https://doi.org/10.71573/dr9zby76

FURBAS - Development and implementation of an efficient and user-friendly model chain for early warning of urban flash floods in Hanover, Germany

Simon Berkhahn [1,*] https://orcid.org/0000-0002-4479-4741, *Insa Neuweiler* [2] https://orcid.org/0000-0002-9297-574X, *Lothar Fuchs*[1] https://orcid.org/0000-0002-9783-6255, *Stefan Krämer*[1] *& Robert Sämann*[1] https://orcid.org/0000-0002-3082-1593

[1]*Institute for Technical and Scientific Hydrology (itwh) GmbH, Hannover, Germany*
[2] *Leibniz Universität Hannover, Faculty of Civil Engineering and Geodetic Science, Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Hannover, Germany*
**Corresponding author email: s.berkhahn@itwh.de*

Abstract
Urban areas are prone to occurrence of pluvial flooding, which has the potential to inflict significant damage on urban infrastructure. At the same time, the real-time prediction of such events remains challenging. Therefore, there is a necessity for the development of advanced predictive models to mitigate potential risks and enhance urban resilience. In the present study, we show a test case for a model chain for predicting urban flooding in the city of Hanover, Germany. The model chain consists of a radar-based rainfall nowcasting model and a data-driven water level prediction model. The training data base for the water level prediction model was generated with a detailed hydrodynamic model. The objective of this study is to evaluate the predictive capabilities of the entire model chain, rather than merely those of its constituent parts. Therefore, we have utilised a historical event from 2021 to evaluate the model chain.

Highlights
- Real-time urban flood prediction with radar-based rainfall nowcasts as input to an ANN model.
- High temporal (5 min) and spatial (5 x 5 m) resolution of water level map predictions.
- The model's memory-efficient structure enables predictions for large catchment areas.

Introduction
Extreme weather events, such as heavy rainfall, are expected to increase due to climate change. These events often cause pluvial urban flooding, one of the most severe natural hazards in cities (Zanchetta & Coulibaly, 2020). Since expanding drainage systems is often impractical due to costs and space constraints, flood forecasting models are critical to mitigating risks (Rözer et al., 2021).
Urban flood prediction models typically use a chain of components, starting with rainfall measurement and nowcasting, which rely on point data (rain gauges), remote sensing (e.g., radar) or a combination of these (Berne et al., 2004). Radar-based models are preferred for capturing storm spatial characteristics, as point data alone is insufficient (Nguyen et al., 2021). The present study uses raw radar reflectivity data (Z) from the German Weather Service, corrected for physical influences such as radar signal attenuation and R-Z conversion to get current rain intensities (R), and rainfall nowcasting with the object orientated storm tracking model HyRaTrac.

Flood maps are typically generated using physically based hydrodynamic models, especially 1D/2D coupled models (Leandro et al., 2009). These models consists of a 1D sewer system model and a 2D flood model. While accurate, their computational demands limit real-time application, leading to the development of machine learning-based surrogate models (e.g., Löwe et al., 2021). These models are using simulation results from physically based models for the training process. In the present study a model based on a combination of conventional neural networks (CNNs) and recurrent neural networks (RNNs) is used to generate flood maps in real-time. This model was first introduced in Berkhahn & Neuweiler (2024).
We deliberately adopt this hybrid modelling approach, which is based on deterministic models, wherever possible, as it aims to model processes in the most transparent and explainable way possible based on known physical laws. This approach guarantees a high degree of trust and traceability, which is particularly important in safety-critical applications such as flood warnings.

So far, individual components of the model chain explained above have been evaluated independently. However, Koltermann da Silva et al. (2024) emphasize the importance of assessing the entire model chain for early warning systems. Therefore, the present study aims in evaluating the entire model chain based on a historical storm event that occurred in Hanover, Germany in 2021.

Methodology

Rainfall Nowcasting
The rainfall nowcast process used for the model chain in the present study involves two main stages: (1) Radar data processing and (2) rainfall nowcasting. The radar raw data provided by the German weather service (DWD) is processed using the NVIS server software to enhance its accuracy (more details in Krämer 2025). This includes the handling of interference echoes, the correction of radome attenuation, and the addressing of radar signal damping. The radar signals are converted (using an R-Z relation), followed by spatial and temporal interpolation at a 1-minute resolution. The processed radar data is then calibrated with ground-based measurements to ensure consistency and reliability. The generated hindcasts on a 500 x 500 metres raster are used for the nowcasting model HyRaTrac, which employs an object-oriented approach to identify hydrologically significant rainfall structures, treating them as objects. Key characteristics of these rainfall structures are derived, and a “lifetime tracking” mechanism is applied by assigning object IDs across consecutive radar data matrices. This enables the calculation of direction vectors and further refinements by incorporating rotational behaviour, improving the accuracy of the rainfall nowcast.

Water Level Prediction
The surrogate for the hydrodynamic flood model used in the present study has been adapted from Berkhahn & Neuweiler (2024), using a spatial resolution of 5x5 metres and a temporal resolution of 5 minutes. This resolution offers a practical balance between computational efficiency and the level of detail required to capture the dynamics of urban flooding on a city-wide scale. The training process uses results from the 1D/2D coupled hydrodynamic model Hystem-Extran 2D, simulating flood events using raster-based rainfall time series as input. The approach uses a simplified U-Net autoencoder architecture (Ronneberger et al., 2015) to compress flood maps into latent vectors, thereby reducing memory demands and enabling efficient processing. These compressed representations are then used in a recursive Nonlinear Autoregressive with Exogenous Inputs (NARX) model, which integrates past and present flood states along with rainfall data to predict future flood maps. Hyperparameters, such as the number of layers, neurons, filter sizes, and the length of the rainfall time series inputs, were carefully optimized for the specific test case to balance model complexity and performance. Loss functions and scaling methods were adapted to account for cells with no flooding potential, ensuring meaningful predictions for relevant areas.
A schematic representation of the entire model chain is shown in **Figure 1**.

Performance evaluation

To assess the overall performance of the model chain we used total flood volume over time as a measure of hydrological response. Furthermore, we compared the water level over time at a test location (marked in **Figure 2**) simulated with the model chain and the hydrodynamic model. To evaluate the spatial distribution of the model chain prediction, the maximum difference with respect to the hydrodynamic model for each cell was analysed. Additionally, two statistical performance metrics were used: the Nash–Sutcliffe Efficiency (NSE) and the Critical Success Index (CSI). These metrics were applied to cells where the water level in the hydrodynamic model exceeded a threshold of 10 cm. The NSE is a widely used efficiency coefficient that quantifies how well simulated values reproduce observed data. It ranges from $-\infty$ to 1, where a value of 1 indicates a perfect match, values above 0.5 are generally considered acceptable, and values below 0 suggest that the model performs worse than the mean of the observed data. The CSI is used for flood occurrence detection. It is defined as the ratio of correctly predicted flooded cells (hits) to the sum of hits, misses, and false alarms, ranging from 0 (no skill) to 1 (perfect prediction).

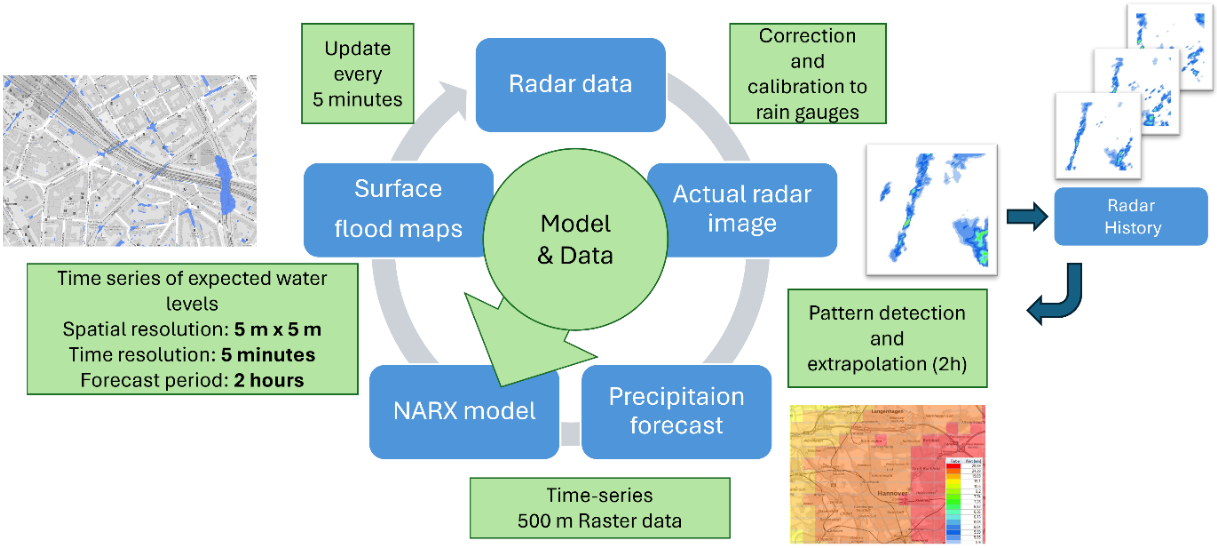


Figure 1. Schematic representation of the model chain for predicting pluvial urban floods.

Case study

The present study shows a model chain for real-time urban flood prediction for the catchment of the city of Hanover, Germany, which covers an area of approximately 260 km². The radar used for rainfall nowcasts is located at the airport north of the catchment. Most of the drainage system (80 %) consists of a separate sewer system, while the city centre area (20 %) is drained by a combined sewer system (see **Figure 2**). In addition to the sewer network, there are ditches, particularly in the outer part of the catchment area and in the urban forest, which contribute to drainage and were modelled in detail. The hydrodynamic model uses approx. 12 million triangles to represent the surface, and approx. 64,000 transport elements represent the sewer network. During this study, data from 2014 to 2022 was analysed for the radar site. From this, 85 storm events were extracted. In addition, 110 synthetic events were created by translation and rotation based on the real events. These events were simulated with the hydrodynamic model. The results of these simulations were then used to train the data-driven surrogate model. The dataset was divided into three subsets: 130 events for training, 40 events for validation of the training process (used for early stopping), and 25 events for the final testing of model performance. The predictive quality of the entire model chain is analysed using a historical event from August 2021 as a case study. It is important to note that this event is part of the test data set and was not utilised in any way for the training of the model chain. Results of the model chain are evaluated in comparison with the results of the hydrodynamic simulation. The NARX model simulation is performed in a closed-loop configuration, where water levels that have been previously predicted are recursively fed back into the model as inputs for future predictions, along with exogenous precipitation nowcasts. This configuration mirrors real-world deployment scenarios in which true future water levels are not available at the time of prediction.

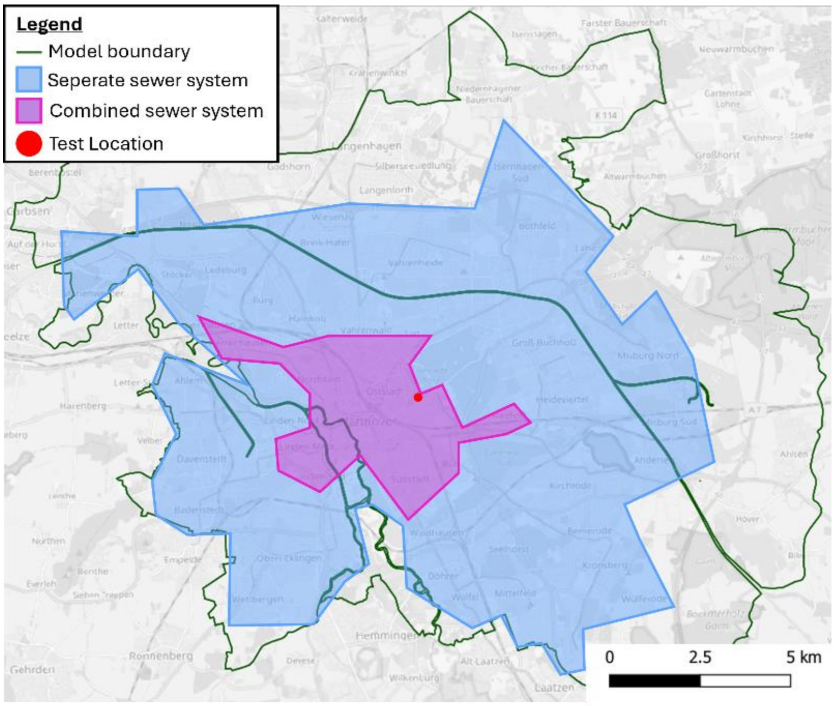
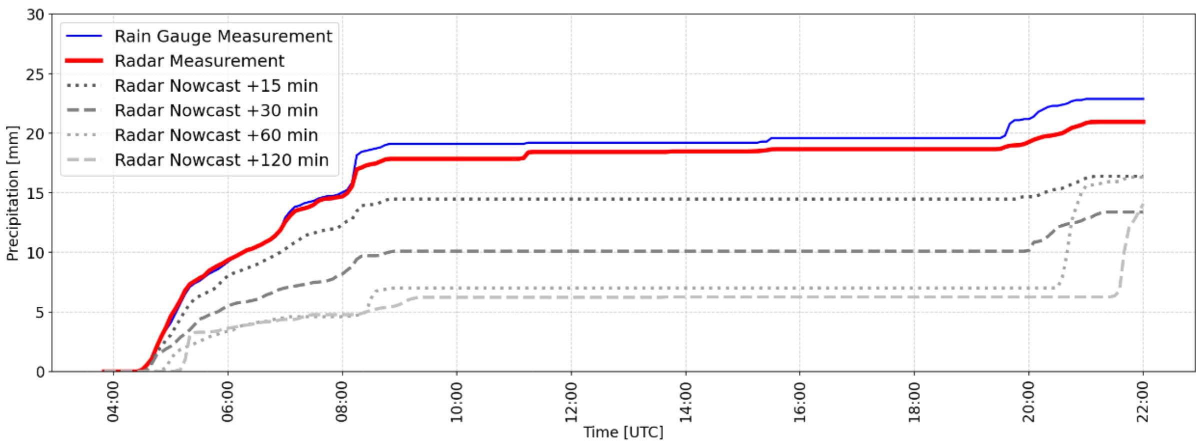


Figure 2. Map of the study catchment (Hanover, Germany), showing the model boundary (green outline), areas served by a separate sewer system (blue shading), areas served by a combined sewer system (purple shading), and a test location (red dot) for evaluation.

Results and discussion

The model chain needs approximately 30 seconds to generate water level predictions for the following 120 minutes. Additionally, around 60 seconds are required by the DWD to provide the raw radar data. In comparison the hydrodynamic model needs about 160 minutes for the same prediction period. However, it is important to note that the model chain result only contains the most essential information, namely precipitation and water level grids.



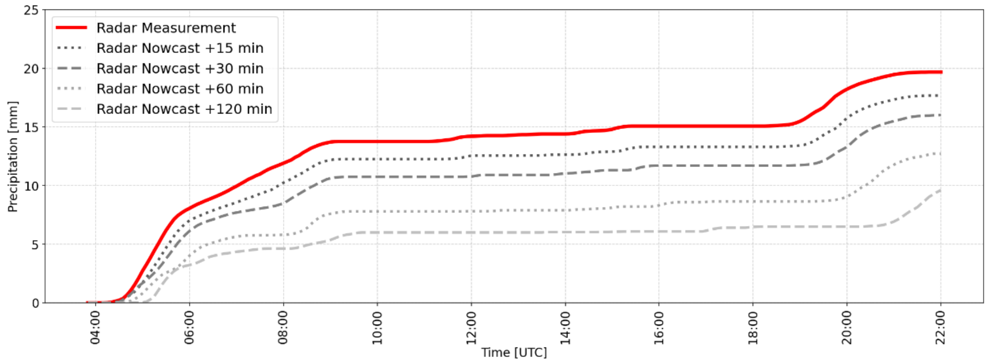


Figure 3. Comparison of cumulative precipitation from rain gauge measurements (blue line), radar-based measurement (red line), and radar nowcasts at different lead times for the rain event from August 22, 2021, for a specific location (top) and area-averaged (bottom) in Hanover, Germany.

The radar measurements shown in the top part of **Figure 3** are in good agreement with the rain gauge measurements during the early and middle stages of the event, showing almost identical accumulation patterns with minor deviations during the peak precipitation at 8:00 and about 20:00. The short-term nowcast (+15 min) also agrees well with the rain gauge data in terms of timing, but it underestimates the amount of rain. This underlines the reliability of the nowcast for short-term precipitation estimates. However, longer lead times show increasing divergence. The +30 min nowcast captures trends but underestimates peaks, while the +60 min and +120 min nowcasts show larger deviations, including overestimates in later phase of the event. The results suggest a growing storm cell at the beginning. However, the forecast model does not take this growth into account, which leads to underestimations. Conversely, in the later phase of the time period shown, the storm cell captured in the +60 and +120 min nowcast had already rained off when reaching the catchment. The area-averaged values for the whole catchment shown in the lower part of **Figure 3** show similar results. These results emphasise the uncertainties of radar-based nowcasting with longer lead times.

The total flood volume in the catchment over time is shown in **Figure 4** for both the hydrodynamic model and the model chain for different lead times. While the early part of the event (before 18:00) is predicted with reasonable low deviation across all lead times, the rising flood volume in the latter part of the event (about 20:00) is predicted to slow and with an overestimation in the end. The model chain results at +0 minutes lead time (red line) closely follows the hydrodynamic model (blue line) across the event, indicating strong short-term predictive accuracy. As the lead time of the forecast increases, the model chain tends to underestimate, a tendency that is in agreement with to rainfall prediction shown in **Figure 3**. Furthermore, increasing lead time results in oscillations. To obtain a water level prediction with higher lead times, all rainfall nowcasts up to this lead time are combined to get the necessary input. These rainfall nowcasts vary strongly in time. This could be the reason for the oscillations in the water level prediction. Nevertheless, the flood prediction seems to be robust as the oscillations are smoothed out during periods of low or no rainfall.

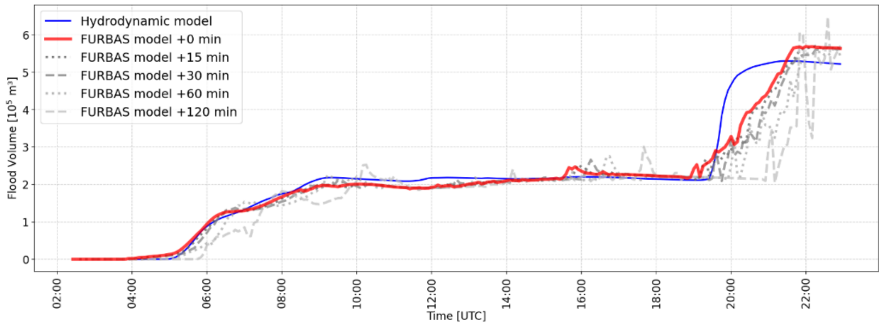


Figure 4. Comparison of total flood volume predicted by the hydrodynamic model (solid blue line) and the model chain with different lead times (solid red line and grey dashed lines) over time for the rain event from August 22, 2021 in Hanover, Germany.

A comparison of the water levels at the test location, a main drainage ditch centrally located in the study area, is shown in **Figure 5**. Both the timing and the height of the water level are predicted with very little deviation. Only in the later course of the event is the decline and rise in the water level not well met. Nevertheless, the NSE value of 0.94 for this specific location is very good. The evaluation of the NSE for all cells in which the water level exceeds a value of 10 cm is given in **Table 1**. In total, 51 percent of the cells are reaching a value of more than 0.5, which means they can be rated as satisfactory or better.

To get an impression of the spatial distribution of the deviations, the maximum difference between model chain results and hydrodynamic model results is shown in **Figure 6**. For a large part of the catchment, the model chain overestimates. The only area with a larger contiguous part of underestimation is in the north. One explanation for this distribution could lie in the division of the area for training the autoencoders used (as shown in Berkhahn & Neuweiler 2024). A further investigation is still pending.

The evaluation of the CSI for the tested event gives a value of 0.575. Comparing this value to benchmarks from similar urban flood studies (e.g. Löwe et al. 2021, Koltermann da Silva et al. 2024), it falls within the range typically considered acceptable.

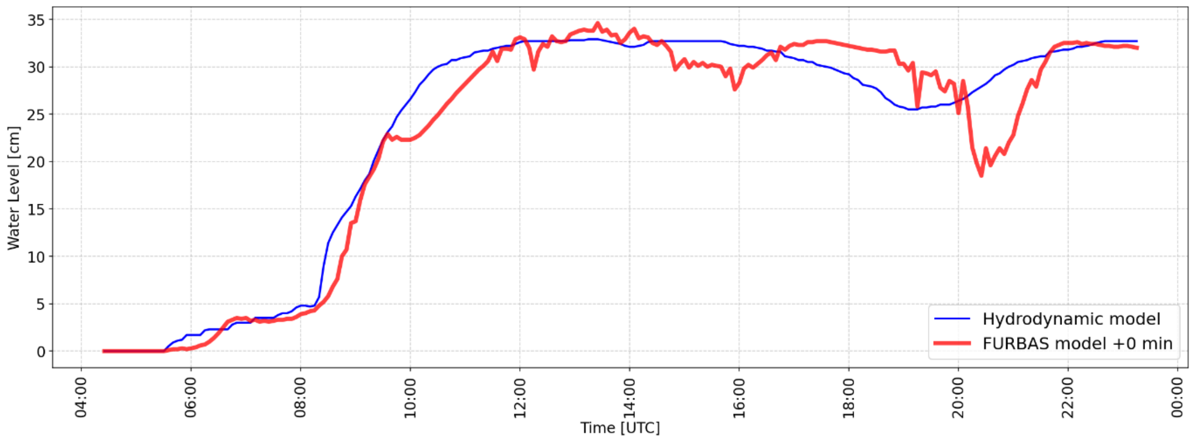


Figure 5. Comparison of water level predicted by the hydrodynamic model (blue line) and the model chain (red line) over time at the test location for the rain event from August 22, 2021 in Hanover, Germany.

Table 1. Classification of model performance based on the Nash–Sutcliffe Efficiency (NSE) according to Moriasi et al. (2015), including the percentage of cells in each category. Only cells that exceeded a value of 10 cm in the hydrodynamic simulation were considered.

NSE Value [-]	number of cells [%]	Performance Rating
NSE > 0.8	15	Very good
0.7 < NSE ≤ 0.8	14	Good
0.5 < NSE ≤ 0.7	22	Satisfactory
0.0 < NSE ≤ 0.5	30	Not satisfactory
NSE ≤ 0.0	19	Poor

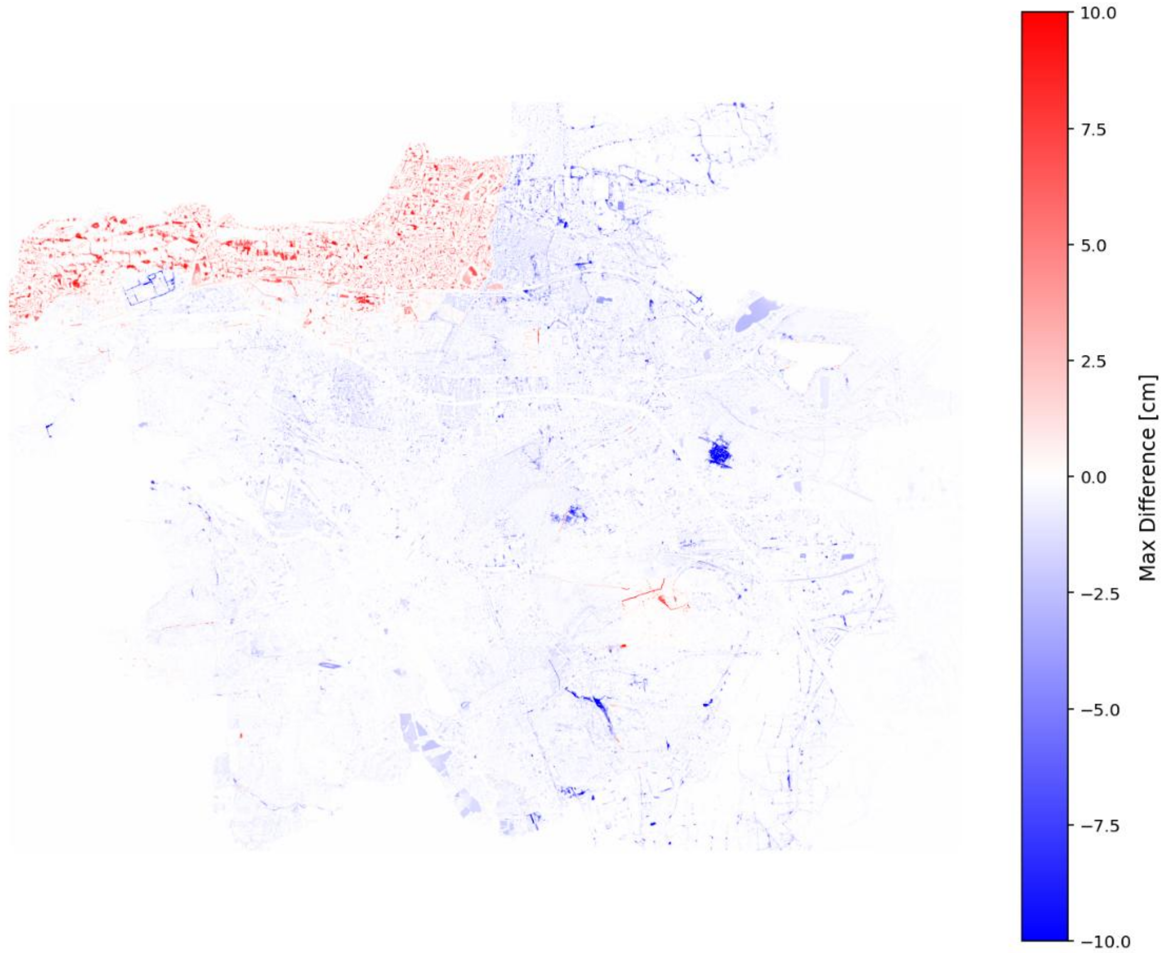


Figure 6. Map of maximal difference between the results of the hydrodynamic model and the model chain for the rain event from August 22, 2021 in Hanover, Germany. Red colour shows underestimation and blue colour shows overestimation.

Conclusions and future work

This study demonstrates that the proposed model chain, combining radar-based nowcasting and a fast data-driven water level prediction model, can provide reliable flood forecasts with significantly reduced computation time compared to conventional hydrodynamic models. With a total processing time of approximately 90 seconds for a 120-minute lead time, the model chain enables real-time applications, making it suitable for early warning systems in urban flood management.

The evaluation of rainfall forecasts shows good agreement with rain gauge data for short lead times, particularly up to +15 minutes. However, accuracy decreases with increasing lead time, leading to a corresponding reduction in flood prediction quality. Despite these limitations, the model chain performs well in predicting overall flood volume and water levels, especially in the early phases of the event. The comparison of predicted and simulated water levels at a key drainage point yielded an NSE of 0.94, indicating very good agreement. Furthermore, over half of the evaluated cells achieved NSE values greater than 0.5, highlighting satisfactory spatial performance across the catchment. The CSI value of 0.575 confirms that the model is capable of capturing inundation extents with an accuracy acceptable for urban flood applications.

Maximal deviations between the model chain and the hydrodynamic reference model were generally small but showed regional patterns of over- and underestimation. These may be influenced by the spatial division used during the training of the autoencoders, which requires further investigation. Validating the approach across multiple events and catchments will be essential steps toward operational deployment.

Improving the radar based nowcasts (particularly for higher lead times) by using data-driven models and taking spatial ground structures into account can lead to a significant improvement in the results of the entire model chain. The small-scale variability of rainfall intensities as a function of ground characteristics and orography has already been analysed for the city of Basel in Krämer et al. (2025).

The incorporation of measurement data (e.g. soil moisture, water levels or traffic data) into the model chain in real time opens up new research questions. Furthermore, the assessment and communication of uncertainties in urban flood prediction must be addressed in future work. One possible approach could be the usage of ensemble predictions.

As the effort required to generate the necessary training data is very high, it would make sense to realise a spatial transferability of the model chain. The precipitation forecast is already spatially transferable. Therefore, the part of the water level predictions must be adapted for spatial transferability. Dividing the area of the given test case into subareas would allow the existing data to be used to train a transferable model.

Acknowledgement

The study is part of the FURBAS project (Forecasting urban floods and strong rainfall events, 2022-2025). The project is funded by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection under grand number 67DAS224.

References

Berkhahn, S., & Neuweiler, I. (2024). Data driven real-time prediction of urban floods with spatial and temporal distribution. *Journal of Hydrology X*, *22*, 100167.

Berne, A., Delrieu, G., Creutin, J. D., & Obled, C. (2004). Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*, *299*(3-4), 166-179.

Koltermann da Silva, J., Burrichter, B., Niemann, A., & Quirmbach, M. (2024). A Systematic Modular Approach for the Coupling of Deep-Learning-Based Models to Forecast Urban Flooding Maps in Early Warning Systems. *Hydrology*, *11*(12)

Krämer, S., Fitzner-Pukade, D., Leberke, H., Günther, H., Lindenberg, M., Friese, F., & Würfel, S. (2025). Radar long term event time series for hydrodynamic discharge modelling. In Proceedings of the *13th Urban Drainage Modelling Conference*. Innsbruck, Austria.

Leandro, J., Chen, A. S., Djordjević, S., & Savić, D. A. (2009). Comparison of 1D/1D and 1D/2D coupled (sewer/surface) hydraulic models for urban flood simulation. *Journal of hydraulic engineering*, *135*(6), 495-504.

Löwe, R., Böhm, J., Jensen, D. G., Leandro, J., & Rasmussen, S. H. (2021). U-FLOOD–Topographic deep learning for predicting urban pluvial flood water depth. *Journal of Hydrology*, *603*, 126898.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, *58*(6), 1763-1785.

Nguyen, D. H., Kim, J. B., & Bae, D. H. (2021). Improving radar-based rainfall forecasts by long short-term memory network in urban basins. *Water*, *13*(6), 776.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.

Rözer, V., Peche, A., Berkhahn, S., Feng, Y., Fuchs, L., Graf, T., Haberlandt, U., Kreibich, H., Sämann, R., Sester, M., Shehu, B., Wahl, J. & Neuweiler, I. (2021). Impact-based forecasting for pluvial floods. *Earth's Future*, *9*(2), 2020EF001851.

Zanchetta, A. D., & Coulibaly, P. (2020). Recent advances in real-time pluvial flash flood forecasting. *Water*, *12*(2), 570.