

A Deep-Learning Model for Rapid Spatiotemporal Prediction of Coastal Water Levels

Ali Shahabi and Navid Tahvildari¹

Department of Civil and Environmental Engineering, Old Dominion University, Virginia,
United States

Abstract

With the increasing impact of climate change and relative sea level rise, low-lying coastal communities face growing risks from recurrent nuisance flooding and storm tides. Thus, timely and reliable predictions of coastal water levels are critical to resilience in vulnerable coastal areas. Over the past decade, there has been increasing interest in utilizing machine learning (ML) based models for emulation and prediction of coastal water levels. However, flood advisory systems still rely on running computationally demanding hydrodynamic models. To alleviate the computational burden, these physics-based models are either run at small scales with high resolution or at large scales with low resolution. While ML-based models are very fast, they face challenges in terms of ensuring reliability and ability to capture any surge levels. In this paper, we develop a deep neural network for spatiotemporal prediction of water levels in coastal areas of the Chesapeake Bay in the U.S. Our model relies on data from numerical weather prediction models as the atmospheric input and astronomical tide levels, while its outputs are time series of predicted water levels at several tide gauge locations across the Chesapeake Bay. We utilized a CNN-LSTM setting as the architecture of the model. The CNN part extracts the features from a sequence of gridded wind fields and fuses its output to several independent LSTM units. The LSTM units concatenate the atmospheric features with respective astronomical tide levels and produce water level time series. The novel contribution

¹ Corresponding Author, E-mail: ntahvild@odu.edu

of the present work is in spatiotemporality and in prioritization of the physical relationships in the model to maintain a high analogy to hydrodynamic modeling, either in the network architecture or in the selection of predictors and predictands. The results show that this setting yields a strong performance in predicting coastal water levels that cause flooding from minor to major levels. We also show that the model stands up successfully to the rigorous comparison with a high-fidelity ADCIRC model, yielding mean RMSE and correlation coefficient of 14.3 cm and 0.94, respectively, in two extreme cases, versus 12.30 cm and 0.96 for the ADCIRC model. The results highlight the practical feasibility of employing fast yet inexpensive data-driven models for resilient coastal management.

Keywords: Storm Surge Prediction, Nuisance Flooding, Machine Learning, Deep Learning, CNN, LSTM

1- Introduction

Climate change has driven a rise in the intensity and frequency of tropical cyclones or hurricanes over the recent decades (Knutson et al., 2010), amplifying the vulnerability of coastal communities to flooding. Future projections suggest that this trend will continue by the end of the century (Karegar et al., 2017). With tropical cyclones as their primary driver, extreme storm surges have been posing an increased threat to coastal communities (Wahl et al., 2015). The economic impacts of major tropical cyclones are of an immense scale. In the U.S., recent named hurricanes such as Hurricanes Katrina (2005), Harvey (2017), and Ian (2022) have caused economic losses exceeding \$100 billion each (Blake & Zelinsky, 2018; Hanna & Cangialosi, 2014; Knabb et al., 2023). Global sea level rise (SLR), on the other hand, is experiencing an accelerating increasing trend as well (Nerem et al., 2018). SLR not only exacerbates the impact of storm surges by facilitating their propagation towards coastal infrastructures (Castrucci and Tahvildari 2018; Tahvildari and Castrucci, 2021) but also contributes to other adverse coastal and environmental impacts such as shoreline erosion

(Leatherman et al., 2000), saltwater intrusion (Werner & Simmons, 2009), and high-tide flooding in low-lying communities, so-called nuisance flooding (e.g. Moftakhari et al., 2015). Looking ahead, studies estimate a notable increase in the economic impacts of coastal flooding, including both less-frequent, high-impact events (Hallegatte et al., 2013; Hinkel et al., 2014), and less severe, recurrent incidents (Dahl et al., 2017; Ezer & Atkinson, 2014). This highlights the importance of developing more advanced methods for high-fidelity simulation, accurate real-time prediction, and reliable risk assessment of coastal flooding along vulnerable coastlines.

Numerical models, based on shallow water equations, have been successfully utilized for over three decades for modeling coastal flooding. ADCIRC (ADvanced CIRculation, Luetlich et al., 1992), SLOSH (Sea, Lake, and Overland Surges from Hurricanes, Jelesnianski et al., 1992), FVCOM (C. Chen et al., 2003), Delft3D (Roelvink & Van Banning, 1995), ROMS (Shchepetkin & McWilliams, 2005), and SCHISM (Zhang et al., 2016) have been widely used by government agencies in the U.S. as well as researchers for prediction and obtaining risk maps of storm surges. These hydrodynamic models are often coupled with spectral wave models such as SWAN (Dietrich et al., 2011), STWAVE (Smith et al., 2001), and WAM (Group, 1988) to account for flow-wave interactions, adding a level of complexity and computational cost. National Hurricane Center (NHC) uses SLOSH to forecast storm surges in the U.S. as a part of its advisory program. The real-time sea level forecast program by Coastal Emergency Risk Assessment (CERA) group runs ADCIRC+SWAN on the Atlantic and the Gulf of Mexico. In the North Atlantic Coastal Comprehensive Study (NACCS) by the U.S. Army Corps of Engineers, coupled ADCIRC and STWAVE were employed to run 1050 and 100 historical synthetic tropical cyclone scenarios, its results have been used for storm surge risk assessment in the North Atlantic (Cialone et al., 2015). The Operational Nowcast and Forecast Hydrodynamic Model Systems (OFS) by National Oceanic and Atmospheric Administration

(NOAA) includes a suite of 15 real-time hydrodynamic models based on ROMS. The OFS models perform nowcast and short-term (48 hr) water level forecasts along most U.S. coasts. Despite improvements in the accuracy of storm surge predictions using hydrodynamic models, high-fidelity large-scale simulations remain time-consuming and computationally expensive. Generating the data for a regional risk analysis usually requires simulating thousands of scenarios which is hard to achieve with physics-based numerical models. Furthermore, the high computational cost has prohibited real-time flood predictions. Even with low-fidelity models, instant flood calculation using meteorological forecasts has been out of reach. For instance, NOAA's OFS models demand considerable computational resources even though they do not account for waves, such that their performance is limited to 4 runs per day on relatively coarse curvilinear grids.

To achieve fast computation, there has been a growing inclination toward utilizing Machine Learning (ML) approaches to develop data-driven or surrogate models for coastal flooding. While ML techniques like random forests, kriging, and PCA have been widely used (Al Kajbaf & Bensi, 2020; Jia et al., 2016; Jia & Taflanidis, 2013; Kyprioti et al., 2021, 2023; Zahura & Goodall, 2022) for coastal flood modeling, neural networks remain the most popular. To this end, different architectures have been used. Earlier studies used conventional Artificial Neural Networks (ANN), while with the evolution of deep learning, more advanced algorithms such as Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) became prevalent. A category of past studies trained neural networks to predict water level at an individual timestamp, using meteorological and tidal data in hand (Kim et al., 2019; T. L. Lee, 2006; Rajasekaran et al., 2008). Most of these models used ANN recurrently to produce a prediction of water level time series. These types of models are not able to account for the lags between sea state and weather conditions or any other temporal relationships. Another category of studies has sought to predict the time series of water levels, for a specific lead time, using

very recent atmospheric and hydrodynamic information. Due to their ability to capture temporal dependencies, RNNs, specifically Long Short-Term Memory (LSTM) method, are used by most of these models (Bai & Xu, 2022; K. Chen et al., 2022; Igarashi & Tajima, 2021; Wang et al., 2021; Wei & Nguyen, 2022). Increasing the lead time in these models profoundly diminishes the accuracy. In another category of studies, neural networks were developed to achieve an accurate prediction of peak surge for specified storm parameters, either as a track or at landfall (Bass & Bedient, 2018; Hashemi et al., 2016; J. W. Lee et al., 2021; Sahoo & Bhaskaran, 2019). These models are potentially powerful tools for risk assessment purposes as well as preliminary decision-making.

Atmospheric data is the primary input to these ML models. The mentioned studies incorporated atmospheric inputs as either storm track or point-wise data. Recently, utilizing 2D gridded wind data is getting attention, where CNN networks are often used to process them. The sources of these data are numerical weather prediction models such as European Centre for Medium-Range Weather Forecasts (ECMWF), Climate Forecast System (CFS), Global Forecast System (GFS), and North American Mesoscale Forecast System (NAM) by NOAA. Xie et al., (2023) developed a CNN network to extract features from a sequence of wind fields sourced from ECMWF and CFS, together with water levels of the past 24 hours to predict the time series of the next 72 hours. Davila Hernandez et al. (2023) used CNN to extract features from the sequence of image-like weather forecasts and LSTM to extract information from the past water levels and fused both extracted features to an ANN to predict water level time series. Zust et al. (2021) developed a network of spatial and temporal encoders to forecast sea level in the northern Adriatic Sea. They fused ECMWF atmospheric data and past measured surges as inputs of their model.

The literature review reveals three gaps in the existing research: (1) The majority of physics-based storm surge models rely on atmospheric data over their grids and harmonic tide values

at the open boundaries to solve shallow water equations governing hydrodynamic variables. In their time-marching numerical schemes, the current sea state is computed using the current weather and hydrodynamic boundary conditions and the sea state in past time step. Consequently, simulating a time window $[t_1, t_2]$ necessitates atmospheric and tidal inputs within the same temporal span. None of the referenced studies exploit this correlation between input and output data in their models. Since this correlation represents the physical reality, developing a similar linkage within the data in an ML flood prediction model should result in accurate learning and prediction. (2) The models described above provide a singular output point. To predict water level at multiple locations, the models must be separately trained for each point, even though the wind's domain of influence remains same. (3) Few prior studies aimed at training ML models with a wide range of flood levels. The development of a model trained for any surge level, including normal-day sea levels, recurrent nuisance flooding, or extreme storm surges, holds a potential for broad practical applications. Such a model could establish cost-effective real-time flood prediction workflows, offering accuracies comparable to hydrodynamic models.

To address the mentioned gaps, we integrate CNN and LSTM neural networks to develop a spatiotemporal predictive model for coastal flooding. The inputs to this model are the gridded wind data over the domain of interest and harmonic tides at multiple tide gauges, while its outputs are time series of predicted water levels at these gauges. The results show that the model is effective in capturing nearly all water levels, spanning from regular sea state to major flooding. As expected, the error is larger for the most extreme storm tides due to their scarcity in the record. The remainder of the paper is organized as follows: Section 2 describes the methodology and data sources, Section 3 explores the performance metrics related to training and testing, Section 4 presents the results of validation and case studies, and Sections 5 and 6 give the discussion and concluding remarks.

2. Methodology

2.1 Study area

The study is focused on the Chesapeake Bay, the largest estuary in the U.S., which is located in the Mid-Atlantic region. The presence of rivers, tidal embayments, coastal marshes, navigation channels, and highly developed and undeveloped coastlines have created a very complex coastal landscape in the Chesapeake Bay. The Bay is home to a diverse range of habitats such as large wetlands, forests, and rivers, all of which play an important role in supporting species and maintaining the ecological balance of the Bay (Baird & Ulanowicz, 1989). Owing to its low-lying topography, the coastal infrastructure is vulnerable to storm surge and recurrent flooding, which is exacerbated by relative SLR (Castrucci & Tahvildari, 2018; Tahvildari & Castrucci, 2021).

Hampton Roads, a major metropolitan region in southeast Chesapeake Bay, experiences the highest relative SLR rate along the U.S. east coast (Boon et al., 2010). This trend is influenced not only by global SLR caused by climate change but also by land subsidence in this area (Sherpa et al., 2023). Several studies have highlighted the growing susceptibility of the transportation infrastructure in Hampton Roads to coastal flooding, including both storm surges and recurrent nuisance flooding, also referred to as high-tide flooding or sunny day flooding (Shen et al., 2019; Tahvildari et al., 2022). The development of a robust and rapid water level prediction model that leverages real-time weather forecast data will be beneficial to addressing the intensifying flooding issue in the region and other coastal communities facing similar challenges. Such a model can be used for early warning, infrastructure planning, and emergency management.

We selected the Chesapeake Bay as the study area to illustrate the capabilities and challenges of the model in a region with a complex coastal landscape, with highly valued assets concentrated in the vicinity of the coastline that is highly vulnerable to sea level rise and climate

change impacts. The model is designed to calculate water level time series at several locations throughout the Chesapeake Bay, trained with historical water level data. The National Oceanic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS) operate numerous tide stations within the Chesapeake Bay region. For this study, ten stations were chosen (see Fig. 1 and Table 1) as the observation points for training and validating the model. Of these stations, four are located near the Bay mouth (CB, KT, SW, and MP), three are in the midsection (LS, WM, CM), and the remaining three are positioned at the north of the Bay (BT, AP, DC).

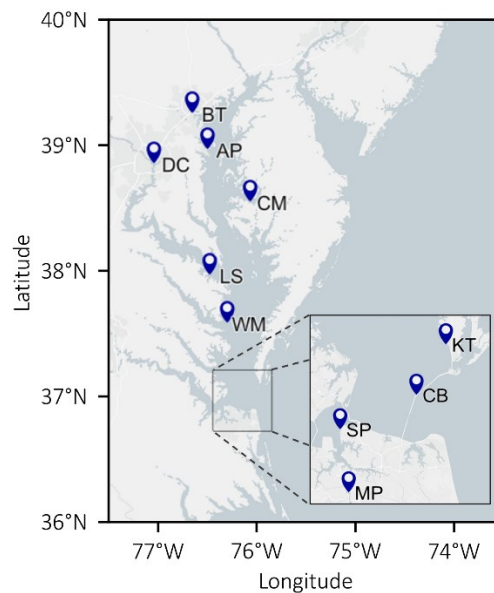


Fig. 1. Map of Chesapeake Bay and the selected observation stations.

Table 1. Names and locations of NOAA stations (NOAA, 2021)

Label	Name	Station ID	LAT. (°)	LON. (°)
AP	Annapolis, MD	8575512	38.9833	-76.4817
BT	Baltimore, MD	8574680	39.2667	-76.6400
CB	CBBT, Chesapeake Channel, VA	8638901	37.0333	-76.0833
CM	Cambridge, MD	8571892	38.5717	-76.0617
DC	Washington, DC	8594900	38.8733	-77.0217
KT	Kiptopeke, VA	8632200	37.1650	-75.9883
LS	Lewisetta, VA	8635750	37.9967	-76.4650
MP	Money Point, VA	8639348	36.7783	-76.3017
SP	Sewells Point, VA	8638610	36.9433	-76.3283
WM	Windmill Point, VA	8636580	37.6150	-76.2900

2.2 Data Preparation

A higher level of inherent relationship within the data in an ML model enhances the potential to gain higher accuracy and dependability. From a physical perspective, total ocean water level is influenced by multiple processes including astronomical tides, storm surge, Coriolis-induced motions, and ocean currents, and the nonlinear interaction between these processes. Among these factors, astronomical tides and wind field are most influential. Astronomical tides are highly predictable and storm surge is mainly driven by wind stress tide, wave setup, and barometric tide, all of which are dependent on the wind field (Dean & Dalrymple, 2001).

The majority of ML-based storm surge models rely on wind data as the primary input. Previous studies have incorporated atmospheric data through various approaches, including point-wise data recorded at meteorological stations, storm track, and more recently, as 2D wind field. The 2D data is preferred as it accounts for more details, and, in contrast to hurricane tracks, it can be used for a wide range of storm events, including minor storms, tropical cyclones, and Nor'easters. In prior studies, tide has been treated differently. Some scholars have removed tide levels from water levels and predicted residual surge levels using atmospheric data. While some others have included tides as their model input. In our proposed model, the input data consists of an hourly sequence of atmospheric data, including velocity components of wind and air pressure (at the sea surface) on a 2D grid, as well as hourly time series of astronomical tide levels at the observation stations. The model's outputs are time series of water level at these stations. The wind velocity vectors at 10 m above sea surface were used as model input. Velocity at this reference elevation is used commonly to parametrize wind stress on water surface because modeling or measuring wind velocity at the sea surface is challenging due to rapid and potentially large changes in surface elevation, and the effect of surface roughness on wind velocity within the atmospheric boundary layer.

209 A dataset spanning 21 years, from 2002 to 2023, served as the training data of the model. Water
210 level measurements, both observed and astronomical, at the observation stations were obtained
211 from the NOAA database (NOAA, 2021), all referenced to the Mean Sea Level (MSL) vertical
212 datum. For the atmospheric inputs, we employed the CFS database. The NCEP Climate
213 Forecast System Reanalysis (CFSR, Saha et al., 2010) was used for the period from 2002 to
214 2010, and The NCEP Climate Forecast System Version 2 (CFSv2, Saha et al., 2014) from 2010
215 to 2023. These datasets have different spatial resolutions, with CFSR at 0.30° resolution and
216 CFSv2 at 0.205° resolution. To ensure consistency, both datasets were resampled using cubic
217 interpolation onto a $5^\circ \times 5^\circ$ grid of 0.205° resolution, covering the entire study region, as
218 illustrated in Figure 2. Within the model, we used the wind velocity components in the form of
219 velocity magnitude and Azimuth direction. The input data were standardized before fusing into
220 the model. In terms of the temporal data length, we selected a window of 84 hours to slide
221 through the data. However, 12 hours of the sliding windows overlap. The first 12 hours of the
222 84-hour period are used to provide the LSTM network with a sufficient memory of the past
223 during the initial time step, but we excluded these 12 hour of predictions from the model output
224 and the calculation of loss and performance parameters. Consequently, each prediction spanned
225 a net length of 72 hours. It is noted that we tested different time windows within 2-4 days,
226 corresponding to typical forecast period, and the model showed slightly better performance for
227 a time window of 72 hours.

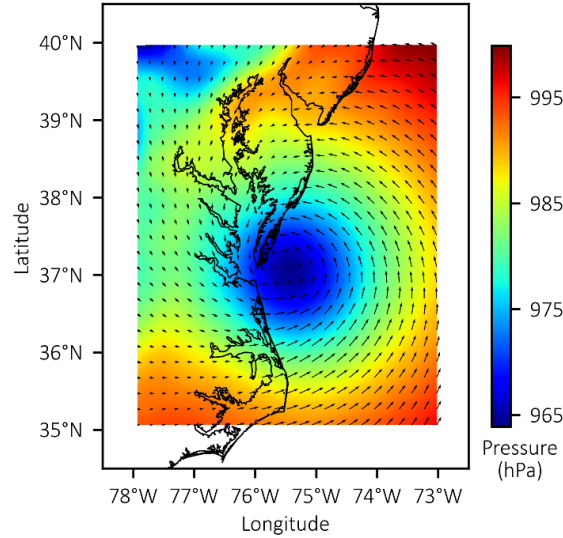


Fig. 2. The extent and resolution of atmospheric data in the model. The figure shows the landfall of Hurricane Irene (2011).

Two approaches exist in terms of the temporal relationship between input and output data within the predictive ML-based models. In one approach, to estimate surge within a time interval $[t, t+T_{lead}]$, very recently recorded data in a time span $[t-T_{min}, t]$ is employed as input. The primary focus of this method is to discern the underlying data patterns and trends, rather than establishing a relationship between input and output grounded in the physics of the problem. This approach is sensitive to lead time, as the accuracy of predictions experiences a decline with increasing lead time. In the second approach, similar to physics-based modeling, a temporal consistency is maintained between input and output. However, in this approach, the model relies on inputs derived from other atmospheric forecast models. This adds an element of uncertainty to the results. Previous models have utilized either of the approaches or their combination. We adopt the second approach here. Our model calculates water level at any time interval given respective inputs from atmospheric models.

To generate training samples, a straightforward method is sliding a fixed time window across the entire dataset, resulting in a substantial number of samples. These samples can either be overlapping or non-overlapping with two samples that do not fully overlap considered as distinct instances. We used a fixed sliding shift of 72 hours to generate non-overlapping

samples. Overall, 2305 samples were created from the collected data. Fig. 3 shows the histogram of the maximum water level in the generated samples.

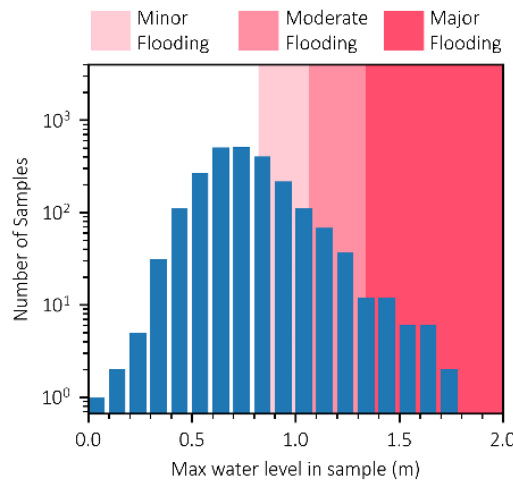


Fig. 3. Distribution of peak water level in the generated samples. The flooding thresholds (minor, moderate, major) are determined by averaging the thresholds used by NOAA across the stations.

2.3. Description of the model

A combination of CNN, LSTM, and fully connected layers was used in the architecture of the model. CNNs excel at processing grid-like data, by effectively capturing local patterns and hierarchies through convolutional layers (Lecun & Bengio, 1995). On the other hand, LSTM networks specialize in sequential data such as time series. While regular RNNs suffer from the vanishing gradient problem, LSTMs capture long-range dependencies in sequential data by utilizing memory cells and a sophisticated gating mechanism (Hochreiter & Schmidhuber, 1997). The combination of CNN and LSTM networks has proven to be highly effective in handling applications that require both spatial and temporal data (Davila Hernandez et al., 2023; Zust et al., 2021). While used in a few past studies, the application of CNN-LSTM networks for spatiotemporal flood prediction holds significant potential.

An ensemble of several trial runs, not detailed here, were conducted to determine the optimal layer dimensions and hyperparameters. The best performance was achieved by manual grid search across the various parameters as well as utilizing random search through hyperparameter

tuning functions. Fig. 4a shows the architecture that delivered the best performance. As shown in this figure, the model incorporates input data in two stages. The process begins with fusing the atmospheric inputs to the CNN layers. The convolution process involves sliding several kernel matrices, typically of the size 2x2 or 3x3, across the input data to produce several feature maps. The kernels initialize with random values and then, are optimized during the backpropagation step. For temporal data, a Time-Distributed option fuses the data frame by frame, enabling the model to process each frame separately (see Fig. 4b).

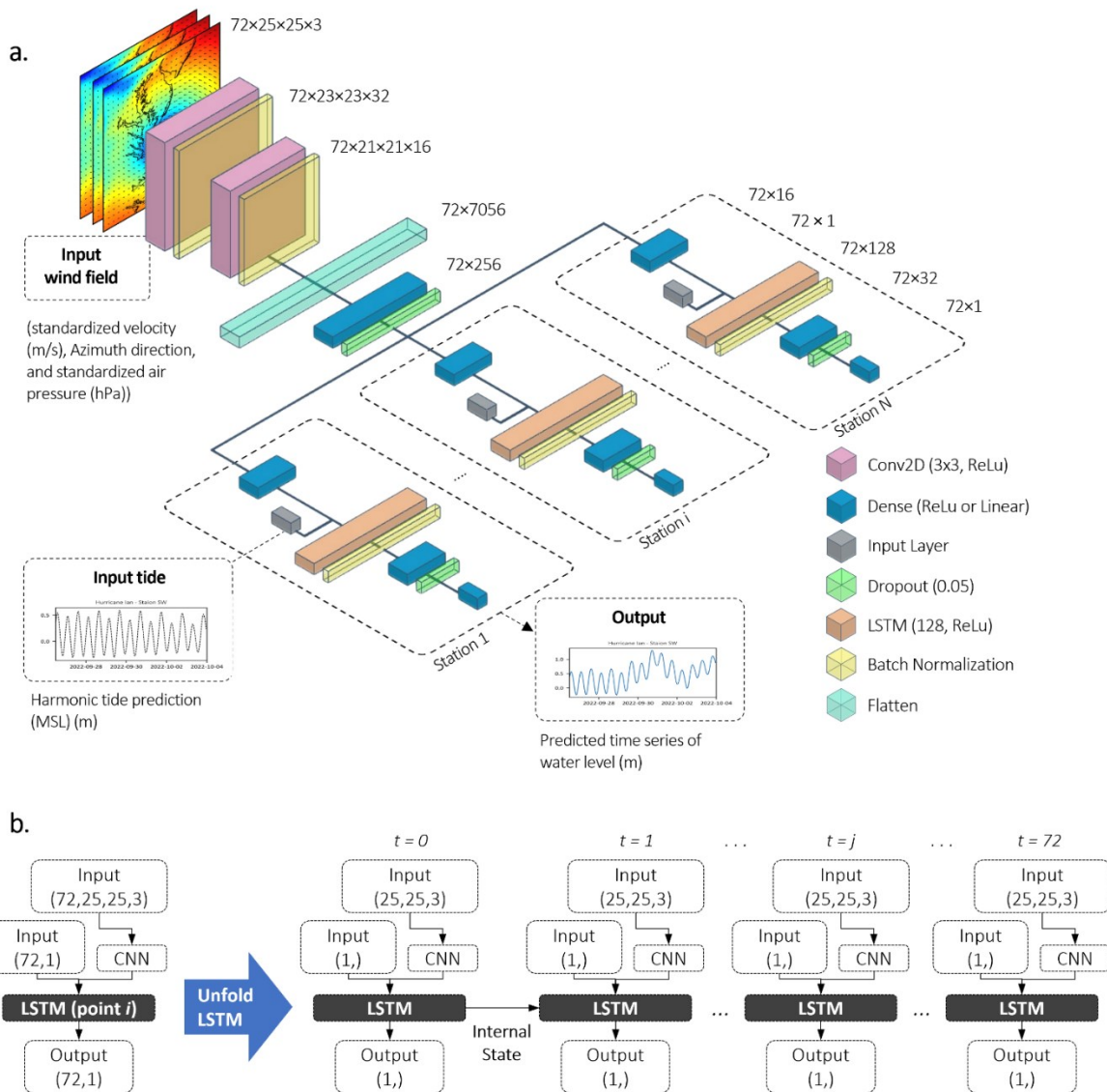


Fig. 4. The architecture of the model. (a) “Conv2D” denotes 2D convolution layers with 32 output channels for the first layer and 16 for the second layer. “Dense” represents fully connected layers. The output dense layers employ a linear activation function while ReLU is used for

the remaining layers. (b) the “Time-Distributed” option enables the model to process one frame of atmospheric input at each timestep of LSTMs.

The input atmospheric data in the model has the dimension (72,25,25,3), including 72 hourly timesteps, 25 by 25 grid points, and 3 wind parameters (velocity, Azimuth direction, and pressure). The initial CNN layer with 3x3 kernels receives the inputs frame by frame and produces 32 output feature maps per frame. This layer is succeeded by the second CNN layer with 32 output channels. Subsequently, one Flatten and one Fully Connected layers condense the output to a 256-element-long vector.

The feature vector derived from the wind field frame undergoes processing within dedicated LSTM units for each of the 10 stations. Before entering each unit, a primary fully connected layer reduces the feature vector dimensions to 16. At this stage, the model incorporates the tide data, concatenating it with the feature vector to enrich the information used for prediction. The LSTM layers, each configured with 128 output parameters at every time step and an equivalent number of hidden parameters, follow this initial processing step. Subsequently, two fully connected layers are employed to further reduce the final size of the outputs, resulting in a 72-hour time series for each observation station. This time series represents the predicted water levels. To improve the model’s performance and prevent overfitting, batch normalization and dropout layers have been used at different stages (see Fig. 4a). A Rectified Linear Unit (ReLU) activation function was used in CNN layers due to its strong empirical success (Zoph & Le, 2018). Table 2 presents the details of hyperparameters used in the model.

Table 2. Summary of hyperparameters used in the model.

Hyperparameter	Value/Method
Loss function	MSE (m ²)
Optimizer	Adam
Learning rate	1e-3
Batch size	64
Early stopping min delta	1e-4 (m ²)
Early stopping patience	15
Dropout rate	0.05
Max epochs	200

3- Model Performance

The model performance was evaluated during both training and testing phases. Mean Squared Error (MSE) was employed as the loss function. MSE is widely employed for regression problems. It quantifies the squared deviation between model predictions and observed data. The formula for calculating MSE is as follows:

$$MSE = \frac{1}{N} \sum_1^N (\eta_{gauge} - \eta_{model})^2, \quad (1)$$

where N is the number of data within a batch, η_{gauge} are the observed water levels, and η_{model} are the output water levels, respectively. In the calculation of MSE, both η_{gauge} and η_{model} are truncated to exclude the first 12 hours of the output time series, as previously discussed in section 2.2.

An 8-fold cross-validation strategy is performed to ensure the reliability of the evaluation and to mitigate potential biases. This approach involves partitioning the dataset into eight subsets, without any shuffling, and training the model on seven of them while validating the remaining subset. This process is repeated eight times, with each subset serving as the validation set once. The validation subsets are partitioned equally into early stopping validation and testing sets. The early stopping validation monitors the model's performance on the validation set and stops the training when further improvement can result in overfitting.

Fig. 5. Shows the learning curve of the model. The steep slope indicates that the model is converging quickly, implying a strong relationship between the data. The dashed lines in the figure, representing the cross-validation data, show a good alignment with the training data. Despite the presence of minor fluctuations, the trends of the two curves are very similar.

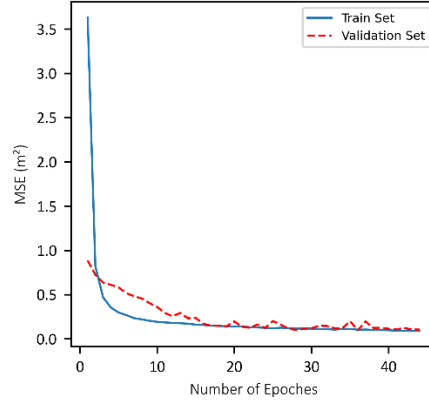


Fig. 5. Learning curve of the model. The blue line represents the training data and the dashed red line denotes the validation set.

The box plots in Fig. 6 show the distribution of both training and testing error across different stations. The errors in this figure are plotted as Root Mean Squared Error (RMSE). Given the model has multiple output layers, each corresponding to a station, the total MSE is computed as the sum of the MSE for all output layers. However, RMSE is calculated by the formula:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\eta_{gauge} - \eta_{model})^2}, \quad (2)$$

where N is the number of data in each station. As shown in Fig. 6, the upper quartile of RMSE for all the stations remains below 12 cm and 18 cm during the training and testing phases, respectively. The station DC shows a more dispersed and larger error, with whisker lines extending to nearly 20 cm during testing. Otherwise, the training and testing errors are comparable across the stations.

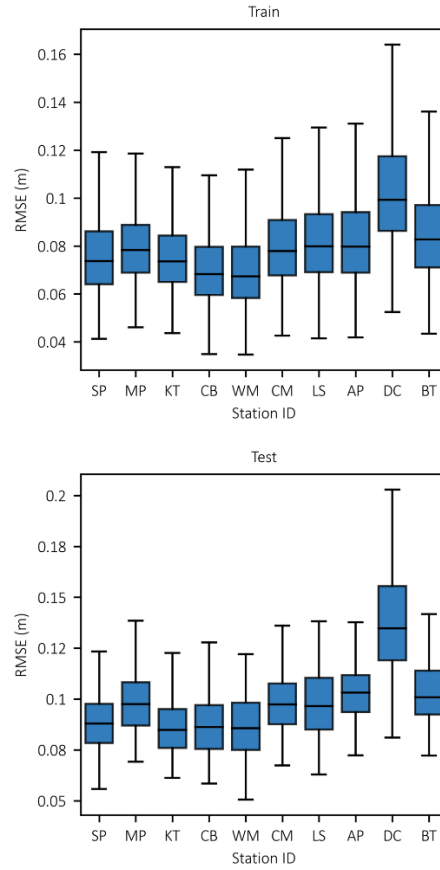


Fig 6. Distribution of RMSE values across stations during training and testing phases. The boxes indicate the median and interquartile ranges (IQR), while the Whiskers extend 1.5 times the IQR.

There are 2026 samples in a training set, each has data for 10 stations, and each station contains 72 hourly water levels. This results in a total of 1,458,720 water level values for the train set, and 103,680 for the test set. Fig. 7 shows the distribution of training and testing Absolute Error (AE), calculated as $|\eta_{gauge} - \eta_{model}|$, for all water levels categorized into four ranges. The box plots show an expected increasing trend in AE with higher water level ranges. For the fourth bin that represents moderate and major flooding in most stations, the median AE is 10.3 cm and 23.2 cm for training and testing stages, respectively.

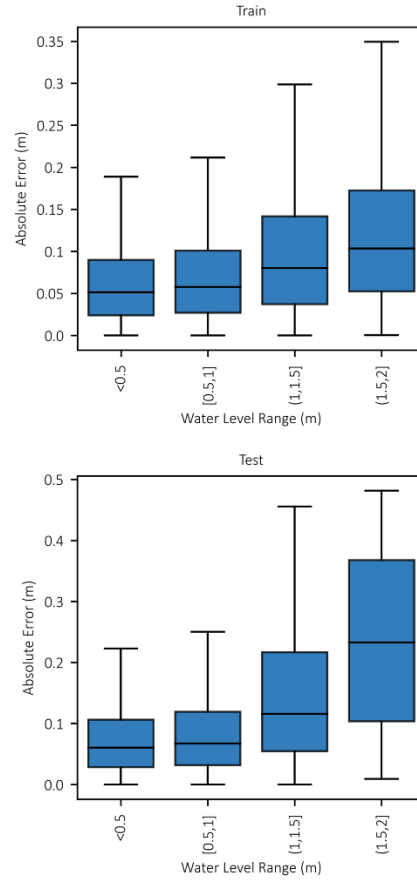


Fig 7. Distribution of absolute error $|\eta_{true} - \eta_{pred}|$ against different water level ranges during training and testing phases. The boxes indicate the median and interquartile ranges (IQR), while the Whiskers extend 1.5 times the IQR.

Fig. 8 compares the RMSE at each station during training and testing steps. The average RMSE during training stands at 8.3 cm, while it slightly increases to 10.1 cm during testing. The model shows a good performance during testing. The close scores during training and testing prove that the model successfully fits the data without overfitting. In trying different architecture settings, we found that the inclusion of batch normalization and drop-out layers are highly effective in preventing overfitting (see Fig. 4a).

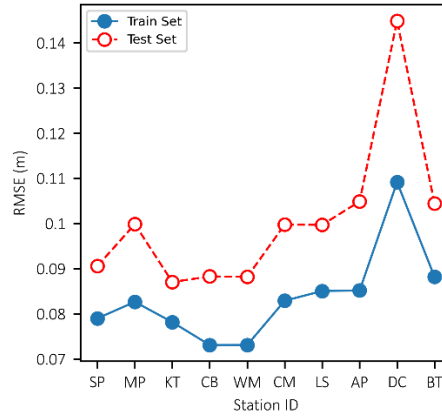


Fig. 8. Comparison of mean RMSE values across the stations between training and testing steps.

4. Storm Tide Prediction

The model's strong performance during testing does not necessarily guarantee its ability in predicting storm tides. This is because the majority of samples in a training set represent normal sea state, where the measured water levels are mainly driven by the tides. We applied the model to nine historical flood events (see Table 2) including major recent hurricanes and two nuisance flooding events to gain a better insight into its performance. The source of data used here was same as in the training step. To predict each event, the respective month was eliminated from the training data to keep its data unseen. The most extreme storm surge, Hurricane Isabel (2003), was kept in the training data.

Table 3. The test cases conducted to show the model's performance. Max Peak Surges are the maximum recorded water level across the stations.

Name	Max Peak Surge (m)	Date
Hurricane Ian	1.32	September-2022
High-Tide flooding	1.26	November-2021
High-Tide flooding	1.30	November-2019
Hurricane Dorian	1.48	September-2019
Hurricane Matthew	1.54	October-2016
Hurricane Joacin	1.66	September-2015
Hurricane Sandy	1.75	October-2012
Hurricane Irene	2.11	August-2011
Nor'Ida 2009	2.11	November-2009

The results herein are presented in terms of RMSE and Coefficient of Correlation. Correlation scores measure the strength of relationship between modeled and observed data. The

coefficient of Determination (R^2) assesses the proportion of the variance in the dependent variable. Adjusted R^2 is commonly used for models with different number of predictors (Basiri et al., 2015). Pearson Coefficient of Correlation (R or CC) quantifies the linear relationship between the observed and modeled data. In this paper, we used CC which is calculated as follows:

$$CC = 1 - \frac{\sum_1^N (\eta_{gauge} - \bar{\eta}_{gauge})(\eta_{model} - \bar{\eta}_{model})}{\sqrt{\sum_1^N (\eta_{gauge} - \bar{\eta}_{gauge})^2 \sum_1^N (\eta_{model} - \bar{\eta}_{model})^2}} \quad (3)$$

where N is the number of data within a prediction.

Fig. 9 shows the time series of the model for Hurricane Ian (2022). It was found that a 6-day time window can capture the duration of storm surge across different storms and different stations sufficiently. This window was used to illustrate the results for Hurricane Ian and other storms, and to calculate the correlation coefficient. Overall, the correlation with the observed time series is satisfactory at all the stations. There is no tidal phase lag between the astronomical data and model results which proves the LSTM units are trained well to follow the tidal phase for various atmospheric inputs. The model shows a very good performance in predicting peak surge for this scenario; the average pick surge in all stations is 0.99 m for both observed and model data.

As an example of model performance for different high water level events, Fig. 10 shows the time series of all the simulated storm surges and king tides in the Sewells Point (SP) tide station. As seen, the model shows a decent performance in predicting water level during major hurricanes in the past two decades. The largest error in the peak storm tide at SP belongs to Hurricane Irene (2011) where the observed and modeled water levels are 1.88 m and 1.45 m, respectively, corresponding to a 23.0% error. On average, the model estimates pick water levels at this station with a 10.6% error.

Fig. 11 shows scatter plots of predicted water levels against gauge data in all the stations. Using MSE as the loss function, a modest underestimation was observed in all stations, but the model showed a decent average RMSE of 11.4 cm in modeling nine historic events, seven of which were extreme storm surges. We examined different loss functions during the training process to minimize the underestimation. Less common loss functions like quantile loss decreased the underestimation but also diminished the overall model performance.

We also compared our model's prediction results with data available in the NACCS database, which includes a suite of high-fidelity numerical simulations of storm surge in the North Atlantic. The effort was conducted to provide information for computing the joint probability of coastal storm forcing parameters for the U.S. Mid-Atlantic and Northeast regions. This information is critical for effective flood risk management, project planning, design, and performance evaluation. The ADCIRC+STWAVE mesh developed for the NACCS study encompasses the western North Atlantic, the Gulf of Mexico, and the western extent of the Caribbean Sea with 3.1 million grid points nodes and 6.2 million grid cells. This mesh is a combination of prior regional meshes developed by FEMA (Cialone et al., 2015). Specifically, for the Chesapeake-Delaware Bay region on the Atlantic coast, the FEMA_R3 mesh was employed, with over 1.7 million nodes and a resolution of up to 30m (Blanton et al., 2011). With this resolution, the NACCS model is considered a reliable high-fidelity flow-wave model for studying storm surge. The initial NACCS model was validated with NOAA tide gauges for several cases. Among the historic events we simulated to show the performance of the model, Hurricanes Irene (2011) and Sandy (2012) were available in the NACCS's historic simulations and were compared to model results. NACCS model generated outputs, including water levels, at approximately 18,000 points, with around 3,000 of these points located within the Chesapeake Bay. For comparison, we used outputs at locations nearly identical to the tide stations used in this study (Table 1).

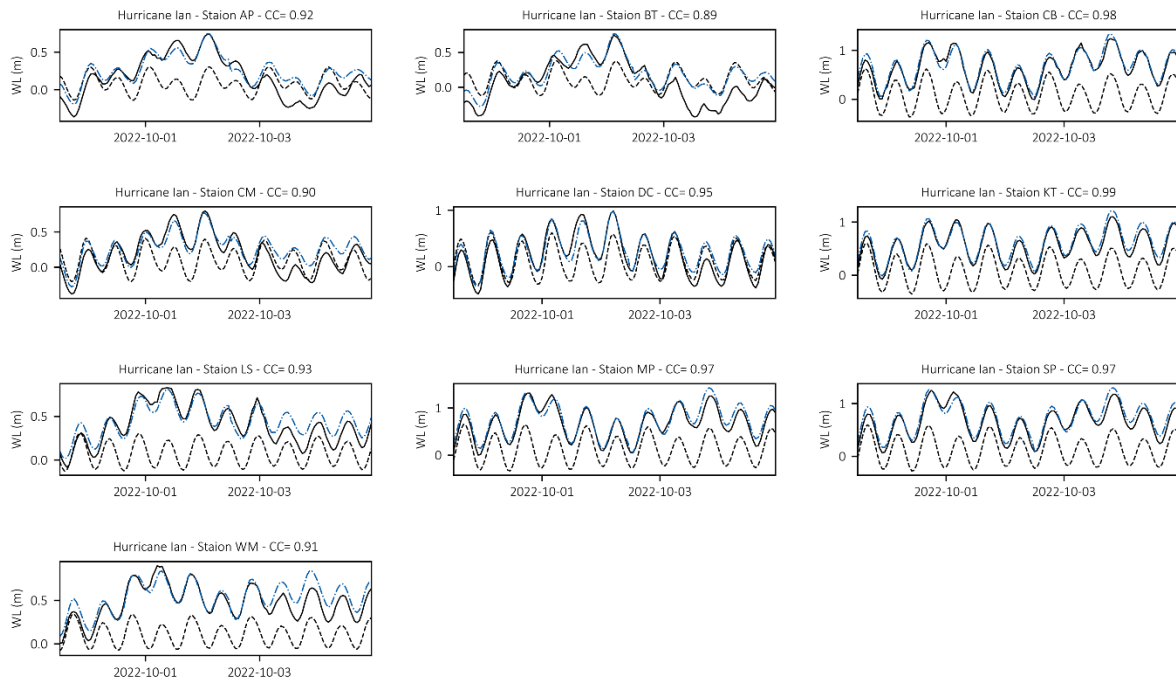


Fig. 9. Time series of water level during Hurricane Ian (2022) at the tide stations under study. Dashed black lines represent astronomical tide levels, the solid black lines are the gauge data and the model's outputs are the blue dash-dot lines.

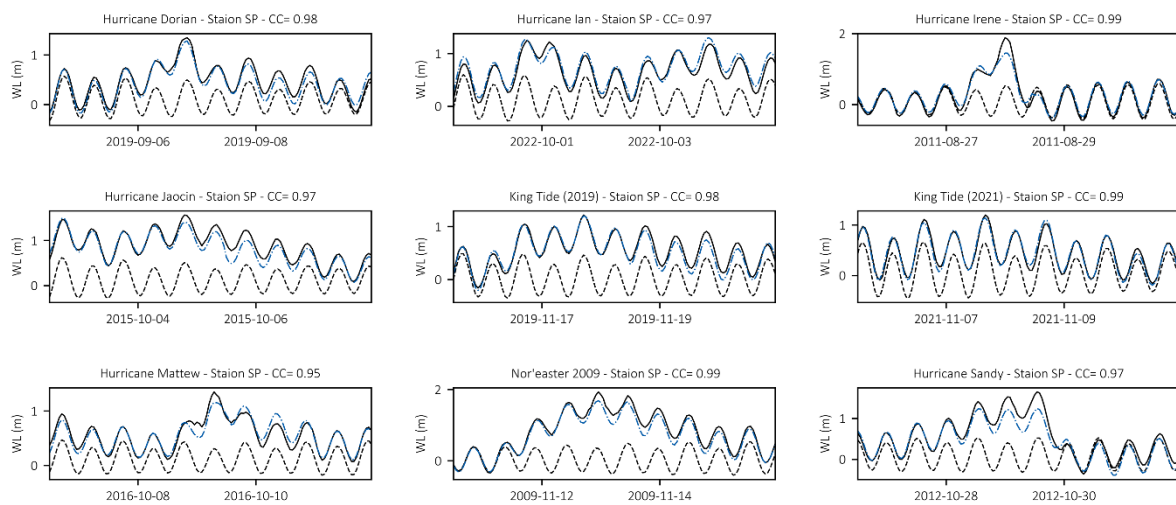


Fig. 10. Time series of simulated storm surges and king tides at the Sewells Point (SP) tide gage. Dashed black lines represent astronomical tide levels, the solid black lines are the gauge data and the model's outputs are the blue dash-dot lines.

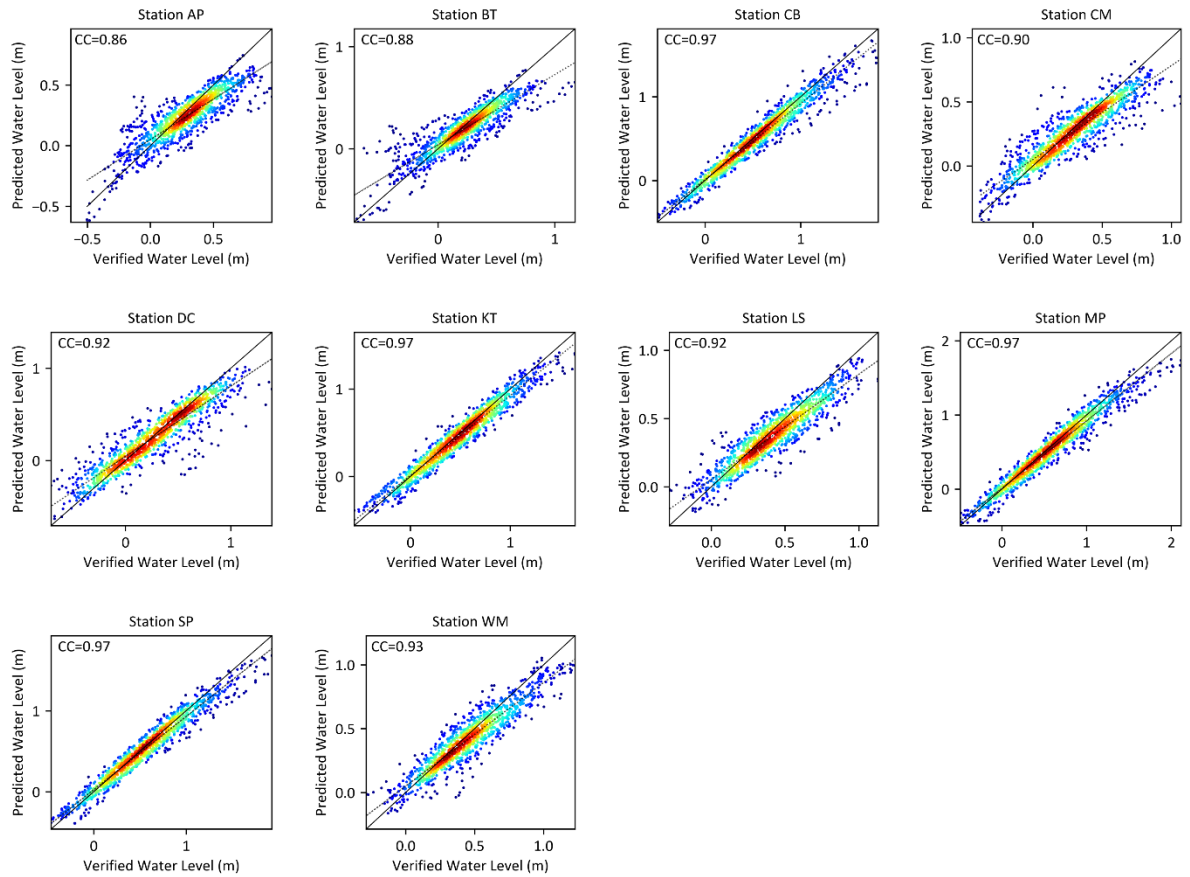


Fig. 11. Scatter plots of predicted against observed water levels across the stations. The color map represents the density of points.

Fig. 12 shows the time series output from the ML model, the ADCIRC model, and the gauge data for hurricanes Irene (2011) and Sandy (2012). The ADCIRC model shows a strong performance in capturing the peak surge in Irene for most stations, while the ML model underestimates the peak. Both models show a weak performance in the northern stations AP, BT, and DC. Excluding these stations, the average errors in peak residual surge values are 21.0% and 12.0% for the present model and the ADCIRC model, while in the northern stations, these errors are 96% and 65%, respectively. The average RMSEs are close; for hurricane Irene, the ML model yields an average RMSE of 13.4 cm, compared to 16.2 cm for ADCIRC. For Hurricane Sandy, the ML model has a 19.5 cm RMSE compared to 15.8 cm for ADCIRC.

Fig. 13 shows the scatter plots of both models and Fig. 14 summarizes the comparison between the two models across the observation stations. Averaging over all the stations for the two

storms used for comparison, the ML model yields an RMSE of 16.4 cm and CC of 0.90, compared to 15.9 cm and 0.88 for the ADCIRC model. Neglecting the uncertain results of the three northern stations, the ML model yields an RMSE of 14.3 cm and CC of 0.94, compared to 12.9 cm and 0.96 for the ADCIRC model. These scores offer a more realistic representation of the comparison between the two models. In summary, the performance of the ML model is quite competitive, with the ADCIRC model performing slightly better in capturing peak surge.

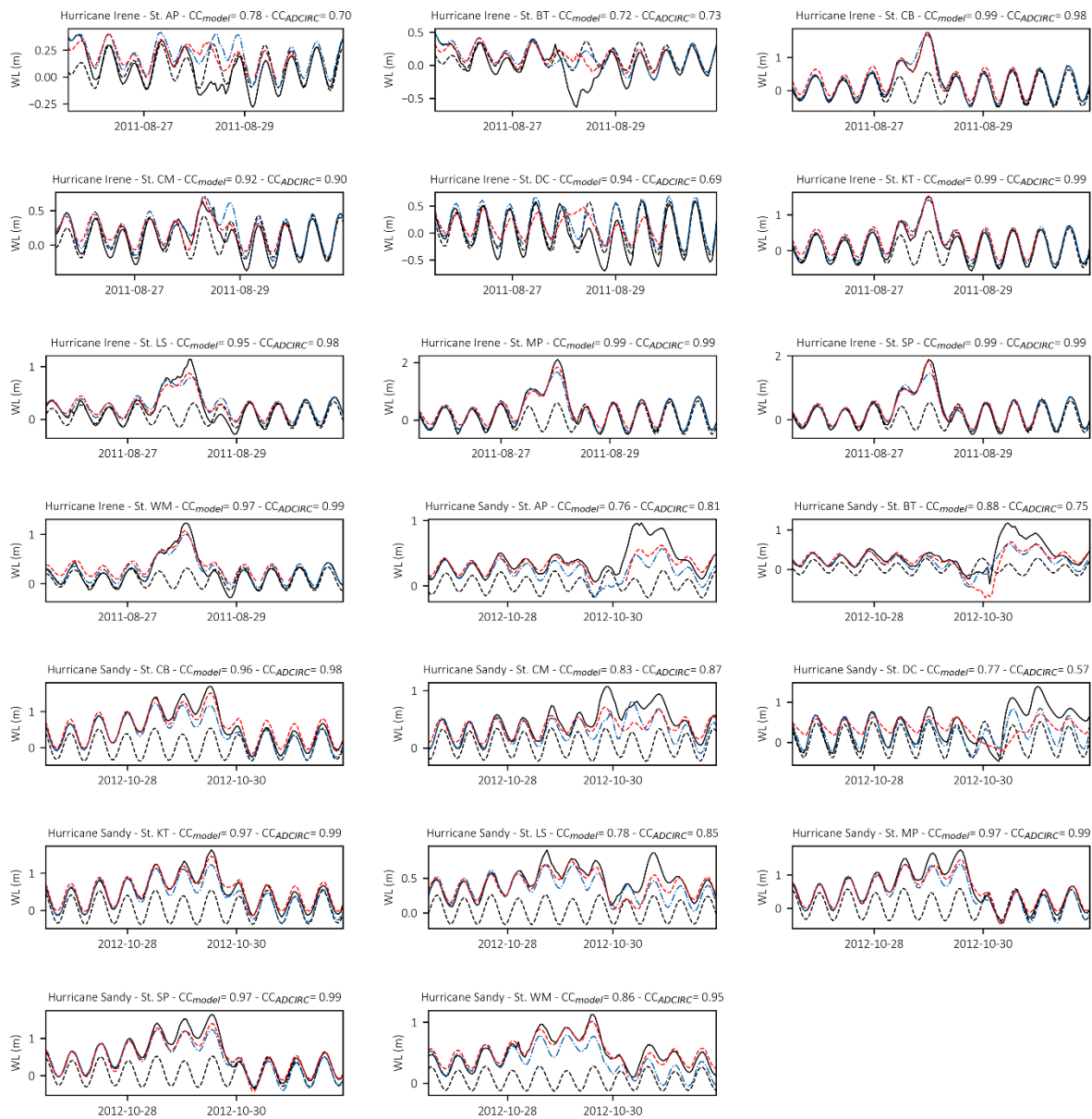


Fig. 12. Time series of water level output using the present ML model, ADCIRC model, and the observations for Hurricanes Irene (2011) and Sandy (2016). Black solid lines represent the observed gauge data, red dashed lines pertain to the ADCIRC model, blue dash-dot lines show the present model's results, and black dashed lines represent harmonic tides.

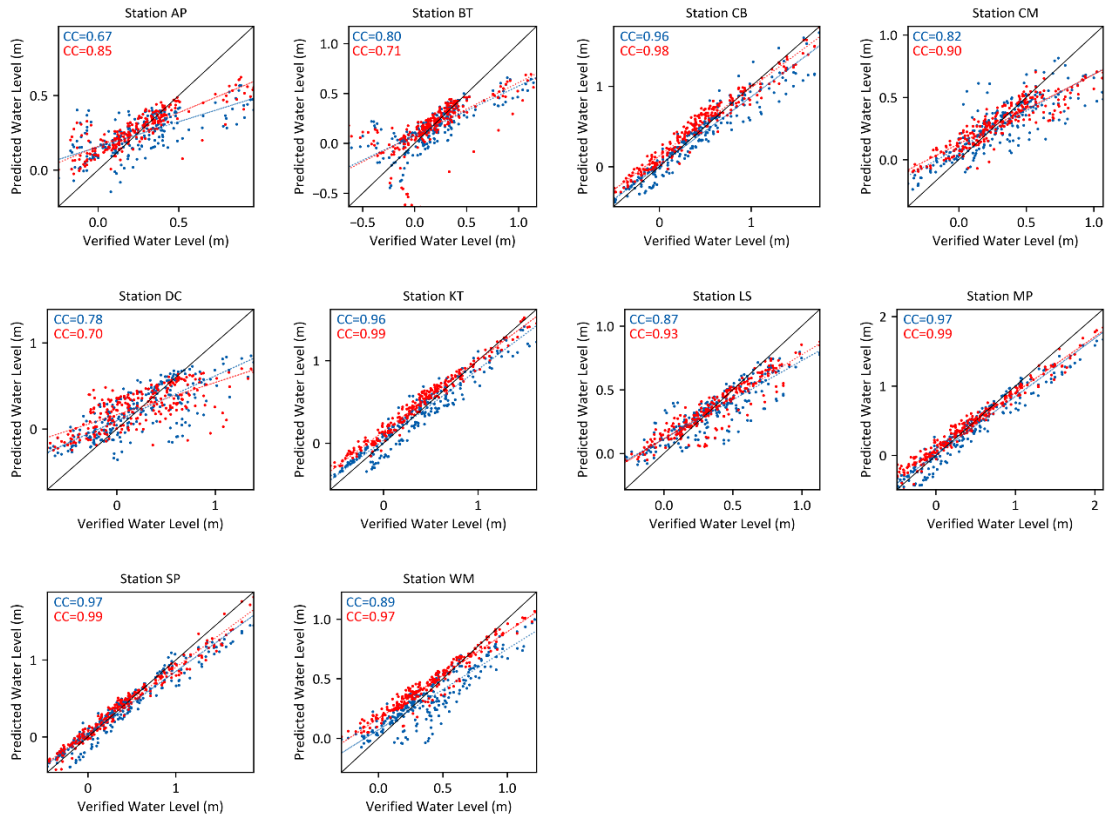


Fig. 13. Scatter plots of observed water levels against the ADCIRC (red dots) and the present ML model (blue dots).

5. Discussion

(1) Overall, the present model shows a strong performance in predicting coastal water levels of different intensity, despite the existence of a few outliers in the predictions. We can notice a trend in the results: the model shows a more reliable performance in prediction of water level near the bay mouth, compared to the stations located at higher latitudes towards the end of the Bay. This behavior is observed in the NACCS model results as well. Though not an objective of this study, discussing the implications of this trend is worthwhile. Several factors may contribute to such disparity. From the ML point of view, difference in correlation means that the network finds a stronger relationship between inputs and outputs in the southern stations compared with the northern ones. This means other factors may affect water level in the northern stations such as more complex hydrodynamics caused by tidal reflection, density gradient circulations, or riverine discharges. From the hydrodynamics point of view, the

performance of the NACCS model (ADCIRC+STWAVE) in storm surge modeling has been widely verified. On the other hand, the circulations within the water column due to density gradients are expected to have a negligible surface signature compared to barotropic tides, wind waves, and wave-induced currents. Neither ADCIRC+STWAVE nor the present ML model accounts for fluvial and pluvial flows. By exploring respective USGS gauges, we found that the discharge from the Potomac and Susquehanna rivers during storm events was not negligible and could affect water levels in the northern Bay. For instance, during Hurricane Sandy (2016), the discharge of the Potomac River increased substantially from 3,000 ft/s to 140,000 ft/s. This underscores the fact that accounting for river discharge in input data in future studies could result in a stronger correlation between ML model output and observations.

(2) The lowest performance of the ML model pertains to the most extreme storms due to the scarcity of that data. Therefore, training the model with more extreme samples, either historical or synthetic, could improve the accuracy of the model in predicting extreme events. In our training data, there are very few major hurricanes with category two or more. The model yields 13.8 cm RMSE for Nor' Ida (2009) considering there is only one more extreme storm in the sample training data (Hurricane Isabel (2003)). While in the case of Hurricanes Ian (2022), since the model had several more extreme storms in the training data, the performance was superior with RMSE at 11.7 cm and CC at 0.97. Synthetic simulations such as NACCS database provide a wide range of hurricanes with rare return periods. Although these data are associated with uncertainties and respective errors, high-fidelity validated data such as NACCS are valuable sources of samples. Future works can combine synthetic storm simulations with real gauge data to improve the model's skill for unseen extreme storms.

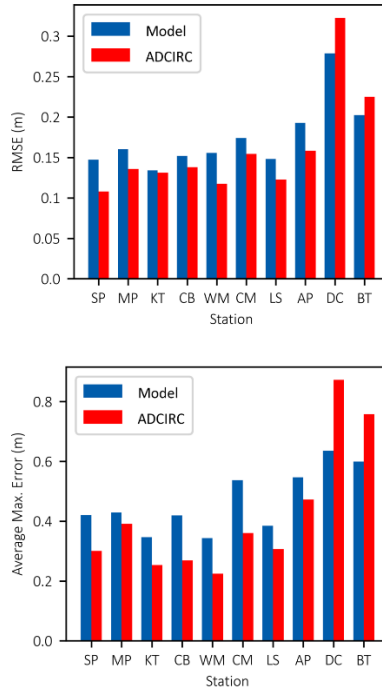


Fig. 14. Summary of performance comparison between the ML model and ADCIRC

The present model substantially reduces the computational time required for water level predictions. Typically, hydrodynamic models take several hours to days, depending on grid resolution and availability of high-performance computation resources, to calculate water levels across regional scales similar to the present study area. The present ML, once trained, produces water level in a few seconds using a personal laptop.

6. Conclusion

In this study, we developed a data-driven model using the deep learning approach to predict water level time series in multiple locations across the Chesapeake Bay. The model is trained with the data from NOAA tide gages and generates water level predictions at the same locations. The data from these tide gages is widely used by local and regional authorities for flood warning and emergency management, and rapid water level predictions at these locations can enable timely precautions and warnings.

The architecture of our model consists of a time-distributed combination of a CNN network with multiple LSTM units. The input data are the hourly sequence of atmospheric data on a 25x25 grid of 0.205° resolution and the time series of astronomical tides at the output points, while the outputs of model are the time series of predicted water level. To maintain a high physical relation, all input and output data pertain to the same time window, with no lead-lag approach. The training data includes samples created from 21 years of continuous data, from 2002 to 2023. An adaptive sliding approach addresses the bias that arose due to the rarity of flooding events against normal sea states.

The trained model is tested with an average RMSE of 10.0 cm. To better evaluate the model's performance predicting storm surge occurrences, nine historical events were studied, where the model scored a CC of 0.94 and RMSE of 11.4 cm. We compared the performance of our method against the NACCS database, the results of high-fidelity hydrodynamic modeling of the North Atlantic. Our model demonstrated a decent accuracy and a comparable performance with the ADCIRC+STWAVE model used in the NACCS study. This study is an effort to address the speed-accuracy trade-off encountered in timely prediction of coastal flooding and suggests that physically related data fused to properly designed deep neural networks can be as accurate as high-resolution hydrodynamic models in real-time forecasting of coastal water levels.

Acknowledgments

This work was supported by the National Science Foundation Grant No. 1951745.

CRedit author statement

Ali Shahabi: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – Original Draft, Visualization, **Navid Tahvildari:** Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition

511 References

- 512 Al Kajbaf, A., & Benssi, M. (2020). Application of surrogate models in estimation of storm surge: A comparative
513 assessment. *Applied Soft Computing Journal*, 91. <https://doi.org/10.1016/j.asoc.2020.106184>
- 514 Bai, L. H., & Xu, H. (2022). Accurate storm surge forecasting using the encoder-decoder long short term memory
515 recurrent neural network. *Physics of Fluids*, 34(1). <https://doi.org/10.1063/5.0081858>
- 516 Baird, D., & Ulanowicz, R. E. (1989). The Seasonal Dynamics of The Chesapeake Bay Ecosystem. *Ecological*
517 *Monographs*, 59(4), 329–364. <https://doi.org/10.2307/1943071>
- 518 Basiri, M. H., Javadnejad, F., & Saeedi, A. (2015). Forecasting crude oil price with an artificial neural network
519 model based on a regular pattern for selecting of the training and testing sets using dynamic command-line
520 functions. *Proceedings of the 24th International Mining Congress of Turkey, IMCET 2015, April*, 732–741.
- 521 Bass, B., & Bedient, P. (2018). Surrogate modeling of joint flood risk across coastal watersheds. *Journal of*
522 *Hydrology*, 558, 159–173. <https://doi.org/10.1016/j.jhydrol.2018.01.014>
- 523 Blake, E. S., & Zelinsky, D. A. (2018). *National Hurricane Center Tropical Cyclone Report: Hurricane Harvey*
524 *(AL092017)*. 2005, 1–77. https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf
- 525 Blanton, B., Stillwell, L., Roberts, H., Atkinson, J., Zou, S., Forte, M., Hanson, J., & Luettich, R. (2011). *Coastal*
526 *Storm Surge Analysis : Computational System*. US Army Engineer Research and Development Center,
527 Coastal and Hydraulics Laboratory.
- 528 Boon, J. D., Brubaker, J. M., & Forrest, D. R. (2010). *Chesapeake Bay Land Subsidence and Sea Level Change*
529 *An Evaluation of Past and Present Trends and Future Outlook*. Virginia Institute of Marine Science Special
530 *Report No. 425 in Applied Marine Science and Ocean Engineering*. 425.
- 531 Castrucci, L., & Tahvildari, N. (2018). Modeling the impacts of sea level rise on storm surge inundation in flood-
532 prone urban areas of Hampton roads, Virginia. *Marine Technology Society Journal*, 52(2), 92–105.
533 <https://doi.org/10.4031/MTSJ.52.2.11>
- 534 Chen, C., Liu, H., & Beardsley, R. C. (2003). An unstructured grid, finite-volume, three-dimensional, primitive
535 equations ocean model: Application to coastal ocean and estuaries. *Journal of Atmospheric and Oceanic*
536 *Technology*, 20(1), 159–186. [https://doi.org/10.1175/1520-0426\(2003\)020<0159:AUGFVT>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)020<0159:AUGFVT>2.0.CO;2)
- 537 Chen, K., Kuang, C., Wang, L., Chen, K., Han, X., & Fan, J. (2022). Storm surge prediction based on long short-
538 term memory neural network in the east China sea. *Applied Sciences (Switzerland)*, 12(1).
539 <https://doi.org/10.3390/app12010181>
- 540 Cialone, M. A., Massey, T. C., Anderson, M. E., Grzegorzewski, A. S., Jensen, R. E., Cialone, A., Mark, D. J.,
541 Pevey, K. C., Gunkel, B. L., & McAlpin, T. O. (2015). *North Atlantic Coast Comprehensive Study Coastal*
542 *Storm Model Simulations: Waves and Water Levels*. US Army Engineer Research and Development Center,
543 Coastal and Hydraulics Laboratory. <http://hdl.handle.net/11681/7339>
- 544 Dahl, K. A., Fitzpatrick, M. F., & Spanger-Siegfried, E. (2017). Sea level rise drives increased tidal flooding
545 frequency at tide gauges along the U.S. East and Gulf Coasts: Projections for 2030 and 2045. *PLoS ONE*,
546 12(2), 1–23. <https://doi.org/10.1371/journal.pone.0170949>
- 547 Davila Hernandez, C., Ho, J., Kim, D., & Oubeidillah, A. (2023). Machine-Learning-Based Model for Hurricane
548 Storm Surge Forecasting in the Lower Laguna Madre. *Algorithms*, 16(5), 232.
549 <https://doi.org/10.3390/a16050232>
- 550 Dean, R. G., & Dalrymple, R. A. (2001). *Coastal Processes with Engineering Applications*. Cambridge University
551 Press. <https://doi.org/10.1017/CBO9780511754500>
- 552 Dietrich, J. C., Zijlema, M., Westerink, J. J., Holthuijsen, L. H., Dawson, C., Luettich, R. A., Jensen, R. E., Smith,
553 J. M., Stelling, G. S., & Stone, G. W. (2011). Modeling hurricane waves and storm surge using integrally-
554 coupled, scalable computations. *Coastal Engineering*, 58(1), 45–65.
555 <https://doi.org/10.1016/j.coastaleng.2010.08.001>
- 556 Ezer, T., & Atkinson, L. P. (2014). Accelerated flooding along the U.S. East Coast: On the impact of sea-level
557 rise, tides, storms, the Gulf Stream, and the North Atlantic Oscillations. *Earth's Future*, 2(8), 362–382.
558 <https://doi.org/10.1002/2014ef000252>
- 559 Group, T. W. (1988). The WAM Model—A Third Generation Ocean Wave Prediction Model. *Journal of Physical*
560 *Oceanography*, 18(12), 1775–1810. [https://doi.org/10.1175/1520-0485\(1988\)018<1775:TWMTGO>2.0.CO;2](https://doi.org/10.1175/1520-0485(1988)018<1775:TWMTGO>2.0.CO;2)
- 561 Hallegatte, S., Green, C., Nicholls, R. J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities.
562 *Nature Climate Change*, 3(9), 802–806. <https://doi.org/10.1038/nclimate1979>

- Hanna, T. S., & Cangialosi, J. P. (2014). *National Hurricane Center Tropical Cyclone Report. October*, 22–28.
- Hashemi, M. R., Spaulding, M. L., Shaw, A., Farhadi, H., & Lewis, M. (2016). An efficient artificial intelligence model for prediction of tropical storm surge. *Natural Hazards*, 82(1), 471–491. <https://doi.org/10.1007/s11069-016-2193-4>
- Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S. J., Marzeion, B., Fettweis, X., Ionescu, C., & Levermann, A. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9), 3292–3297. <https://doi.org/10.1073/pnas.1222469111>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Igarashi, Y., & Tajima, Y. (2021). Application of recurrent neural network for prediction of the time-varying storm surge. *Coastal Engineering Journal*, 63(1), 68–82. <https://doi.org/10.1080/21664250.2020.1868736>
- Jelesnianski, C. P., Chen, J., & Shaffer, W. A. (1992). *Sea, Lake, and Overland Surges from Hurricanes* (Issue NOAA technical report NWS; 48). https://repository.library.noaa.gov/view/noaa/7235/noaa_7235_DS1.pdf
- Jia, G., & Taflanidis, A. A. (2013). Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment. *Computer Methods in Applied Mechanics and Engineering*, 261–262, 24–38. <https://doi.org/10.1016/j.cma.2013.03.012>
- Jia, G., Taflanidis, A. A., Nadal-Caraballo, N. C., Melby, J. A., Kennedy, A. B., & Smith, J. M. (2016). Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms. *Natural Hazards*, 81(2), 909–938. <https://doi.org/10.1007/s11069-015-2111-1>
- Karegar, M. A., Dixon, T. H., Malservisi, R., Kusche, J., & Engelhart, S. E. (2017). Nuisance Flooding and Relative Sea-Level Rise: The Importance of Present-Day Land Motion. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-11544-y>
- Kim, S., Pan, S., & Mase, H. (2019). Artificial neural network-based storm surge forecast model: Practical application to Sakai Minato, Japan. *Applied Ocean Research*, 91(June), 101871. <https://doi.org/10.1016/j.apor.2019.101871>
- Knabb, R. D., Rhome, J. R., Brown, D. P., & National Oceanic and Atmospheric Administration National Hurricane Center. (2023). Hurricane Katrina Tropical Cyclone Report (23 - 30 August 2005). *Tropical Cyclone Report, August 2005*. https://www.nhc.noaa.gov/data/tcr/AL122005_Katrina.pdf
- Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J. P., Srivastava, A. K., & Sugi, M. (2010). Tropical cyclones and climate change. *Nature Geoscience*, 3(3), 157–163. <https://doi.org/10.1038/ngeo779>
- Kyprioti, A. P., Irwin, C., Taflanidis, A. A., Nadal-Caraballo, N. C., Yawn, M. C., & Aucoin, L. A. (2023). Spatio-temporal storm surge emulation using Gaussian Process techniques. *Coastal Engineering*, 180(June 2022), 104231. <https://doi.org/10.1016/j.coastaleng.2022.104231>
- Kyprioti, A. P., Taflanidis, A. A., Nadal-Caraballo, N. C., & Campbell, M. (2021). Storm hazard analysis over extended geospatial grids utilizing surrogate models. *Coastal Engineering*, 168(August 2020), 103855. <https://doi.org/10.1016/j.coastaleng.2021.103855>
- Leatherman, S. P., Zhang, K., & Douglas, B. C. (2000). Sea level rise shown to drive coastal erosion. *Eos*, 81(6), 55–57. <https://doi.org/10.1029/00EO00034>
- Lecun, Y., & Bengio, Y. (1995). Convolutional Networks for Images, Speech, and Time-Series. *The Handbook of Brain Theory and Neural Networks*, 255–258. https://www.researchgate.net/profile/Yann_Lecun/publication/2453996_Convolutional_Networks_for_Images_Speech_and_Time-Series/links/0deec519dfa2325502000000.pdf?%0Ahttps://www.researchgate.net/profile/Yann_Lecun/publication/2453996_Convolutional_Networks_for
- Lee, J. W., Irish, J. L., Bensi, M. T., & Marcy, D. C. (2021). Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning. *Coastal Engineering*, 170(September), 104024. <https://doi.org/10.1016/j.coastaleng.2021.104024>
- Lee, T. L. (2006). Neural network prediction of a storm surge. *Ocean Engineering*, 33(3–4), 483–494. <https://doi.org/10.1016/j.oceaneng.2005.04.012>
- Luettich, R. A., Westerink, J. J., & Scheffner, N. W. (1992). ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves Coasts and Estuaries, Report 1: Theory and Methodology of ADCIRC-2DDI

- and ADCIRC-3DL, Dredging Research Program Technical Report DRP-92-6. In *Coastal Engineering Research Center (U.S.), Engineer Research and Development Center (U.S.)*. (Issue 32466, pp. 1–137). <https://erdc-library.erdcresearchcenter.dren.mil/jspui/handle/11681/4618>
- Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Feldman, D. L., Sweet, W., Matthew, R. A., & Luke, A. (2015). Increased nuisance flooding along the coasts of the United States due to sea level rise: Past and future. *Geophysical Research Letters*, 42(22), 9846–9852. <https://doi.org/10.1002/2015GL066072>
- Nerem, R. S., Beckley, B. D., Fasullo, J. T., Hamlington, B. D., Masters, D., & Mitchum, G. T. (2018). Climate-change-driven accelerated sea-level rise detected in the altimeter era. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9), 2022–2025. <https://doi.org/10.1073/pnas.1717312115>
- NOAA. (2021). *Tides and currents*. <https://tidesandcurrents.noaa.gov/stations.html?type=Water+Levels>
- Rajasekaran, S., Gayathri, S., & Lee, T. L. (2008). Support vector regression methodology for storm surge predictions. *Ocean Engineering*, 35(16), 1578–1587. <https://doi.org/10.1016/j.oceaneng.2008.08.004>
- Ris, R., Holthuijsen, L., & Booij, N. (2011). A Spectral Wave Model for the Coastal Zone. *Coastal Engineering Proceedings*, 1(24), 68–78.
- Roelvink, J., & Van Banning, G. (1995). Design and development of DELFT3D and application to coastal morphodynamics. *Oceanographic Literature Review*, 11(42), 925. <https://www.infona.pl/resource/bwmetal.element.elsevier-1ca19bb6-25b9-3bf5-bfe9-e96a7027c553>
- Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., ... Goldberg, M. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8), 1015–1057. <https://doi.org/10.1175/2010BAMS3001.1>
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H. Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Van Den Dool, H., Zhang, Q., Wang, W., Chen, M., & Becker, E. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, 27(6), 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Sahoo, B., & Bhaskaran, P. K. (2019). Prediction of storm surge and coastal inundation using Artificial Neural Network – A case study for 1999 Odisha Super Cyclone. *Weather and Climate Extremes*, 23(September 2018), 100196. <https://doi.org/10.1016/j.wace.2019.100196>
- Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9(4), 347–404. <https://doi.org/10.1016/j.ocemod.2004.08.002>
- Shen, Y., Morsy, M. M., Huxley, C., Tahvildari, N., & Goodall, J. L. (2019). Flood risk assessment and increased resilience for coastal urban watersheds under the combined impact of storm tide and heavy rainfall. *Journal of Hydrology*, 579(September), 124159. <https://doi.org/10.1016/j.jhydrol.2019.124159>
- Sherpa, S. F., Shirzaei, M., & Ojha, C. (2023). Disruptive Role of Vertical Land Motion in Future Assessments of Climate Change-Driven Sea-Level Rise and Coastal Flooding Hazards in the Chesapeake Bay. *Journal of Geophysical Research: Solid Earth*, 128(4), 1–18. <https://doi.org/10.1029/2022JB025993>
- Smith, J. M., Sherlock, A. R., & Resio, D. T. (2001). *STWAVE: Steady-State Spectral Wave Model User's Manual for STWAVE, Version 3.0*. February, 66.
- Tahvildari, N., Abi Aad, M., Sahu, A., Shen, Y., Morsy, M., Murray-Tuite, P., Goodall, J. L., Heaslip, K., & Cetin, M. (2022). Quantification of Compound Flooding over Roadway Network during Extreme Events for Planning Emergency Operations. *Natural Hazards Review*, 23(2), 1–18. [https://doi.org/10.1061/\(asce\)nh.1527-6996.0000524](https://doi.org/10.1061/(asce)nh.1527-6996.0000524)
- Tahvildari, N., & Castrucci, L. (2021). Relative Sea Level Rise Impacts on Storm Surge Flooding of Transportation Infrastructure. *Natural Hazards Review*, 22(1), 1–17. [https://doi.org/10.1061/\(asce\)nh.1527-6996.0000412](https://doi.org/10.1061/(asce)nh.1527-6996.0000412)
- Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, 5(12), 1093–1097. <https://doi.org/10.1038/nclimate2736>
- Wang, B., Liu, S., Wang, B., Wu, W., Wang, J., & Shen, D. (2021). Multi-step ahead short-term predictions of storm surge level using CNN and LSTM network. *Acta Oceanologica Sinica*, 40(11), 104–118. <https://doi.org/10.1007/s13131-021-1763-9>
- Wei, Z., & Nguyen, H. C. (2022). Storm Surge Forecast Using an Encoder–Decoder Recurrent Neural Network Model. *Journal of Marine Science and Engineering*, 10(12). <https://doi.org/10.3390/jmse10121980>

- Werner, A. D., & Simmons, C. T. (2009). Impact of sea-level rise on sea water intrusion in coastal aquifers. *Ground Water*, 47(2), 197–204. <https://doi.org/10.1111/j.1745-6584.2008.00535.x>
- Xie, W., Xu, G., Zhang, H., & Dong, C. (2023). Developing a deep learning-based storm surge forecasting model. *Ocean Modelling*, 182(November 2022), 102179. <https://doi.org/10.1016/j.ocemod.2023.102179>
- Zahura, F. T., & Goodall, J. L. (2022). Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model. *Journal of Hydrology: Regional Studies*, 41(March), 101087. <https://doi.org/10.1016/j.ejrh.2022.101087>
- Zhang, Y. J., Ye, F., Stanev, E. V., & Grashorn, S. (2016). Seamless cross-scale modeling with SCHISM. *Ocean Modelling*, 102, 64–81. <https://doi.org/10.1016/j.ocemod.2016.05.002>
- Zoph, B., & Le, Q. V. (2018). Searching for activation functions. *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 1–13. <https://doi.org/https://doi.org/10.48550/arXiv.1710.05941>
- Zust, L., Fettich, A., Kristan, M., & Licer, M. (2021). HIDRA 1.0: Deep-learning-based ensemble sea level forecasting in the northern Adriatic. *Geoscientific Model Development*, 14(4), 2057–2074. <https://doi.org/10.5194/gmd-14-2057-2021>