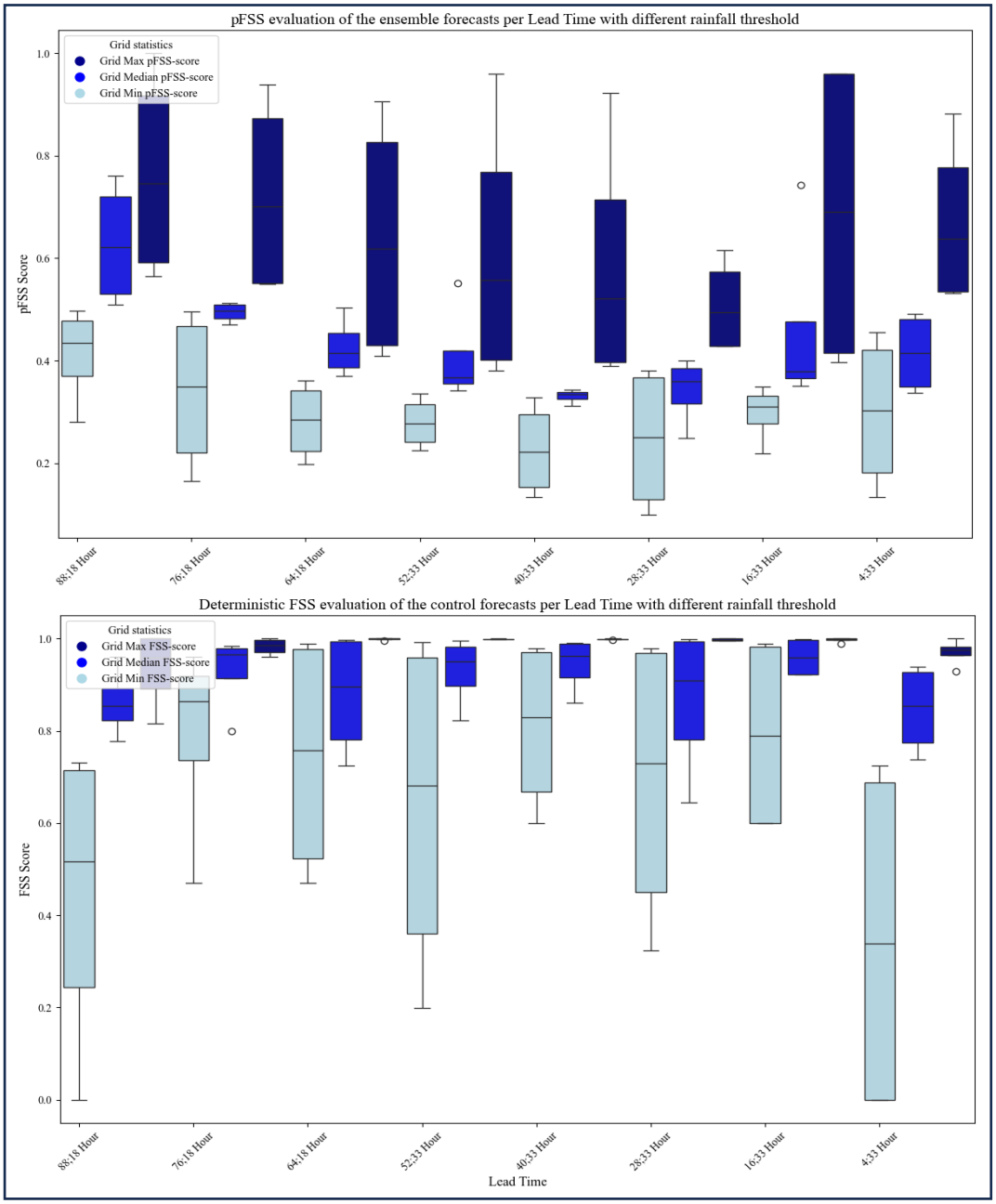


*Figure C.4: Boxplots of the Fraction Skill Score (FSS) for ensemble percentile forecasts and control forecasts across different lead times. The plots display the range of minimum (light blue), median (blue), and maximum (dark blue) FSS values per lead time dependent on the rainfall intensity thresholds of 0.5, 1, 3, and 4 mm/h.*

# Appendix D: Cross sectional analysis for improving the SFINCS output

In this phase, two key adjustments were made to improve the outcomes of SFINCS. First, the river network was refined to better represent the flow pathways. Second, a projection mismatch was identified and corrected, which previously led to unrealistic flood maps due to spatial misalignment of input data and local impoundments because of present waterworks.

In the initial model configuration, the river network only consisted of the primary branches of the Geul river. To improve the channel network, smaller tributaries were delineated and integrated using QGIS to ensure a continuous flow path and a more accurate river bathymetry (derived from Bril et al. (2025)). This refinement included smoothing of right-angle turns. Figure D.1 shows the transition of the river network in the model.

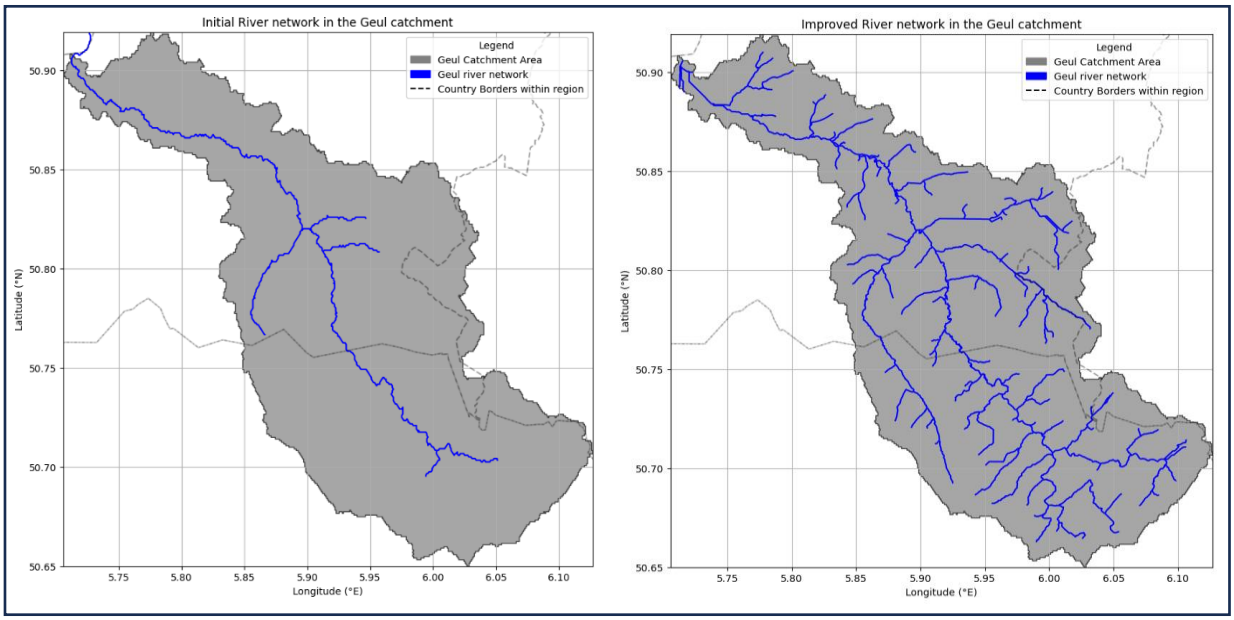


*Figure D.1: The transition of the river network within the SFINCS model by elaborating the tributaries of the Geul River.*

Overall, this refinement improved the accuracy of the simulated flood extent and depth, reducing extreme and unrealistic water levels from approximately +20 meters to water depths of twelve meters. Smoothing the tributaries was crucial to prevent artificial impoundments between grid cells. However, following the implementation of the updated river network, a remarkable flooding pattern was observed, as illustrated in Figure D.2.

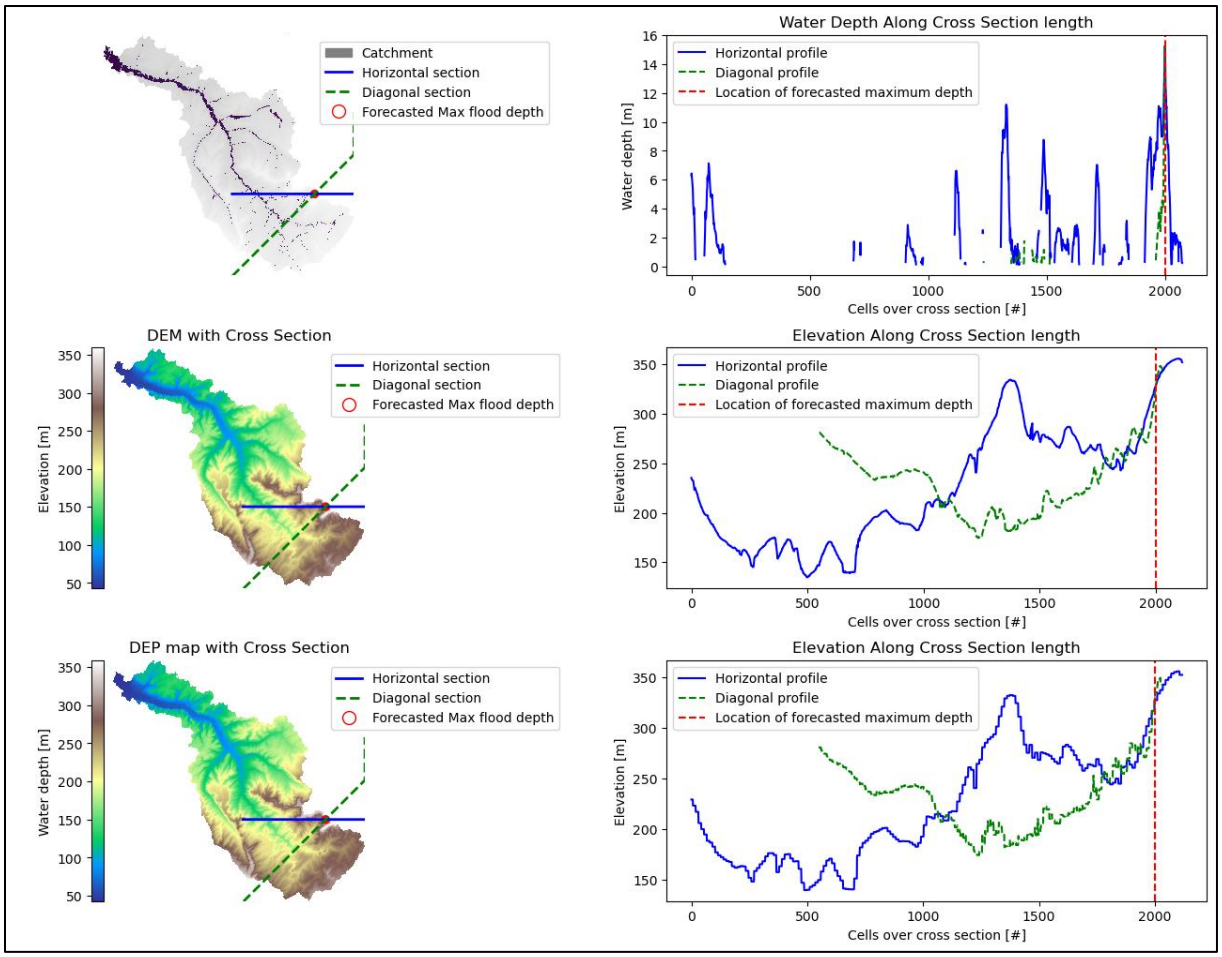


*Figure D.2: A cross-sectional analysis of a strange line of flooding around the Dutch border with water depths of over 18 meters. The blue and green lines represent two cross sections in two directions over the grid cells. In the figure the profile of the water depth, DEM and DEP is visualized where the location of the cross sections is specified at the left from the line plots. The red line represents the location of the maximum forecasted flood depth in the catchment.*

This flooding pattern occurred at the highest elevation points in the cross section, making it remarkable. Ultimately, the pattern was caused by a mismatch in the projection of the model as the line was located at the boundaries of the coordinate system Amersfoort RD new. By specifying the coordinate system the misprojection was solved and the strange line disappeared.

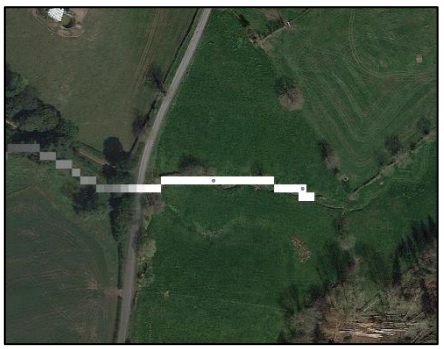

*Figure D.3: An example of a local impoundment caused by (in this case) infrastructure, with white representing the water depths of approximate 17 meters.*

Subsequently, a cross sectional analysis is conducted in order to ensure that the simulated flood depths are more realistic, as presented in Figure D.4. Looking at the figure, the flood depths are more realistic ranging from 0.15 meters till 2.5 meters. However, an exception is the cross section at Plombieres, at the cross section the water depth ranges till 16 meters indicating that there are single cells along the cross section with these flood depths. It appears that at the end of the tributaries the model simulates local impoundments. These impoundments are caused by infrastructure where in reality waterworks such as culverts provide this runoff. Figure D.3 illustrates an example of a local impoundment. Interestingly, when the river network is removed

at the location of the impoundment, the impoundment does not occur. As a result, the unrealistic water levels were identified and the river network was shortened. For the example in Figure D.3, it meant that the river network was shortened till the road to remove the blocking effect.

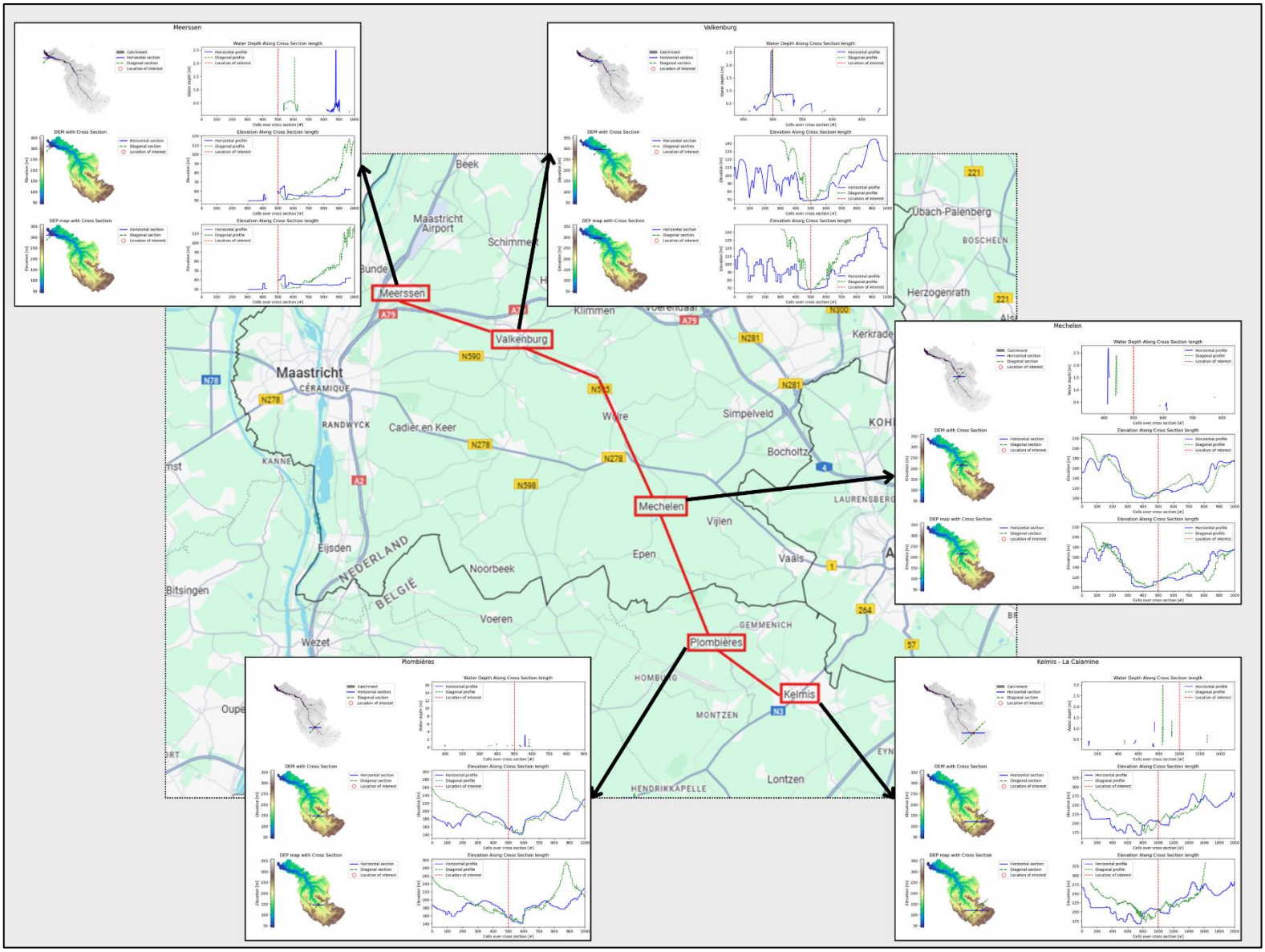


*Figure D.4: A cross-sectional analysis along the primary branch of the Geul River. The blue and green lines represent two cross sections in two directions over the grid cells. In the figure the profile of the water depth, DEM and DEP is visualized where the location of the cross sections is specified at the left from the line plots. The red line represents the location of the maximum forecasted flood depth in the catchment.*

# Appendix E: Additional plots for the Impact analysis

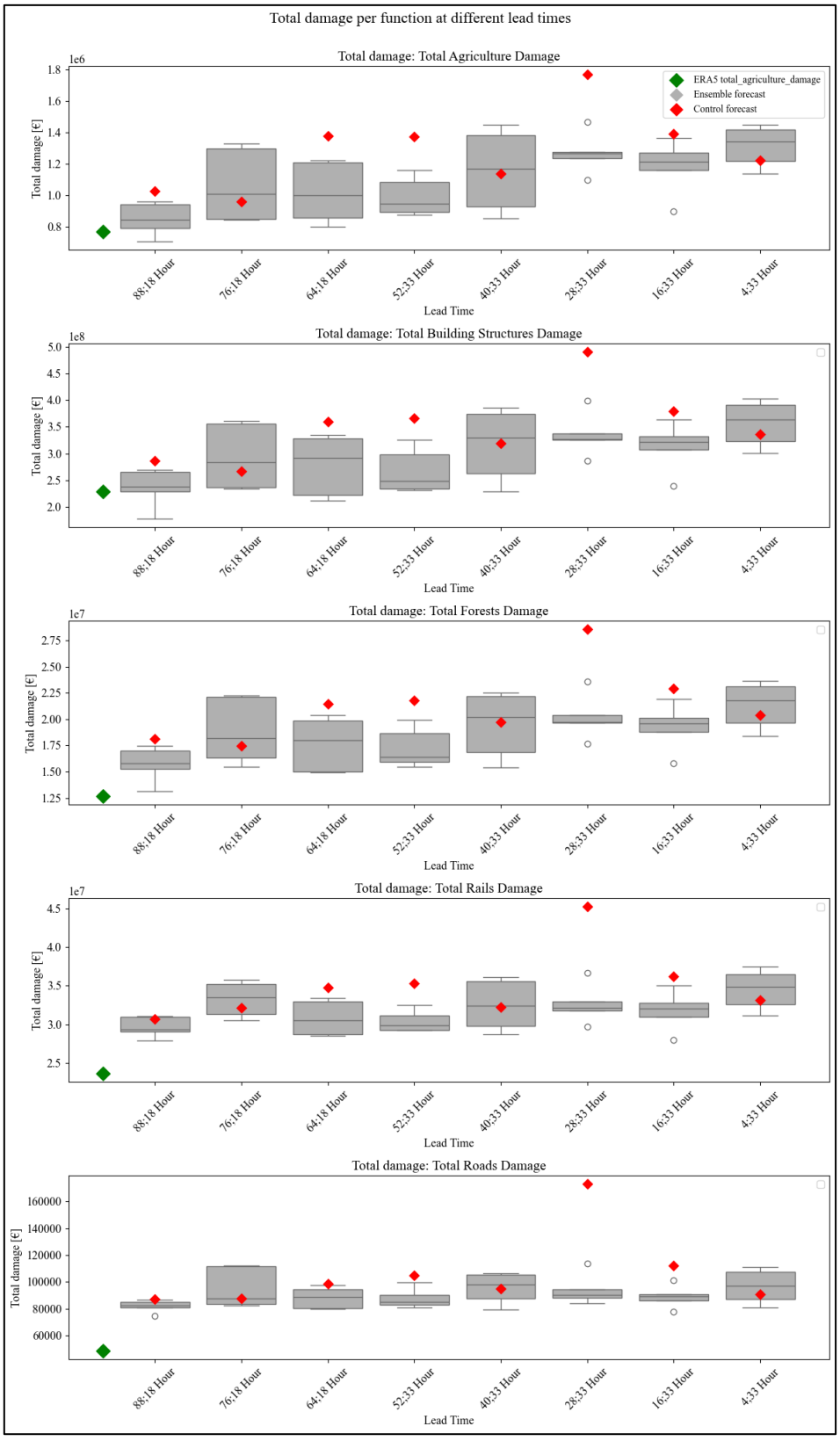


*Figure E.1: The total damage per function across the forecast types and lead times with green representing the ERA5 data, red the control forecast, and the grey boxplots the ensemble percentile forecasts.*

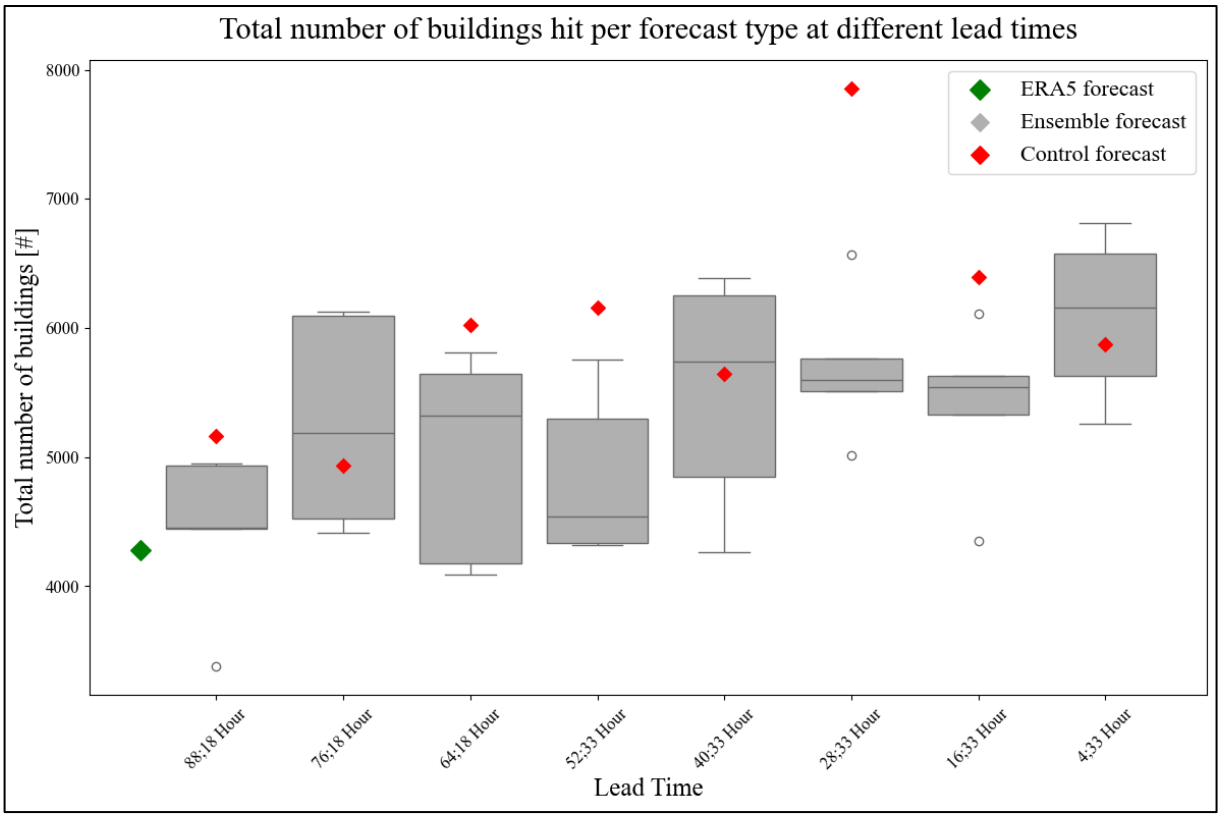


*Figure E.2: The total number of buildings hit across the forecast types and lead times with green representing the ERA5 data, red the control forecast, and the grey boxplots the ensemble percentile forecasts.*

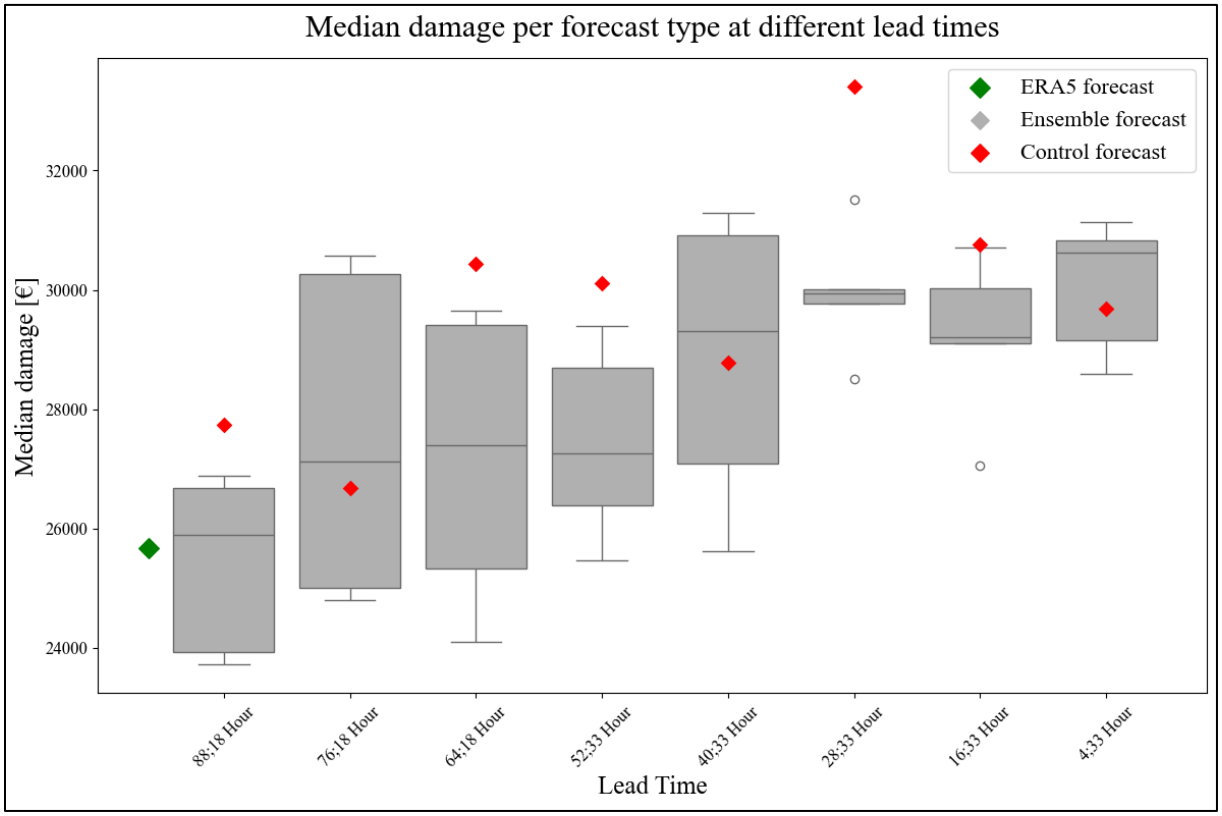


*Figure E.3: The median damage per building across the forecast types and lead times with green representing the ERA5 data, red the control forecast, and the grey boxplots the ensemble percentile forecasts. Here, the number of buildings with damages above 0 €.*

# Appendix F: Additional plots for the in-depth Impact analysis in Valkenburg

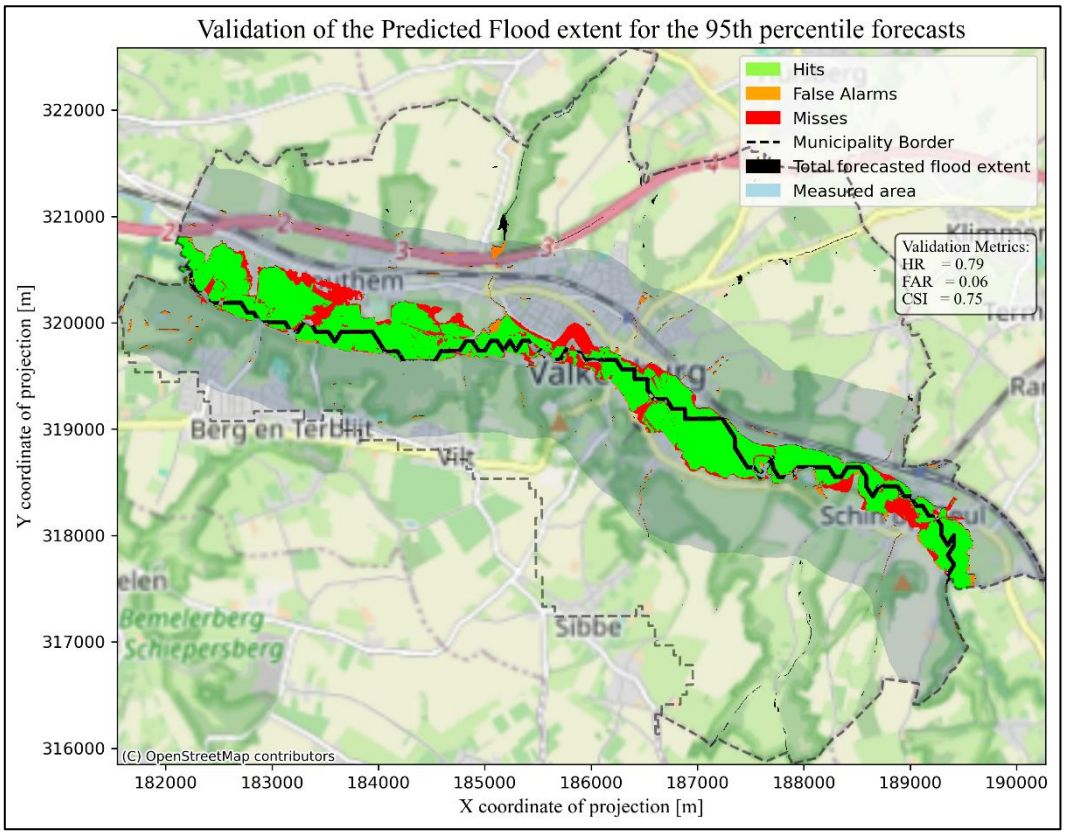


*Figure F.1: Spatial evaluation of the simulated flood extent inside the Valkenburg municipality for the 95$^{th}$ percentile ensemble forecast with 88-hour lead time as input against the observed flood extent of Slager et al. (2021). The hits are represented in light green, the false alarms in orange, the misses in red, the measured area in light blue, and the total flood extent in black.*

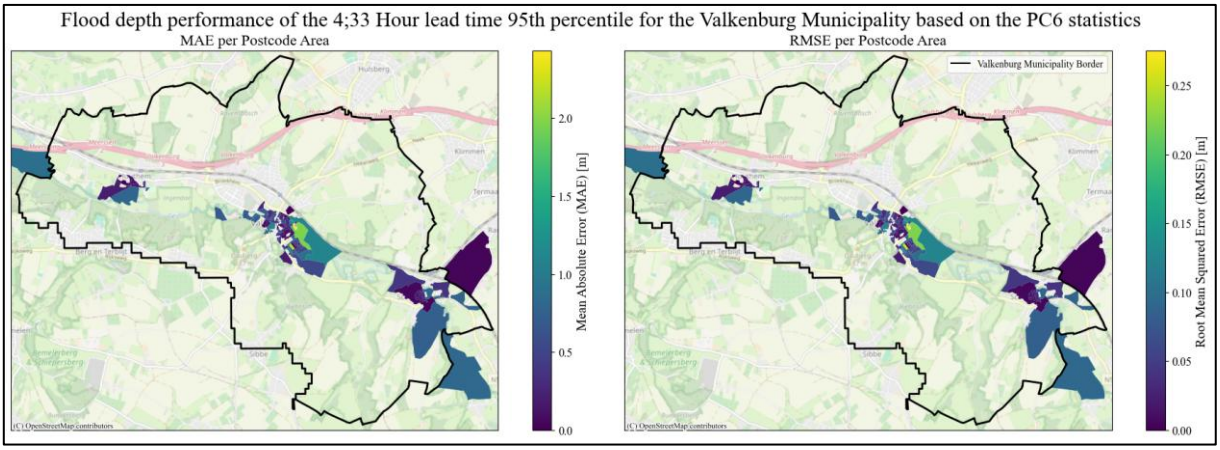


*Figure F.2: The weighted spatial evaluation of the simulated flood depths for the Valkenburg municipality with the 95$^{th}$ percentile ensemble forecast of the 88h lead time against the observations from the survey of Endendijk et al. (2023). This evaluation is based on the aggregation of the flood depths 6-digit postal code area detail level.*

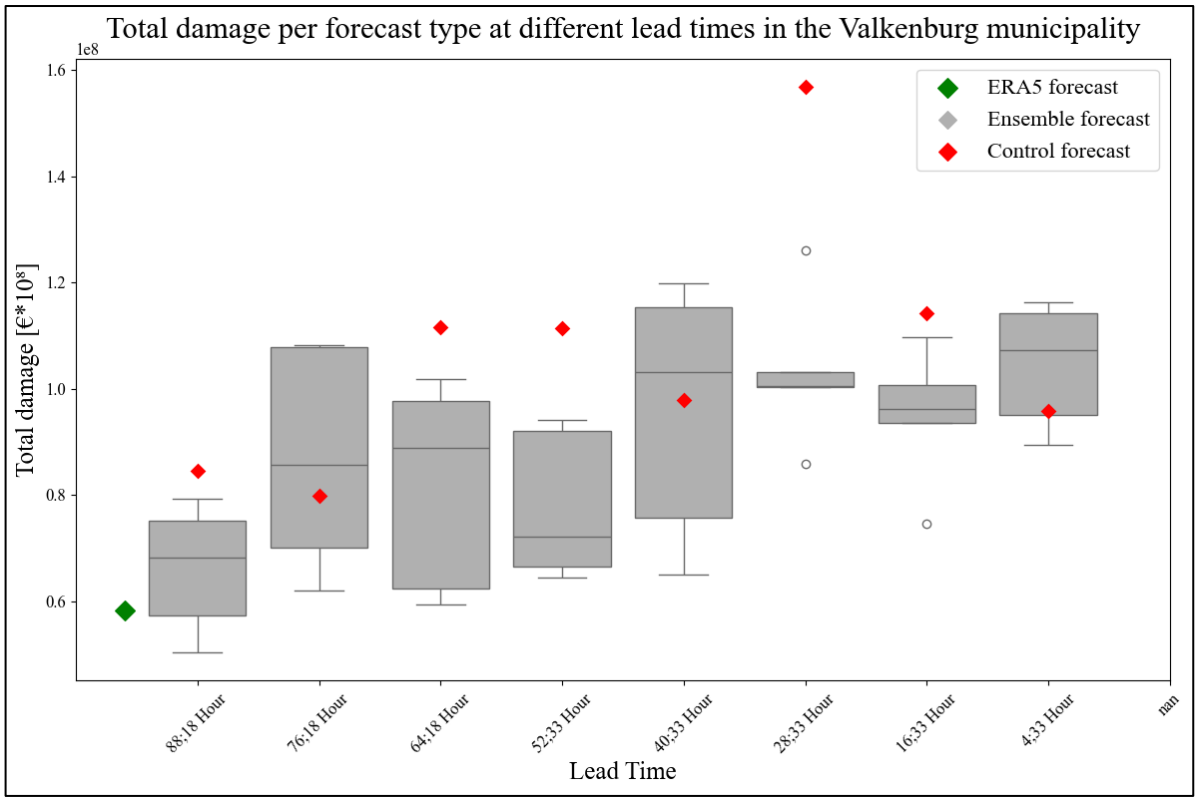


*Figure F.3: The total damages across the municipality of Valkenburg displayed for each forecast type and lead time. The total damages include the damage to forests, agriculture, infrastructure and building structures, which are derived from the vulnerability curves. The control forecast is represented in Red, the ERA5 data in green, and the ensemble percentile forecasts with the grey boxplots.*