Real-Time Data-Driven Urban Flood Forecasting

DISSERTATION
zur Erlangung des Grades eines Doktors
der Ingenieurwissenschaften

vorgelegt von
Felix Schmid, M.Sc.

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät
der Universität Siegen
Siegen 2025

Betreuer und erster Gutachter
Prof. Dr. habil. Jorge Leandro
Universität Siegen

Zweiter Gutachter
Prof. Dr.-Ing. Markus Disse
Technische Universität München

Tag der mündlichen Prüfung
28. Juli 2026

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer, nicht angegebener Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Es wurden keine Dienste eines Promotionsvermittlungsinstituts oder einer ähnlichen Organisation in Anspruch genommen.

Vorwort

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Hydromechanik und Wasserbau des Forschungsinstituts Wasser und Umwelt (fwu) am Department Bauingenieurwesen der Naturwissenschaftlich-Technischen Fakultät der Universität Siegen. An dieser Stelle möchte ich den nachfolgend genannten Personen meinen besonderen Dank aussprechen. Ohne ihre Unterstützung wäre diese Promotionsschrift nicht möglich gewesen.

Zunächst möchte ich meinem Doktorvater Herrn Prof. Dr. habil. Jorge Leandro ganz herzlich danken. Er hat die Entstehung dieser Arbeit überhaupt erst ermöglicht gemacht. Die mir gewährte Zeit, seine wertvollen Anregungen, Offenheit und kritischen Rückmeldungen bildeten eine ausgezeichnete Grundlage für die Bearbeitung der in dieser Dissertation behandelten Forschungsschwerpunkte. Ebenso danke ich meinem Zweitgutachter und Betreuer, Herrn Prof. Dr.-Ing. Markus Disse von der Technischen Universität München, für die angenehme und konstruktive Zusammenarbeit.

Meiner Kollegin Leonie Müller sowie meinen Kollegen Taylor Johnson und Sebastian Ramsauer gilt mein herzlicher Dank. Sie haben meine fachliche wie auch persönliche Weiterentwicklung maßgeblich begleitet und mich stets unterstützt. Zudem möchte ich Dr. Simon Berkhahn von Itwh Hanover für den angehnemen und fachlich sehr hilfreichen Austausch danken.

Mein großer Dank gilt zudem meinen Eltern, für ihre bedingungslose Unterstützung in allen Lebensbereichen. Sie haben meinen bisherigen Lebensweg ermöglicht und damit in erheblichem Maße zum Entstehen dieser Arbeit beigetragen. Zuletzt möchte ich meiner Frau Sarah von Herzen Danke sagen: Wegen ihrer Geduld, Liebe und Ermutigung konnte diese Arbeit erst entstehen. Ihr ist diese Dissertation gewidmet.

Kurzfassung

Hochwasservorhersagen sind ein entscheidender Bestandteil zur Risikominimierung für Mensch, Infrastrukturen und Ökosysteme, insbesondere da der Klimawandel die Häufigkeit und Intensität extremer Wetterereignisse erhöht. Vor allem urbane Gebiete sind aufgrund der hohen Versiegelungsraten, dichter Bebauung und Bevölkerungskonzentration besonders anfällig. Traditionelle Vorhersagen basieren auf physikalischen hydrodynamischen Modellen, die präzise Informationen über Ort, Zeitpunkt und Ausmaß von Hochwasserereignissen liefern. Obwohl diese Modelle den Stand der Technik darstellen, sind sie für eine Echtzeitvorhersage und den operationellen Einsatz aufgrund ihres hohen Rechenaufwands ungeeignet. Eine Alternative bieten datengetriebene Modelle, die Muster und Zusammenhänge direkt aus den Daten erlernen. Durch die Approximation der zugrunde liegenden Eingangs-Ausgangs-Dynamik, ohne die physikalischen Gleichungen explizit zu lösen, erreichen datengetriebene Ansätze deutlich verkürzte Rechenzeiten und eignen sich damit für die Echtzeitanwendung in der Hochwasservorhersage.

Diese Dissertation entwickelt und erforscht datengetriebene Vorhersagesysteme, die speziell für urbane Umgebungen ausgelegt sind, und behandelt dabei die drei Haupt-Überflutungstypen: (1) Kanalnetzüberlastungen (Überstau), (2) Flusshochwasser (fluviale Ereignisse) und (3) Starkregenereignisse (pluviale Ereignisse). Die Forschung und Entwicklung folgen dabei einer strukturierten Entwicklung: von lokalisierten Kurzfristprognosen über großskalige dynamische Vorhersagen bis hin zu gebietsunabhängigen, übertragbaren und physikalisch informierten Vorhersagen.

Für Kanalnetzüberlastungen wurde ein Ensemble-Vorhersagesystem entwickelt, welches auf einem nichtlinearen autoregressiven Netz mit exogenen Eingaben basiert. Das System prognostiziert den Beginn und die Dauer von Überstau-Ereignissen in Kanalschächten und implementiert sowohl strukturelle Modell- als auch Parameterunsicherheiten explizit durch Ensemblemodellierung. Mit einer konsistenten Erfassung beobachteter Ereignisse innerhalb des Unsicherheitsbands stellt das System ein zuverlässiges Frühwarnsystem dar, das zudem modell- oder echtzeitbasierte Steuerung moderner Kanalnetze unterstützen kann.

Für Flusshochwasser wurde ein Feature-informierts Vorhersagesystem entwickelt, welches die Vorhersage als Bild-zu-Bild-Übersetzung formuliert. Durch die Integration von faltenden neuronalen Netzen mit räumlichen Merkmalen wie der Distanz zum Fluss, sagt das Feature-informierts Vorhersagesystem die maximalen Überflutungstiefen innerhalb weniger Sekunden voraus und generiert eine Karte zur Unterstützung von Echtzeit-Risikobewertungen, Verkehrssteuerung und Katastrophenschutzmaßnahmen. Es erreicht Genauigkeiten, die mit hydrodynamischen Modellen vergleichbar sind, und demonstriert damit die Skalierbarkeit datengetriebener Ansätze für Flusshochwasser in urbanen Gebieten. Aufbauend darauf wurde ein dynamisches Vorhersagesystem entwickelt, das mithilfe rekursiver faltender neuronaler Netze zeitlich aufgelöste Überflutungskarten für bis zu 24 Stunden im Voraus erzeugt. Im Vergleich zu physikalisch basierten Modellen weist das dynamische System eine hohe Genauigkeit und Zuverlässigkeit auf und schließt damit die Lücke zwischen statischen Maximalen-Wasserstandskarten und der dynamischen Echtzeitüberwachung. Da konventionelle datengetriebene Modelle im Allgemeinen standortspezifisch sind, ist ihre Übertragbarkeit auf andere Gebiete stark eingeschränkt.

Zur Überwindung dieser Limitation wurde im Kontext von pluvialen Ereignissen ein physikalisch informiertes räumlich und zeitlich aufgelöstes System entwickelt. Dieses integriert faltende neuronale Netze mittels Bild-zu-Bild-Übersetzung mit physikalischen Bedingungen, indem die Kontinuitätsgleichung sowie eine kinematische Wellenapproximation in die Verlustfunktion eingebettet werden. Dieses Framework verbessert die physikalische Plausibilität, erhöht die Vorhersagege-

nauigkeit und ermöglicht standortunabhängige Prognosen, die auf bisher unbekannten Flächen zuverlässig angewendet werden können.

Zusammenfassend zeigt diese Dissertation, wie datengetriebene und im Speziellen physikalisch informierte Modelle, echtzeitfähige, skalierbare und übertragbare Werkzeuge für die Hochwasservorhersage bereitstellen können. Die entwickelten Systeme erweitern sowohl den Stand der Forschung, als auch den der Technik, indem sie Recheneffizienz, Prognosegenauigkeit und physikalische Konsistenz vereinen. Damit bilden sie eine Grundlage für weitere Forschungen im der Bereich Hochwasservorhersage- und Frühwarnsysteme für urbane Räume.

Abstract

Flood forecasting is a critical component of mitigating risks to human life, infrastructure, and ecosystems, particularly as climate change increases the frequency and severity of extreme weather events. Especially urban areas are vulnerable due to high sealing rates, dense infrastructure, and concentrated populations. Traditional forecasting approaches rely on physically based hydrodynamic models that provide accurate information on the location, timing, and extent of flood events. While these models represent the state of the art, they are computationally too demanding for real-time forecasting and operational decision support. An alternative is offered by data-driven models, which learn patterns and relationships directly from the data. By approximating the underlying input–output dynamics without solving the governing physical equations explicitly, data-driven approaches achieve substantially reduced computation times, making them suitable for real-time flood forecasting.

This dissertation develops and advances data-driven forecast systems tailored for urban environments, addressing three major flood types: (1) sewer surcharge events, (2) fluvial floods, and (3) pluvial floods. The research follows a structured progression, advancing from localized short-term predictions to large-scale dynamic forecasts, and finally to domain-independent transferable, physically informed frameworks.

For sewer surcharge events, an Ensemble Forecast System based on nonlinear autoregressive networks with exogenous inputs was developed. The Ensemble Forecast System predicts the onset and duration of surcharge events in manholes while explicitly quantifying structural model- and parameter uncertainty through ensemble modeling. With observed events consistently captured within uncertainty bounds, the system provides a reliable framework for early warnings and can support model-based or real-time control in modern sewer networks.

For fluvial floods, a Feature-Informed Forecast System was introduced, framing flood prediction as an image-to-image translation task. By integrating convolutional neural networks (CNNs) with spatial features, such as distance-to-river, this system predicts maximum inundation depths within seconds, generating inundation maps that support real-time risk assessment, traffic management, and emergency response planning. The Feature-Informed Forecast System achieves accuracy comparable to that of hydrodynamic models, thus demonstrating the scalability of data-driven approaches to river flooding in urban areas. Building on this, to capture flood dynamics over time, a Multi-Step Dynamic Forecast System was developed. Employing recursive CNN-based predictions, the Multi-Step Dynamic Forecast System produces temporally resolved inundation maps for up to 24 hours ahead. Compared to physically based models, it maintains high accuracy and reliability, thereby bridging the gap between static maximum flood maps and real-time dynamic monitoring.

Since these and, in general, conventional data-driven models are often case-site specific and therefore domain-dependent, their applicability in different sites is limited. To overcome this constraint, this dissertation proposes a physically informed spatial-temporal forecast framework, in the context of pluvial floods. This approach integrates CNN-based image-to-image translation with physical constraints by embedding the continuity equation and a kinematic wave approximation into the loss function. This framework improves physical plausibility, enhances predictive accuracy, and enables domain-independent forecasts that generalize to new and unknown areas.

Collectively, the contributions of this dissertation demonstrate how data-driven, especially physically informed, models transform flood forecasting into a real-time, scalable, and transferable application. The developed systems advance the state of the art by combining computational efficiency, predictive accuracy, and physical consistency, laying a new foundation for further research in the field of flood forecasting and early warning systems in urban areas.

Contents

Eidesstattliche Erklärung . . . 1
Vorwort . . . 2
Kurzfassung . . . 3
Abstract . . . 5
1 Introduction . . . 8
1.1 Frequencies of flooding and flood damages . . . 8
1.2 Urban flooding . . . 9
1.3 Urban flooding modeling . . . 10
1.4 Problem statement: Urban flood forecasting and its limitations . . . 12
1.5 Data-driven modeling and real-time forecasting . . . 13
1.5.1 Deep learning . . . 13
1.5.2 Strengths and challenges . . . 15
2 Research objectives and questions . . . 18
2.1 Research objective 1 . . . 18
2.2 Research objective 2 . . . 20
2.3 Research objective 3 . . . 21
2.4 Research objective 4 . . . 22
3 An ensemble data-driven approach for incorporating uncertainty in the forecasting of stormwater sewer surcharge . . . 24
4 A feature-informed data-driven approach for predicting maximum flood inundation extends . . . 53
5 A data-driven multi-step flood inundation forecast system . . . 72
6 A physically informed domain-independent data-driven inundation forecast model . . . 94
7 Outcome and further research . . . 114
Literature . . . 120
Appendix . . . 129
List of further publications . . . 129
List of conference presentations . . . 130
List of students supervisions . . . 131

Introduction

1 Introduction

1.1 Frequencies of flooding and flood damages

Climate change constitutes one of the most pressing environmental challenges of the 21st century, fundamentally altering the global climate system and associated hydrological processes. Anthropogenic emissions of greenhouse gases have driven a rapid increase in global mean surface temperatures since the mid-20th century (Calvin et al., 2023), with profound implications for the earth's water cycle. Warmer air has a greater capacity to hold moisture, approximately 7% more per degree Celsius of warming, according to the Clausius-Clapeyron relationship (Estrella & Menzel, 2013; Knutti et al., 2016; Tradowsky et al., 2023), thus increasing the potential for more frequent and intense precipitation events. Observational records and climate model simulations indicate that these hydrological changes are already occurring (Szymczak et al., 2022). The Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) states with very high confidence that the intensity and frequency of heavy precipitation events have increased in most land regions since the mid-20th century (Calvin et al., 2023). In particular, short-duration high-intensity rainfall associated with convective storms is becoming more common (Gründemann et al., 2022). This trend has been documented in multiple regions, including western and central Europe, where extreme daily and sub daily rainfall has intensified at rates consistent with or exceeding Clausius-Clapeyron scaling (Calvin et al., 2023; Fowler et al., 2021; Knutti et al., 2016).

This intensification of extreme precipitation leads to an increased risk of flooding, as both natural catchments and engineered drainage systems struggle to convey and store incoming water. As a result, flood events have become more frequent and severe across many regions (Calvin et al., 2023; Gründemann et al., 2022), posing significant threats to human settlements, critical infrastructure, and ecosystems. Floods now represent the most common and economically damaging natural hazard globally (Razavi et al., 2020), with consequences that extend well beyond immediate physical destruction, including disruptions to livelihoods, supply chains, and national economies. For example in 2021, the global economic losses caused by flooding amounted to USD 280 billion, the highest among all natural catastrophes (Swiss RE Institue, 2022).

A prominent and tragic example of these increasing flood risks occurred in July 2021 in Europe, when extreme precipitation affected Germany, Belgium, Luxembourg, and the Netherlands. Between the 12th and 15th of July, intense rainfall, estimated to have a return period of approximately 400 years (Kreienkamp et al., 2021), triggered catastrophic flooding, particularly within the Ahr, Erft, and Vesdre river basins. The event caused over 200 fatalities and extensive damage to infrastructure, including residential areas, transportation networks, and railway lines (Szymczak et al., 2022; Tradowsky et al., 2023). In many locations, communities were isolated for several days, severely hindering emergency response and recovery operations (Szymczak et al., 2022). Insured losses were estimated at approximately USD 13 billion, while the total economic impact is believed to have been substantially higher, at about USD 48 billion (Munich, 2022). The magnitude of the 2021 event, together with projections of increasing frequency and intensity of extreme precipitation, underscores the need for integrated flood risk management. This includes not only the implementation of protective infrastructure but also adaptive strategies aimed at reducing societal vulnerability, enhancing the resilience of populations, and mitigating economic, social, and environmental impacts in the coming decades.

1.2 Urban flooding

Urban areas are particularly vulnerable to flooding due to rapid urbanization and associated characteristics such as high population density, dense accumulation of buildings, and extensive impervious surfaces (Balaian et al., 2024; Razavi et al., 2020). This dense infrastructure reduces natural infiltration and increases runoff, while the arrangement and form of buildings can channel or impede floodwaters, amplifying local hazard. Consequently, urbanization not only increases exposure to floods but also modifies flow pathways, often intensifying flood impacts compared to rural or less-developed areas (Balaian et al., 2024). Moreover, in addition to the climate change, the urban heat island effect, where the densely impervious infrastructure and surfaces absorb and retain heat, can locally enhance atmospheric instability and convective activity, contributing to a higher frequency and intensity of extreme precipitation events (Huong & Pathirana, 2013; Y. Wu et al., 2021). This increases the flood risk in cities even further (Salinas Rodriguez et al., 2018).

In urban hydrology, flooding can originate from several sources, including rainfall, snow melt, and coastal processes such as storm surges or tidal inundations (Alves et al., 2024). While snow melt and coastal flooding are primarily driven by thermal or marine influences (Kirezci et al., 2020), rainfall-induced mechanisms dominate most urban environments. These include sewer surcharge flooding, fluvial flooding, as well as pluvial flooding (Bulti & Abebe, 2020). This dissertation focuses on these rainfall-induced flood types.

The urban drainage system (UDS) is engineered to convey stormwater and wastewater, providing protection from flooding and wastewater pollution (Fu et al., 2011). However, due to technical and economic constraints, it cannot accommodate all potential runoff volumes. Consequently, when precipitation exceeds the system's design capacity, surcharge events become unavoidable (Qin et al., 2013; Schmitt et al., 2004). A surcharge event occurs when the hydraulic head within the sewer pipes exceeds the crown level of a manholes, causing water from the UDS to flow back into the urban environment (Martins et al., 2018). This backflow results in flooding of streets, properties, and the other connected areas. Such surcharge events are particularly severe in combined sewer systems, where both stormwater and wastewater are conveyed in the same pipes, since beside of surcharged water volume, pathogens contained in the wastewater may be further released into the urban environment, posing additional public health risks (Leandro et al., 2022). Hence, the UDS plays a crucial role in urban flooding, as it can either mitigate or exacerbate events depending on its capacity and operational condition (Martins et al., 2018).

Pluvial flooding, is generated locally and often develops rapidly following short-duration, high-intensity precipitation events. It occurs when rainfall exceeds both the infiltration capacity of the soil and the conveyance capacity of the urban drainage system (UDS), leading to surface pounding and overland flow (Alves et al., 2024). In highly impervious urban areas, rainfall is quickly converted into runoff. When the stormwater cannot be managed effectively, water initially accumulates in streets and low-lying zones such as underpasses, before spreading to other surrounding areas. The severity and extent of pluvial flooding are influenced by rainfall intensity and duration, surface topography, land use, and the condition of drainage infrastructure (Alves et al., 2024; Bulti & Abebe, 2020). While surcharge events and pluvial flooding are distinct phenomena, they often occur simultaneously during intense precipitation events, compounding the risk of urban inundation (Martins et al., 2018). The localized nature and short response time make pluvial floods particularly challenging to predict and manage, posing significant risks to urban infrastructure, traffic safety, and public health.

Unlike pluvial flooding, fluvial flooding originates from rivers or streams when watercourses exceed their channel capacity due to long duration or intense rainfall over upstream catchment

areas (Alves et al., 2024). In urban environments located along river corridors, fluvial flooding can inundate extensive areas, particularly when natural floodplains are constrained by development or when river channels have been modified. The timing and magnitude of fluvial floods depend on catchment characteristics, antecedent soil moisture, and the spatial distribution of precipitation. Fluvial flooding involves catchment-scale hydrological processes and can affect larger downstream areas (Martins et al., 2018). Elevated river levels during such events can worsen urban flooding by restricting stormwater drainage, potentially triggering sewer surcharges and amplifying pluvial flooding (Martins et al., 2018).

1.3 Urban flooding modeling

Physically-based models are widely used for simulating urban floods and are considered state-of-the-art tools in flood modeling (Bulti & Abebe, 2020; Rözer et al., 2021). They rely on fundamental hydraulic equations, such as the shallow water equations (SWE) or their simplified forms, to represent the movement of water across surfaces and through drainage systems. These models provide a mechanically understanding of flood dynamics and can capture the complex interactions between precipitation, discharges, surface runoff, and urban drainage infrastructure, allowing for more accurate predictions of flood extent, depth, and timing.

UDS are typically simulated using one-dimensional hydraulic equations, such as the Saint-Venant equations, which describe the conservation of mass and momentum within the respective pipes and manholes (Leandro et al., 2011; Martins et al., 2018; Schmitt et al., 2004). The one-dimensional Saint-Venant equations can be expressed as:

Continuity equation (conservation of mass):

$$\frac{\partial A}{\partial t} + \frac{\partial (Au)}{\partial x} = 0, \tag{1}$$

Momentum equation (conservation of momentum):

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + g\frac{\partial h}{\partial x} + g(S_f - S_0) = 0, \tag{2}$$

where A is the cross-sectional area of flow [m^2], Q is the discharge [m^3/s], h is the water depth [m], x is the spatial coordinate along the conduit [m], t is time [s], u represents the flow velocity [m/s], S_0 is the bed slope, S_f is the friction slope, and g is the acceleration due to gravity [m/s^2].

Thus, these models capture the dynamic behavior of flow in sewer networks, including surcharges events through manholes. Widely used software such as EPA SWMM (Storm Water Management Model) implements these equations to simulate stormwater and combined sewer systems (Rossman, 2007; Rossman, 2016), allowing for the assessment of hydraulic capacity, surcharge events, and potential flood risk in urban areas.

Pluvial flooding is commonly simulated using two-dimensional surface flow models on fine-resolution grids (typically 1×1 m to 5×5 m). These models solve the SWE or simplified overland flow formulations using numerical schemes to simulate the propagation of water across the urban landscape (Leandro et al., 2014).

The two-dimensional SWE can be written in conservative form as follows:

Continuity equation (mass conservation):

$$\frac{\partial h}{\partial t} + \nabla \cdot (hu) = R, \tag{3}$$

Momentum equations (momentum conservation in x and y directions):

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u + \frac{\mu_t}{h}\nabla(h\nabla u) + g\nabla(h + z) = gS_f, \tag{4}$$

where h is the water depth, $u = [u_x \; u_y]^T$ is the depth-averaged velocity vector, g is the acceleration due to gravity, z is the bed elevation, μ_t is the turbulent eddy viscosity, R is the rainfall or external inflow term, and $S_f = [S_{f,x} \; S_{f,y}]^T$ is the bed friction vector.

This approach provides highly detailed spatial resolution for water depth calculations, which is essential in urban environments where flow is strongly influenced by underlying geometries (Balaian et al., 2024; Leandro et al., 2016). Most recent advantages allow these two-dimensional surface models to be coupled with one-dimensional sewer network models, capturing interactions between surface runoff and drainage infrastructure. Such coupling enables the simulation of scenarios where stormwater is partially conveyed by the sewer system while simultaneously accounting for surcharge events. This one-dimensional/two-dimensional coupled modeling approach is currently considered the most accurate representation of urban flood dynamics (Hammond et al., 2015; Leandro & Martins, 2016).

Fluvial flooding on the other hand is simulated by integrating catchment-scale hydrological models with hydraulic models of river channels. Rainfall-runoff models are used to estimate river discharge based on precipitation, soil moisture, and catchment characteristics (Alves et al., 2024). The resulting river discharge is then often applied as a boundary condition to a two-dimensional surface overland flow model, allowing the simulation of inundation across connected areas. In many applications, the river channel itself is modeled using a one-dimensional approach, while the floodplain is represented in two dimensions, enabling a detailed representation of flow dynamics where water overtops the banks and spreads into the terrain (Hammond et al., 2015). This approach captures both the longitudinal flow in the river and the lateral inundation across the floodplain, providing a more realistic representation of fluvial flooding in urban and rural environments.

Since these physically based hydrodynamic models, provide the most accurate representation of surface and subsurface water dynamics (Hammond et al., 2015; Leandro & Martins, 2016), they are therefore widely used in the design and assessment of structural flood mitigation measures such as levees, retention basins, and sewer networks. These structures are typically engineered to store or redirect water up to a defined design capacity; however, they are inherently limited and may fail when flows exceed this threshold (Schmitt et al., 2004). Furthermore, structural measures are fixed in both location and capacity, restricting their adaptability to changing boundary conditions, evolving urban morphologies, and intensifying extreme events. Consequently, while structural measures can substantially reduce flood risk, they cannot eliminate it entirely, highlighting the importance of complementary, flexible, and adaptive non-structural strategies, including flood forecasting, early warning systems, and operational flood management.

1.4 Problem statement: Urban flood forecasting and its limitations

Traditional urban flood forecasting relies also on these physically based numerical models simulating the hydraulic behavior of surface runoff and sewer flow. Such model-based management approaches could enable real-time control of drainage infrastructure and optimized operation of retention and storage facilities, thereby attenuating incoming discharge peaks (Jean et al., 2022; Xu et al., 2024). Although one-dimensional sewer models are, in principle, capable of real-time simulation, their computational performance declines significantly for large and complex sewer networks due to the high spatial and temporal resolution required (F.-J. Chang et al., 2014; Schmid & Leandro, 2023a). Moreover, explicit representation of uncertainty through ensemble simulations of inputted rainfall is often computationally infeasible (Schmid & Leandro, 2023a). As a result, despite their physical interpretability and proven reliability, traditional one-dimensional hydraulic solvers remain computationally demanding and poorly suited for operational real-time forecasting in large urban catchments.

Flood maps from coupled (sewer/overland flow models) or pure overland flow hydrodynamic models, providing spatially distributed inundation depth and extent. Such maps are essential for impact-based forecasting, supporting intelligent traffic management, targeted warnings, and optimized allocation of emergency resources (Crotti et al., 2019; Rözer et al., 2021). Nevertheless, their use in operational contexts remains limited due to excessive computational demands. Conventional numerical solvers for hydrodynamic simulations typically require sub-second time steps to ensure numerical stability and accuracy, resulting in simulation times that can extend to several hours even for a single event (Berkhahn & Neuweiler, 2024). To alleviate these limitations, recent research has leveraged graphical processing units (GPUs) to accelerate the numerical solution process. GPU-based implementations exploit the inherent parallelism of hydrodynamic equations, achieving computational speedups of up to an order of magnitude, approximately 15 times faster than CPU-based solvers (Hou et al., 2020; Morales-Hernández et al., 2021). Despite these advances, physically based models remain too computationally intensive for real-time flood forecasting applications, especially when uncertainty and ensemble predictions are required (Berkhahn, 2019; Berkhahn & Neuweiler, 2024; Henonin et al., 2013; Schmid & Leandro, 2023b).

To overcome the computational constraints many authorities have adopted pre-simulated flood scenarios to estimate inundation extent and depth in the context of forecasting. In operational practice, this often involves the use of a static map, generated for a predefined design return period (for example: a 100-year flood) (Berkhahn, 2019). Although such a map provides valuable insights, as the map is derived from synthetic boundary conditions, it cannot capture the spatio temporal dynamics of an actual flood event. Recognizing this limitation, an intermediate approach was developed that integrates a pre-simulated database of flood maps for different return periods and events enabling the selection of the most representative pre-simulated inundation scenario (Crotti et al., 2019). Nevertheless, the reliability of such frameworks depends heavily on the diversity and completeness of the pre-simulated database, as well as the representativeness of the selected scenario under changing hydrological and meteorological conditions. Consequently, these approaches can only approximate real flood dynamics, as the selected inundation map will never reflect the evolving situation, and temporal changes in flooding cannot be captured explicitly.

Given these limitations, particularly in the context of urban flooding, where rainfall patterns or incoming discharges evolve rapidly and boundary conditions must be continuously updated, there is a growing need for modeling frameworks capable of producing time-dependent, dynamically adaptive forecasts (Berkhahn & Neuweiler, 2024; Schmid & Leandro, 2024). Recent research has

therefore increasingly focused on developing computationally efficient and flexible forecasting approaches that enable real-time prediction and decision support.

1.5 Data-driven modeling and real-time forecasting

In recent years, data-driven modeling has become a key part of modern scientific and engineering applications. Instead of explicitly solving physical laws or empirical equations, data-driven approaches learn patterns, relationships, and dependencies directly from data using specified algorithms (Géron, 2018; Solomatine et al., 2008). These models have been successfully applied across diverse domains and research fields, including robotics (Z. Li et al., 2022), climate prediction (Asanjan), medical diagnostics (Ronneberger et al., 2015), image or video generation, and environmental modeling (Lin, Leandro, Gerber, & Disse, 2020; Thapa et al., 2020), owing to their ability to capture complex nonlinear behaviors that are difficult to express analytically.

In the context of urban flood forecasting, data-driven models have emerged as computationally efficient alternatives to the physically based hydrodynamic models (Bentivoglio et al., 2022; X. Li et al., 2010). While the traditional models explicitly solve the governing hydrodynamic equations, data-driven approaches approximate the underlying hydrological and hydraulic processes through supervised learning, where the model is trained on historical input–output pairs. Inputs may include rainfall or upstream discharges x_t and catchment properties c (e.g., slope, land use, imperviousness) while outputs y_t typically represent water depths or downstream discharges.

Formally, a data-driven model approximates an unknown mapping function:

$$y_t = f_\theta(x_t, x_{t-1}, \ldots, x_{t-n}, c) + \varepsilon_t, \tag{5}$$

where $f_\theta(\cdot)$ is a nonlinear function parameterized by learnable weights θ, and ε_t represents the residual modeling error (Géron, 2018).

This mapping from inputs to outputs represents the central principle of data-driven modeling. The model parameters θ are learned iteratively during the so-called training by minimizing a loss function $\mathcal{L}$, which quantifies the difference between observed (y_t^{obs}) and predicted (y_t^{pred}) outputs. A common choice is the Mean Squared Error (MSE):

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left\| y_t^{\text{obs}} - y_t^{\text{pred}} \right\|^2 \tag{6}$$

To minimize the loss function, optimization algorithms such as stochastic gradient descent (SGD) or its variants (e.g., Adams algorithm or the Root Mean Squared Propagation) are employed (Kingma & Ba, 2017; Wan, 2019). This described training approach (minimizing the error) is called supervised learning. The model is trained using these input-output pairs (x_t, y_t^{obs}), where the observed outputs provide a reference for learning the mapping. Once trained, data-driven models can generate predictions on the order of milliseconds to seconds per simulation, enabling real-time forecasting and efficient ensemble simulations (Géron, 2018).

1.5.1 Deep learning

In addition, deep learning has further advanced the field of data-driven modeling by enabling automatic extraction of hierarchical and highly nonlinear features from large datasets. Deep learning models are composed of multiple interconnected processing layers, allowing them to learn

increasingly abstract representations of the input data (Géron, 2018; Z. Li et al., 2022). This multi-level feature learning capability makes deep neural networks particularly effective for complex tasks, such as, sequence modeling and spatiotemporal prediction. As a result, deep learning has become a central methodology in modern data-driven analysis, offering unprecedented predictive accuracy and flexibility (Géron, 2018).

Artificial Neural Networks (ANNs) are among the earliest and most widely applied data driven/deep learning techniques in flood forecasting and real-time modeling by mainly focusing on the prediction of the expected discharges (Mosavi et al., 2018; Rezaeianzadeh et al., 2014). They consist of interconnected layers of neurons that transform input signals through nonlinear activation functions. Each neuron in a hidden layer computes a weighted sum of its inputs:

$$h_j = \sigma\left(\sum_{i=1}^{m} w_{ij} x_i + b_j\right), \tag{7}$$

where x_i are input features, w_{ij} are learnable weights, b_j is a bias term, and $\sigma(\cdot)$ is a nonlinear activation function (e.g., Rectified Linear Unit (ReLU), Sigmoid, or Tangent hyperbolic ($\tanh$)), which is used to represent the relevance of the respective output. The output layer then maps the final hidden representation h to the target variable y:

$$y = g_\phi(h) = \phi\left(\sum_j v_j h_j + b\right), \tag{8}$$

where v_j and b are output weights and bias, respectively, and $\phi(\cdot)$ is typically a linear or bounded activation function (Solomatine et al., 2008). The limitation of ANNs, however, lies in their lack of temporal memory, as each prediction depends solely on the current input and not on preceding time steps.

To address temporal dependencies, Recurrent Neural Networks (RNNs) extend ANNs by incorporating recurrent connections that retain information from previous time steps. At each time t, the hidden state h_t is computed as

$$h_t = \sigma_h\left(w_x x_t + w_h h_{t-1} + b_h\right), \tag{9}$$

and the output as

$$f y_t = \sigma_y\left(w_y h_t + b_y\right), \tag{10}$$

where w_x, w_h, and w_y are weight matrices, and b_h and b_y are bias terms (Sherstinsky, 2020).

This recurrent formulation allows the model to maintain a temporal memory of past inputs, enabling the prediction of dynamic processes such as runoff or water depth accumulation, delayed flow responses, and storage effects. In particular, the Long Short-Term Memory (LSTM) networks, an advanced form of RNN designed to mitigate vanishing-gradient issues, are widely used in flood forecasting for predicting one-dimensional water levels or river discharges due to their ability to capture long-range temporal dependencies in hydrological time series (Kratzert et al., 2018; Kratzert et al., 2024; Nagesh Kumar et al., 2004).

While ANNs and RNNs are effective for learning nonlinear relationships and temporal dependencies, respectively, they are limited in their ability to represent spatially distributed processes. Flood inundation patterns depend not only on local conditions but also on the interaction of neighboring cells, such as water movement driven by topography, land cover, or local flow connectivity. Neither standard ANNs nor classical RNNs explicitly model these spatial correlations, and their

fully connected structures become computationally inefficient for high-resolution flood maps, especially during the training face (C. Chen et al., 2021).

To address these limitations, Convolutional Neural Networks (CNNs) have become a central architecture in flood mapping and inundation modeling. Originally developed for image recognition, CNNs are designed to capture spatial patterns through localized convolution operations. A CNN applies a set of learnable filters K that convolve over an input image (e.g., rainfall intensity or digital elevation):

$$h_{i,j} = \sigma\left(\sum_{m,n} K_{m,n} \cdot x_{i+m,\,j+n} + b\right), \tag{11}$$

where $\sigma(\cdot)$ denotes a nonlinear activation function, K represents the convolutional kernel, and b is a bias term (Z. Li et al., 2022). This operation captures local spatial correlations, such as how water flow is influenced by neighboring cells' elevation or rainfall intensity. By stacking multiple convolutional layers, CNNs learn hierarchical spatial features that can describe the structure and propagation of inundation patterns, making them well suited for real-time inundation forecasting (Bentivoglio et al., 2022; Z. Guo et al., 2021; Kabir et al., 2020; Schmid & Leandro, 2023b) and flood susceptibility mapping (Seleem et al., 2022; Towfiqul Islam et al., 2021; Zhao et al., 2020).

Following standard data-driven learning practice, the overall modeling workflow typically includes dataset preparation, training, validation, and testing. Training datasets may consist of historical observations or simulation outputs, while independent validation and testing sets are used to monitor generalization performance. The skill of the model is commonly assessed using established evaluation metrics such as the Root Mean Squared Error, Mean Absolute Error, or the Critical Success Index, providing quantitative measures of predictive accuracy and robustness.

In recent years, these architectures have been further adapted and extended, giving rise to more advanced deep learning paradigms. Notable developments include attention-based models such as Transformers (Vaswani et al., 2023), generative frameworks like Generative Adversarial Networks (Goodfellow et al., 2014), and probabilistic generative approaches such as diffusion models (Ho et al., 2020). These methods offer powerful mechanisms for capturing long-range dependencies, generating high-fidelity spatial fields, and modeling complex uncertainty structures. Despite their rapid development, these models largely originate from fields such as natural language processing, computer vision, and image synthesis, and have therefore only recently appeared in the field of hydrology, hydrodynamics and flood modeling. Their application in the field of flood prediction and mapping, and their suitability for operational forecasting is under active investigation and research.

1.5.2 Strengths and challenges

In the broader context of supervised learning, which has so far been the predominant paradigm in flood forecasting, such data-driven models rely on paired inputs-outputs to learn the underlying mapping between forcing variables (rainfall or discharges) and hydrodynamic response variables (water levels or inundation depths). On the one hand, this dependence on labeled data offers a clear advantage: when high-quality measurements or outputs from physically based hydrodynamic models are available, supervised deep learning models, including CNNs, RNNs, LSTMs, ANNs can approximate flood dynamics with high accuracy while providing orders-of-magnitude speed-ups compared to numerical solvers. This enables real-time forecasting and allows the generation of large ensembles that would be infeasible with traditional models (Bentivoglio et al., 2022; Schmid & Leandro, 2023a; Schmid & Leandro, 2024).

On the other hand, it also introduces challenges, as preparing suitable training datasets often requires extensive pre-simulations, measurement campaigns, or data collection, which can be time-consuming and computationally demanding (Schmid et al., 2026). Moreover, measurement campaigns are not always as beneficial as intended, since extreme events rarely occur during these periods, and for inundation modeling in particular, point-based measurements provide little value because spatially distributed information is required. As a result, data-driven models must still rely heavily on simulated scenarios to capture the full range of hydrological extremes and their spatial dynamics. Additionally, the scalability of such systems remains limited: as catchment size, spatial resolution, and input complexity increase, the required training data and computational resources grow disproportionately. A further challenge is transferability, as models trained on one catchment fail to generalize to others. Since they primarily learn catchment-specific patterns embedded in the training data, their predictive skill does not readily extend to areas with different hydrological, topographical, or infrastructural characteristics. Consequently, substantial additional data generation, retraining, or model adaptation is required for each new catchment. In addition, classical data-driven models typically neglect the underlying physical laws that govern hydrodynamic processes. Since these models learn statistical relationships solely from input–output pairs, they operate as black-box systems, providing predictions without explicit guarantees of physical consistency—such as mass conservation, energy balance, or realistic flow routing (Berkhahn & Neuweiler, 2024; Z. Guo et al., 2021; Schmid & Leandro, 2024). This can enable strong performance when training data are abundant and representative, however, it also limits explanation and extrapolation capabilities during extreme or previously unseen events. Thus, the performance of data-driven models depends strongly on the quantity, quality, and representativeness of the training data and strategy, as well as on the design of the network architecture. Hence, assembling such datasets and configuring robust models remains a challenging task.

To conclude, data-driven flood modeling offers significant advantages, but also comes with notable challenges. The key points can be summarized as follows:

- Benefits: Capture complex nonlinear hydrodynamic processes, enable real-time predictions, and integrate diverse data sources efficiently.
- Computation and real-time prediction: Provide orders-of-magnitude faster simulations than physically based models, allowing rapid ensemble generation and operational early warning.
- Data and scalability challenges: Require large, high-quality, and spatially distributed training datasets; high-resolution models increase computational and storage demands.
- Transferability limitations: Models are often catchment-specific and do not generalize well, requiring retraining or data generation for new regions.

Research objectives and questions

2 Research objectives and questions

Building on this background, the overall objective of this dissertation is the development of real-time, data-driven models for the respective types of urban flooding (sewer surcharge, fluvial and pluvial flooding), providing computationally efficient alternatives to traditional physically based models. The dissertation is presented in cumulative form, comprising four scientific studies that have been published in international peer-reviewed journals. Each study deals with specific research questions and has undergone rigorous evaluation, including independent peer-review by three external experts.

Furthermore, this dissertation addresses the key challenges inherent to data-driven modeling, providing insights into its limitations, particularly with respect to scalability and transferability. Through these studies, valuable knowledge is gained not only on the pure data-driven prediction but also on the current limitations of data-driven models and potential strategies to overcome them. The following sections introduce each research subject and its corresponding questions, followed by chapters presenting the original published studies. A comprehensive synthesis and critical evaluation of the findings, including reflections on model limitations and avenues for improvement, is provided in the outcome and further research chapter of the dissertation.

2.1 Research objective 1

As already mentioned, the urban drainage system (UDS) plays a crucial role in urban flooding since it can mitigate or enhance flooding (Leandro & Martins, 2016; Martins et al., 2018). Since it is engineered to convey stormwater and wastewater and due to technical and economic constraints, surcharge events (where the hydraulic head within sewer pipes exceeds the crown level of a manhole) are unavoidable (Schmitt et al., 2004). Thus, accurate forecasting of surcharge events could support early warnings and operational decision-making.

State-of-the-art sewer modeling is predominantly carried out using the one-dimensional physically based hydrodynamic models such as EPA SWMM (Storm Water Management Model) (Rossman, 2007; Rossman, 2016). When sewer infrastructure is designed to support operational flexibility, these models can be integrated into Real-Time Control (RTC) or Model-based Control (MBC) frameworks to enhance system performance. RTC has become an essential strategy for mitigating sewer surcharge, combined sewer overflows, and localized flooding in densely urbanized catchments (García et al., 2015; Jean et al., 2022). Traditional operational strategies are typically based on static rule sets, such as fixed weir configurations or predefined pump activation schedules, which lack the ability to respond to rapidly evolving hydrometeorological conditions. More modern RTC approaches overcome this limitation by coupling the hydrodynamic models with real-time sensor data, optimization routines, and short-term forecasting algorithms, moving to MBC. This integration enables dynamic system operation through actions such as adjusting gate positions, modulating pump speeds, utilizing in-line and off-line storage tanks, or redistributing flows across interconnected subcatchments (García et al., 2015). However, the effectiveness of MBC systems is fundamentally constrained by the computational burden of the underlying hydrodynamic model and the quality of the short-term forecasts on which they depend (Jean et al., 2022). Although one-dimensional hydrodynamic models are, in principle, capable of real-time computation, their performance significantly drops for large and highly interconnected UDSs, where model runtimes can prohibit the rapid update cycles required for operational control (F.-J. Chang et al., 2014). More-

over, rainfall typically is a driving force of sewer surcharge, it is subject to substantial uncertainty. Hence, the hydrodynamic simulations must be rerun frequently to update control actions.

In addition, input, parameter, and model structural uncertainty are difficult to quantify and propagate within classical physically based models during real-time operation. Classical ensemble-based approaches quantify these uncertainties by generating multiple equally plausible members realizations through perturbations of inputs or parameter sets, or by sampling within likelihood-based frameworks such as Generalized Likelihood Uncertainty Estimation (GLUE) providing a probabilistic forecast (Leandro et al., 2019). Providing multiple outputs, decision-making can consider explicitly the likelihood of the associated event. However, the computational expense of physically based models limits this approach.

These limitations have motivated research into data-driven approaches for sewer surcharge prediction. Early studies employed artificial neural networks (ANNs) to predict water levels in sewer systems, where data-driven models were trained on rainfall time series to forecast temporal water-level dynamics in sewer chambers, demonstrating notable improvements in operational performance (Fernando et al., 2007; Mounce et al., 2014). Other studies increasingly adopted recurrent neural networks (RNNs), which are better suited for time-series prediction due to their ability to retain temporal information through internal memory states. RNN-based approaches have been successfully used to forecast water levels at pumping stations during storm events (F.-J. Chang et al., 2014; Chiang et al., 2010) and to predict surcharge conditions at selected critical manholes in complex urban drainage systems (Duncan et al., 2013; Rjeily et al., 2017). These studies either used pre-hydrodynamic simulations for training or had measurement from the drainage system available to directly train on measured data.

Building on the background presented, the first research subject of this dissertation concerns the prediction of sewer surcharge events. The objective is to develop an ensemble forecasting system capable of predicting the onset and duration of stormwater sewer surcharge events in real time. The proposed approach employs an ensemble of nonlinear autoregressive models with exogenous inputs (NARX) neural networks, enabling rapid forecasting while explicitly accounting for structural and parameter uncertainties. The NARX network is a specialized type of recurrent neural network that incorporates feedback connections from the output back to the input layer (Duncan et al., 2013; Rjeily et al., 2017; Wunsch et al., 2018). This structure allows for teacher forcing during training, where the network is optimized using the true target values rather than its own predictions, and facilitates efficient parameter optimization through well-established gradient-based algorithms (Sherstinsky, 2020). The following research questions guide this investigation:

- How can urban sewer surcharge events be predicted in real time, accounting for the nonlinear and time-dependent dynamics of urban drainage systems using NARX-based data-driven models?
- How can uncertainties arising from model structure and parameters be quantified and incorporated into real-time forecasts to support operational decision-making?

These research questions are addressed in the study titled, *"An Ensemble Data-Driven Approach for Incorporating Uncertainty in the Forecasting of Stormwater Sewer Surcharge"*. It is published in **Urban Water Journal** (https://doi.org/10.1080/1573062X.2023.2240309) and presented in Chapter 3. Due to copyright restrictions, the manuscript version submitted to and accepted by the journal is included in this dissertation.

2.2 Research objective 2

Advancing from the localized short-term sewer surcharge event predictions, the second research subject of this dissertation focuses on flood inundation forecasting at the scale of entire catchments. Inundation maps, which show spatially distributed water depths allow for supporting intelligent traffic management, localized specific warnings, and optimized allocation of emergency resources (Crotti et al., 2019; Rözer et al., 2021). The generation of these inundation maps within real-time operational forecasting is, as previously discussed, computationally infeasible when relying solely on traditional physically based hydrodynamic models. Consequently, data-driven approaches offer a promising alternative. A data-driven model that has been thoroughly trained on pre-simulated hydrodynamic scenarios can effectively replace the full hydrodynamic model during operation and produce inundation maps in real time. Such event-specific predictions provide substantially greater value than static inundation maps or selecting the closest matching pre-simulated scenario from a database (Crotti et al., 2019), as the data-driven model learns the catchment-specific hydraulic behavior and can therefore generate a tailored inundation map for each new event.

Recent advances in flood inundation forecasting have demonstrated the strong potential of data-driven models, particularly artificial neural networks (ANNs) (Berkhahn, 2019; L.-C. Chang et al., 2014; Lin, Leandro, Wu, et al., 2020), convolutional neural networks (CNNs) (C. Chen et al., 2021; Kimura et al., 2019), and specialized generative CNN architectures (D. Chen et al., 2020; Hofmann & Schüttrumpf, 2021), to approximate computationally intensive hydrodynamic simulations. Further improvements in prediction accuracy have been achieved by augmenting the neural networks with physically and geographically relevant information, such as high-resolution Geographic Information System (GIS) datasets (e.g. elevation, land use, or flow directions) enabling models to better reflect local topographic and hydrologic conditions (Burrichter et al., 2023; Z. Guo et al., 2021).

Nevertheless, existing data-driven approaches often require discretizing the study area into multiple sub-models, either because CNN-based methods operate on image-like spatial grids or because ANN architectures must be trained separately for individual locations. This discretizing substantially increases computational complexity, limits scalability, and complicates model training and hyperparameter optimization. Furthermore, methods that incorporate detailed physical or geographical information typically depend on large volumes of high-resolution input data and substantial computational resources, making them impractical for many local authorities that lack access to high-performance computing infrastructure.

Therefore, the second research objective of this dissertation addresses these challenges by developing a data-driven inundation forecasting model capable of operating at the scale of an entire catchment while incorporating supplementary GIS-based physical information. A key requirement is that the model remains computationally tractable. The overarching aim is to design an architecture that integrates physically meaningful features into the learning process without resorting to spatial subdivision of the domain, thereby enabling end-to-end prediction for the full catchment. The research is guided by the following questions:

- How can real-time flood inundation forecasting be achieved for entire study areas without the need for spatial discretization into multiple sub-models?
- How can prior physical knowledge be effectively integrated into the learning process without requiring extensive additional GIS input layers as direct input?

- How can the forecasting framework be designed to remain computationally feasible on standard hardware commonly available to local authorities?

These questions are addressed in Chapter 4, which presents the study published as *"A Feature-Informed Data-Driven Approach for Predicting Maximum Flood Inundation Extends"* in **Geosciences** (https://doi.org/10.3390/geosciences13120384).

2.3 Research objective 3

While the research on maximum inundation prediction (Schmid & Leandro, 2023b) has demonstrated the potential of data-driven models to capture entire catchments and advance scalability, static maximum inundation maps fail to represent the temporal evolution of flooding. However, operational early-warning systems require forecasts that not only delineate the maximum extent but also capture how inundation develops over time. The temporal dynamics of flood propagation are critical, as the timing of water arrival directly affects emergency response, evacuation planning, and operational decision-making. To address this limitation, the third research subject of this dissertation investigates how the existing data-driven inundation modeling framework can be extended to generate fully spatiotemporal flood forecasts. The goal is to enable predictions that are both temporally and spatially resolved, thereby providing a more comprehensive basis for real-time flood risk management.

Previous studies have employed either separate networks for individual timesteps (Lin, Leandro, Gerber, & Disse, 2020) or a sequential training strategy (Burrichter et al., 2023; Cao et al., 2025; Pianforini et al., 2025; Shi et al., 2015) in which a single network predicts multiple consecutive maps corresponding to targeted timesteps. Both approaches achieve temporally and spatially resolved predictions, producing dynamic inundation maps. However, networks trained for separate timesteps are fixed to their respective steps and lack flexibility, requiring multiple networks to be trained for a single event. In contrast, the sequence-based approach is more effective, as it allows the network to implicitly learn temporal error propagation by observing intermediate steps up to the final prediction. Nonetheless, this method can be less flexible when intermediate updates are needed (Berkhahn & Neuweiler, 2024) and can become computationally demanding when applied to entire catchments.

Building on these insights, the current network training strategy (Schmid & Leandro, 2023b) is adjusted by the knowledge gained from the sewer surcharge data-driven architecture (Schmid & Leandro, 2023a) and analogous convolutional approaches (Berkhahn & Neuweiler, 2024), resulting in a recursive model capable of representing the entire catchment while producing temporally consistent predictions. Accordingly, the study is framed around the following research questions:

- How can the existing data-driven model be adapted to provide multi-step forecasts that capture the temporal dynamics of flooding?
- How can such dynamic forecasts be designed for real-time operational use while retaining the developed computational efficiency?

These questions are answered in Chapter 5, which presents the study published as *"A Data-Driven Multi-Step Flood Inundation Forecast System"* in **forecasting** (https://doi.org/10.3390/forecast6030039).

2.4 Research objective 4

Despite the promising progress of the previous studies regarding the implementation and scalability of data-driven flood forecasting approaches, a major limitation remains their strong domain dependence. Since these models typically learn only the hydraulic behavior represented in their training datasets, their applicability is restricted to the specific catchment or urban area for which they were developed. Setting up a data-driven forecast system in a new catchment typically requires generating an extensive set of physically based hydraulic simulations and retraining the model from scratch. This process is computationally demanding, time-consuming, and requires substantial storage capacity. Moreover, most existing approaches disregard underlying physical laws and instead learn purely data-driven patterns, rendering the models essentially black boxes with limited interpretability. This lack of transferability and explainability represents a major barrier to operational deployment, where flexible, robust, and catchment-independent forecasting systems are essential. These challenges are addressed in the fourth research subject of this dissertation, investigated in the context of pluvial flooding.

Some first studies achieved success in generalizing by incorporating rainfall and catchment characteristics of multiple locations or catchments (Do Lago et al., 2023; Z. Guo et al., 2022; Löwe et al., 2021; Seleem et al., 2023). However, these approaches face substantial challenges when applied to large catchments, as their CNN-based architectures require spatial inputs to be represented as fixed-size images. Large catchments often exceed the feasible input dimensions of standard CNN models, forcing downscaling. Moreover, they provide maximum inundation maps, without resolving the temporal evolution of flooding, an aspect that is critical for operational early-warning systems, as previously discussed. Additionally, even when models are trained on multiple catchments and supplemented with geospatial information describing catchment-specific physical characteristics, they may still neglect fundamental physical laws because their learning process is driven purely by data. This limits their physical consistency and constrains their reliability in real-world applications.

To enhance physical consistency, physics-informed neural networks (PINNs) have been developed. Classical PINNs (Raissi et al., 2019) embed the governing partial differential equations (PDEs) directly into the loss function and enforce their residuals at collocation points in space and time. This constrains the model to satisfy physical laws while learning from data. Several studies have applied PINNs to hydrodynamics, incorporated the one-dimensional Saint-Venant equations into the loss function (Feng et al., 2023; Mahesh et al., 2022), yielding promising results but restricted to one-dimensional flows. Recently a two-dimensional shallow-water-equation–governed PINN has been developed that outperformed state-of-the-art data-driven models (Donnelly et al., 2024). However this one is not designed for domain-independent prediction or generalization to unseen areas. Further a key limitation of classical PINNs is their need for all dependent variables and derivatives (e.g., water depth and velocities), which becomes computationally demanding.

Building on these limitations and knowledge, the fourth study of this dissertation develops a physically informed, domain-independent, data-driven forecast system for pluvial flooding. The proposed framework is based on a convolutional neural network trained on geospatial and temporal inputs and incorporates physical principles through a mass-conservation loss derived from the continuity equation under the kinematic-wave assumption. This loss is evaluated over the full prediction grid and does not require explicit velocity fields or collocation points, enabling physically consistent predictions with minimal additional information.

The study is guided by the following research questions:

- How can a data-driven flood inundation forecast system be designed to provide physically consistent predictions?
- How can the forecast system achieve domain-independent predictions, enabling generalization to previously unseen areas?
- How can the system provide temporally resolved, inundation maps in real time while remaining computationally efficient for large-scale areas?

These research questions are addressed in Chapter 6, which presents the study published as *"A Physically Informed, Domain-Independent Data-Driven Inundation Forecast Model"* in **Water Research** (https://doi.org/10.1016/j.watres.2025.124819).

An ensemble data-driven approach for incorporating uncertainty in the forecasting of stormwater sewer surcharge

An ensemble data-driven approach for incorporating uncertainty in the forecasting of stormwater sewer surcharge

Felix Schmid[α*] and Jorge Leandro[α]

[α] Chair of Hydromechanics and Hydraulic Engineering, Research Institute Water and Environment, University of Siegen, 57076 Siegen, Germany

Corresponding Author: Felix.Schmid@uni-siegen.de

Abstract

Stormwater surcharge events are unavoidable above a certain rainfall intensity. Thus, for protection and damage mitigation, forecast systems are of outstanding importance. This study develops an ensemble forecast system (EFS) to predict the beginning and end of sewer surcharge events. It applies a nonlinear autoregressive with exogenous inputs (NARX) network to each member of the ensemble, making it suitable for real-time predictions. The fundamental idea is the forecast of water depth time series within manholes based on the given rainfall. The novelty lies in consideration of uncertainty through the ensemble structure for which the numbers of neurons in the hidden layer, the weights, and biases are considered to be uncertain. Results are evaluated based on observed values captured within the uncertainty band (P_{factor}), and the width of the band (R_{factor}). The P_{factor} varied between 74% and 94% and the R_{factor} between 1.36 and 10.68.

Keywords: Urban drainage systems; Surcharge; Ensemble flood forecasting; Nonlinear autoregressive network with exogenous inputs (NARX); Neural networks

Subject classification codes: Artificial intelligence, Flood forecasting, Urban flooding, Drainage, Uncertainty

1 Introduction

Floods are one of the most common natural disasters posing a threat to human lives, properties, and the environment itself. Especially in urban areas, the large percentage of imperviousness and high precipitation intensities result in large surface runoff volumes (Henonin et al. 2013; F.-J. Chang et al. 2014; L.-C. Chang et al. 2018). The recent events from July 2021 in West Germany, with billions of euros in property damage and even an unexpectedly high number of fatalities, have again demonstrated the immense damage potential of such extreme rainfall events. Intensities of this magnitude exceed the capacity of structural protective measures, highlighting the need for improving our preparedness with fast and accurate forecast systems. Knowing when a protective measure is at risk of failure or evacuation needs to be started is critical for protecting the population and mitigating the damage.

The urban drainage system (UDS) is designed to convey waste- and stormwater and, thus, to control and protect urban areas from flooding and wastewater pollution (Fu et al. 2011; Qin, Li, and Fu 2013). From a technical and economic point of view, the system cannot be built to collect the total amount of runoff, meaning flooding is unavoidable above a certain design protection level (Schmitt, Thomas, and Ettrich 2004). At the beginning of a storm event higher than the design protection level surcharge is unlikely to occur. However, as the event progresses, the pipe flow may become pressurized (Djordjević et al. 2005). If the hydraulic head becomes higher than the water level atop the sewer manholes (or the crown level of the manholes in case of non-existent overland flow) surcharge will occur flooding the overland surface (Rubinato, Martins, and Shucksmith 2018). Especially during these surcharge events, significant adverse economic, physical, environmental, and social impacts are expected, particularly in the case of combined sewer systems (Fu et al. 2011; Rjeily et al. 2017; Mark et al. 2018). Knowing when and where the UDS is at risk of surcharging is essential information for reducing and mitigating the damage arising.

That is the reason why this study focuses on three main challenges: The need for accurate forecast systems to prevent and mitigate damage, the high computational requirement in order to provide a sufficient lead time for warning and countermeasures, and the consideration of uncertainty to facilitate decision-making.

The first step is to provide a forecast of where and when a surcharge will occur. In the past decade, one-dimensional numerical models became the state-of-the-art of simulation hydraulic behavior in the UDS (Hsu, Chen, and Chang 2000; Mark et al. 2004; Leandro et al. 2011; F.-J. Chang et al. 2014). There are several software/models available which are capable of simulating the hydrological and hydrodynamic behavior of the UDS in a satisfactory way, such as the Storm Water Management Model (SWMM) or Infoworks (Martins, Leandro, and Djordjević 2018). A simulation-oriented approach for forecasting is simply to run the hydrodynamic model of the UDS to simulate the response of the incoming rainfall event.

However, a simulation-oriented approach is sometimes and especially in large UDSs, not suitable for real-time or short-term predictions due to the long computational times (Duncan et al. 2013; Chiang et al. 2010; F.-J. Chang et al. 2014; L.-C. Chang et al. 2018). Having enough time to prepare mitigation measurements and reduce the damage is crucial. Thus, alternative approaches such as data-driven models like neural networks have gained special attention among the research community, which is also the primary motivation of this study.

Neural networks are computationally more efficient than conventional deterministic physical-based models. In recent studies, neural networks have been used to upgrade the physical model (Bermúdez et al. 2018; Li 2020) to generate inundation maps for rainfall or discharge (L.-C. Chang et al. 2018; Berkhahn, Fuchs, and Neuweiler 2019; Lin et al. 2020) or to map urban flood susceptibility

(Falah et al. 2019). In the field of UDS, Mounce et al. (2014) applied neural networks to predict water depth in sewer chambers using rainfall time series, thereby enhancing drainage system performance. Chiang et al. (2010) and F.-J. Chang et al. (2014) applied recurrent neural networks to predict the water level at pumping stations for given storm events. The results from F.-J. Chang et al. (2014) showed that especially the nonlinear autoregressive with exogenous input (NARX) algorithm outperformed conventional artificial networks due to its unique feedback connections. Duncan et al. (2013) and Rjeily et al. (2017) showed similar results also for more complex UDS.

Besides the need for accurate and computationally efficient forecasts, uncertainties also need to be quantified. Multiple sources of uncertainty, such as input-, model structure- or parameter uncertainty, affect the simulation and the forecasts. Since the transformation from rainfall to runoff or inundation is a complex task, structural model- and parameter uncertainty should be included in the forecast (Dietrich et al. 2009). Model structure refers to a specific mathematical formulation (normally applied in a numerical model) used to describe a phenomenon (e.g., flooding). Model structure uncertainty refers to the unexplained variability in the outputs of a model describing a phenomenon that arises from the choice of different mathematical formulations. For example, Quintana-Romero and Leandro (2022) developed a methodology for deriving different flood model structures to obtain probabilistic flood uncertainty maps. Another approach while communicating and accounting for prediction uncertainty is the use of ensemble forecasts.

These forecasts aim to capture the uncertainties of the possible future development based on different multiple equally likely members of the ensemble. The members are often generated by small changes in the inputs or model parameters, for example, by applying perturbation (Leandro, Leitão, and de Lima 2013) or selecting members following the Generalized Likelihood Uncertainty Estimation (GLUE) framework from Beven and Binley (1992). This methodology is used to quantify uncertainty while assuming equality among different parameter sets incorporated as members of the ensemble via a predefined measure of likelihood (Beven and Binley 1992). Leandro et al. (2019), for example, used a Nash-Sutcliffe Efficiency (NSE) value larger than 0.3 for incorporating members into the ensemble. By delivering multiple outputs, decision-making can consider explicitly the likelihood associated with the model outcomes and consider events with lower probability, unlike deterministic approaches (Leandro, Leitão, and de Lima 2013). This strategy, in combination with the more efficient computational times of neural networks, allows the consideration of uncertainty in real-time and offers, therefore, a major advantage over physically based oriented approaches.

This study draws on these three challenges and focuses on developing an ensemble forecast system (EFS) for predicting the beginning and end of stormwater sewer surcharge. The EFS will use an ensemble of nonlinear autoregressive with exogenous input (NARX) neural networks to enable real-time forecasts and consider structural and parameter uncertainty. Due to the time dependency of this task, a neural network that features some dynamic feedback connections or memory effects is necessary. Recurrent neural networks or a long short-term memory (LSTM) network are suitable for this kind of task. The NARX network is particularly selected because of its unique architecture and training strategy. The recurrent connection is given by the output itself rather than from the internal stages of the network. This allows the use of well-established backpropagations algorithms (Bianchi et al. 2017), and it accounts for time dependencies, making it useful for hydrological modeling and flood forecasting (F.-J. Chang et al. 2014; Rjeily et al. 2017; Abbas et al. 2018; L.-C. Chang et al. 2018; Wunsch, Liesch, and Broda 2018).

2 Methods

2.1 Framework of the ensemble forecast system (EFS)

The aim is to build an EFS able to predict the beginning- and end-time of sewer surcharge. This is achieved by the ensemble prediction of the water depth time series in M_n manholes caused by a given rainfall within a time L_t. Based on this ensemble, the starting- and end-times of a surcharge event can be forecasted with a corresponding confidence interval.

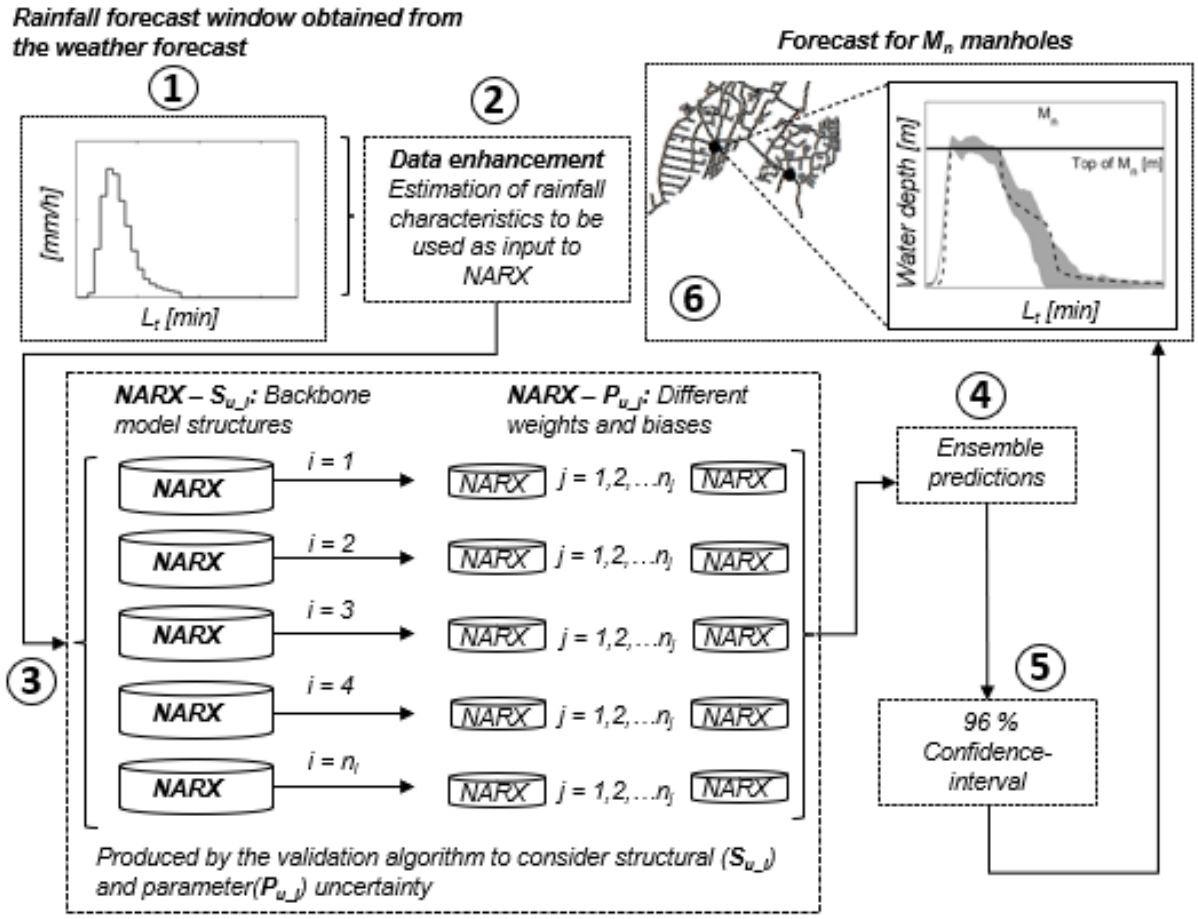


Figure 1: The architecture of the ensemble forecast system (EFS). The rainfall within the forecast window with the time L_t is taken after a data enhancement process as input to different NARX structures ($\mathrm{NARX} - S_{u_i}$), followed by subordinate NARX networks ($\mathrm{NARX} - P_{u_j}$) with different weights and biases. The final forecast is done by applying a probability distribution to the predictions of the ensemble.

The architecture of the EFS is presented in Figure 1. Six different steps are processed to generate the output confidence interval for the respective manholes.

1. The forecast window (obtained from weather forecasts) is a rainfall-intensity time series with a flexible length of L_t.

2. Neural networks learn the relationship between input and output data sets. To ensure a unique and clearly learnable relationship, information describing this relationship should

be provided to the network. The precipitation event in the forecasting window must have a unique characteristic for each time step to be well predicted. Therefore, the procedure of Duncan et al. (2013) is used to define the rainfall characteristics in the input vector. Hence, time series for rainfall volume, cumulative rainfall, and time accumulated for precipitation duration (in integer values) are generated. Since these data types have different units, a standard scaling is applied. This is necessary to ensure comparability and smooth optimization. This step proved necessary after training only with the intensity time series, which resulted in significantly worse performance.

3. These four time series are then fed into an ensemble of NARX neural networks. The ensemble is designed to describe structural and parameter uncertainty. Thus, the members are grouped in different $(i, i+1, \ldots, i+n)$ NARX structures ($\mathrm{NARX} - S_{u_i}$), which serve as backbone model structures and consist of different numbers of neurons in the hidden layer. From each backbone structure, different $(j, j+1, \ldots, j+n)$ subordinated NARX networks ($\mathrm{NARX} - P_{u_j}$) are assigned with different weights and biases. Section 2.2 provides the details of how each member (i.e., the $\mathrm{NARX} - P_{u_j}$ networks) of the ensemble is generated.

4. Depending on the number of backbone structures i and subsequent networks j, the networks produce an ensemble of water depth time series in the respective manhole.

5. From the ensemble runs, the forecast is outputted with a confidence interval (e.g., 96%).

6. The EFS is able to forecast the water depth time series for multiple manholes M_n simultaneously.

2.2 Algorithm for the validation of each NARX network of the ensemble

Forecasting of a network generally follows a training and validation routine (Wunsch, Liesch, and Broda 2018). Weights are adjusted with the training data, and the generalization ability is only evaluated later with the validation part (Bergmeir and Benítez 2012). Besides that, supervised neural networks often tend to overfit or get stuck in a local minimum (Atakulreka and Sutivong 2007), which may lead to poor performance. Herein, and in order to obtain a more robust estimate of predictive network performance, a blocked k-fold cross-validation sequence combined with an out-of-sample strategy is developed. Some of the basic ideas behind this strategy follow the works of (Bergmeir and Benítez 2012; Bergmeir, Costantini, and Benítez 2014; Wunsch, Liesch, and Broda 2018).

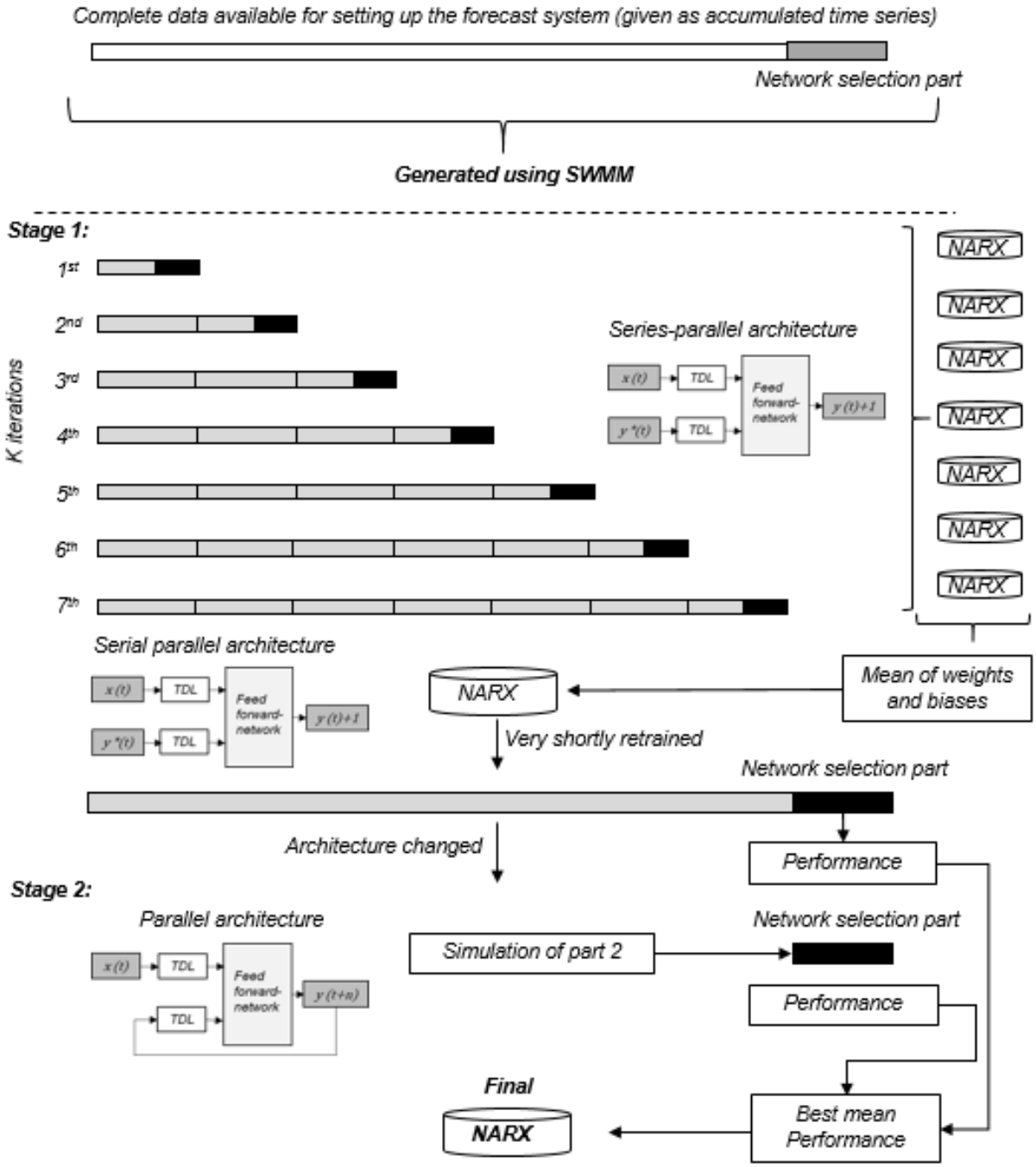


Figure 2: Architecture and strategy for network validation. The black blocks symbolize the validation parts. Stage 1 includes the complete k iterations phase and the training for the mean NARX is computed using the Serial-parallel architecture. The final network validation is done in stage 2. It is based on the simulation of the validation part within the Parallel architecture of the gained network from stage 1.

The complete dataset available for setting up the EFS is divided into two parts. 90% are used for training, and the last 10% are used for the validation part following the out-of-sample strategy. The 90% are split again into seven subsets for seven iterations, where the last 10% of each subset serves for validation. The first iteration includes one subset, while the following iteration includes

the second subset. This is done until the whole 90% is consumed. The approach ensures for each iteration, more data for the training part and a new (smaller) data set for the network validation.

The upper part of Figure 2 shows this process. Besides using more data for validation, the weights and biases are calculated several times on different data, which offers the possibility to treat them as stochastic variables (Bianchi et al. 2017; Wunsch, Liesch, and Broda 2018). For each of the seven iterations, a new NARX network is created where the weights are randomly initialized, which ensures statistical independence among the iterations. Once all iterations are finished, the weights and biases are extracted and averaged. This prevents overfitting and provides prior weights- and bias- knowledge. After that, a new NARX network, with the averaged weights and biases, is generated. This network is very shortly retrained on the total dataset. As such, the prior knowledge of the weights and biases helps to avoid a local minimum. Afterward, the performance in the network validation part of the last 10% is saved, and the first stage is concluded.

This whole process is done in the so-called Series-parallel architecture, where the actual target output, namely the water depth, is used as an input. This configuration allows the use of backpropagation training algorithms, and the network learns on the desired values, which results in good performances in the training stage (Bianchi et al. 2017). Since in the operational and testing mode, the desired values are not available anymore, the so-called Parallel architecture is generated, where the output of the network is fed back as new input. The NARX network, generated in the first stage, is transformed into the parallel architecture, and a simulation of the last 10% is computed. The best performances at this second stage serve as the final validation. This stage is necessary since the Series-parallel architecture often has poor performance when it comes to the operational- or testing phase.

The algorithm described above is used then for two purposes:

1. To find the best network structures which serve as the backbone for the EFS regarding the number of neurons in the hidden layer, namely the ($\mathrm{NARX} - S_{u_i}$).

2. To generate the subsequential networks with different weights and biases, namely the (NARX-P$_{u_j}$).

2.3 The Nonlinear autoregressive with exogenous input network (NARX)

NARX is a supervised neural network with dynamic feedback connections, which is a popular tool in time series modeling and prediction (F.-J. Chang et al. 2014; Rjeily et al. 2017; Abbas et al. 2018; L.-C. Chang et al. 2018; Wunsch, Liesch, and Broda 2018). Mathematically the network, given by Equation (1), calculates the output for the next time step as a function of multiple inputs and outputs of the actual and previous time steps.

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-q}, x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-p}) \tag{1}$$

where y is the output of the modeling time series; x is the input time series; t is the time; and p and q are time delay unit sizes, which account for the time-dependencies and serve as a memory effect. In the present configuration, x is given by the rainfall input, and y is given by the water depth time series, which is fed back. Determining the time delay unit sizes p and q can be a crucial task (Wunsch, Liesch, and Broda 2018). In this study, correlation analysis is applied to investigate the significant lags. As shown by other studies, this provides efficient values for the delay units (Fernando, Zhang, and Kinley 2007; Mounce et al. 2014; Wunsch, Liesch, and Broda 2018). The x delay is determined by the cross-correlation between the rainfall intensity time series and the

water depth time series, while the y delay is gained by the results of the autocorrelation of the water depth series.

2.4 Performance measures

The evaluation and testing of the EFS is linked to a more operational approach. Running the EFS will create the confidence band for the water depth, and decision-making is based on these results. In any case, a decision is more difficult if the forecast delivers a wide interval (Wetterhall et al. 2013). Accordingly, the evaluation underlies two considerations: First, how many target values are hit, and second the width of the confidence band. This can be described by the P_{factor} and R_{factor} in Abbaspour et al. (2007) and (2009). The P_{factor} gives the percentage of the observed values covered/bracketed by the uncertainty band and the R_{factor} gives the width of the uncertainty band divided by the standard deviation of the observed values (Abbaspour et al. 2007, 2009). Thus, a high-quality forecast can be defined as a high P_{factor} and low R_{factor} simultaneously.

$$P_{\text{factor}}(f_u, f_l) = \frac{\sum_{t}^{D_f=180} \text{Obs}_t \subset [(y_t f_u,\ y_t f_l)]}{D_f} \tag{2}$$

$$R_{factor}\,(f_u, f_l) = \frac{|\ (y_t f_u - \ y_t f_l)\ |_{t=Df}}{\delta_{obs}} \tag{3}$$

where f_u and f_l are the upper and lower bounder of the uncertainty band; y_t are the observed values; D_f is the number of predicted minutes; and δ_{obs} is the standard deviation of the observed values. In this study, the observed values are given by SWMM time series. Considering the assumption that rainfall is not changing over time, the P_{factor} is calculated over the whole 180 min forecast window. Since the standard deviation of water-depth values from a UDS normally provides very small values, the R_{factor} is not applicable in this range. To improve and stabilize the factor, a threshold value is incorporated. Hence, the R_{factor} is only calculated when the upper boundary of the confidence interval reaches a surcharge rate of 70%, which ensures a suitable value for the standard deviation. The calculation is then done with a lead time of every 30 min, and the highest value is taken for the evaluation. Since the uncertainty band is generated by the members of the ensemble a measure for assessing the performance of a single member is also necessary. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), or Mean Squared Error (MSE) are often used for this purpose. That is the reason why we use MSE for selecting the members of the ensemble (training and validation). Thus each member of the ensemble is selected based on their MSE performance.

2.5 Synthetic generation of rainfall intensity distributions

The EFS needs to not only predict the timing of surcharge in extreme events but also to distinguish events where no surcharge occurs. Furthermore, the complex hydraulic behavior in the UDS, as well as the different design standards of different locations, demand a wide range of distributed data to be used for training the neural networks. Therefore, rainfall volumes of the return periods 1, 2, 3, 5, 10, 20, 30, 50, and 100 yrs considering durations of 5, 10, 15, 20, 30, 45, 60, 90, 120, and 180 min are taken from the KOSTRA table. This study applies a 5 min rainfall resolution with homogenous and variable intensity distributions. The variable intensity is described with EULER,

EULER TYPE II, POISSON, and for the Germany-based OTTER/KÖNIGER distributions. The two Euler types are described in Worksheet 118 (DWA 2006). EULER TYPE II follows the same values as the EULER, except that the peak intensity is not occurring at the beginning of the event. The POISSON assumes that the highest intensity is in the middle of the event. The OTTER/KÖNIGER storms are given as standardized intensity gradients (Bayerisches Landesamt für Wasserwirtschaft 2019). The latter rainfall hydrographs consist of a pre-rain, main rainfall, and an after-rain whereby the sum of the main rainfall equals the relevant precipitation volume of the "Coordinated heavy precipitation regionalization and evaluation of the German Weather Service" (in German: "Koordinierte Starkniederschlagsregionalisierung und –auswertung" – KOSTRA) table for the respective duration.

2.6 Storm Water Management Model (SWMM)

In this study, the physically based model SWMM is used to generate training, validation, and testing data for the neural networks: SWMM is a one-dimensional (1D) fully dynamic rainfall-runoff simulation model designed by USEPA. It consists of two major components, namely the Rainfall-Runoff and the Routing component. The former operates on a collection of sub-catchments, which receive precipitation and generate runoff. The Routing component transports, according to the Saint-Venant equations, the generated runoff through a system of pipes, channels, storage/treatment devices, pumps, and regulators (Chiang et al. 2010; Rossman 2017). At every calculation time step, a combination of the momentum and the continuity equation (4) must be solved for each link to compute the time-dependent behavior of the flow (Rossman 2017).

$$\frac{\partial Q}{\partial t} = 2U\frac{\partial A}{\partial t} + U^2\frac{\partial A}{\partial x} - gA\frac{\partial H}{\partial x} - gAS_f \tag{4}$$

where Q is the flow rate; t is the time; U is the flow velocity; A is the cross-sectional flow area; g is the acceleration of gravity; H is the hydraulic head of water in the conduit; x is the distance; and S_f is the friction slope (head loss per unit length). SWMM is commonly applied and widely used among the research community (Baffaut and Delleur 1990; Tsihrintzis and Hamid 1998; Ackerman and Schiff 2003; Denault, Millar, and Lence 2006; Maharjan et al. 2009; Chiang et al. 2010; Rossman 2017). Further information on the governing equations and the underlying solving methods can be found in Rossman (2017). This study applies SWMM for the calculation of the water-depth time series in the manholes for all events.

3 Study area, data, and default settings

The city of Baiersdorf, located in Bavaria, Germany, and its UDS, was chosen to be the study site (Figure 3). The combined system conveys the waste- and stormwater runoff from the city itself and, in addition, from six small villages acting as satellites. About 16,000 inhabitants are connected to the UDS, and the wastewater treatment plant has a maximum capacity of 25,000 PE (Population Equivalent). The sewer network data, containing the location and geometry of the components, are provided by the Municipal Drainage Authority of Baiersdorf. The model includes 3,475 nodes (representing manholes, retention storages, and outfalls), 3,436 conduits, six pumps, 16 weirs, and 19 orifices. Since the UDS is a combined sewer system, a dry-weather wastewater baseflow of 130.5 l/(PE*d) is additionally integrated (Brombach and Dettmar 2019). A digital elevation model (DEM) with 5 m resolution and land use data also provided by the authority of Baiersdorf

are used herein. Sealed surfaces and building areas are connected to the system via the nearest manhole. Hence, each manhole has its own subcatchment. The subcatchments drain a total area of about 290 ha. The Manning values for the sealed surfaces are taken from (HIOS 2019) and set to 0.013 s/[$m^{1/3}$], and a depression storage coefficient value of 4 mm is assumed as recommended in Computational Hydraulics International – CHI (2019). Figure 3 illustrates the classification and shows the UDS in the study site. Since only sealed surfaces are connected to the sewer system, any change in infiltration rate or groundwater recharge from multiple consecutive rain events is not considered.

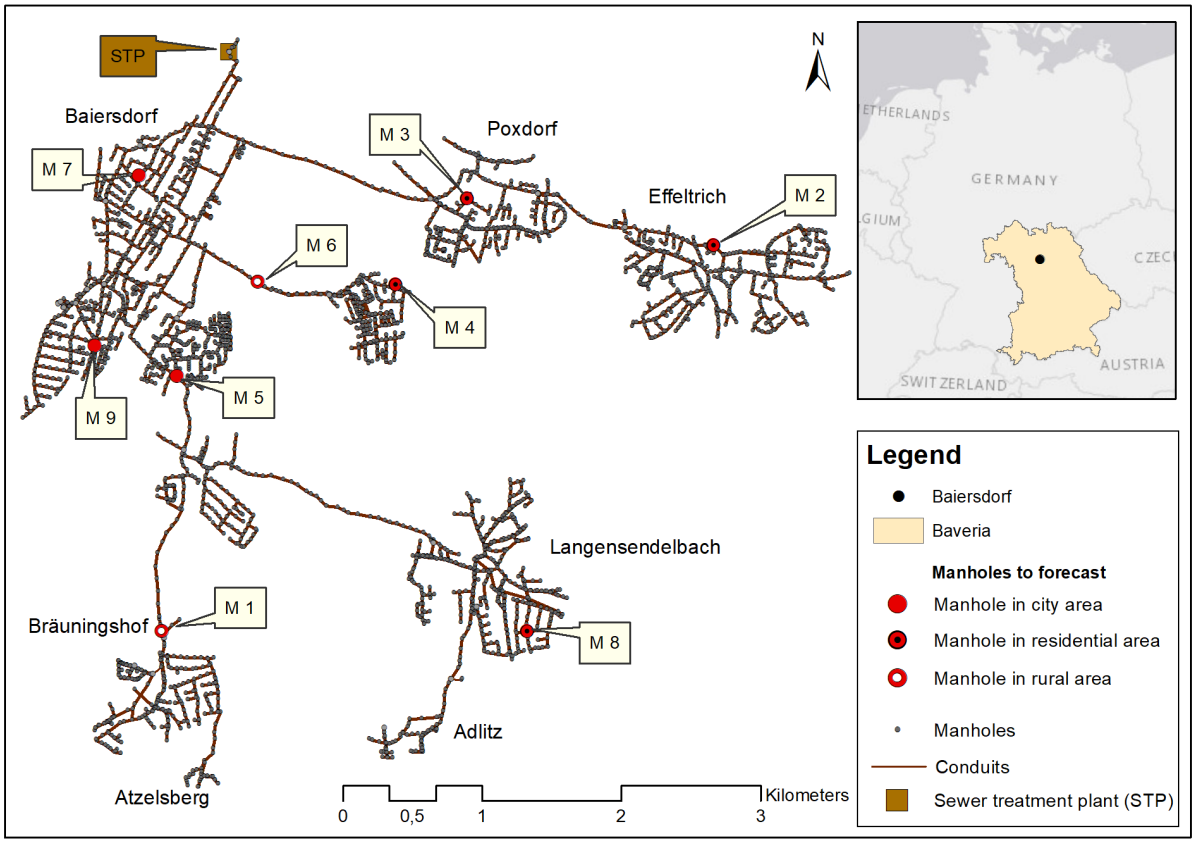


Figure 3: UDS of the city of Baiersdorf and satellites. The red highlighted manholes are the selected manholes for the EFS.

The rainfall is given at one gauge station, assuming a homogeneous rainfall distribution over the study area. The respective precipitation volumes for the different scenarios are taken from KOSTRA. KOSTRA data offer precipitation volumes depending on rainfall duration (varying from 5 min to 72 h) and return period (varying from 1 yrs to 100 yrs) extrapolated to a Germany-wide grid. These data are commonly used for dimensioning water and wastewater management structures. In this study, the latest version is applied.

For the sake of demonstrating the new methodology, only nine manholes ($M_n = 9$) are selected for the EFS in this study. Figure 3 shows the UDS in the study area and the selected manholes. These manholes are among the critical manholes that surcharge often. Since the selected ones are hydraulically independent of each other and distributed over the entire UDS, it is reasonable to assume that the methodology is transferable to other manholes. Having, by aim, the forecast of short-duration high-intensity rainfall events, and given the small size of the catchment ($A = 280$ ha),

the synthetic data for deriving the EFS is made of rainfall events with a maximum duration of 180 min. The maximum lead time is, therefore, also set for $L_t = 3$ hours.

Besides the synthetic rainfall data, the observed event from the year 2007, shown in Figure 4, which caused major flooding, is used for testing the EFS.

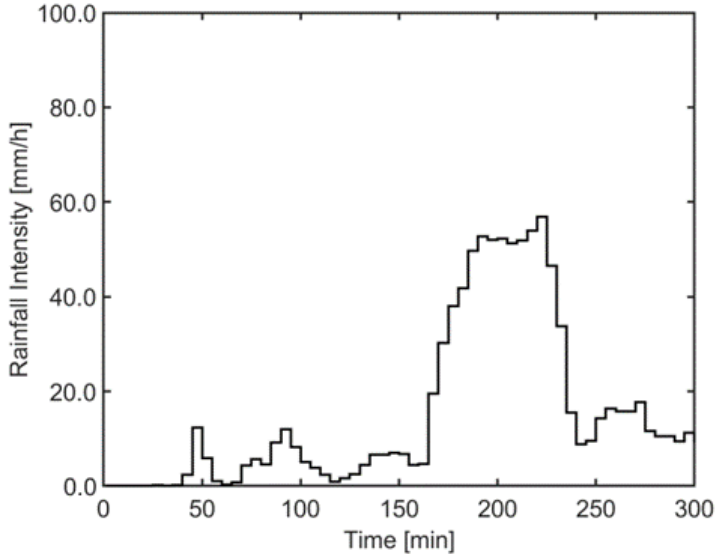


Figure 4: Single observed rainfall from the year 2007.

The neural networks are trained by the scaled conjugate gradient descent (SCG) (Møller 1993) using the mean squared error (MSE) as the loss function. It is decided to consider only one hidden layer and a combination of the hyperbolic tangent sigmoid as a transfer function in the hidden layer and a linear transfer function for the output layer. This combination also delivered promising results in other studies (Chiang et al. 2010; Rjeily et al. 2017; Bianchi et al. 2017). Also, due to computational times and for the sake of demonstrating the new methodology, the $\mathrm{NARX} - P_{u_j}$ are obtained by repeating the validation algorithm 100 times for each backbone structure and the best 20 ($j = 20$) are taken as members into the EFS.

4 Model verification following DIN EN 752 and DWA-A 118

A classical calibration and validation approach based on actual measured data for specific events was not possible. In fact, this lack of measured field data is a recognized problem in urban flood modeling (Leandro et al. 2011). Thus, a different validation approach to the sewer network is used based on regulation. In this study, the model is validated based on the European standard regulation DIN-EN 752 (2017) and the German regulation DWA-A 118 (DWA 2006). These regulations define the design standards for sewer systems and provide recommendations for the frequencies of a surcharge event. Accordingly, it was verified that the sewer network did not surcharge for a 5 yrs return period in city areas, 3 yrs for residential areas, and 2 yrs for rural areas (DWA 2006). The duration of the design storm of 15 min is recommended in DWA-A 118 (DWA 2006). The rainfall volume is taken from KOSTRA. A widespread sewer surcharge was only observed from a 20 yrs return period.

5 Results

5.1 Synthetic generation of rainfall intensity distributions and prior data division

397 synthetic rainfall events build up the database. Figure 5 shows the covered intensity rainfall patterns.

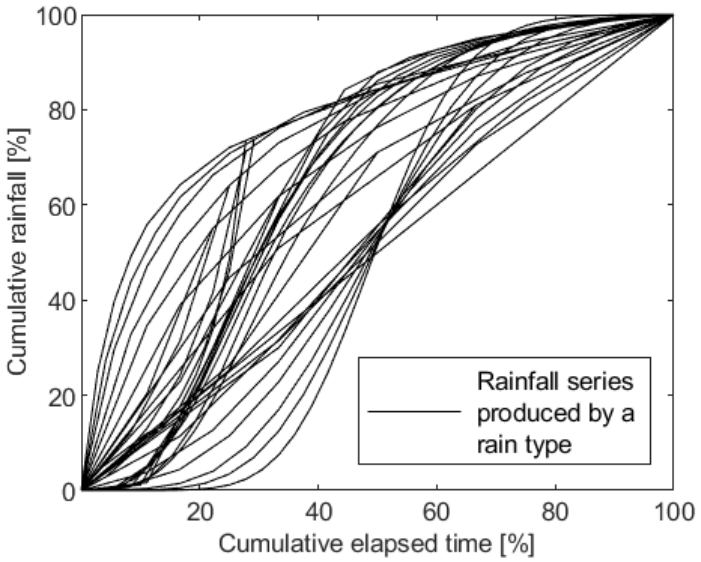


Figure 5: Covered intensity patterns by all rainfall types, presented by the relationship between the cumulative values of elapsed time and rainfall volume.

Ten synthetic events and the observed rainfall from 2007 are used for testing the developed EFS and are not considered in the training phase of the ensemble. Table 1 shows the types, return periods, and durations covered. The remaining 387 events are used for the training phase of the ensemble of the EFS.

Table 1: Test events for the developed EFS.

Rain type	Return period [yrs]	Duration [min]
Homogenous	2	10
Euler	3	20
Homogenous	5	15
Otter-Königer	5	120
Poisson	10	180
Euler type II	10	60
Otter-Königer	20	45
Euler type II	30	20
Poisson	50	90
Otter-Königer	100	30
Observed event from 2007	Approx. 100	320 (peak interval $\sim$120)

5.2 Algorithm for validation of the NARX networks and time delays

The determination of the time delay unit sizes p and q concludes after the cross-correlation between the rainfall intensity time series and the water depth series with a seven-time steps delay for p (referring to rainfall x), while the autocorrelation of the water depth series comes to a value of 12 timesteps for q (referring to y).

The application of the newly developed algorithm for validation of the NARX network for the backbone structures ($\mathrm{NARX} - S_{u_i}$), (see chapter 2.2; application point 1) lead to the results presented in Table 2. It indicates the use of $i = 5$ backbone structures since the sixth-best performance (0.0408 with 16 neurons) is about 40% worse than the fifth place (0.0300 with 14 neurons). So it is decided to use 14, 15, 18, 19, and 20 neurons in the hidden layer for the respective i networks (S_{u_i}).

Table 2: Results are obtained by running the network validation algorithm concerning the number of neurons in the hidden layer. A comparison with the performance of the first stage illustrates the need for stage 2.

Number of neurons in HL	MSE after stage 1	MSE after stage 2
3	0.0105	0.3455
4	0.0046	1.0250
5	0.0033	1.0968
6	0.0025	2.1136
7	0.0018	0.0795
8	0.0019	0.0425
9	0.0012	0.5479
10	0.0017	0.0843
11	0.0012	0.0721
12	0.0015	0.1767
13	0.0018	1.5524
14	0.0011	0.0300
15	0.0014	0.0227
16	0.0010	0.0408
17	0.0011	0.0841
18	0.0010	0.0270
19	0.0013	0.0182
20	0.0011	0.0212

Afterward, the algorithm is applied for generating the respective ensemble runs $\mathrm{NARX} - P_{u_j}$. The results for the training and validation accuracy are shown as a mean MSE value of the respective 20 runs in Table 3.

Table 3: Results are obtained by running the network validation algorithm concerning the subsequential NARX networks for describing the parameter uncertainty through the ensemble runs.

Backbone $\mathrm{NARX} - S_{u_i}$	Training Set Mean MSE	Validation Set Mean MSE
	Calculated by the best 20 $\mathrm{NARX} - P_{u_j}$ generated by the 100 times iterations of the validation algorithm	
14	0.0369	0.0389
15	0.0266	0.0286
18	0.0282	0.0352
19	0.0207	0.0267
20	0.0236	0.0276

5.3 Evaluation and testing of the ensemble forecast system (EFS)

5.3.1 Testing of the EFS with Synthetic events

For testing the developed EFS special focus is set on the performance of the uncertainty band generated by the 100 ensemble runs. Hence, the measures proposed by Abbaspour 2007, namely P_{factor} and R_{factor} as justified in chapter 2.4 are used. The results show a good performance of the forecast system for the ten synthetic events tested. The EFS delivers a P_{factor} at a minimum of 74% and a maximum (best) of 94%. With an increasing return period, a small performance increase is also visible. The R_{factor} delivers values between a minimum (best) of 1.36 and a maximum of 3.08. Table 4 shows the respective results. Figures 6 and 7 show the visual output of the EFS for two of the ten tests.

Besides the high prediction performance, the EFS delivers these predictions within 56 sec, instead of 200 min taken by SWMM (2 min for one single ensemble member) for running the 100 ensemble members of the EFS. Thus, the EFS is about 214 times more efficient than SWMM.

Table 4: Results for the P and R factors for the ten synthetic test events.

Rain type	Return period [yrs]	Duration [min]	P_{factor} [%]	R_{factor}
Homogenous	2	10	77.96	3.08
Euler	3	20	74.69	2.86
Homogenous	5	15	80.12	2.81
Otter-Königer	5	120	93.33	2.26
Poisson	10	180	94.94	2.90
Euler type II	10	60	89.23	2.66
Otter-Königer	20	45	89.20	1.36
Euler type II	30	20	93.89	2.58
Poisson	50	90	92.04	1.97
Otter-Königer	100	30	91.11	2.73

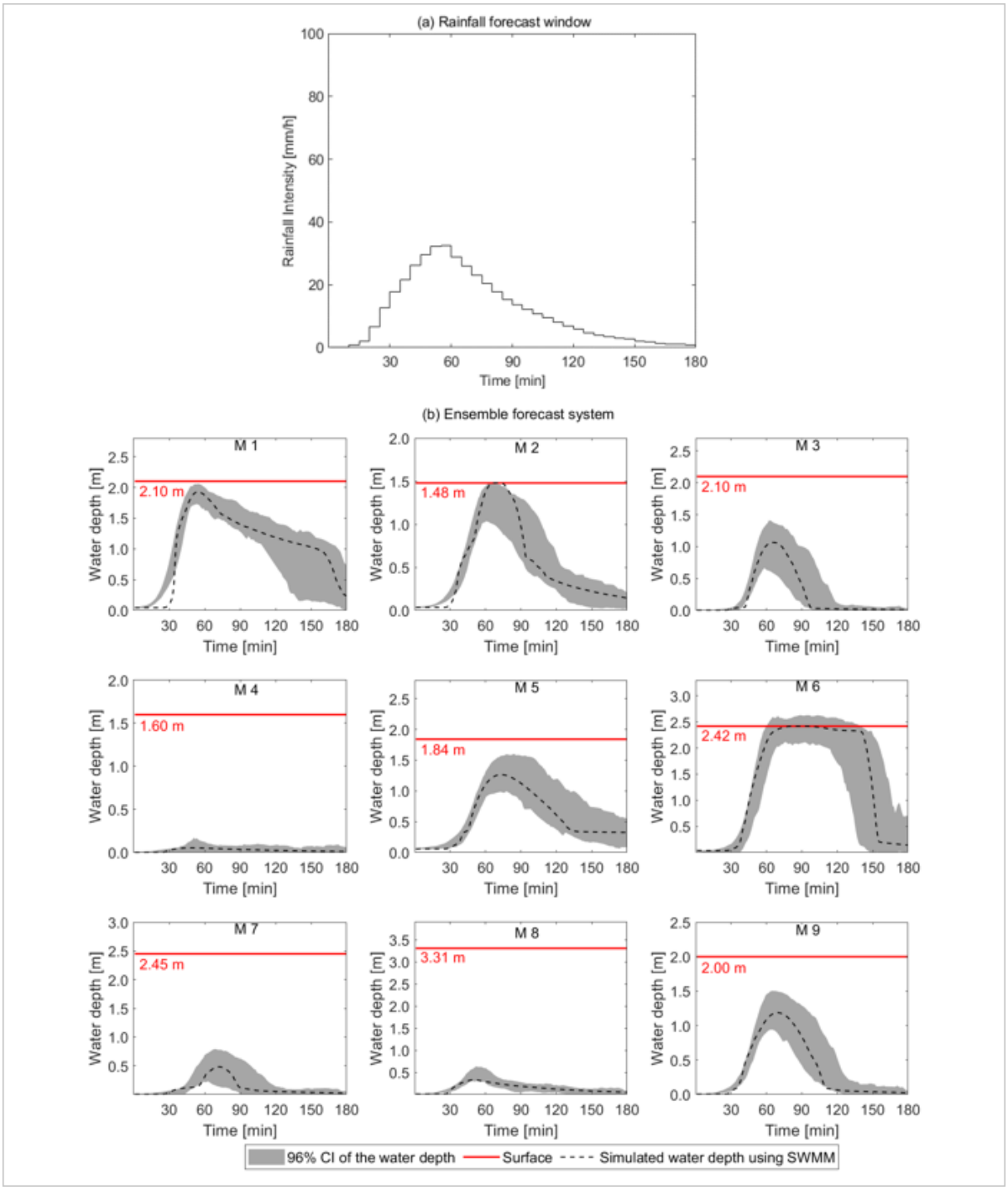


Figure 6: Forecast run of a 5 yrs Otter-Königer rainfall. Rainfall prediction (a) and results of running the forecast system (b), where the water depth is given by a 96% confidence interval.

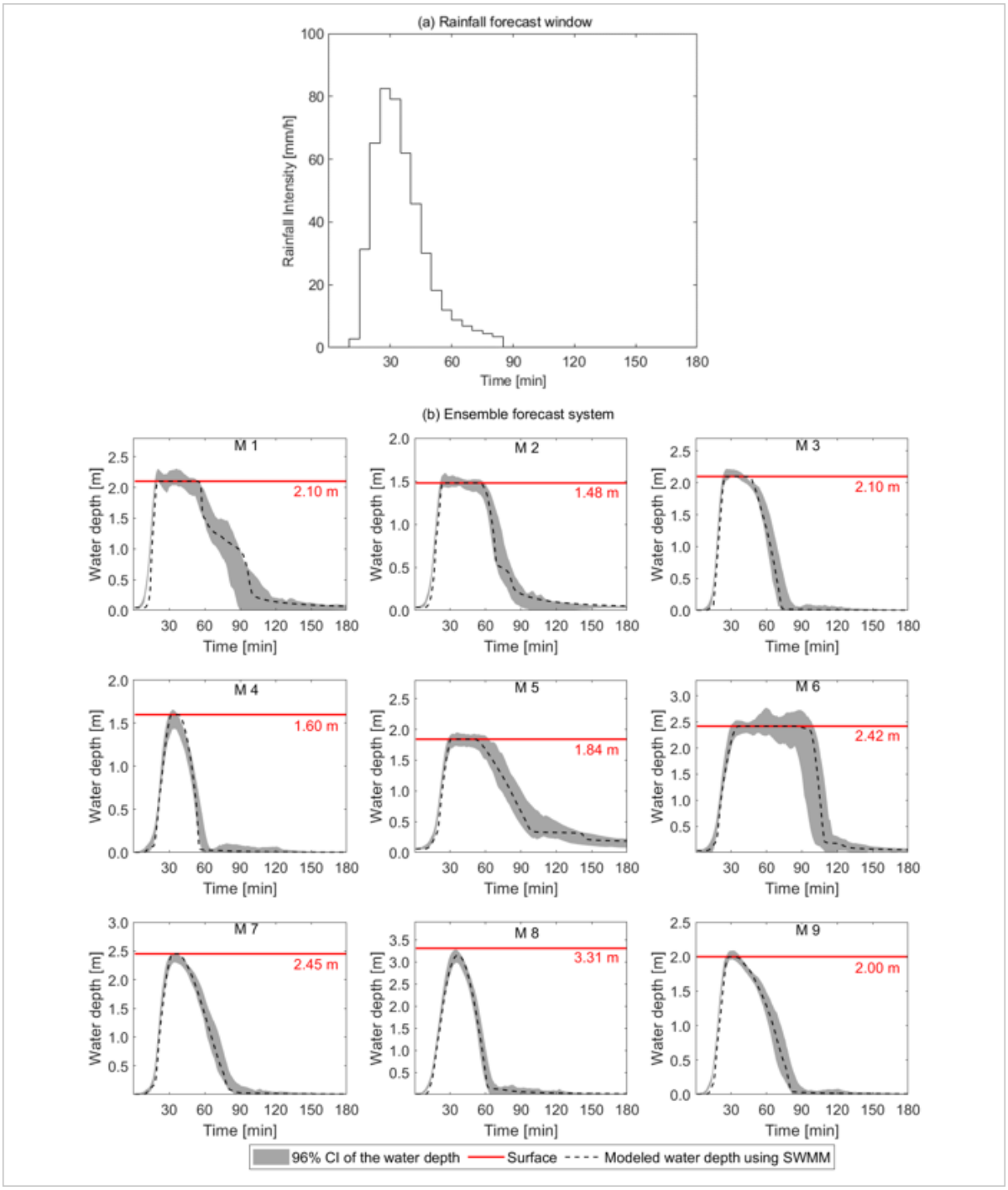


Figure 7: Forecast run of a 20 yrs Otter-Königer rainfall. Rainfall prediction (a) and results of running the forecast system (b), where the water depth is given by a 96% confidence interval.

5.3.2 Evaluation of the EFS with the observed single rainfall event from July 2007

Since the duration of the observed event from 2007 exceeds the forecasting time window of 180 min, the forecast was split into two parts. The results show that the P_{factor} of the entire forecast reaches 85% in the first part and 90% in the second part. The forecast for the first part provides a nearly perfect hit rate. The second part follows a similar pattern but with a somewhat larger uncertainty

band, particularly toward the end of the forecast window. The starting times of surcharging, as obtained by SWMM, are bracketed by the uncertainty interval of 96% of the EFS. Nevertheless, the R_{factor} are higher and reach the whole EFS of 10.69 in the second part. Figures 8 and 9 show the results graphically, where detailed values are provided in Tables 5 and 6.

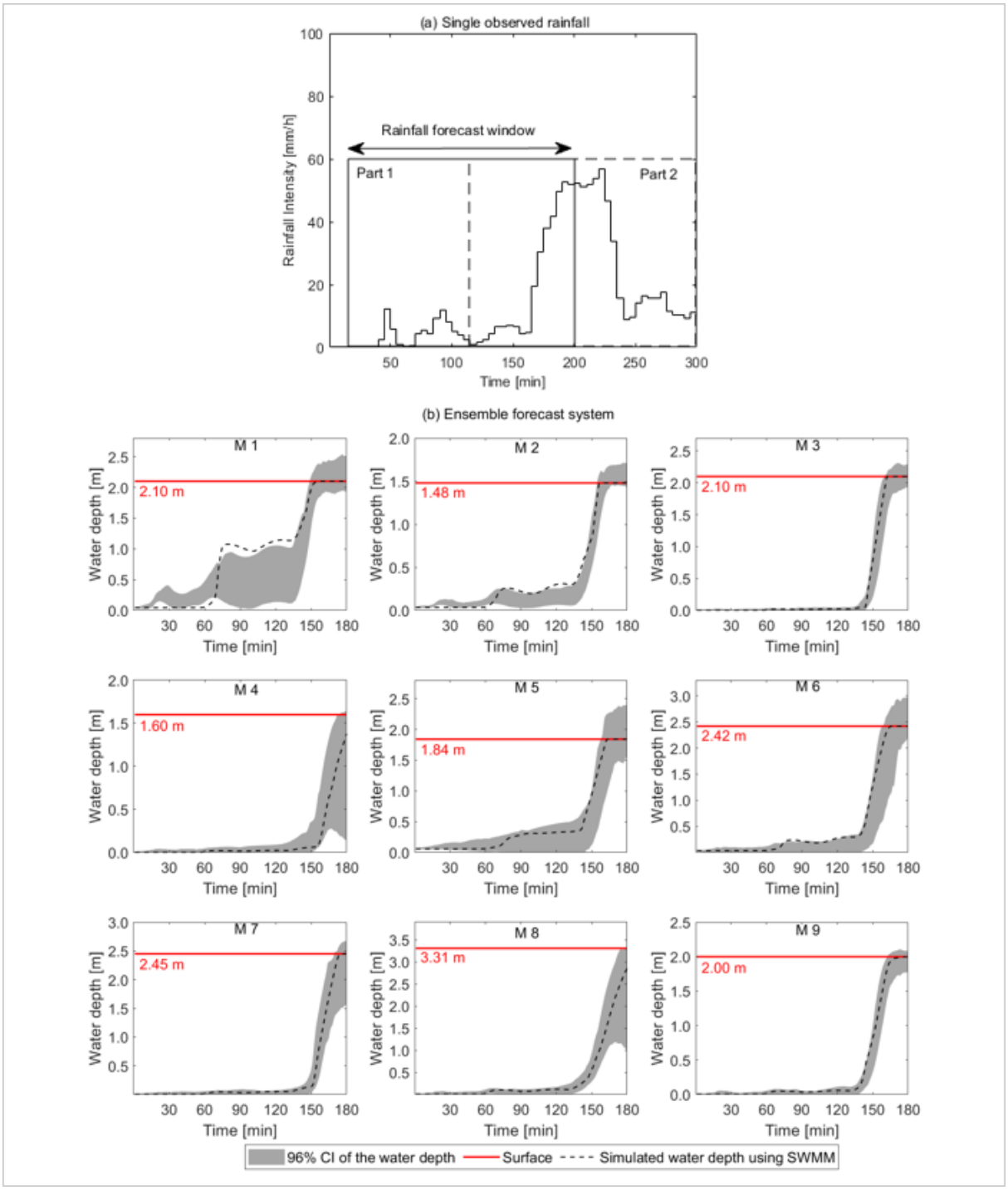


Figure 8: Forecast results of the first part of the observed rainfall event from July 2007.

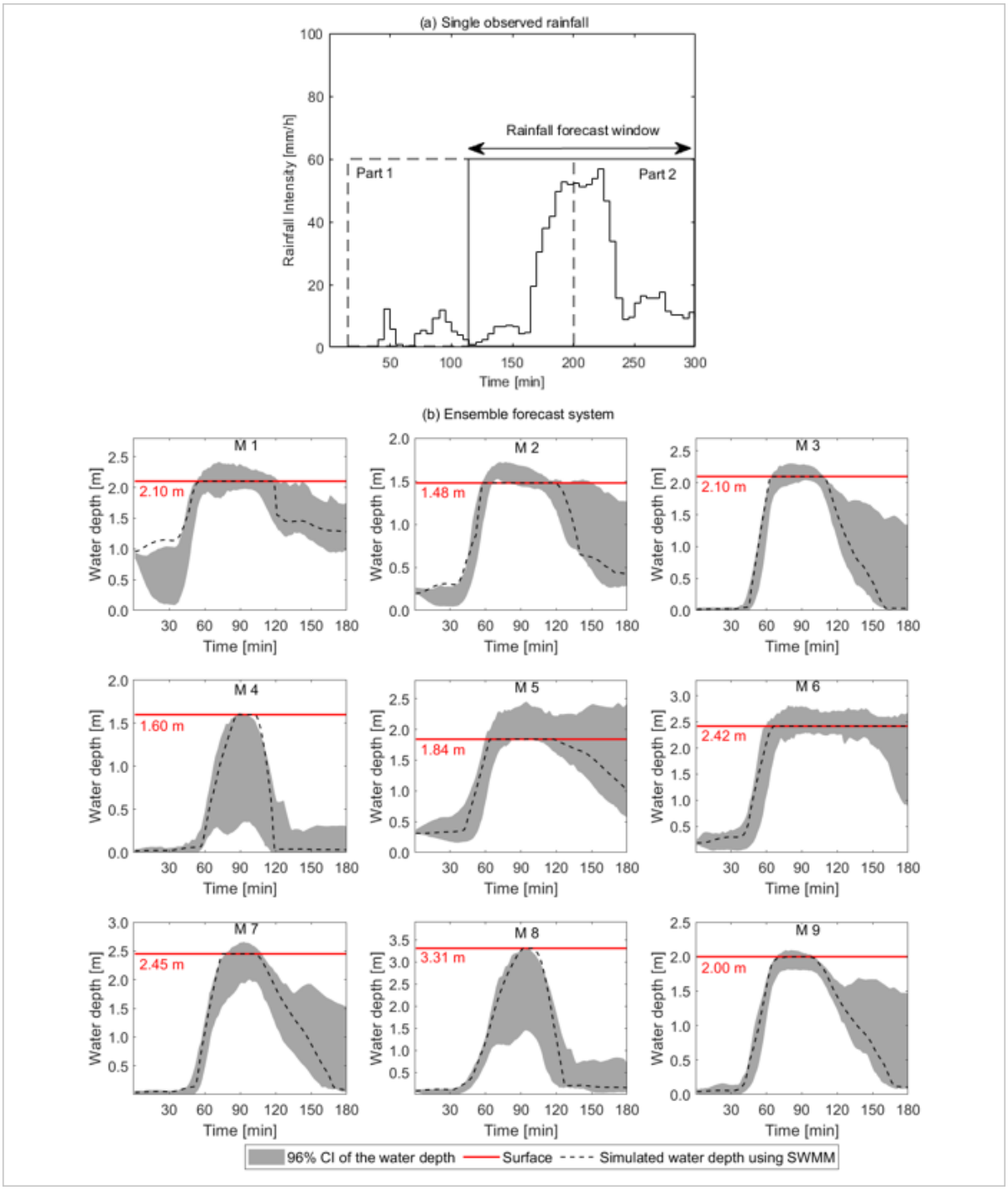


Figure 9: Forecast results of the second part of the observed rainfall event from July 2007.

Table 5: Summary of the forecast system performance for the real event: Part 1. P_{factor} is calculated over the entire 180 min window, while the R_{factor} is calculated for each lead time when the upper boundary of the CI reaches a surcharge rate of 70%. The highest value of this set is printed.

	Forecast system		
P_{factor} [%]	85.49		
R_{factor}	5.01		
Manhole	1	2	3
P_{factor} [%]	42.22	48.89	100.00
R_{factor}	0.78	0.58	0.54
Manhole	4	5	6
P_{factor} [%]	100.00	100.00	84.44
R_{factor}	5.01	1.72	1.08
Manhole	7	8	9
P_{factor} [%]	100.00	96.67	97.22
R_{factor}	1.89	3.47	0.49

Table 6: Summary of the forecast system performance for the real event: Part 2. P_{factor} is calculated over the entire 180 min window, while the R_{factor} is calculated for each lead time when the upper boundary of the CI reaches a surcharge rate of 70%. The highest value of this set is printed.

	Forecast system		
P_{factor} [%]	90.80		
R_{factor}	10.50		
Manhole	1	2	3
P_{factor} [%]	77.78	70.00	98.33
R_{factor}	1.79	1.94	1.68
Manhole	4	5	6
P_{factor} [%]	91.11	93.33	100.00
R_{factor}	10.50	3.014	2.36
Manhole	7	8	9
P_{factor} [%]	97.22	89.44	100.00
R_{factor}	3.68	2.06	1.81

6 Discussion

6.1 Synthetic generation of rainfall intensity distributions and prior data division

The rainfall intensity distributions cover a wide range of scenarios, as can be seen in Figure 5. Table 1 presented the selected rainfall types for testing the EFS. Care was taken that the test dataset consisted of all constructed rainfall types. The range of return periods and duration match the values mentioned in KOSTRA. Some pre-testing with SWMM showed no surcharge or notable water depths for 5 min durations at each return period, as such, this duration was not included in the testing set.Since UDSs are designed in the ranges equally to 1 to 10 yrs return periods, for the critical (in case of surcharging) years 5 and 10, four test scenarios were selected to achieve a deeper investigation of these ranges. This can be visually verified by looking at the number of manholes surcharging from the selected ones (Figure 7).

The good overall performance, shown in Table 4, indicates that the amount of synthetic rainfall intensity distribution was sufficient and effective for constructing the EFS. The intensity pattern and training data even showed high performance for the single observed rain (see Figures 8 and 9; Tables 5 and 6), although the intensity distribution was different from that used for training.

6.2 Algorithm for validation of the NARX networks and time delays

The time delay unit sizes p and q, which serve as the network memory for the feedback connection of the water depth y and the rainfall input x, are determined by correlation analysis. Since the results demonstrated a nearly perfect hit rate for the beginning of the surcharge, it can be concluded that the unique recurrent network architecture and the delays selected can maintain the hydrodynamic characteristics and the input-output relationships faithful to the hydrodynamic model. This can be visually verified by Figures 6, 7, 8, and 9. Therefore, the correlation analysis undertaken provided a good estimation for these parameters.

The application of the developed algorithm for validation of a single NARX network with the purpose of selecting the backbone $\mathrm{NARX}-S_{u_i}$ structures and $\mathrm{NARX}-P_{u_j}$ was shown in Tables 2 and 3. Table 2 shows very low MSE values for the number of neurons 14, 15, 18, 19, and 20. Table 2 also shows the importance of validating the NARX parallel architecture (stage 2). Since nearly all configurations of neurons offered very low prediction errors in stage 1, a clear and optimal network selection could only be made after stage 2. In summary, the low MSE values of stage 2 in Tables 2 and 3 demonstrated the effectiveness of the developed algorithm. The performance results, as they are shown in the previous section, clearly verify that the algorithm was able to validate the networks, produce a robust prediction estimator, and overcome the overfitting or local minimum issues of supervised networks in the desired manner.

6.3 Evaluation and testing of the ensemble forecast system (EFS)

It is visible that the EFS is performing well and is able to predict the timing of the surcharge of the nine manholes investigated. The results demonstrate that the NARX network is a suitable tool for modeling and predicting time series multi-steps ahead, even for 3 h, as in this case. Similar results and conclusions were also drawn regarding the ability of NARX in other studies (F.-J. Chang et al. 2014; Rjeily et al. 2017; Boussaada et al. 2018; L.-C. Chang et al. 2018; Wunsch, Liesch, and Broda

2018). The current framework of the EFS delivers a suitable forecast which offers the user also information on the included uncertainty, and therefore the task of decision-making is made easier. For all test events, at least P_{factor} values of 74% and 94% were reached, as it was shown in Table 4.

The investigation of the width of the confidence band and the R_{factor} showed a range between 1.36 and 10.68. For all test events, the band starts narrower at the beginning and delivers an almost perfect hit rate at the beginning of the surcharge (see Figures 6, 7, 8, and 9). As the forecast continues, the uncertainty band becomes wider, which can be explained by the divergence nature of an ensemble, but the end of the surcharge can be predicted sufficiently. Especially the results of the single observed rainfall showed a higher uncertainty (Figures 8 and 9; Tables 5 and 6). This is likely due to the lack of observed rainfall events included in the training set. In comparison to the synthetic events, the observed rainfall, especially in part 2, does not show a strictly rising limb, which was responsible for the higher uncertainty in the EFS.

The width of uncertainty results from the different ensemble runs $\mathrm{NARX} - P_{u_j}$. Each member is generated after a random initialization of weights and biases and the iteration of the validation algorithm. A simple measure to change the uncertainty band is to change the confidence interval or the statistical distribution used to describe it.

The results indicate a small tendency toward a better prediction for higher return periods (see Table 4). This is because the MSE (which was used as the loss function for training) is sensitive to high values, and the training algorithm follows the gradient with the steepest descent. This is the reason for the lower P_{factor} values in the lower range of the return periods and consequently explains the excellent performance in the upper range. Similar behavior was also found by Rjeily et al. (2017). Root mean square error or the mean absolute error can be used to evaluate models' performance. Herein P and R_{factor} are preferred since we are specifically interested in the description of the uncertainty bands for flood forecasting (Abbaspour et al. 2009; Leandro et al. 2019). Within one type of rain, this tendency does not quite apply. For example, the POISSON events show a bit worse results going from 10 to 50 years. However, the results vary only by about 3%. This represents only five timesteps outside of the uncertainty band. Furthermore, this is likely to occur in the first time steps, for which the uncertainty band is extremely thin. In general, all test events showed that the EFS is capable of forecasting the beginning and end of a surcharge of the stormwater sewer satisfactorily.

6.4 Limitations, optimization, and recommendations

The present EFS is only trained with synthetic rainfall events, which limits the performance on real observed values. One of the first strategies to improve the quality of EFS is, therefore, simply to enlarge the training data subset, meaning to include more real observed rainfall events. Other studies used a hybrid database for training the network resulting already in higher efficiency (Crotti, Leandro, and Bhola 2019). For a real application, it is therefore recommended to implement this extension. It should be, however, noted that this extension cannot be applied everywhere. Indeed, in the case of data scarcity, it may not be possible to extend the synthetic datasets to a hybrid data set.

Additionally, the rainfall intensity distributions are based on a homogenous spatial rainfall over the study area. For small areas, such as in our case study, this is a reasonable assumption; however, for large UDS, this may not be the case. In those cases, the main characteristics of the rainfall should be first investigated and then provided additionally to the network as input.

Finally, the EFS was trained based on the hydrodynamic model SWMM. Thus, all structural and parameter uncertainties arising from the use of SWMM are not included in the EFS. In any case, it should be mentioned that manhole and pipe measurements in stormwater sewers are not available, hence replacing the physically based model with such data is not possible. As such, including this source of uncertainty in the forecast requires further research to be adequately quantified.

Besides the included uncertainty, it is worth mentioning that SWMM solves the Saint-Venant equation along the sewer system. The 2D flood inundation of the surcharged water at the surface is not considered. For a more realistic flow exchange between surface and sewer, a 1D/2D dual drainage model should be used, coupling a 1D sewer system with a 2D surface flow model. However, it should be noted that such a model is more challenging to set up and computationally more expensive than SWMM. Also, it should be mentioned that our study focuses on developing a data-driven approach for predicting the stormwater sewer surcharge. In this sense, the method developed is general and applicable to both 1D sewer models and 1D/2D dual drainage models.

6 Conclusion

This study presented the development of an ensemble forecast system for predicting the beginning and end of stormwater sewer surcharge. This was achieved by the forecast of the water depth time series occurring in selected manholes using an ensemble of NARX neural networks.

Based on the results, the following conclusions can be drawn:

1. The EFS and the included NARX networks deliver a very good forecast in a computationally efficient way. It was demonstrated that water depths in manholes distributed throughout an entire UDS can be predicted within 56 sec, making the developed EFS suitable for real-time operational applications even for very large stormwater sewers.

2. Supervised neural networks often tend to overfit or often get stuck in a local minimum. The developed novel algorithm for validation of a NARX network showed that these issues can be overcome.

3. Assessing the uncertainty in the forecasts is essential for decision-making. The ensemble of neural networks with different numbers of neurons in the hidden layer and different weights and biases was able to produce a useful uncertainty band. The uncertainty band is added to the water depth time series of the manhole. The band reached goodnesses of fit of P_{factor} between 74% and 94% and R_{factor} between 1.36 and 10.68.

Acknowledgments

The research presented in this paper has been carried out as part of the HiOS project (Hinweiskarte Oberflächenabfluss und Sturzflut) funded by the Bayerisches Staatsministerium für Umwelt und Verbraucherschutz (StMUV) [Bavarian State Ministry of the Environment and Consumer Protection] (69-0270-92086/2017) and supervised by the Bayerisches Landesamt für Umwelt (LfU) [Bavarian Environment Agency].

References

Abbas, Farukh, Donghan Feng, Salman Habib, Usama Rahman, Aazim Rasool, and Zheng Yan. 2018. Short Term Residential Load Forecasting: An Improved Optimal Nonlinear Auto Regressive (NARX) Method with Exponential Weight Decay Function. *Electronics* 7 (12): 432. doi:10.3390/electronics7120432.

Abbaspour, Karim C., Monireh Faramarzi, Samaneh Seyed Ghasemi, and Hong Yang. 2009. Assessing the Impact of Climate Change on Water Resources in Iran. *Water Resources Research* 45 (10). doi:10.1029/2008WR007615.

Abbaspour, Karim C., Jing Yang, Ivan Maximov, Rosi Siber, Konrad Bogner, Johanna Mieleitner, Juerg Zobrist, and Raghavan Srinivasan. 2007. Modelling Hydrology and Water Quality in the Pre-Alpine/Alpine Thur Watershed Using SWAT. *Journal of Hydrology* 333 (2–4): 413–430. doi:10.1016/j.jhydrol.2006.09.014.

Ackerman, Drew, and Kenneth Schiff. 2003. Modeling Storm Water Mass Emissions to the Southern California Bight. *Journal of Environmental Engineering* 129 (4). American Society of Civil Engineers: 308–317. doi:10.1061/(ASCE)0733-9372(2003)129:4(308).

Atakulreka, Akarachai, and Daricha Sutivong. 2007. Avoiding Local Minima in Feedforward Neural Networks by Simultaneous Learning. In *AI 2007: Advances in Artificial Intelligence*, edited by Mehmet A. Orgun and John Thornton, 4830:100–109. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-76928-6_12.

Baffaut, Claire, and Jacques W. Delleur. 1990. Calibration of SWMM Runoff Quality Model with Expert System. *Journal of Water Resources Planning and Management* 116 (2): 247–261. doi:10.1061/(ASCE)0733-9496(1990)116:2(247).

Bayerisches Landesamt für Wasserwirtschaft. 2019. Aufstellung von Modellregengruppen Nach OTTER/KOENIGER Mittels Normiertem Intensitaetsverlauf: Anlage 2 Zum Merkblatt 4.3/3 [WWW Document]. https://www.lfu.bayern.de/wasser/merkblattsammlung/teil4_oberirdische_gewaesser/doc/nr_433_anlage2.pdf.

Bergmeir, Christoph, and José M. Benítez. 2012. On the Use of Cross-Validation for Time Series Predictor Evaluation. *Information Sciences* 191 (May): 192–213. doi:10.1016/j.ins.2011.12.028.

Bergmeir, Christoph, Mauro Costantini, and José M. Benítez. 2014. On the Usefulness of Cross-Validation for Directional Forecast Evaluation. *Computational Statistics & Data Analysis* 76 (August): 132–143. doi:10.1016/j.csda.2014.02.001.

Berkhahn, Simon, Lothar Fuchs, and Insa Neuweiler. 2019. An Ensemble Neural Network Model for Real-Time Prediction of Urban Floods. *Journal of Hydrology* 12. doi:10.1016/j.jhydrol.2019.05.066.

Bermúdez, María, Victor Ntegeka, Vincent Wolfs, and Patrick Willems. 2018. Development and Comparison of Two Fast Surrogate Models for Urban Pluvial Flood Simulations. *Water Resources Management* 32 (8): 2801–2815. doi:10.1007/s11269-018-1959-8.

Beven, Keith, and Andrew Binley. 1992. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrological Processes* 6 (3): 279–298. doi:10.1002/hyp.3360060305.

Bianchi, Filippo Maria, Enrico Maiorino, Michael C. Kampffmeyer, Antonello Rizzi, and Robert Jenssen. 2017. An Overview and Comparative Analysis of Recurrent Neural Networks for Short Term Load Forecasting. *ArXiv:1705.04378 [Cs]*. doi:10.1007/978-3-319-70338-1.

Boussaada, Zina, Octavian Curea, Ahmed Remaci, Haritza Camblong, and Najiba Mrabet Bellaaj. 2018. A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. *Energies* 11 (3): 620. doi:10.3390/en11030620.

Brombach, Hansjörg, and Joachim Dettmar. 2019. Im Spiegel der Statistik: Abwasserkanalisation und Regenwasserbehandlung in Deutschland. *KA-Korrespondenz Abwasser* 2019 (5): 354–364. doi:10.3242/kae2019.05.001.

Chang, Fi-John, Pin-An Chen, Ying-Ray Lu, Eric Huang, and Kai-Yao Chang. 2014. Real-Time Multi-Step-Ahead Water Level Forecasting by Recurrent Neural Networks for Urban Flood Control. *Journal of Hydrology* 517 (September): 836–846. doi: 10.1016/j.jhydrol.2014.06.013.

Chang, Li-Chiu, Mohd Amin, Shun-Nien Yang, and Fi-John Chang. 2018. Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. *Water* 10 (9): 1–18. doi: 10.3390/w10091283.

CHI. 2019. PCSWMM Support [WWW Document]. https://support.chiwater.com.

Chiang, Yen-Ming, Li-Chiu Chang, Meng-Jung Tsai, Yi-Fung Wang, and Fi-John Chang. 2010. Dynamic Neural Networks for Real-Time Water Level Predictions of Sewerage Systems-Covering Gauged and Ungauged Sites. *Hydrology and Earth System Sciences*.

Crotti, Giampaolo, Jorge Leandro, and Punit Kumar Bhola. 2019. A 2D Real-Time Flood Forecast Framework Based on a Hybrid Historical and Synthetic Runoff Database. *Water* 12 (1): 114. doi:10.3390/w12010114.

Denault, Catherine, Robert G. Millar, and Barbara J. Lence. 2006. Assessment of Possible Impacts of Climate Change in an Urban Catchment. *Journal of the American Water Resources Association* 42 (3): 685–697. doi:https://doi.org/10.1111/j.1752-1688.2006.tb04485.x.

Dietrich, J., A. H. Schumann, M. Redetzky, J. Walther, M. Denhard, Y. Wang, B. Pfützner, and U. Büttner. 2009. Assessing Uncertainties in Flood Forecasts for Decision Making: Prototype of an Operational Flood Management System Integrating Ensemble Predictions. *Natural Hazards and Earth System Sciences* 9 (4): 1529–1540. doi:10.5194/nhess-9-1529-2009.

DIN-EN 752. 2017. Entwässerungssysteme Außerhalb von Gebäuden – Kanalmanagement.

Djordjević, S., D. Prodanović, Č. Maksimović, M. Ivetić, and D. Savić. 2005. SIPSON – Simulation of Interaction between Pipe Flow and Surface Overland Flow in Networks. *Water Science and Technology* 52 (5): 275–283. doi:10.2166/wst.2005.0143.

Duncan, Andrew P., Albert S. Chen, Edward C. Keedwell, Slobodan Djordjević, and Dragan A. Savić. 2013. RAPIDS: Early Warning System for Urban Flooding and Water Quality Hazards. Conference Paper.

DWA, Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall, ed. 2006. *Hydraulische Bemessung und Nachweis von Entwässerungssystemen*. März 2006. DWA-Regelwerk, A 118. Hennef: DWA.

Falah, Fatemeh, Omid Rahmati, Mohammad Rostami, Ebrahim Ahmadisharaf, Ioannis N. Daliakopoulos, and Hamid Reza Pourghasemi. 2019. Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, 323–336. Elsevier. doi:10.1016/B978-0-12-815226-3.00014-4.

Fernando, Achela K., Xiujuan Zhang, and Peter F. Kinley. 2007. Combined Sewer Overflow Forecasting with Feed-Forward Back-Propagation Artificial Neural Network. *International Journal of Civil and Environmental Engineering* 1 (12): 7.

Fu, Guangtao, David Butler, Soon-Thiam Khu, and Si'Ao Sun. 2011. Imprecise Probabilistic Evaluation of Sewer Flooding in Urban Drainage Systems Using Random Set Theory. *Water Resources Research* 47 (2). doi:10.1029/2009WR008944.

Henonin, Justine, Beniamino Russo, Ole Mark, and Philippe Gourbesville. 2013. Real-Time Urban Flood Forecasting and Modelling – a State of the Art. *Journal of Hydroinformatics* 15 (3): 717–736. doi:10.2166/hydro.2013.132.

HIOS. 2019. Hinweiskarte Oberflächenabfluss Und Sturzflut. Chair of Hydrology and River Basin Management, Technical University of Munich. LookUpTable Stricklerwerte Überlandabfluss Tatsächliche Nutzung. Technical University of Munich.

Hsu, M. H., S. H. Chen, and T. J. Chang. 2000. Inundation Simulation for Urban Drainage Basin with Storm Sewer System. *Journal of Hydrology* 234 (1–2): 21–37. doi:10.1016/S0022-1694(00)00237-7.

Leandro, J., S. Djordjević, A. S. Chen, D. A. Savić, and M. Stanić. 2011. Calibration of a 1D/1D Urban Flood Model Using 1D/2D Model Results in the Absence of Field Data. *Water Science and Technology* 64 (5): 1016–1024. doi:10.2166/wst.2011.467.

Leandro, J., A. Gander, M. N. A. Beg, P. Bhola, I. Konnerth, W. Willems, R. Carvalho, and M. Disse. 2019. Forecasting Upper and Lower Uncertainty Bands of River Flood Discharges with High Predictive Skill. *Journal of Hydrology* 576 (September): 749–763. doi:10.1016/j.jhydrol.2019.06.052.

Leandro, J., J. P. Leitão, and J. L. M. P. de Lima. 2013. Quantifying the Uncertainty in the Soil Conservation Service Flood Hydrographs: A Case Study in the Azores Islands. *Journal of Flood Risk Management* 6 (3): 279–288. doi:10.1111/jfr3.12010.

Li, Jiada. 2020. A Data-Driven Improved Fuzzy Logic Control Optimization-Simulation Tool for Reducing Flooding Volume at Downstream Urban Drainage Systems. *Science of The Total Environment* 732 (August): 138931. doi:10.1016/j.scitotenv.2020.138931.

Lin, Qing, Jorge Leandro, Wenrong Wu, Punit Bhola, and Markus Disse. 2020. Prediction of Maximum Flood Inundation Extents With Resilient Backpropagation Neural Network: Case Study of Kulmbach. *Frontiers in Earth Science* 8 (August): 332. doi:10.3389/feart.2020.00332.

Maharjan, M., A. Pathirana, B. Gersonius, and K. Vairavamoorthy. 2009. Staged Cost Optimization of Urban Storm Drainage Systems Based on Hydraulic Performance in a Changing Environment. *Hydrology and Earth System Sciences* 13: 481–489. doi:10.5194/hess-13-481-2009.

Mark, O., C. Jørgensen, M. Hammond, D. Khan, R. Tjener, A. Erichsen, and B. Helwigh. 2018. A New Methodology for Modelling of Health Risk from Urban Flooding Exemplified by Cholera - Case Dhaka, Bangladesh. *Journal of Flood Risk Management* 11 (January): S28–S42. doi:10.1111/jfr3.12182.

Mark, O., S. Weesakul, C. Apirumanekul, S. Aroonnet, and S. Djordjevic. 2004. Potential and Limitations of 1D Modelling of Urban Flooding. *Journal of Hydrology* 299 (3–4): 284–299. doi:10.1016/S0022-1694(04)00373-7.

Martins, R., J. Leandro, and S. Djordjević. 2018. Influence of Sewer Network Models on Urban Flood Damage Assessment Based on Coupled 1D/2D Models. *Journal of Flood Risk Management* 11 (February): S717–S728. doi:10.1111/jfr3.12244.

Møller, Martin Fodslette. 1993. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks* 6 (4): 525–533. doi:10.1016/S0893-6080(05)80056-5.

Mounce, S. R., W. Shepherd, G. Sailor, J. Shucksmith, and A. J. Saul. 2014. Predicting Combined Sewer Overflows Chamber Depth Using Artificial Neural Networks with Rainfall Radar Data. *Water Science and Technology* 69 (6): 1326–1333. doi:10.2166/wst.2014.024.

Qin, Hua-peng, Zhuo-xi Li, and Guangtao Fu. 2013. The Effects of Low Impact Development on Urban Flooding under Different Rainfall Characteristics. *Journal of Environmental Management* 129 (November): 577–585. doi:10.1016/j.jenvman.2013.08.026.

Quintana-Romero, Teresa, and Jorge Leandro. 2022. A Method to Devise Multiple Model Structures for Urban Flood Inundation Uncertainty. *Journal of Hydrology* 604 (January): 127246. doi:10.1016/j.jhydrol.2021.127246.

Rjeily, Yves Abou, Oras Abbas, Marwan Sadek, Isam Shahrour, and Fadi Hage Chehade. 2017. Flood Forecasting within Urban Drainage Systems Using NARX Neural Network. *Water Science and Technology* 76 (9): 2401–2412. doi:10.2166/wst.2017.409.

Rossman, Lewis A. 2017. Storm Water Management Model Reference Manual Volume II – Hydraulics.

Rubinato, M., R. Martins, and J. D. Shucksmith. 2018. Quantification of Energy Losses at a Surcharging Manhole. *Urban Water Journal* 15 (3): 234–241. doi:10.1080/1573062X.2018.1424217.

Schmitt, T., M. Thomas, and N. Ettrich. 2004. Analysis and Modeling of Flooding in Urban Drainage Systems. *Journal of Hydrology* 299 (3–4): 300–311. doi:10.1016/S0022-1694(04)00374-9.

Tsihrintzis, Vassilios A., and Rizwan Hamid. 1998. Runoff Quality Prediction from Small Urban Catchments Using SWMM. *Hydrological Processes* 12 (2): 311–329. doi:10.1002/(SICI)1099-1085(199802)12.

Wetterhall, F., F. Pappenberger, L. Alfieri, H. L. Cloke, J. Thielen-del Pozo, S. Balabanova, J. Daňhelka, et al. 2013. HESS Opinions: Forecaster Priorities for Improving Probabilistic Flood Forecasts. *Hydrology and Earth System Sciences* 17 (11): 4389–4399. doi:10.5194/hess-17-4389-2013.

Wunsch, Andreas, Tanja Liesch, and Stefan Broda. 2018. Forecasting Groundwater Levels Using Nonlinear Autoregressive Networks with Exogenous Input (NARX). *Journal of Hydrology* 567 (December): 743–758. doi:10.1016/j.jhydrol.2018.01.045.

A feature-informed data-driven approach for predicting maximum flood inundation extends

geosciences

MDPI

Article

A Feature-Informed Data-Driven Approach for Predicting Maximum Flood Inundation Extends

Felix Schmid and Jorge Leandro *

Chair of Hydromechanics and Hydraulic Engineering, Research Institute Water and Environment, University of Siegen, 57076 Siegen, Germany; felix.schmid@uni-siegen.de
* Correspondence: jorge.leandro@uni-siegen.de

Abstract: As climate change increases the occurrences of extreme weather events, like flooding threaten humans more often. Hydrodynamic models provide spatially distributed water depths as inundation maps, which are essential for flood protection. Such models are not computationally efficient enough to deliver results before or during an event. To ensure real-time prediction, we developed a feature-informed data-driven forecast system (FFS), which interpreted the forecasting process as an image-to-image translation, to predict the maximum water depth for a fluvial flood event. The FFS combines a convolutional neural network (CNN) and feature-informed dense layers to allow the integration of the distance to the river of each cell to be predicted into the FFS. The aim is to ensure training for the whole study area on a standard computer. A hybrid database with pre-simulated scenarios is used to train, validate, and test the FFS. The FFS delivers predictions within seconds making a real-time application possible. The quality of prediction compared with the results of the pre-simulated physically-based model shows an average root mean square error (RMSE) of 0.052 for thirty-five test events, and of 0.074 and 0.141 for two observed events. Thus, the FFS provides an efficient alternative to hydrodynamic models for flood forecasting.

Keywords: flood forecasting; inundation forecasting; artificial neural network; convolutional neural network; real-time forecasting

check for updates

Citation: Schmid, F.; Leandro, J. A Feature-Informed Data-Driven Approach for Predicting Maximum Flood Inundation Extends. *Geosciences* **2023**, *13*, 384. https://doi.org/10.3390/geosciences13120384

Academic Editors: Sabina Porfido and Jesus Martinez-Frias

Received: 18 October 2023
Revised: 4 December 2023
Accepted: 12 December 2023
Published: 14 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Flood events are one of the most common natural disasters posing a threat to human lives and society. In 2021, global economic losses caused by flooding amounted to USD 82 billion, the highest among all natural catastrophes [1]. This amount is about to rise in the upcoming years, as climate change increases the likelihood of the occurrences of extreme weather events, causing even more severe floods [2–4]. As a non-structural measure, flood forecasting can reduce both economic losses and loss of life in all types of floods by providing information about the impending flood wave so that authorities and society are better prepared for this natural disaster [5]. Besides this hydrological forecast, which focuses on predicting the flood hydrographs, an additional and important information is the flood inundation extent [6]. The aim is to predict the inundation extent based on the flood hydrograph by providing a map with spatially distributed water depths in the respective study site. Such information makes it possible to quickly identify risk zones, adapt and optimize rescue routes and strategies in terms of traffic, and allow spatial-specific warnings and evacuations.

Flood inundation maps are commonly obtained using physically based numerical models, which solve the shallow water equations on a discretized grid over a study area. This approach allows a good spatial resolution for the water depth calculation. However, physically based models are computationally expensive, making them not suitable for real-time flood forecasting [7]. Providing enough lead time to prepare mitigation measurements is crucial for protecting the population and reducing the damage. Hence, alternative

approaches like the applications of data-driven models gained immense attention in recent years [8].

The key idea behind data-driven is to take advantage of the much better computation efficiency of these approaches. Herein, pre-simulated hydrodynamic scenarios are used as training data to allow the model to predict inundation maps. In the best configuration, these maps match those of the hydrodynamic model. Generally, this approach allows for the prediction of the maximum water depth for a flood event in the study area, while maintaining the resolution of the hydrodynamic model. Thus, the hydrodynamic model can be replaced by the data-driven model. In recent years, several studies have been presented using this methodology [9–15]. Some of these studies use artificial neural networks (ANNs). ANNs take one-dimensional input data and feed the information through a network of interconnected units (neurons) organized into layers. By adjusting the weights within the neurons based on the difference between predicted and actual outputs during the iterative training process, they learn to capture complex patterns in the data and make predictions on new examples [16]. For an inundation forecast, it is often not suitable to keep the total number of cells to be predicted within a single ANN, studies additionally discretize the study area by multiple ANNs [9,11]. This approach in turn makes training and optimizing a computational and challenging task.

Other studies however make use of different types and architectures of data-driven models such as convolutional neural networks (CNN) [10,12–15]. CNNs are specifically designed for processing structured grid-like data, such as images. They have been widely successful in various computer vision tasks, including image classification, object detection, and image segmentation [17,18]. The key idea behind CNNs is the use of multiple convolutional layers. A convolutional layer applies a set of learnable filters (also known as kernels) to the input image. Each filter slides over the image in a window (matrix) and performs element-wise multiplication and summation to produce a feature map. The purpose of this operation and the feature map is to capture and store local patterns, characteristics, and properties of the image. Typically, after convolutional layers, a pooling layer is often applied. This pooling layer is responsible for reducing the spatial dimensions of the feature maps while preserving the most important information. Max-pooling is commonly used for such an operation, which selects the maximum value within the window and discards the rest. Pooling helps to make the representations more robust to small spatial translations and reduces the number of parameters in the network, making training time and prediction performance more effective [16,17].

The latest studies, besides the CNN structure, used physical information like, for example, the digital elevation model or roughness values as additional inputs to the data-driven model [13,15,19]. The idea is the additional data provided will give the data-driven model a deeper knowledge about the flood inundation to improve performance. However, this approach involves a huge amount of data, so again discretization of the study area with multiple CNNs is required, and training on standard computers, such as the ones often used by local authorities, is not applicable. Therefore, access to a data center is necessary.

The present study focuses on the challenge of developing a feature-informed data-driven flood inundation forecast system (FFS) for fluvial events for real-time predictions. To make use of the robust prediction and training efficiency, the FFS is based on a convolutional architecture. Furthermore, the study builds on the idea of providing additional knowledge to the data-driven model. Unlike previous studies, our aim is to ensure the training for the whole study area on a standard computer. Therefore, we developed an innovative feature-informed strategy and a new optimization workflow, to integrate (prior) additional knowledge about the flooding into the network, without providing additional multiple layers (GIS-data) as input as in Löwe et al., 2021 [19].

2. Materials and Methods

2.1. Framework of the Feature-Informed Forecast System (FFS)

The challenge of developing the FFS for fluvial events is achieved by the feature-informed data-driven approach, which interprets the forecasting as an image-to-image procedure. The FFS collects hydrographs time series at the entrance locations from three rivers flowing into the study site, converts these time series into images, and feeds them as input through the data-driven model. In our case study, the hydrographs have a maximum length of 50 h and an hourly resolution. This could be increased or reduced for other case studies. The data-driven model translates these images into the respective inundation map using our feature-informed strategy with the convolutional architecture. The result is the maximum inundation extends and depths within a study area resulting from the three hydrographs. Alike other studies [9,11,14], the FFS is trained, validated, and tested with pre-simulated hydrodynamic scenarios, specifically with a hybrid database including syntactic, real-based events, and historical observed events. Figure 1 shows the framework of the FFS which can be split into the following three steps.

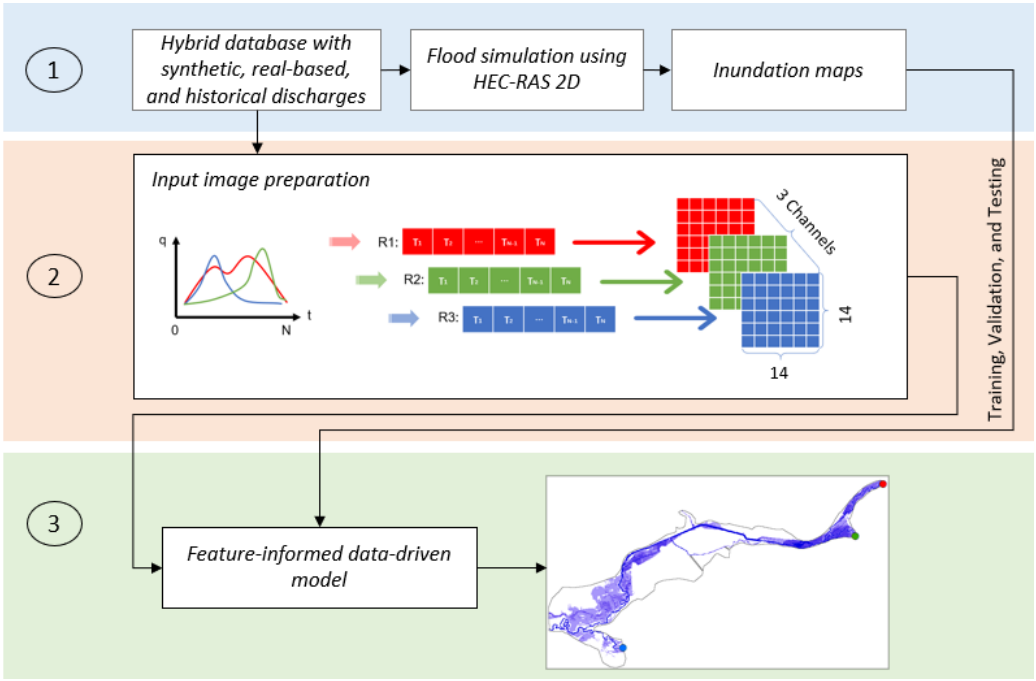


Figure 1. The framework of the forecast system (FS). The numbers on the right side refer to the explanation in the numbered list below.

1. The hybrid database of this study was generated by Crotti et al. [6] (see Section 2.6). The hydrological model LARSIM (Large Area Runoff Simulation Model) [20] was used by the authors to generate synthetic, real-based, and historical discharge scenarios for three rivers, which were flowing into and through the study area. These scenarios were simulated with the hydrodynamic model HEC-RAS 2D (Hydrologic Engineering Center—River Analysis System, Davis, CA, USA) [21] to obtain the corresponding maximum water levels in the catchment area as an inundation map.
2. As we follow an image-to-image strategy as a forecasting process, the time series of the hydrographs have to be converted so they can be presented to the feature-based data-driven model as images, which are later then translated into the corresponding inundation map. Therefore, the time series, are converted into a tensor with the size $14 \times 14 \times R$, where R describes the numbers of hydrographs. In our study, R is equal to three. Nevertheless, other study areas may require different values

of R. The 14 × 14 size, is the minimum size, necessary to make effective use of the convolutional operations. Smaller sizes lead to a 1 × 1 image size during the convolutional downsampling operations. Larger image sizes can be used without loss of accuracy or relevance of the modeling. For filling the image cells, the first timestep of the series is then placed into position 1 × 1 × 1, where the second timestep is placed into 1 × 2 × 1, and so on until the entire hydrograph is used up. A similar approach is used in the study of Kimura et al., 2019 [10];

3. The images from stage 2 are then fed into the feature-informed data-driven model, described in Section 2.2. This model translates these images into the upcoming inundation map. The architecture allows the integration of spatial and physical prior knowledge into the network. The model is trained, validated, and tested with the hybrid database from step 1, whereby the resolution of the prediction within the study area is given by the resolution of the respective hydrodynamic model used for generating the database. After training and validation, the data-driven model is able to replace the 2-dimensional flood model and can be used for real-time forecasting.

2.2. Feature-Informed Data-Driven Approach

The feature-informed data-driven model is responsible for generating an inundation map. Figure 2 shows the architecture of the model.

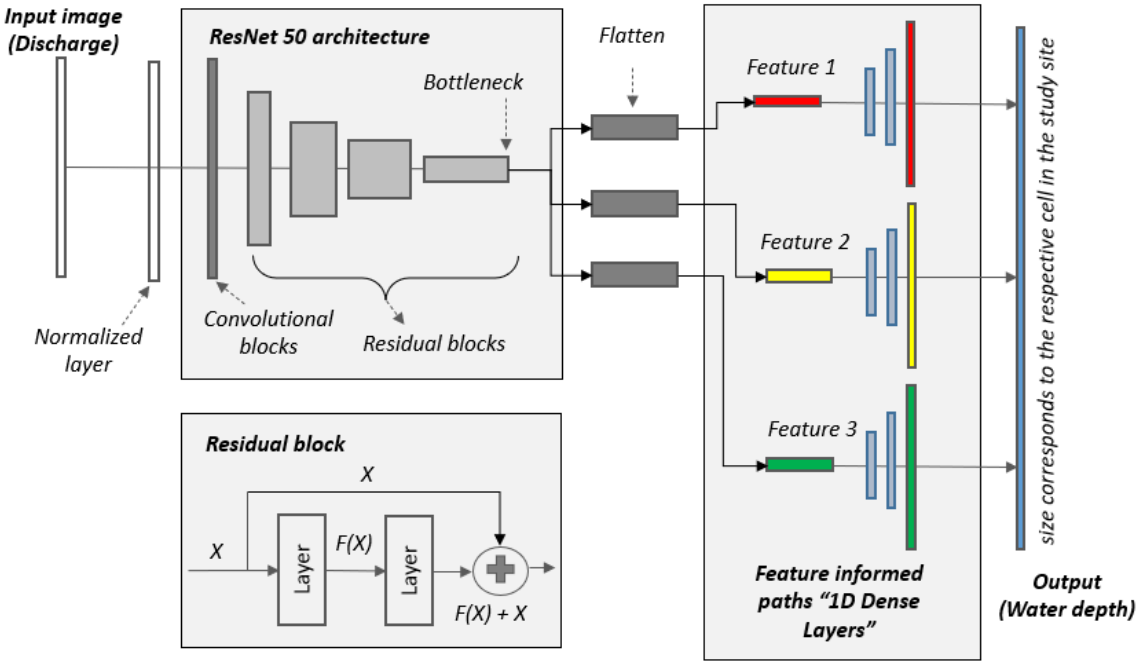


Figure 2. The architecture of the feature-informed data-driven model.

The novelty of this architecture is the feature-informed part. This part allows, in combination with our optimization workflow, the integration of additional information into the model without providing new or different datasets. The overall idea of this model is the image-to-image strategy and consists of two parts:

- residual Convolutional Neural Network with 50 layers (ResNets50),
- the feature-informed Dense Layers.

Before the images are presented to the model, the dataset needs to be normalized. A normalization layer that scales the data in a range from 0 to 1 is therefore applied. Afterward, the images are processed through the ResNet50 [22]. The idea behind deploying a convolutional architecture is to make effective use of its ability to extract useful patterns,

properties, and characteristics of the image. The ResNet50 is a special type of convolutional network. This type of network works by using skip connections, also known as residual connections, to enable the training of very deep neural networks. After a first/initial convolutional block, four residual blocks are followed. A residual block contains convolutional layers and a skip connection that adds the original input back to the output of the convolutional layers. The skip connection helps the network to learn the difference (residual) between the input and convolutional layer output, allowing it to refine features effectively [22]. Mathematically the residual block can be described as follows:

$$y = \mathrm{F}(x, \{W_i\}) + x \tag{1}$$

where x and y are the input and output data of the residual blocks. The function $\mathrm{F}(x, \{W_i\})$ represents the residual mapping to be learned, while W is the respective convolutional operation [22].

We applied the classical ResNet50 architecture until the flattened layer. As activation functions within the convolutional operations, we used leaky ReLu. The activation function is used to determine whether the information is important. For a detailed ResNet description refer to He et al. [22].

After the final residual block, the extracted features are stored in the so-called bottleneck layer, this layer can be seen as compressed and dense information of the input image. After the flattening operation, which generates a one-dimensional vector out of the bottleneck layer, the information is used to rebuild a new image by the feature-informed dense layers. A dense layer, also known as a fully connected layer, consists of multiple neurons. Each neuron in the layer is connected to every neuron in the previous layer. Mathematically the output (out) of a neuron is given by:

$$\mathrm{out} = f_t(\mathrm{i} \times \mathrm{w} + \mathrm{b}) \tag{2}$$

where i is the input and information from the flattened bottleneck, w is the vector of weights, b is a neuron-specific bias value, f_t is the transfer function to control whether the information of the neuron is important and to consider, and out is the output of the neuron. In this study, the sigmoid transfer function for the first two dense layers and the linear function are applied. This is a common application for dense layers [23,24]. The output from the feature-informed dense layers is the predicted inundation image/map. Thus, the size of the neurons in the last dense layers matches exactly the number of cells to be predicted and a cell in the image can be interpreted by the output of a neuron.

The feature-informed strategy groups the information from the bottleneck into three paths. Each path consists of three dense layers and represents one feature. By creating three paths we focus the forecast of the cells within each feature, based on the information contained in each corresponding area. This allows the integration of additional knowledge into the network. For the sake of testing, we investigate in our study the use of three features. Since cells closer to the river are likely to be flooded first than cells far away from the river, features are defined herein as the distance of each cell to be predicted to the river. Therefore, we decided to integrate cells within features according to the distance to the river:

1. feature 1: cells that are in a range of 150 m to a river;
2. feature 2: cells that are in a range of 150 to 400 m to a river;
3. feature 3: cells that are farther than 400 m away from a river.

By grouping the cells into these features, the neuron in each path generates cells of the inundation map that share the same inundation pattern, and thus these neurons have a specific knowledge about the position of the cell to be predicted in the study area. Other studies must define the number of features and output size of the dense layers before setting up the FFS. Also, a GIS analysis is necessary to classify the cells into these three features. Herein, we decided to use three features after a trial-and-error procedure.

2.3. Optimization and Ensemble Approach

By adjusting the feature maps in the ResNet50 architecture and the weights and biases in the feature-informed part, the data-driven model can be trained to optimize its prediction. The objective is to find the water depth predictions that agree the most with the physically based inundation maps. The agreement and thus quality are defined by the loss function (*LF*).

We applied the mean squared error (*MSE*) as *LF*, where: *P* are the predicted values, *T* the observed values, and *n* is the number of predictions:

$$LF = MSE = \frac{1}{n}\sum_{i=1}^{n}(P - T)^2 \quad (3)$$

During the training procedure, a minimum value of this *LF* is searched by the optimization algorithm. We used Adams' optimizer, which is typically used and well-known [25]. To ensure an optimal prediction/minimum error, we developed a novel workflow for the training procedure. The workflow is shown in Figure 3 and consists of 2 algorithms. In the first algorithm, the complete feature-informed data-driven model is trained on the training data. To set the focus of the prediction based on the features, after training we first freeze all weights and feature maps within the FFS, and then we generate three retraining subsets. Each subset includes the input discharge images (input) and the cells (true output) of each respective feature. After that, algorithm 1 is complete, providing as output a trained FFS and three retraining subsets.

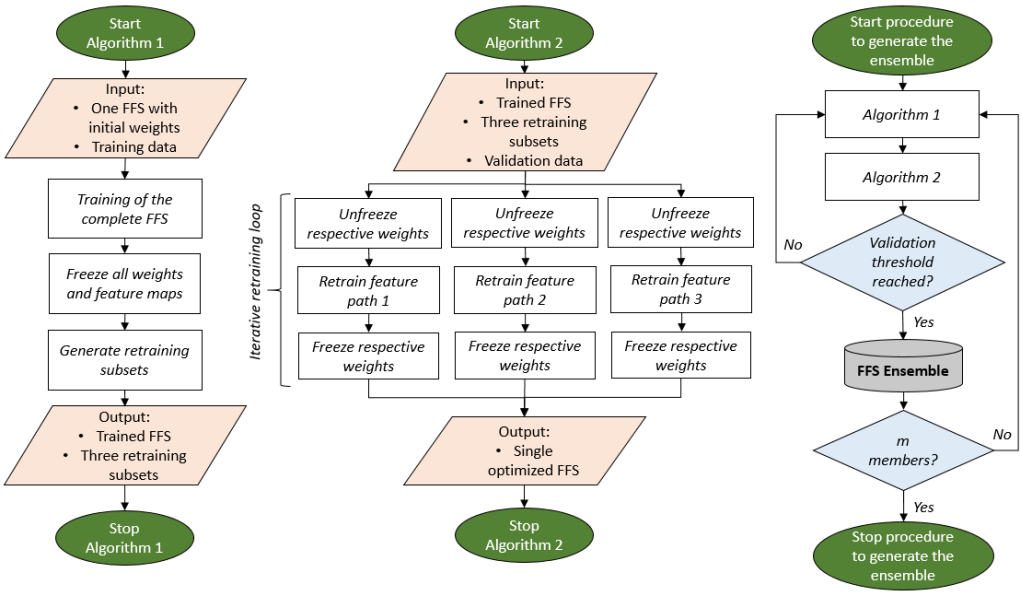


Figure 3. Workflow for the optimization and the generation of the ensemble with m members.

Algorithm 2 uses the outputs of algorithm 1 and additionally the validation data as inputs. First, an iterative retraining loop around the three features is applied. As such the FFS is retrained separately for each feature path with the respective subset, while only the weights in the respective feature path are optimized. Therefore, only the weights of the corresponding feature path (either 1, 2, or 3) must be unfrozen in the path to be trained and have to be frozen again after training. This approach is necessary since algorithm 1 trains all features at once and therefore it is not possible to set a special focus on a single feature (and gain the optimal weights within this feature). Hence, to make sure the weights in each feature path are optimized to describe the flood depths and characteristics of the respective feature in the best possible way, this retraining strategy is applied. The output of algorithm 2 is then a single optimized FFS. This concludes algorithm 2.

It is common to use an ensemble approach for accounting for uncertainties in the forecast. Herein, and to provide multiple equally likely members (m) within the ensemble, the steps tasks of algorithms 1 and 2 are repeated. With this procedure, the final water depth for each cell to be predicted is given by the average value calculated by the members of the ensemble. This procedure makes it possible to capture the uncertainty caused by the random weight initialization at the start of the training [9,26]. A similar approach is used by Berkhahn and Neuweiler (2019) [9] and by Schmid and Leandro (2023), to account for uncertainty [24].

To ensure equal probability among the members while maintaining high prediction quality, a member is accepted only if the prediction error on the validation data is less than five centimeters. This is quantified by the root mean squared error (RMSE) (see Equation (4) in Section 2.4), which takes the square root of the LF. The procedure is sequential and starts with algorithms 1 and 2. After, it checks if the single optimized FFS reaches the validation threshold (RMSE below 0.05 m). If that is not the case, the optimized FFS is dismissed and algorithms 1 and 2 are repeated. If the validation threshold is reached, an ensemble member is accepted. A final condition checks whether the ensemble already consists of (at least) the desired members. The procedure is stopped if this condition is met. Otherwise, algorithms 1 and 2 are repeated. The model was trained on an NVIDIA Quadro M4000 graphics card with a memory of 8 GB. Such a card (or similar) is found commonly in many PCs. Python 3.9 and Tensorflow 2.6 were used for the development of the FFS.

2.4. Model Evaluation Criteria

To assess the performance of the FFS, two criteria are used, namely the mean squared error (MSE) (Equation (3)) and the root mean squared error (RMSE) (Equation (4)). The MSE is used to compare and validate the innovative methodology with the results obtained by Lin et al., 2020 [11]. The RMSE is used because the results are more intuitive to interpret as the error value has the same units as the water level. The observed values are taken as the results from the 2D hydrodynamic model HEC-RAS. The evaluation criteria are calculated for each event and each cell in the study area, resulting in a map representing the spatial distribution of the prediction error. Additionally, the average value is calculated, by dividing the RMSE by the number of cells to be predicted, to obtain a value for the entire study area.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P-T)^2} \quad (4)$$

where: P is the by the forecast system predicted value, T is the observed value, n is the number of predictions.

2.5. Study Site

The study area, Kulmbach (Figure 4), is located in Bavaria, Germany. The area has a size of about 12 km^2 and nearly 27,000 people live in the area. The town center is located in the middle of the area. Three main rivers flow through the study area, namely the Schorgast, the Weißer Main, and the Roter Main. The flow direction is from northeast to southwest, where the area is narrow and surrounded by hills in the northeast. In the eastern part, the study site turns wider into a river floodplain.

As described in Section 2.2, the feature of the network is represented by the distance to the river lines. Therefore, for each cell to be forecasted, a GIS analysis was performed to calculate the distances to the rivers. The resulting distances were then classified into the three classes. Figure 5 shows this classification. Subsequently, the classes and respective cells were assigned to the three different paths within the network. The number of outputs within the last dense layers matches the number of cells of each class.

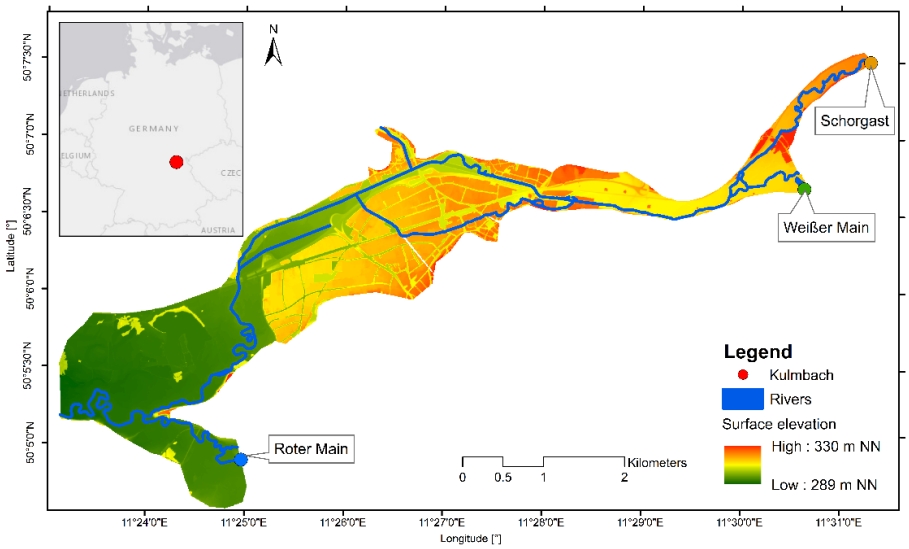


Figure 4. The study site: Kulmbach. The subplot in the upper left shows the location of Kulmbach within Germany.

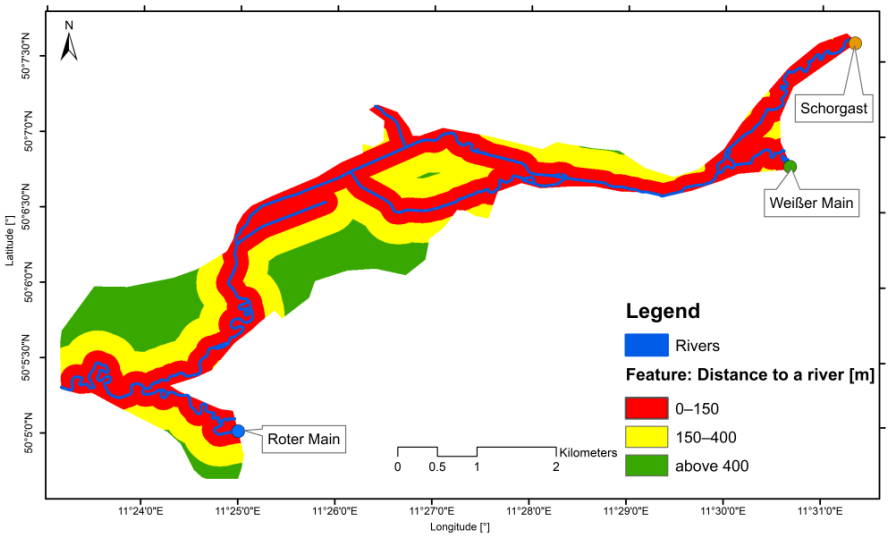


Figure 5. Visual representation of the integrated feature: distance to the rivers.

2.6. Dataset for Training, Validation, and Testing the Forecast System

A hybrid database containing 270 different events with their associated discharges and corresponding inundation maps is used to develop and test the FFS. The database was generated by Crotti et al. [6]. The authors have used a two-step procedure in which, first, the discharges for the three main rivers are generated by applying the hydrological model LARSIM (Large Area Runoff Simulation Model) [20], and second, the generated discharges are simulated with the hydrodynamic model HEC-RAS 2D [21] to obtain the corresponding maximum water levels in the catchment area as a map. The validation of the inundation model was performed by Bhola et al. in 2018 [27]. The hydrographs have an hourly resolution and 50 timesteps. To shape these hydrographs into the input tensor of $14 \times 14 \times 3$, the 50 timesteps are extrapolated to fit into this frame. The maps are projected to a spatial resolution of 4×4 m, resulting in approximately 500,000 cells. One half (135 events) of the database consists of synthetic events produced by using return periods (durations: 5, 30, 90, 360, 720, and 1140 min and periods: 1, 50, 100, 200, 300, and 1000 years),

and the other half (135 events) consists of real-based events. The real-based events are generated by scaling individual observed discharge events from time series analyses from 1970 to 2017 of the three rivers to higher return periods so that flooding occurred. The remaining two events are historically observed events from the years 2005 and 2013, which lead to major flooding within the study area. Figure 6 shows these hydrographs.

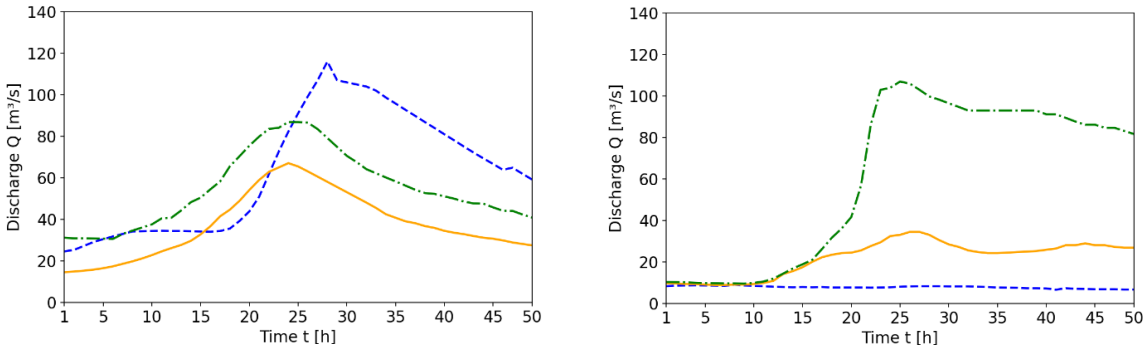


Figure 6. Observed hydrographs time series. Orange (continuous) represents the Schorgast, green (dashed-dotted) the Weißer Main, and blue (dotted) the Roter Main: (**a**) hydrograph from the year 2005; (**b**) hydrograph from the year 2013.

The database (excluding the historical events) is randomly split into the three subsets training (70%), validation (15%), and testing (15%). The two observed historical events are used to test the FFS on real observed data. For further details on generating the hybrid database please refer to Crotti et al. [6]. We like to state the main focus of the study is the development of a data-driven approach. The validation and setup of the flood inundation model was carried out by Bhola et al., 2018 [27]. Hence grid size, hydrodynamic parameters, rainfall duration, and intensity were already given. Our methodology is in any case flexible and can be applied to any database or flood model in which parameters and rainfall events of interest may differ.

3. Results

3.1. Training Durations and Validation Accuracy

The application of the developed optimizing workflow (Figure 3) on the feature-informed architecture resulted in a candidate for the ensemble. Validation was based on a set of 40 random events from the hybrid database and as described in Section 2.3 we applied an ensemble strategy to consider uncertainty. In case the output of algorithm 2 reaches the validation threshold of 0.05 m, the candidate is added as a member of the ensemble. We set the number of members to three (m = 3). In our study, the workflow (Figure 3) was repeated 17 times until the final ensemble was reached. Table 1 shows the results of the first three integrated ensemble members and the best four candidates overall as well as the respective training times.

Table 1. Training time and validation accuracy for the feature-informed forecast data-driven model, for the top four procedures. Only the first three are selected for the ensemble. Procedure 4 is the first to go over the selection threshold.

	Validation Accuracy [RMSE]	Training Time [min]
Output of algorithm 2: 1	0.039 m	126
Output of algorithm 2: 2	0.042 m	131
Output of algorithm 2: 3	0.048 m	116
Output of algorithm 2: 4	0.051 m	125

3.2. Performance of the Feature-Informed Forecast System

3.2.1. Overall Performance

To evaluate the overall performance of the developed FFS, the average RMSE values are calculated. Table 2 shows the results of the FFS on the test dataset and the two observed events. The FFS delivers very good results, achieving the aim of a high prediction quality. For the 35 test events, the FFS makes an average prediction error of only 5 cm for each cell in the entire study area. The results of the observed events also show a good and sufficient prediction quality, as they are under 10 cm and 15 cm errors, respectively.

Table 2. The average accuracy of the forecast system on the test dataset and the observed events.

	Accuracy [RMSE]
Test dataset (35 events)	0.052 m
Observed event 2005	0.074 m
Observed event 2013	0.141 m

Besides the high prediction performance, the FFS delivers these predictions within 19 s (the pure prediction of a member is 3 s).

3.2.2. Comparison of the Individual Feature Performance on the Observed Events

We investigate the performance of the individual features on the observed events. Table 3 summarizes these results. It can be detected feature 2 performs nearly exactly like the overall average RMSE, whereas feature 1 shows around 40% better, and feature 3 40% worse results than the overall average prediction RMSE.

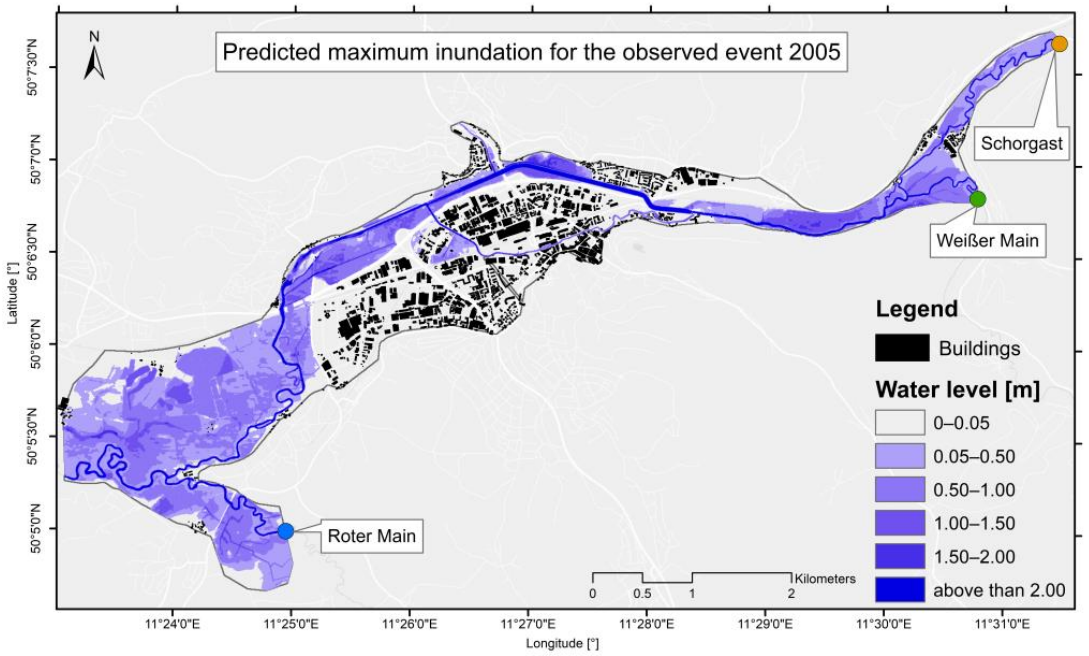
Table 3. Comparison of the average MSE performance of the individual feature.

	RMSE of Feat. 1 (Cells Up to 150 m Away)	**RMSE of Feat. 2 (Cells Up to 400 m Away)**	**RMSE of Feat. 3 (Cells > 400 m Away)**
Observed event 2005	0.063 m	0.070 m	0.078 m
Observed event 2013	0.101 m	0.141 m	0.152 m

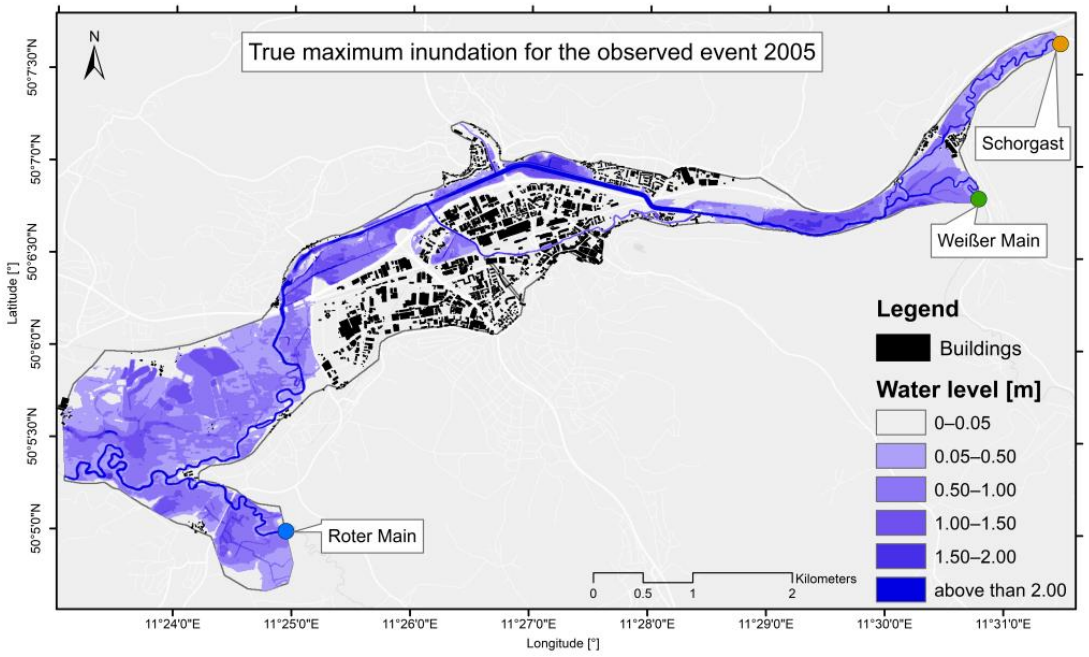
3.2.3. Inundation Maps and Spatially Distributed Prediction Error of the Observed Events

The inundation maps of the observed events from the years 2005 and 2013 are presented in Figures 7a and 8a. Additionally, the respective true inundation gained by the HEC-RAS 2D model is visualized (Figures 7b and 8b). It is clearly visible both maps (predicted and true) are very similar.

To further investigate the performance of the observed events, the MSE value is calculated for each cell to visualize the spatial prediction error. Since Lin et al., 2020 also developed a data-driven approach following multiple ANNs for predicting the inundation in the same study area, we can compare the results of their MSE maps with ours [11]. Figures 9 and 10 visualize this comparison. It is seen our feature-informed architecture produces better predictions.

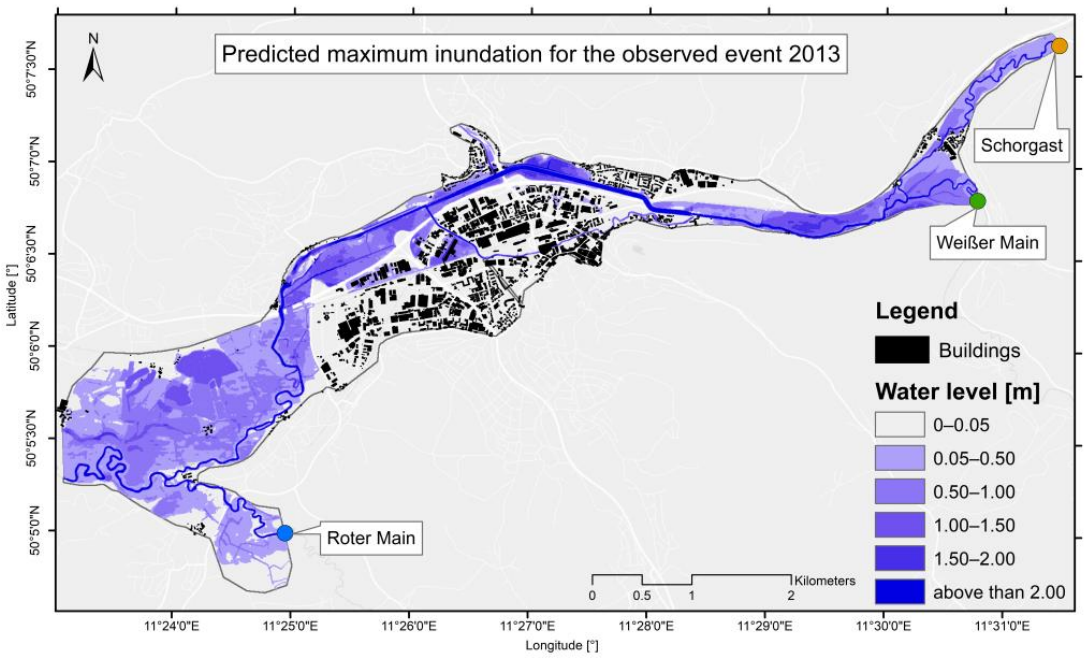


(**a**)

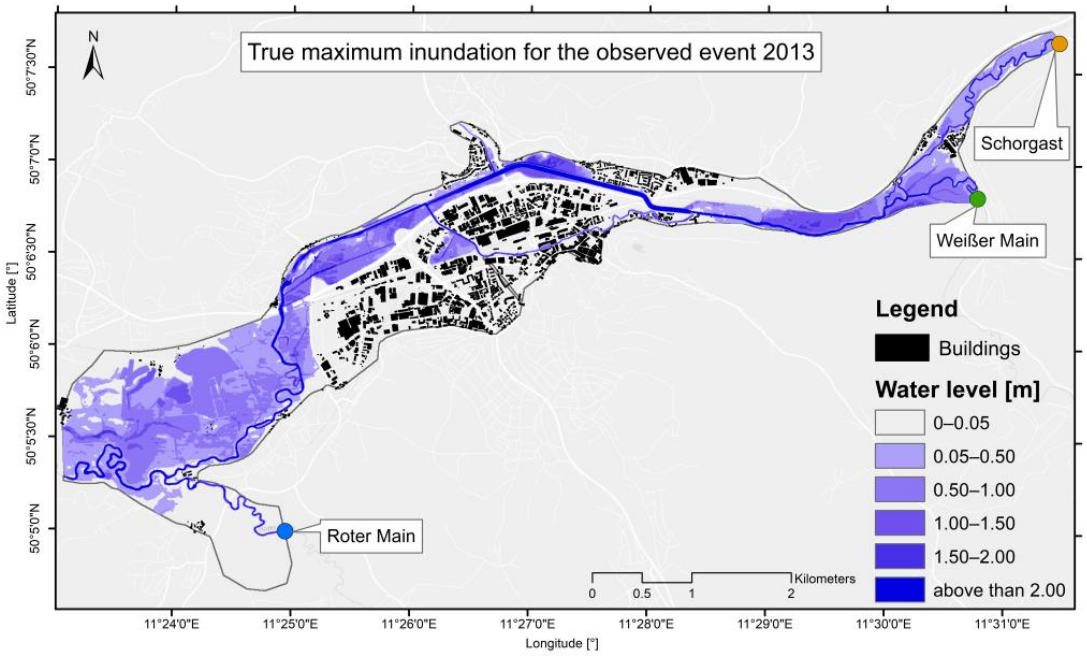


(**b**)

Figure 7. Inundation maps for the observed flood event from 2005. The input hydrographs are shown in Figure 6a: (**a**) prediction of the feature-informed forecast system; (**b**) true inundation delivered by the physically based model HEC-RAS 2D.

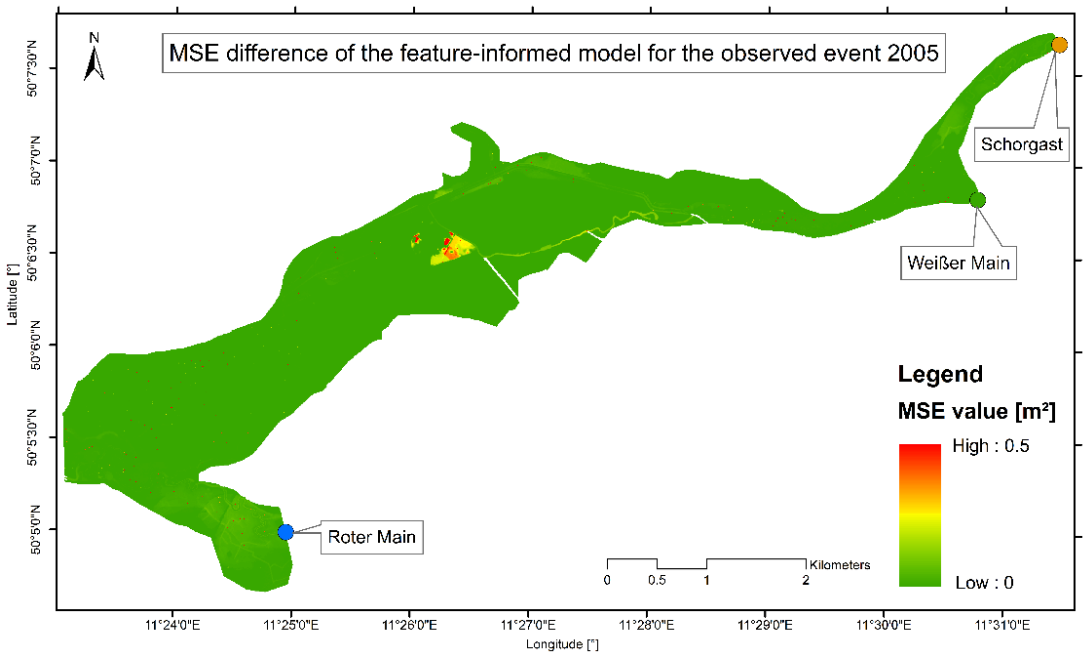


(**a**)

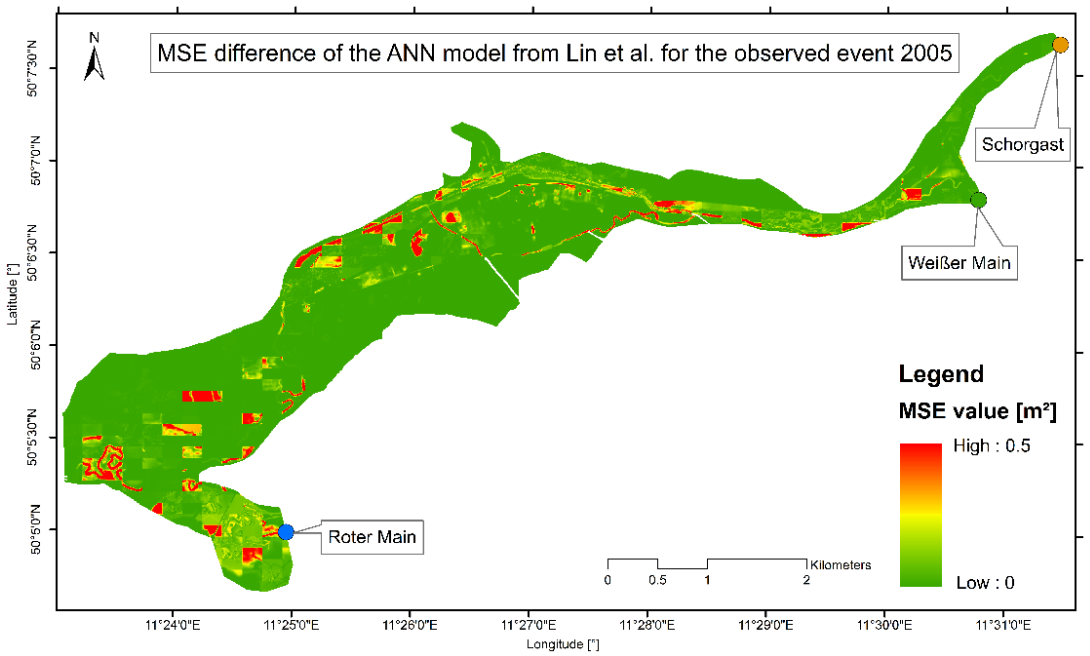


(**b**)

Figure 8. Inundation maps for the observed flood event from 2013. The input hydrographs are shown in Figure 6b: (**a**) prediction of the feature-informed forecast system; (**b**) true inundation delivered by the physically based model HEC-RAS 2D.

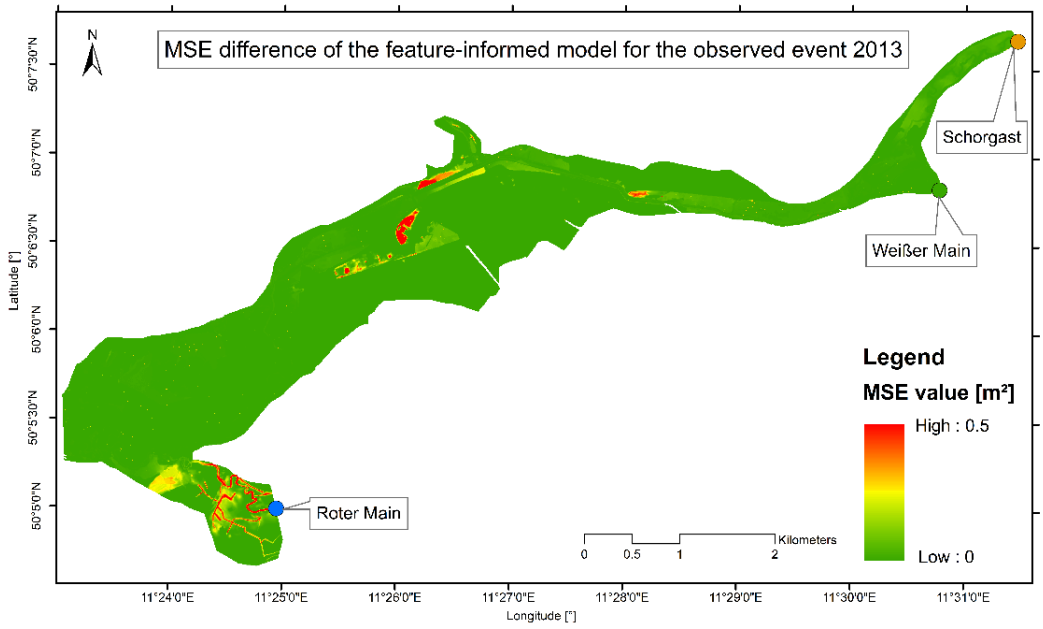


(**a**)

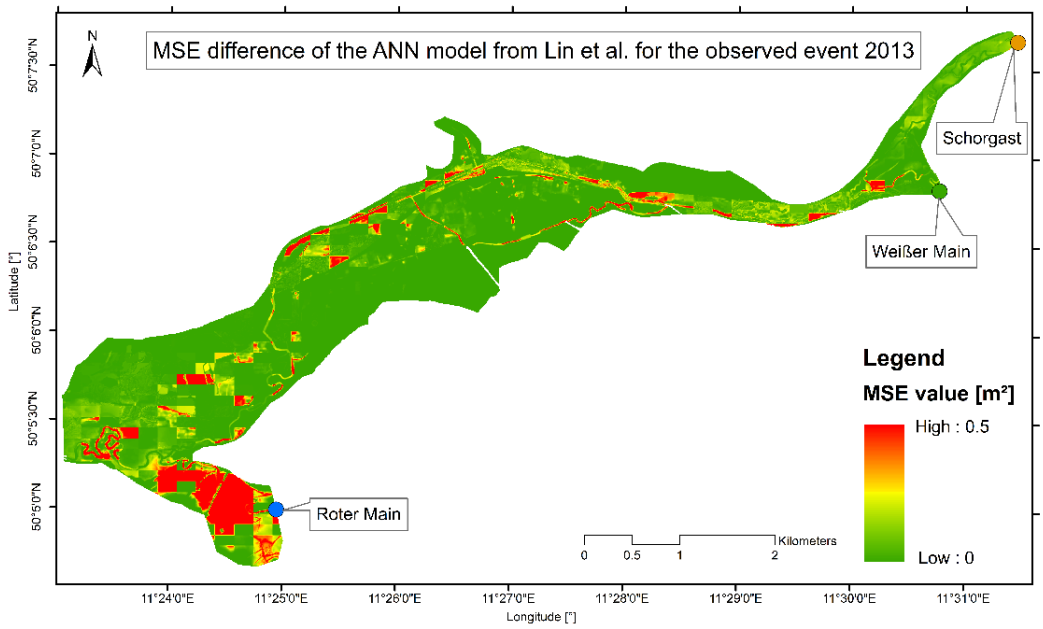


(**b**)

Figure 9. MSE comparison of the ANN approach from Lin et al., 2020 [11] and our feature-informed approach. The MSE is calculated by the difference between the neural approach and the hydrodynamic model of the observed event in 2005: (**a**) MSE difference of the feature-informed forecast system; (**b**) MSE difference of the ANN approach from Lin et al., 2020 [11].



(**a**)



(**b**)

Figure 10. MSE comparison of the ANN approach from Lin et al., 2020 [11] and our feature-informed approach. The MSE is calculated by the difference between the neural approach and the hydrodynamic model of the observed event in 2013: (**a**) MSE difference of the feature-informed forecast system; (**b**) MSE difference of the ANN approach from Lin et al., 2020 [11].

4. Discussion

4.1. Training Durations and Validation Accuracy

The results presented in Table 1 are very good. This is due to the newly developed optimization workflow. The iterative training process ensures every path is optimized

separately and specialized to its respective feature. This procedure ensures the integration of additional knowledge (in our case the distance to the river) into the model. Furthermore, by following this approach there is no need to present the network with an additional GIS-layer dataset like [19], and on the other hand, there is no need to make compromises in the prediction quality since flood-influencing data are missing. This also allows training the model on a standard computer and predicting the entire study area using a single data-driven model. Besides that, the effective computational times for the training duration are attributed to the use of a ResNet50 in the classification part, as this is a convolutional network. This was responsible for the short training durations and thus made the optimization of the model more efficient. Therefore, it was possible to consider the uncertainty origin from the initial weights and integrate a threshold value for the validation, which in turn improved the prediction quality. The selection of a small number of the total members of the ensemble is done to ensure the time required for a single prediction is feasible on a standard computer and therefore is kept within a few seconds. Nevertheless, our workflow and methodology are also applicable to high-performance computing facilities for which the number of members does not impact the performance of the FFS.

4.2. Performance of the Feature-Informed Forecast System

4.2.1. Overall Performance

Table 2 shows the overall performance results of the FFS to the test dataset and the two observed events. These very good performances highlight the effectiveness of the image-to-image translation strategy of the feature-informed architecture (Figure 2) and the developed optimization workflow (Figure 3). The average RMSE values for our test data are better than those of Lin et al., 2020 (RMSE value of 0.316) [11]. The performances of the FFS for the observed events are of sufficient quality as well. The slightly worse performance for the event from 2013 can be explained by the lack of similar hydrograph shapes found in the training set. In the training set, all three rivers contribute to the flood volume (as well as rising or falling limbs), which is in contrast to the hydrograph shapes from the 2013 event (Figure 6b). In the latter, only the Schorgast rises rapidly and remains high throughout the forecast period. The integration of more events like this into the training set would enhance, therefore, the performance of the FFS also in this event. Besides the performance quality, the prediction time of 19 s makes the FFS suitable for real-time applications.

4.2.2. Comparison of the Individual Feature Performance on the Observed Events

Table 3 shows a 40% worse performance of feature 3. This can be explained since these are the cells the furthest away from the river (compare Figure 5). They are only inundated during relatively higher discharge events. The FFS could benefit from more scenarios being included in the training data set, for the relationship between inundation and discharges to be better captured. On the other hand, Table 3 shows a 40% better performance for feature 1, which represents the cells closest to the river with the highest inundation.

The overall and individual performance of each feature leads us to conclude splitting the cell into the distance from the river and providing this information feature-wise to the data-driven model is a good strategy.

4.2.3. Inundation Maps and Spatially Distributed Prediction Error of the Observed Events

The predicted inundation maps for the observed events (Figures 7 and 8) show only minimal differences compared to the true inundation. Small differences between predicted and true inundation can be found for the event from 2013 in the southwestern part of the study area (Figure 8). This issue can be attributed partly to the fact that the majority of these cells are in feature 3, which performs somewhat worse than the other two features (described in Section 4.2.2), and also to the fact that the training set does not contain similar hydrograph shapes to the 2013 event (see Section 2.6). Either expanding the training set or splitting the features further based on different categories (as described in Section 4.2.2) are possible strategies.

By evaluating the results and prediction error, cells with an error based on the MSE larger than 0.2 m^2 were only 0.11% for the event in 2005. The comparison of the prediction error with the results achieved by Lin et al. in 2020, where the authors stated 8.97% of the cells have an MSE larger than 0.2 m^2 [11], concluded therefore in a great prediction quality improvement. This can also be visually verified in Figure 9. The analysis of the observed event in 2013 concluded only 2.14% of the cells have an MSE larger than 0.2 m^2. In comparison, Lin et al., 2020 achieved 13.62% [11]. This can once again be visually verified in Figure 10. This comparison demonstrates the effectiveness of our feature-informed data-driven approach.

Nevertheless, it should be also mentioned, besides the architecture and optimization workflow, the use of a hybrid database ensures this high prediction quality, especially on the observed events. Schmid and Leandro also found this to be true when they developed a prediction system based only on synthetic events [24].

4.3. Limitations and Future Research

It should be mentioned, for running the FFS three hydrographs are necessary and are given in this study. For real-time applications, this hydrograph has to be predicted beforehand. For this, there are multiple strategies for delivering a flood hydrograph in real time [28,29]. The FFS is trained based on the hydrodynamic model HEC-RAS 2D. Thus, structural and parameter uncertainties arising from using this model are included in the FFS. In any case, physically based models like HEC-RAS 2D are considered state-of-the-art for generating spatially distributed water depths and inundation maps and therefore the best choice for training [7].

As for the features, we integrated only the distance to the river for the sake of testing and demonstrating our methodology. However, there are more flood-influencing factors, like elevation, land use, or slope. Future studies could investigate the integration of more features into the forecast system. Besides that, in our approach, since we focused on the maximum water depth, we neglected the temporal resolution of the inundation. Future research should focus on providing a temporal multi-step forecast of the inundation in the study area.

5. Conclusions

This study presented the development of a forecast system to predict the maximum inundation for flood hydrographs. The FFS consists of a feature-informed data-driven model, which interprets the forecasting process as an image-to-image translation and combines a convolutional classification architecture (ResNet 50) with three paths of informed dense layers to generate the inundation map. The dense layers are informed by the distance of the respective cell to be predicted to a river. Based on our results, the following conclusions can be drawn.

1. The FFS delivers the prediction within 19 s, thus making the system usable for real-time applications.
2. The innovative training workflow with pre-simulated inundation scenarios from a hybrid database ensures a high prediction quality. The accuracy of the FFS compared with the physically based model HEC-RAS shows average RMSE values of 0.052 for 35 test events. A test on two observed flood events showed RMSE values of 0.074 and 0.141.
3. The feature-informed architecture makes it possible to integrate additional knowledge about flooding into the model without providing an additional dataset. This not only enhances the prediction performance, since it also allows training on a standard computer, making such an application accessible to a wider range of users. The prediction performance on two observed test events showed only 0.11% and 2.14% of the cells forecasted have an MSE value larger than 0.2 m^2, an improvement as compared to the work of Lin et al., 2020 [11] which reached values of 8.97% and 13.62%.

Author Contributions: Conceptualization, F.S.; methodology, F.S.; software, F.S; validation, F.S; data curation, F.S.; writing—original draft preparation, F.S.; writing—review and editing, J.L.; visualization, F.S.; supervision, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research presented in this paper has been carried out as part of the HiOS project (Hinweiskarte Oberflächenabfluss und Sturzflut) funded by the Bayerisches Staatsministerium für Umwelt und Verbraucherschutz (StMUV) [Bavarian State Ministry of the Environment and Consumer Protection] (69-0270-92086/2017) and supervised by the Bayerisches Landesamt für Umwelt (LfU) [Bavarian Environment Agency].

Data Availability Statement: The datasets generated for this study are available upon request to the corresponding author.

Acknowledgments: We would like to thank Qing Lin (Justus Liebig University Gießen—Center for International Development and Environmental Research (ZEU)) for sharing the data from his article: Prediction of Maximum Flood Inundation Extents With Resilient Backpropagation Neural Network: Case Study of Kulmbach.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Swiss RE Institue Sigma—Natural Catastrophes in 2021: The Floodgates Are Open. Available online: https://www.swissre.com/institute/research/sigma-research/sigma-2022-01.html (accessed on 11 December 2023).
2. Arnell, N.W.; Lloyd-Hughes, B. The Global-Scale Impacts of Climate Change on Water Resources and Flooding under New Climate and Socio-Economic Scenarios. *Clim. Chang.* **2014**, *122*, 127–140. [CrossRef]
3. Bevacqua, E.; Maraun, D.; Vousdoukas, M.I.; Voukouvalas, E.; Vrac, M.; Mentaschi, L.; Widmann, M. Higher Probability of Compound Flooding from Precipitation and Storm Surge in Europe under Anthropogenic Climate Change. *Sci. Adv.* **2019**, *5*, eaaw5531. [CrossRef] [PubMed]
4. Calvin, K.; Dasgupta, D.; Krinner, G.; Mukherji, A.; Thorne, P.W.; Trisos, C.; Romero, J.; Aldunce, P.; Barrett, K.; Blanco, G.; et al. *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Core Writing Team, Lee, H., Romero, J., Eds.; IPCC: Geneva, Switzerland, 2023.
5. Rözer, V.; Peche, A.; Berkhahn, S.; Feng, Y.; Fuchs, L.; Graf, T.; Haberlandt, U.; Kreibich, H.; Sämann, R.; Sester, M.; et al. Impact-Based Forecasting for Pluvial Floods. *Earth's Future* **2021**, *9*, 2020EF001851. [CrossRef]
6. Crotti, G.; Leandro, J.; Bhola, P.K. A 2D Real-Time Flood Forecast Framework Based on a Hybrid Historical and Synthetic Runoff Database. *Water* **2019**, *12*, 114. [CrossRef]
7. Henonin, J.; Russo, B.; Mark, O.; Gourbesville, P. Real-Time Urban Flood Forecasting and Modelling—A State of the Art. *J. Hydroinform.* **2013**, *15*, 717–736. [CrossRef]
8. Sun, W.; Bocchini, P.; Davison, B.D. Applications of Artificial Intelligence for Disaster Management. *Nat. Hazards* **2020**, *103*, 2631–2689. [CrossRef]
9. Berkhahn, S. An Ensemble Neural Network Model for Real-Time Prediction of Urban Floods. *J. Hydrol.* **2019**, *12*, 743–754. [CrossRef]
10. Kimura, N.; Yoshinaga, I.; Sekijima, K.; Azechi, I.; Baba, D. Convolutional Neural Network Coupled with a Transfer-Learning Approach for Time-Series Flood Predictions. *Water* **2019**, *12*, 96. [CrossRef]
11. Lin, Q.; Leandro, J.; Wu, W.; Bhola, P.; Disse, M. Prediction of Maximum Flood Inundation Extents With Resilient Backpropagation Neural Network: Case Study of Kulmbach. *Front. Earth Sci.* **2020**, *8*, 332. [CrossRef]
12. Kabir, S.; Patidar, S.; Xia, X.; Liang, Q.; Neal, J.; Pender, G. A Deep Convolutional Neural Network Model for Rapid Prediction of Fluvial Flood Inundation. *J. Hydrol.* **2020**, *590*, 125481. [CrossRef]
13. Guo, Z.; Leitão, J.P.; Simões, N.E.; Moosavi, V. Data-driven Flood Emulation: Speeding up Urban Flood Predictions by Deep Convolutional Neural Networks. *J. Flood Risk Manag.* **2021**, *14*, e12684. [CrossRef]
14. Hofmann, J.; Schüttrumpf, H. floodGAN: Using Deep Adversarial Learning to Predict Pluvial Flooding in Real Time. *Water* **2021**, *13*, 2255. [CrossRef]
15. Burrichter, B.; Hofmann, J.; Koltermann da Silva, J.; Niemann, A.; Quirmbach, M. A Spatiotemporal Deep Learning Approach for Urban Pluvial Flood Forecasting with Multi-Source Data. *Water* **2023**, *15*, 1760. [CrossRef]
16. Russell, S.J.; Norvig, P.; Chang, M.; Devlin, J.; Dragan, A.; Forsyth, D.; Goodfellow, I.; Malik, J.; Mansinghka, V.; Pearl, J.; et al. *Artificial Intelligence: A Modern Approach*, 4th ed.; Pearson Series in Artificial Intelligence; Pearson: Harlow, UK, 2022; ISBN 978-1-292-40113-3.
17. O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. *arXiv* **2015**, arXiv:1511.08458.
18. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [CrossRef]
19. Löwe, R.; Böhm, J.; Getreuer Jensen, D.; Leandro, J.; Højmark Rasmussen, S. U-FLOOD—Topographic Deep Learning for Predicting Urban Pluvial Flood Water Depth. *J. Hydrol.* **2021**, *603*, 126898. [CrossRef]

20. Ludwig, K.; Bremicker, M. *The Water Balance Model LARSIM—Design, Content and Applications*; Freiburger Schriften zur Hydrologie: Freiburg, Germany, 2006; Volume 22.
21. US Army Corps of Engineers, Hydrologic Engineering Center. HEC-RAS River Analysis System—2D Modeling Users Manual 2016. Available online: https://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS%205.0%202D%20Modeling%20Users%20Manual.pdf (accessed on 11 December 2023).
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
23. Chiang, Y.-M.; Chang, L.-C.; Tsai, M.-J.; Wang, Y.-F.; Chang, F.-J. Dynamic Neural Networks for Real-Time Water Level Predictions of Sewerage Systems-Covering Gauged and Ungauged Sites. *Hydrol. Earth Syst. Sci.* **2010**, *11*, 1309–1319. [CrossRef]
24. Schmid, F.; Leandro, J. An Ensemble Data-Driven Approach for Incorporating Uncertainty in the Forecasting of Stormwater Sewer Surcharge. *Urban Water J.* **2023**, *20*, 1140–1156. [CrossRef]
25. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
26. Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; Wierstra, D. Weight Uncertainty in Neural Networks. *arXiv* **2015**, arXiv:1505.05424.
27. Bhola, P.; Leandro, J.; Disse, M. Framework for Offline Flood Inundation Forecasts for Two-Dimensional Hydrodynamic Models. *Geosciences* **2018**, *8*, 346. [CrossRef]
28. Tayfur, G.; Singh, V.; Moramarco, T.; Barbetta, S. Flood Hydrograph Prediction Using Machine Learning Methods. *Water* **2018**, *10*, 968. [CrossRef]
29. Tayfur, G. Real-Time Flood Hydrograph Predictions Using Rating Curve and Soft Computing Methods (GA, ANN). In *Handbook of Hydroinformatics*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 325–338. ISBN 978-0-12-821962-1.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

A data-driven multi-step flood inundation forecast system

forecasting

MDPI

Article

A Data-Driven Multi-Step Flood Inundation Forecast System

Felix Schmid and Jorge Leandro *

Chair of Hydromechanics and Hydraulic Engineering, Research Institute Water and Environment, University of Siegen, 57076 Siegen, Germany; felix.schmid@uni-siegen.de
* Correspondence: jorge.leandro@uni-siegen.de

Abstract: Inundation maps that show water depths that occur in the event of a flood are essential for protection. Especially information on timings is crucial. Creating a dynamic inundation map with depth data in temporal resolution is a major challenge and is not possible with physical models, as these are too slow for real-time predictions. To provide a dynamic inundation map in real-time, we developed a data-driven multi-step inundation forecast system for fluvial flood events. The forecast system is based on a convolutional neural network (CNN), feature-informed dense layers, and a recursive connection from the predicted inundation at timestep t as a new input for timestep t + 1. The forecast system takes a hydrograph as input, cuts it at desired timesteps (t), and outputs the respective inundation for each timestep, concluding in a dynamic inundation map with a temporal resolution (t). The prediction shows a Critical Success Index (CSI) of over 90%, an average Root Mean Square Error (RMSE) of 0.07, 0.12, and 0.15 for the next 6 h, 12 h, and 24 h, respectively, and an individual RMSE value below 0.3 m, for all test datasets when compared with the results from a physically based model.

Keywords: real-time forecasting; urban flooding; artificial neural network; convolutional neural network; temporal and spatial distribution

Citation: Schmid, F.; Leandro, J. A Data-Driven Multi-Step Flood Inundation Forecast System. *Forecasting* **2024**, *6*, 761–781. https://doi.org/10.3390/forecast6030039

Academic Editor: Ajay Kalra

Received: 1 July 2024
Revised: 11 September 2024
Accepted: 12 September 2024
Published: 13 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Extreme weather events pose major challenges for humanity. Especially due to climate change, such events are likely to occur more often in the future [1,2]. In this context, flood events are one of the most common natural disasters and lead to the highest damages among all possible natural catastrophes [3]. In recent years, flood management has evolved from traditional structural measures to include non-structural measures such as early warning and forecast systems [4]. Such systems are of great interest to protect the population and infrastructures as well as to prevent or mitigate economic damage [5].

Flood forecasting is typically conducted by rainfall-runoff modeling, providing information about the impending hydrograph. Once a specific discharge threshold value is reached, a flood warning is communicated. In addition to this hydrological forecast, the impact of the impending hydrograph on a possible inundation and its extent is essential information [6]. The inundation extent is commonly provided by a map, showing spatially distributed water depths in the respective study site, helping to quickly identify dangerous regions and allowing, for example, the optimization of rescue routes [7].

Such inundation maps are produced by hydrodynamic modeling using physically based models, which solve the shallow water equations for given water input based on numerical schemes. The advantage of these models lies in the consideration of a detailed spatial resolution (e.g., solving on a grid of 4 by 4 m) and thus in high-resolution spatially distributed water depths.

However, their disadvantage lies in long calculation times. The numerical schemes that solve the equations require timesteps of seconds, resulting in an enormous amount of

calculations [7–9]. To speed up these calculation times, recent studies have developed numerical solvers for the shallow water equation that can make use of Graphical Processing Units (GPUs) [10–12]. Although the run times could be reduced by a factor of 15 in some cases [12,13], they are still not suitable in the context of real-time forecasting, especially if the inundation model has to be run multiple times in the context of uncertainty estimation or in the case of a changing input when new timesteps are available [7,9].

To consider inundation extent and depth in the context of forecasting despite the long calculation times of physically based models, nowadays, some authorities use one pre-simulated static flood map for a specific return period (e.g., 100 years) or, in rare cases, a complete database of pre-simulated flood events [6]. This is undoubtedly an improvement, but above all, it improves flood preparedness in advance of an event. Static maps based on a synthetic return period are not able to describe the inundation of a real event. Operational early warning requires real-time forecasts and especially information about the times (e.g., what will the situation be like in 3 or 6 h). Considering a temporal resolution of the flood event could significantly improve flood and early warning management, as time strongly influences decision making. Such information about time could be achieved by a dynamic inundation map, which shows the respective inundation for several timesteps. Nevertheless, providing a dynamic inundation map remains a great challenge.

Due to the excessive computing times of physically based models for real-time predictions, other approaches are being researched. Crotti et al. (2019), for example, have developed an inundation forecast framework based on a database with multiple pre-simulated inundation maps, where the similarity of the input data is used to select the most likely inundation map [6]. Along with inundation mapping, flood susceptibility mapping based on flood-influencing factors has also gained more attention in recent years, where spatial geographical data [14] and even social media data [15] are used to generate a map that shows the susceptibility towards inundation and flooding.

Besides these approaches, data-driven applications have gained importance in the research community in recent years. Data-driven models, as the name suggests, rely on datasets and attempt to establish a relationship between a set of input and corresponding output data [16]. They offer very good computational efficiency and originate from the fields of statistical modeling, artificial intelligence, and machine learning. Data-driven models are, for example, applied in computer vision tasks in the field of medical engineering [17] or object detection [18].

To make use of the benefit, in the field of flood forecasting, data-driven approaches are applied in multiple studies to predict the flood hydrograph [19,20] or to replace physically based models and to generate flood inundation maps [7,9,21–29]. Recent advances in data-driven flood forecasting models aim to incorporate multiple uncertainty sources within the input data [20,28] and physical laws [19].

The generation of inundation maps, showing the maximum water level occurring during a flood event, is presented for pluvial events in a study by Berkhahn et al. (2019) [21] and for fluvial events by Lin et al. (2020) [22]. Both authors used artificial neural networks (ANNs) and a supervised learning strategy with pre-simulated physically based scenarios. The output resolutions of the ANNs match the grid resolution of the physically based models, and after successful training, the ANNs were able to replace the physically based model. However, both studies did not consider a temporal variation of the inundation, outputting a single (maximum) water depth. Follow-up studies focused on the improvement of the data-driven model architecture, applying convolutional neural networks (CNNs) following an image-to-image translation strategy for a single part of the catchment [26] or for the total catchment in multiple parts presented as subcatchments [24,25,27]. CNNs are specialized data-driven models for handling structured grid data, and, in contrast to ANNs, they take into account the information of their neighboring cells for the prediction of one cell. This improved the forecast quality and made CNNs the standard procedure. For further improvement, some studies added additional

geographical information (GIS) data into the model, like elevation or slope [24,25]. This approach improved the prediction quality but made training on a standard computer nearly impossible. Besides that, these studies did not consider the temporal resolution of the inundation. To address the issue of training on a standard computer while keeping the idea of integrating further knowledge into the network via additional datasets, Schmid and Leandro (2023) developed a feature-informed network for the prediction of fluvial flooding extents [7]. This approach made it possible to integrate information about the distance for each cell to be predicted to a river into the network without providing the network with additional datasets. This strategy improved the forecast but did also not consider the temporal resolution of the flood event.

An approach for incorporating temporal resolution with the prediction for fluvial flood events is presented by Lin et al. (2020) [23]. The authors used separate ANNs for the timesteps 3, 6, 9, and 12 h, and each ANN was trained with hydrographs up to this timestep. This configuration allowed the integration of a time, but for each timestep, a separate ANN has to be set up and trained. Furthermore, no CNNs were used, and the inundation is only dependent on the discharge. Another approach for pluvial flooding is presented by Burrichter et al. (2023), where the output of the data-driven model is the complete time sequence [29]. However, the complete time sequence is not always feasible, especially at the beginning of an event. To resolve these issues, Berkhahn and Neuweiler (2024) developed a data-driven inundation model with a spatial and temporal resolution for pluvial flooding [9]. The authors applied a recursive modeling architecture, which is also often used in multiple studies in the fields of multi-step time series forecasting [28,30,31]. These recursive structures feed the prediction from the precious timestep back and use it as new input. Thus, the prediction of the next timestep is influenced by some current input and the output of the previous timestep. This data-driven model showed promising results but did not consider the inclusion of additional knowledge into the data-driven model to improve the forecast.

The present study builds on the described knowledge and methods and aims to develop a data-driven multi-step flood inundation forecast system for fluvial events to provide a dynamic inundation map for the respective study site in real time. To ensure high prediction quality for both spatial and temporal dimensions, the forecast system is implemented with a recursive connection, but unlike other studies, we follow the idea of a feature-informed network to further integrate additional knowledge into the forecast system.

2. Materials and Methods

2.1. General Overview

Like other studies [7,9,21–23,25,26], we applied a supervised learning strategy using pre-simulated physically based inundation scenarios as true and, therefore, desired inundation values to achieve the aim of developing a multi-step forecast system for real-time prediction. In this case, a supervised strategy ensures better prediction quality since the model has a clear desired target value. Once the optimization is finished, the data-driven model is able to replace the physically based model by providing a dynamic inundation map in real time. Figure 1 describes the general workflow for the development, which consists of three blocks: (I) the operational design, (II) data-driven modeling, and (III) the physically based modeling. By following this workflow, a data-driven multi-step forecast system can be developed for any study area.

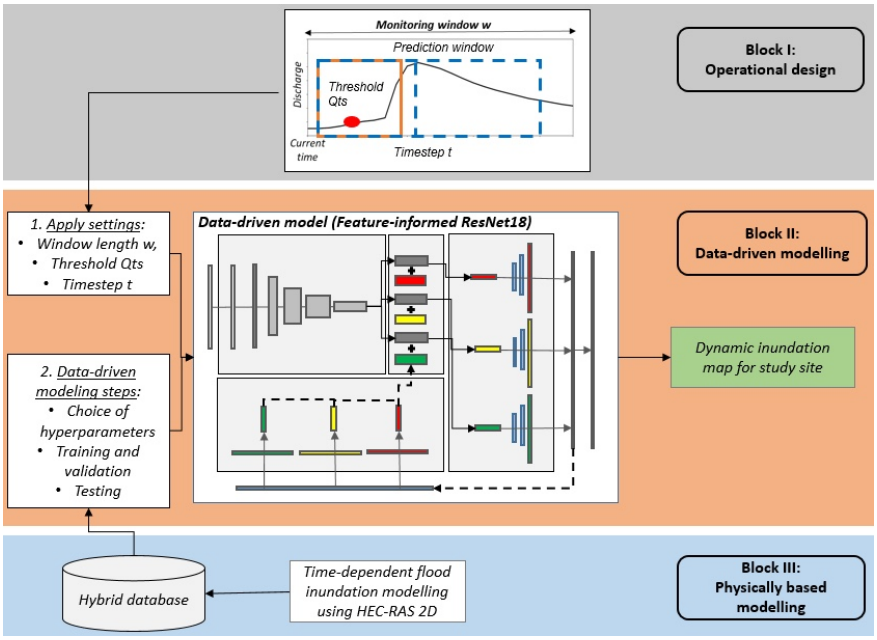


Figure 1. Workflow for the development of a data-driven multi-step forecast system.

The first block deals with the operational design. A specific time period is defined within a monitoring window w. Within this window, a future/predicted hydrograph is listed, which forms the basic input for the inundation forecast later. The length of this window is also the maximum lead time for the multi-step forecast. The hydrograph within this window moves with a specific timestep t towards the current time. As the inundation forecast does not have to be carried out continuously, a threshold value Qts can be set for the start of the forecast. These three parameters (length of the window w, timestep t, and threshold Qts) provide the operational settings for the forecast system. Once a forecast is started, for each possible timestep within the window, the respective inundation map will be generated. It is worth mentioning that the operating design must be defined in advance but can be adapted to meet the requirements of the application in any study area. In this study, the length of the forecasting window w is set to 24 h, the discharge threshold Qts is set to 25 m^3/s, and the timestep t is set to 1 h. A detailed description of the forecast strategy is presented in Section 2.2.

The second block deals with the data-driven modeling (block II). In this, the architecture of the model itself is set, and the operational settings from block I (monitoring window length w, discharge threshold Qts, and the timestep t) are applied to the model. Afterward, the hyperparameters of the models are set, and training, validation, and testing are carried out in this part. We follow an image-to-image approach [7,9,26] since this allows the use of a convolutional neural network, and, additionally, we apply a feature-informed strategy since the integration of additional knowledge into the network has been proven beneficial [7]. Furthermore, we add a recursive feedback connection from the predicted inundation of the previous timestep as an additional new input besides the hydrograph. The final developed forecast system is able to produce a dynamic inundation map for the maximum lead time (length of the forecast window) with a temporal resolution of timestep t. Section 2.3 presents further details about the model structure and choice of hyperparameters as well as the complete data-driven modeling approach.

The third block deals with physical modeling. Since we aim to replace the physical-based model with a data-driven model, it is necessary to generate a database, which is used for setting up the data-driven model later. Therefore, different discharge events are simulated with a two-dimensional (2D) inundation model to gain physical-based inundation maps. Here, the 2D-inundation model must output these maps at the same

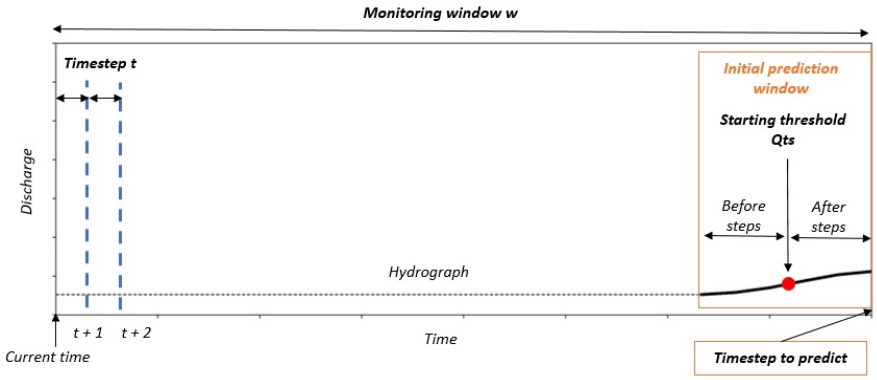
timestep t as in the operational design. In general, any database that contains hydrographs and respective inundation maps can be used for development. In this study, we used a hybrid database generated by Crotti et al. [6], where the hydrological model LARSIM (Large Area Runoff Simulation Model) [32] was used to generate the discharge scenarios, and the hydrodynamic model HEC-RAS 2D (Hydrologic Engineering Center—River Analysis System, Davis, CA, USA) [33] to obtain the corresponding inundation map. This database was also used by Schmid and Leandro [7] for the development of a feature-informed data-driven model to predict the maximum inundation extent. A detailed description of the database and the physical-based modeling is presented in Section 2.4.

2.2. Operational Design

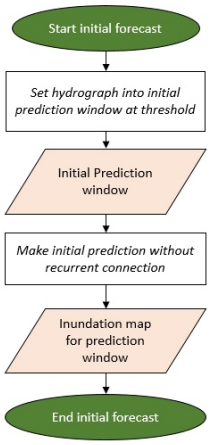
The operational design builds the frame around the forecast system and is controlled by three parameters: (i) the monitoring window w, (ii) the timestep t, and (iii) the discharge threshold value Qts. These parameters will be applied later to the data-driven model. We would like to mention that these parameters are user-defined and need to be set beforehand; however, by introducing them, the methodology is customizable and, therefore, easy to transfer for the development of a forecast system in other study areas. A maximum prediction lead time is decided by the length of the monitoring window w, in which the future hydrograph is listed from the current time $(t = 0)$ towards the maximum prediction lead time and end of the monitoring window w $(t = t + \max lead\ time)$. Once time evolves, the hydrograph moves with a specific timestep t towards the current. T also gives the temporal resolution of the dynamic inundation map later. The third parameter, the threshold Qts, is responsible for starting an inundation forecast. Figure 2a shows these parameters for an event.

Once the discharge threshold Qts is reached, the inundation forecast is started. If the threshold is reached for the first time, the inundation is still multiple steps ahead, and only the last timesteps of the hydrograph in the monitoring window w are responsible for the incoming inundation. Therefore, an initial single-step forecast is carried out (Figure 2a), and an initial prediction window is set three timesteps before and after the discharge threshold Qts value are taken into the prediction window (visualized by the orange box in Figure 2a). This is necessary for obtaining the shape of the hydrograph. The rest of the hydrograph (timesteps before the initial prediction window) is neglected (shown by the dotted points in Figure 2b). Afterward, the initial prediction window is used to carry out the first inundation forecast. Since this inundation is only influenced by the hydrograph itself, no inundation of the previous timestep is necessary. The initial single-step forecast outputs an inundation as a map occurring at the maximum lead time. Figure 2b shows the algorithm behind the initial forecast.

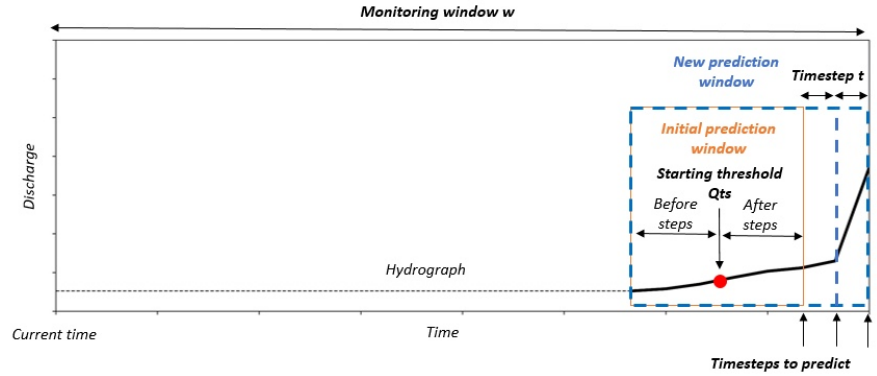
Once time passes by, new timesteps of the hydrograph are available, and the event comes closer to the current time (visualized in Figure 2c). Now, it is important to also know how the inundation will behave over time. At this stage, the multi-step forecast is started, and a first dynamic inundation map can be generated. The time of reaching the discharge threshold value Qts can be identified again, and the initial prediction window (orange) can be set, now only closer to the current time. Since the shape of the hydrograph normally also changes once new timesteps are available, the initial prediction window is forecasted again following the initial forecast. Afterward, the initial prediction window is extended by every newly available timestep. This new prediction window (visualized by the blue box in Figure 2c) is used to forecast the new timesteps via consideration of recursive connections.



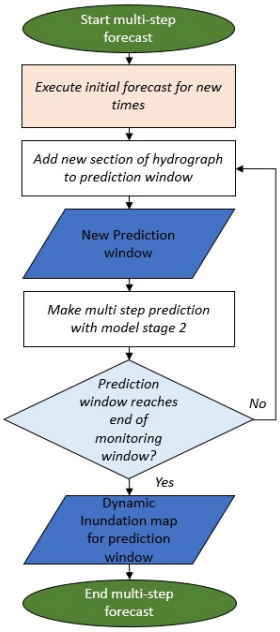
(**a**) Initial single-step forecast

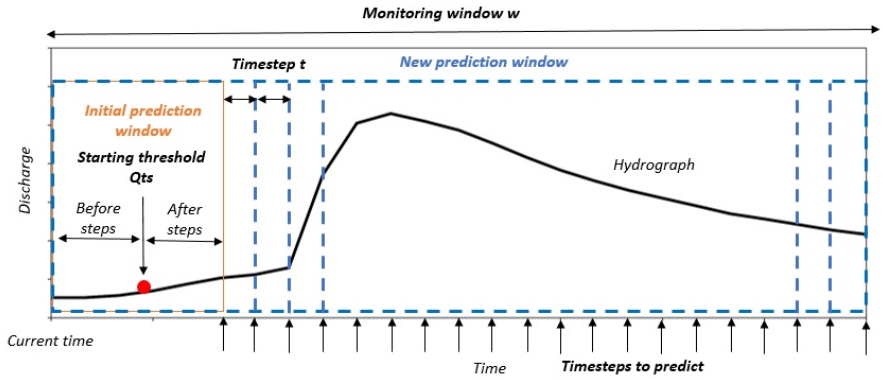


(**b**) Algorithm for initial forecast



(**c**) Multi-step forecast for the new prediction window including in initial prediction window and two more timesteps





(**d**) Final multi-step forecast

(**e**) Algorithm for multi-step forecast

Figure 2. Visual presentation of the forecast system in an operational mode. (**a**) Representation of a monitoring window w once the discharge threshold value Qts is reached for the first time and an initial single-step forecast is carried out; (**b**) algorithm that is necessary to carry out the initial forecast; (**c**) the monitoring window w once two more timesteps after the initial single-step forecast are available and a multi-step forecast is carried out for the new prediction window; (**d**) the monitoring window w once the initial prediction window reaches the current time, which marks the final forecast; (**e**) algorithm that is necessary to carry out the multi-step forecast.

This multi-step forecast is carried out until the new prediction window is at the current time of the monitoring window w, concluding with a dynamic inundation map, starting with the initial forecast, and adding more inundation maps with the timely resolution t until the maximum lead time is reached (shown in Figure 2d). The necessary algorithm and its steps for running a multi-step forecast are listed in Figure 2e. An overall algorithm for running the forecast system in operational mode is provided in Figure 3. This algorithm describes the application of the forecast system for real-world implementation.

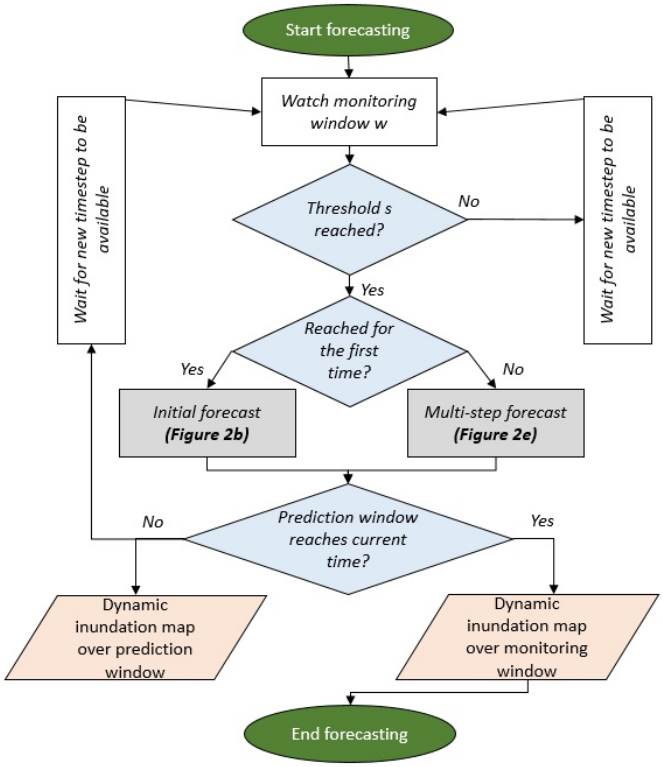


Figure 3. Algorithm for running the forecast system.

2.3. The Data-Driven Modeling

The data-driven modeling part is responsible for generating the dynamic inundation map in real time. We applied the image-to-image translation strategy following a feature-informed architecture.

2.3.1. Architecture

The data-driven modeling approach to enable time-dependent predictions presented herein builds upon the work of the authors on feature-informed architecture [7]. The feature-informed architecture combines a convolutional neural network with feature-informed dense layers, in which the additional information of the distance to the river for each cell to be predicted is integrated.

To address the challenge of providing time-dependent multi-step predictions, this architecture is enhanced with a recursive connection in which the predictions of the feature-informed path are fed back as a new input; thus, the prediction of timestep t is dependent on the input of the hydrograph and its own prediction of t − 1. Figure 4 shows this architecture. The architecture can mainly be split into three different parts: (i) convolution, (ii) feature-informed paths, and (iii) recursive feature-informed paths. Each part will be explained in the following paragraphs.

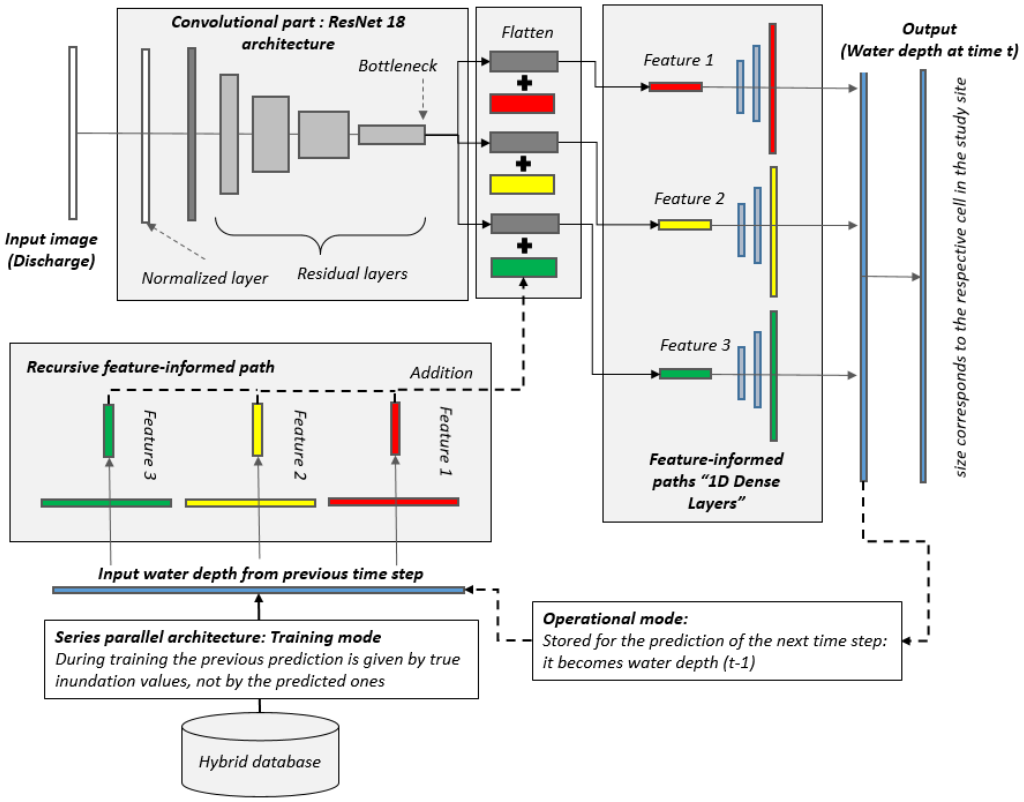


Figure 4. The architecture of the data-driven model. The recursive connection is represented with the dotted black line and is activated only in the operational mode, whereas during training, it is replaced by the true inundation values.

The convolution part consists of a residual convolutional network with 18 layers and is responsible for taking the input hydrograph as an image. This adjustment from the original residual network 50 (ResNet50) from Schmid and Leandro 2023 [7] was re-adjusted to minimize the computational amount. Convolutional networks (CNNs) are specifically designed for handling structured grid-like data, such as images. They have demonstrated widespread success in a range of computer vision tasks, like image classification, object detection, and image segmentation [34,35]. The fundamental concept behind CNNs lies in their utilization of multiple convolutional layers. A convolutional layer applies learnable filters (also known as kernels, a smaller matrix with the respective kernel size) to the input image. Each filter/kernel slides over the image, conducting element-wise multiplication and summation, resulting in a feature map. The primary aim of this process and the resultant feature map is to identify and retain local patterns, characteristics, and attributes of the image. A mathematical description of this process is given by Equation (1) [34]:

$$FM_{ijk} = \sum_{z}^{c} \sum_{x}^{f} \sum_{y}^{f} K_{zxyk} I_{(i*S+x)(j*S+y)z} \quad (1)$$

where FM_{ijk} is the respective value at index (i, j) in the kth output feature map, k refers to the number of filters applied, c is the number of channels of the input image, f is the size of the kernel, K_{xyzk} is the kernel and respective value at index (z, x, y, k) while z iterates over the input channels of the provided image, $I_{(i*S+x)(j*S+y)z}$ is the input image and the respective value at the position $(i * S + x)(j * S + y)$ in the channel z, and the stride S is the step size at which the kernel moves over the image. After such a

convolutional operation, batch normalization (Equations (2) and (3)) is applied to the produced feature maps, where x_{BN} is the normalized activation, x is the input, μ is the mean of the batch, σ^2 is the variance, ϵ is a small constant for stability, γ and β are learnable scales and shift parameters, respectively, and y_{BN} is the output of the batch normalization.

$$x_{BN} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{2}$$

$$y_{BN} = \gamma x_{BN} + \beta \tag{3}$$

Batch normalization is performed because it accelerates training, reduces overfitting, stabilizes gradients, and makes the network more robust [36]. An activation function, which is responsible for determining the importance of information passed through the network, is applied to the output of the batch normalization. We applied the rectified linear activation function (ReLU). It is a piecewise linear function that will output the input directly if it is positive; otherwise, it will output zero. The ReLu function is commonly used in data-driven modeling [9,37].

These three steps can be summarized into a convolutional block; thus, the output of a convolutional block can be described as follows (Equation (4)), where $Out_{ConvBlock}$ is the output of a single block; f_a the activation function; BN the batch normalization, produced by Equations (2) and (3); and FM the feature map, produced by the convolutional operation from Equation (1).

$$Out_{ConvBlock} = f_a(BN(FM)) \tag{4}$$

After one convolutional block, four residual blocks follow. A residual block always contains two convolutional blocks and a skip connection that adds the original input back to the output of this block. The skip connection forces the network to learn the difference (residual) that is left over when you subtract input from output, allowing it to refine features effectively [37]. Mathematically, the residual block can be described as follows (Equation (4)), where y_{Res} is the output of the residual block, x_{in} is the input to the block, $\mathrm{F}(x, \{W_i\})$ represents the residual mapping to be learned by the layers in the block, and W_i are the weights/values in the kernel of the layer, which is responsible for the generation of the feature maps.

$$y_{Res} = \mathrm{F}(x_{in}, \{W_i\}) + x_{in} \tag{5}$$

After the last residual block, the extracted feature maps are stored in the flattened layer. This layer can be interpreted as compressed and dense information from the input, and it generates a one-dimensional vector. Afterward, the feature-informed part follows.

The feature-informed part is responsible for rebuilding the image and translating it into the inundation map. Each path represents a feature, which should be integrated into the network. The paths consist of three dense layers, which hold a specific number of neurons in it. One neuron can be described as follows (Equation (6)), where y is the output of a dense layer, x_{dense} the input to the dense layer, and b an added bias.

$$y = f_a(w x_{dense} + b) \tag{6}$$

In this study, we applied the ReLu activation function until the last dense layer, in which we applied a linear activation function. This is also a common strategy [9,21,28]. The output size from the feature-informed dense layers matches exactly the number of cells to be predicted within the study area. After the feature-informed path, the first inundation map for the timestep t is produced. This map is outputted, and, if a multi-step prediction is aimed, this map is provided to the recursive part. If it is the first prediction, no time resolution for the inundation is necessary, since the inundation will take place at the end of the monitoring window/max lead time.

The recursive part stores the prediction from the last timestep and serves as a memory. Once more, it feeds the information through one feature-informed dense layer with the size of the flattened layer and mathematically adds the information to this layer. Thus, at this stage, the model is thereby influenced by the information of the hydrograph at the time t, which runs in the convolutional part, and the inundation from $t-1$, which was processed through the feature-informed paths, allowing the model to perform multi-timestep into the future.

2.3.2. Hyperparameter and Optimization

Before the data-driven model is optimized in the training process, the hyperparameters need to be set. The term hyperparameter covers all parameters/settings within the data-driven structure that are not trainable. Since this type of data-driven model has already proven suitable for flood prediction [7], in this study, we only made the filters of the convolution part and the number of neurons in the recursive paths variable, as we changed the architecture from the ResNet50 of Schmid and Leandro (2023) [7] to a ResNet18 and added the recursive connection. Additionally, we considered in the recursive connection the inundation only from the previous timestep $(t-1)$. Taking more timesteps ($t-2$ *or* 3) into account might sound reasonable, but it makes training on a standard computer impossible.

To ensure the high prediction quality of the forecast system, the prediction of the data-driven model should match the physically based inundation map as closely as possible. This agreement between prediction and true values is measured by the loss function (LF). We applied the mean squared error (MSE) (Equation (7)) as LF, where the true values T are given by the physically based model, P is the prediction value of the data-driven model, and n is the number of samples.

$$\mathrm{LF} = \mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(P-T)^2 \tag{7}$$

During the training process, the weights in the data-driven model are adjusted so that a minimum value of the LF is found by the optimization algorithm. This is carried out by iterating over the number of epochs. In each epoch, the model sees all training data once and tries to improve the prediction performance. Since our model features a recursive connection and the output is dependent on its own prediction, we had to find a training strategy. To still make use of the common optimization algorithms, the recursive connection is cut during training and replaced with the true inundation from the previous timestep provided by the physically based model. This is known as series-parallel architecture. Only in the operational forecasting mode the recursive connection is re-installed. This is also visualized in Figure 4. Such a strategy is also applied in multiple other studies [9,28,31,38]. Furthermore, this is beneficial since the network and weights are trained on true values and the prediction error of each timestep is not carried on.

For optimization, we used the Adam algorithm [39]. Adam is a stochastic gradient descent algorithm that incorporates momentum to past gradients into account, which supports avoiding becoming stuck in a local minimum. Besides this training optimization algorithm, we applied the workflow for the optimization of the feature-informed architecture developed by the authors [7]. This workflow consists of two separate algorithms. The first algorithm trains the total data-driven model. The second algorithm takes the previously trained model, freezes all trainable weights, and then starts an iterative retraining loop around the feature paths. In each iteration, the weights in the respective feature path are unfrozen and retrained. This retraining strategy ensures that the weights in each feature path are optimized to describe the flood depths and characteristics of this path in the best possible way. For more details on the optimization algorithm, please refer to [7]. Furthermore, all input data for the model are scaled between

0 and 1 as a preprocessing step. This improves the prediction quality as well. It is a standard approach in many other studies [21,28].

2.3.3. Evaluation Criteria

Evaluating flood prediction of data-driven models is critical since, in real cases, scenario decisions will be based on the predictions. The outcome of the forecast system is an inundation map, which shows the water depths within the study area. These water depths are typically provided in certain classes, and only cells that are higher than 0.05 m are highlighted, indicating cells below 0.05 m as safe areas. Therefore, we applied the Critical Success Index (CSI) as the first evaluation criterion. The CSI indicates that an alarm is given if a certain limit value is reached; thus, we set the limit to 0.05 m to assess the quality of the inundation extent. Such a limit is also used in Löwe et al. (2021) [24].

Equation (8) presents the calculation of the CSI, where $Hits$ refers to the number of correct positive predictions (values classified correctly into the classes below and above the limit), $Misses$ refers to the number of positive events that were missed by the model (water-depth predictions, which should be higher than 0.05 m but are less), and $False\ Alarm$ refers to the number of false positive predictions (water-depth predictions, which should be less than 0.05 m but are higher). An optimal value of 100% would indicate a perfect prediction/classification around the limit value.

$$\mathrm{CSI} = \left(\frac{Hits}{Hits + Misses + Flase\ Alarm}\right) * 100 \tag{8}$$

Furthermore, to assess the prediction quality of the forecast system, the root mean squared error (RMSE) (Equation (9)) is used. The true values T hereby are taken as the results from the physical-based hydrodynamic model, while the predicted values P are provided by the developed forecast system. The small n accounts for the number of predictions. The RMSE is used because the results are intuitive to interpret, as the error value has the same units as the water level. An optimal score of 0 would indicate a perfect agreement between prediction and true values. We also apply a threshold value, which should be reached to assure high prediction quality [7]. The threshold is set to 0.05 m, which can be interpreted as a water depth error of around 5 cm.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P - T)^2} \tag{9}$$

To summarize the evaluation, we assess the performance of the multi-step forecast system in three different ways:

- For all values within a test dataset, the CSI is calculated.
- For each cell, the RMSE value is calculated. This concludes with a map showing the spatial distribution of the prediction error.
- For each timestep in the monitoring window (t_1, t_2, t_3, …, t_n), an average RMSE error (Equation (10)) is calculated, where j is the respective RMSE of a cell and N refers to the number of cells in the study site. The lead-time prediction error is given by the following equation:

$$\overline{RMSE} = \frac{1}{N}\sum_{i=1}^{N} RMSE_j \tag{10}$$

2.4. Physical-Based Modeling and Dataset for Training, Validation, and Testing

A hybrid database was used for the development of the multi-step inundation forecast system. The database was generated by Crotti et al. (2019) [6] and consists of 270 different scenarios, where each scenario includes discharges of the three rivers and

corresponding inundation maps. One-half of the database (135) consists of synthetic events, where the authors applied the hydrological model LARSIM (Large Area Runoff Simulation Model) [32] on rainfall return periods provided by the German Meteorological Services Deutscher Wetterdienst (DWD) KOSTRA (Koordinierte Starkniederschlags Regionalisierungs Auswertungen) to generate the discharge hydrographs at the inflow location of the three rivers. The other half (135) consists of real-life-based events, where the authors applied a rescaling process to individual observed discharge events from time series analyses to higher return periods so that flooding in the study site occurred. The hydrographs have an hourly resolution and were then used as boundary conditions for two-dimensional inundation modeling using the hydrodynamic model HEC-RAS 2D [33] to obtain respective inundation maps. The maps have a spatial resolution of 4 m × 4 m, resulting in approximately 500,000 cells for the whole study site and a timely resolution of one hour. We would like to state that hydrodynamic modeling was not our focus. For details regarding the generation or validation of the 2D inundation model, please refer to Bhola et al. (2018) [40] or Crotti et al. (2019) [6]. The database was already used for multiple studies [22,23,28]. Besides this database, the three observed events from the years 2005, 2006, and 2013 were additionally used to test the developed forecast system on observed events.

The database was randomly split into three subsets of training (70%), validation (15%), and testing data (15%). The validation subset was used to gain an unbiased evaluation of the model during fitting and hyperparameter tuning, while the test dataset was used to provide an unbiased evaluation of the final model. Since we followed an image-to-image approach, the discharges had to be converted into images. We used a tensor with the size 14 × 14 × 3. Each of the three rivers was represented within its own channel, while the first timestep of the hydrographs was placed in the first position of the image matrix (1, 1), the second timestep in the second one (1, 2), and so on.

3. Study Area

We implemented our developed methodology in the study area Kulmbach (Figure 5). The area has a size of about 12 km^2 and is located in the federal state of Bavaria, Germany. The catchment area has a flow direction from northeast to southwest. Three main rivers flow through the area, namely the Schorgast, the White Main, and the Red Main. The first two flow into the study area in the northeast, while the Red Main enters in the southern part and, unlike the other rivers, flows only briefly through the area. The northeastern part is narrow and surrounded by hills, while in the western part, the catchment area becomes wider and merges into floodplains.

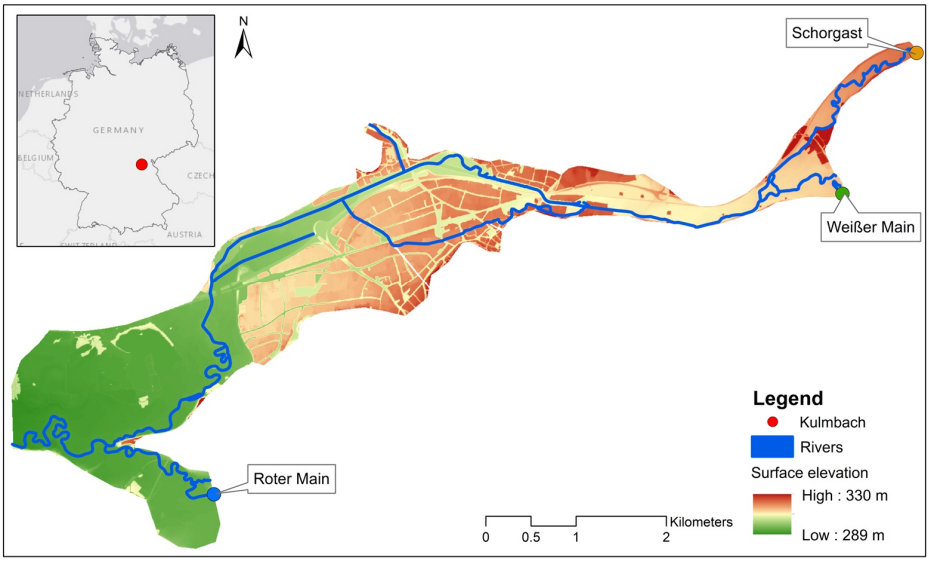


Figure 5. The study area: Kulmbach. The subplot in the upper left shows the location of Kulmbach within Germany.

The soil consists of gravels and sands with a higher proportion of fine material originating from stream or river deposits ranging from the Pleistocene to the Holocene [41]. This mixture forms a porous aquifer. The city center is located in the central part of the catchment area and is home to 27,000 people [22]. Depending on the discharge of the three rivers, the study area can become flooded and pose a threat to its residents. In the years 2005, 2006, and 2013, the study area was heavily flooded, which led to an investigation of flood protection measures for the city. As feature information for the data-driven model, we applied the distance to the river of each cell and grouped them into three groups: cells with a distance of a maximum of 150 m, a maximum of 400 m, and more than 400 m.

4. Results

4.1. Operational Design

Since the study area is influenced by three rivers, we had to implement these rivers into the monitoring window w. For the sake of demonstrating our methodology, we applied a monitoring window w of 24 h and a discharge threshold Qts of 19 m^3/s, as inundation typically starts at this value. Since the resolution of the hydrographs from the hybrid database was one hour, the choice to set the timestep t also to one hour is reasonable. As previously mentioned above, for the first forecast, the initial prediction window consists of three timesteps before and after the threshold; hence, the initial forecast accounts for 6 h. This configuration takes 19 timesteps to predict the final forecast in (the next 6 h as the initial forecast, following t + 7, t + 8, t + 9, …t + 24). The model was trained and tested on an NVIDIA Quadro M4000 graphics card with a memory of 8 GB. Such a card (or similar) is found commonly in many PCs. Python 3.9 and TensorFlow 2.6 were used for the development of the multi-step forecast system.

4.2. Hyperparameter and Optimization

We considered the filters in the convolutional part and the neurons in the recursive paths variable hyperparameters, and, thus, we performed a sensitivity analysis on these hyperparameters (the filter and neurons). To decide on the final model configuration, we relied on the validation results of this configuration and checked if the threshold (RMSE of 0.05 m) was reached. Table 1 shows the tested configurations, and Table 2 shows the

evaluation and results of the sensitivity analysis for these configurations. Since configuration 2 showed the best validation performance and reached the threshold, we decided on this one.

Table 1. Tested configurations for the number of filters in the convolutional part of the model.

Block	Configuration 1 [Number of Filters/Neurons]	Configuration 2 [Number of Filters/Neurons]	Configuration 3 [Number of Filters/Neurons]
Convolutional block	32	64	64
Residual block 1	32, 32	64, 64	128, 128
Residual block 2	64, 64	128, 128	256, 256
Residual block 3	128, 128	256, 256	512, 512
Residual block 4	256, 256	512, 512	1024, 1024
Flatten layer	256	512	1024
Feature path 1 *	256, 512, 210,666	256, 512, 210,666	256, 512, 210,666
Feature path 2 *	256, 512, 115,292	256, 512, 115,292	256, 512, 115,292
Feature path 3 *	256, 512, 83,642	256, 512, 83,642	256, 512, 83,642
Recursive feature path 1	256	512	1024
Recursive feature path 2	256	512	1024
Recursive feature path 3	256	512	1024

* Features taken from [7].

Table 2. Training and validation results of the three different configurations.

Configuration	Training RMSE [m]	Validation RMSE [m]
1	0.035	0.052
2	0.032	0.046
3	0.029	0.086

4.3. Performance Evaluation of the Forecast System

The forecast system was tested with the test dataset, including 35 events from the hybrid database. Additionally, three observed events from the years 2005, 2006, and 2013 were also used for testing. Since, in our case, the hydrographs are not changing over time, all tests were performed over the entire monitoring window. The prediction of one dynamic inundation map over the monitoring window was carried out within 47 s instead of the hours required by the physically based model, achieving the aim of real-time prediction.

As already mentioned, the performance of the forecast system was evaluated in three different ways: the calculation of the CSI with a limit value of 0.05 m, the calculation of the average RMSE value for each timestep and event, and the calculation of the spatial distribution of the prediction error. Section 4.3.1 presents the CSI values, Section 4.3.2 shows the overall performance and lead time prediction error, and Section 4.3.3 shows exemplarily the observed event from the year 2006, the inundation maps, and spatial error distribution.

4.3.1. Critical Success Index (CSI)

Table 3 shows the CSI values for all values in all timesteps to predict. For prediction purposes, we also took a look at the CSI values for each respective timestep. It turned out that there is no real difference between the respective timesteps, and, therefore, the CSI value for all timesteps is a sensible and meaningful approach. It can be seen that the CSI is always above 90%, meaning that, more or less, all cells are classified correctly in cells with water higher than 0.05 m (which can be interpreted as flooded areas) and safe cells (water depth < 0.05). Regarding this evaluation criteria, the forecast system delivers very good results and achieves the aimed high prediction quality.

Table 3. Critical Success Index (CSI) in percentage for the test datasets over all timesteps.

Test Dataset	Critical Success Index (CSI) [%]
Test dataset (56 events)	95.71
Observed event 2005	93.31
Observed event 2006	94.63
Observed event 2013	92.12

4.3.2. Overall Performance and Lead Time Prediction Error

Figure 6 shows the average RMSE values for these test datasets plotted over the entire monitoring window, concluding with a prediction error for each timestep. It can be seen that all sets have an RMSE for the first 6 h below 0.07 m, which can be interpreted as an average water depth error below 7 cm over the entire study site. This error increases over time since the prediction error from t − 1 is carried on for each timestep. Nevertheless, even after 12 h, the error is still acceptable and around 0.08 m and 0.12 m. Except for the observed event from 2013, the worst predictions are at 24 h. At this stage, the error ranges from 0.1 m to 0.14 m. Overall, the prediction errors are good and always below 0.15 m.

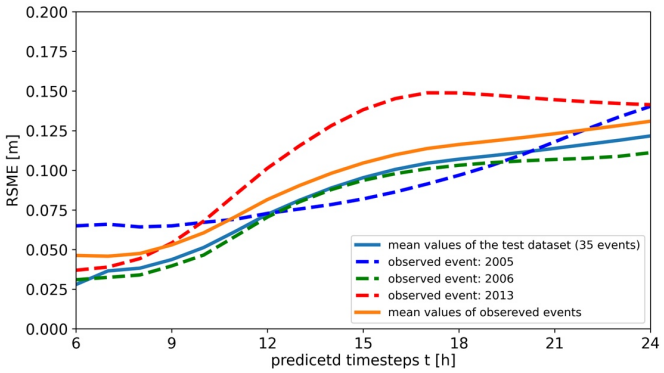
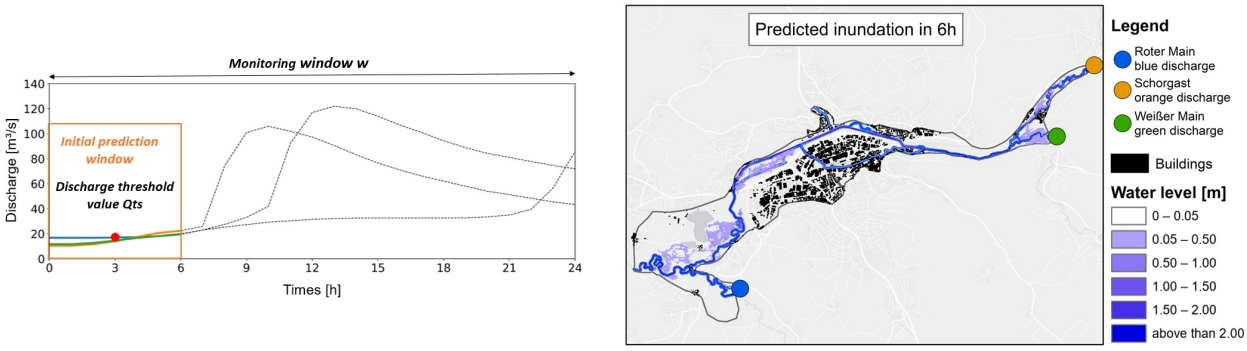


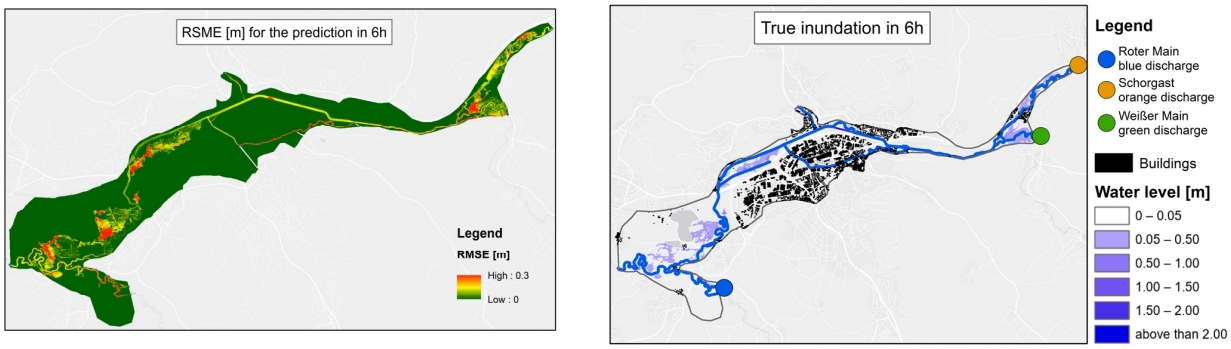
Figure 6. Average RMSE prediction errors for the test datasets over the timesteps to predict.

4.3.3. Inundation Maps and Spatial Prediction Error Distribution

The evaluation of the forecast system for three timesteps (6 h, 12 h, and 24 h) of the observed events from the year 2006 is presented in Figures 7–9, where Figure part (a) shows the used section of the hydrographs within the monitoring window, part (b) the respective predicted inundation map, (c) the true inundation, and part (d) the RMSE value of the respective cells. It is clearly visible that all predicted inundation maps are very similar to the true inundation maps, which were calculated by HEC-RAS 2D. Furthermore, the RMSE map shows that most of the respective cells have very small RMSE values.

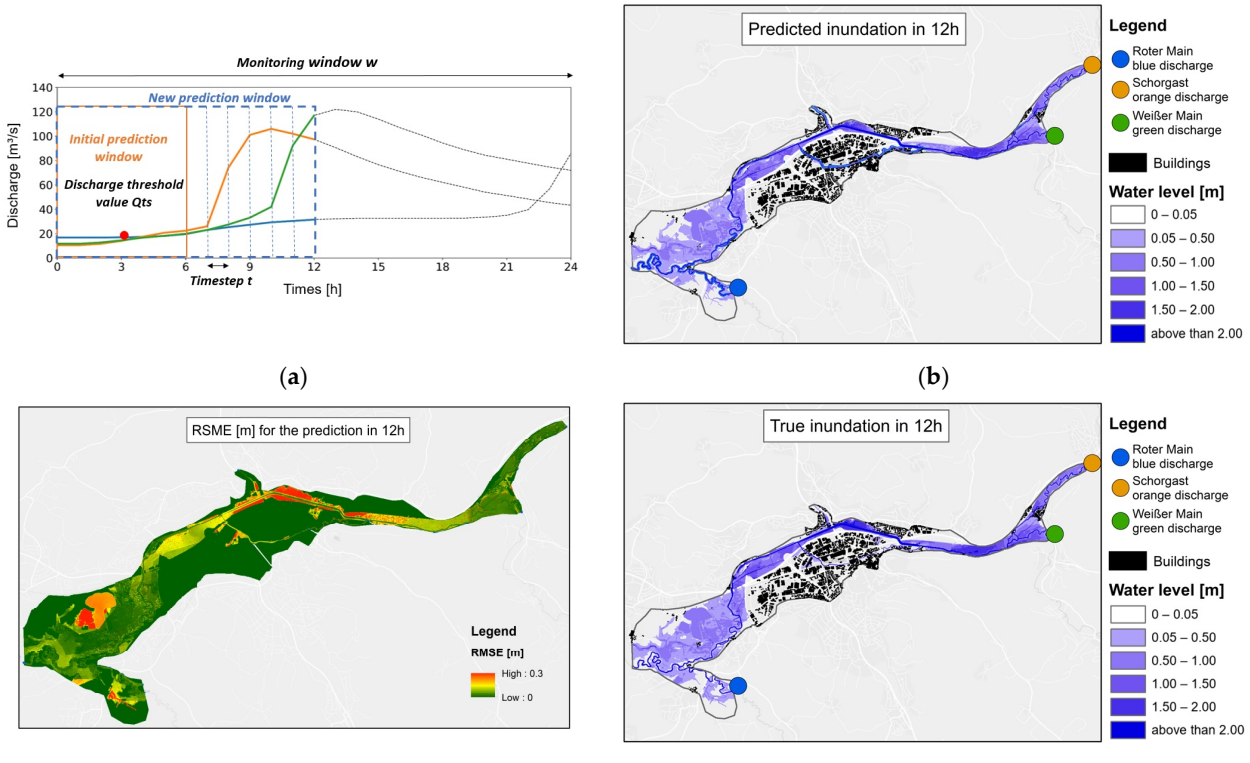


(**a**) (**b**)



(**c**) (**d**)

Figure 7. Visualization of the prediction for the next 6 h of the observed event from 2006. (**a**) Hydrographs within the monitoring window *w*, dotted timesteps are not used for the initial forecast, the orange box represents the initial prediction window, and the red dot represents the discharge threshold *Qts*; (**b**) predicted inundation map; (**c**) true inundation for this timestep; (**d**) spatially distributed RMSE prediction error.



(**a**) (**b**)

(**c**) (**d**)

Figure 8. Visualization of a multi-step prediction for the next 12 h of the observed event from 2006. (**a**) Hydrographs within the monitoring window *w*, dotted timesteps are not used for the forecast, the orange box represents the initial prediction window, the red dot represents the discharge threshold *Qts*, and the blue box represents the new prediction window including the timesteps t (visualized by the thin blue lines); (**b**) predicted inundation map; (**c**) spatially distributed RMSE prediction error; (**d**) true inundation for this timestep.

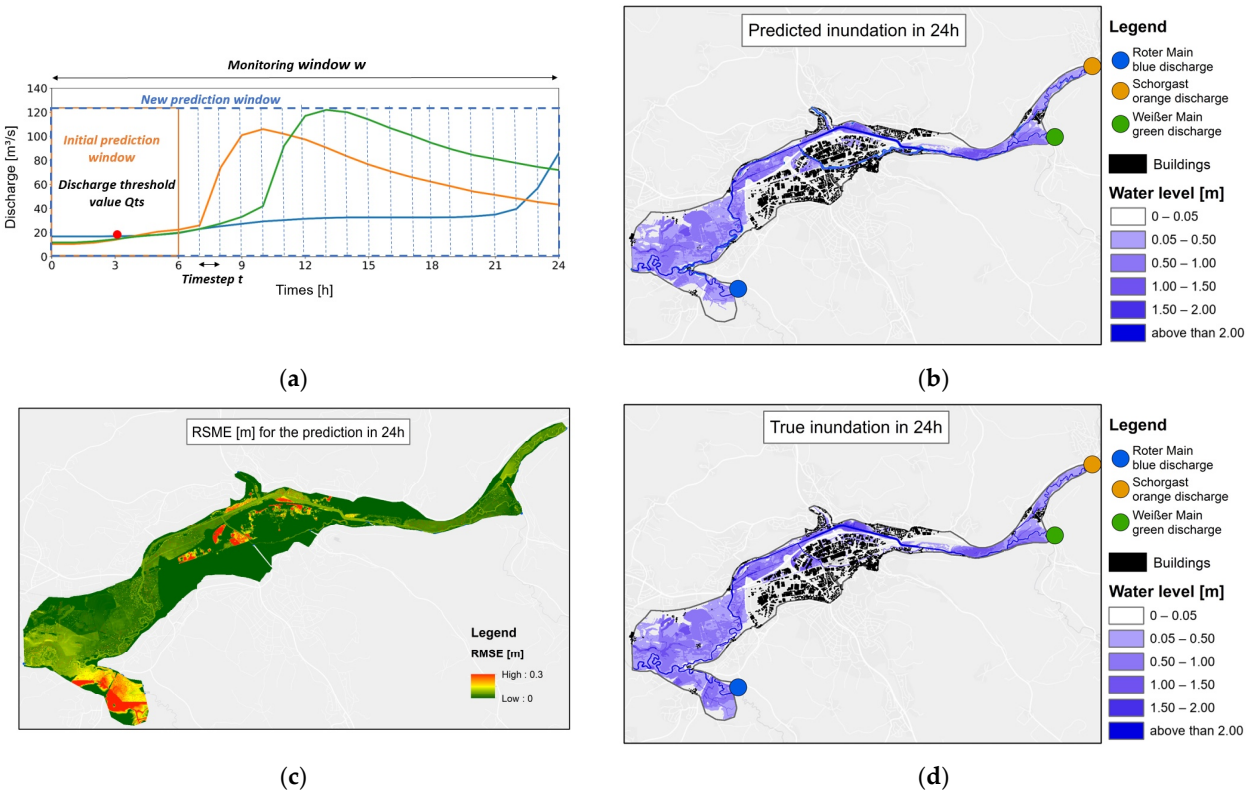


Figure 9. Visualization of a multi-step prediction for the next 24 h of the observed event from 2006. (**a**) Hydrographs within the monitoring window w, dotted timesteps are not used for the forecast, the orange box represents the initial prediction window, the red dot represents the discharge threshold *Qts*, and the blue box represents the new prediction window including the timesteps t (visualized by the thin blue lines); (**b**) predicted inundation map; (**c**) spatially distributed RMSE prediction error; (**d**) true inundation for this timestep.

5. Discussion

5.1. Operational Design

Since we focused on the development of the data-driven model and implemented our method for the sake of demonstration based on the available dataset, no critical statement on the implementation of the operational design can be made. However, we would like to mention that the operational design is extremely flexible; thus, any requirement for any study area can be implemented. For a longer lead time prediction, the monitoring window just needs to be increased. If someone aims for a higher temporal resolution, the timestep within the monitoring window can be lower. Thus, our method is clearly beneficial for the implementation of real-world applications.

5.2. Hyperparameters and Optimization

In this study, the number of filters and number of neurons in the recursive feature path were considered flexible hyperparameters. Table 1 shows the tested configurations, while Table 2 shows the validation results of these configurations. The three configurations add more parameters (filters/neurons) to the model. In general, each additional parameter in a data-driven model increases the number of parameters that need to be learned during training. This leads to higher computational requirements during both training and inference phases, which can be a problem, especially for resource-constrained environments. Nevertheless, we were able to train all of these configurations and could investigate adding more parameters.

Adding more parameters also increases the model's capacity to memorize the training data, which can lead to overfitting. Overfitting occurs when the model performs well on the training data but fails to generalize to unseen data. This is exactly what happened while using configuration 3. From Table 2, a better training accuracy can be seen, while the validation accuracy becomes worse in comparison to the other configurations. This affirmed us to take configuration 2, since in comparison to configuration 1, it had better validation accuracy.

The training and validation results for configuration 2 shown in Table 2 are very good. This indicates the effectiveness, firstly, of the feature-informed architecture and, secondly, of the optimizing workflow (retraining each path) as well as the omission of the recursive path and training only on the true flooding values. We would also like to mention that the recurve path, which is used in the operational forecasting mode, allows the integration of updating procedures. For example, real-time measurements taken at the study site could be integrated to update the predicted inundation map before the map is fed back into the model. This circumstance is also beneficial for real-world applications.

5.3. Performance Evaluation of the Forecast System

5.3.1. Critical Success Index (CSI)

Table 3 shows the results of the CSI evaluation, where the limit value is considered to be 0.05 m. This threshold is also critical since values of this magnitude are typically not visualized in flood maps. The table shows that all test data reach values of over 90%. This very high performance originates from the feature-informed architecture and the recursive informed loop, which also maintains the feature information for the next input. The high values across the test datasets indicate that the predictive model performs very well in classifying cells correctly in flooded cells (cells with water depth higher than 0.05 m) and safe cells (water depth < 0.05 m); thus, the model is highly accurate in predicting the flooding extent. This also indicates that the forecast system has achieved its goal of providing high-quality predictions of water depth. Stakeholders and decision-makers can have confidence in using the forecast system to make decisions based on the flooding extent.

5.3.2. Overall Performance and Lead Time Prediction Error

Figure 6 shows the average prediction quality as RMSE for each timestep to predict. All timesteps have RMSE values below 0.15 m, which indicates high prediction quality. Of course, the overall prediction becomes worse over the prediction timesteps and reaches a maximum in general at 24 h, since the new predictions also include the prediction error of the previous timestep fed back via the recursive path. Nevertheless, the feature-informed recursive path is able to integrate the inundation from the previous timestep in a satisfactory manner. One could argue that an hourly flood map (new map every hour) is not suitable for decision making in case of an emergency, but this resolution also keeps the uncertainty within the maps low and ensures that there are no bigger jumps in the inundations. For practical reasons, the timestep for a communicated map could be set to a higher interval (for example, 3 h).

The observed event for the year 2005 performs the worst at the beginning. The reason for this is that the available discharge data for this event already start higher than the threshold value of 25 m^3/s. The results of the observed event from 2013 are the worst in total since this event is so different from the training dataset. Either expanding the training set or splitting the features in the informed layers into more categories are possible strategies.

5.3.3. Inundation Maps and Spatial Prediction Error Distribution

Figures 7–9 show inundation for the timesteps 6, 12, and 24 h for the observed event in 2006. The inundation maps in figure parts (b and d) show no real visual difference from

the true inundation maps from the HEC-RAS 2D model. This shows that the data-driven approach is a real alternative to the physically based model. Part (c) of these figures also shows the spatially distributed prediction error as RMSE. It can be seen that the predictions have an RMSE below 0.3 m, meaning that the prediction error is never worse than 30 cm. In Figure 9c, it can be seen that the southwestern part of the study area has the worst predictions. This is due to the fact that the hydrograph at this entry point is rising, while the other two are falling. Such a scenario is not covered within the training data. Expanding the training data would be an appropriate approach to deal with this issue.

Lin et al. (2020) [23] also developed a multi-step prediction model for this study area using the same dataset. However, the authors of this study only applied forecasts for the next 3, 6, 9, and 12 h with a network without a recursive path. For the observed event from 2006, they concluded that 89,63% and 82,40% of the predicted values had an RMSE below 0.3 m for the timesteps t = 6 h and t = 12 h, respectively. Hereby, the recursive and informed strategy of our approach shows its benefit since all predicted values for all timesteps have RMSE values below 0.3 m.

5.4. Limitations and Outlook

It should be mentioned that for running our multi-step inundation forecast, at least one hydrograph is required; otherwise, no forecast is possible. This hydrograph must be predicted in advance, and, thus, it also has to be mentioned that all of the uncertainties and data quality issues, which come hand in hand with the hydrograph, are also transferred into the data-driven model, leading to potentially poor performance. Furthermore, since the forecast system is trained on the hydrodynamic HEC-RAS 2D model results, all sources of uncertainty within this model are also transferred into the data-driven model. Nevertheless, HEC-RAS 2D is considered to be state of the art for the generation of inundation maps.

Also, if the forecast system has to be implemented within a different study area, the database for training, validation, and testing has to be generated beforehand by running a physically based model, and training has to be performed again, since the data-driven model is validated only to this specific study area. However, to date, there is no data-driven model available that can be transferred into a new study while keeping high prediction quality without retraining. Future research should investigate strategies for optimizing transferability.

In this study, we applied a monitoring window of 24 h; thus, it allows only discharge time series for the next 24 h. Longer series require a longer monitoring window. Nevertheless, this could be achieved with the current architecture by adjusting the parameters of the operational design and a re-training process. Additionally, in the current configuration of the data-driven multi-step forecast system, the predictions have to be stopped once the event reaches the current time. Further research should investigate strategies for also providing predictions during the event.

6. Conclusions

This study presented the development of a data-driven multi-step inundation forecast system for fluvial flood events, which provides a dynamic inundation map. The forecast system is set up over a monitoring window, where a hydrograph is listed. Once the hydrograph reaches a certain pre-defined discharge threshold, the inundation prediction is started. Therefore, the timesteps of the hydrograph responsible for the inundation are used as input for the data-driven model. The data-driven model consists of a convolutional neural network with the architecture of a residual network (ResNet18), feature-informed dense layers, and a recursive path. For the first prediction, the recursive path is not used. The recursive path is added once more timesteps of the hydrograph are relevant for the inundation. This path allows the model to predict multiple steps into the future. This procedure outputs a dynamic inundation map over the entire monitoring window. The

temporal resolution of the dynamic inundation map is given the timestep at which the hydrograph is updated, and the maximum lead time is dependent on the length of the monitoring window. Test results show very high prediction quality; for example, for the next 6 h, 12 h, and 24 h, RMSE values of 0.07 m, 0.12 m, and 0.15 m are, respectively, reached. Furthermore, the predictions are performed within seconds, making the forecast system capable of real-time predictions in comparison to traditional physically based models. Unlike other studies, our data-driven model combines the recursive path with the input of additional knowledge about flooding into the data-driven model by the use of the feature-informed architecture. This led to high prediction quality even for multiple timesteps into the future. Additionally, the algorithm for running the whole forecast system is designed from an operation point of view; thus, it can be transferable to other real-world applications.

Author Contributions: Conceptualization, F.S.; methodology, F.S.; software, F.S.; validation, F.S., data curation, F.S.; writing—original draft preparation, F.S.; writing—review and editing, J.L.; visualization, F.S.; supervision, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research presented in this paper has been carried out as part of the HiOS project (Hinweiskarte Oberflächenabfluss und Sturzflut) funded by the Bayerisches Staatsministerium für Umwelt und Verbraucherschutz (StMUV) [Bavarian State Ministry of the Environment and Consumer Protection] (69-0270-92086/2017) and supervised by the Bayerisches Landesamt für Umwelt (LfU) [Bavarian Environment Agency].

Data Availability Statement: The datasets generated for this study are available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arnell, N.W.; Lloyd-Hughes, B. The Global-Scale Impacts of Climate Change on Water Resources and Flooding under New Climate and Socio-Economic Scenarios. *Clim. Chang.* **2014**, *122*, 127–140. https://doi.org/10.1007/s10584-013-0948-4.
2. Calvin, K.; Dasgupta, D.; Krinner, G.; Mukherji, A.; Thorne, P.W.; Trisos, C.; Romero, J.; Aldunce, P.; Barrett, K.; Blanco, G.; et al. IPCC, 2023: Climate Change 2023: Synthesis Report. In *Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Core Writing Team, Lee, H., Romero, J., Eds.; First Intergovernmental Panel on Climate Change (IPCC): Geneva, Switzerland, 2023.
3. Swiss RE Institue Sigma. Natural Catastrophes in 2021: The Floodgates Are Open. Available online: https://www.swissre.com/institute/research/sigma-research/sigma-2022-01.html (accessed on: 13 September 2024).
4. Kuller, M.; Schoenholzer, K.; Lienert, J. Creating Effective Flood Warnings: A Framework from a Critical Review. *J. Hydrol.* **2021**, *602*, 126708. https://doi.org/10.1016/j.jhydrol.2021.126708.
5. Rözer, V.; Peche, A.; Berkhahn, S.; Feng, Y.; Fuchs, L.; Graf, T.; Haberlandt, U.; Kreibich, H.; Sämann, R.; Sester, M.; et al. Impact-Based Forecasting for Pluvial Floods. *Earth's Future* **2021**, *9*, 2020EF001851. https://doi.org/10.1029/2020EF001851.
6. Crotti, G.; Leandro, J.; Bhola, P.K. A 2D Real-Time Flood Forecast Framework Based on a Hybrid Historical and Synthetic Runoff Database. *Water* **2019**, *12*, 114. https://doi.org/10.3390/w12010114.
7. Schmid, F.; Leandro, J. A Feature-Informed Data-Driven Approach for Predicting Maximum Flood Inundation Extends. *Geosciences* **2023**, *13*, 384.
8. Henonin, J.; Russo, B.; Mark, O.; Gourbesville, P. Real-Time Urban Flood Forecasting and Modelling–a State of the Art. *J. Hydroinformatics* **2013**, *15*, 717–736. https://doi.org/10.2166/hydro.2013.132.
9. Berkhahn, S.; Neuweiler, I. Data Driven Real-Time Prediction of Urban Floods with Spatial and Temporal Distribution. *J. Hydrol. X* **2024**, *22*, 100167. https://doi.org/10.1016/j.hydroa.2023.100167.
10. Sanders, B.F.; Schubert, J.E.; Detwiler, R.L. ParBreZo: A Parallel, Unstructured Grid, Godunov-Type, Shallow-Water Code for High-Resolution Flood Inundation Modeling at the Regional Scale. *Adv. Water Resour.* **2010**, *33*, 1456–1467. https://doi.org/10.1016/j.advwatres.2010.07.007.
11. Delestre, O.; Darboux, F.; James, F.; Lucas, C.; Laguerre, C.; Cordier, S. FullSWOF: Full Shallow-Water Equations for Overland Flow. *J. Open Source Softw.* **2017**, *2*, 448. https://doi.org/10.21105/joss.00448.
12. Morales-Hernández, M.; Sharif, M.B.; Kalyanapu, A.; Ghafoor, S.K.; Dullo, T.T.; Gangrade, S.; Kao, S.-C.; Norman, M.R.; Evans, K.J. TRITON: A Multi-GPU Open Source 2D Hydrodynamic Flood Model. *Environ. Model. Softw.* **2021**, *141*, 105034. https://doi.org/10.1016/j.envsoft.2021.105034.
13. Hou, J.; Kang, Y.; Hu, C.; Tong, Y.; Pan, B.; Xia, J. A GPU-Based Numerical Model Coupling Hydrodynamical and Morphological Processes. *Int. J. Sediment Res.* **2020**, *35*, 386–394. https://doi.org/10.1016/j.ijsrc.2020.02.005.
14. Seleem, O.; Ayzel, G.; De Souza, A.C.T.; Bronstert, A.; Heistermann, M. Towards Urban Flood Susceptibility Mapping Using Data-Driven Models in Berlin, Germany. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1640–1662. https://doi.org/10.1080/19475705.2022.2097131.

15. Li, Y.; Martinis, S.; Wieland, M. Urban Flood Mapping with an Active Self-Learning Convolutional Neural Network Based on TerraSAR-X Intensity and Interferometric Coherence. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 178–191. https://doi.org/10.1016/j.isprsjprs.2019.04.014.
16. Solomatine, D.; See, L.M.; Abrahart, R.J. Data-Driven Modelling: Concepts, Approaches and Experiences. Water Science and Technology Library. In *Practical Hydroinformatics*; Abrahart, R.J., See, L.M., Solomatine, D.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; Volume 68, pp. 17–30; ISBN 978-3-540-79880-4.
17. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* **2015**, arXiv:1505.04597.
18. Liu, C.; Tao, Y.; Liang, J.; Li, K.; Chen, Y. Object Detection Based on YOLO Network. In Proceedings of the 018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 14–16 December 2018; pp. 799–803.
19. Shao, P.; Feng, J.; Lu, J.; Zhang, P.; Zou, C. Data-Driven and Knowledge-Guided Denoising Diffusion Model for Flood Forecasting. *Expert Syst. Appl.* **2024**, *244*, 122908. https://doi.org/10.1016/j.eswa.2023.122908.
20. Xu, C.; Zhong, P.; Zhu, F.; Xu, B.; Wang, Y.; Yang, L.; Wang, S.; Xu, S. A Hybrid Model Coupling Process-Driven and Data-Driven Models for Improved Real-Time Flood Forecasting. *J. Hydrol.* **2024**, *638*, 131494. https://doi.org/10.1016/j.jhydrol.2024.131494.
21. Berkhahn, S. An Ensemble Neural Network Model for Real-Time Prediction of Urban Floods. *J. Hydrol.* **2019**, *575*, 743–754.
22. Lin, Q.; Leandro, J.; Wu, W.; Bhola, P.; Disse, M. Prediction of Maximum Flood Inundation Extents with Resilient Backpropagation Neural Network: Case Study of Kulmbach. *Front. Earth Sci.* **2020**, *8*, 332. https://doi.org/10.3389/feart.2020.00332.
23. Lin, Q.; Leandro, J.; Gerber, S.; Disse, M. Multistep Flood Inundation Forecasts with Resilient Backpropagation Neural Networks: Kulmbach Case Study. *Water* **2020**, *12*, 3568. https://doi.org/10.3390/w12123568.
24. Löwe, R.; Böhm, J.; Jensen, D.G.; Leandro, J.; Rasmussen, S.H. U-FLOOD–Topographic Deep Learning for Predicting Urban Pluvial Flood Water Depth. *J. Hydrol.* **2021**, *603*, 126898. https://doi.org/10.1016/j.jhydrol.2021.126898.
25. Guo, Z.; Leitão, J.P.; Simões, N.E.; Moosavi, V. Data-driven Flood Emulation: Speeding up Urban Flood Predictions by Deep Convolutional Neural Networks. *J. Flood Risk Manag.* **2021**, *14*, e12684. https://doi.org/10.1111/jfr3.12684.
26. Hofmann, J.; Schüttrumpf, H. floodGAN: Using Deep Adversarial Learning to Predict Pluvial Flooding in Real Time. *Water* **2021**, *13*, 2255. https://doi.org/10.3390/w13162255.
27. Guo, Z.; Moosavi, V.; Leitão, J.P. Data-Driven Rapid Flood Prediction Mapping with Catchment Generalizability. *J. Hydrol.* **2022**, *609*, 127726. https://doi.org/10.1016/j.jhydrol.2022.127726.
28. Schmid, F.; Leandro, J. An Ensemble Data-Driven Approach for Incorporating Uncertainty in the Forecasting of Stormwater Sewer Surcharge. *Urban Water J.* **2023**, *20*, 1140-1156. https://doi.org/10.1080/1573062X.2023.2240309.
29. Burrichter, B.; Hofmann, J.; Koltermann da Silva, J.; Niemann, A.; Quirmbach, M. A Spatiotemporal Deep Learning Approach for Urban Pluvial Flood Forecasting with Multi-Source Data. *Water* **2023**, *15*, 1760. https://doi.org/10.3390/w15091760.
30. Chiang, Y.-M.; Chang, L.-C.; Tsai, M.-J.; Wang, Y.-F.; Chang, F.-J. Dynamic Neural Networks for Real-Time Water Level Predictions of Sewerage Systems-Covering Gauged and Ungauged Sites. *Hydrol. Earth Syst. Sci.* **2010**, *11*, 1309–1319.
31. Chang, F.-J.; Chen, P.-A.; Lu, Y.-R.; Huang, E.; Chang, K.-Y. Real-Time Multi-Step-Ahead Water Level Forecasting by Recurrent Neural Networks for Urban Flood Control. *J. Hydrol.* **2014**, *517*, 836–846. https://doi.org/10.1016/j.jhydrol.2014.06.013.
32. Ludwig, K.; Bremicker, M. The Water Balance Model LARSIM-Design, Content and Applications. In *Freiburger Schriften zur Hydrologie*; Institut für Hydrologie: Amsterdam, The Netherlands, 2006; p. 22.
33. US Army Corps of Engineers. *H.E.C. HEC-RAS River Analysis System-2D Modeling Users Manual, Version 5.0*; US Army Corps of Engineers: Davis, CA, USA, 2016.
34. O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. *arXiv* **2015**, arXiv:1511.08458.
35. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827.
36. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
38. Wunsch, A.; Liesch, T.; Broda, S. Forecasting Groundwater Levels Using Nonlinear Autoregressive Networks with Exogenous Input (NARX). *J. Hydrol.* **2018**, *567*, 743–758. https://doi.org/10.1016/j.jhydrol.2018.01.045.
39. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
40. Bhola, P.K. Dynamic Flood Inundation Forecasting in Real-Time Including Associated Uncertainties for Operational Flood Risk Management. Dissertation TU München, München, Germany, 2019.
41. Landsamt für Umwelt (LfU). Hydrogeologische Karte. Available online: https://www.lfu.bayern.de/geologie/hydrogeologie_karten_daten/hk50/index.htm (accessed on: 13 September 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

A physically informed domain-independent data-driven inundation forecast model

Water Research 289 (2026) 124819

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

A physically informed domain-independent data-driven inundation forecast model

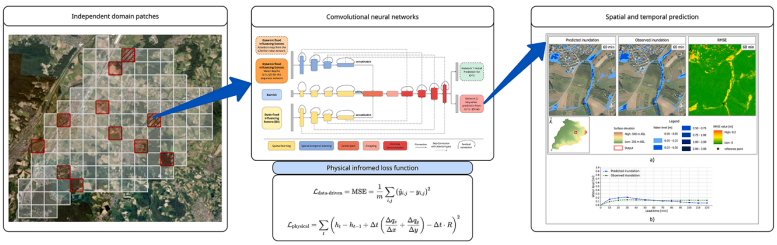
Felix Schmid *, Leonie Müller, Jorge Leandro

Department of Civil Engineering, Chair of Hydromechanics and Hydraulic Engineering, Research Institute Water and Environment, University of Siegen, 57076 Siegen, Germany

HIGHLIGHTS

- Domain-independent data-driven inundation forecasts for unseen areas.
- Physics-informed learning based on 2D continuity & kinematic wave assumption.
- Improved prediction quality by about 25 % compared to classical data-driven loss.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Urban flooding
Spatio-temporal prediction
Physics-informed neural network
Continuity & kinematic wave loss
Generalization to unseen catchments
Deep learning

ABSTRACT

Inundation maps with spatial and temporal distribution of the water depths are essential for protecting the population in case of pluvial flood events. Creating these maps in operational forecasting is currently not possible with traditional physically-based numerical models, as these are too slow for real-time predictions. Data-driven models are able to produce predictions in real-time, however, due to their domain-specific training, they are only applicable to the respective study site. Therefore, in this study, we propose a physically informed data-driven forecast system to overcome this limitation and provide spatial and temporal forecasts of water depth inundations in unknown areas. Our data-driven model is developed based on data from the catchment of Baiersdorf in Germany. It follows a Convolutional Neural Network (CNN) based on an image-to-image translation process and is trained on various flood-influencing factors, which represent catchment characteristics. We proposed a specific spatiotemporal prediction framework that: (1) enables temporal time-stepping of 10 min, higher than physically based hydraulic models with seconds, (2) data-driven domain-independent forecasts, tested on 23 unknown areas by a cross-validation, and (3) eliminates the need for downsampling for larger catchments (typical of data-driven forecast systems). Further, we integrate a 2-dimensional continuity equation together with a kinematic wave formulation for estimating the velocity in the loss function to enforce physically informed forecasts. Results on unknown areas produce Critical Success Index (CSI) values of about 74 % and mean Root Mean Squared Error (RMSE) values of 0.045 m. Our physically informed loss function was able to outperform a classical data-driven loss function and improved the RMSE by about 25 %.

* Corresponding author.
E-mail addresses: Felix.Schmid@uni-siegen.de (F. Schmid), Leonie.Mueller@uni-siegen.de (L. Müller), Jorge.Leandro@uni-siegen.de (J. Leandro).

https://doi.org/10.1016/j.watres.2025.124819
Received 15 June 2025; Received in revised form 6 October 2025; Accepted 14 October 2025
Available online 15 October 2025
0043-1354/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Extreme rainfall events pose a major challenge. Particularly in urban areas characterized by high sealing rates, precipitation quickly turns into runoff followed by pluvial flooding, once structural measures are overwhelmed (Henonin et al., 2013). Such flood events are one of the most severe natural hazards for urban life and are likely to occur more often due to climate change (Calvin et al., 2023). Thus, early warning systems and flood forecasts are of great interest to protect the population and infrastructure, and to prevent or mitigate damage (Henonin et al., 2013).

In recent years, flood forecasting has started to provide information on possible inundation extents based on the incoming hydrograph or precipitation. Such an inundation extent is typically provided by a map with a fine resolution (e.g. 1×1 m or 5×5 m), showing spatially distributed water depths in the respective study site, helping to quickly identify dangerous regions (Crotti et al., 2019) and allowing, for example, the optimization of rescue routes. These maps are commonly produced by hydrodynamic models, which solve the Shallow-water equations based on numerical solvers. However, such numerical solvers require time steps smaller than a second, leading to very large computational times (e.g. hours). To reduce the high computation time, recent studies have developed numerical solvers that utilize graphical processing units (GPUs). This approach leads to improvements in reducing computational times by a factor of up to 15 times (Morales-Hernández et al., 2021). Although the run times could be reduced, physically based models are still not suitable in the context of real-time forecasting. Especially for urban flooding, where lead times are crucial and precipitation patterns can change quickly (Berkhahn and Neuweiler, 2024).

Hence, data-driven inundation modelling is becoming increasingly important in the research community, as it is more computationally efficient and enables real-time predictions. A data-driven generation of inundation maps showing maximum water levels in one catchment and for different events has been demonstrated for pluvial events by Berkhahn et al. (2019) using multi-layer perceptron neural networks and for fluvial events by using a resilient backpropagation neural network (Lin et al., 2020a). Schmid and Leandro (2023a) improved the performance of such a system by using a feature-informed convolutional neural network that provides the distance to water bodies as a GIS feature in the architecture. The output inundation map of these studies matches the spatial resolution of the hydrodynamic model used beforehand to generate the maps. Thus, these models are able to replace the hydrodynamic one and provide for real-time forecasting. Follow-up studies focused on the integration of temporal time-stepping by training multiple networks for different timesteps (Lin et al., 2020b), by recursive connections (Berkhahn and Neuweiler, 2024; Schmid and Leandro, 2024) or a sequenced approach (Cao et al., 2025) for providing a dynamic flood inundation map. However, these developed models are domain-specific, meaning that the forecast system can only be used in the respective site since it only learned the hydraulic behaviour represented in the dataset. For setting up such a forecast system in a different site, the hydraulic dataset for training has to be created beforehand, and training has to be carried out again, making such a methodology time-consuming and increasing the data-storage demand considerably.

A step further towards domain independence was reached by Löwe et al. (2021). The authors trained a data-driven model to predict maximum water depths on many different inundation image frames by providing rainfall and catchment characteristics as input parameters, enabling a domain-independent model for the first time. Guo et al. (2022) and Seleem et al. (2022) also followed this approach but trained on total catchments. Training on whole catchments improved the performance, as flow paths were not divided as in Löwe et al. (2021). Nevertheless, the size of the catchments played a crucial role, since bigger catchments cannot be predicted due to computational storage limits and thus downsampling was applied. Even by providing physical catchment parameters, those models sometimes lacked prediction quality, required large datasets, and none of the models considered temporal time-stepping.

As data-driven approaches may also produce physically unrealistic predictions, physics-informed neural networks (PINNs) have gained increasing attention. Classical PINNs (Raissi et al., 2019) are neural networks in which the governing partial differential equations (PDEs) are explicitly embedded into the loss function, and their residuals are enforced at collocation points in space and time. This forces the model to approximately satisfy known physical laws while learning from data, thereby improving physical consistency. Feng et al. (2023) and Mahesh et al. (2022) developed a 1D PINN by incorporating the Saint-Venant equation into the loss function. The results were promising, but only applicable to one-dimensional stream-flows. Donnelly et al. (2022) showed that Gaussian Process emulation can act as a surrogate for 2D flood models, providing efficient spatio-temporal predictions but with limitations in transferability and capturing full nonlinear dynamics. Building on this line of research, Donnelly et al. (2024) developed a model governed by the 2D shallow water equation, which outperformed state-of-the-art data-driven models, however, the authors did not design their model for domain-independent prediction and transferability to unseen areas.

Since classical PINNs explicitly embed the residuals of the governing PDEs into the loss function, they typically require access to all dependent variables and their derivatives (water depth and velocity), making them computationally intensive in higher dimensions. In contrast, we proposed a physically informed forecast system (PI-FS) based on a convolutional neural network that is trained with rasterised geospatial and temporal input data. It integrates physical knowledge via a mass-conservation-based loss function, which is derived from the continuity equation under the kinematic wave assumption and is evaluated over the full prediction grid, without the need for explicit velocity inputs or collocation points.

The objective of this study is therefore to develop and evaluate a pluvial 2D, domain-independent, physically informed, data-driven forecast system (PI-FS) with temporal time-stepping, and to test its generalization capability in unseen areas. Once established, the PI-FS willprovide time-dependent predictions of inundation extent and depth from rainfall in any area, making it, to the best of our knowledge, the first study of its kind.

2. Materials and methods

Following our aim of providing spatial and temporal inundation depths in any area, we first need to deal with multi-grid-wise data. Therefore we utilize a convolutional neural network (CNN) architecture based on a surrogate image-to-image translation process, where a cell in the image represents one cell from the physically-based numerical model. Like other studies (Berkhahn and Neuweiler, 2024; Cao et al., 2025; Löwe et al., 2021), we apply an encoder-decoder setup. However, we enhance this approach by using multiple encoders, each of which processes different types of data and later combines them in the network. In addition, we use similar to Cao et al. (2025) novel attention gates (Oktay et al., 2018) in the skip connection to help the network learn the most relevant features. As flood inundation in unknown areas is predicted, we provide suitable static flood-influencing factors that describe the physical area and dynamic-influencing factors, which describe flood dynamics. Based on these inputs inundation characteristics can be learned and physical domain knowledge is considered in the prediction. To further ensure physically reasonable learning, which is crucial when the network is deployed in a new unknown study area, we introduce similar to Donnelly et al. (2024), a physically informed loss function into the optimization process. This enables the network to learn the physical mechanism of flooding and establish a better generalization ability between the physical input data and the desired inundation data. Finally, we implement our physically informed, data-driven forecast

system (PI-FS) into a specific spatial and temporal prediction framework, where the input frame to the PI-FS is larger than the prediction output, which lies in the centre of the input frame. In this way, we introduce information from outside of the area that will be predicted into the network. Section 2.1 outlines the spatial and temporal prediction framework, followed by a description of the PI loss function and flood-influencing factors in Section 2.2. The network architecture is detailed in Section 2.3.

2.1. Spatiotemporal prediction framework

The PI-FS utilizes a CNN architecture. To enhance the model's ability to capture spatial dynamics, each input frame is provided at a larger input frame, while the model outputs predictions, focusing on the central subregion. This spatial prediction framework is presented in Fig. 1. The central area is selected to avoid edge effects and to ensure that predictions benefit from the surrounding context, which helps capture incoming water fluxes.

This design choice is particularly beneficial, where the movement of water across terrain boundaries plays a critical role. By including a larger spatial context around the prediction area, the model gains access to important upstream and surrounding information that can influence fluxes into the forecasted region. It allows the network to learn how external water in/outflows impact inundation behaviour within the core prediction zone. Essentially, the larger input acts as a buffer zone, enabling the model to better simulate boundary-driven flood propagation and improve the physical correctness of the predictions. To apply this approach across an entire hydrological study area, we operate this framework like a sliding scanner, where the model processes the study area in overlapping tiles. For each tile, it receives the wider input frame (visualized in blue in Fig.1), centered around the region of interest, and generates a smaller output prediction for the central area (visualized in red in Fig.1). By systematically sliding/iterating across the domain in a grid-like fashion, the model generates predictions for the entire area (like the white "Further outputs" in Fig.1 are indicating). The overlap between input tiles allows for smooth transitions and consistency at the boundaries of each predicted subregion (Berkhahn and Neuweiler, 2024). This tiling strategy makes the framework scalable to large geographic extents while still leveraging the benefits of high-resolution local predictions enriched by contextual information from adjacent areas. We applied an input resolution of 512×512 and predicted the 256×256 cell in the centre of the input.

The forecasting system is integrated into a specific temporal framework (Fig. 2). We apply two networks within the forecast system. The first timestep is predicted by an initial network, which takes the rainfall, catchment-describing flood influencing factors, and an attention map. The attention map comes from a cellular automata submodel from Guidolin et al. (2016) and produces a first intermediate flood inundation depth map, which helps the network improve peak localization, while its influence on subsequent timesteps is minimal, as it is applied only to the initial network. The output from the initial network, in conjunction with the initial water depths at time t(0) (representing existing water in the domain), serves as input to a sequence-based network. This subsequent network leverages the provided past information to predict multiple future timesteps sequentially.

Our approach is especially beneficial since the initial network can focus only on the first prediction, while the sequence model is able to predict the subsequent timesteps in one forward pass. This not only improves computational efficiency but also allows the model to learn temporal dependencies in a more structured manner. To demonstrate

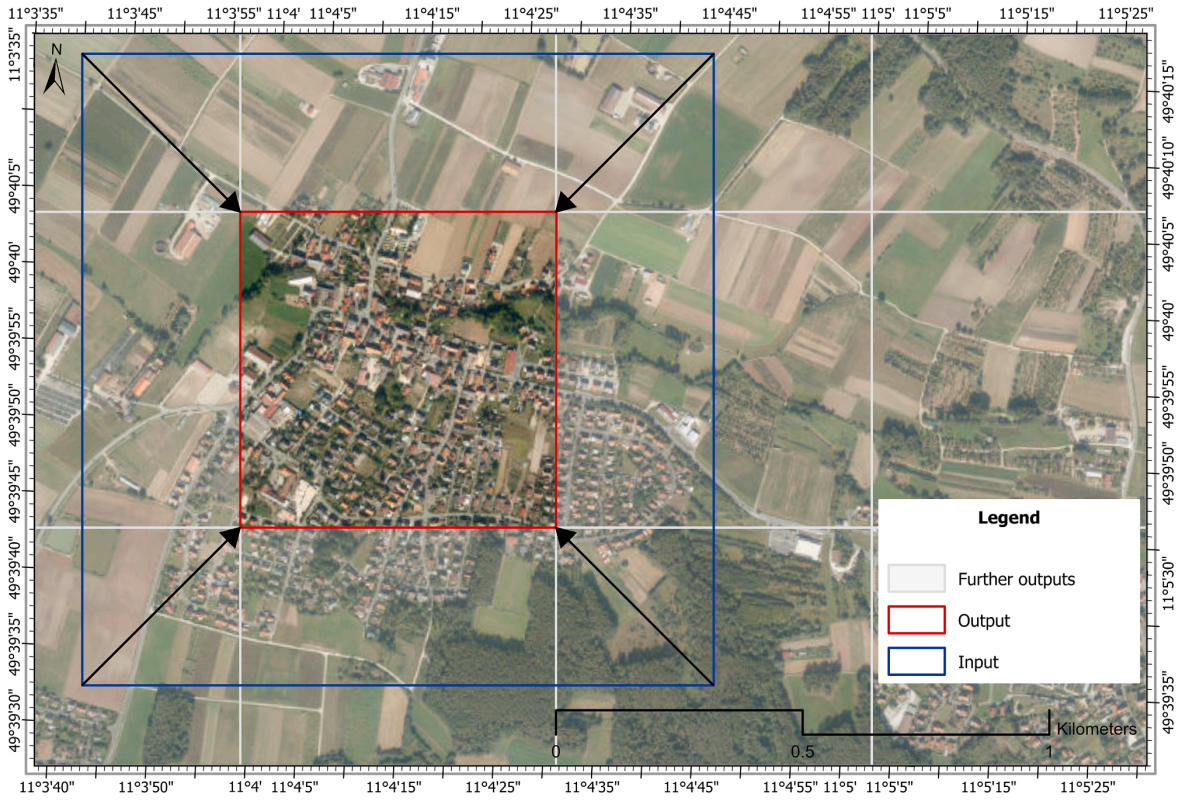


Fig. 1. Spatial prediction framework. The input (blue) to our CNN model covers a larger area than the output (red), allowing the model to integrate surrounding information. The "Further outputs" indicated the sliding scanner operating strategy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

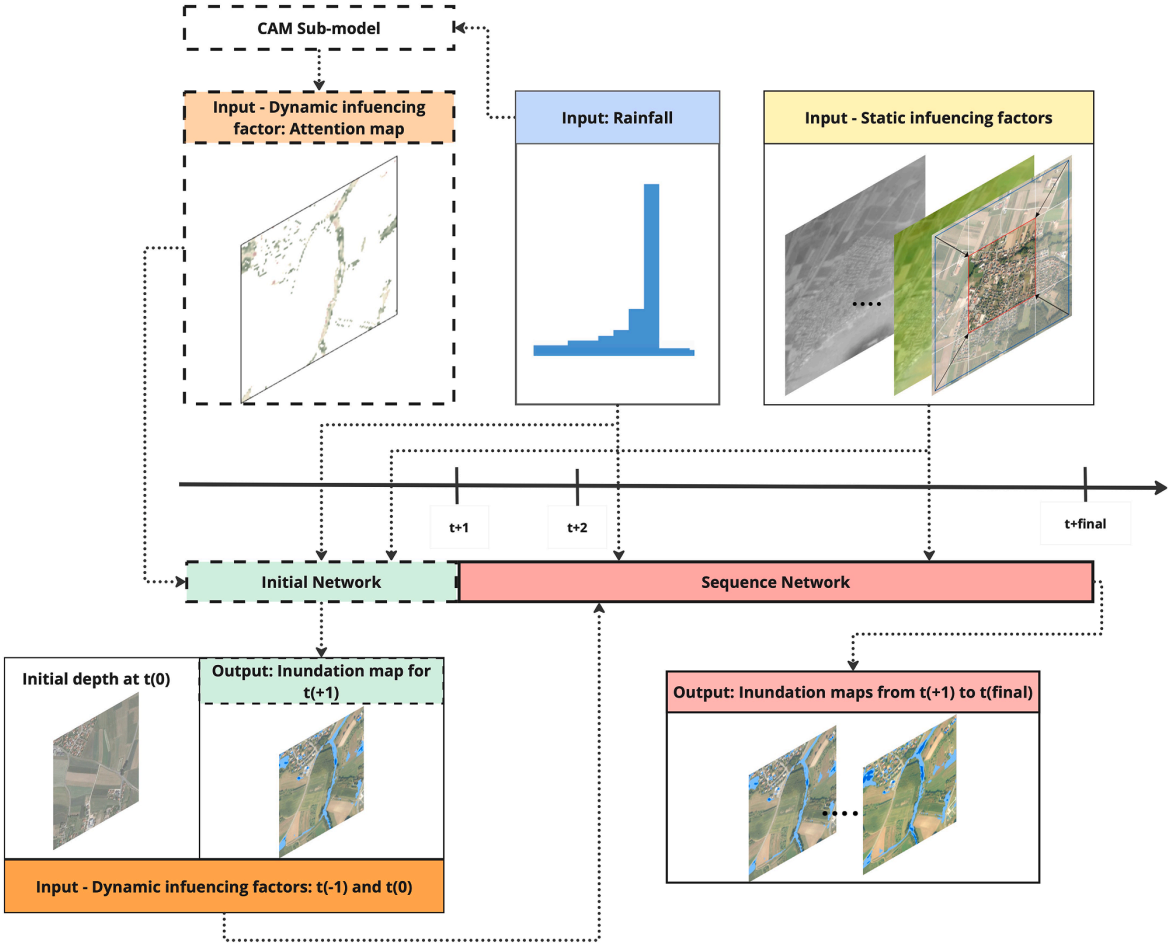


Fig. 2. Temporal prediction framework. The initial network (dotted green) uses static (catchment describing) flood influencing factors, rainfalls, and an attention map (generated by a cellular automata model) to predict the first timestep; The sequential network (solid red) predicts the upcoming timesteps to the final one at once, using the result from the initial network, initial depths, rainfall, and the static influencing factors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

our methodology, we do predictions up to 2 h t(final) and test different temporal time-stepping (10, 15, 20 min).

2.2. Physically informed generalization

To ensure transferability to an unknown area, the network is provided with physical input data (flood influencing factors), and a physically informed loss (PI) function is implemented in the optimization process to enable (enforce) the network to learn the physical mechanism of flood inundation.

2.2.1. Physically informed learning

We follow a supervised learning approach, where the desired output is given by the hydrodynamic model, and the prediction quality is measured by the agreement between the prediction and observed values. This agreement is calculated by the loss function (LF). Typically, a mean squared error (MSE) (Eq. (1)) is used, where $\widehat{y_{i,j}}$ is the water depth prediction in the respective cells, $y_{i,j}$ the observed depth given by the numerical model, and m is the number of events predicted.

$$LF_{data-driven} = MSE = \frac{1}{m} * \sum_{i,j} \left(\widehat{y_{i,j}} - y_{i,j}\right)^2 \tag{1}$$

During the training process, the weights in the network are adjusted so that a minimum value of the LF is found by the optimization algorithm. This LF, however, is purely data-oriented and is learning patterns mainly from the data alone. To ensure the transferability of the network however an understanding of the underlying physical mechanism is necessary. From a physical perspective, inundation extents and water depth are typically obtained by solving the shallow water equations, which apply the principles of mass continuity and momentum conservation ensuring water volume is conserved over time and water flows according to external forces like gravity, friction, and pressure gradients.

The 2-dimensional continuity equation is given by Eq. (2), where h is the water depth at location (x, y) at time t, (u, v) are velocity vector components, and R is a source term like rainfall. Following an explicit approximation and rearranging leads to Eq. (3) in which q_x and q_y describe the flux movement in the respective directions (Eq. (4)):

$$\frac{\partial h}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} = R \tag{2}$$

$$h_t - h_{t-1} + \Delta t * \left[\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y}\right] - \Delta t R = 0 \tag{3}$$

$$q_x = h_t u_t;\ q_y = h_t v_t \tag{4}$$

These equations are able to describe the water movement in a 2-dimensional grid. The PI-FS only forecasts water depths at each time-step, hence, the information on the velocity components (u, v) is not available. However, the flux movement q_x and q_y can be roughly approximated with the kinematic wave (in this case the Manning-Strikler formula) given by Eq. (5) where n in the Manning roughness, β the surface slope in the respective x and y direction, and h the water depth.

$$q = \frac{1}{n}\sqrt{\tan(\beta)}h^{5/3} \tag{5}$$

This equation is integrated into Eq. (3). Performing a final summation over the study area leads to the PI-LF (Eq. (6)).

$$LF_{physical} = \sum\left(h_t - h_{t-1} + \Delta t * \left[\frac{\Delta q_x}{\Delta x} + \frac{\Delta q_y}{\Delta y}\right] - \Delta t R\right) = 0 \tag{6}$$

By replacing h with $\widehat{y_{i,j}}$ (water depth predictions from the network) this equation forces the network to predict water depths, which leads to a velocity field that satisfies the mass conservation and therefore delivers physically meaningful predictions. Furthermore, the velocity is based on slope and roughness values, which therefore incorporates gravity and friction into the desired learning process and integrates the input parameters roughness and slope directly into the LF. Like other studies (Donnelly et al., 2024; Mahesh et al., 2022), we implemented our physically informed $LF_{physical}$ in combination with the data-driven $LF_{data-driven}$ from Eq. (1). To ensure the same units in the overall loss function, and since the data-driven has the units squared, we apply a power of two to the physical part. Thus, the final LF is given by Eq. (7), which enforces the network to learn from both physical laws and image patterns at the same time.

$$LF_{overall} = LF_{data-driven} + \left(LF_{pyhsical}\right)^2 = 0 \tag{7}$$

2.2.2. Flood influencing factors and input parameters

The network must be provided with suitable data that describes the underlying process. These are the pluvial flood influencing factors. For a better learnable relationship, we divide these influencing factors into three groups: Static flood influencing factors, which offer catchment characteristics based on the ideas of Yang et al. (2024), Löwe et al. (2021), and Guo et al. (2021), dynamic influencing factors, which provide information on the inundation in previous timesteps, and the rainfall itself.

We selected these parameters based on literature and expert reasoning. Common for all input parameters is that they all can be derived from geodata using standard raster processing tools. Table 1 summarizes the input parameters for our data-driven network and further explains the reasoning behind each parameter.

It is worth mentioning that we did not conduct correlation analysis, since our objective was domain-independent prediction. While certain predictors may correlate within specific catchments, they can capture distinct hydrological or topographic processes in others. Excluding them based on local correlations would risk losing physically meaningful descriptors necessary for generalization across diverse basins. We would like to mention that our study site focuses on pluvial and not on fluvial flooding. Water bodies in our area are made of small channels that typically fall dry in the absence of rainfall and therefore do not exhibit classical fluvial dynamics with upstream and down-stream boundary conditions.

2.3. Architecture of the data-driven networks

The data-driven networks are responsible for generating the respective inundation maps. Both networks, the initial and the sequence, follow the same architecture and only differ in their input and output data. As mentioned, we apply multiple encoders that extract different information about the inundation. After combining this information in the centre part, a decoder with attention gate skip connections is applied. Fig. 3 shows the architecture of these networks.

These inputs are explicitly separated from each other, as all data contains different information about the flooding. By separating them,

Table 1
The three groups of influencing factors used as input to the network.

Food influencing factors	
Static factors	**Reasoning**
Digital elevation model	Surface elevation including buildings, since it contains information on flow directions. Used in: Cao et al., 2025; Guo et al., 2021, 2022; Löwe et al., 2021; Seleem et al., 2023.
Slope	Slope angles, since they influence the flow velocities and are necessary for the physical loss. Used in: Cao et al., 2025; Guo et al., 2021, 2022; Löwe et al., 2021; Seleem et al., 2023.
Roughness	Manning roughness values, since they influence the water depths and are necessary for the physical loss function. Used in: Seleem et al., 2023
Aspect	Characterizes the flow direction in the terrain. Used in: Cao et al., 2025; Guo et al., 2021, 2022; Löwe et al., 2021; Seleem et al., 2023.
Curvature	Influences the flow direction in the terrain. Used in: Cao et al., 2025; Guo et al., 2021, 2022; Löwe et al., 2021; Seleem et al., 2023.
Channels	A raster layer containing 0 (no-channel) and 1 (channel-cell). Thus, it indicates small channels that have a storage of 10 cm (e.g. streets) in the area and therefore informs the network where inundation starts. Used in: Seleem et al., 2023.
Water bodies	A raster layer containing 0 (no water) and 1 (water-cell). Thus, it indicates initial water in the area and therefore informs the network where inundation starts or is already present. Used in: Seleem et al., 2023
Distance to a channel	A raster layer that contains distances of each cell to the nearest channel, indicating a higher tendency to flood. Used in: Schmid and Leandro, 2023a; Seleem et al., 2023.
Dynamic factors	
Attention map[1]	Assumed water depth for 5 min based on the CAM sub-model. This layer gives extra attention to the cells, which should be inundated at the first timestep.
Water depth t(−1): Output from the initial network	Used to inform the sequential network on the first inundations Used in: Berkhahn and Neuweiler, 2024; Schmid and Leandro, 2024.
Water depth at t(0) [2]	Initial water depth, since it firstly informs the model on existing water (e.g. depths in the channels or water bodies). Secondly, in combination with the depth of 10 min., temporal dependencies of the upcoming inundation process can already be estimated. Used in: Berkhahn and Neuweiler, 2024.
Rainfall	Is the driving inundation force and input to the physical loss function.

[1] Used in the initial network;.
[2] Used in the sequence network.

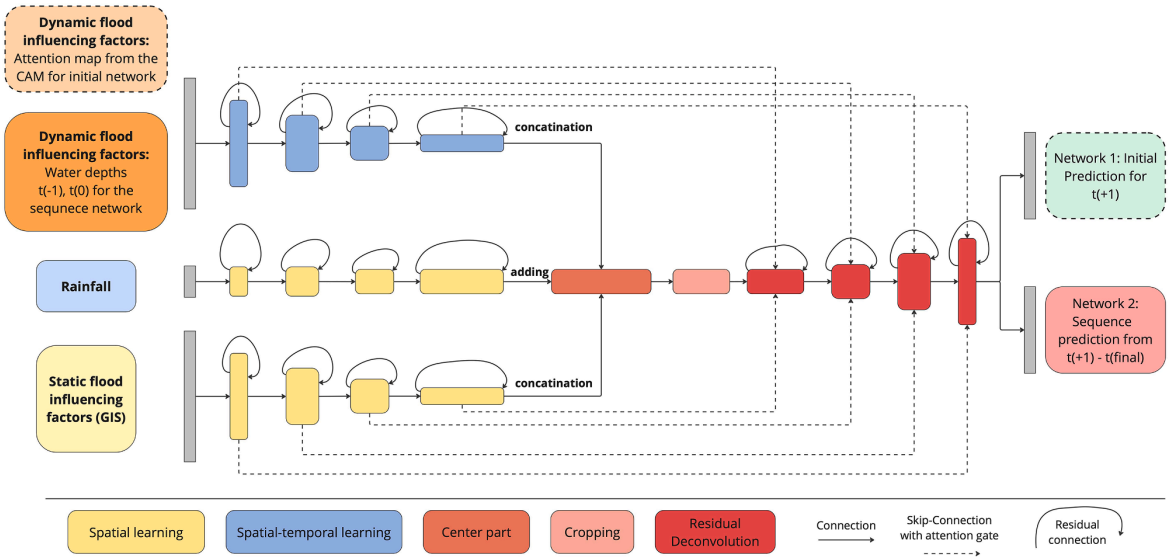


Fig. 3. Architecture of the networks. Both apply the same operation and only differ in the input and output data. The three input datasets are downsampled. After concatenation and addition of the rainfall, the current tensor is cropped, and the images are reconstructed by decoding. The first timestep is given by the initial network, while the remaining predictions are given by the sequence-to-sequence model based on the selected temporal time-stepping. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

we ensure careful information extraction and storage from these sources, which are then combined in a later phase of the network. The rainfall and static flood-influencing factors are computed based on spatial learning with residual convolutional blocks. We stacked four residual convolutional blocks, each containing three sequences of a convolutional operation, batch normalization, and a Rectified Linear Unit (ReLU) activation, chosen for its computational efficiency and ability to mitigate vanishing gradients (Agarap, 2019). After the last activation, the input to this block is added by a residual connection. Residual connections within each block improve gradient flow and training efficiency by learning residual mappings (He et al., 2015). MaxPooling follows each block to reduce spatial dimensions and extract multi-scale features. The dynamic influencing factors (water depths from t(−1) and t(0)) are computed based on spatial-temporal learning with residual convolutional long-short-term memory (LSTM) blocks. These blocks receive the inundation at t(0) and t(−1) as a sequence in separate channels. We selected this lag order (t(–1) and t(0)) for all dynamic input features to capture short-term dynamics, as with longer lag sequences, our available computational resources failed for training. The Convolutional LSTM is explicitly used to capture the time dependency within this sequence, which cannot be learned properly with a standard convolution operation. In comparison to spatial learning, the classical convolutional operations are replaced by Convolutional LSTM operations, which replace the dense operations in standard LSTMs with convolution ones, preserving spatial structure while modelling temporal dependencies (Shi et al., 2015). We note that alternatives such as Gated Recurrent Units (GRUs) or Bidirectional Long Short-Term Memory (BiLSTMs) were not considered, since these approaches rely on vectorized representations and therefore do not maintain the 2D spatial context of inundation maps. After the downsampling, the static and dynamic influencing factors are concatenated, and the rainfall is added. Again, the rainfall is explicitly added since it holds the information on the volume of water. After that, the current features are cropped, which can be understood as zooming into the centre of the image, representing the prediction area based on the spatial prediction framework (Fig. 1).

From this stage, the residual decoder is responsible for recreating the respective inundation maps. The decoder reconstructs the output using transpose convolutions, which reverse the encoder's downsampling by progressively increasing spatial resolution. At each stage, the attention gate skip connections are used to restore and transfer information from the encoder to the decoder (Ronneberger et al., 2015). Attention gates are neural network components that help the model focus on the most relevant spatial features during training and prediction (Oktay et al., 2018). They are designed to filter out irrelevant background information and highlight important features, essentially teaching the network where to look. While classical skip connections carry rich spatial information from the encoder to the decoder (Ronneberger et al., 2015), they may also bring unnecessary noise or background features. In inundation forecasting, for example, not all spatial features contribute equally to a correct prediction. For example, a steep channel: although it has a high potential for directing rapid water flow, it may be completely irrelevant if no rainfall occurs upstream or if it is currently dry. In such a case, the model might pick up on the steep slope as a strong feature (due to its shape or elevation), but this may become noise if no actual flooding occurs there. Without a mechanism to filter such noise from the inputs, the network could end up focusing on processing features that do not influence the current flood dynamics. The hierarchical decoding structure, in combination with the attention gates, refines predictions and maintains spatial detail until the final prediction area and stage are reached.

As described in Section 2.2 for t(+1) the initial network is used, while after that, the second network predicts the remaining timesteps t (+2,3, final) at once in a sequence based on t(0) and t(+1) up to two hours. The final output is a dynamic inundation map for the next two hours with a temporal time-stepping.

2.4. Experiments and optimization

The overall goal of this study is to develop a domain-independent PI-FS that can predict time-dependent water depth in unknown areas. To this end, we conduct the following experiments:

(1) Architecture evaluation: We test the two networks in the PI-FS under different model configurations, where the main tunable hyperparameter is the number of filters in the encoder and decoder. We compare three configurations [32,64,128,256],

[64,128,256,512], and [128,256,512,1024], each with (i) and without attention gates (ii).

(2) Spatio-temporal framework assessment: We analyse the effect of spatial and temporal input design by (i) testing different time-stepping strategies, (ii) training with and without outside information, and (iii) evaluating the impact of two patch sizes, 512 and 1024.

(3) Channel proximity evaluation: To address the potential influence of fluvial processes, we conducted an additional experiment where we excluded both the water bodies and the distance to a channel from the model inputs. This allowed us to explicitly evaluate the effect of channel proximity on flood predictions and clarify the role of local channel structures, which in the study area represented drainage channels rather than rivers.

Based on these findings (presented in chapter 4.1), we apply the PI-FS to unknown areas in the test dataset to prove domain independence. We aim to reach a proper prediction quality and prove the benefit of the physics-informed loss function. Therefore, we carried out the following experiments:

(1) Domain independence: We test on unknown areas and further conduct a cross-validation with new randomly selected areas (presented in Chapters 4.2 and 4–3).

(2) Physically based learning: We trained the model and input configurations found in (1), (2), and (3) twice. (i) by using the informed loss function (Eq. (7)), (ii) using the classical MSE (Eq. (1)) approach (presented in Chapter 4.4).

All training operations apply the Adam optimizer (Kingma and Ba, 2017) with mini-batch updates. Adam is chosen for its ability to adapt learning rates and incorporate momentum from past gradients, which helps the model converge faster and avoid local minima.

2.5. Evaluation criteria

The results of our network are inundation maps with different time-stepping. We evaluate how well the inundation extent is predicted. Therefore, we used the Critical Success Index (CSI). The CSI indicates that an alarm is triggered when a certain threshold is reached. The threshold is set at 0.05 m, which could be interpreted as meaning that the respective cell is classified as wet. As the second criterion, the Relative Volume Error (RVE) is used. The RVE measures the percentage difference between observed and predicted inundated volume in the study area. Thus, negative values indicate underestimation, whereas positive values indicate overestimation. The third criterion is the root mean squared error (RMSE). This metric compares each predicted cell individually with the observed inundation given by the hydraulic model. The RMSE is used because it allows individual cell-wise evaluation. Further, a mean RMSE is calculated to gain one criterion, which can be linked to a prediction time. Thus, by comparing the RMSE values over time, a lead-time prediction error can be calculated. All metrics used are shown in Table 2.

Table 2
Evaluation metrics used for comparing the data-driven predictions with the hydraulic model.

Evaluation metric	Equation	Range	Optimal score
CSI [%]	$\left(\frac{H}{H+M+FA}\right)*100$	[0,100]	100 %
RVE [%]	$\left(\frac{\sum\left(\widehat{y_{i,j}}-y_{i,j}\right)*A}{\sum y_{i,j}*A}\right)*100$	[−100,100]	0 %
RMSE [m]	$\sqrt{\frac{1}{n}*\sum_{i,j}^{n}\left(\widehat{y_{i,j}}-y_{i,j}\right)^2}$	[0,∞)	0m
Mean RMSE [m]	$\frac{1}{N}*\sum RMSE_{i,j}$	[0,∞)	0m

Where H refers to the number of correct positive predictions (values correctly assigned to the classes below and above the threshold), M refers to the number of positive events missed by the model (water depth predictions higher than 0. 05 m), FA refers to the number of false positive predictions (water depth predictions that should be lower than 0.05 m but are higher), $\widehat{y_{i,j}}$ are the water depth prediction in the respective cells, $y_{i,j}$ the observed water depth given by the hydraulic model, A is the cell size, n is the number of events predicted, and N is the number of cells predicted.

2.7. Physically based modelling and rainfall data

To generate the target inundation values, we applied the hydraulic model P-DWave and simulated design rainfalls for the return periods of 100, 50, and 30 years with durations of 10–120 min. The rainfall is given by the German Weather Service (DWD) and the product KOSTRA ("Koordinierte Starkniederschlags-Regionalisierungs-Auswertung"), which contains regionally derived intensity duration frequency (IDF) relationships for heavy rainfall across Germany. For each event, the temporal distribution of rainfall was represented at a 5-min resolution, while the spatial distribution was assumed to be uniform and homogeneous across the study domain. This setup ensured that model training, validation, and testing were based on consistent rainfall scenarios while maintaining reproducibility of the experiments.

2.7.1. 2D inundation modelling using P-DWave

P-DWave is a two-dimensional, parallel diffusive wave model with variable time steps developed by (Leandro et al., 2014). The model is governed by the diffusive form of the shallow water equations (SWE), comprising the continuity Eq. (8) and a simplified momentum (9) equation, expressed as:

$$\frac{\partial h}{\partial t} = \nabla(uh) = R \tag{8}$$

$$g\,\nabla(h+z) = gS_f \tag{9}$$

Where, h is the water depth, $u = \left[u_x u_y\right]^T$ is the depth-averaged flow velocity vector, g is the acceleration due to gravity, z is the bed elevation, R is the source term, and $S_f = \left[S_{fx} S_{fy}\right]$ is the bed friction. For more details on the P-Dwave, please refer to Leandro et al. (2014).

2.7.2. Cellular-Automata approach

To generate the attention map for the initial prediction network, we implemented a cellular automata (CA) algorithm from Guidolin et al. (2016). In this approach, the inundation in each cell is based on simple transition rules that consider water depth, elevation, and neighbouring cell states. The CA model simulates the physical movement of water by updating the water level in each cell according to mass conservation and local topography, enabling a fast and interpretable approximation of flood dynamics. For more details on the CA algorithm, please refer to Guidolin et al. (2016).

3. Study area and implementation

The methodology developed in this study is applied to the catchment area of the city of Baiersdorf in Germany, which has a size of 49.45 km^2. The main urban areas within the study site are, from west-north to south-east, Baiersdorf, Poxdorf, Effeltrich, Hagenau, Igelsdorf, Langensendelbach, Bräuningshof, and Adilz (Fig. 4). Fig. 4 also visualizes the natural channels that flow to the downstream area, the city centre, and discharge into the river Regnitz. The highest elevation point in the catchment is 548 m, and the lowest one is 256 m.

In this study, we applied the model setup validated by Quintana-Romero and Leandro (2022) and simulated the respective return periods.

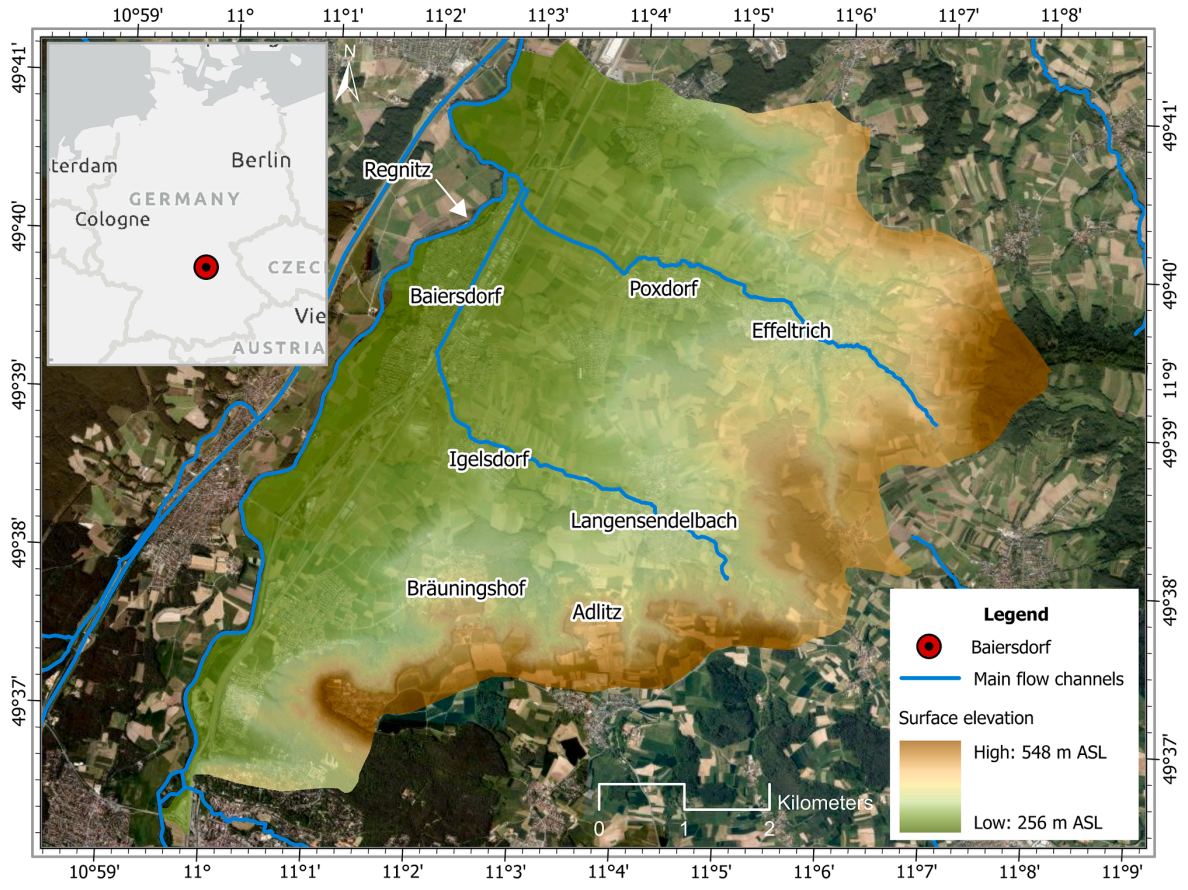


Fig. 4. Study area Baiersdorf (national boundary originates from the ESRI Basemap).

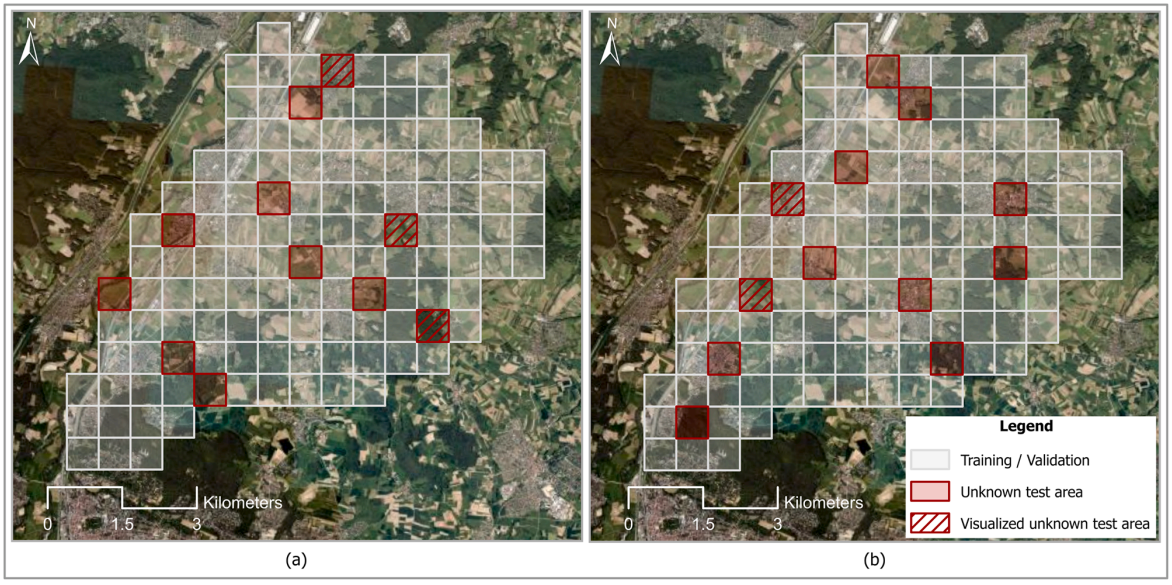


Fig. 5. Dataset for training, validation, and testing. (a) First fold of the cross-validation. (b) The second fold of the cross-validation. The red areas are removed from the training and validation sets to test our forecast system in unknown areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

The CA model is applied for the first 5 min with a fixed timestep of 30 s and outputs an approximation of the upcoming inundation, which is used as an attention map of the initial network prediction. To demonstrate our methodology and generate diverse spatial samples, we divided the total catchment into 124 frames (Fig. 5). We applied a cross-validation strategy in which, at each fold (Fig. 5a and b), a subset of areas was kept entirely unseen during training and validation to rigorously assess the model's predictive performance on unknown regions.

The data-driven model was implemented in Python 3.9 using TensorFlow 2.10 and trained on an NVIDIA RTX A6000 GPU (48 GB VRAM). Training required approximately 73 h. Model loading into GPU memory takes about 32 s. Inference is highly efficient: a single tile prediction requires <1 s, while batch prediction of 64 tiles (operational setting) requires about 3 s per timestep. For the entire study area, a complete prediction step across all tiles takes about 54 s. These runtimes confirm that the framework can deliver real-time forecasts at operational scales, overcoming the limitations of traditional physics-based models.

4. Results

4.1. Networks setup

The first experiment is the selection of the number of filters in the network, evaluated based on the RMSE with and without the attention gates. Table 3 shows the results of the total validation dataset. It can be seen that the filter configuration set 2: [64, 128, 256, 512] filters in the respective encoder and decoder blocks, in combination with attention gates, performed the best. Further, Table 3 shows the influence of the spatial and temporal prediction framework. It can be seen that providing outside information (a larger input frame than output prediction frame) (iii) in combination with a time-stepping of 10 min benefits the quality and performs the best. Additionally, the experiments show that a patch size of 512 outperforms 1024, and a similar effect is observed in the channel experiment, where including channel-related information improves prediction accuracy.

4.2. Overall prediction quality and lead time error

The overall performance of the total test data of the PI-FS with attention gates and the spatiotemporal framework (based on the best configuration found in Section 4.1) is shown in Table 4. It can be seen that for Cross-Validation Fold 1, the mean CSI lies at 74 % where the lowest and highest are 67 % and 89 %. By evaluating the RVE, it can be concluded that the FS tends to overestimate the flood volume since the mean RVE concludes with a 10 % overestimation. The RMSE values range from 0.012 to 0.175 m with a mean of 0.045 m, indicating a high prediction quality. In Cross-Validation Fold 2, the model shows very strong spatial accuracy with CSI values reaching up to 97 %, highlighting its ability to delineate flood extents reliably. Although RMSE values and RVE are somewhat higher than in Fold 1, the results still fall within an acceptable range.

Fig. 6 shows a confidence interval of the mean RMSE values for the test dataset over the forecasting period. Further, the mean of this interval is presented. It can be seen that RMSE increases over time, but the mean value stays for the entire 120 min below 0.06 m, which can be interpreted as a mean water depth error of about 6 cm over the entire study site.

Fig. 7 shows the error distributions and error-boxplots. It reveals consistently low RMSE values across all lead times, with medians ranging from about 0.02–0.04 m for short horizons (10–40 min) and only increasing moderately to around 0.05–0.07 m for longer horizons (90–120 min). Even at the maximum forecast horizon, most predictions remain well below 0.1 m.

Table 3
Results of the hyperparameter section process, presented as RMSE [m] values on the validation dataset. Set 1 has the filter numbers [32, 64, 128, 256], set 2 [64, 128, 256, 512], and set 3 [128, 256, 512, 1024]. Grey-marked fields represent the best performances.

(1) Architecture evaluation			
	Set 1	Set 2	Set 3
(i) Without attention gates	0.089	0.054	0.092
(ii) Attention gates	0.075	0.036	0.165
(2) Spatio-temporal framework assessment (with best model architecture)			
(i) Time stepping	10 min	15 min	20 min
(ii) No outside information*	0.091	0.164	0.354
(iii) Patch size of 512 and spatial prediction framework	0.035	0.040	0.043
(iii) Patch size of 1024 and spatial prediction framework	0.042	**0.051**	**0.058**
(3) Channel proximity evaluation (with best architecture and best configuration from (2)			
(i) Complete input data	0.035		
(ii) Without water bodies and distance to a channel	0.040		

* Input images are 256 × 256, same as the output size.

4.3. Inundation maps and spatial prediction error

The following presents inundation maps of five example events (Figs. 8–12). Additionally, the respective flood inundation obtained by the hydraulic P-DWave model is visualized. For all events, it is clearly visible that both maps (predicted and true) are very similar. Further, the spatial prediction error provided as RMSE is visualized as well. It can be seen that the RMSE is only for some cells close to 0.2 m. Furthermore, we include a time series plot of a reference point in each figure to illustrate the model's performance over the entire event. These plots show a very close alignment with the observed inundation, underscoring the accuracy of the predictions. Table 4 provides the results of the other evaluation criteria for these three events.

4.4. Influence of the physical informed loss

To verify our approach and concept of physically informed learning, we further train the network only with the data-driven loss (MSE) and test it on the same unknown areas as the PI-FS for comparison. Table 6 shows this comparison for the total test data and one event. This event is also presented visually in Fig. 9. We further calculate the RMSE only for the wet cell to highlight the performance improvement. It can be seen that PI-FS outperforms the data-driven forecast system, proving our methodology. While the RMSE values and CSI values are much better for the PI-FS, it can also be seen that the MSE-based forecast underestimates the flood volume, which is indicated by the negative RVE. In contrast, the PI-FS only slightly overestimates the flood volume, with its predictions closer to zero, indicating a better overall fit.

5. Discussion

5.1. Network setup

Table 3 shows that all configurations incorporating attention gates outperformed those without, demonstrating the effectiveness of the developed attention skip connections. These mechanisms help filter irrelevant information and enhance feature interpretation. This is consistent with the findings from Cao et al. (2025) and Oktay et al. (2018), supporting their use in future studies. Among the filter configurations, the second performs best, suggesting it reaches a global minimum in prediction error. The first configuration underperforms, indicating insufficient network capacity to fully capture input-output relationships. The third, with the lowest performance, likely suffers from overfitting or vanishing gradients due to excessive network

Table 4
Overall evaluation of the forecast system.

	Evaluation criteria								
	CSI [%]			RVE [%]			RMSE [m]		
	Min. CSI	Mean CSI	Max. CSI	Min. RVE	Mean. RVE	Max. RVE	Min. RMSE	Mean RMSE	Max. RMSE
Test data: Corr-Validation Fold 1:	67	74	89	−6	11	31	0.012	0.045	0.175
Test data: Corr-Validation Fold 2:	68	75	97	−12	19	33	0.003	0.063	0.210

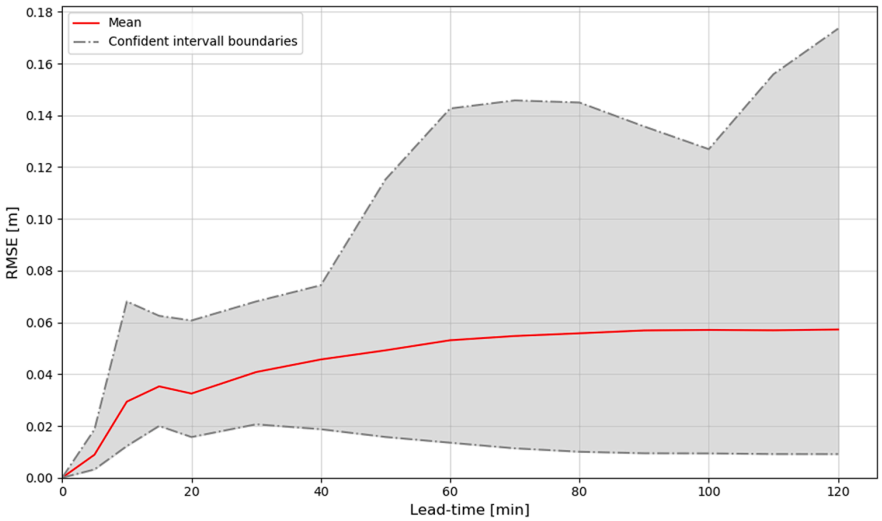


Fig. 6. Lead time error of the forecast system based on the mean RMSE.

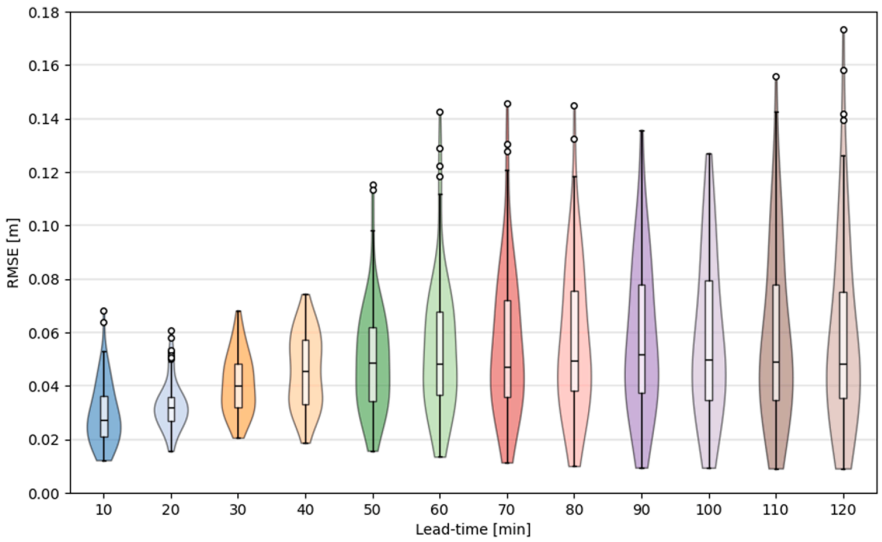


Fig. 7. Error distribution and histogram for the respective forecasting horizon.

parameters. These observations also align with our/and other previous studies (Schmid and Leandro, 2023b, 2024), underscoring the need for careful hyperparameter tuning.

Moreover, the spatial prediction framework, which integrates external contextual information, significantly outperforms configurations that exclude such information. By incorporating data on external water inflows and outflows, the model captures the influence of upstream and surrounding hydrological features on local flow dynamics, thereby enhancing predictive accuracy. This confirms the benefit of our framework.

Among the time-steppings tested, the 10-min interval yields the best results. Higher-frequency observations allow the model to capture rapid changes in water levels and flow patterns more accurately than the coarser 15- and 20-min intervals. At these coarser intervals, the effective

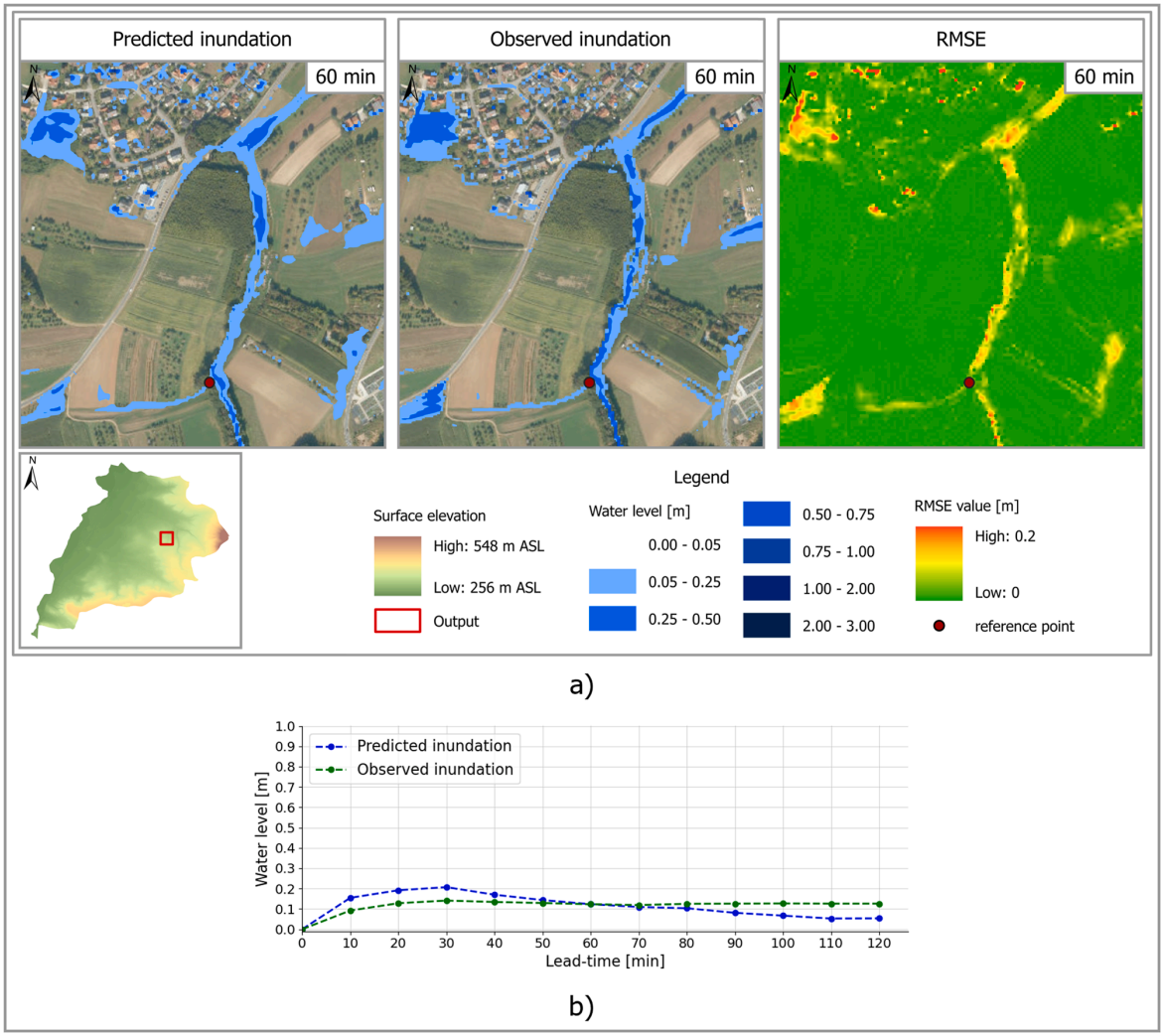


Fig. 8. Example event 1 from the cross-validation fold 1 (a) Water depth predicted in an unknown area at 60 min, compared to the observed inundation modelled by P-DWave and evaluated spatially by the RMSE (b) Timeseries over entire forecasting horizon for the referenc point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

lag order increases, meaning the network requires more temporal steps to correctly propagate information through the sequence, which reduces accuracy at the last predicted timesteps. Additionally, error accumulation over time is more pronounced with coarser resolutions, as deviations at each timestep compound along the sequence. Similar findings were also reported by Cao et al. (2025). As for the patch size differences between 512 and 1024, the 512 setup performed better. This improved performance can be explained by the smaller patches capturing local flood dynamics more effectively, whereas larger patches (1024) tend to smooth out fine-scale spatial variations.

The results of the channel proximity experiment demonstrate a benefit of including channel-related information. With the complete input data, the model achieves a better RMSE, whereas removing water bodies and distance-to-channel information results in a higher error. This indicates that, although the study site does not contain rivers, the channels still add valuable information to improve the flood prediction accuracy.

5.2. *Overall prediction quality and lead time error*

The overall performance of the PI-FS model on unseen test areas demonstrates strong domain-independent predictive capabilities. As shown in Table 4, the model achieves a mean CSI of 74 % (67–89 %), indicating consistent spatial accuracy across diverse events. The mean RMSE remains low at 0.045 m (0.012–0.175 m), while the RVE shows only a slight overestimation of flood volumes (average 11 %, range –6 % to 31 %).

The overall performance of the PI-FS model on unseen test areas demonstrates strong domain-independent predictive capabilities. As shown in Table 4, the model achieves over both Cross-validation folds an average CSI of about 74 % (67–97 %), indicating consistent spatial accuracy across diverse events. The mean RMSE remains low at about 0.045 for fold 1 and 0.063 m (0.003–0.210 m), while the RVE shows only a slight overestimation of flood volumes (mean for fold 1: 11 % and 19 % for fold 2, total range –12 % to 33 %). Compared to prior studies, PI-FS achieves clear quantitative improvements: Löwe et al. (2021) reported CSI values around 50 % and RMSE values near 0.1 m, Seleem et al. (2023) achieved CSI values of about 60 % with mean RMSEs

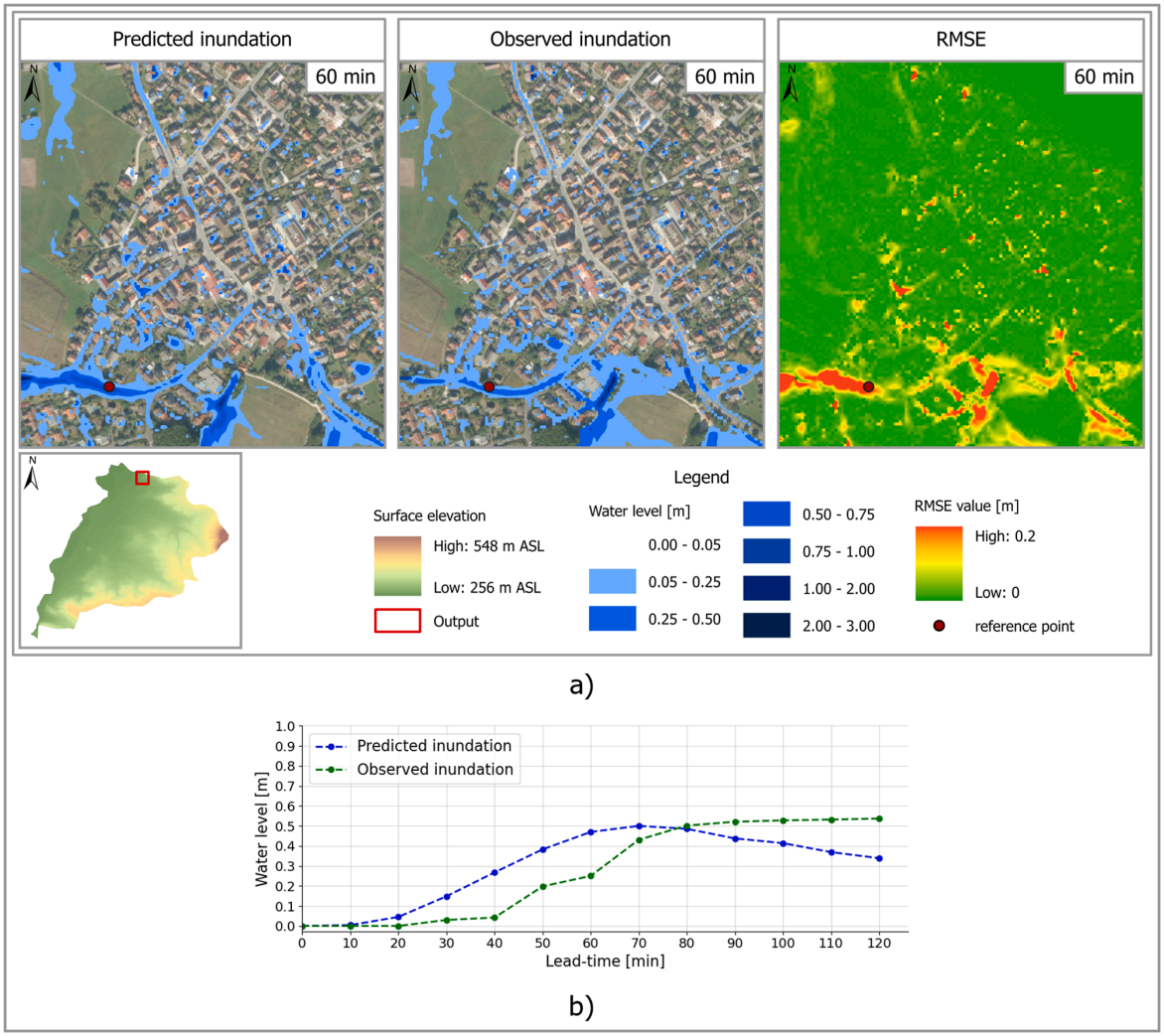


Fig. 9. Example event 2 from cross-validation fold 1 (a) Water depth predicted in an unknown area at 60 min, compared to the observed inundation modelled by P-DWave and evaluated spatially by the RMSE (b) Timeseries over the entire forecasting horizon for the reference point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

between 0.08 m and 0.2 m, and Guo et al. (2022) reported RMSEs of about 0.2 m for wet cells without CSI evaluation. Thus, our model reduces RMSE by about half relative to Guo et al. (2022) and improves CSI by about 20 percentage points compared to Seleem et al. (2023) and Löwe et al. (2021). In comparison, Donnelly et al. (2024) reported a mean RMSE of 0.086 m using a PINN approach, however, their framework did not incorporate transferability to unseen areas, which is a central strength of PI-FS. Nevertheless, it is important to note that direct comparisons of model performance across data-driven studies are inherently constrained by differences in the underlying datasets and event characteristics.

These gains stem from two key innovations: (i) the integration of a physically informed loss function with a spatial prediction framework, which enhances both spatial accuracy and hydrological consistency, and (ii) the preservation of full spatial resolution, avoiding the downsampling used in other works and thereby improving local detail in flood extents. Furthermore, unlike previous approaches, PI-FS explicitly incorporates time-stepping, enabling more realistic temporal dynamics in flood propagation. Taken together, these advances demonstrate that PI-FS not only improves predictive accuracy but also enhances physical interpretability, offering clear advantages for both scientific applications and operational flood management.

Fig. 6 illustrates the benefit of our dual-network design. The forecasting error flattens toward the 120 min horizon, reflecting the strength of the sequence network, which implicitly learns temporal error propagation by predicting the full sequence of flood maps in a single forward pass. As expected in forecasting scenarios, the confidence band widens, which was also found by Schmid and Leandro (2023b). However, the upper bound remains within an acceptable range, with a maximum value of 0.17 m.

The error distributions across lead times demonstrate that the model achieves consistently low RMSE values, with the majority of predictions well below 0.1 m. Fig. 7 highlights that for short lead times (10–40 min), the error distributions are narrow and centered around very low values, indicating highly reliable near-term forecasts. As the lead time increases, the median errors rise only moderately, while the interquartile ranges remain stable, reflecting that the model maintains robustness and predictive skill even at longer horizons. The overall symmetric shapes of the distributions and the limited number of outliers further emphasize the stability of the predictions. These results underline the model's ability to

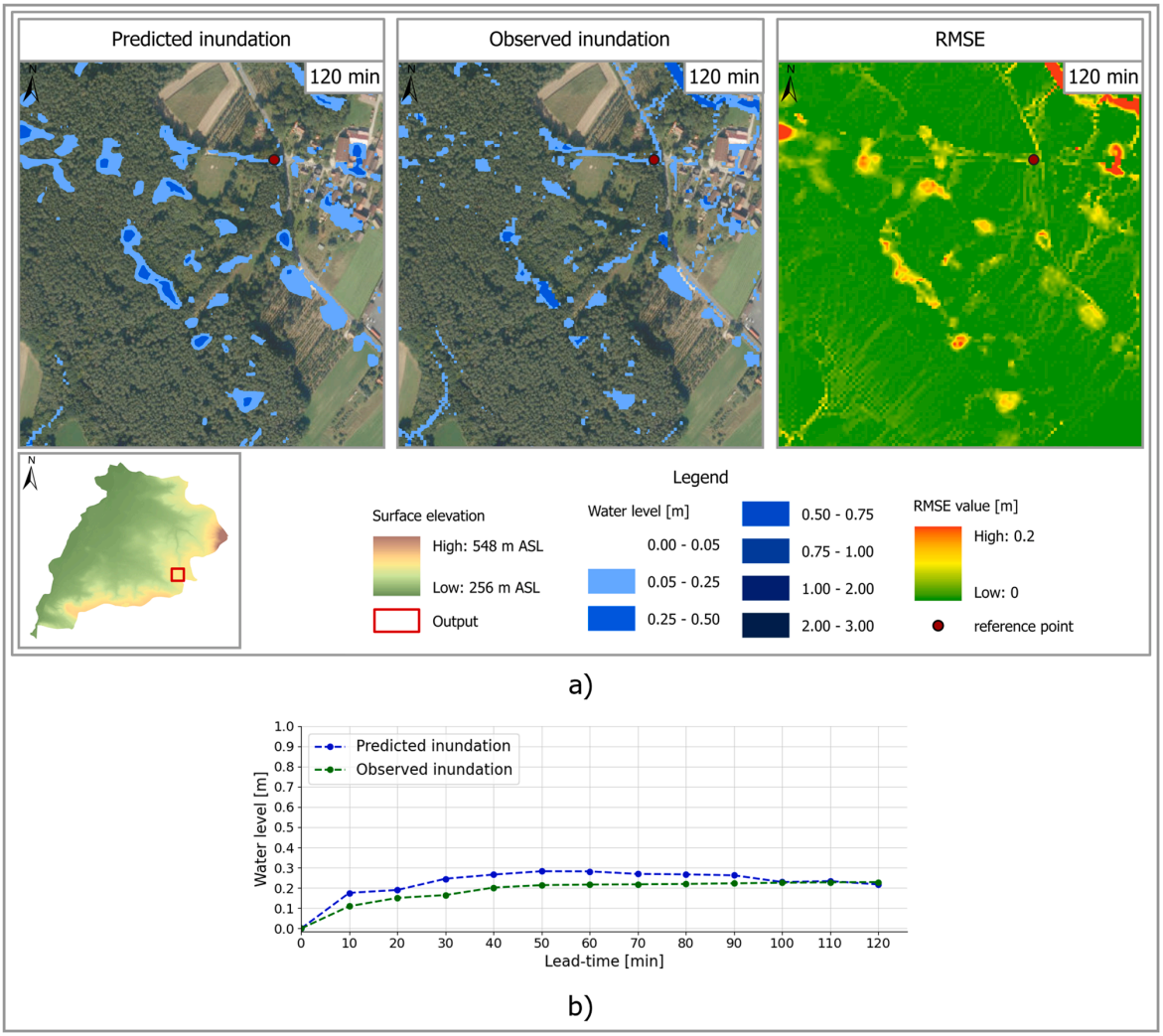


Fig. 10. Example event 3 from cross-validation fold 1 (a) Water depth predicted in an unknown area at 120 min, compared to the observed inundation modelled by P-DWave and evaluated spatially by the RMSE (b) Timeseries over the entire forecasting horizon for the reference point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

provide accurate and dependable forecasts across a range of lead times, making it well-suited for practical flood prediction applications.

5.3. Inundation maps and spatial prediction error

Table 5 and Figs. 8–12 illustrate the performance across five representative flood events, each occurring in areas characterized by distinct land use types. Example 1 (mixed rural–urban, CV Fold 1) achieves the best overall performance with the lowest mean RMSE (0.032 m), the highest CSI (83.7 %), and a moderate RVE of 10.4 %, indicating both high spatial accuracy and reasonable volume estimation. Example 2 (urban, CV Fold 1) shows a similarly low RMSE (0.033 m) and a slight underestimation of flood volume (RVE:6.5 %), but its CSI is the lowest at 67.7 %, reflecting difficulties in capturing complex flow dynamics in urbanized areas. In contrast, Example 3 (rural, CV Fold 1) has the highest RMSE (0.055 m) and the largest overestimation of flood volume (RVE: 30.8 %), with a CSI of 71.2 %, highlighting the challenges of predicting flood extents in open, less structured terrains. The additional cross-validation results confirm these patterns: Example 4 (urban, CV Fold 2) achieves a balanced performance with RMSE of 0.039 m, CSI of 75.5 %, and near-perfect volume estimation (RVE:1.1 %), while Example 5 (rural, CV Fold 2) shows a moderate RMSE (0.042 m), lower CSI (68.0 %), and an RVE of 17 %, again demonstrating the influence of land use and topographic complexity on model accuracy.

These differences in RMSE and CSI can be explained by variations in land use: urban surfaces tend to channel water into narrow pathways, amplifying local errors in spatial extent, whereas rural areas allow more diffuse flooding, increasing uncertainty in flood volume and extent. Across all examples, the highest prediction errors tend to appear along main flow paths, which are typically more dynamic and sensitive to small topographic changes, as was also observed by other studies (Guo et al., 2022; Löwe et al., 2021).

These results demonstrate that the proposed PI-FS can support practical flood protection and management by enabling the evaluation of mitigation measures during the planning phase, such as the placement of embankments or retention areas. In real-time applications, the system can provide high-resolution inundation maps to emergency services, allowing the assessment of mobile protection measures and informed decision-making during flood events.

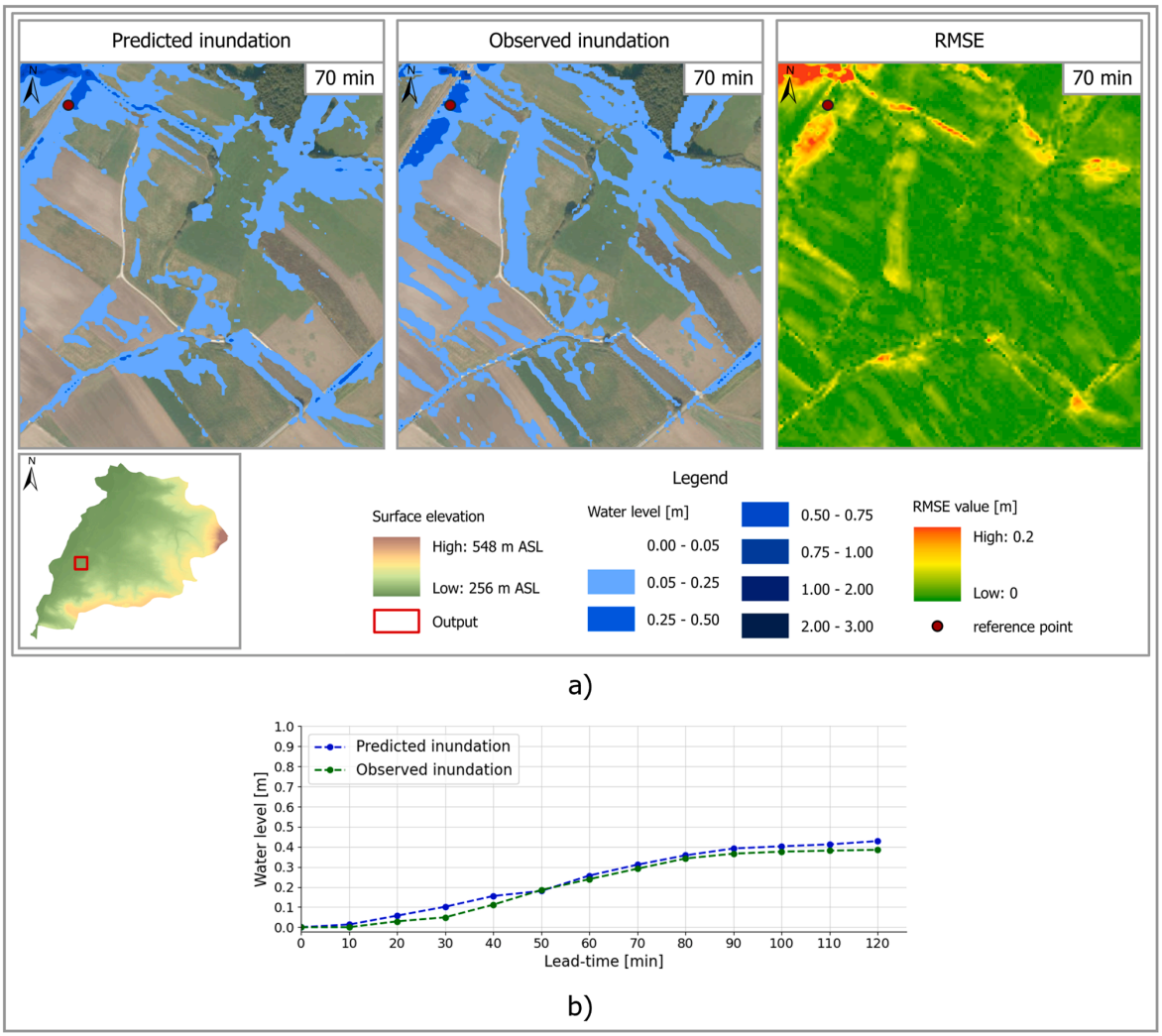


Fig. 11. Example event 4 from cross-validation fold 2 (a) Water depth predicted in an unknown area at 50 min, compared to the observed inundation modelled by P-DWave and evaluated spatially by the RMSE (b) Timeseries over the entire forecasting horizon for the reference point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

5.4. *Influence of the physical informed loss*

The results presented in Table 6 and Fig. 13 demonstrate the clear advantages of incorporating physical knowledge into the loss function. The physically informed loss Function (PI-LF) consistently outperforms the purely data-driven Mean Squared Error (MSE) loss across all evaluation metrics on both the test data and representative example events. A similar, however not directly comparable result (since the aim of the study was different) was found by Donnelly et al. (2024). The PI-LF achieves lower RMSE values and improves the prediction for the total data by about 25 % for the entire area and 52 % for the wet cells. A performance improvement for both the entire area 25 % and wet cells, about 35 % is highlighted by the event chosen. The PI-LF also yields a notably 20 % higher CSI than the MSE-based prediction. Importantly, PI-LF reduces the RVE to a small overestimation of 10 %, while the MSE approach results in a substantial underestimation of 20 % or, as is the case with the chosen event, about 35 %. These findings highlight the value of embedding domain-specific physical constraints into the training process. By guiding the networks toward physically plausible predictions, the PI-LF not only improves RMSE accuracy but also enhances the reliability of spatial and volumetric flood characteristics, which are key aspects of inundation forecasting and follow-up decision-making.

5.5. *Limitations and further research*

We acknowledge that complexity in our modelling approach is added by introducing multiple encoders, attention gates, and input data. Hence, we conducted experiments evaluating the influence of some components, as shown in Table 3. These experiments provided valuable insight into the relative contributions of different design choices to prediction performance and to the effectiveness of our approach. Nevertheless, a more detailed investigation using formal input sensitivity analyses could be useful for a deeper understanding of the relative importance of input features.

The overall loss function (Eq. (7)) combines the pixel-wise MSE, which ensures local accuracy in reproducing observed flood depths, with the physics-informed component, which constrains the predictions to be physically plausible. These two terms complement each other during training: while the MSE penalizes deviations at the pixel level,

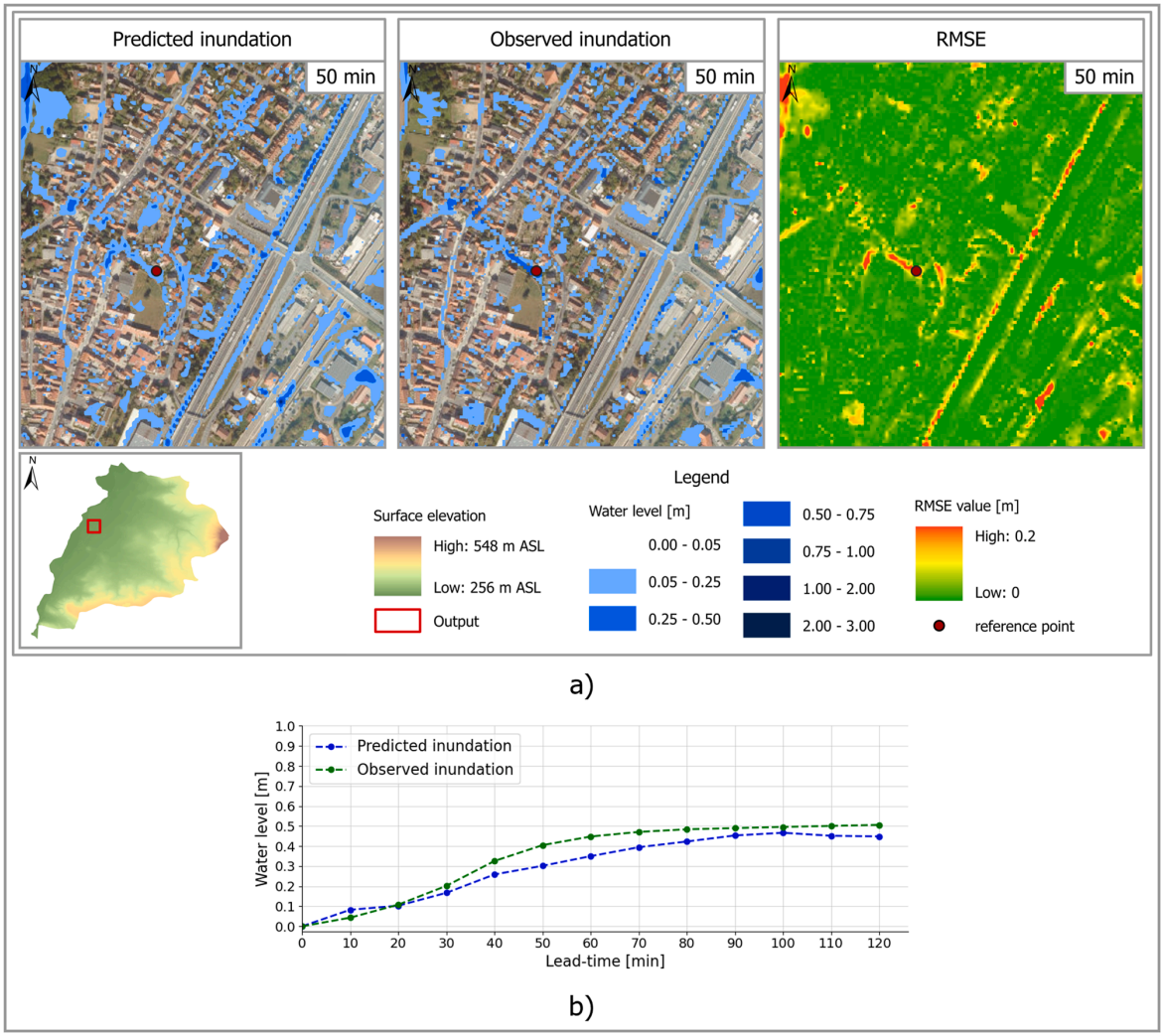


Fig. 12. Example event 5 from cross-validation fold 2 (a) Water depth predicted in an unknown area at 70 min, compared to the observed inundation modelled by P-DWave and evaluated spatially by the RMSE (b) Timeseries over the entire forecasting horizon for the reference point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 5
Performance of the example events shown in this section.

	Mean RMSE [m]	CSI [%]	RVE [%]
Example 1 / CV Fold 1 (Fig 8)	0.032	83.7	10.4
Example 2 / CV Fold 1 (Fig 9)	0.055	71.2	30.8
Example 3 / CV Fold 1 (Fig 10)	0.033	67.7	−6.5
Example 4 / CV Fold 2 (Fig 11)	0.039	75.5	−1.1
Example 5 / CV Fold 2 (Fig 12)	0.042	68.0	17

the physics-based term provides for consistency with hydrological principles. However, in this study, both components were combined in a fixed manner. A dynamic weighting factor between the data-driven and physics-based parts could be applied and systematically tested in future research to better balance accuracy and physical consistency.

The proposed physics-informed loss function is derived from the continuity equation under the kinematic wave assumption. As such, it enforces conservation of mass and incorporates slope and roughness effects, but neglects local and convective acceleration and pressure terms in the full shallow water equations. While this simplification reduces physical completeness, it makes the loss computationally efficient

Table 6
Comparison of the physically informed loss function and the data-driven loss (MSE). Grey-marked fields represent the best performances.

	Mean RMSE [m]		RMSE wet cells [m]		CSI [%]		RVE [%]	
	PI-FS	MSE	PI-FS	MSE	PI-FS	MSE	PI-FS	MSE
Test data	0.045	0.056	0.082	0.125	74.0	58.3	11.0	−19.8
Example event 1	0.032	0.042	0.078	0.118	83.7	68.7	10.4	−32.4

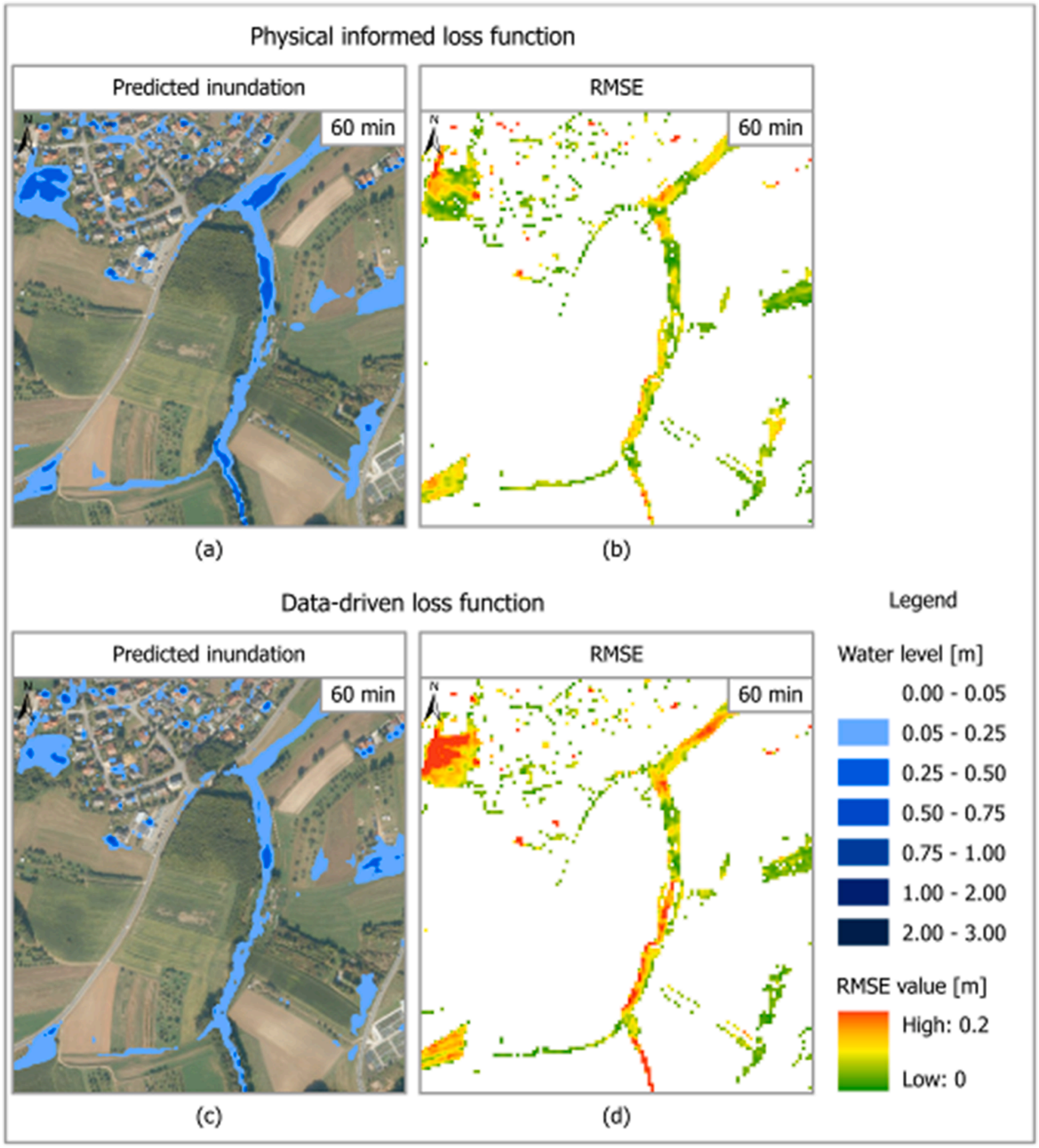


Fig. 13. Visual comparison of example event 1. Prediction based on the physical informed loss (a) and the belonging RMSE for wet cells (b). Prediction based on the data-driven MSE (c) and the belonging RMSE for wet cells (d). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

and directly integrable into a CNN framework, while allowing the $LF_{physical}$ to retain mass conservation and a correct kinematic flood wave propagation.

We acknowledge that pluvial flooding processes are more adequately represented by diffusive or dynamic wave formulations. However, integrating such higher-order formulations directly into a deep learning framework remains challenging, as they require explicit velocity fields and higher-order derivatives. In this study, the kinematic wave is therefore used as a first-order approximation, which provides an acceptable way to embed fundamental hydrodynamic principles, like mass conservation and slope-roughness-driven flow, into the learning process. It should be noted that the kinematic wave is still used for flood wave propagation studies (Tayfur and Moramarco, 2022). We also wish to acknowledge that our research focuses more on the rising limb phase of the hydrograph and not on the recession phase. By doing so, our approach effectively captures peak-dominated flow regimes, providing a reliable representation of the initial flood response. Because capturing both phases is not straightforward, and since we set the training and evaluation window to two hours, we will also focus our future work on this research question.

5. Conclusion

Our study presented a physically informed, data-driven prediction system for floods that enabled both domain-independent and spatial-temporal predictions. Based on the results, the following conclusions could be drawn: (1) The attention gates in the convolutional architecture benefited prediction quality because irrelevant background information was effectively filtered during the training process. (2) The spatial-temporal prediction framework integrated upstream and surrounding contextual information. This enabled the network to effectively learn consistent flow paths and flood inundation patterns. (3) Our method, unlike previous studies, avoided downsampling mechanisms required for large catchments. This ensured applicability across different sizes of study areas without requiring changes in the architecture. (4) Time-stepping was explicitly incorporated, supporting operational real-time forecasting. (5) The physics-informed loss function, derived from the governing equations (2D continuity equation and Manning-Strickler velocity formulation), guided the network towards physically consistent predictions.

Our results supported the model's strong predictive capabilities across multiple spatial and temporal configurations, enabling the forecast system to be domain-independent. Notably, the model maintained high spatial accuracy, as evidenced by a mean CSI of 74 % and low RMSE values. The comparative analysis against existing state-of-the-art methods further demonstrated the benefit of the proposed approach. Crucially, the introduction of the physically informed loss function significantly enhanced prediction reliability, improving RMSE, CSI, and RVE metrics across the entire domain. These results affirmed the value of embedding domain and physical knowledge into the training process, guiding the model toward physically consistent forecasts. Overall, the proposed framework offered a very promising direction for real-time inundation forecasting.

Funding

The research presented in this paper has been partially carried out as part of the PuwaSTAR project (Pumpwerkswarnung für Starkregen und Hochwasser im urbanen Raum) funded by the Bundesministerium für Forschung und Bildung (BMBF) [Federal Ministry of Research and Education] (13N16802) and supervised by the Verein Deutscher Ingenieure (VDI) [The Association of German Engineers].

CRediT authorship contribution statement

Felix Schmid: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Leonie Müller:** Visualization, Software. **Jorge Leandro:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The dataset was provided by the Bavarian State Ministry of the Environment and Consumer Protection (StMUV) and supervised by the Bavarian Environment Agency (LfU), as part of the HiOS project (Hinweiskarte Oberflächenabfluss und Sturzflut).

Data availability

The code will be available on: https://github.com/FXSchmid/physically_informed_data_driven_prediction.

References

Agarap, A.F. (2019). *Deep learning using rectified linear units (ReLU)* (arXiv:1803.08375). arXiv. https://doi.org/10.48550/arXiv.1803.08375.

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

Berkhahn, S., Neuweiler, I., 2024. Data driven real-time prediction of urban floods with spatial and temporal distribution. J. Hydrol. X 22, 100167. https://doi.org/10.1016/j.hydroa.2023.100167.

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P.W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W.W.L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., Péan, C., 2023. IPCC, 2023: Climate Change 2023: Synthesis Report. (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647.

Cao, X., Wang, B., Yao, Y., Zhang, L., Xing, Y., Mao, J., Zhang, R., Fu, G., Borthwick, A.G. L., Qin, H., 2025. U-rnn high-resolution spatiotemporal nowcasting of urban flooding. J. Hydrol. 659. https://doi.org/10.2139/ssrn.4935234.

Crotti, G., Leandro, J., Bhola, P.K., 2019. A 2D real-time flood forecast framework based on a hybrid historical and synthetic runoff database. Water (Basel) 12 (1), 1. https://doi.org/10.3390/w12010114.

Donnelly, J., Abolfathi, S., Pearson, J., Chatrabgoun, O., Daneshkhah, A., 2022. Gaussian process emulation of spatio-temporal outputs of a 2D inland flood model. Water Res. 225, 119100. https://doi.org/10.1016/j.watres.2022.119100.

Donnelly, J., Daneshkhah, A., Abolfathi, S., 2024. Physics-informed neural networks as surrogate models of hydrodynamic simulators. Sci. Total Environ. 912, 168814. https://doi.org/10.1016/j.scitotenv.2023.168814.

Feng, D., Tan, Z., He, Q., 2023. Physics-informed neural networks of the saint-venant equations for downscaling a large-scale river model. Water Resour. Res. 59. https://doi.org/10.1029/2022WR033168.

Guidolin, M., Chen, A.S., Ghimire, B., Keedwell, E.C., Djordjević, S., Savić, D.A., 2016. A weighted cellular automata 2D inundation model for rapid flood analysis. Environ. Modell. Softw. 84, 378–394. https://doi.org/10.1016/j.envsoft.2016.07.008.

Guo, Z., Leitão, J.P., Simões, N.E., Moosavi, V., 2021. Data-driven flood emulation: speeding up urban flood predictions by deep convolutional neural networks. J. Flood Risk Manag. 14 (1), 1. https://doi.org/10.1111/jfr3.12684.

Guo, Z., Moosavi, V., Leitão, J.P., 2022. Data-driven rapid flood prediction mapping with catchment generalizability. J. Hydrol. 609, 127726. https://doi.org/10.1016/j.jhydrol.2022.127726.

He, K., Zhang, X., Ren, S., & Sun, J, 2015. *Deep residual learning for image recognition* (Issue arXiv:1512.03385). http://arxiv.org/abs/1512.03385.

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

Kingma, D.P., & Ba, J. (2017). *Adam: a method for stochastic optimization* (Issue arXiv:1412.6980). http://arxiv.org/abs/1412.6980.

Leandro, J., Chen, A.S., Schumann, A., 2014. A 2D parallel diffusive wave model for floodplain inundation with variable time step (P-DWave). J. Hydrol. 517, 250–259. https://doi.org/10.1016/j.jhydrol.2014.05.020.

Lin, Q., Leandro, J., Gerber, S., Disse, M., 2020a. Multistep flood inundation forecasts with resilient backpropagation neural networks: kulmbach case study. Water (Basel) 12 (12), 3568. https://doi.org/10.3390/w12123568.

Lin, Q., Leandro, J., Wu, W., Bhola, P., Disse, M., 2020b. Prediction of maximum flood inundation extents with resilient backpropagation neural network: case study of Kulmbach. Front. Earth Sci. 8, 332. https://doi.org/10.3389/feart.2020.00332.

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. J. Hydrol. 603, 126898. https://doi.org/10.1016/j.jhydrol.2021.126898.

Mahesh, R.B., Leandro, J., Lin, Q., 2022. Physics informed neural network for spatial-temporal flood forecasting. In: Kolathayar, S., Mondal, A., Chian, S.C. (Eds.), Climate Change and Water Security, Climate Change and Water Security, 178. Springer, Singapore, pp. 77–91. https://doi.org/10.1007/978-981-16-5501-2_7.

Morales-Hernández, M., Sharif, M.B., Kalyanapu, A., Ghafoor, S.K., Dullo, T.T., Gangrade, S., Kao, S.-C., Norman, M.R., Evans, K.J., 2021. TRITON: a multi-GPU open source 2D hydrodynamic flood model. Environ. Modell. Softw. 141, 105034. https://doi.org/10.1016/j.envsoft.2021.105034.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). *Attention U-Net: learning where to look for the Pancreas*. https://doi.org/10.48550/arXiv.1804.03999.

Quintana-Romero, T., Leandro, J., 2022. A method to devise multiple model structures for urban flood inundation uncertainty. J. Hydrol. 604, 127246. https://doi.org/10.1016/j.jhydrol.2021.127246.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. http://arxiv.org/abs/1505.04597.

Schmid, F., Leandro, J, 2023a. A Feature-Informed Data-Driven Approach for Predicting Maximum Flood Inundation Extends. Geosciences 13, 384. https://doi.org/10.3390/forecast6030039.

Schmid, F., Leandro, J., 2023b. An ensemble data-driven approach for incorporating uncertainty in the forecasting of stormwater sewer surcharge. Urban Water J. 1–17. https://doi.org/10.1080/1573062X.2023.2240309.

Schmid, F., Leandro, J., 2024. A data-driven multi-step flood inundation forecast system. Forecasting 6 (3), 761–781. https://doi.org/10.3390/forecast6030039.

Seleem, O., Ayzel, G., Bronstert, A., Heistermann, M., 2023. Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany. Nat. Hazards Earth Syst. Sci. 23 (2), 809–822. https://doi.org/10.5194/nhess-23-809-2023.

Seleem, O., Ayzel, G., De Souza, A.C.T., Bronstert, A., Heistermann, M., 2022. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. Geomatics Nat. Hazards Risk 13 (1), 1640–1662. https://doi.org/10.1080/19475705.2022.2097131.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2015). *Convolutional LSTM Network: a machine learning approach for precipitation nowcasting.* https://doi.org/10.48550/arXiv.1506.04214.

Tayfur, G., Moramarco, T., 2022. Kinematic reverse flood routing in natural rivers using stage data. Appl. Water Sci. 12 (8), 182. https://doi.org/10.1007/s13201-022-01707-2.

Yang, F., Ding, W., Zhao, J., Song, L., Yang, D., Li, X., 2024. Rapid urban flood inundation forecasting using a physics-informed deep learning approach. J. Hydrol. 643, 131998. https://doi.org/10.1016/j.jhydrol.2024.131998.

Outcome and further research

7 Outcome and further research

The overarching objective of this dissertation was to develop real-time, data-driven models for different types of urban flooding, including sewer surcharge, fluvial, and pluvial events, providing computationally efficient alternatives to traditional physically based hydrodynamic models. In parallel, the thesis investigated key challenges inherent to data-driven modeling, with particular emphasis on scalability and transferability, and explored strategies to address these limitations.

The first study investigated how urban sewer surcharge events can be predicted in real time by developing an ensemble forecast system based on nonlinear autoregressive neural networks with exogenous inputs (NARX). By forecasting water-depth time series at selected manholes, the system captures the nonlinear and time-dependent dynamics of urban drainage networks with high computational efficiency. The ensemble of NARX networks produced accurate predictions across the entire sewer system within approximately 56 seconds, demonstrating its suitability for real-time operational applications, even in large catchments. To address the question of how uncertainties arising from model structure and parameters can be quantified and incorporated into operational forecasting, a novel validation and ensemble-based uncertainty quantification framework was introduced. The ensemble consisted of NARX networks with different random initialisations and hidden-layer configurations, allowing the derivation of probabilistic forecast bands around the predicted water-depth time series.

The quality of these uncertainty bounds was assessed using two metrics: the *Pfactor* and the *Rfactor*. The *Pfactor* represents the fraction of observed values that lie within the forecast uncertainty band, and thus quantifies how well the uncertainty range captures the true system behavior. The *Rfactor* measures the average width of the uncertainty band, providing an indication of the sharpness and practical usefulness of the interval. In this study, the uncertainty bands reached *Pfactor* values between 74% and 94%, and *Rfactor* values between 1.36 and 10.68, demonstrating that the ensemble approach delivers informative uncertainty estimates suitable for operational decision-making. Collectively, the results show that NARX-based ensemble modeling enables accurate, efficient, and uncertainty-informed real-time prediction of sewer surcharge dynamics, addressing both the nonlinear behavior of urban drainage systems and the need for rigorous uncertainty assessment in operational forecasting.

Nevertheless, the presented methodology lacks scalability, as only selected manholes were considered, and transferring the model to another drainage system is impractical because the network learned only the underlying hydraulics of these specific locations. In recent years, however, rapid advances in data-driven modeling have significantly transformed the prediction of urban sewer surcharge events. Methodological progress since the time of the study's publication has shifted the state of the art toward Graph Neural Networks (GNNs) (Q. Guo & Wang, 2024; Kerimov et al., 2024; J. Wu et al., 2025).

GNNs are a class of deep learning models specifically designed to operate on graph-structured data. Unlike conventional neural networks, which assume grid-based or independent inputs, GNNs explicitly incorporate relational dependencies between entities through message passing. In this process, each node aggregates information from its neighbors along the graph's edges, updates its hidden state, and iteratively propagates these learned representations throughout the network. This enables GNNs to capture spatial dependencies, directional influences, and structural patterns inherent in complex systems (Zhou et al., 2020). These properties make GNNs particularly well suited for modeling urban drainage systems, which naturally exhibit a graph topology: manholes correspond to nodes, while pipes represent edges that encode hydraulic connectivity. In contrast to NARX-based approaches, which require separate predictive models for individually selected

locations, GNNs operate on the entire sewer system as a single integrated graph and can predict surcharge states for all manholes simultaneously. This system-wide modeling capability provides substantial advantages in terms of scalability, generalization, and computational efficiency. Furthermore, recent investigation had begun exploring the transferability of graph-based models to unseen drainage networks, with promising initial results (Kerimov et al., 2024). These findings mark an important step toward developing generalizable, fully data-driven models for urban sewer systems. Nonetheless, the uncertainty prediction framework and ensemble-based approach developed in this thesis remain a novel contributions, offering a robust and operationally valuable means of quantifying forecast confidence that is still absent in current GNN-based methods.

The second study developed a real-time flood forecasting system capable of predicting maximum inundation extents for entire study areas using a single, unified deep learning model. By formulating inundation prediction as an image-to-image translation task and employing a feature-informed architecture, the proposed forecast system overcame the common need to partition the spatial domain into multiple sub-models, which were necessary in classical data-driven approaches (Berkhahn, 2019; Burrichter et al., 2023; Hofmann & Schüttrumpf, 2021; Lin, Leandro, Wu, et al., 2020). The model integrated prior physical knowledge through three informed dense-layer paths that encode the distance of each grid cell to the river network. This design allowed the forecast system to incorporate essential hydraulic behavior without requiring extensive additional GIS input layers, thereby embedding meaningful physical information directly into the learning process.

The results demonstrated that the system predicts full-domain inundation maps in approximately 19 seconds, enabling real-time applicability. Trained on a hybrid database of pre-simulated scenarios, the system achieves high predictive accuracy, with an average root mean squared error of 0.052 m across 35 test events and RMSE values of 0.074 m and 0.141 m for two observed floods. The proportion of cells exceeding an MSE threshold of 0.2 m^2 was limited to 0.11% and 2.14%, substantially improving upon related work (Lin, Leandro, Wu, et al., 2020). Furthermore, the feature-informed architecture ensured that the model remains computationally effective and trainable on standard hardware. By integrating physical knowledge directly into the network structure rather than expanding input dimensionality, the workflow preserves computational feasibility while enhancing prediction skill. In summary, the study showed that real-time inundation forecasting for complete study regions can be achieved without spatial model subdivision, that physical knowledge can be embedded into the learning process without additional GIS layers, and that the resulting framework remained accessible and efficient for operational flood forecasting environments.

A key limitations of this study is that the proposed model focused on predicting maximum inundation extents and therefore lacks temporal resolution. Additionally, the architecture was trained for the specific catchments, which limits its transferability to new regions. The application of this system in a new area would need to run the hole process, pre-hydraulic modeling for data generation and training of the network. Subsequent research has addressed these limitations by incorporating temporally resolved forecasts, bringing temporal resolution in early warning and forecast systems to the state of the art (Berkhahn & Neuweiler, 2024; Cao et al., 2025; Schmid & Leandro, 2024). In addition, more advanced architectures such as graph neural networks (GNNs) (Bentivoglio et al., 2025) and transformers (Song et al., 2025) have been explored for inundation modeling. In particular, GNNs provide a powerful framework for capturing the structural information inherent in flow accumulation and hydrological connectivity, enabling predictions that naturally account for spatial dependencies across the network. This makes graph-based approaches highly effective for accurate and scalable inundation forecasting in complex catch-

ments (Bentivoglio et al., 2025). Nevertheless, the feature-informed, full-domain architecture presented in this thesis remains a novel contribution in terms of scalability, allowing efficient prediction across entire catchments and demonstrating the potential for highly effective real-time flood forecasting.

The third study presented the development of a data-driven, multi-step inundation forecast system for fluvial flood events, capable of generating dynamic, temporally resolved inundation maps. The system operates over a monitoring window, where hydrographs are tracked, and inundation predictions are initiated once the flow exceeds a pre-defined discharge threshold. The model used the relevant timesteps of the hydrograph as input and consists of a convolutional neural network architecture, feature-informed dense layers, and a recursive path. While the recursive path is initially inactive, it is activated as additional timesteps become relevant, allowing the model to forecast multiple steps into the future.

This approach produced a continuous dynamic inundation map across the monitoring window, with a temporal resolution corresponding to the hydrograph update interval and a maximum lead time dependent on the length of the monitoring period. Test results demonstrated high predictive accuracy, with RMSE values of 0.07 m, 0.12 m, and 0.15 m for forecasts over 6 h, 12 h, and 24 h, respectively. Predictions are generated within seconds, enabling real-time operational application, a clear advantage over conventional physically based models. The combination of the recursive path with the feature-informed layers allows the model to integrate prior knowledge about flooding into the learning process, maintaining high accuracy across multiple forecast horizons. Furthermore, the operational design of the forecast system ensures that it is transferable to other real-world catchments, highlighting its potential for broad applicability in flood early warning and management.

However, a key limitation of the proposed system is its domain dependence. The model is trained on specific catchments and hydrographs, meaning that applying it to a new area requires the full process of data generation, pre-simulated hydraulic modeling, and retraining, which is computationally intensive and time-consuming. To overcome this limitation, subsequent research has focused on developing domain-independent approaches, which aim to eliminate the need for site-specific training and data generation while maintaining real-time prediction capabilities (Do Lago et al., 2023; Z. Guo et al., 2022; Löwe et al., 2021; Schmid et al., 2026; Seleem et al., 2022). These efforts are critical for enabling scalable and transferable flood forecasting systems suitable for operational use across multiple regions.

For multi-step inundation prediction, more advanced architectures, such as transformer-based networks (Pianforini et al., 2025) or attention-enhanced recurrent neural networks (Cao et al., 2025), have been explored further improving prediction performance. These models are specifically designed to capture long-range temporal dependencies and contextual relationships, improving predictive accuracy and generalization. Nevertheless, the architectural design of the data-driven model presented in this study remains novel, as it integrates the entire catchment domain, a temporally recursive path, and an operationally focused workflow, ensuring both real-time performance and practical applicability.

The fourth study presented a physically informed, data-driven flood prediction system capable of producing domain-independent and spatio-temporally resolved forecasts for pluvial flooding. Unlike previous approaches (Z. Guo et al., 2022; Seleem et al., 2022), the model avoids downsampling mechanisms, ensuring applicability across catchments of varying sizes without requiring architectural modifications. The explicit incorporation of time-stepping supports operational real-time forecasting, while the physics-informed loss function, derived from the two-dimension continuity equation and Manning-Strickler velocity formulation, ensures physically consistent

predictions. Evaluation across multiple spatial and temporal configurations confirmed strong predictive performance, achieving a mean critical success index of 74% and a mean root mean squared error of 0.045 m. Compared to classical data-driven loss functions, the physics-informed loss outperformed state-of-the-art data-driven models by approximately 25%, demonstrating enhanced reliability and physical consistency. Overall, these results underscore the value of integrating domain knowledge and physical principles into data-driven models, providing a robust, scalable, and operationally viable framework for real-time flood forecasting.

Nevertheless, the current model is based on the kinematic wave approximation and primarily focuses on the early stages of inundation, providing forecasts up to two hours into the future. While this approach captures the initial dynamics effectively, the reliance on the kinematic wave may limit predictive performance for later stages of flooding, where diffusion effects become more pronounced. Therefore, future investigations should consider incorporating the diffusion wave formulation to better capture the full temporal evolution of inundation events. Furthermore, the framework was developed and tested within the context of a single catchment, effectively making the dataset a quasi-regional one. As a result, the transferability of the model to completely unknown catchments, or its applicability on a global scale, has not yet been demonstrated. Despite these limitations, the outcomes of this study represent a significant and novel step toward reliable, real-time, data-driven flood forecasting. The approach demonstrates the potential for developing more generalizable and scalable data-driven models, laying the foundation for operationally viable, high-resolution inundation prediction across diverse catchments.

Collectively, this dissertation presented a comprehensive investigation into data-driven approaches for real-time flood forecasting across multiple spatial and temporal scales and flood types. The four studies demonstrated that data-driven forecast systems can achieve high predictive accuracy while substantially surpassing traditional hydrodynamic models in computational efficiency. Consequently, the methods developed in this work enable real-time predictions and provide a robust foundation for integration into operational flood forecasting and early-warning systems, while also showing the way for the next generation of advanced, data-driven flood forecasting tools.

A key insight emerging from this dissertation is the potential of integrating physical knowledge directly into the loss function of data-driven models to enhance both interpretability and transferability. By embedding principles such as mass conservation, flow routing, or other governing hydrodynamic equations, models are guided toward physically consistent predictions, even in areas or scenarios that differ from the original training data. This approach represents a significant step beyond purely empirical, black-box models, producing forecasts that are not only accurate but also more explainable in the context of underlying physical processes. Embedding physics into the learning process enhances prediction quality and robustness while offering a framework for scalable and potentially transferable real-time flood forecasting.

Future research should build on this direction by embedding additional or alternative physical principles directly into the learning process and by further developing meta-models that are inherently generalizable. Incorporating the diffusive wave approximation, for example, represents a promising next step, as it more accurately reflects momentum diffusion and backwater effects than the simplified kinematic wave formulation used in this study. Such an enhancement could guide the learning process toward even more physically consistent inundation dynamics. Moreover, early-warning systems must consider not only water depth but also flow velocity, as velocity is a critical driver of flood hazard, impact, and structural damage. None of the presented studies, nor most existing data-driven approaches, provide reliable velocity predictions. Developing models that jointly estimate water depth and velocity, ideally in combination with a physically informed

loss function, therefore represents an important avenue for future work and would significantly improve the practical relevance of data-driven flood forecasting.

Another essential direction concerns the transferability of these models. The generalization capabilities of meta-models should be investigated by applying them to completely new catchments at a lager or even global scale, or by exploring transfer strategies that require only minimal hydrodynamic simulation effort, potentially as a single calibration run. Such approaches would help overcome the current limitations where models remain to a certain way constrained by their training datasets and struggle to extrapolate beyond them. Achieving robust transferability would not only strengthen real-time flood forecasting, but also enable the rapid development of generic models that support the efficient design and assessment of structural mitigation measures during early pre-design phases.

Another promising research avenue lies in the coupling of physically informed two-dimensional inundation models with graph-based representations of sewer systems. The physically informed model presented in this thesis already incorporates hydrodynamic principles for overland flow, while GNNs have demonstrated strong potential for modeling sewer surcharge dynamics. Integrating both components into a unified framework, where surface flooding and subsurface sewer dynamics interact bidirectionally, could enable more holistic urban flood forecasting. Such a coupled data-driven approach would allow the representation of key processes such as sewer overflow, surface–subsurface exchange, and dynamic exchange between drainage systems and overland flow.

An alternative research direction could involve rethinking model complexity: as models grow deeper and more data-hungry, they become increasingly computationally demanding. Future studies could therefore explore the development of simpler models that require less data while achieving comparable predictive performance. Such approaches would not only reduce training costs but also make high-quality flood forecasting accessible in data-scarce regions and on resource-constrained hardware. Moreover, more compact architectures may offer greater interpretability, helping practitioners understand model behavior and identify failure modes more easily. Exploring these trade-offs between complexity, performance, and practicality could open the door to a new class of lightweight yet reliable forecasting systems.

Literature

Literature

Alves, P. B. R., Amanguah, E., McNally, D., Espinoza, M., Ghaedi, H., Reilly, A. C., & Hendricks, M. D. (2024): Navigating the definition of urban flooding: A conceptual and systematic review of the literature. *Water Science & Technology*, *90*(10), 2796–2812. https://doi.org/10.2166/wst.2024.351

Asanjan, A. A. (n.d.): An advanced deep learning framework for short-term precipitation forecasting from satellite information, 113.

Balaian, S. K., Sanders, B. F., & Abdolhosseini Qomi, M. J. (2024): How urban form impacts flooding. *Nat Commun*, *15*(1), 6911. https://doi.org/10.1038/s41467-024-50347-4

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2025): Multi-scale hydraulic graph neural networks for flood modelling. *Nat. Hazards Earth Syst. Sci.*, *25*(1), 335–351. https://doi.org/10.5194/nhess-25-335-2025

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

Berkhahn, S. (2019): An ensemble neural network model for real-time prediction of urban floods. *Journal of Hydrology*, 12.

Berkhahn, S., & Neuweiler, I. (2024): Data driven real-time prediction of urban floods with spatial and temporal distribution. *Journal of Hydrology X*, *22*, 100167. https://doi.org/10.1016/j.hydroa.2023.100167

Bulti, D. T., & Abebe, B. G. (2020): A review of flood modeling methods for urban pluvial flood application. *Model. Earth Syst. Environ.*, *6*(3), 1293–1302. https://doi.org/10.1007/s40808-020-00803-z

Burrichter, B., Hofmann, J., Koltermann da Silva, J., Niemann, A., & Quirmbach, M. (2023): A spatiotemporal deep learning approach for urban pluvial flood forecasting with multi-source data. *Water*, *15*. https://doi.org/10.3390/w15091760

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Péan, C. (2023, July 25): *IPCC, 2023: Climate change 2023: Synthesis report.* Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647
Edition: First.

Cao, X., Wang, B., Yao, Y., Zhang, L., Xing, Y., Mao, J., Zhang, R., Fu, G., Borthwick, A. G. L., & Qin, H. (2025): U-rnn high-resolution spatiotemporal nowcasting of urban flooding. *Journal of Hydrology*, *659*. https://doi.org/10.2139/ssrn.4935234

Chang, F.-J., Chen, P.-A., Lu, Y.-R., Huang, E., & Chang, K.-Y. (2014): Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, *517*, 836–846. https://doi.org/10.1016/j.jhydrol.2014.06.013

Chang, L.-C., Shen, H.-Y., & Chang, F.-J. (2014): Regional flood inundation nowcast using hybrid SOM and dynamic neural networks. *Journal of Hydrology*, *519*, 476–489. https://doi.org/10.1016/j.jhydrol.2014.07.036

Chen, C., Hui, Q., Xie, W., Wan, S., Zhou, Y., & Pei, Q. (2021): Convolutional neural networks for forecasting flood process in internet-of-things enabled smart city. *Computer Networks*, *186*, 107744. https://doi.org/10.1016/j.comnet.2020.107744

Chen, D., Gao, X., Xu, C., Chen, S., Fang, J., Wang, Z., & Wang, Z. (2020): FlowGAN: A conditional generative adversarial network for flow prediction in various conditions [Place: Baltimore,

MD, USA]. *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, 315–322. https://doi.org/10.1109/ICTAI50040.2020.00057

Chiang, Y.-M., Chang, L.-C., Tsai, M.-J., Wang, Y.-F., & Chang, F.-J. (2010): Dynamic neural networks for real-time water level predictions of sewerage systems-covering gauged and ungauged sites. *Hydrol. Earth Syst. Sci.*, 11.

Crotti, G., Leandro, J., & Bhola, P. K. (2019): A 2d real-time flood forecast framework based on a hybrid historical and synthetic runoff database. *Water*, *12*(1), 114. https://doi.org/10.3390/w12010114
Number: 1.

Do Lago, C. A., Giacomoni, M. H., Bentivoglio, R., Taormina, R., Gomes, M. N., & Mendiondo, E. M. (2023): Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. *Journal of Hydrology*, *618*, 129276. https://doi.org/10.1016/j.jhydrol.2023.129276

Donnelly, J., Daneshkhah, A., & Abolfathi, S. (2024): Physics-informed neural networks as surrogate models of hydrodynamic simulators. *Science of The Total Environment*, *912*, 168814. https://doi.org/10.1016/j.scitotenv.2023.168814

Duncan, A. P., Chen, A. S., Keedwell, E. C., Djordjević, S., & Savić, D. A. (2013): RAPIDS: Early warning system for urban flooding and water quality hazards. *Conference Paper*, 6.

Estrella, N., & Menzel, A. (2013): Recent and future climate extremes arising from changes to the bivariate distribution of temperature and precipitation in bavaria, germany: Bivariate distribution of temperature and precipitation in bavaria. *Int. J. Climatol.*, *33*(7), 1687–1695. https://doi.org/10.1002/joc.3542
Number: 7.

Feng, D., Tan, Z., & He, Q. (2023): Physics-informed neural networks of the saint-venant equations for downscaling a large-scale river model. *Water Resources Research*, *59*. https://doi.org/https://doi.org/10.1029/2022WR033168

Fernando, A. K., Zhang, X., & Kinley, P. F. (2007): Combined sewer overflow forecasting with feed-forward back-propagation artificial neural network. *International Journal of Civil and Environmental Engineering*, *1*(12), 7
Number: 12.

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., Guerreiro, S., Haerter, J. O., Kendon, E. J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., & Zhang, X. (2021): Anthropogenic intensification of short-duration rainfall extremes. *Nat Rev Earth Environ*, *2*(2), 107–122. https://doi.org/10.1038/s43017-020-00128-6

Fu, G., Butler, D., Khu, S.-T., & Sun, S. (2011): Imprecise probabilistic evaluation of sewer flooding in urban drainage systems using random set theory: IMPRECISE PROBABILISTIC EVALUATION OF SEWER FLOODING. *Water Resour. Res.*, *47*(2). https://doi.org/10.1029/2009WR008944
Number: 2.

García, L., Barreiro-Gomez, J., Escobar, E., Téllez, D., Quijano, N., & Ocampo-Martinez, C. (2015): Modeling and real-time control of urban drainage systems: A review. *Advances in Water Resources*, *85*, 120–132. https://doi.org/10.1016/j.advwatres.2015.08.007

Géron, A. (2018): *Neural networks and deep learning* (1st edition). O'Reilly Media, Inc.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, June 10): Generative adversarial networks. https://doi.org/10.48550/arXiv.1406.2661

Gründemann, G. J., Van De Giesen, N., Brunner, L., & Van Der Ent, R. (2022): Rarest rainfall events will see the greatest relative increase in magnitude under future climate change. *Commun Earth Environ*, *3*(1), 235. https://doi.org/10.1038/s43247-022-00558-8

Guo, Q., & Wang, W. (2024): HydroNet: A spatio-temporal graph neural network for modeling hydraulic dependencies in urban wastewater systems. *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, 717–718. https://doi.org/10.1145/3678717.3695761

Guo, Z., Leitão, J. P., Simões, N. E., & Moosavi, V. (2021): Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *J Flood Risk Management*, *14*(1). https://doi.org/10.1111/jfr3.12684
Number: 1.

Guo, Z., Moosavi, V., & Leitão, J. P. (2022): Data-driven rapid flood prediction mapping with catchment generalizability. *Journal of Hydrology*, *609*, 127726. https://doi.org/10.1016/j.jhydrol.2022.127726

Hammond, M., Chen, A., Djordjević, S., Butler, D., & Mark, O. (2015): Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, *12*(1), 14–29. https://doi.org/10.1080/1573062X.2013.857421
Number: 1.

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

Ho, J., Jain, A., & Abbeel, P. (2020, December 16): Denoising diffusion probabilistic models. https://doi.org/10.48550/arXiv.2006.11239

Hofmann, J., & Schüttrumpf, H. (2021): floodGAN: Using deep adversarial learning to predict pluvial flooding in real time. *Water*, *13*(16), 2255. https://doi.org/10.3390/w13162255
Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

Hou, J., Kang, Y., Hu, C., Tong, Y., Pan, B., & Xia, J. (2020): A GPU-based numerical model coupling hydrodynamical and morphological processes. *International Journal of Sediment Research*, *35*(4), 386–394. https://doi.org/10.1016/j.ijsrc.2020.02.005

Huong, H. T. L., & Pathirana, A. (2013): Urbanization and climate change impacts on future urban flooding in can tho city, vietnam. *Hydrol. Earth Syst. Sci.*, *17*(1), 379–394. https://doi.org/10.5194/hess-17-379-2013
Number: 1.

Jean, M.-È., Morin, C., Duchesne, S., Pelletier, G., & Pleau, M. (2022): Real-time model predictive and rule-based control with green infrastructures to reduce combined sewer overflows. *Water Research*, *221*, 118753. https://doi.org/10.1016/j.watres.2022.118753

Kabir, S., Patidar, S., Xia, X., Liang, Q., Neal, J., & Pender, G. (2020): A deep convolutional neural network model for rapid prediction of fluvial flood inundation. *Journal of Hydrology*, *590*, 125481. https://doi.org/10.1016/j.jhydrol.2020.125481

Kerimov, B., Taormina, R., & Tscheikner-Gratl, F. (2024): Towards transferable metamodels for water distribution systems with edge-based graph neural networks. *Water Research*, *261*, 121933. https://doi.org/10.1016/j.watres.2024.121933

Kimura, N., Yoshinaga, I., Sekijima, K., Azechi, I., & Baba, D. (2019): Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions. *Water*, *12*(1), 96. https://doi.org/10.3390/w12010096
Number: 1.

Kingma, D. P., & Ba, J. (2017, January 29): Adam: A method for stochastic optimization. Retrieved September 6, 2023, from http://arxiv.org/abs/1412.6980
Issue: arXiv:1412.6980.

Kirezci, E., Young, I. R., Ranasinghe, R., Muis, S., Nicholls, R. J., Lincke, D., & Hinkel, J. (2020): Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st century. *Sci Rep*, *10*(1), 11629. https://doi.org/10.1038/s41598-020-67736-6

Knutti, R., Rogelj, J., Sedláček, J., & Fischer, E. M. (2016): A scientific critique of the two-degree climate change target. *Nature Geosci*, *9*(1), 13–18. https://doi.org/10.1038/ngeo2595

Kratzert, F., Gauch, M., Klotz, D., & Nearing, G. (2024): HESS opinions: Never train a long short-term memory (LSTM) network on a single basin. *Hydrol. Earth Syst. Sci.*, *28*(17), 4187–4201. https://doi.org/10.5194/hess-28-4187-2024

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018): Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S. C., Ciavarella, A., Cruz, L. D., de Vries, H., Demuth, N., Ferrone, A., Fischer, M., Fowler, H. J., Goergen, K., Heinrich, D., Henrichs, Y., … van, M. (2021): Rapid attribution of heavy rainfall events leading to the severe flooding in western europe during july 202. *World Weather Attribution*. https://www.worldweatherattribution.org/wp-content/uploads/Scientific-report-Western-Europe-floods-2021-attribution.pdf

Leandro, J., Chen, A., & Schumann, A. (2014): A 2d parallel diffusive wave model for floodplain inundation with variable time step (p-DWave). *Journal of Hydrology*, *517*, 250–259. https://doi.org/10.1016/j.jhydrol.2014.05.020

Leandro, J., Djordjević, S., Chen, A. S., Savić, D. A., & Stanić, M. (2011): Calibration of a 1d/1d urban flood model using 1d/2d model results in the absence of field data. *Water Science and Technology*, *64*(5), 1016–1024. https://doi.org/10.2166/wst.2011.467
Number: 5.

Leandro, J., Gander, A., Beg, M., Bhola, P., Konnerth, I., Willems, W., Carvalho, R., & Disse, M. (2019): Forecasting upper and lower uncertainty bands of river flood discharges with high predictive skill. *Journal of Hydrology*, *576*, 749–763. https://doi.org/10.1016/j.jhydrol.2019.06.052

Leandro, J., Schumann, A., & Pfister, A. (2016): A step towards considering the spatial heterogeneity of urban key features in urban hydrology flood modelling. *Journal of Hydrology*, *535*, 356–365. https://doi.org/10.1016/j.jhydrol.2016.01.060

Leandro, J., Hotta, C. I., Pinto, T. A., & Ahadzie, D. K. (2022): Expected annual probability of infection: A flood-risk approach to waterborne infectious diseases. *Water Research*, *219*, 118561. https://doi.org/10.1016/j.watres.2022.118561

Leandro, J., & Martins, R. (2016): A methodology for linking 2d overland flow models with the sewer network model SWMM 5.1 based on dynamic link libraries. *Water Science and Technology*, *73*(12), 3017–3026. https://doi.org/10.2166/wst.2016.171

Li, X., Zhou, F., & Lodewyk, S. (2010): Applications of artificial neural networks in urban water system. *Watershed Management*, 12.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022): A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827
Number: 12 Conference Name: IEEE Transactions on Neural Networks and Learning Systems.

Lin, Q., Leandro, J., Gerber, S., & Disse, M. (2020): Multistep flood inundation forecasts with resilient backpropagation neural networks: Kulmbach case study. *Water*, *12*(12), 3568. https://doi.org/10.3390/w12123568

Lin, Q., Leandro, J., Wu, W., Bhola, P., & Disse, M. (2020): Prediction of maximum flood inundation extents with resilient backpropagation neural network: Case study of kulmbach. *Front. Earth Sci.*, *8*, 332. https://doi.org/10.3389/feart.2020.00332

Löwe, R., Böhm, J., Getreuer Jensen, D., Leandro, J., & Højmark Rasmussen, S. (2021): U-FLOOD – topographic deep learning for predicting urban pluvial flood water depth. *Journal of Hydrology*, *603*(126898). https://doi.org/10.1016/j.jhydrol.2021.126898
Number: 126898.

Mahesh, R. B., Leandro, J., & Lin, Q. (2022): Physics informed neural network for spatial-temporal flood forecasting. In S. Kolathayar, A. Mondal, & S. C. Chian (Eds.), *Climate change and water security* (pp. 77–91, Vol. 178). Springer Singapore. https://doi.org/10.1007/978-981-16-5501-2_7
Series Title: Lecture Notes in Civil Engineering.

Martins, R., Leandro, J., & Djordjević, S. (2018): Influence of sewer network models on urban flood damage assessment based on coupled 1d/2d models: Influence of sewer models on urban flood damage assessment. *J Flood Risk Management*, *11*, S717–S728. https://doi.org/10.1111/jfr3.12244

Morales-Hernández, M., Sharif, M. B., Kalyanapu, A., Ghafoor, S., Dullo, T., Gangrade, S., Kao, S.-C., Norman, M., & Evans, K. (2021): TRITON: A multi-GPU open source 2d hydrodynamic flood model. *Environmental Modelling & Software*, *141*, 105034. https://doi.org/10.1016/j.envsoft.2021.105034

Mosavi, A., Ozturk, P., & Chau, K.-w. (2018): Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536
Number: 11.

Mounce, S. R., Shepherd, W., Sailor, G., Shucksmith, J., & Saul, A. J. (2014): Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. *Water Science and Technology*, *69*(6), 1326–1333. https://doi.org/10.2166/wst.2014.024
Number: 6.

Munich, R. (2022): *Munich re NatCatSERVICE natural catastrophes in 2021*. Retrieved October 5, 2025, from https://www.munichre.com/%20content/dam/munichre/mrwebsiteslaunches/natcat-2022/2021_Figures-of-the-year.pdf

Nagesh Kumar, D., Srinivasa Raju, K., & Sathish, T. (2004): River flow forecasting using recurrent neural networks. *Water Resources Management*, *18*(2), 143–161. https://doi.org/10.1023/B:WARM.0000024727.94701.12
Number: 2.

Pianforini, M., Dazzi, S., Pilzer, A., & Vacondio, R. (2025): FloodSformer: A transformer-based data-driven model for predicting the 2-d dynamics of fluvial floods. *Environmental Modelling & Software*, *193*, 106599. https://doi.org/10.1016/j.envsoft.2025.106599

Qin, H.-p., Li, Z.-x., & Fu, G. (2013): The effects of low impact development on urban flooding under different rainfall characteristics. *Journal of Environmental Management*, *129*, 577–585. https://doi.org/10.1016/j.jenvman.2013.08.026

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019): Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Razavi, S., Gober, P., Maier, H. R., Brouwer, R., & Wheater, H. (2020): Anthropocene flooding: Challenges for science and society. *Hydrological Processes*, *34*(8), 1996–2000. https://doi.org/10.1002/hyp.13723

Rezaeianzadeh, M., Tabari, H., Arabi Yazdi, A., Isik, S., & Kalin, L. (2014): Flood flow forecasting using ANN, ANFIS and regression models. *Neural Comput & Applic*, *25*(1), 25–37. https://doi.org/10.1007/s00521-013-1443-6

Rjeily, Y. A., Abbas, O., Sadek, M., Shahrour, I., & Hage Chehade, F. (2017): Flood forecasting within urban drainage systems using NARX neural network. *Water Science and Technology*, *76*(9), 2401–2412. https://doi.org/10.2166/wst.2017.409
Number: 9.

Ronneberger, O., Fischer, P., & Brox, T. (2015, May 18): U-net: Convolutional networks for biomedical image segmentation. Retrieved July 24, 2024, from http://arxiv.org/abs/1505.04597

Rossman, L. A. (2007): Storm water management model reference manual volume II – hydraulics, 190.

Rossman, L. A. (2016): Storm water management model reference manual volume i – hydrology.

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. https://doi.org/10.1029/2020EF001851
Number: 2.

Salinas Rodriguez, C., Gersonius, B., Zevenbergen, C., Serrano, D. J., & Ashley, R. (2018): A semi risk-based approach for managing urban drainage systems under extreme rainfall. *Water*, *10*, 384. https://doi.org/10.3390/w10040384

Schmid, F., & Leandro, J. (2023a): An ensemble data-driven approach for incorporating uncertainty in the forecasting of stormwater sewer surcharge. *Urban Water Journal*, 1–17. https://doi.org/10.1080/1573062X.2023.2240309

Schmid, F., & Leandro, J. (2023b): A feature-informed data-driven approach for predicting maximum flood inundation extends. *Geosciences*, (13(12)), 384. https : / / doi . org / 10 . 3390 / geosciences13120384

Schmid, F., & Leandro, J. (2024): A data-driven multi-step flood inundation forecast system. *Forecasting*, *6*(3), 761–781. https://doi.org/10.3390/forecast6030039

Schmid, F., Müller, L., & Leandro, J. (2026): A physically informed domain-independent data-driven inundation forecast model. *Water Research*, *289*, 124819. https://doi.org/10.1016/j.watres.2025.124819

Schmitt, T., Thomas, M., & Ettrich, N. (2004): Analysis and modeling of flooding in urban drainage systems. *Journal of Hydrology*, *299*(3), 300–311. https://doi.org/10.1016/S0022-1694(04)00374-9
Number: 3-4.

Seleem, O., Ayzel, G., Bronstert, A., & Heistermann, M. (2023): Transferability of data-driven models to predict urban pluvial flood water depth in berlin, germany. *Nat. Hazards Earth Syst. Sci.*, *23*(2), 809–822. https://doi.org/10.5194/nhess-23-809-2023

Seleem, O., Ayzel, G., De Souza, A. C. T., Bronstert, A., & Heistermann, M. (2022): Towards urban flood susceptibility mapping using data-driven models in berlin, germany. *Geomatics, Natural Hazards and Risk*, *13*(1), 1640–1662. https://doi.org/10.1080/19475705.2022.2097131

Sherstinsky, A. (2020): Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306. https://doi.org/10.1016/j.physd.2019.132306

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2015): Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *arXiv:1506.04214 [cs]*. Retrieved April 29, 2021, from http://arxiv.org/abs/1506.04214

Solomatine, D., See, L., & Abrahart, R. (2008): Data-driven modelling: Concepts, approaches and experiences. In R. J. Abrahart, L. M. See, & D. P. Solomatine (Eds.), *Practical hydroinformatics* (pp. 17–30, Vol. 68). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79881-1_2 ISSN: 0921-092X Series Title: Water Science and Technology Library.

Song, W., Guan, M., & Yu, D. (2025): SwinFlood: A hybrid CNN-swin transformer model for rapid spatiotemporal flood simulation. *Journal of Hydrology*, *660*, 133280. https://doi.org/10.1016/j.jhydrol.2025.133280

Swiss RE Institue. (2022): Sigma - natural catastrophes in 2021: The floodgates are open. *01/2022*.

Szymczak, S., Backendorf, F., Bott, F., Fricke, K., Junghänel, T., & Walawender, E. (2022): Impacts of heavy and persistent precipitation on railroad infrastructure in july 2021: A case study from the ahr valley, rhineland-palatinate, germany. *Atmosphere*, *13*(7), 1118. https://doi.org/10.3390/atmos13071118

Thapa, S., Zhao, Z., Li, B., Lu, L., Fu, D., Shi, X., Tang, B., & Qi, H. (2020): Snowmelt-driven streamflow prediction using machine learning techniques (LSTM, NARX, GPR, and SVR). *Water*, *12*(6), 1734. https://doi.org/10.3390/w12061734
Number: 6.

Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., & Linh, N. T. T. (2021): Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, *12*(3), 101075. https://doi.org/10.1016/j.gsf.2020.09.006
Number: 3.

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), 90. https://doi.org/10.1007/s10584-023-03502-7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023, August 2): Attention is all you need. https://doi.org/10.48550/arXiv.1706.03762
Comment: 15 pages, 5 figures.

Wan, X. (2019): Influence of feature scaling on convergence of gradient iterative algorithm. *J. Phys.: Conf. Ser.*, *1213*, 032021. https://doi.org/10.1088/1742-6596/1213/3/032021

Wu, J., Ma, L., Chen, X., & Broyd, T. (2025): Research on interpretable graph neural network agent model for siltation diagnosis in urban-scale sewer systems. *Tunnelling and Underground Space Technology*, *162*, 106666. https://doi.org/10.1016/j.tust.2025.106666

Wu, Y., Sun, J., Ying, Z., Xue, L., Chen, D., & Lin, W. (2021): Effects of local-scale orography and urban heat island on the initiation of a record-breaking rainfall event. *JGR Atmospheres*, *126*(16), e2021JD034839. https://doi.org/10.1029/2021JD034839

Wunsch, A., Liesch, T., & Broda, S. (2018): Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). *Journal of Hydrology*, *567*, 743–758. https://doi.org/10.1016/j.jhydrol.2018.01.045

Xu, C., Zhong, P.-a., Zhu, F., Xu, B., Wang, Y., Yang, L., Wang, S., & Xu, S. (2024): A hybrid model coupling process-driven and data-driven models for improved real-time flood forecasting. *Journal of Hydrology*, *638*, 131494. https://doi.org/10.1016/j.jhydrol.2024.131494

Zhao, G., Pang, B., Xu, Z., Peng, D., & Zuo, D. (2020): Urban flood susceptibility assessment based on convolutional neural networks. *Journal of Hydrology*, *590*, 125235. https://doi.org/10.1016/j.jhydrol.2020.125235

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020): Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Appendix

Appendix

As part of my work at the Chair of Prof. Leandro (Hydromechanics and Hydraulic Engineering), I was extensively involved in teaching activities. I independently designed and supervised exercises as well as course materials, particularly for the subjects Urban Flood Modelling (UFM), Flood Risk and Resilience (FRR), and Hydromechanics and Hydraulic Engineering. In addition, I actively contributed to the organization and implementation of the Summer School Programming for Engineers. Furthermore, I made substantial contributions to the development, preparation, and execution of several research projects. My responsibilities included drafting and writing project proposals as well as contributing to the research work itself. The projects were:

- RiskAware – Development of a Generative Planning Tool for Stormwater Risk-Aware Urban Planning. Duration: 01.03.2025 – 29.02.2028. Coordinator: University of Siegen, Chair of Hydromechanics and Hydraulic Engineering. Partners: Okeanos Smart Data Solutions GmbH. Funding: State of North Rhine-Westphalia through the EFRE/JTF programme.
- PuwaSTAR – Pump Station Warning for Heavy Rainfall and Flooding in Urban Areas. Duration: 01.10.2023 – 30.09.2025. Coordinator: Emschergenossenschaft Lippeverband (EGLV). Partners: Lippeverband, Hydrotec Ingenieurgesellschaft für Wasser und Umwelt mbH, City of Dorsten, Dorsten Fire Department, THW Local Chapter Gladbeck/Dorsten, THW Regional Office Gelsenkirchen, City of Bottrop, Emschergenossenschaft. Funding: German Federal Ministry of Education and Research (BMBF)
- CoDiLe – Collaborative Digital Learning. Duration: 01.01.2022 – 31.12.2022. Funding: Ministry of Culture and Science of the State of North Rhine-Westphalia (MKW NRW), Stifterverband. Programme: Fellowships for Innovations in Digital Higher Education (digiFellows). Focus: Digitalization of higher education.

Furthermore, throughout my research activities at the chair and within the framework of these projects, I contributed to a variety of conference presentations, supervised student work, authored additional scientific publications (as outlined in the sections below), and performed peer review for the *Journal of Hydrology* and the Journal of *Stochastic Environmental Research and Risk Assessment*, completing four reviews between 2024 and 2025.

List of further publications

- Ramsauer, S., Schmid, F., Eckers, H., Falter, D., Johann, G., Leandro, J. (2025): Bayesian analysis of stormwater pump failures and flood inundation extents. Water 2025, 17(19), 2876; https://doi.org/10.3390/w17192876.
- Schmid, F., Ramsauer, S., Eckers, H., Falter, D., Johann, G., Leandro, J. (2025): Vorhersagesystem für Pumpwerke bei Hochwassergefahr. wwt Wasserwirtschaft Wassertechnik, Jahrgang 74, Heft 1-2, Seite 16 - 19. doi.org/10.51202/1438-5716-2025-1-2-016.
- Janicke, L., Schmid, F., James Shucksmith, J., Carvalho, R., F., Datta, S., Leandro, J. (2022): Codile - Collaborative Digital Learning. 39th IAHR World Congress, Granada. DOI:10.3850/IAHR-39WC2521711920221262.

List of conference presentations

The titles of the presentations are given in their original language, which is why some appear in German.

- Ramsauer, S., Schmid, F., Leandro, J., Johann, G., Falter, D., Eckers, H., (2025): A Novel Data-Driven Approach for the Dynamic Prediction of Maximum Flood Inundation Considering Pump Station Failures. Innsbruck. 13th Urban Drainage Modelling (UDM) conference.
- Eckers, H., Buchholz, O., Falter, D., Johann, G., Leandro, J., Nafo, I., Nijzink, J., Pfister, A., Ramsauer, S., and Schmid, F. (2025): Flood forecasting and pumping station warning in urban areas to improve flood-resilience: The project PuwaSTAR, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-16233, https://doi.org/10.5194/egusphere-egu25-16233.
- Schmid, F., Althoff, I., Müller, L., Oppel, H., Leandro, J. (2025): Zur Entwicklung eines generativen Planungstool zur Starkregenrisiko bewussten Stadtplanung. Augsburg, Tag der Hydrologie 2025.
- Ramsauer, S., Schmid, F., Leandro, J., Johann, G., Falter, D., Eckers, H. (2025): Entwicklung eines neuartigen datengetriebenen Überflutungsvorhersagemodells für Pumpwerke (Poster). Augsburg, Tag der Hydrologie 2025.
- Schmid, F., Leandro, J. (2024): Zur Entwicklungen hinsichtlich der Übertragbarkeit eines datengetriebenen Überflutungsvorhersagemodells auf unbekannte Einzugsgebiete. Berlin, Tag der Hydrologie 2024.
- Falter, D., Schmid, F., Ebener, A., Leandro, J., Pfister A., Teichgräber, B., Johann, G. (2024): Pumpwerkswarnung für Starkregen und Hochwasser im urbanen Raum: Das Projekt PuwaSTAR. Berlin, Tag der Hydrologie 2024.
- Schmid, F., Leandro, J. (2023): KI basierte „multi-time step“ Überflutungsvorhersage von fluvialen Ereignissen. Essen, DWK-Workshop 2023.
- Schmid, F., Leandro, J. (2023): Zur Entwicklung eines zeitlich dynamischen „multi-step“ Echtzeit-Überflutungsvorhersagesystems. Bochum, Tag der Hydrologie 2023.
- Schmid, F., Leandro, J., (2022): Towards the development of innovative data-driven inundation forecast systems. Medellín-Semiar 2022.
- Schmid, F., Leandro, J., Disse, M. (2022): Zur Entwicklung und Innovation datenbasiert 2D-Überflutungsvorhersagesystem. München, Tag der Hydrologie 2022.
- Schmid, F., Leandro, J. (2021): Zur Entwicklung eines flexiblen Frühwarnsystems für kanalnetzbedingte Überschwemmungen (Poster). Potsdam, Tag der Hydrologie 2021.

List of students supervisions

The titles of the theses are given in their original language, which is why some appear in German.

- Leonie Müller (2025) - Masterarbeit (M.Sc. Bauingenieurwesen): Entwicklung einer generalisierten KI für dynamische Hochwasservorhersagen. Universität Siegen.
- Daniel Lisowski (2025) - Master thesis (M.Sc. Environmental engineering): Development of a Temporally Dynamic Real-Time Flood Forecasting Model Utilizing Convolutional Neural Networks. Technical University of Munich.
- Leonie Müller (2025) - Studienarbeit (M.Sc. Bauingenieurwesen): Zur Entwicklung einer generalisierten KI für dynamische Hochwasservorhersagen. Universität Siegen.
- Leonie Müller (2024) - Studienarbeit (M.Sc. Bauingenieurwesen): Ein 1D/2D gekoppeltes Überflutungsmodell vom Einzugsgebiet des Pumpwerks Dorsten-Hammbach unter Verwendung von P-DWave und SWMM. Universität Siegen.
- Lars Erner (2024) - Studienarbeit (M.Sc. Bauingenieurwesen): Hydronumerische Untersuchungen im Einzugsgebiet des Sottebachs in Freudenberg, Nordrhein-Westfalen. Universität Siegen.
- Florian Spieß (2023) - Bachelorarbeit (B.Sc. Bauingenieurwesen): Erstellung eines Entwässerungssystems für Regenwasser unter der Verwendung von SWMM am Beispiel der Stadt Siegen – Gemarkung Weidenau. Universität Siegen.
- Emine Baykal-Kara (2023) - Masterarbeit (M.Sc. Bauingenieurwesen): Ensemblebasierte Quantifizierung von Unsicherheiten des Hochwasserrisikos für die Stadt Baiersdorf. Universität Siegen.
- Annalena Sporer (2022) - Master thesis (M.Sc. Environmental engineering): Image-To-Image Translation for Pluvial Flood Prediction Using Two Different Architectures: U-Net and cGAN. Technical University of Munich.
- Emine Baykal-Kara (2022) - Studienarbeit (M.Sc. Bauingenieurwesen): 1D/2D gekoppelte Überflutungsanalyse des Kanalnetzes Baiersdorf mittels SWMM und P-DWave. Universität Siegen.
- Julia Feilke (2021) - Bachelorarbeit (B.Sc. Bauingenieurwesen): Hochwasserrisikobewertung der Stadt Baiersdorf unter Berücksichtigung von Unsicherheiten in den Niederschlagsintensitätsverteilungen. Universität Siegen.
- Hongfei Zhu (2021) - Master thesis (M.Sc. Environmental engineering): CNN-based Approaches for Forecasting Maximum Fluvial Flood Inundation: Case Study of Kulmbach. Technical University of Munich.
- During 2022 and 2025 I was responsible for the supervision of student assistants, including delegation and coordination of tasks.