



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

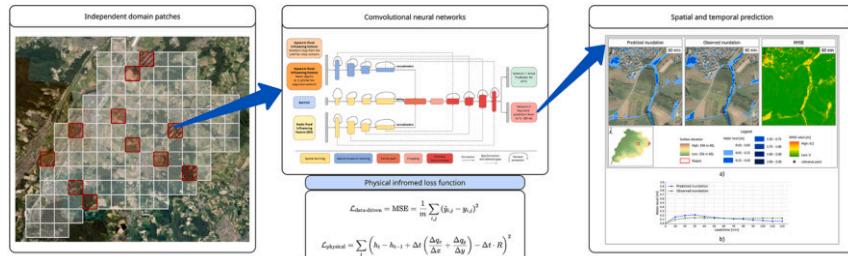
Felix Schmid <sup>\*</sup> , Leonie Müller , Jorge Leandro

*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](mailto:Felix.Schmid@uni-siegen.de) (F. Schmid), [Leonie.Mueller@uni-siegen.de](mailto:Leonie.Mueller@uni-siegen.de) (L. Müller), [Jorge.Leandro@uni-siegen.de](mailto:Jorge.Leandro@uni-siegen.de) (J. Leandro).

## 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.  $1 \times 1$  m or  $5 \times 5$  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 will provide 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

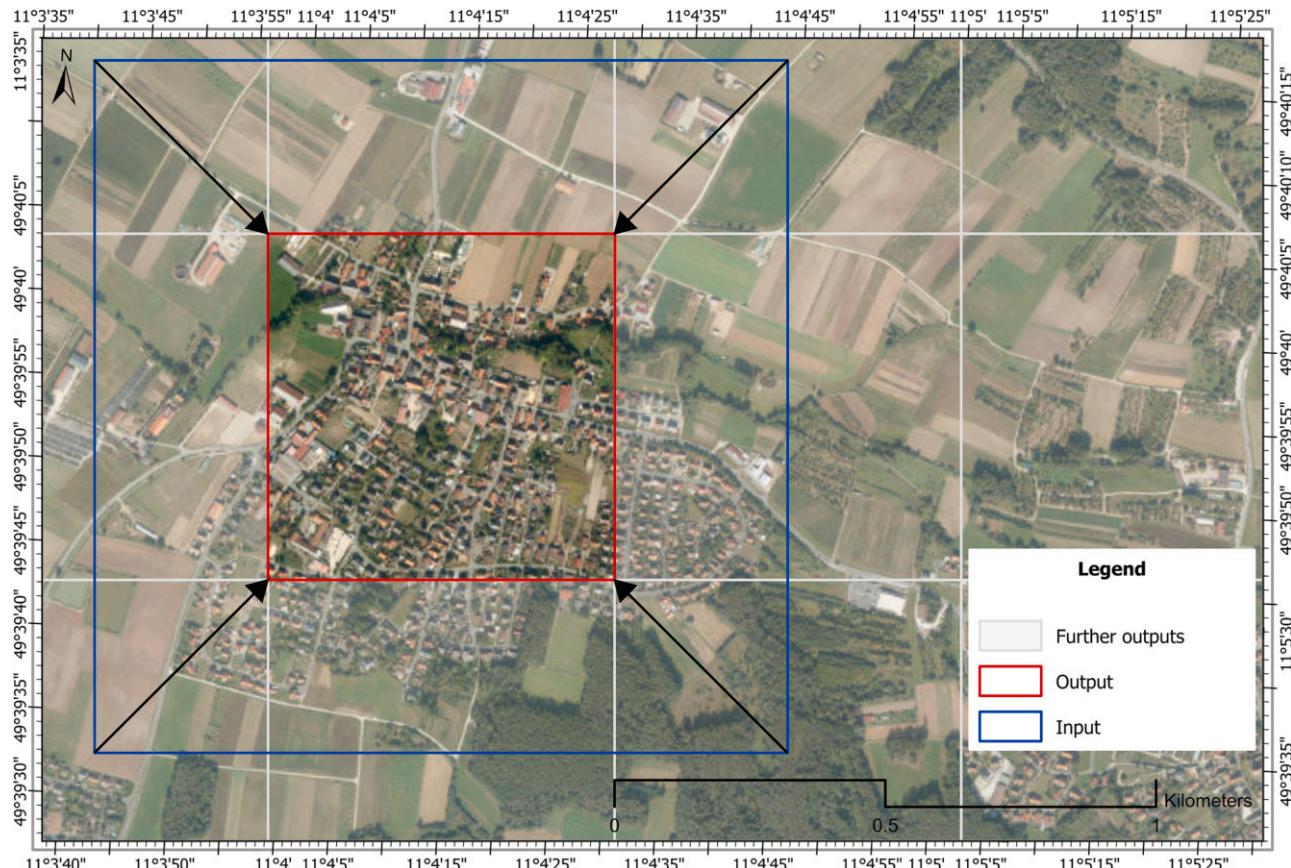
The PI-FS utilizes a CNN architecture. To enhance the model's ability to capture spatial dynamics, each input frame is provided at a larger input frame, while the model outputs predictions, focusing on the central subregion. This spatial prediction framework is presented in Fig. 1. The central area is 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

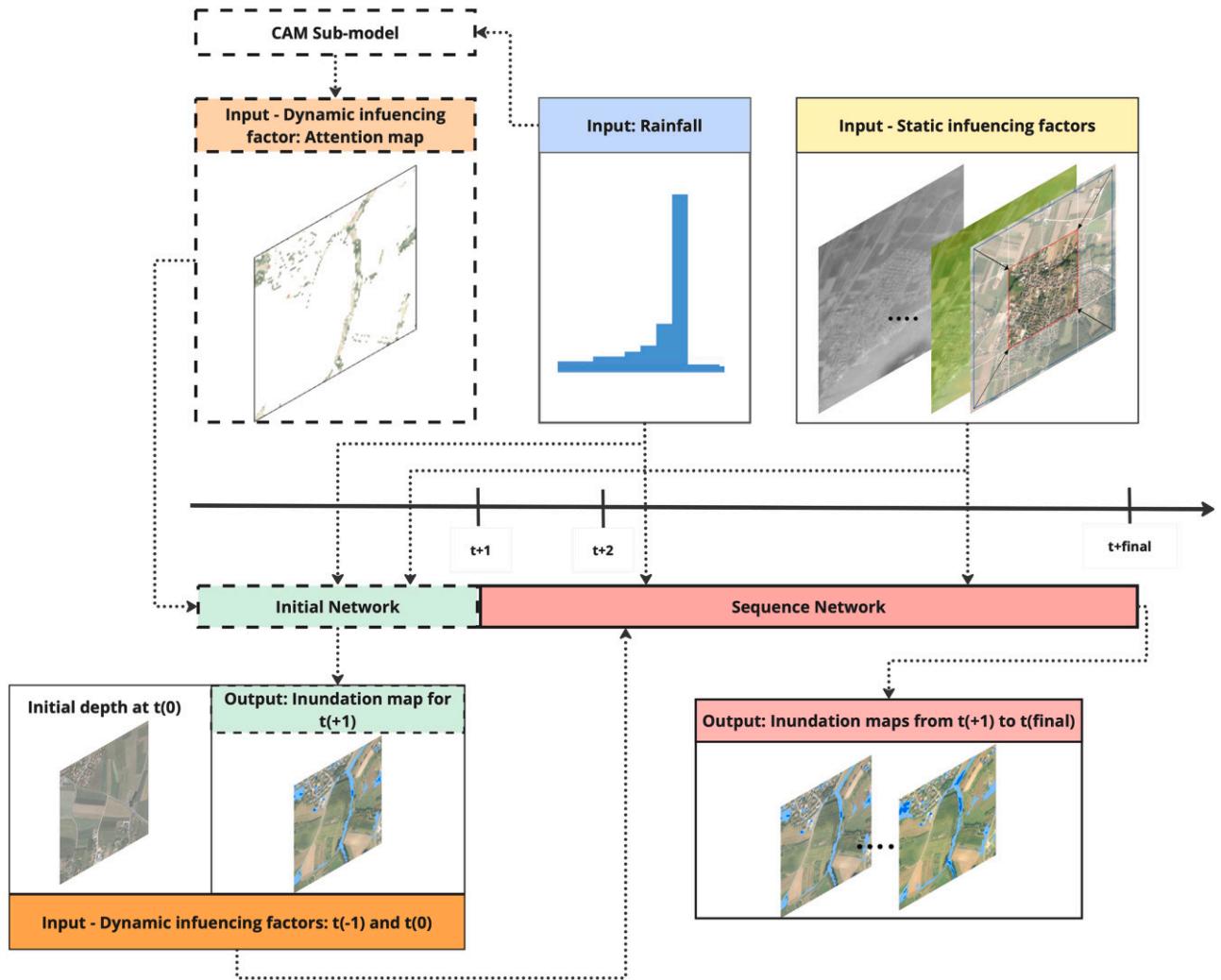
(visualized in blue in Fig. 1), centered around the region of interest, and generates a smaller output prediction for the central area (visualized in red in Fig. 1). By systematically sliding/iterating across the domain in a grid-like fashion, the model generates predictions for the entire area (like 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  $512 \times 512$  and predicted the  $256 \times 256$  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  $t(\text{final})$  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.

$$LF_{\text{data-driven}} = MSE = \frac{1}{m} * \sum_{ij} (\widehat{y}_{i,j} - y_{i,j})^2 \quad (1)$$

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)):

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

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

$$q_x = h_t u_t; \quad q_y = h_t v_t \quad (4)$$

These equations are able to describe the water movement in a 2-dimensional grid. The PI-FS only forecasts water depths at each timestep, 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  $n$  is the Manning roughness,  $\beta$  the surface slope in the respective  $x$  and  $y$  direction, and  $h$  the water depth.

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

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

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

By replacing  $h$  with  $\widehat{y}_{ij}$  (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.

$$LF_{overall} = LF_{data-driven} + (LF_{physical})^2 = 0 \quad (7)$$

### 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,

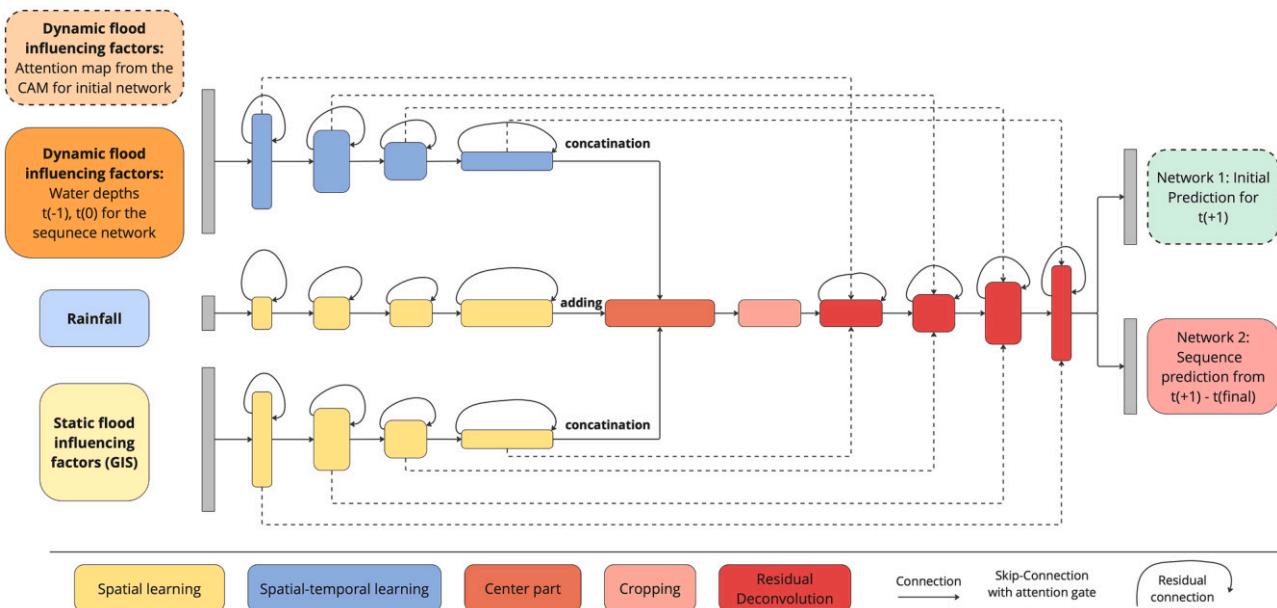
**Table 1**

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.
<b>Dynamic factors</b>	
Attention map <sup>1</sup>	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: Berkahn and Neuweiler, 2024; Schmid and Leandro, 2024.
Water depth at t(0) <sup>2</sup>	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: Berkahn and Neuweiler, 2024.
Rainfall	Is the driving inundation force and input to the physical loss function.

<sup>1</sup> Used in the initial network;

<sup>2</sup> 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  $t(+2,3, \dots, \text{final})$  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

**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 [%]	$\left( \frac{\sum (\widehat{y}_{ij} - y_{ij}) * A}{\sum y_{ij} * A} \right) * 100$	[-100,100]	0 %
RMSE [m]	$\sqrt{\frac{1}{n} * \sum_{ij}^n (\widehat{y}_{ij} - y_{ij})^2}$	[0,∞)	0m
Mean RMSE [m]	$\frac{1}{N} * \sum RMSE_{ij}$	[0,∞)	0m

shown in Table 2.

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),  $\widehat{y}_{ij}$  are the water depth prediction in the respective cells,  $y_{ij}$  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:

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

$$g \nabla(h + z) = gS_f \quad (9)$$

Where,  $h$  is the water depth,  $u = [u_x u_y]^T$  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  $S_f = [S_{fx} S_{fy}]$  is the bed friction. For more details on the P-Dwave, please refer to Leandro et al. (2014).

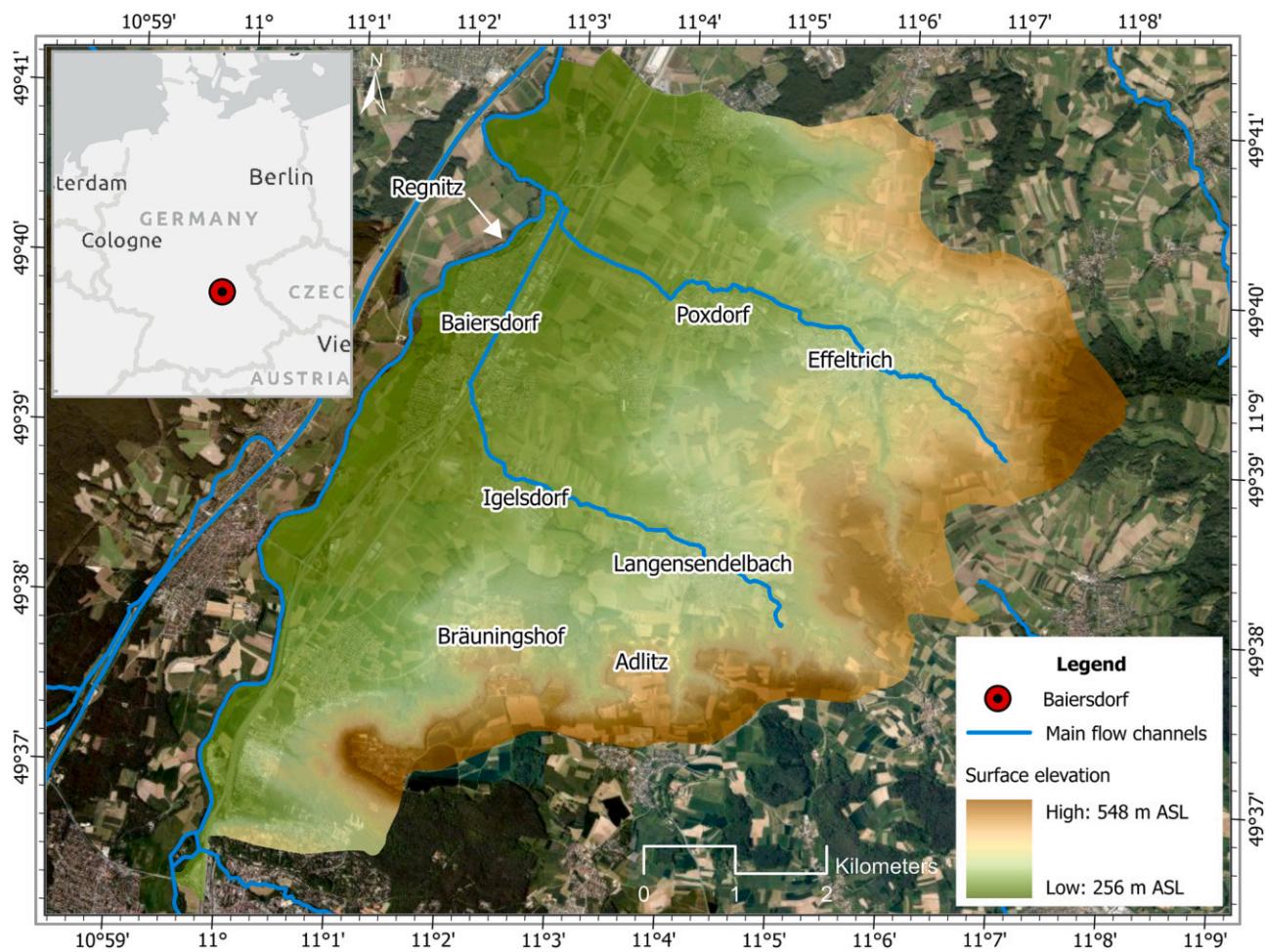
### 2.7.2. Cellular-Automata approach

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).

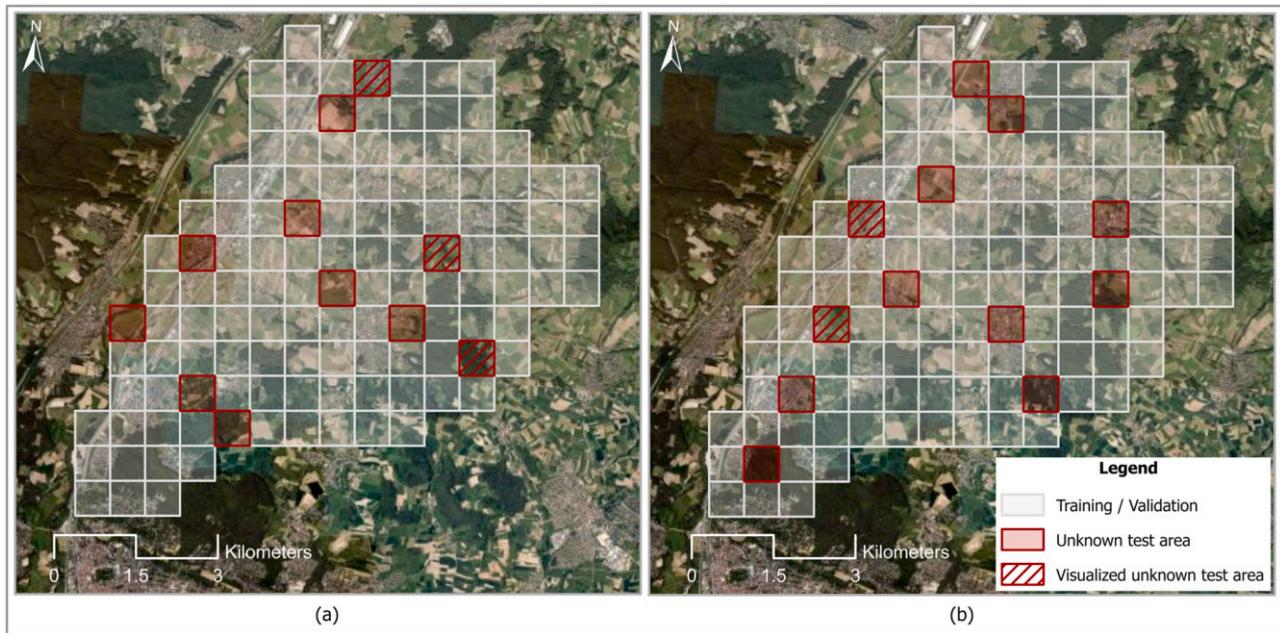
## 3. Study area and implementation

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<sup>2</sup>. 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

**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.

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.

### 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

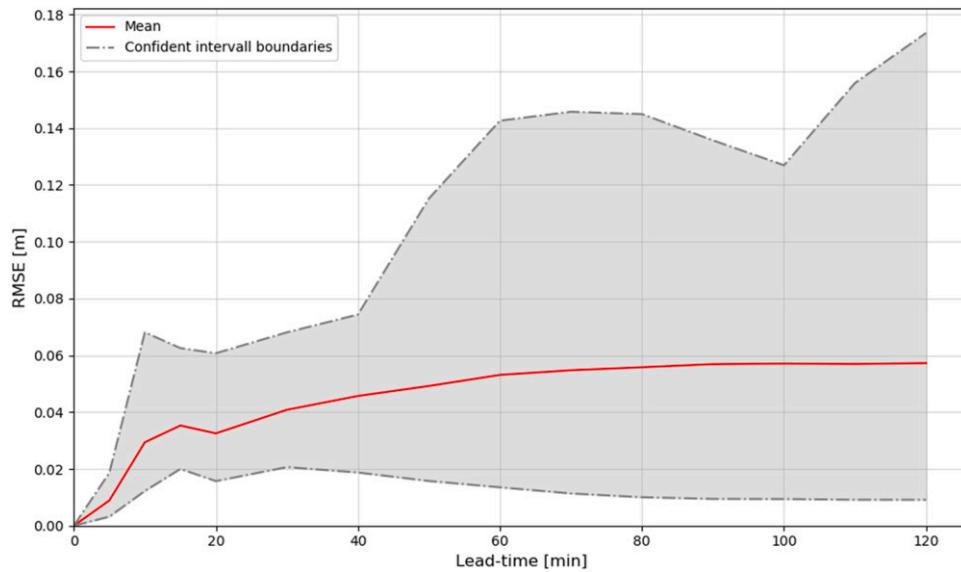
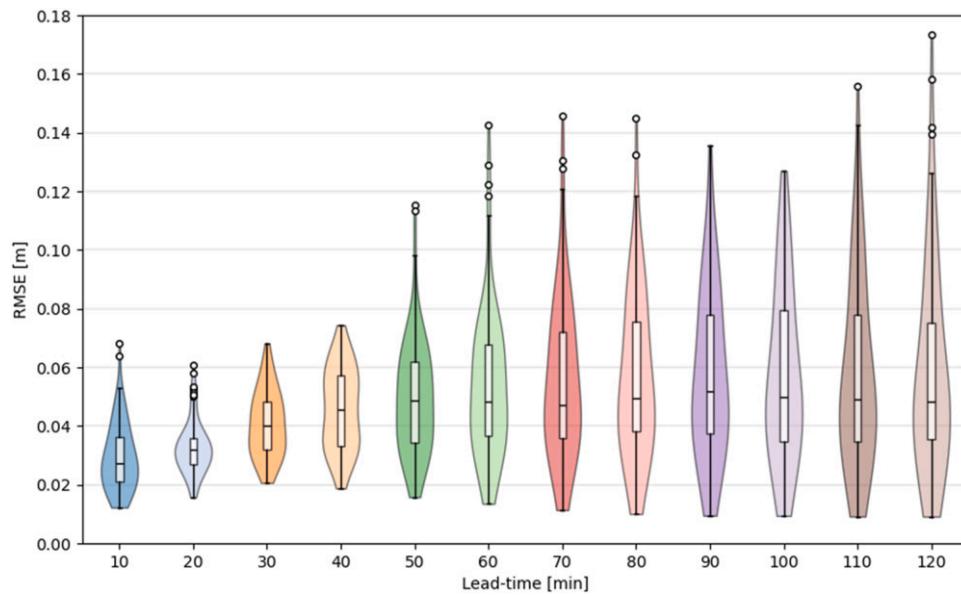
### 5.1. Network setup

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						Min. RMSE	Mean RMSE	Max. RMSE			
	CSI [%]			RVE [%]								
	Min. CSI	Mean CSI	Max. CSI	Min. RVE	Mean. RVE	Max. RVE						
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			

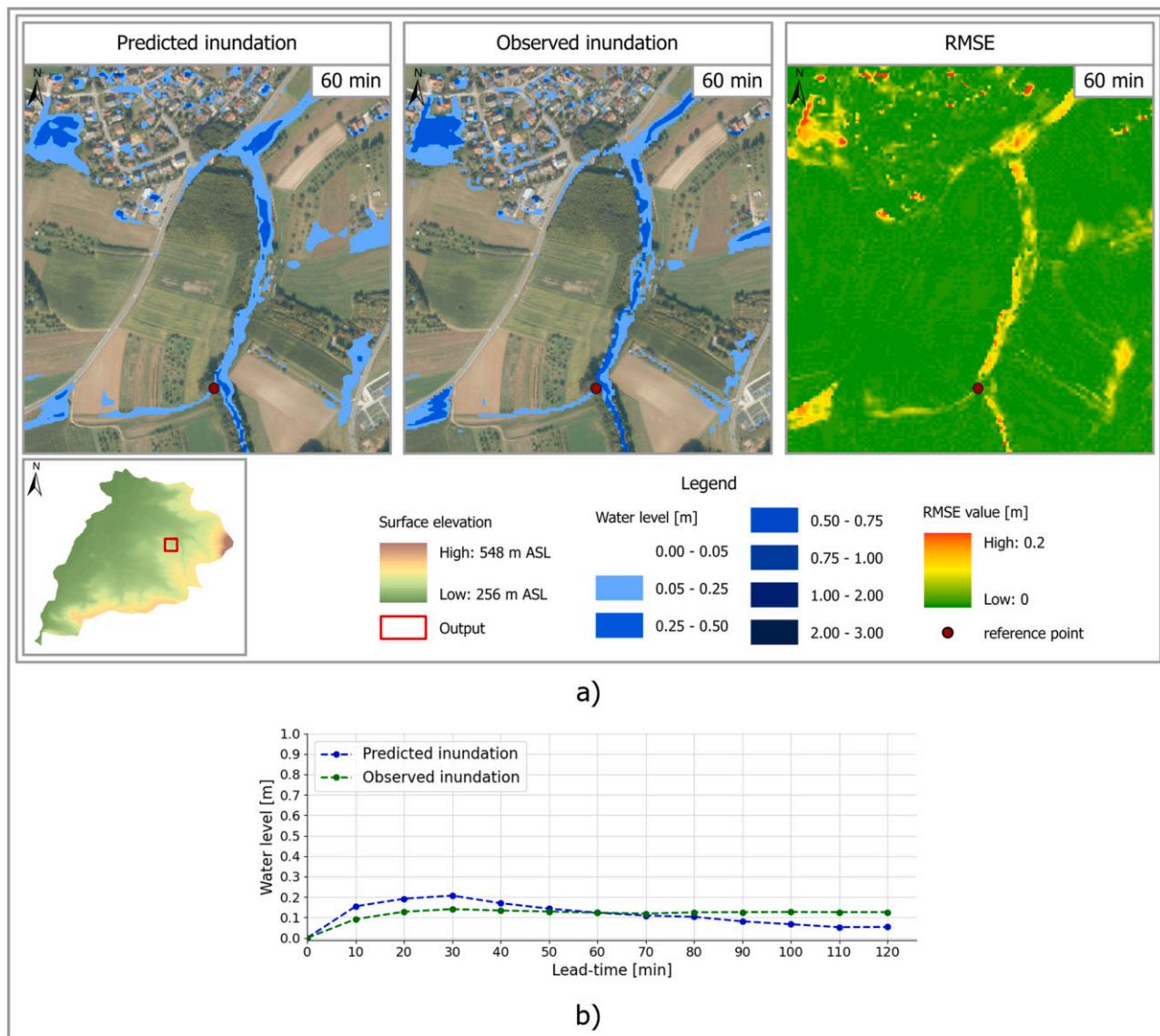
**Fig. 6.** Lead time error of the forecast system based on the mean RMSE.**Fig. 7.** Error distribution and histogram for the respective forecasting horizon.

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.).

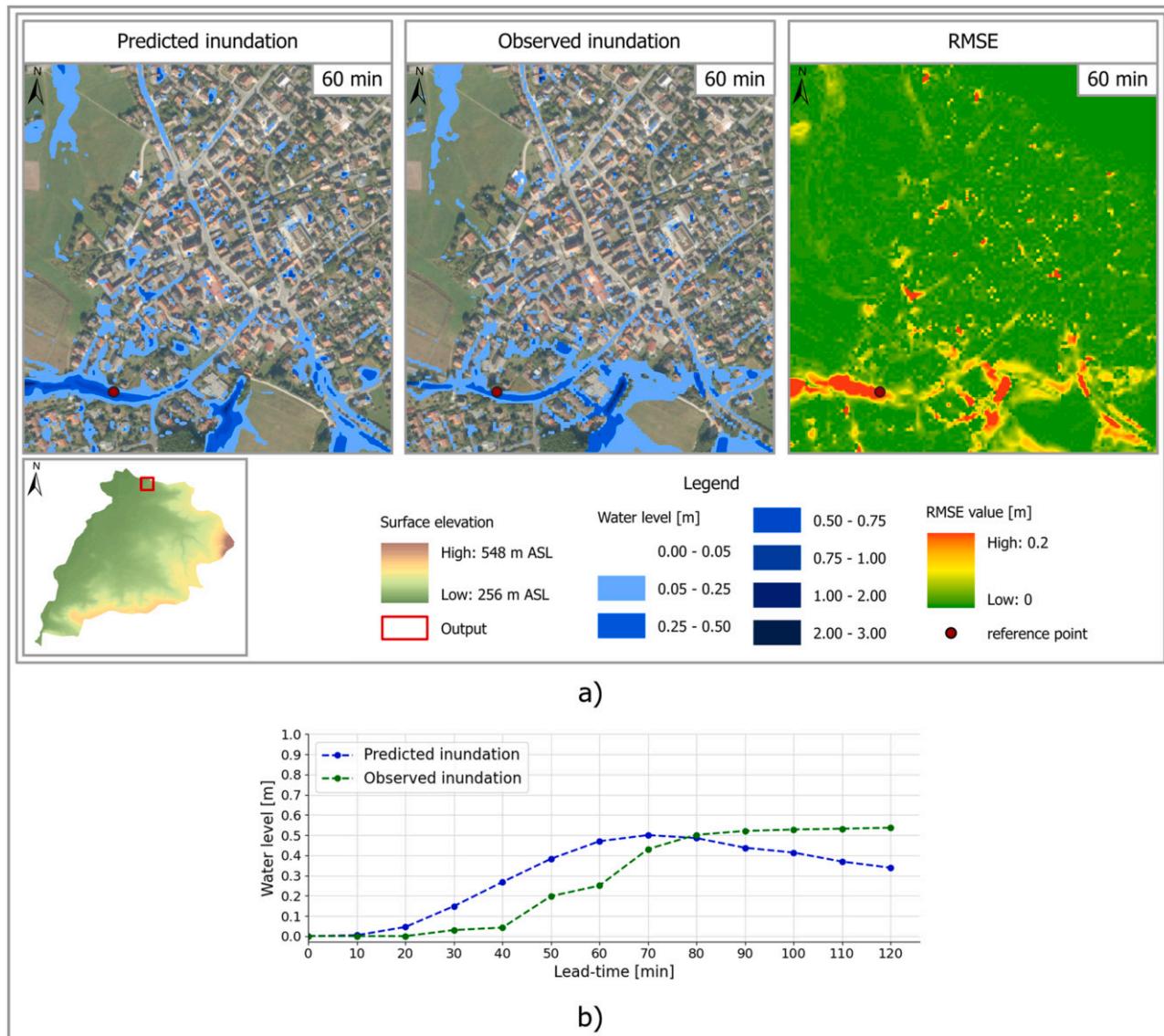
lag order increases, meaning the network requires more temporal steps to correctly propagate information through the sequence, which reduces accuracy at the last predicted timesteps. Additionally, error accumulation over time is more pronounced with coarser resolutions, as deviations at each timestep 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.).

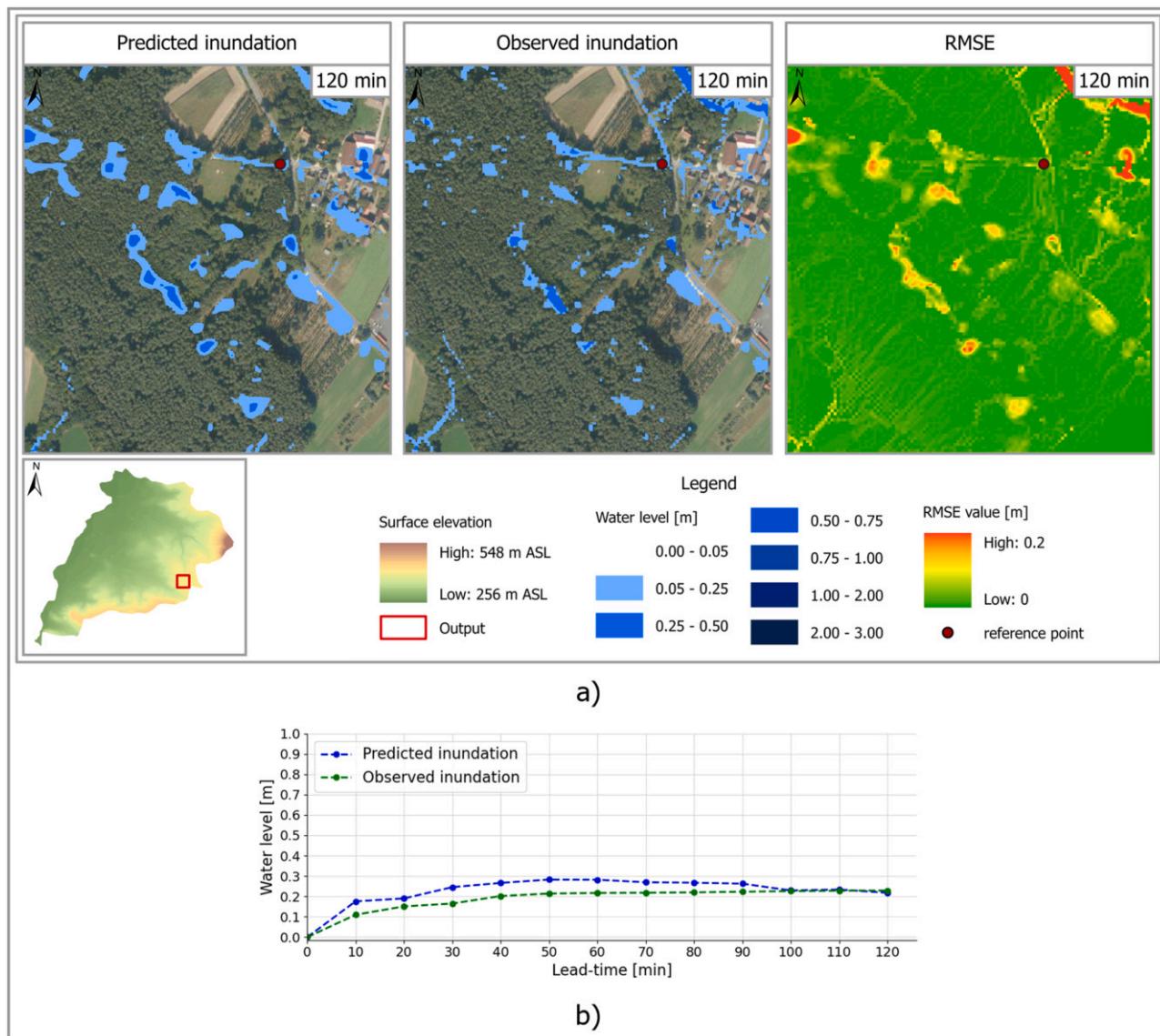
between 0.08 m and 0.2 m, and Guo et al. (2022) reported RMSEs of about 0.2 m for wet cells without CSI evaluation. Thus, our model reduces RMSE by about half relative to Guo et al. (2022) and improves CSI by about 20 percentage points compared to Seleem et al. (2023) and Löwe et al. (2021). In comparison, 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 down-sampling 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.

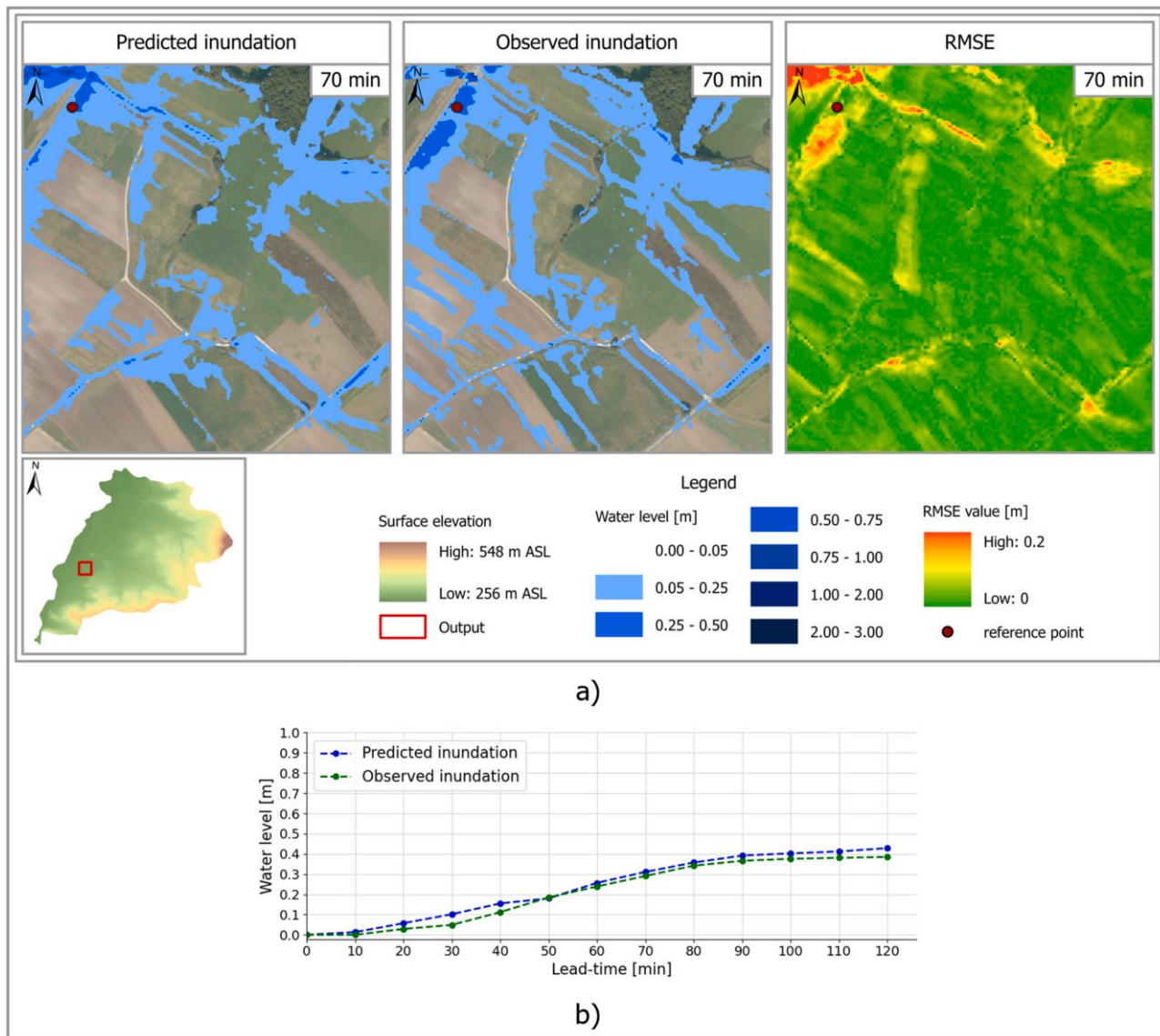
### 5.3. Inundation maps and spatial prediction error

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

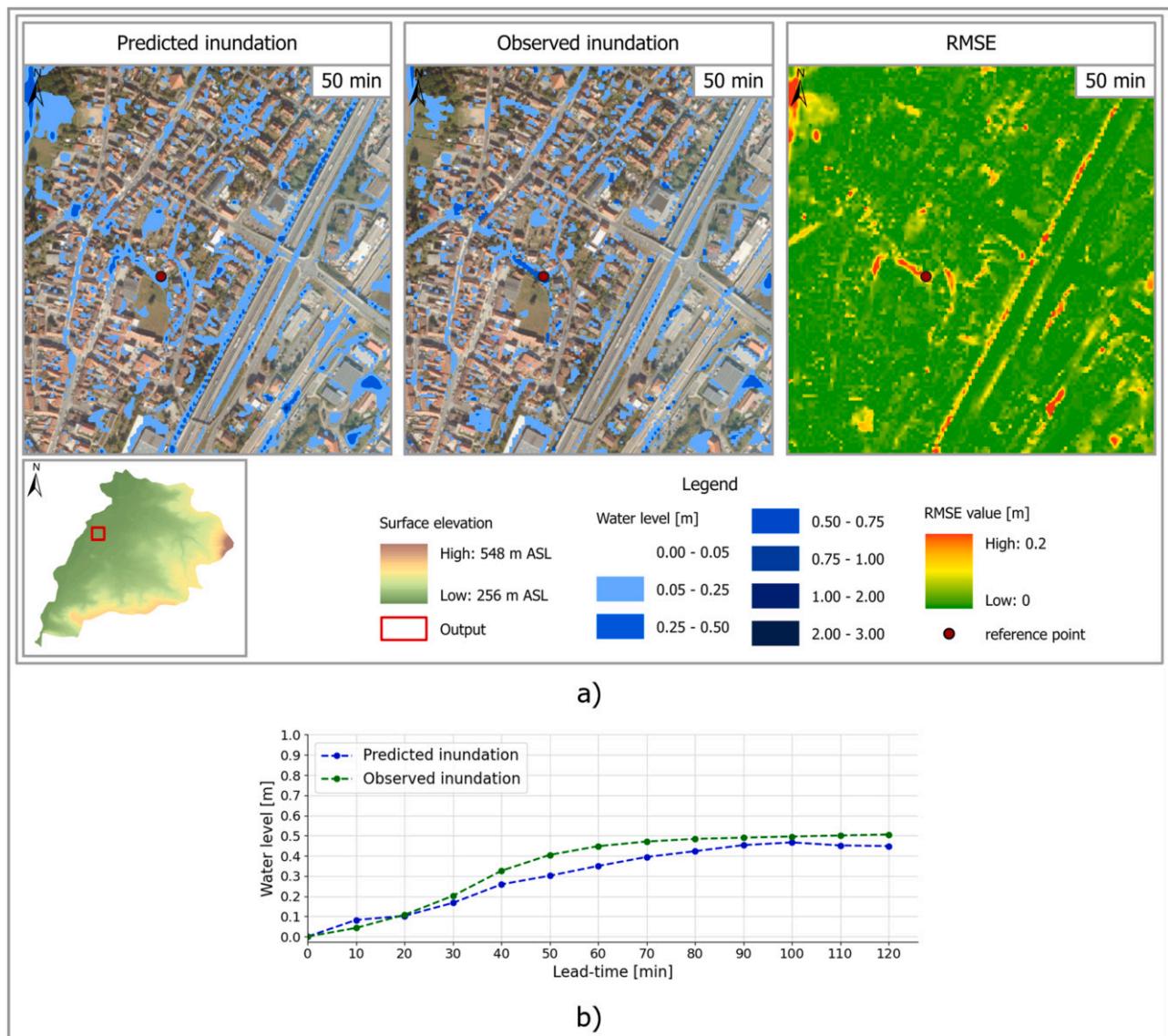
The results presented in Table 6 and Fig. 13 demonstrate the clear advantages of incorporating physical knowledge into the loss function. The physically informed loss Function (PI-LF) consistently outperforms the purely data-driven Mean Squared Error (MSE) loss across all evaluation metrics on both the 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.

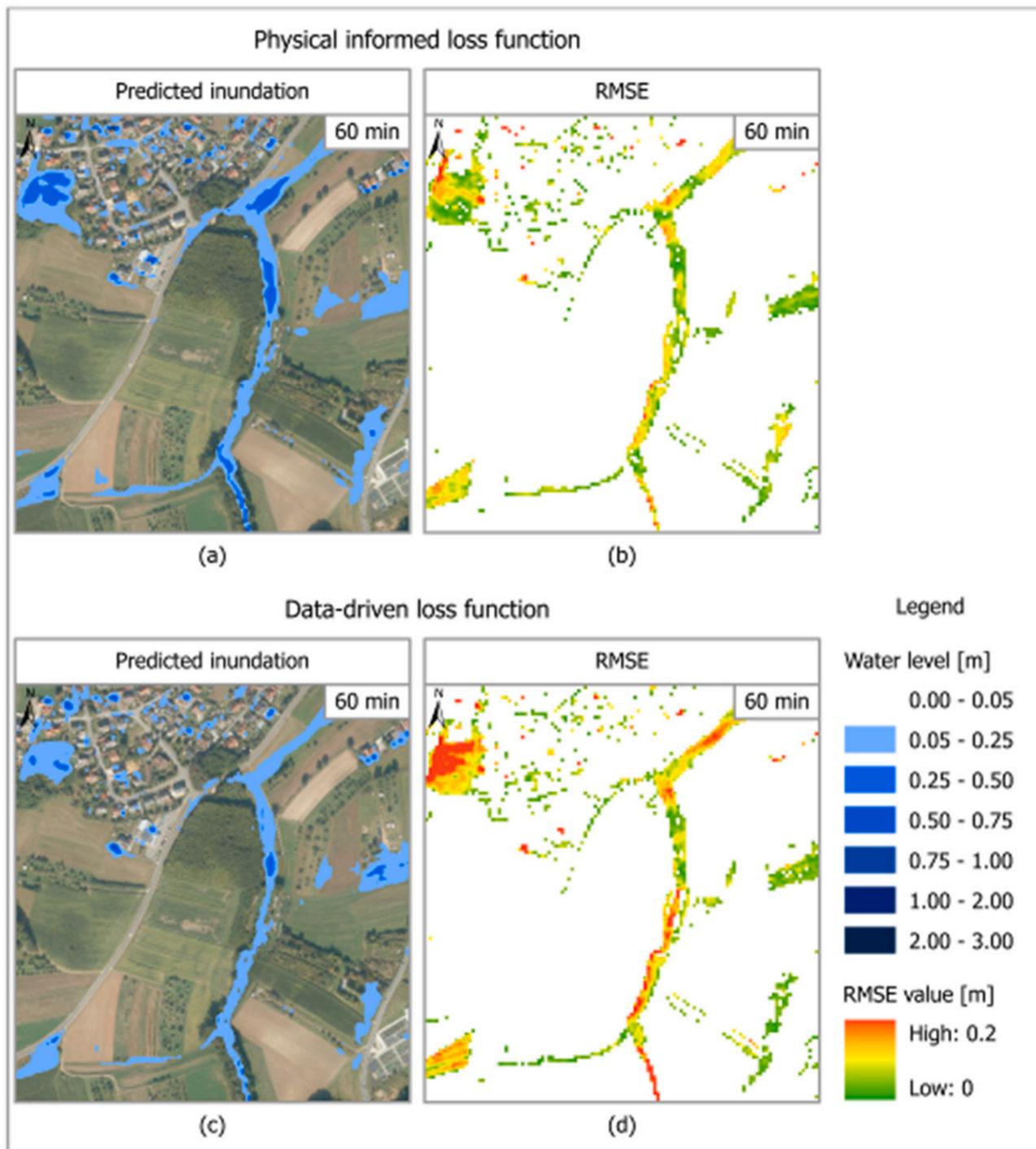
	Mean RMSE [m]	CSI [%]	RVE [%]
Example 1 / CV Fold 1 (Fig 8)	0.032	83.7	10.4
Example 2 / CV Fold 1 (Fig 9)	0.055	71.2	30.8
Example 3 / CV Fold 1 (Fig 10)	0.033	67.7	-6.5
Example 4 / CV Fold 2 (Fig 11)	0.039	75.5	-1.1
Example 5 / CV Fold 2 (Fig 12)	0.042	68.0	17

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](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.jhydro.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., Daneshkhan, 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., Daneshkhan, 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. Model. 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: Kolathayor, 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](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. Model. Softw. 141, 105034. <https://doi.org/10.1016/j.envsoft.2021.105034>.
- Oktay, O., Schlempert, 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>.
- Raiissi, 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>.