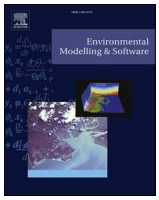
Environmental Modelling and Software 191 (2025) 106499

Contents lists available at ScienceDirect

Environmental Modelling and Software

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



A cluster-based temporal attention approach for predicting cyclone-induced compound flood dynamics

Samuel Daramola [a,*], David F. Muñoz [a], Hamed Moftakhari [b], Hamid Moradkhani [b]

[a] Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, 24061, USA
[b] Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL, 35487, USA

ARTICLE INFO

Keywords:
Compound flood dynamics
Cyclones
ConvLSTM
Attention mechanism
Hurricane Beryl

ABSTRACT

Deep learning (DL) models have been used for rapid assessments of environmental phenomena like mapping compound flood hazards from cyclones. However, predicting compound flood dynamics (e.g., flood extent and inundation depth over time) is often done with physically-based models because they capture physical drivers, nonlinear interactions, and hysteresis in system behavior. Here, we show that a customized DL model can efficiently learn spatiotemporal dependencies of multiple flood events in Galveston, TX. The proposed model combines the spatial feature extraction of CNN, temporal regression of LSTM, and a novel cluster-based temporal attention approach to assimilate multimodal inputs; thus, accurately replicating compound flood dynamics of physically-based models. The DL model achieves satisfactory flood timing (±1 h), critical success index above 60 %, RMSE below 0.10 m, and nearly perfect error bias of 1. These results demonstrate the model's potential to assist in flood preparation and response efforts in vulnerable coastal regions.

1. Introduction

Coastal areas are increasingly vulnerable to compound flood events, which occur when storm surges, river overflow, and heavy rainfall occur simultaneously or in close succession (Eilander et al., 2023; Muñoz et al., 2021; Xu et al., 2024). Although these events have historically been rare, their frequency and intensity have risen significantly in recent years (Taherkhani et al., 2020; Wing et al., 2022). This increase is primarily attributed to climate change, which drives sea level rise, marine heatwaves, torrential rainfall, intense winds from tropical cyclones, glacier melting, land subsidence, and altered ocean currents (Bloomfield et al., 2023; Levy et al., 2024; Ohenhen et al., 2023; Radfar et al., 2024; Santiago-Collazo et al., 2019; Thiéblemont et al., 2024; Volkov et al., 2023). Currently, approximately 700 million people and an estimated $13 trillion in assets located in areas less than 10 m above mean sea level worldwide are at risk of flooding (Kirezci et al., 2023). In the United States alone, tropical cyclones intensify rainfall and river overflow, contributing to over $13.8 billion in flood damages and 86 flood fatalities out of a total $280 billion in cyclone-related damages and 683 deaths over the past five years (NOAA-NCEI, 2024). Hence, the imperative to understand flood dynamics caused by compound events is driving the exploration of advanced modeling approaches that can enhance current prediction methods and support flood management practices.

Methods for predicting compound and coastal flood dynamics rely on both physically-based and data-driven approaches (Chiang et al., 2024; Fraehr et al., 2022; Hu et al., 2019; Löwe et al., 2021; Marsooli and Wang, 2020; Sampurno et al., 2022; Shahabi and Tahvildari, 2024). Physically-based approaches, such as hydrodynamic and hydraulic models, are particularly effective at estimating the interaction of multiple flood drivers such as rainfall, river discharge, and storm surges (Bates, 2022, 2023; Leijnse et al., 2021; Santiago-Collazo et al., 2024). These approaches demonstrate strong predictive skills across multiple flood scenarios by complying with physical constraints such as the conservation of mass and momentum over time and space (Alipour et al., 2022; Camus et al., 2021; Jibhakate et al., 2023; Peña et al., 2022; Zhong et al., 2024). However, they require sufficient computational resources and observational (forcing) data, even for short-term (hourly to daily) predictions over large-scale domains (Bilskie et al., 2021; Nezhad et al., 2023). Data-driven models are generally efficient in this regard since they leverage statistical and machine learning techniques to learn from nonlinear associations and infer data patterns, thus rapidly generating flood predictions without relying on complex physical simulations (Gomez et al., 2024; Lewis et al., 2024; Moftakhari et al., 2017).

* Corresponding author.
E-mail address: samueldaramola@vt.edu (S. Daramola).

https://doi.org/10.1016/j.envsoft.2025.106499
Received 3 February 2025; Received in revised form 23 April 2025; Accepted 26 April 2025
Available online 27 April 2025

Statistical approaches typically require explicit assumptions about data distribution to predict flood data associated with a given return period (e.g., peak storm surge, river flow, and rainfall) (Boumis et al., 2023; Maduwantha et al., 2024; Moftakhari et al., 2021; Zhong et al., 2024).

Machine learning techniques, particularly deep learning (DL) models, can capture complex nonlinear associations and hidden patterns from input data features; thus, enhancing the prediction accuracy (Foroumandi et al., 2024; Fu et al., 2022; Sattari et al., 2025). Recently, transfer learning techniques have been incorporated into DL model architectures to further enhance the predictive capability in areas with limited or scarce data (Obara and Nakamura, 2022; Seleem et al., 2023). Previous studies have noted accurate prediction of storm surges, extreme water levels, river flow, among other flood drivers (Daramola et al., 2025; Green et al., 2025; Hussain and Khan, 2020; Mahakur et al., 2025; McKeon and Piecuch, 2025; Muñoz et al., 2021; Samantaray et al., 2025; Tang et al., 2025; Tiggeloven et al., 2021). Nevertheless, DL models are mostly used for station-based flood predictions with hazard and risk map generation focused on the peak flood extent (Ayyad et al., 2022; Tedesco et al., 2024). This is because many conventional DL models excel at either spatial (e.g., Graph Convolution Networks, Convolutional Neural Networks) or temporal analysis of flood events (e.g., Long Short-Term Memory networks, Gated Recurrent Units). As a result, they often lack the ability to process flood dynamics in both space and time. To overcome this limitation, studies tend to use DL frameworks that combine spatial and temporal architectures, preferably employing static datasets for spatial analyses and dynamic features for temporal contribution (Farahmand et al., 2023; Fathi et al., 2025). Additionally, DL models designed for spatiotemporal analysis like Convolutional Long Short-Term Memory (ConvLSTM) may effectively capture general spatiotemporal patterns but struggle to accurately represent the varying conditions across the spatial domain. The latter results from the lack of inherent physical constraints in DL models to guide flood dynamics in contrast to those present in hydrodynamic models. Furthermore, regional variations in spatial flood dependencies are ignored when conducting current and future flood risk assessments (Brunner et al., 2020).

Given the challenges described, a key question arises: which DL techniques can effectively capture hysteresis in system behavior across spatiotemporal domains? In flood dynamics, hysteresis occurs when water levels depend not only on current conditions but also on prior states, producing lagged effects that vary in space and time (Wu et al., 2023). Spatially, water level variations at one location can induce changes in water levels of adjacent areas due to factors like bottom friction (roughness) effects on the propagation of flood waves as well as morphological features that either attenuate or amplify water levels in coastal and estuarine systems (Hoitink and Jay, 2016; Prandle, 1985; Talke and Jay, 2020). Temporally, peak water level timing might differ across inland, transition, and coastal regions, resulting in peak water levels occurring earlier or later depending on distance from the flood source, local topography and bathymetry, or flood connectivity. This complex interplay challenges traditional DL models, which often fail to capture these lagged dependencies effectively (Brunner et al., 2020). As a result, such models may erroneously assume a single peak water level across a large domain (Yu et al., 2024). This oversight can miss critical timing differences in peak water levels, such as those along tidally-influenced rivers or estuaries driven by storm surges, river flow, and backwater effects (Hoitink and Jay, 2016; Sandbach et al., 2018). To model hysteresis accurately, advanced DL frameworks must integrate spatiotemporal dependencies across the model domain including the interconnected nature of flooding processes and different peak water level timings.

To address these limitations, we propose architectural modifications that enable DL models to better capture the spatiotemporal dynamics of compound flood events, as simulated in hydrodynamic models. Specifically, our approach involves three key modifications: (i) implement mechanisms that enable the model to selectively focus on and weigh the importance of features, enhancing their ability to understand and replicate intricate spatiotemporal dependencies; (ii) cluster the spatial domain regionally and modulate with temporally varying contributions based on water level dynamics at observational stations to reflect water level variability over time and space; and (iii) leverage dynamic input data from all flood drivers (just like in hydrodynamic models) to facilitate the learning process of nonlinear interactions and improve the model's predictive accuracy. Our proposed DL model is applied to Galveston, TX, on the Gulf Coast of the United States, a region that has been significantly affected by torrential rainfall and cyclone-induced flood events.

2. Study area

The Galveston Bay (G-Bay) is the seventh largest estuary in the US and serves as a vital hydrological link between Houston, Texas, and the Gulf of Mexico through a complex network of bayous, interior bays, and rivers (Muñoz et al., 2024). Extending 56 km in length and averaging 31 km in width, G-Bay, with a shallow depth of approximately 2 m, encompasses an expansive surface area of about 1600 km^2 (Fig. 1a). The combination of substantial freshwater discharge and dynamic tidal forces creates a sensitive hydrodynamic environment susceptible to rapid changes in water levels. Additionally, G-Bay's complex morphology resembles a bottle-like neck that connects the Buffalo Bayou River with the Bay. Such a morphologic feature exacerbates the impacts from coastal and inland flood drivers, including storm surges and rainfall-runoff (Muñoz et al., 2022; Valle-Levinson et al., 2020), and makes it particularly susceptible to compound flood events affecting both Galveston and Harris Counties.

For this study, we selected six of the most recent flooding events that have affected G-Bay, including Hurricanes Ike (2008), Harvey (2017), Nicholas (2021), and Beryl (2024), as well as the torrential rainfall events of Memorial Day (2015) and Tax Day (2016). Making landfall near Galveston on September 13, 2008, as a Category 2 hurricane, Hurricane Ike brought maximum sustained winds of approximately 175 km/h and a storm surge exceeding 4.5 m (National Weather Service, 2008). The hurricane caused extensive flooding in Galveston and Harris Counties. Torrential rainfall on May 23–24, 2015, around Memorial Day resulted in widespread flooding across South-Central Texas. Southern Blanco County received a record of 0.25–0.33 m leading to Blanco River at Wimberley rising from nearly 1.5 m to over 12.5 m within hours, causing severe flooding (National Weather Service, 2015). Similarly, about 0.61 m of rain fell in Houston and in Waller County from April 17–18, 2016, filling local reservoirs to full capacity and hitting record high water levels (ABC News, 2019).

Harvey made landfall near Rockport, Texas, on August 25, 2017, as a Category 4 hurricane with winds of 215 km/h. The storm stalled over Southeast Texas, dropping over 1.27 m of rain in some areas, resulting in catastrophic flooding, particularly in the Houston metropolitan area (National Environmental SatelliteDataand Information Service, 2024). Nicholas made landfall on September 14, 2021, near Sargent Beach, Texas, as a Category 1 hurricane with maximum sustained winds of 120 km/h, causing significant rainfall and flooding in Southeast Texas and impacting communities around Galveston Bay (National Hurricane Center, 2022). Beryl made landfall near Matagorda, Texas, on July 8, 2024, as a Category 1 hurricane with winds of 129 km/h, leading to extensive flooding in Galveston and Harris Counties (NASA earth observatory, 2024). The six events resulted in over 210 fatalities and approximately $200 billion in economic damages. This high susceptibility to compound flooding renders Galveston Bay an ideal site for applying the proposed DL model.

3. Methodology

We developed a DL framework that incorporates information from a physically-based model, namely Delft3D-FM, to estimate compound

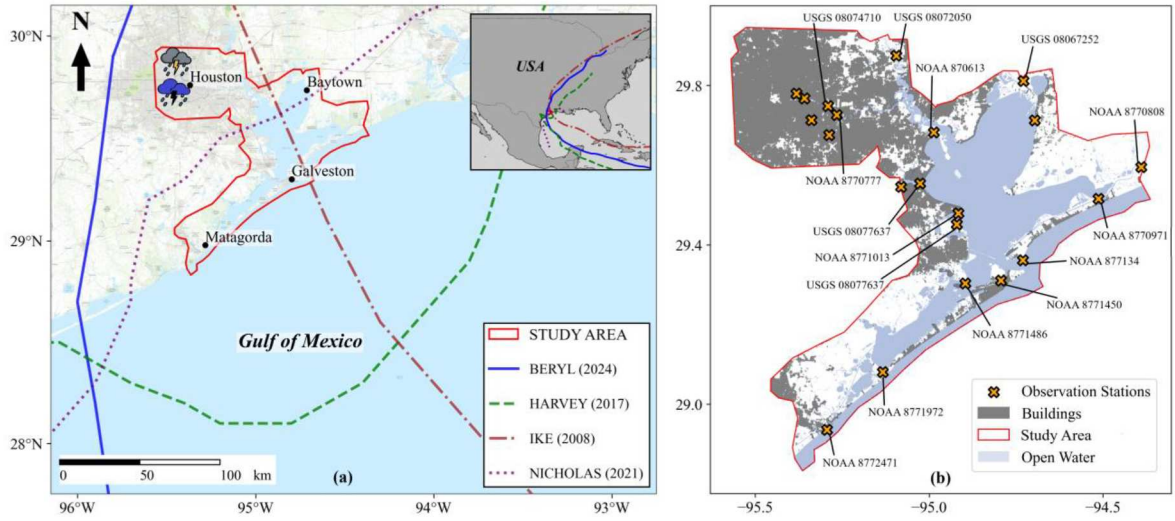


Fig. 1. (a) Storm tracks of Hurricanes Ike, Harvey, Nicholas, and Beryl making landfall near the Galveston Bay within the Gulf of Mexico, USA. Torrential rainfall-runoff events occurred around Memorial Day, 2015; Tax Day, 2016, at the northwestern side of Houston. (b) List of observational water level stations at Galveston Bay.

flood dynamics during extreme events. Both the DL and the Delft3D-FM models utilize the same input features: observed water levels and river discharge at specific stations, along with spatially distributed variables, such as precipitation, digital elevation model (DEM), atmospheric pressure, and wind speed (Fig. 2). These input features capture key spatiotemporal patterns from variables driving compound flood dynamics, enabling both models to predict flood depth and inundation extent. Since the DL model aims to replicate the model predictions of Delft3D-FM, the simulated water depth data of the latter model is the target feature of the proposed framework during training. The framework employs the LSTM model to analyze water level data from multiple tide-gauge stations (Fig. 2a and b). Through an attention mechanism, which is a means to solve the problem of information overload (Niu et al., 2021), the model identifies critical time steps that reflect complex flood dynamics across the area. The attention vectors generated by this step are then integrated into the ConvLSTM model, which captures spatially distributed flood patterns, allowing for nuanced time-dependent interactions across the domain. Enhanced by its own attention mechanism, the ConvLSTM model learns both the unique contributions of individual spatial features and their interactions over time and space (Fig. 2c and d). The attention vectors from the stations are then combined with the ConvLSTM output, in an attempt for the framework to improve flood predictions (Fig. 2b and d). The framework's final output is the spatiotemporal evolution of flood depth and inundation extent (Fig. 2e).

3.1. Physically-based model

We leverage a previously developed Delft3D-FM model of G-Bay that simulates compound flood dynamics associated with hurricane events (Muñoz et al., 2024).

3.1.1. Model setup

The model employs an unstructured finite-volume mesh grid composed of triangular cells that vary in size. The grid resolution ranges from 3 km at the ocean boundary to 5 m within Harris County, enabling high-resolution simulation of the complex interactions between G-Bay's physical morphologic and hydrodynamic processes. The model is driven by multiple external forcing conditions, including tidal harmonic constituents at the open ocean boundary, derived from the TPXO 8.0 global inverse tide model; hourly riverine inflows from USGS river gauges; and hourly wind speed and atmospheric pressure data from the ERA5 reanalysis dataset with a spatial resolution of ~31 km. Since rainfall data is a key flood driver in inland areas, we interpolated these data using a dense network of rain gauges in Harris County. Following Sebastian et al. (2021) and Muñoz et al. (2024), we set the "inverse distance weight" as the interpolation method in ArcGIS with an output cell size of 1 km (e.g., shortest Euclidean distance between existing rain gauges), a search radius of 5 points, and a power function of 2. Moreover, we supplemented rain gauge data with "total precipitation" from ERA5 to estimate rainfall patterns in coastal areas beyond Harris County and over the Gulf of Mexico.

3.1.2. Model calibration and validation

The G-Bay model has been rigorously calibrated and validated using all the above-mentioned compound flood events as shown in Muñoz et al. (2024) as well as Figs. S3–S6 in the supplementary material. Legacy DEMs, including the 2006 Galveston, Texas Coastal Digital Elevation Model and the Continuously Updated Digital Elevation Model (CUDEM), are used to simulate the conditions during past and recent hurricane and storm events. Additionally, the National Land Cover Database (NLCD) provides annual land cover maps at a 30-m resolution, which are crucial for inferring spatially distributed roughness values during the calibration process. Model calibration involves adjusting the Manning's roughness coefficient for different land cover types to ensure that simulated water levels closely align with those of observed data. To further optimize the model's performance, 50 ensemble simulations were conducted using high-performance computing resources to find the optimal combination of roughness coefficients that minimize errors of simulated water level and inundation depth in terms of RMSE, KGE, and NSE altogether. Observed water levels and ground-truth inundation depths were obtained from NOAA's Tide & Currents and high-water marks from the USGS's Flood Event Viewer. The ensemble members consisted of a unique combination of plausible roughness values according to the NLCD land cover class. Such combinations were obtained from the Latin Hypercube Sampling technique and considered for the compound flood events (Helton and Davis, 2003; Muñoz et al., 2022). This comprehensive calibration ensures that the Delft3D-FM model accurately reproduces water levels, inundation depth, and flood extent in G-Bay under a variety of scenarios.

3.2. Deep learning model architecture

The proposed DL framework integrates LSTM and ConvLSTM to

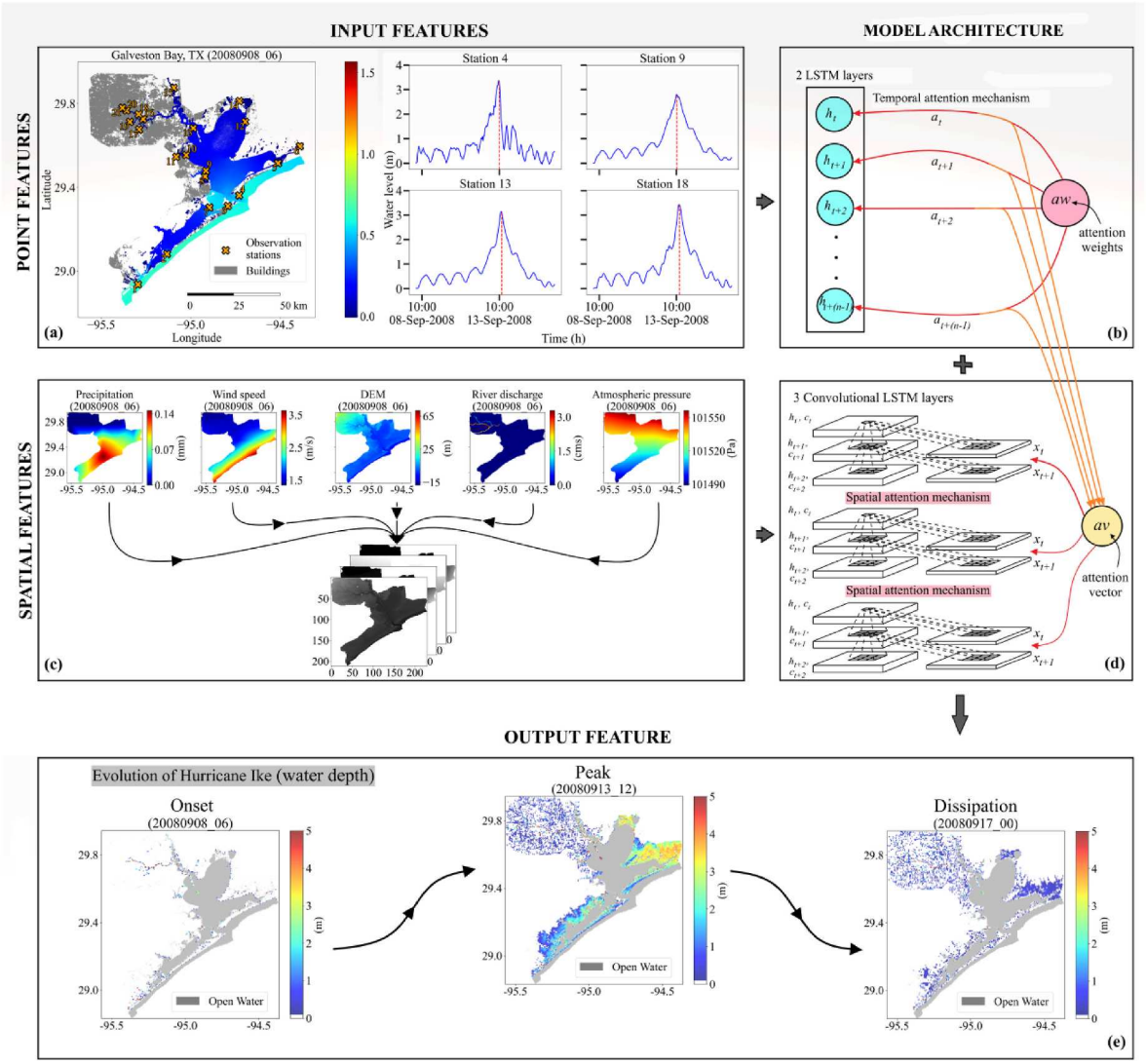


Fig. 2. Overview of the DL framework for flood prediction using multi-modal inputs and multi-model NN architecture. (a) The map of Galveston Bay, TX shows observation stations that provide water level input to the LSTM model. (b) The LSTM layers compute attention weights for each station which are then converted to attention vectors to guide the ConvLSTM model's predictions. (c) Gridded data of precipitation, wind speed, DEM, river discharge and atmospheric pressure are processed as inputs for the ConvLSTM model. (e) The model predicts compound flood dynamics in terms of flood extent and inundation depth.

process point-based and spatially distributed data, respectively. These different data types are initially processed separately to allow the combined model to exert varying localized influences across broader spatial features. The proposed model is trained and validated on sequential and representative flooding events including Hurricane Ike (2008), Memorial Day (2015), and Tax Day (2016). Both point and spatial data of these events are processed in a 6-h sequence to effectively learn patterns over meaningful and practical time intervals without adding excessive complexity in the model. In this regard, the U.S. National Hurricane Center provides forecasts every 6 h that are eventually used to run physically-based models. We use a Bayesian optimization to first select the best combination of hyperparameters after 50 tuning trials, while the model training for the best model is configured for up to 300 epochs. We utilize the Adam optimizer with mean squared error as the loss function. The model validation is monitored with two critical callbacks (model checkpoint and early stopping). The model checkpoint saves the best model whenever the validation loss improves, ensuring that the final model retains the best weights achieved during training. Early stopping halts the training process if the validation loss does not improve after 10 consecutive epochs, preventing overfitting and ensuring that the best-performing model is saved. L2 regularization is also incorporated to prevent overfitting by adding a penalty to the loss function based on the magnitude of the model's weights (Xie et al., 2022).

3.2.1. *Point feature processing*

Hourly water level data from 21 observation stations provide localized temporal dynamics of the extreme flood events. The time-indexed water level data are loaded into structured arrays, with each array corresponding to a different observation station. For each station *i*, the time series data is processed through two sequential LSTM layers, with an attention mechanism that captures unique temporal patterns and computes the relative importance of each timestep at the observation stations (Chaudhari et al., 2021).

i. ***LSTM and hidden states:*** The first LSTM layer processes the input sequence, generating hidden states for each timestep (Equation (1)). These hidden states serve as input to the second LSTM layer, which produces a final hidden state that encapsulates the model's comprehensive understanding of the entire sequence (Equation (2)) (Chaudhari et al., 2021). The hidden states at each timestep for both layers are computed as follows:

$$h_{t,i}^{(1)} = LSTM^{(1)}\left(h_{t-1,i}, x_{t,i}\right) \tag{1}$$

$$h_{T,i}^{(2)} = LSTM^{(2)}\left(h_{T,i}^{(1)}\right) \tag{2}$$

where $h_{t-1,i}$ is the hidden state from the previous timestep, $x_{t,i}$ is the input (water level data) at timestep t, $h_{t,i}$ is the hidden state at the current timestep, and $h_{T,i}^{(2)}$ is the final hidden state of the second LSTM layer for station i.

ii. ***Temporal attention mechanism***: A unique custom attention mechanism is applied to each observation station. For each station i, the attention score $e_{t,i}$ at timestep t is computed using Equation (3):

$$e_{t,i} = \tanh\left(h_{t,i}^{(1)} \cdot W + b\right) \tag{3}$$

where W and b are learnable parameters within the attention mechanism.

Rather than performing a simple dot product between the hidden states of LSTM layers, the attention mechanism increases the attention score of the highest attention weights to enhance the model's focus on the most critical timesteps (e.g., peak water level). The attention weights $aw_{t,i}$, are calculated by applying a softmax function to the attention scores. This is then followed by amplifying the influence of the top 10 % of the attention scores using an emphasis factor (Daramola et al., 2025) (Equation (4)):

$$aw_{t,i} = \begin{cases} \dfrac{exp\ (e_{t,i})}{\sum_{k=1}^{T} exp\ (e_{k,i})} \times emphasis\_factor & if\ t \in top\ 10\% \\ \dfrac{exp\ (e_{t,i})}{\sum_{k=1}^{T} exp\ (e_{k,i})} & otherwise \end{cases} \tag{4}$$

Both the percentage and emphasis factor are iteratively explored until the best model performance is achieved. The process improves the model's ability to accurately capture magnitude and timing of extreme water levels (Daramola et al., 2025). The attention weights $aw_{t,i}$ are utilized to compute the attention vector av_i in Equation (5). This vector represents a weighted sum of the LSTM hidden states across all timesteps (Chaudhari et al., 2021):

$$av_i = \sum_{t=1}^{T} aw_{t,i} \cdot h_{t,i}^{(1)} \tag{5}$$

Since each hidden state is a vector, the attention vector will have the same dimensionality as the individual hidden states at all stations. Finally, the attention vectors are extracted and used to modulate the output of the spatial features processed by the ConvLSTM. It is important to note that each attention vector is not a single, fixed value applied across the entire modeling period. Instead, it dynamically represents the importance of each timestep within a sequence of data, capturing the evolving conditions of the event.

3.2.2. Spatial feature processing

Spatial features, including atmospheric pressure, wind speed, DEM, precipitation, and water depth, are processed as 50-m resolution GeoTIFF images at hourly timesteps. The DEM, which remains static for the duration of each event (e.g., often 1-week), is replicated to match the temporal resolution of the other features. These spatial data are loaded into structured arrays while maintaining the spatial coordinate reference system (CRS) and transformation metadata, which are critical for conducting a geospatial analysis. To address the varying scale among features, *minmaxscaler* normalization is applied. Subsequently, the input data for the ConvLSTM is stacked along the channel dimension, where each channel corresponds to a spatial feature (i.e., atmospheric pressure, wind speed, DEM, precipitation), with water depth designated as the target or predicted variable. A mask is used to track invalid cells containing NaN values within the spatial feature map. This mask is utilized in a custom loss function and during the spatial attention mechanism to ensure that missing data do not adversely affect the model's training (Fig. S2 in supplementary material). After the *minmaxscaler* normalization, the values of valid cells across all features are scaled between 0.1 and 1.0, with all NaN cells set to 0. In the denormalization process, the predicted output is similarly masked, restoring the NaN values to ensure consistency with the original data structure.

i. ***Convolutional and recurrent operation***: At each timestep t, the ConvLSTM applies convolutional filters to the input spatial data to extract local spatial features (Equation (6)), using a kernel size of 3 × 3 pixels (or 150 × 150 m). The ConvLSTM updates its hidden states over time, like those of the LSTM (Equation (7)), but it also processes spatial information through convolutions (Gavahi et al., 2021, 2023). The convolution operation involves sliding a kernel (or filter) over the input feature maps to produce a new output (Muñoz et al., 2021). Specifically, we use ConvLSTM consisting of three layers for balance of complexity and efficiency (Table 1). Note that adding too many layers can lead to overfitting or excessive computational cost. Layer normalization is applied after the first two ConvLSTM layers to stabilize the training process (Equation (8)). Unlike batch normalization, it operates independently of batch size by normalizing across the features of a single timestep, making it more robust where memory constraints result into a smaller batch size. By treating each timestep independently, layer normalization ensures that the model effectively learns from the unique characteristics of each timestep.

$$X_t^{(l)} = Conv\left(X_t^{(l-1)}; W^{(l)}, b^{(l)}\right) \tag{6}$$

$$h_t^{(l)} = ConvLSTM\left(h_{t-1}^{(l)}, X_t^{(l-1)}; W^{(l)}, b^{(l)}\right) \tag{7}$$

$$\widehat{h}_t^{(l)} = \frac{h_{t-1}^{(l)} - \mu}{\sqrt{\sigma^2 + \epsilon}} \cdot \gamma + \beta \tag{8}$$

where $X_t^{(l)}$ is the output feature map of the l-th layer at timestep t, $h_t^{(l)}$ and $\widehat{h}_t^{(l)}$ are the unnormalized and normalized hidden state at timestep t in layer l, $W^{(l)}$ and $b^{(l)}$ are the filter weights and biases of the l-th convolutional layer, μ and σ^2 are the mean and variance of the activations in layer l, γ and β are learnable scaling and shifting parameters, and ϵ is a small constant for numerical stability.

ii. ***Spatial attention mechanism:*** To enhance the model's focus on flooded regions, a Convolutional Block Attention Module (CBAM) (Woo et al., 2018) is integrated after each of the first two ConvLSTM layers. The CBAM applies both channel (feature) and spatial attention modules. First, the channel attention module enables the model to emphasize relevant features effectively (see supplementary material). Then, the spatial attention module emphasizes relevant spatial locations by computing attention weights $\alpha_t(x, y)$ for each spatial location at timestep t (Equation (9)), which is used to modulate the output feature map (Equation (10)). The attention-modulated feature maps from both CBAM modules are

Table 1
Hyperparameter tuning configurations.

Hyperparameters	Range of tested values
ConvLSTM Filters (3 layers)	16–64 (step of 16)
LSTM Units (3 layers)	32–64 (step of 16)
Learning Rate (Adam Optimizer)	1e-6 – 1e-3 (log sampling)
L2 regularization	1e-6 – 1e-3 (log sampling)
Dense (1 layer)	32–64 (step of 16)
Dropout Rate	0.2–0.5 (step of 0.1)

averaged to produce the final modulated feature map (Equation (11)).

$$\alpha_t(x,y) = sigmoid(W^* X_t[x,y] + b) \tag{9}$$

$$X_t^{(l),att}(x,y) = \alpha_t(x,y) \cdot X_t^{channel\_att}(x,y) \tag{10}$$

$$X_{att}(x,y) = \frac{1}{2}\left(X_t^{(1),att}[x,y] + X_t^{(2),att}[x,y]\right) \tag{11}$$

where $X_t^{channel\_att}$ is the channel-attended feature map at location (x,y), * denotes convolution, $X_t^{(1),att}(x,y)$, and $X_t^{(2),att}(x,y)$ are the outputs from the first and second CBAM modules, respectively.

The ConvLSTM model outputs these spatiotemporal feature maps that encode both the spatial patterns via convolutions as well as temporal dependencies via the recurrent updates. The output from the final ConvLSTM layer (X_T) is flattened into a vector (s) that contains all the spatial-temporal information learned by the ConvLSTM up to the final timestep T.

iii. ***Cluster-based temporal attention modulation:*** The study domain is partitioned into clusters using Voronoi tessellation (Fukami et al., 2021), where the number of clusters (N) corresponds to number of observation stations $(i_1, i_2, \ldots, i_N)$ (Table S1 in the supplementary material). This segmentation enables the model to effectively capture the varying flood dynamics associated with each station and its immediate surrounding region (Fig. 3). This assumption is based on the expectation that cells surrounding the stations within each cluster are likely to experience water level peaks at nearly the same time. For each observation station i, the Voronoi cluster V_i is defined as:

$$V_i = \{p \in \mathbb{R}^2 : \|p - s_i\| \leq \|p - s_j\| \forall i_n \neq i\} \tag{12}$$

where p is a point in the spatial domain, $\mathbb{R}^2$ refers to the 2-dimensional Euclidean space, $\|\cdot\|$ is the Euclidean distance, s_i is the location of station i, and s_j are the locations of other stations.

Binary cluster masks are created to identify the cells belonging to each Voronoi cluster (Equation (13)). Using a mutual exclusivity technique, we ensure that spatial cells are associated with at most one cluster, preventing overlapping influences. These masks are crucial for applying the temporal attention vectors on the spatial feature map in a localized manner during the model training.

$$mask_i(x,y) = \begin{cases} 1 & if\ (x,y)\ belongs\ to\ cluster\ V_i \\ 0 & otherwise \end{cases} \tag{13}$$

iv. ***Applying localized temporal attention through selective assignment:*** After extracting the temporal attention vectors for each station, they are transformed and applied to modulate the spatial feature map (Table S2 in the supplementary material). This approach selectively assigns the attention vectors to modulate their specific corresponding clusters, which could lead to spatially varying localized influences. Each station provides an attention vector, denoted as av_i, which captures key temporal patterns from water level dynamics (Equation (5)). Recall that the individual attention vector is only within a cell in spatial feature map. Hence, each attention vector is passed through a fully connected (dense) layer that projects it into a 1D tensor, $flattened\_vector_i$, having a length equal to the total number of cells in the spatial feature map $(H \times W)$ (Equation (14)). The $flattened\_vector_i$ is then reshaped into a 2D array, $reshaped\_vector_i$, which aligns attention information with the grid-like shape of the spatial map (Equation (15)).

$$flattened\_vector_i = (Dense[H \times W][av_i]) \tag{14}$$

$$reshaped\_vector_i = Reshape(H, W)(flattened\_vector_i) \tag{15}$$

The Voronoi cluster mask, $mask_i(x,y)$, corresponding to a specific spatial region of each station is now used to confine the influence of the attention vector on each $reshaped\_vector_i$ into a $localized\_vector_i$ (Equation (16)). This ensures that the $localized\_vector_i$ only has attention vector within the cluster, effectively masking out areas outside the cluster. As a result, only the values at points (x,y) within each cluster V_i is retained.

$$localized\_vector_i(x,y) = reshaped\_vector_i(x,y) \times mask_i(x,y) \tag{16}$$

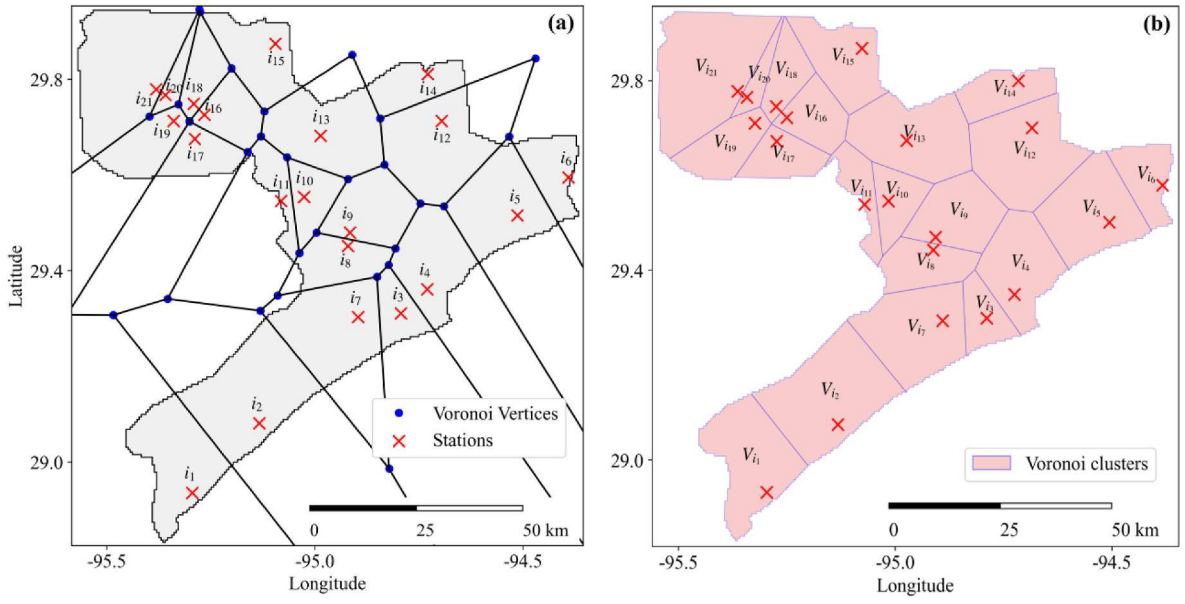


Fig. 3. Cluster generation using Voronoi tessellation. (a) The study domain is divided into clusters based on the available observation stations (red crosses). (b) Voronoi clusters within the model domain (shaded polygons) are then used to highlight the flood event's influence on the specific observation station and its immediate surroundings. The region V_{i_1} represents the set of all cells that are closer to station i_1 than to any other station i_n. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Lastly, a $combined\_vector(x,y)$ is initialized with zero-values to match the shape of the spatial feature map. For each station i, we assign its $localized\_vector_i(x,y)$ to the corresponding cluster locations in $combined\_vector(x,y)$ using a function (Equation (17)). Since the Voronoi clusters are mutually exclusive, each cell in the spatial feature map is influenced by exactly one attention vector.

$$combined\_vector(x,y) = \begin{cases} localized\_vector_i(x,y) & if\ (x,y) \in V_i \\ 0 & otherwise \end{cases} \quad (17)$$

v. ***Modulating the ConvLSTM output:*** The $combined\_vector(x,y)$ is used to modulate the ConvLSTM output s through element-wise multiplication (Equation (18)). This modulation emphasizes or de-emphasizes the ConvLSTM output based on high or low learned attention weights (Chenmin et al., 2024) across the spatial regions, respectively.

$$modulated\_output(x,y) = s(x,y) \times combined\_vector(x,y) \quad (18)$$

The modulated output is flattened and passed through fully connected layers z to generate the final output $\hat{y}$ (Equation (19)), which is then reshaped back to the spatial feature map's dimensions (Equation (20)):

$$z = flatten(modulated\_output[x,y]) \quad (19)$$

$$\hat{y}.final\_output = Reshape([H,W])(z) \quad (20)$$

The proposed framework offers a significant advancement in modeling the spatiotemporal complexities of cyclone-induced flood dynamics, surpassing the capabilities of conventional ConvLSTM models and existing hybrid approaches (Fig. 4). While conventional ConvLSTM models effectively capture general spatiotemporal patterns, they are not suitable to account for the region-specific temporal variability and spatial heterogeneity inherent in flood events. This limitation becomes particularly evident in large and/or diverse regions, such as Galveston Bay, where flood propagation and peak water level timing exhibit substantial variation between coastal, transition, and inland zones. To overcome these shortcomings, our approach incorporates a novel cluster-based temporal attention mechanism. This framework partitions the spatial domain into distinct clusters, assigning localized water level dynamics to each cluster. By aligning temporal observations with region-specific spatial predictions, the model accurately captures the varying timing of peak water levels across extensive areas. Moreover, the proposed framework leverages dynamic inputs that evolve over time to provide comprehensive spatial and temporal information. In contrast, many hybrid models rely heavily on static features, such as elevation, floodplain, land use, etc., to ensure spatial consistency (Farahmand et al., 2023; Muñoz et al., 2024; Valle-Levinson et al., 2020), often at the expense of capturing temporally varying flood behaviors. This innovation enables the model to precisely replicate the varied timing of peak water levels across extensive spatial domains, potentially improving its robustness and predictive accuracy over conventional ConvLSTM and hybrid methods.

3.2.3. Model prediction and analysis

We evaluate the trained model's performance on unseen data to assess its ability to generalize and make accurate predictions of water depth and extent during flooding events. Testing is conducted using process-based model simulations of Hurricane Harvey (2017), Hurricane Nicholas (2021) and Hurricane Beryl (2024). The input and output feature processing follow the same steps as in the model training, ensuring consistency in data handling. The cluster-based training approach and the application of localized attention through selective assignment are already built into the model. Nevertheless, the accuracy of the prediction relies on using the attention model to extract attention vectors from the test data (Equations (21) and (22)). This provides insight into how the model weighs different timesteps from the test events. This cluster-based attention mechanism allows the model to focus on different spatial regions during prediction, providing more refined, localized predictions of inundation depth and flood extent.

$$attention\_vectors = attention\_model\left(water\_level_{test\_sequences}\right) \quad (21)$$

$$\hat{y}_{pred} = best\_model\left(X_{test\_sequences}, water\_level_{test\_sequences}\right) \quad (22)$$

Recommended performance metrics for predicting inundation depth are computed for each flood event, including the Root-Mean Square Error (RMSE), coefficient of determination (R^2), Kling-Gupta Efficiency (KGE) and Nash-Sutcliffe Efficiency (NSE) (Gupta et al., 2009; Mahakur et al., 2025; Nash and Sutcliffe, 1970; Samantaray et al., 2025). In addition, hit rate (H), false alarm ratio (F), critical success index (C), and error bias (E) are used to compute the accuracy of flood map replication (Wing et al., 2022). H measures the ratio of correctly predicted flood instances (true positives) to the total number of actual flood instances,

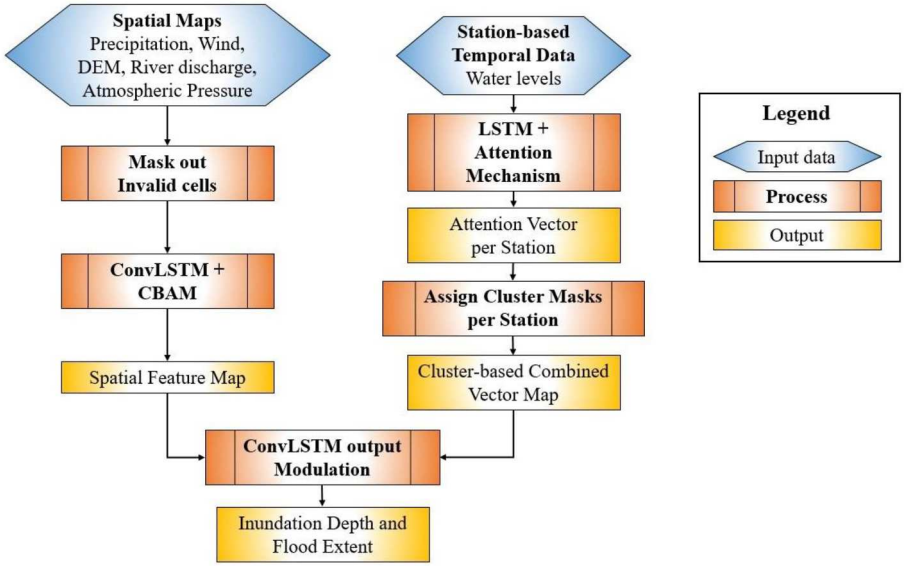


Fig. 4. Schematic illustrating the flow from input features through both the temporal and spatial feature processing, attention vector extraction, cluster-based modulation, and the final modulated ConvLSTM output.

indicating the model's ability to identify flooding when they occur. F indicates the proportion of false positives among all instances where the model predicted a flood, measuring how often the model incorrectly predicts flooding when there is none. C assesses the proportion of correctly predicted flood events by considering all hits, false alarms, and misses, offering a comprehensive measure of the model's accuracy in predicting floods. E reflects the ratio of predicted flood instances (both true positives and false positives) to actual flood instances (both true positives and false negatives), revealing whether the model has a tendency to overpredict or underpredict flooding. Best results are closer to 1 except for F, which should be closer to 0. We calculated the residuals by taking the direct difference between the actual and predicted flood maps of the water depth to evaluate prediction errors.

4. Results and discussion

4.1. Hyperparameter tuning process for the best model

The Bayesian optimization process systematically searches the hyperparameter space across 50 trials, successfully identifying configurations that minimize the validation loss. This approach iteratively tunes the hyperparameter values based on prior trial outcomes. For each trial, the optimal model is derived using the Adam optimizer with mean squared error as the loss function, achieving convergence within the 300-epoch limit. The model checkpointing callback saves the best model weights based on the lowest validation loss achieved during training, and the early stopping callback terminates training after 10 consecutive epochs without improvement in validation loss. This approach resulted in models that not only achieved the lowest validation loss per trial but also demonstrated strong generalization to unseen data, effectively capturing the underlying patterns in the training set. The consistent improvement in validation loss across trials demonstrates the effectiveness of this method in identifying a robust model configuration. Fig. 5 illustrates the convergence of the optimization process, showcasing the performance of the optimal hyperparameters (i.e., models) obtained across all 50 trials.

Most trials exhibit satisfactory training and validation loss curves, with rapid decreases during the initial epochs that stabilize near-zero around 50 epochs (Fig. 4a and b). Generally, trials with steeper initial declines and lower asymptotic losses tend to converge better and may have less risk of overfitting (Singh et al., 2024). Most trials also stop early, around 40 to 50 epochs (Fig. 5c), suggesting they reach optimal validation loss relatively quickly, though a few continue up to 200 and 300 epochs, indicating more complex convergence patterns. The validation losses of most trials fall within a narrow range, with a significant concentration between 5E-5 and 1E-4 m^2 (Fig. 5d). The trial with the lowest validation loss (Trial 11) is achieved after 50 epochs (Fig. 5e). In this trial, the training-to-validation loss ratio initially oscillates above 1 in the early epochs (Fig. 5f), signaling some expected early divergence between the losses. However, this ratio stabilizes around epoch 10, trending closer to 1, which suggests a balanced performance between training and validation. Therefore, all trial ratios trending towards 1 underscore the model's consistency in generalizing across datasets (Fig. 5f), while ratios below 1 indicate overfitting. Finally, we apply all models and evaluate their ability to accurately capture flood events.

4.2. Attention modulation

The attention vectors extracted by the LSTM inform how well the model integrates the influence of water level variability in predicting flood dynamics across the study area. Timesteps of high-water level variability (or abrupt changes) receive higher weights, indicating potential flooding around the observation stations, while lower magnitudes suggest reduced flood risk (Fig. 6). Based on the modulation approach, attention weights from individual stations are projected over the clusters containing those stations, allowing distinct and varying influences that correlate with the station's water levels. For instance, at timestep 130 h, a higher attention weight is applied to the cluster associated with station 4 (Galveston Bay Entrance, TX – NOAA 8771341) since the peak water level for Hurricane Ike occurs approximately 5 h earlier than at station 18 (Manchester, TX – NOAA 8770777) (Fig. 6a). However, attention weights for both station clusters at timestep 188 h during the Memorial Day flooding are quite similar, reflecting the relatively low water levels at that time (Fig. 6b). This approach effectively accounts for lag-times in flood dynamics across the larger domain, enhancing the model's ability to predict regional flood variations accurately. It is important to note that the spatiotemporal analysis is conducted at the cell-size scale using the ConvLSTM model. Furthermore, the attention vector does not imply uniform flooding across the entire cluster; rather, its influence is applied to the spatially varied

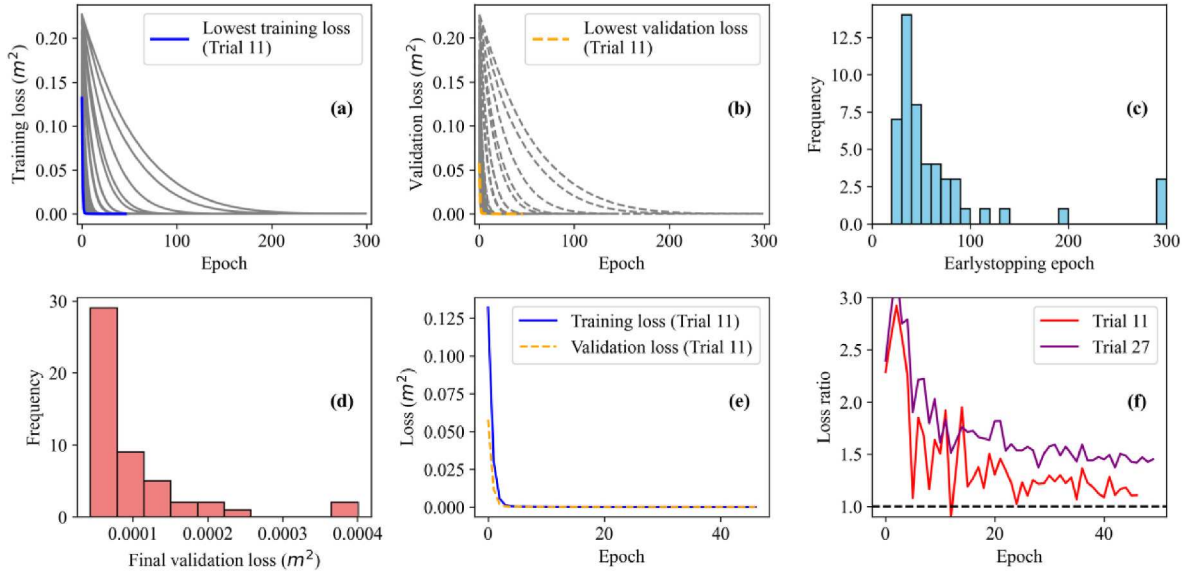


Fig. 5. Summary of the hyperparameter tuning process across 50 trials analyzing (a) training loss across epoch, (b) validation loss across epochs, (c) early stopping epoch frequency, (d) final validation loss frequency in terms of mean square error, (e) training loss and validation loss of the best trial, and (f) loss ratio across epochs of the best two trials.

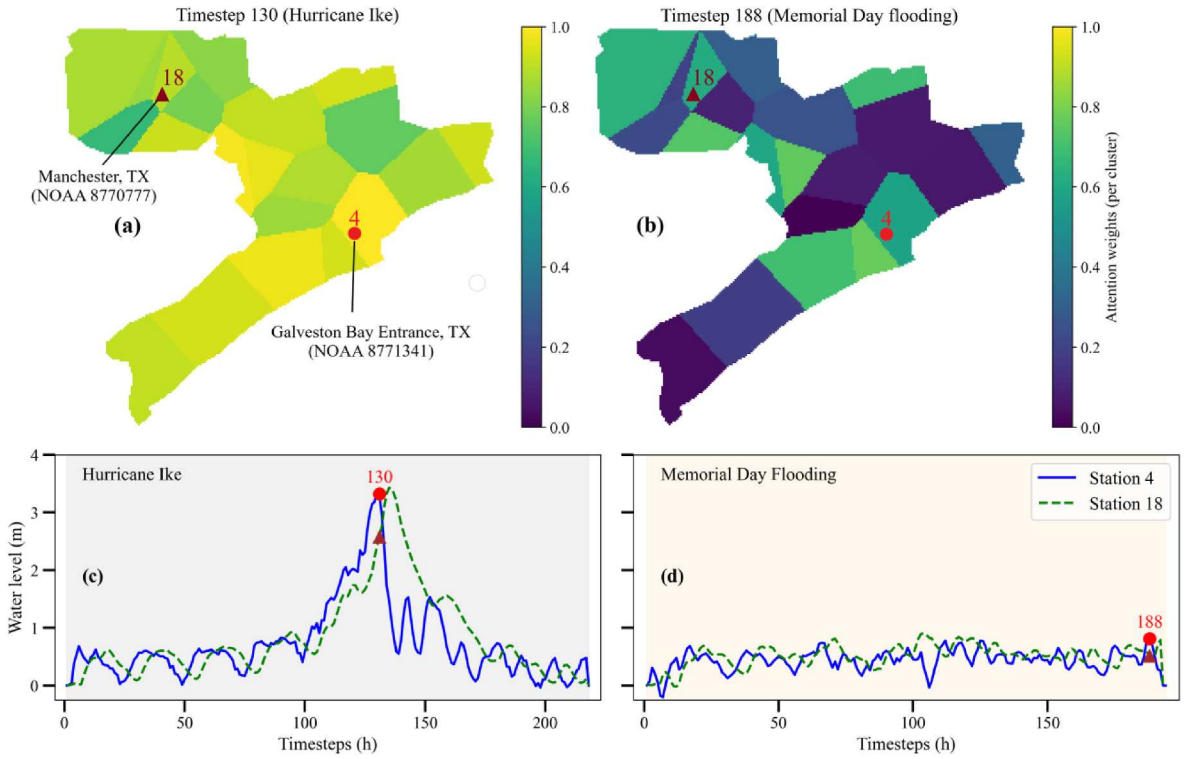


Fig. 6. Temporal attention vector maps for two selected training timesteps reflecting how water level variability contributes to flood prediction.

values per timestep already estimated at individual cells by the ConvLSTM model. This means that each individual cell in the spatial map has a calculated value that varies across cells and at each timestep. The attention vector, although uniform over its cluster, emphasizes these varying values per cell with information about the events' evolution.

In addition, we compare the proposed framework with a model applying an attention vector to the entire domain to highlight its ability to accurately account for the timing differences in flood dynamics over a large domain. The single attention vector reflects the combined influence of all observation stations. The cell with the highest actual peak water depth across all timesteps within each cluster is selected and compared with its predicted data to observe any lag- or lead-times. Results of this analysis show a significant lag-time of about 5 h for the majority of clusters and a lead-time of about 2 h for the clusters when a single attention vector is used. However, correct peak timing is observed in two clusters without a color shade (Fig. 7a), suggesting where the maximum attention weight aligns with the peak flood timing. This indicates that averaging attention across stations may only capture flood dynamics correctly in clusters where timing coincides by chance. In contrast, incorporating cluster-based application of attention vectors demonstrates a significantly improved match with the actual flood timing, with only 5 clusters within 1 h lead- and lag-times (Fig. 7b).

In a previous study, we demonstrated that deep learning models can

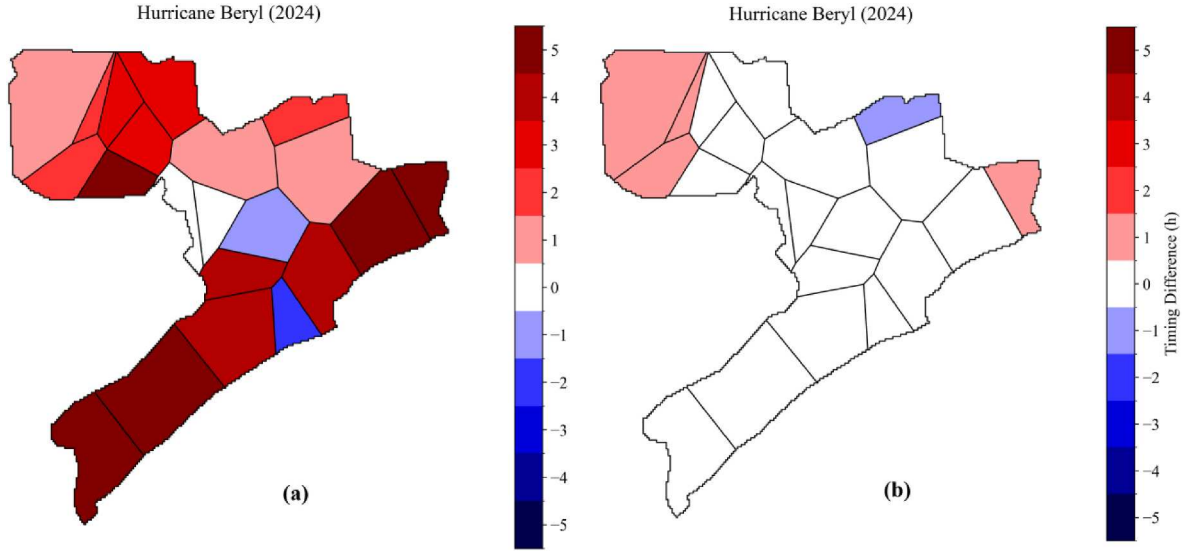


Fig. 7. Timing difference between actual and predicted peak water depth at each cluster. Positive and negative values indicate lead- and lag-time (a) without cluster-based application, and (b) with cluster-based application of the attention vector.

be significantly enhanced by incorporating an attention mechanism (Daramola et al., 2025). This approach helps emphasize indicators that signal the presence and severity of flooding events, rather than relying solely on repetitive patterns in training datasets or overlapping characteristics among various flood drivers. Consequently, attention mechanisms were integrated into the architecture of the proposed model in this study to ensure its effectiveness during both the training and testing phases. Given the model's satisfactory performance, this technique is shown to bolster its ability to effectively identify flood patterns across space and time. This improvement enables the proposed model to generalize effectively, indicating that the framework is robust and can be applied to other regions with varying distributions of observation stations or diverse hydrodynamic conditions.

4.3. Deep learning predictions and process-based model simulations

Most of the DL models indicate that the proposed framework achieves satisfactory accuracy in predicting flood dynamics in terms of flood extent and inundation depth (Fig. S4 in the supplementary material). To showcase the model's ability, we present its performance for predicting the peak flood depth maps of the three events. For Hurricane Beryl (Fig. 8), the peak flood depth occurred on July 8, 2024, corresponding to the timestep with the maximum RMSE value. The best model slightly overpredicts flood depth in coastal areas and underpredicts in inland areas, resulting in an overall slight underprediction with an error bias (E) lesser than 1. The false alarm ratio (F) identifies only 14 % of non-flooded areas are falsely classified as flooded, with a critical success index (C) over 70 % of flood maps (percentage of accurately predicted depth and extent). The model also achieved a high accuracy in identifying flooded areas, with a hit rate (H) of 80 %. For the evolution of flood depth in selected areas, the model demonstrates satisfactory performance across all timesteps with KGE, NSE, and R^2 metrics over 0.70.

At the peak flooding on August 29, 2017, for Hurricane Harvey, the best model tends towards slight overprediction of the flood map, with an error bias (E) greater than 1 (Fig. S10a and b, supplementary material), especially along the coast. However, flood depth is underestimated in the inland areas (Fig. S10c). Similar to Hurricane Beryl, the flood map has the maximum RMSE value. Nevertheless, C is over 60 %, indicating that most of the flood map's depth and extent are accurately predicted. Additionally, over 80 % (H) of the flooded areas are correctly identified. F is below 30 %, which means that the model is less likely to classify non-flooded areas as flooded. Similarly, the model demonstrates satisfactory performance across all timesteps with KGE, NSE metrics over 0.70 for the evolution of flood depth at selected areas. Just like the first two results, the time of peak flooding (on September 14, 2021) corresponds to the timestep with the maximum RSME for Hurricane Nicholas. The best model tends to slightly underestimate flooding in the inland region while overestimating flood depth along the coast especially at the eastern area of the model domain (Fig. S11a and b, supplementary material). The model accurately predicts most of flood map's depth and extent with a C of 64 %, while correctly identifying flooded areas with 80 % accuracy (H). F is below 25 %, i.e., low tendency to incorrectly classify non-flooded areas as flooded.

Regarding the peak flood maps in Fig. 7, S10 and S11, the underestimation in inland areas is due to the omission of many interconnected rivers that could contribute to flooding during these events. In contrast, the overestimation in coastal areas is caused by high attention vectors projected onto the clusters, as the associated stations are exposed to the ocean. Nevertheless, the model performances metrics are satisfactory for all flood events and across all clusters (Fig. 9, S12 and S13).

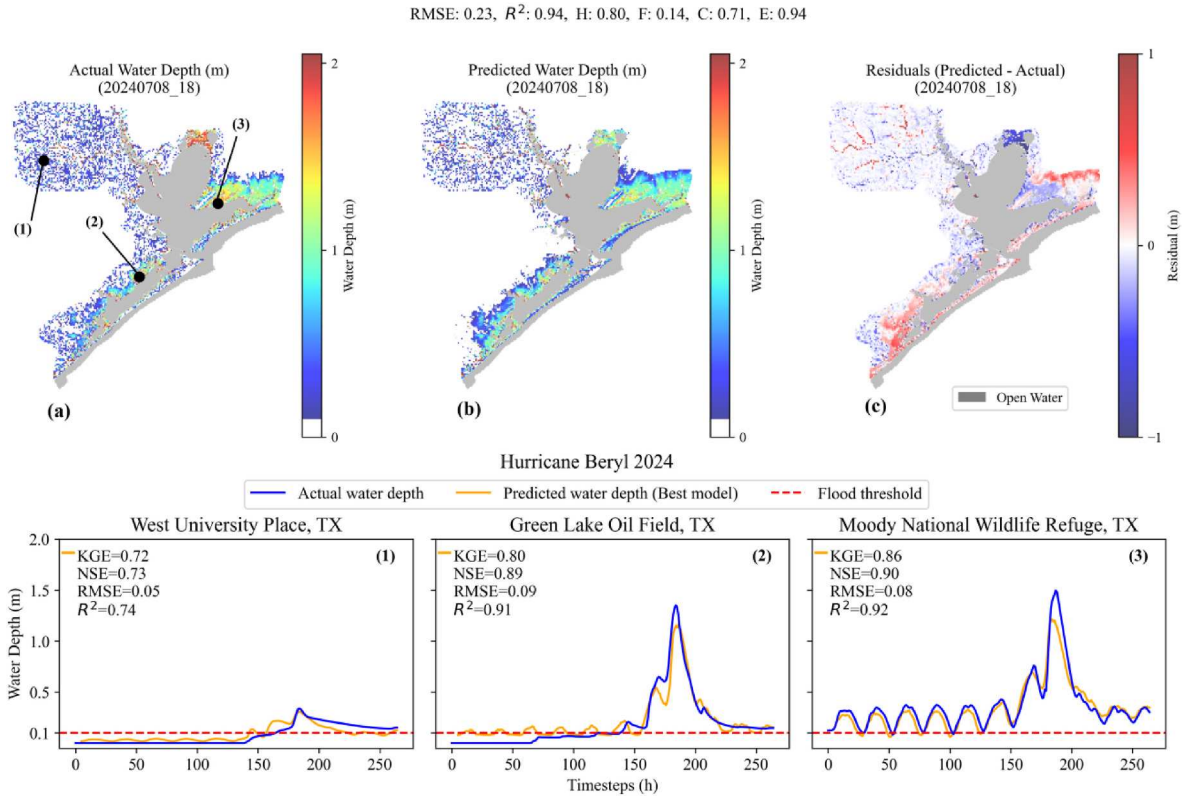


Fig. 8. Predicted water depth for Hurricane Beryl (2024) across Galveston Bay, Texas with a flood threshold of 0.10 m. (a) Actual water depth on July 8, 2017, with three marked observed flooded locations. (b) Predicted water depth from the best-performing model for the same flood map, showing a close match with actual water depths. (c) Residuals of actual and predicted water depth, indicating spatial variations in prediction accuracy. Positive (red) and negative (blue) areas represent overprediction and underprediction, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4.4. Limitation and future work

The model utilizes a physically-based model simulation as the "ground truth" flood extent and inundation depth during the compound flood events. However, these model simulations slightly differ from actual observations introducing errors despite the thorough model calibration (Muñoz et al., 2024). The Delft3D-FM model is subject to several sources of uncertainty, including the initial condition, the forcing (or boundary) conditions, model parameters, and model structure (Abbaszadeh et al., 2022; Muñoz et al., 2022). Additionally, there may exist interpolation errors when resampling the spatially varying cell sizes of Delft3D-FM to a uniform 50-m DL model input's cell size. The resampling procedure was needed to enable correct model training and validation of the CNN architecture. Next, river discharges recorded at upstream river-gauge stations are applied along the river branches and used as an input feature in the training process. Ignoring the lag-time in peak flow that may occur along the river branches might contribute to prediction errors, especially in large river deltas, estuaries, and bays. These uncertainties in the Delft3D-FM outputs can propagate to the DL model and affect its performance evaluation due to inherited biases, evident by the slight over- and underestimation of the comparison plots (Figs. S3–S6 in the supplementary material).

While the cluster-based attention vector for water level variability can account for the lag-time, some clusters may encompass areas with slight time variations, potentially affecting the model's accuracy. Nevertheless, all cells contained in each cluster are still spatially varying in magnitude, even though they follow the same timing pattern for the attention vector modulation during an event's evolution. When the average timing of all cells within the cluster is estimated, all the clusters have lead- and lag-times between 0 to approximately 2 h with respect to the actual peak timing. In other words, the more observation points for water levels, the more spatial timing variation will be captured; leading to a substantial improvement of the model's accuracy. Although incorporating attention mechanisms emphasizes certain weights and enhances generalization, some models produce straight-line predictions across all time steps due to L2 regularization effects in the architecture that penalize large weights (Figs. S7–S9). The latter occurs because a higher regularization coefficient can overly smooth the attention emphasis during predictions, preventing the model from capturing rapid fluctuations or noise in the data. Therefore, it is crucial to balance the strength of regularization technique within the architecture to enhance generalization without inducing underfitting. To further reduce the error between model predictions and actual observations, future work will focus on introducing residual learning techniques (Tedesco et al., 2024; Zou et al., 2023). Residual learning enables the model to concentrate on learning the differences or residuals between its predictions and the true observations rather than attempting to model the entire mapping directly.

The performance metrics used in this study also have limitations. For example, RMSE is highly sensitive to large errors, so extreme flood events can disproportionately inflate its value, potentially skewing perceptions of the model's overall performance. R^2 indicates explained variance but does not account for prediction bias, which is crucial for accurate flood depth estimation. KGE and NSE, which are standard metrics in hydrology, help capture these biases during extreme events. Metrics like H, F, and CSI effectively assess flood extent but may overlook the magnitude of depth predictions. E, however, is particularly significant for extreme events because it captures even small systematic over- and underpredictions. Despite their limitations, these metrics are used in a complementary manner and remain the most valid metrics for flood analysis. Future studies could consider comparison with other *state-of-the-art* DL architectures and explore additional metrics or weighting schemes that prioritize performance during extreme conditions. The latter will ensure a balanced evaluation across all flood scenarios.

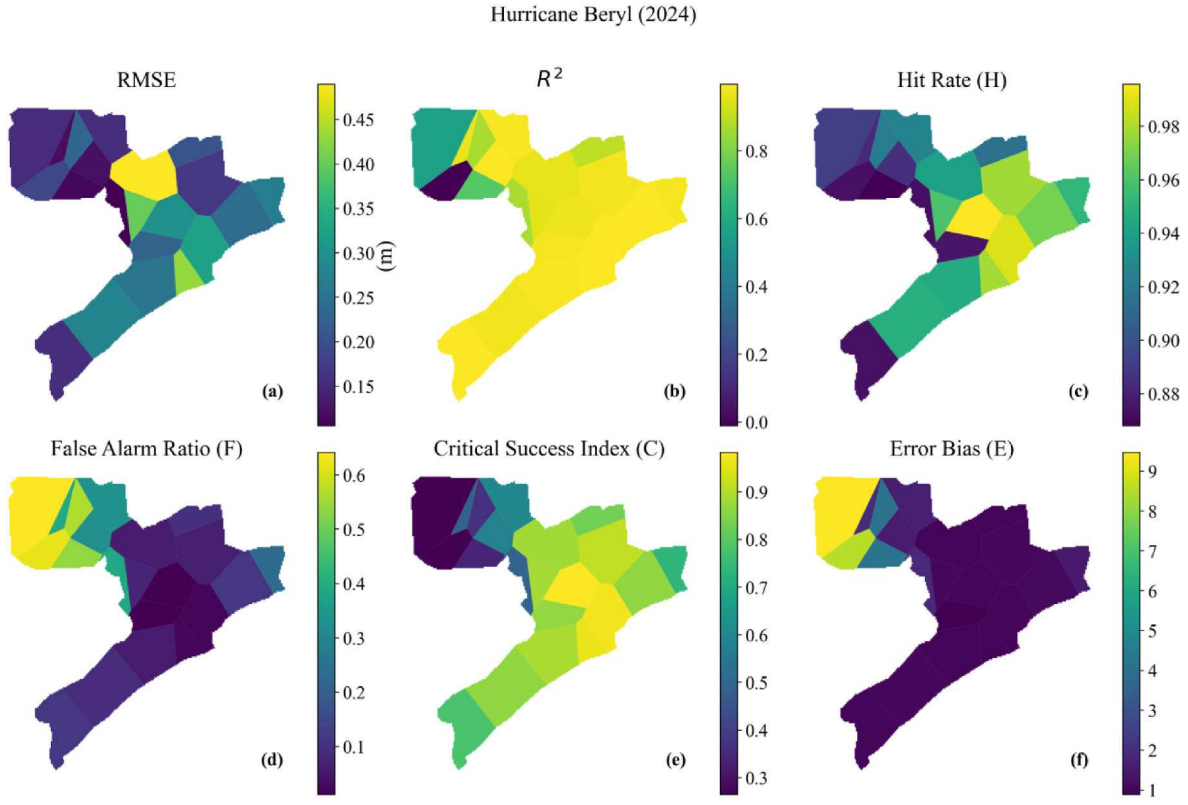


Fig. 9. Spatial distribution of flood prediction metrics for Hurricane Beryl (2024). Each map illustrates the model's performance across various regions based on the cluster-based application. Optimal scores of RMSE, R^2, H, F, C, and E are 0 m, 1, 1, 0, 1, and 1, respectively.

5. Conclusion

In this study, we developed a DL framework to address the complexities of cyclone-induced compound flood dynamics in Galveston Bay, TX. The framework combined LSTM and ConvLSTM model architectures, integrating spatial and temporal attention mechanisms to enhance the model's ability to capture nonlinear associations among flood drivers such as precipitation, river discharge, and storm surge. By leveraging output data from a coastal hydrodynamic model (Delft3D-FM) and incorporating cluster-based attention vectors, the framework achieved high accuracy in replicating flood dynamics in terms of inundation depth and flood extent. Importantly, it accounts for the hysteresis in system behavior, leading accurate timing of flood dynamics, particularly the propagation of floods across a large domain, caused by extreme cyclonic events. We also show the capability of the DL framework to handle the spatial variability of dominant flood drivers in coastal (storm surge) and inland areas (rainfall-runoff). This in turn demonstrates the potential of coupling DL architectures to enhance the prediction accuracy of cyclone-induced compound flood dynamics, specifically in complex coastal systems.

The model demonstrates a satisfactory performance in identifying flooded areas, achieving an average hit rate of over 80 % while maintaining average false alarm rate below 25 %, even with the large domain size and resolution. Average KGE and NSE for the prediction of all events evolution are above 0.7. The study demonstrates that applying cluster-based temporal attention allows the model to focus selectively on different spatial regions, becoming more effective at capturing localized timing variations in flood dynamics across a large domain. Despite these performances, incorporating more water level observation points and discharge observations along river channels could enhance spatial variability and improve the model's accuracy. The proposed coupled DL model holds promise for disaster preparedness and flood mitigation in vulnerable coastal regions, offering a scalable approach adaptable to other cyclone-prone areas.

CRediT authorship contribution statement

Samuel Daramola: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **David F. Muñoz:** Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Hamed Moftakhari:** Writing – review & editing, Supervision. **Hamid Moradkhani:** Writing – review & editing, Supervision.

Software

Delft3D -FM 2021.03 (Physically-based modeling) is an open-source software that can be accessed at https://oss.deltares.nl/web/delft3d/downloads.

Linux OS, NVIDIA A100 GPU (Machine learning)

Python scripts: The codes that support this study are publicly available at https://github.com/CoRAL-Lab-VT/FloodDepthDL.git.

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.

Acknowledgement

The authors would like to thank financial support from the National Science Foundation via awards: 480948 and 2223893.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2025.106499.

Data availability statement

All data used in this study are hourly time series, except for the static Digital Elevation Models (DEMs), and all are publicly available. Specifically, the legacy "Galveston, Texas Coastal Digital Elevation Model" was obtained from NOAA's National Geophysical Data Center (https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id = gov.noaa.ngdc.mgg.dem:403, NOAA-NCEI, 2024), CUDEM was obtained from NOAA's Data Access Viewer (https://coast.noaa.gov/, NOAA-NCEI, 2024), land cover maps were obtained from the Multi-Resolution Land Characteristics Consortium (https://www.mrlc.gov/data, NLCD, 2024), data time series were obtained from NOAA's Tides & Currents portal (https://tidesandcurrents.noaa.gov/, NOAA – Tides and Currents, 2024) and the USGS National Water Dashboard (https://dashboard.waterdata.usgs.gov/app/nwd/en/, USGS-NWD, 2024), barotropic tides were obtained from the TPXO 8.0 global inverse tide model (https://www.tpxo.net/global/tpxo8-atlas, TPXO 8.0, 2024), post-flood high-water marks were obtained from the USGS Flood Event Viewer (https://stn.wim.usgs.gov/fev/, USGS-FEV, 2024), rainfall data were obtained from the rain-gauge network of the Harris County Flood Warning System (https://www.harriscountyfws.org/, Harris County FWS, 2024), and the ERA5 reanalysis dataset was obtained from the European Centre for Medium-Range Weather Forecasts (https://cds.climate.copernicus.eu/#!/home, ERA5, 2024).

References

ABC News, 2019. Remembering the Tax day flood of 2016. https://abc13.com/tax-day-flood-flooding-houston-rain/3361014/ accessed 12.18.24.

Alipour, A., Jafarzadegan, K., Moradkhani, H., 2022. Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping. Environ. Model. Software 152, 105398. https://doi.org/10.1016/j.envsoft.2022.105398.

Ayyad, M., Hajj, M.R., Marsooli, R., 2022. Machine learning-based assessment of storm surge in the New York metropolitan area. Sci. Rep. 12, 19215. https://doi.org/10.1038/s41598-022-23627-6.

Bates, P., 2023. Fundamental limits to flood inundation modelling. Nat Water 1, 566–567. https://doi.org/10.1038/s44221-023-00106-4.

Bates, P.D., 2022. Flood inundation prediction. Annu. Rev. Fluid Mech. 54, 287–315. https://doi.org/10.1146/annurev-fluid-030121-113138.

Bilskie, M.V., Zhao, H., Resio, D., Atkinson, J., Cobell, Z., Hagen, S.C., 2021. Enhancing flood hazard assessments in coastal Louisiana through coupled hydrologic and surge processes. Front. Water 3. https://doi.org/10.3389/frwa.2021.609231.

Bloomfield, H.c., Hillier, J., Griffin, A., Kay, A.l., Shaffrey, Pianosi, F., James, R., Kumar, D., Champion, A., Bates, P.D., 2023. Co-occurring wintertime flooding and extreme wind over Europe, from daily to seasonal timescales. Weather Clim. Extrem. 39, 100550. https://doi.org/10.1016/j.wace.2023.100550.

Boumis, G., Moftakhari, H.R., Moradkhani, H., 2023. Storm surge hazard estimation along the US Gulf Coast: a Bayesian hierarchical approach. Coast. Eng. 185, 104371. https://doi.org/10.1016/j.coastaleng.2023.104371.

Brunner, M.I., Gilleland, E., Wood, A., Swain, D.L., Clark, M., 2020. Spatial dependence of floods shaped by spatiotemporal variations in meteorological and land-surface processes. Geophys. Res. Lett. 47, e2020GL088000. https://doi.org/10.1029/2020GL088000.

Camus, P., Haigh, I.D., Nasr, A.A., Wahl, T., Darby, S.E., Nicholls, R.J., 2021. Regional analysis of multivariate compound coastal flooding potential around Europe and environs: sensitivity analysis and spatial patterns. Nat. Hazards Earth Syst. Sci. 21, 2021–2040. https://doi.org/10.5194/nhess-21-2021-2021.

Chaudhari, S., Mithal, V., Polatkan, G., Ramanath, R., 2021. An attentive survey of attention models. ACM Trans. Intell. Syst. Technol. 12 (53), 1–53. https://doi.org/10.1145/3465055, 32.

Chenmin, N., Marsani, M.F., Pei Shan, F., 2024. Flood prediction based on feature selection and a hybrid deep learning network. J. Water Clim. Change 15, 1245–1261. https://doi.org/10.2166/wcc.2024.559.

Chiang, S., Fu, H.-S., Hsiao, S.-C., Hsiao, Y.-H., Chen, W.-B., 2024. An efficient 2-D flood inundation modelling based on a data-driven approach. J. Hydrol.: Reg. Stud. 52, 101741. https://doi.org/10.1016/j.ejrh.2024.101741.

Daramola, S., Muñoz, D.F., Muñoz, P., Saksena, S., Irish, J., 2025. Predicting the evolution of extreme water levels with long short-term memory station-based approximated models and transfer learning techniques. Water Resour. Res. 61, e2024WR039054. https://doi.org/10.1029/2024WR039054.

Eilander, D., Couasnon, A., Leijnse, T., Ikeuchi, H., Yamazaki, D., Muis, S., Dullaart, J., Haag, A., Winsemius, H.C., Ward, P.J., 2023. A globally applicable framework for compound flood hazard modeling. Nat. Hazards Earth Syst. Sci. 23, 823–846. https://doi.org/10.5194/nhess-23-823-2023.

Farahmand, H., Xu, Y., Mostafavi, A., 2023. A spatial–temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features. Sci. Rep. 13, 6768. https://doi.org/10.1038/s41598-023-32548-x.

Fathi, M.M., Liu, Z., Fernandes, A.M., Hren, M.T., Terry, D.O., Nataraj, C., Smith, V., 2025. Spatiotemporal flood depth and velocity dynamics using a convolutional neural network within a sequential Deep-Learning framework. Environ. Model. Software 185, 106307. https://doi.org/10.1016/j.envsoft.2024.106307.

Foroumandi, E., Gavahi, K., Moradkhani, H., 2024. Generative adversarial network for real-time flash drought monitoring: a deep learning study. Water Resour. Res. 60, e2023WR035600. https://doi.org/10.1029/2023WR035600.

Fraehr, N., Wang, Q.J., Wu, W., Nathan, R., 2022. Upskilling low-fidelity hydrodynamic models of flood inundation through spatial analysis and Gaussian process learning. Water Resour. Res. 58, e2022WR032248. https://doi.org/10.1029/2022WR032248.

Fu, G., Jin, Y., Sun, S., Yuan, Z., Butler, D., 2022. The role of deep learning in urban water management: a critical review. Water Res. 223, 118973. https://doi.org/10.1016/j.watres.2022.118973.

Fukami, K., Maulik, R., Ramachandra, N., Fukagata, K., Taira, K., 2021. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. Nat. Mach. Intell. 3, 945–951. https://doi.org/10.1038/s42256-021-00402-2.

Gavahi, K., Abbaszadeh, P., Moradkhani, H., 2021. DeepYield: a combined convolutional neural network with long short-term memory for crop yield forecasting. Expert Syst. Appl. 184, 115511. https://doi.org/10.1016/j.eswa.2021.115511.

Gavahi, K., Foroumandi, E., Moradkhani, H., 2023. A deep learning-based framework for multi-source precipitation fusion. Rem. Sens. Environ. 295, 113723. https://doi.org/10.1016/j.rse.2023.113723.

Gomez, F.J., Jafarzadegan, K., Moftakhari, H., Moradkhani, H., 2024. Probabilistic flood inundation mapping through copula Bayesian multi-modeling of precipitation products. Nat. Hazards Earth Syst. Sci. 24, 2647–2665. https://doi.org/10.5194/nhess-24-2647-2024.

Green, J., Haigh, I.D., Quinn, N., Neal, J., Wahl, T., Wood, M., Eilander, D., de Ruiter, M., Ward, P., Camus, P., 2025. Review article: a comprehensive review of compound flooding literature with a focus on coastal and estuarine regions. Nat. Hazards Earth Syst. Sci. 25, 747–816. https://doi.org/10.5194/nhess-25-747-2025.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Saf. 81, 23–69. https://doi.org/10.1016/S0951-8320(03)00058-9.

Hoitink, A.F., Jay, D.A., 2016. Tidal river dynamics: implications for deltas. Rev. Geophys. 54, 240–272.

Hu, R., Fang, F., Pain, C., Navon, I., 2019. Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. J. Hydrol. 575, 911–920.

Hussain, D., Khan, A.A., 2020. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. Earth Sci Inform 13, 939–949. https://doi.org/10.1007/s12145-020-00450-z.

Jibhakate, S.M., Timbadiya, P.V., Patel, P.L., 2023. Flood hazard assessment for the coastal urban floodplain using 1D/2D coupled hydrodynamic model. Nat. Hazards 116, 1557–1590. https://doi.org/10.1007/s11069-022-05728-7.

Kirezci, E., Young, I.R., Ranasinghe, R., Lincke, D., Hinkel, J., 2023. Global-scale analysis of socioeconomic impacts of coastal flooding over the 21st century. Front. Mar. Sci. 9. https://doi.org/10.3389/fmars.2022.1024111.

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

Levy, R., Naish, T., Lowry, D., Priestley, R., Winefield, R., Alevropolous-Borrill, A., Beck, E., Bell, R., Blick, G., Dadic, R., Gillies, T., Golledge, N., Heine, Z., Jendersie, S., Lawrence, J., O'Leary, K., Paulik, R., Roberts, C., Taitoko, M., Trayling, N., 2024. Melting ice and rising seas – connecting projected change in Antarctica's ice sheets to communities in Aotearoa New Zealand. J. Roy. Soc. N. Z. 54, 449–472. https://doi.org/10.1080/03036758.2023.2232743.

Lewis, M., Moftakhari, H., Passalacqua, P., 2024. Challenges for compound coastal flood risk management in a warming climate: a case study of the Gulf Coast of the United States. Front. Water 6. https://doi.org/10.3389/frwa.2024.1405603.

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.

Maduwantha, P., Wahl, T., Santamaria-Aguilar, S., Jane, R., Booth, J.F., Kim, H., Villarini, G., 2024. A multivariate statistical framework for mixed storm types in compound flood analysis. Nat. Hazards Earth Syst. Sci. 24, 4091–4107. https://doi.org/10.5194/nhess-24-4091-2024.

Mahakur, V., Mahakur, V.K., Samantaray, S., Ghose, D.K., 2025. Prediction of runoff at ungauged areas employing interpolation techniques and deep learning algorithm. HydroResearch 8, 265–275. https://doi.org/10.1016/j.hydres.2024.12.001.

Marsooli, R., Wang, Y., 2020. Quantifying tidal phase effects on coastal flooding induced by hurricane sandy in manhattan, New York using a micro-scale hydrodynamic model. Front. Built Environ. 6. https://doi.org/10.3389/fbuil.2020.00149.

McKeon, K., Piecuch, C.G., 2025. Compound minor floods and the role of discharge in the Delaware river estuary. J. Geophys. Res.: Oceans 130, e2024JC021716. https://doi.org/10.1029/2024JC021716.

Moftakhari, H., Muñoz, D.F., Song, J.Y., Alipour, A., Moradkhani, H., 2021. Challenges for Appropriate Characterization of Compound Coastal Hazards, vols. 58–68. https://doi.org/10.1061/9780784483695.007.

Moftakhari, H.R., Salvadori, G., AghaKouchak, A., Sanders, B.F., Matthew, R.A., 2017. Compounding effects of sea level rise and fluvial flooding. Proc. Natl. Acad. Sci. 114, 9785–9790. https://doi.org/10.1073/pnas.1620325114.

Muñoz, D.F., Abbaszadeh, P., Moftakhari, H., Moradkhani, H., 2022. Accounting for uncertainties in compound flood hazard assessment: the value of data assimilation. Coast. Eng. 171, 104057. https://doi.org/10.1016/j.coastaleng.2021.104057.

Muñoz, D.F., Moftakhari, H., Moradkhani, H., 2024. Quantifying cascading uncertainty in compound flood modeling with linked process-based and machine learning models. Hydrol. Earth Syst. Sci. 28, 2531–2553. https://doi.org/10.5194/hess-28-2531-2024.

Muñoz, D.F., Muñoz, P., Moftakhari, H., Moradkhani, H., 2021. From local to regional compound flood mapping with deep learning and data fusion techniques. Sci. Total Environ. 782, 146927. https://doi.org/10.1016/j.scitotenv.2021.146927.

NASA earth observatory, 2024. Beryl makes landfall in Texas. https://earthobservatory.nasa.gov/images/153035/beryl-makes-landfall-in-texas accessed 12.18.24.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

National Environmental Satellite, Data, and Information Service, 2024. Hurricane Harvey: a look back seven years later. https://www.nesdis.noaa.gov/news/hurricane-harvey-look-back-seven-years-later accessed 12.18.24.

National Hurricane Center, 2022. Hurricane Nicholas. https://www.nhc.noaa.gov/data/tcr/AL142021_Nicholas.pdf.

National Weather Service, 2015. 2015 memorial day weekend flooding. https://www.weather.gov/media/ewx/wxevents/ewx-20150524.pdf.

National Weather Service, 2008. Hurricane Ike - september 2008. https://www.weather.gov/hgx/projects_ike08 accessed 12.18.24.

Nezhad, S.K., Barooni, M., Velioglu Sogut, D., Weaver, R.J., 2023. Ensemble neural networks for the development of storm surge flood modeling: a comprehensive review. J. Mar. Sci. Eng. 11, 2154. https://doi.org/10.3390/jmse11112154.

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091.

NOAA-NCEI, 2024. U.S. Billion-Dollar weather and climate disasters. https://www.ncei.noaa.gov/archive/accession/0209268 accessed 7.15.24.

Obara, Y., Nakamura, R., 2022. Transfer learning of long short-term memory analysis in significant wave height prediction off the coast of western Tohoku, Japan. Ocean Eng. 266, 113048. https://doi.org/10.1016/j.oceaneng.2022.113048.

Ohenhen, L.O., Shirzaei, M., Ojha, C., Kirwan, M.L., 2023. Hidden vulnerability of US Atlantic coast to sea-level rise due to vertical land motion. Nat. Commun. 14, 2038.

Peña, F., Nardi, F., Melesse, A., Obeysekera, J., Castelli, F., Price, R.M., Crowl, T., Gonzalez-Ramirez, N., 2022. Compound flood modeling framework for surface–subsurface water interactions. Nat. Hazards Earth Syst. Sci. 22, 775–793. https://doi.org/10.5194/nhess-22-775-2022.

Prandle, D., 1985. Classification of tidal response in estuaries from channel geometry. Geophys. J. Int. 80, 209–221. https://doi.org/10.1111/j.1365-246X.1985.tb05086.x.

Radfar, S., Moftakhari, H., Moradkhani, H., 2024. Rapid intensification of tropical cyclones in the Gulf of Mexico is more likely during marine heatwaves. Commun. Earth Environ. 5, 1–13. https://doi.org/10.1038/s43247-024-01578-2.

Samantaray, S., Sahoo, A., Yaseen, Z.M., Al-Suwaiyan, M.S., 2025. River discharge prediction based multivariate climatological variables using hybridized long short-term memory with nature inspired algorithm. J. Hydrol. 649, 132453. https://doi.org/10.1016/j.jhydrol.2024.132453.

Sampurno, J., Vallaeys, V., Ardianto, R., Hanert, E., 2022. Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. Nonlinear Process Geophys. 29, 301–315. https://doi.org/10.5194/npg-29-301-2022.

Sandbach, S.D., Nicholas, A.P., Ashworth, P.J., Best, J.L., Keevil, C.E., Parsons, D.R., Prokocki, E.W., Simpson, C.J., 2018. Hydrodynamic modelling of tidal-fluvial flows in a large river estuary. Estuar. Coast Shelf Sci. 212, 176–188. https://doi.org/10.1016/j.ecss.2018.06.023.

Santiago-Collazo, F.L., Bilskie, M.V., Bacopoulos, P., Hagen, S.C., 2024. Compound inundation modeling of a 1-D idealized coastal watershed using a reduced-physics approach. Water Resour. Res. 60, e2023WR035718. https://doi.org/10.1029/2023WR035718.

Santiago-Collazo, F.L., Bilskie, M.V., Hagen, S.C., 2019. A comprehensive review of compound inundation models in low-gradient coastal watersheds. Environ. Model. Software 119, 166–181. https://doi.org/10.1016/j.envsoft.2019.06.002.

Sattari, A., Foroumandi, E., Gavahi, K., Moradkhani, H., 2025. A probabilistic machine learning framework for daily extreme events forecasting. Expert Syst. Appl. 265, 126004. https://doi.org/10.1016/j.eswa.2024.126004.

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

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, 809–822. https://doi.org/10.5194/nhess-23-809-2023.

Shahabi, A., Tahvildari, N., 2024. A deep-learning model for rapid spatiotemporal prediction of coastal water levels. Coast. Eng. 190, 104504. https://doi.org/10.1016/j.coastaleng.2024.104504.

Singh, A., Chan, S., Moskovitz, T., Grant, E., Saxe, A., Hill, F., 2024. The transient nature of emergent in-context learning in transformers. Adv. Neural Inf. Process. Syst. 36.

Taherkhani, M., Vitousek, S., Barnard, P.L., Frazer, N., Anderson, T.R., Fletcher, C.H., 2020. Sea-level rise exponentially increases coastal flood frequency. Sci. Rep. 10, 6466. https://doi.org/10.1038/s41598-020-62188-4.

Talke, S.A., Jay, D.A., 2020. Changing tides: the role of natural and anthropogenic factors. Ann. Rev. Mar. Sci 12, 121–151. https://doi.org/10.1146/annurev-marine-010419-010727.

Tang, B., Nederhoff, K., Gallien, T.W., 2025. Quantifying compound coastal flooding effects in urban regions using a tightly coupled 1D–2D model explicitly resolving flood defense infrastructure. Coast. Eng. 199, 104728. https://doi.org/10.1016/j.coastaleng.2025.104728.

Tedesco, P., Rabault, J., Sætra, M.L., Kristensen, N.M., Aarnes, O.J., Breivik, Ø., Mauritzen, C., Sætra, Ø., 2024. Bias correction of operational storm surge forecasts using Neural Networks. Ocean Model. 188, 102334. https://doi.org/10.1016/j.ocemod.2024.102334.

Thiéblemont, R., Le Cozannet, G., Nicholls, R.J., Rohmer, J., Wöppelmann, G., Raucoules, D., de Michele, M., Toimil, A., Lincke, D., 2024. Assessing current coastal subsidence at continental scale: insights from europe using the European ground motion Service. Earths Future 12, e2024EF004523. https://doi.org/10.1029/2024EF004523.

Tiggeloven, T., Couasnon, A., van Straaten, C., Muis, S., Ward, P.J., 2021. Exploring deep learning capabilities for surge predictions in coastal areas. Sci. Rep. 11, 17224. https://doi.org/10.1038/s41598-021-96674-0.

Valle-Levinson, A., Olabarrieta, M., Heilman, L., 2020. Compound flooding in houston-galveston bay during hurricane Harvey. Sci. Total Environ. 747, 141272. https://doi.org/10.1016/j.scitotenv.2020.141272.

Volkov, D.L., Zhang, K., Johns, W.E., Willis, J.K., Hobbs, W., Goes, M., Zhang, H., Menemenlis, D., 2023. Atlantic meridional overturning circulation increases flood risk along the United States southeast coast. Nat. Commun. 14, 5095. https://doi.org/10.1038/s41467-023-40848-z.

Wing, O.E.J., Lehman, W., Bates, P.D., Sampson, C.C., Quinn, N., Smith, A.M., Neal, J.C., Porter, J.R., Kousky, C., 2022. Inequitable patterns of US flood risk in the Anthropocene. Nat. Clim. Change 12, 156–162. https://doi.org/10.1038/s41558-021-01265-6.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: convolutional Block attention module. Presented at the Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19.

Wu, J., Zhang, Q., Li, Y., Xu, C.-Y., Ye, X., 2023. Spatial-temporal variations of stage-area hysteretic relationships in large heterogeneous lake–floodplain systems. J. Hydrol. 620, 129507. https://doi.org/10.1016/j.jhydrol.2023.129507.

Xie, X., Xie, M., Moshayedi, A.J., Noori Skandari, M.H., 2022. A hybrid improved neural networks algorithm based on L2 and dropout regularization. Math. Probl Eng. 2022, 8220453. https://doi.org/10.1155/2022/8220453.

Xu, K., Han, Z., Bin, L., Shen, R., Long, Y., 2024. Rapid forecasting of compound flooding for a coastal area based on data-driven approach. Nat. Hazards. https://doi.org/10.1007/s11069-024-06846-0.

Yu, M., Huang, Q., Li, Z., 2024. Deep learning for spatiotemporal forecasting in Earth system science: a review. International Journal of Digital Earth 17, 2391952. https://doi.org/10.1080/17538947.2024.2391952.

Zhong, M., Xiao, L., Li, X., Mei, Y., Jiang, T., Song, L., Chen, X., 2024. A study on compound flood prediction and inundation simulation under future scenarios in a coastal city. J. Hydrol. 628, 130475. https://doi.org/10.1016/j.jhydrol.2023.130475.

Zou, Y., Wang, J., Lei, P., Li, Y., 2023. A novel multi-step ahead forecasting model for flood based on time residual LSTM. J. Hydrol. 620, 129521. https://doi.org/10.1016/j.jhydrol.2023.129521.