https://doi.org/10.71573/jqeepf03

Insight into the digital twin of Hanover: real-time modelling of flooded areas

Alrun Jasper-Tönnies[1] * https://orcid.org/0000-0001-5020-2573, *Bruno Castro*[1], *Melissa Pimiento Castañeda*[1], *Thomas Einfalt* [1] https://orcid.org/0000-0003-1280-6813, *Erik Ristenpart*[2] *& Alida Reinsch*[2]

[1] *hydro&meteo GmbH, Breite Str. 6-8, 23552 Lübeck, Germany*
[2] *ifs Ingenieurgesellschaft für Stadthydrologie mbH, Stiftstr. 12, Hannover, Germany*
**Corresponding author email: jasper-toennies@hydrometeo.de*

Abstract

The Digital Twin platform developed for the “ZwillE” research project in Germany serves different purposes, covering a wide range of urban water management tasks of the wastewater utility of Hanover in Germany. Fundamental concepts, such as model integration issues and different use cases with different levels of complexity are tackled within the project. This contribution focuses on a new approach for the fast estimation of maximum water levels for use in relatively flat urban environments. The approach combines pre-calculated flood maps based on design rainfall for fixed return periods with real-time radar measurements and radar nowcasts, including uncertainty information. An evaluation with 10 extreme precipitation events in Hanover, between 2002 and 2023, was performed, where an observed flooding, and a high number of fire brigade calls were registered throughout the city. Forecast and Maximum forecast precipitation conditions are also included and validated. Results of the new approach for fast flooding estimation are assessed in comparison to fire brigade calls and to a coupled 1D sewer and 2D surface runoff simulation using the model Hystem-Extran/Hystem-Extran2D. The fast estimation shows good results of the maximum water level in comparison to the reference simulation, with absolute differences mostly below 0.1 m or 0.2 m respectively. The approach is suitable for real-time application in the newly developed digital Twin for the city Hanover.

Highlights

- Fast estimation of flooded areas using city-wide flood hazard maps and radar data suitable for real-time application.
- Good results in comparison to the reference simulations with the coupled 1D sewer and 2D surface runoff model Hystem-Extran/Hystem-Extran2D.
- The fast-flooding estimation (RadEF) is part of a newly developed digital Twin for water and wastewater for the city Hanover.

Introduction

The use of digital twins is gaining increasing popularity also in stormwater management. A digital twin is understood in this context as virtual representation of the physical water management system aiming at reflecting the current state of the real system and providing support for decision making (Schütze et al., 2023). The concept of a digital twin forms the core of the ZwillE project funded by the German Federal Ministry of Education and Research (BMBF). The project aims to support the management of water extreme events building up a digital twin of the drainage infrastructure for the city of Hanover in Germany. It is supposed to give an integrated view of the current state of the infrastructure with sewer system and sewage treatment plants under the respective current hydrological and meteorological conditions and under future extreme scenarios. The aim is to better anticipate impacts of extreme rain events in order to improve the management of such events and to support planning decisions for an increased resilience. This contribution focuses on a new approach for the fast estimation of flooded areas during extreme rain events as part of the digital twin for Hanover.

Methodology

Radar data in many places provide the best data source for current precipitation estimates. Particularly in real-time systems, compromises in data quality often have to be accepted, since rapid data availability is more important than the best possible preparation. In Germany, there are single site radar data with a resolution of 250m x 1° from the German Weather Service DWD, which are available via DWD Open Data and provide a good data basis for real-time systems. With the help of correction procedures prepared and adjusted using past data, a good correction of errors such as clutter and beam blockage can be achieved in real time (Michelson et al., 2004). The real-time (quasi-) adjustment with rain gauge data is problematic, since these can be supplied only for a past period, and many station measurements are not available in time, so that it is necessary to work with a smaller number of stations compared to subsequent processing. The radar data is corrected and quasi-adjusted in real-time using the software SCOUT from hydro & meteo GmbH and is used as input to calculate high resolution nowcasts and nowcast ensembles (Jasper-Tönnies et al., 2018). Further information on the processing of the data within the ZwillE project can be found in Jasper-Tönnies et al. (2023).

Flood hazard maps exist in many municipalities. They are based on design storms of different return periods, e.g. 10 years or 100 years, or based on an extreme event. These maps are static and, therefore, only give a rough estimate of what may happen in the real world. Within the project, flooding simulations for Hanover are conducted both for design storms and for real precipitation events using gauge-adjusted radar data as input. City-wide flood maps are produced through 1D-2D coupled simulations. For the 1D sewer simulation, the software packages Hydrological Urban Drainage Model (HYSTEM) and Explicit TRANSport model (EXTRAN) from ITWH GmbH, in the 8.6 version, were implemented. Regarding the surface runoff, HYSTEM-EXTRAN 2D was applied to obtain two-dimensional hydrodynamic calculations of the runoff taking into account realistic terrain conditions. As input, parameters such as terrain heights, building bodies, break edges, building structures and other special structures (such as walls), that could possible hinder runoff, were included for the 3D surface model creation, as well as runoff roughness parameters of the area. The sewer network model was provided by the Hanover City Drainage Authority (SEH).

The simulation results of the real precipitation events serve as reference to validate a newly developed method for a Fast estimation of the flooded areas (RadEF) within the city. The method has as input radar precipitation data and the previously described flood maps based on design storms for return periods from 1 to 100 years. The precipitation input is used to calculate the maximum accumulated rain per grid point and duration, on a 500m x 500m grid. Then, the values are related to the equivalent highest KOSTRA return period classification using the corresponding values from the KOSTRA dataset (KOSTRA-DWD-2020) for the city. The KOSTRA dataset (DWD, 2023) is the Germany-wide atlas of rainfall return periods for all duration intervals to be used in urban drainage applications. Using Geographical Information Systems (QGIS) combined with Python programming, a specific flood map for the rain event is created, extracting the correspondent piece from the design storm flood map that corresponds to the maximum return period from the KOSTRA analysis. The result is a specific flood map showing the maximum estimated water level city-wide at each point with a resolution of 3 m.

In addition to the measured and RadEF data, Forecast ($\mathrm{RadEF_f}$) and Maximum Forecast ($\mathrm{RadEF_{mf}}$) data sets were produced and included in the processing and validation phases. As the measured flood maps described before are based on the precipitation measurements of the past 24 hours, the forecast data was based on the measurements of the past 22 hours plus nowcasts for the next 2 hours and the maximum forecast data was calculated based on the maximum precipitation forecast (maximum of 10 ensemble nowcasts). The Forecast and Maximum forecast were calculated for the time steps of 120 minutes, 90 minutes, 60 minutes and 30 minutes before the end of the event.

The forecast data was also implemented and validated in a catchment-scale analysis. For this analysis, the KOSTRA data was used to classify the catchments of the area with flood related risk classes based on the maximum return period of the precipitation. The maximum return period reached in each catchment was classified as follows: 2 and 3 years return period were considered as Low-risk class, 5 and 10 years were classified as Moderate-risk, 20 and 30 years were interpreted as High-risk and return periods of 100 years or more were considered as Very-High risk category. Figure 1 illustrates the study area implementing this classification. Catchments with a return period below 2 years are considered as no risk.

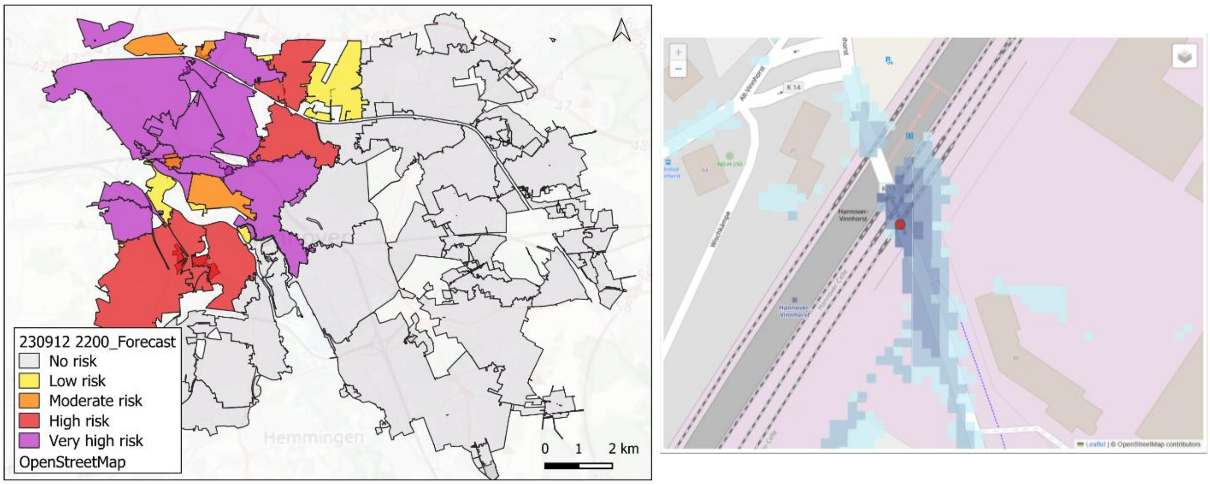

Figure 1. On the left, flood risk categories on catchment analysis level based on KOSTRA return periods. Grey colour symbolizes no risk in the catchment. On the right, a zoomed view from a section of Hanover, with the RadEF flood map in terms of maximum water height (m).

Results and discussion

RadEF estimations and validation

For the validation of the fast-flooding estimation, a random sampling point method is applied using the QGIS software, where the performance and quality of the products can be assessed through direct comparison between the generated data and the reference data from the coupled 1D/2D HYSTEM-EXTRAN simulation. The sampling, where around 100.000 random points were used, is complemented by statistical analysis using the metrics of Pearson Correlation, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), in addition to the evaluation parameters of the Probability of Detection (POD), the False Alarm Ratio (FAR), where a false alarm represents a threshold exceedance in RadEF but not in the reference simulation, and the Critical Success Index (CSI).

According to the previous information, a more detailed description of the extreme precipitation event from the 15th Oct 2019 is illustrated. For this case, the location of flood-related fire brigade calls around the city were analysed in parallel to the spatial precipitation distribution. While flood emergency reports depend on additional factors such as population density and type and state of infrastructure, for this event, a direct relation between the higher return period values and the location of most of the flood emergency reports was observed.

As a result, Figure 2 shows the maximum water level from the fast estimation (RadEF) compared to the reference simulation for a section of the city. The fast estimation shows good results in comparison to the reference simulation, with absolute differences in a wide range below 0.1 m and in the remaining areas mostly below 0.2 m. The Pearson Correlation, RMSE, and MAE reach values of 0.82, 0.06 m and 0.02 m respectively. A closer look reveals that the fast estimation tends to slightly overestimate the flood height as a result of the procedure based on evaluating the maximum accumulated precipitation for every duration without considering the temporal course.

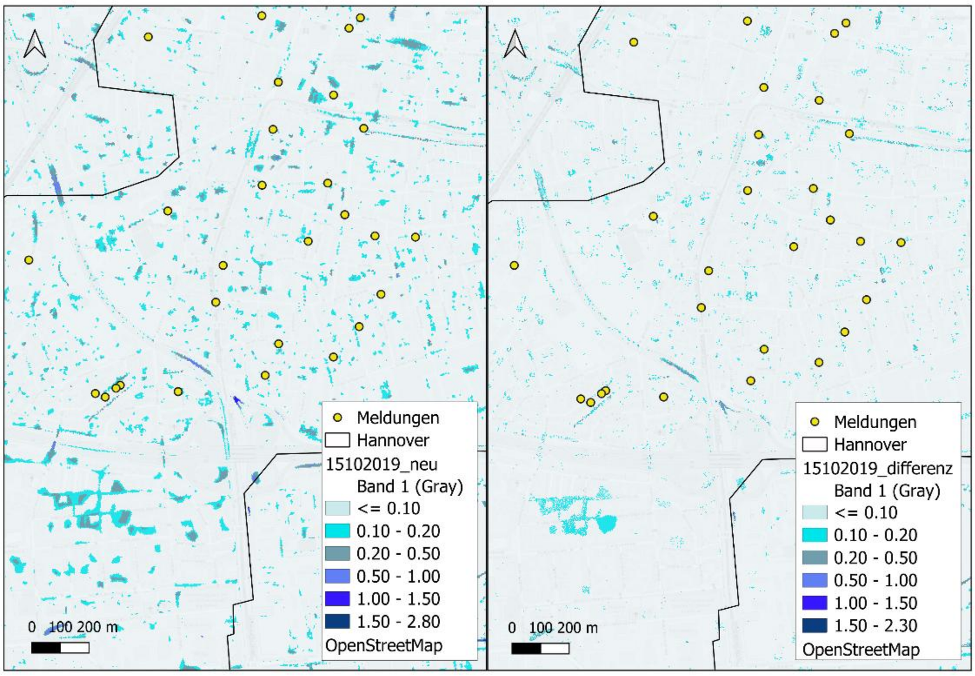

Figure 2. Maximum water level from the fast estimation (left) and the difference map (right) showing the differences between the results from the fast estimation and the reference simulation for the event on 15th Oct 2019.

The RadEF calculation method is then applied to a total of 10 extreme events between 2002 and 2023, including short convective events and long-lasting events. The results can be observed in Table 1. In this case, a high correlation between the maximum simulated water levels from the RadEF method and the reference water levels was obtained, reaching a mean Pearson correlation of 0.81 for all events, and a lowest of 0.72 as independent value. For the RMSE and MAE metrics, resulting values have mean results of 0.05 m and 0.01 m respectively, with a mean RadEF water level of 0.02m.

Table 1. Quality evaluation metrics, events between 2002 and 2023, RadEF water prediction.

Event	Duration	Pearson Correlation	MAE	RMSE	RadEF mean water high	Probability of Detection (POD)	False Alarm Ratio (FAR)	Critical Success Index (CSI)
	(h)		(m)	(m)	(m)	wh > 0.5 m	wh > 0.5 m	wh > 0.5 m
22 05 2002	2	0.81	0.02	0.05	0.03	0.87	0.41	0.55
17 07 2002	48	0.72	0.01	0.06	0.01	0.45	0.11	0.43
25 06 2006	4	0.86	0.01	0.05	0.01	0.63	0.24	0.53
26 08 2010	24	0.84	0.02	0.06	0.02	0.52	0.12	0.49
22 06 2017	1	0.81	0.01	0.04	0.02	0.76	0.29	0.58
28 07 2018	12	0.81	0.01	0.04	0.01	0.92	0.57	0.42
15 10 2019	2	0.82	0.02	0.06	0.03	0.99	0.66	0.34
14 08 2020	2	0.83	0.01	0.05	0.02	0.92	0.62	0.37
16 06 2020	3	0.82	0.02	0.05	0.03	0.73	0.30	0.55
12 09 2023	24	0.80	0.01	0.05	0.01	0.39	0.04	0.38
Mean		0.81	0.01	0.05	0.02	0.72	0.34	0.46

Regarding the POD, FAR and CSI parameters, rates of 72%, 34% and 46% were obtained for conditions above 0,5m water height (wh). Based on this, it can be said that water levels tend to be slightly overestimated for short duration events (less than three hours) and underestimated for longer events lasting 24 hours or more.

Forecast validation (RadEF$_f$)

Once the validation for the RadEF results was performed, an additional phase is implemented using forecast data mentioned before, referred as RadEF$_f$. This section only focused on the short duration events (4 hours or less), reducing the number of evaluated events from 10 to 6 extreme cases. The results in comparison to the HYSTEM-EXTRAN reference simulation can be observed in Table 2, were the performance metrics and additional parameters are computed for a forecast time step of 60 minutes before the end of the event.

Table 2. Quality evaluation metrics, forecast data (RadEF$_f$) for short duration events between 2002 and 2020. POD, FAR and CSI computed for 60 minutes forecast time step.

Event	**Duration**	**Pearson Correlation**	**MAE**	**RMSE**	**RadEF$_f$ mean water high**	**Probability of Detection (POD)**	**False Alarm Ratio (FAR)**	**Critical Success Index (CSI)**
	(h)		**(m)**	**(m)**	**(m)**	**wh > 0.5 m**	**wh > 0.5 m**	**wh > 0.5 m**
22 05 2002	2	0.72	0.01	0.04	0.01	0.85	0.63	0.45
25 06 2006	4	0.36	0.01	0.03	0.002	0.22	0.56	0.17
22 06 2017	1	0.57	0.01	0.04	0.01	0.72	0.69	0.28
15 10 2019	2	0.76	0.01	0.05	0.02	0.87	0.71	0.27
16 06 2020	3	0.71	0.01	0.03	0.01	0.37	0.23	0.33
14 08 2020	2	0.45	0.01	0.04	0.01	0.83	0.88	0.12
Mean		0.60	0.01	0.04	0.01	0.64	0.62	0.27

It was observed that the forecast quality varies from event to event; however, the hit rate (POD) for a threshold exceedance of 0.5 m is, on average, relatively close to the value for the real-time case (64% versus 72%), whereas the proportion of false alarms is significantly higher (62% versus 34%). For the CSI, a mean value of 0.28 was reached, meaning that the forecast quality "at the point" is similar to the precipitation forecast over one hour in case of heavy rain events of current precipitation forecasting methods. Comparable values were determined, for example, in a study by Ayzel & Heistermann (2025) for the forecast accuracy of aggregated precipitation over one hour, with thresholds of 5-10 mm/h. For the performance metrics, mean values of 0.60, 0.01 and 0.04 were achieved for the Person correlation, the MAE and RMSE respectively, considered as good performance results in comparison to other recent studies that use machine learning algorithms (ML) such as Random Forest (RF) or K-nearest neighbor (KNN), or ML-based prediction methods (Hou, J., Zhou, N., Chen, G. et al., 2021; Bao, D., George Xue, Z., Hiatt, M. et al., 2025).

Forecast validation (RadEF$_f$), catchment level analysis

As mentioned in the methodology section, the quality of the produced forecast data was also assessed on a catchment-level, with a total of 83 evaluated catchments with areas between 1 km^2 and 20 km^2. Figure 3 illustrates the distribution of the accuracy for the evaluated data averaged over all the short duration events showed in Table 2. In this case, results are considered accurate when the prediction of a flood risk category in the forecast data set (RadEF$_f$), corresponds with the measured category (RadEF) in the same reference point. Outcomes showed whether the predicted category was precisely the same as the measured one (accurate), or if it was shifted by one class (Neighbor) or multiple classes (Higher), indicating the degree of discrepancy. This analysis was done in dependency of the lead time over the entire forecast window, from 120 minutes until the end of the event.

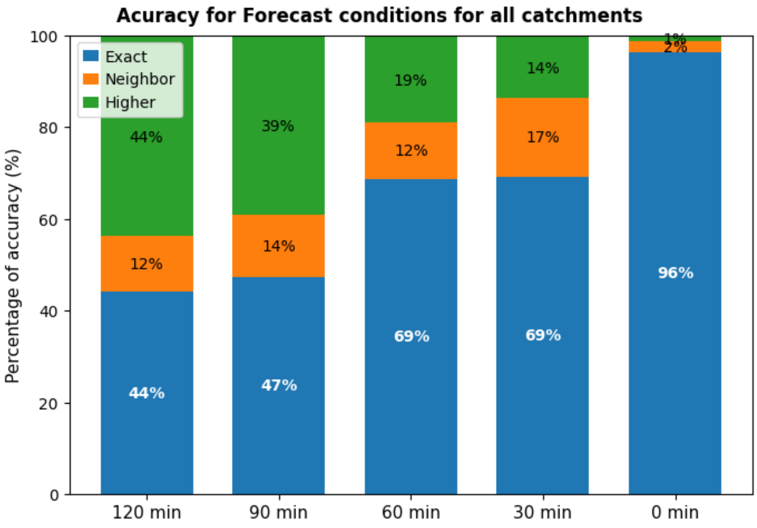

Figure 3. Accuracy results for forecast evaluation (RadEF$_f$) in dependency of the lead time, catchment-scale analysis, average over all events.

It is observed that the prediction of the exact class improves from 44% in the 120 min forecast, to almost 70% in the 60 minutes and 30 minutes forecasts. This is comparable to the POD result mentioned in Table 2, which reached a value of 64% for water heights above 0.5m in the 60 min forecast. Furthermore, less than 15% of the predictions for all forecast time steps fell in a neighbouring class, and less than 19% of the predictions from the 60 minutes time forecast onwards were not accurate and were classified in another category.

In order to address the warning perspective, where one of the most important elements is the production and availability of relevant and reliable information on time, an additional evaluation is performed, comparing the forecast quality of the Maximum Forecast data sets (RadEF$_{mf}$), which consider worst precipitation scenarios in the city, with the Mean Forecast (RadEF$_f$). The quality analysis was performed specifically to assess the Detection Rate (DR) of RadEF$_f$ and RadEF$_{mf}$ data sets. We evaluate the percentage of catchments with correctly predicted classes (correct risk class or a higher risk class) over all classes with observed risk class > 0 (DR class > 0). In addition, we evaluate the percentage of correctly predicted classes without flooding risk (DR class = 0). Figure 4 illustrates the results of the detection rate percentages from RadEF$_f$ and RadEF$_{mf}$ in dependency of the lead time.

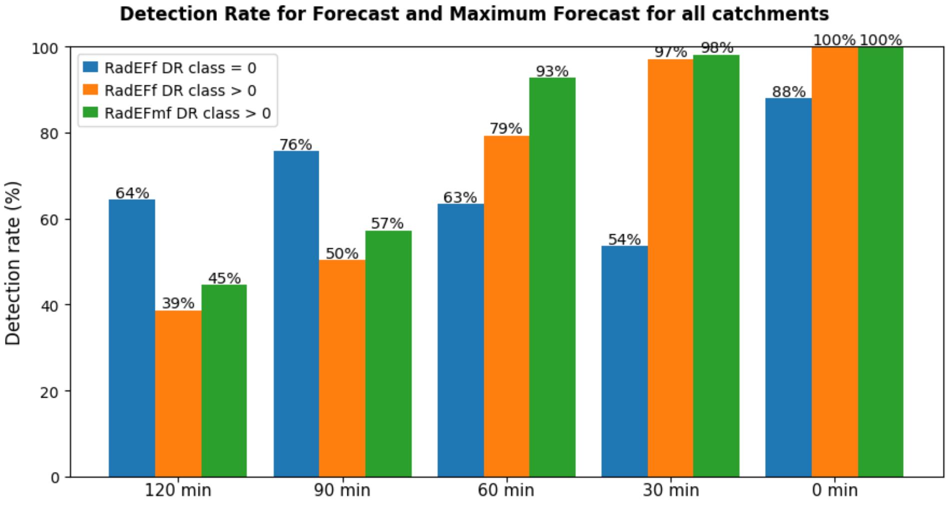

Figure 4. Detection rate (DR) for Forecast (RadEF$_f$) and Maximum Forecast (RadEF$_{mf}$) conditions, on catchment-scale, for all time steps for all events. In blue, detection rate for risk class 0. In orange and green, detection rate for risk classes > 0 (correct or higher) for Forecast and Maximum Forecast.

From the RadEF$_f$ results for the 60 minutes time step, 63% were correctly predicted as class 0 (no risk), while from the total catchments that were measured as classes > 0 (Low, Moderate, High and Very High), 79% were correctly detected. For the RadEF$_{mf}$, a value of 93% is reached for all classes above 0 at the same time interval. These last percentages increase considerable for classes above 0, both for RadEF$_f$ and RadEF$_{mf,}$ in the 30 minutes time step, reaching detection rates values of 97% and 98% respectively. Taking this into account, it can be said that the Maximum forecast predictions have a very high confidence level and are ideal to be used as input in warning systems and flood management initiatives, due to its high detection rate for all risk classes.

It is important to mention that when evaluating the results, uncertainties related to heavy rain events, resulting from precipitation measurement and forecasting, in addition to possible uncertainties related to runoff behaviour (e.g., clogged gullies or culverts), should be considered. In this case, results indicate that the inaccuracies of the RadEF method compared to the reference simulation are not significant, for example, when it comes to short-term decisions during a heavy rainfall event. At the same time, the results confirm the overall high uncertainties in the prediction of heavy rainfall and flooding, as described in other studies (e. g. Hofmann & Schüttrumpf, 2021; Ayzel & Heistermann, 2025).

Conclusions and outlook

The RadEF method offers several advantages over AI-based approaches such as neural networks (NNs), Random Forest (RF), K-nearest neighbor (KNN), or convolutional neural networks (CNNs), with comparable performance results. Transferability and updating are significantly easier, as extensive training with large amounts of data is not required. While the method is not considering long flow paths, it is ideal for city-wide flood-prediction processes in flat environments, with short processing times below 5 min and a simple set up. The method is transparent and can be replicated step by step, which facilitates the understanding of simulation results, implementing as inputs radar precipitation data, KOSTRA maps and general flood maps based on design storms. Especially in the case of heavy rain events, where inaccurate forecasts and false alarms must always be considered, its explainability represents an advantage over black-box models such as NNs or CNNs.

Results prove to have high confidence and quality outcomes for different extreme rain events in Hanover, especially for short duration related convective rainfall. This represents an alternative, but also a useful complement, to AI-based methods for generating comparative data and creating a more robust basis for decision-making in the occurrence of an event, especially when maximum forecast conditions are taken into consideration, illustrating a clear relationship between precipitation and flooding. This illustrates potential ways for how heavy rain hazard maps can be used and interpreted when combined with current precipitation data.

Currently, the RadEF method is being implemented at the Hanover Municipal Authority (Stadtentwasserung Hannover, SEH), as part of the digital twin developed in the ZwillE project. It allows a simultaneous view of the flooding situation for current, forecast, and maximum forecast conditions, both at the catchment level, in terms of flood risk categories, and with a detailed level showing the maximum water height city-wide at a 3 m-resolution.

As future work, the catchment scale risk classification applied in this method that includes the Low, Moderate, High and Very High classes, both for forecast and maximum forecast conditions, will be extended with an Urban flood risk index which is related to the build-up areas from the city and the expected water levels at particularly affected areas, in order to obtain and visualize a Combined flood risk index. The integration of the combined index will represent a further step in supporting the management and decision-making during heavy rain events.

Acknowledgement

The Federal Ministry of Research, Space and Technology (BMFTR) is funding the project „ZwillE – Digitaler Zwilling zum KI-unterstützten Management von Wasser-Extremereignissen im urbanen Raum“ (FK 02WEE1627E) within the “Hydrological extreme events (WaX)” funding measure as part of the federal research program on water “Wasser: N”. Wasser: N contributes to the BMFTR “Research for Sustainability (FONA) Strategy”.

References

Ayzel, G. & M. Heistermann (2025): Brief communication: Training of AI-based nowcasting models for rainfall early warning should take into account user requirements. – Natural Hazards and Earth System Sciences, 25(1), 41-47. DOI: https://doi.org/10.5194/nhess-25-41-2025.

Bao, D., George Xue, Z., Hiatt, M. et al. A machine learning-based prediction-to-map framework for rapid and accurate spatial flood prediction. npj Nat. Hazards 2, 71 (2025). https://doi.org/10.1038/s44304-025-00122-2

DWD (2023) Koordinierte Starkregenauswertung und -regionalisierung des Deutschen Wetterdienstes. https://dwd.de/kostra.

Hofmann, J. & H. Schüttrumpf (2021): floodGAN: Using Deep Adversarial Learning to Predict Pluvial Flooding in Real Time. – Water, 13(16), 2255. DOI: https://doi.org/10.3390/w13162255

Hou, J., Zhou, N., Chen, G. *et al.* Rapid forecasting of urban flood inundation using multiple machine learning models. *Nat Hazards* 108, 2335–2356 (2021). https://doi.org/10.1007/s11069-021-04782-x

Jasper-Tönnies, A., Hellmers, S., Einfalt, T., Strehz, A., & Fröhle, P. (2018). Ensembles of radar nowcasts and COSMO-DE-EPS for urban flood management. Water Science and Technology, 2017(1), 27-35.

Jasper-Tönnies, A.; Einfalt, T.; Schütze, M.; Ristenpart, E.; Strehz, A.: Advanced real-time precipitation components for urban hydrological applications as part of a digital twin for the city of Hanover, Posterpräsentation, 11th Novatech 2023, Lyon, verfügbar auf: https://hal.science/hal-04185043/document

Michelson, D., T. Einfalt, I. Holleman, U. Gjertsen, K. Friedrich, G. Haase, M. Lindskog, A. Jurczyk (2004) Weather Radar Data Quality in Europe: Quality Control and Characterization, Report of COST 717, EU, 91 pages.

Schütze, M., Förster, L., Hübner, C., Alex, J. (2023): A Digital Twin platform for purpose-driven modelling; Novatech 2023: 11e Conférence internationale sur l'eau dans la ville, Jul 2023, Lyon, France. ⟨hal-04176903⟩