Characterizing the Evolution of Extreme Water Levels with Long Short-Term Memory Station-based Approximated Models and Transfer Learning Techniques

Samuel Daramola[1], Paul Muñoz[2], Jennifer Irish[1], Siddharth Saksena[1], David F. Muñoz[1]

[1]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24060, United States

[2]Department of Water and Climate, Vrije Universiteit Brussel (VUB), 1050 Brussels

Correspondence to: Samuel Daramola (samueldaramola@vt.edu) and David F. Muñoz (davidmunozpauta@vt.edu)

Abstract

Extreme water levels (EWLs) resulting from tropical and extratropical cyclones pose significant risks to coastal communities and their interconnected ecosystems. To date, physically-based models have enabled accurate characterization of EWLs despite their inherent high computational cost. However, the applicability of these models is limited to data-rich sites with diverse morphologic and hydrodynamic characteristics. The dependence on high quality spatiotemporal data, which is often computationally expensive, hinders the applicability of these models to regions of either limited or data-scarce conditions. To address this challenge, we present a computationally efficient deep learning framework, employing Long Short-Term Memory (LSTM) networks, to predict the evolution of EWLs beyond site-specific training stations. The framework, named LSTM-Station Approximated Models (LSTM-SAM), consists of a collection of bidirectional LSTM models enhanced with a custom attention layer mechanism embedded in the model architecture. Moreover, the LSTM-SAM framework incorporates a transfer learning approach that is applicable to target (tide-gage) stations along the U.S. Atlantic Coast. The LSTM-SAM framework demonstrates satisfactory performance with “transferable” models achieving average Kling-Gupta Efficiency (KGE), Nash-Sutcliffe Efficiency (NSE), and Root-Mean Square Error (RMSE) ranging from 0.78 to 0.92, 0.90 to 0.97, and 0.09 to 0.18 at the target stations, respectively. Following these results, the LSTM-SAM framework can accurately predict not only EWLs but also their evolution over time, i.e., onset, peak, and dissipation, which could assist in large-scale operational flood forecasting, especially in regions with limited resources to set up high fidelity physically-based models.

Keywords: long short-term memory networks, transfer learning, extreme water level, tropical cyclones

1. Introduction

About 11% of the world's population (890 million people) currently resides in low-lying areas, and according to the Intergovernmental Panel on Climate Change, this number is projected to exceed 1 billion by the year 2050 (Pörtner et al. 2019; Glavovic et al. 2022). Low-lying areas are particularly vulnerable to weather and climate disasters which are responsible for severe socioeconomic and environmental impacts (Zscheischler et al. 2020; Rainey et al. 2021). The United States, accounting for 1.6% of the current global population (129 million people) in low-lying areas (Office for Coastal Management, 2024), has reported more than 377 weather and climate disasters since 1980 (NOAA-NCEI 2024). In the same period, total reported losses exceed \$2.67 trillion when adjusted for the 2024 Consumer Price Index (NOAA-NCEI 2024). Among these disasters, six of the world’s costliest hurricane events resulted in over \$50 billion in damages in the United States (Douris and Kim 2021; Sanders et al. 2022). Hurricanes (or tropical cyclones) are responsible for coastal flood hazards characterized by extreme water levels (EWLs) and exacerbated by climate-related impacts, including regional sea level rise and anthropogenic activities (Hino and Nance 2021; Khojasteh et al. 2021).

EWLs in estuarine and coastal systems arise from various flood drivers, including precipitation, river discharge, storm surge, tides, and waves. Yet, these drivers do not necessarily act in isolation but rather synergize resulting in compound flooding (Wahl et al. 2017; Muis et al. 2019; Parker et al. 2023). Compound flood (CF) hazards, and their associated risk to coastal communities, are particularly severe when flood drivers co-occur or unfold in close succession (Arns et al. 2020; Almar et al. 2021). For example, storm surge can co-occur with extreme precipitation events during tropical cyclones (TCs) (Wahl et al. 2015; Bevacqua et al. 2019), high tide can coincide with the peak of a storm surge (Thomas et al. 2019; Marsooli and Wang 2020), peak river flow and storm surge can co-occur along estuarine systems (Moftakhari et al. 2019; Muñoz et al. 2020), and waves and storm surge can interact nonlinearly (Rueda et al. 2016; Serafin et al. 2017a); thereby amplifying the effects of CF events. It has been noted that changes in storminess would also play a major role in future EWLs (Santiago-Collazo et al. 2019). This is corroborated by the increasing frequency and intensity of TCs along with the rise of sea levels and ocean temperatures over the past 35 years (Anderson et al. 2021; Ghanbari et al. 2021; Bloemendaal et al. 2022).

TCs have been responsible for 60% of flood-induced population displacements in the United States (1985 to 2021), especially in densely inhabited coastal cities along the Gulf of Mexico and the Atlantic Coast (Brakenridge 2021; Tate et al. 2021; Wing et al. 2022). Recognizing the heightened risks to coastal communities, it has become imperative for researchers and practitioners to rely on either physically-based or data-driven modeling approaches to characterize EWLs in terms of peak magnitude and timing. Physically-based models are commonly used to estimate EWLs based on simplified hydrometeorological processes governed by the conservation of mass and momentum equations (Santiago-Collazo et al. 2019; Bates 2023). The accuracy of these models depends on the availability and quality of several spatiotemporal datasets to appropriately characterize input and forcing conditions, topography and bathymetry, land surface roughness, and other key morphologic characteristics (Jafarzadegan et al. 2021; Alipour et al. 2022; Bates 2022a). Nevertheless, such models are often constrained by limited spatial scope and/or high computational demands necessary to solve large-scale flood dynamics (Bilskie et al. 2021; Muñoz et al. 2021). While physically-based models developed with a lower spatial resolution (e.g., cell-grid size and digital elevation model resolution) can cover broader areas and reduce computational time, they can lead to less accurate predictions due to a lack of detailed spatiotemporal information around key morphological and hydrodynamic variables in narrow tidal inlets and river channels (Saksena and Merwade 2015; Fraehr et al. 2022).

In contrast, data-driven models such as neural networks (NN) can discern intricate or hidden patterns in large datasets and predict storm surges and EWLs with reduced computational demands when compared to those of physically-based models (Muñoz et al. 2021). Importantly, data-driven models offer rapid and efficient forecasting solutions at large scales (Lee et al. 2021; Hamitouche and Molina 2022; Hamidi et al. 2023) and have the ability to generalize or identify patterns from the data they are trained on. In addition, these models can be updated over time which improves their predictions as more and new information becomes available. The fact that NN models are inherently adept at capturing nonlinear associations in complex systems makes them reliable candidates for EWL prediction (Tedesco et al. 2023). Particularly, deep learning approaches like long short-term memory (LSTM) networks, a variant of the recurrent neural network (RNN), learn nonlinear relationships and patterns from sequential time-series data, to enhance prediction accuracy in hydrological and coastal contexts (Li et al. 2021; Zhang et al. 2022).

A growing body of research is demonstrating that LSTM networks can predict EWLs and aid in flood susceptibility assessments, barrage integrity, riverine flood level forecast, and surge prediction (Tiggeloven et al. 2021; Fang et al. 2021; Kardhana et al. 2022; Merizalde et al. 2023a; Liu et al. 2023). LSTM networks are designed to recognize sequence-to-sequence patterns and selectively retain information over time, which in turn enhances its predictive accuracy by utilizing memorized patterns (Hewamalage et al. 2021; Lindemann et al. 2021). On the global scale, LSTM networks outperformed other NN models designed for surge prediction at 92% of 1,276 tide stations across regions of Europe, Africa, Australia, the Pacific, and the United States (Tiggeloven et al. 2021). In addition, LSTM networks can be integrated with spatial NN algorithms to improve the modeling of geographical correlations (Gavahi et al. 2021) and even designed to capture spatial flood characteristics (Fang et al. 2021). Unlike physically-based models that are typically confined and developed using site-specific information, LSTM networks can be trained using geographical characteristics, morphological and hydrodynamic features, and forcing drivers at different scales. The resulting learned patterns can then be generalized and applied to neighboring regions through transfer learning (TL) techniques.

TL addresses the challenge of either data scarce or insufficient training data by leveraging gained knowledge from data-rich training domains and applying it to other (target) domains that share similar characteristics or features (Shen 2018; Tan et al. 2018; F. Zhuang et al. 2021). Moreover, TL can be used to expedite decision-making processes and circumvent time constraints associated with the development and training of NN models. Several studies have implemented TL techniques in NN models to estimate urban flood levels (Zhao et al. 2021; Seleem et al. 2023), predict significant wave height (Obara and Nakamura 2022), conduct land cover mapping (Mahdianpari et al. 2018), and compound flood hazard characterization of nearby regions to the training domain (Muñoz et al. 2021). However, creating a data-driven model with effective generalization capabilities beyond its training domain still remains a significant hurdle (Bentivoglio et al. 2022; Bates 2022b). While maintaining consistency in location enhances the accuracy and lead-time of model predictions (Altunkaynak and Kartal 2021), this limits the geographical areas suitable for effectively applying TL techniques. Therefore, NN models should learn patterns from nonlinear interactions among inputs features and further benefit from mechanisms that ensure accurate model predictions at target domains.

In the present study, we introduce a comprehensive framework that (i) accurately predicts the evolution of extreme water levels beyond training domains, and (ii) addresses the underlying limitations attributed to transfer learning techniques. The proposed framework, named LSTM - Station Approximated Models (LSTM-SAM), achieves these two objectives by gathering learned patterns from neighboring tide-gage stations of the U.S. Atlantic Coast and optimizing the LSTM models with an attention layer mechanism during the training phase. The remainder of the manuscript continues as follows. Section 2 presents the study area, data availability, data processing, and model architecture. Results of the proposed LSTM-SAM framework are shown in section 3 and discussed in section 4. Lastly, section 5 presents the conclusions of this study as well as future work.

2. Methods

2.1 Study area

The proposed LSTM-SAM framework is trained using time-series data from 5 strategically selected tide-gage stations located along the U.S. Atlantic Coast. These stations are Atlantic City, NJ (NOAA ID: 8534720), Sewells Point, VA (8638610), Beaufort, NC (8656483), Fort Pulaski, GA (8670870) and Trident Pier, FL (8721604) (Figure 1). The training stations are selected based on two criteria: (i) they have been hit by hurricane events within a radius of 60 km of the landfall location, and (ii) they have over 70% consecutive water level (WL) data spanning at least 40 years. The latter ensures that the training stations contain EWLs attributed to either TCs (hurricanes) or extra-TCs (Nor'easter winter storms) to effectively train and validate the LSTM-SAM framework. We then implement a TL approach in the framework and transfer nonlinear patterns from training to target stations in order to predict the evolution of EWLs. Most of the target stations are directly exposed to the Atlantic Ocean and located in-between the training stations (Figure 1). Those stations include: 1) Montauk, NY (NOAA ID: 8510560), 2) Sandy Hook, NJ (8656483), 3) Lewes, DE (8557380), 4) Ocean City, MD (8570283), 5) Kiptopeke, VA (8632200), 6) Duck, NC (8656483), 7) Oregon Inlet Marina, NC (8652587), 8) USCG Station Hatteras, NC (8654467), 9) Wrightsville Beach, NC (8658163), 10) Springmaid Pier, SC (8661070), 11) Charleston, SC (8665530), 12) Mayport, FL (8720218), 13) Lake Worth Pier, FL (8722670), and 14) Virginia Key, FL (8723214).

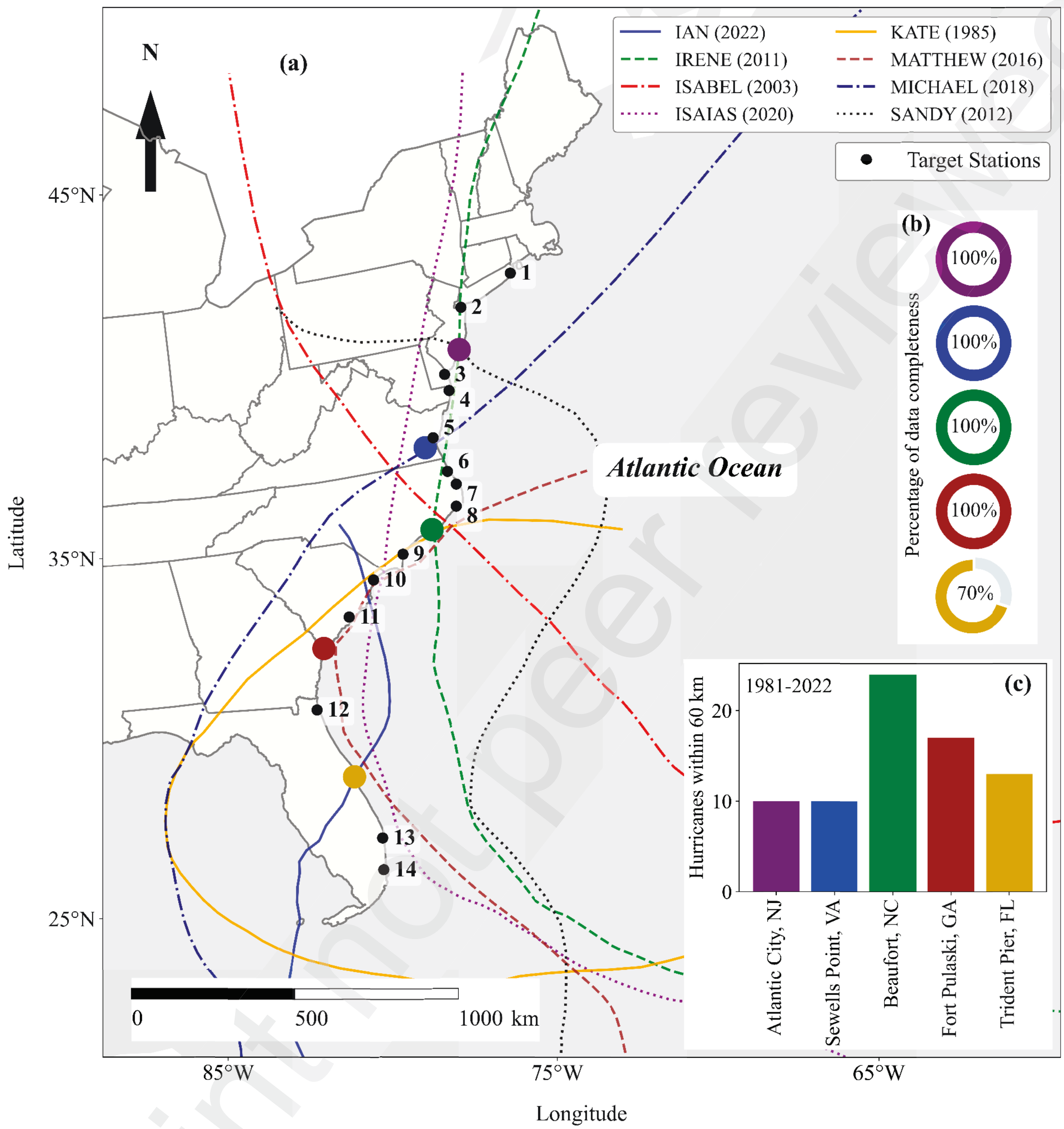

Figure 1: Location of training and target stations along the U.S. Atlantic Coast. (a) Selected training and target stations (numbered from 1 to 14) are shown with colored and black circles, respectively. (b) For each training station, percentage of water level data completeness obtained from the NOAA's Tides & Current portal. (c) Relevant hurricane's best tracks within a 60 km radius of the hurricane's landfall locations.

2.2 Data availability

We retrieve WL data from the National Oceanic and Atmospheric Administration (NOAA)'s Tides & Currents portal (https://tidesandcurrents.noaa.gov/map/index.html) and complement these with legacy data from the University of Hawaii Sea Level Center (https://uhslc.soest.hawaii.edu/data/?rq), particularly for stations where WL records are not available. Meteorological and wave data are obtained from the European Centre for Medium-Range Weather Forecasts Reanalysis dataset (ERA5, version 5) produced by the Copernicus Climate Change Service (https://cds.climate.copernicus.eu/). ERA5 dataset has a spatial resolution of 31 km that allows for accurate representation of extreme climate events at large scale including those driven by TCs (Bian et al. 2021). Specifically, we use hourly wind speed and direction at 10 m elevation, atmospheric pressure, sea level pressure, sea surface temperature, air temperature, precipitation, wave direction, and wave height. In addition, we retrieve data from the U.S. Army Corps of Engineers (USACE)'s Wave Information Studies (WIS) portal that provides consistent, hourly, and long-term wave climatology along the U.S. coastlines (https://wisportal.erdc.dren.mil/#). These aforementioned datasets have been successfully applied to other NN models that predict hourly non-tidal residuals at tide stations on a global scale with satisfactory results (Bruneau et al. 2020).

2.3 Data processing

The required data length to effectively train NN models depends on the response time of the system under analysis. For coastal systems, previous studies recommend at least six years of training data consisting of complete consecutive sequences (10 days) in order to achieve consistent proficiency in NN models (Bruneau et al. 2020; Tiggeloven et al. 2021). Following this, we conduct data quality control over the training stations and ensure that the time-series contain complete data sequences to train the LSTM models. Then, we decompose the time-series data of WL into seasonality, trend, predicted tides, and non-tidal residual (NTR) components using the Seasonal-Trend decomposition using LOESS (STL) and Unified Tidal Analysis and Prediction (UTide) packages in Python (Cleveland et al. 1990; Codiga 2011). The STL analysis, adept at time-series analysis for its outlier resilience, flexible seasonal adjustment, and trend adaptability, provides comprehensive insights into long-term and seasonal dynamics (Chen et al. 2020). UTide employs a decision tree algorithm, a recognized method for automatically selecting the most relevant constituents from 147 tidal constituents, and offers tide prediction correction for records spanning up to one full (18.6-year) nodal cycle (Codiga 2011; Tiggeloven et al. 2021; Tedesco et al. 2023). We consider a window size of 40 days and a time step of 3 days for time-series decomposition in order to ensure that at least one full lunar cycle is covered (Figure S1, Supplementary material), including both spring and neap tides and the independence of large storm events by selecting the maximum NTR on a stepped basis (Serafin and Ruggiero 2014; Rashid et al. 2024; Moftakhari et al. 2024). The time-series decomposition aids to improve deep learning by distinguishing clear, recurring patterns from irregular variations; thereby refining the models' ability to learn from the data and enhancing the accuracy of their predictions (Parker et al. 2023).

In addition to the WL components, we extract meteorological and wave data from the closest grid pixel of ERA-5 dataset to tide-gage stations. For this, we calculate the minimum square difference between the latitudes and longitudes of the data points and the specified location, that is, within a radius of 15.5 km. Next, we use the time-series of WL components, meteorological, and wave data as relevant input features to the LSTM-SAM framework in order to predict the target variable (e.g., EWLs and their evolution over time). Both input and target variables are first scaled using the "minmaxscaler" function from the sklearn library in Python. This technique normalizes the range of multisource data and ensures that all features have an identical scale, typically between 0 and 1 (de Amorim et al. 2023). Also, we create a RNN dataset function that preprocesses data for LSTM by taking normalized input features (X_norm), corresponding target values (y_scaled), and a specified look-back period to construct a dataset suitable for sequence prediction. Moreover, we consider two look-backs of 6 and 24 hours to train the LSTM and evaluate the effects of different time steps on the model's prediction performance, i.e., the number of previous time steps in hours used to predict the next time step.

X_norm and their corresponding y_scaled are sequentially split into a training size of 80% and a testing size of 20%. Finally, we capture the evolution of EWLs in the training and testing datasets by focusing our analysis on historic hurricane events and Nor'easter winter storms within a 7-day window centered around the peak WL. Results of a sensitivity analysis show that longer time-windows favor model's performance metrics due to multiple non-extreme WLs being accounted for, whereas shorter ones could not effectively capture the evolution of EWLs across all stations as they focused more on the peak WL.

2.4 Model architecture

The LSTM-SAM framework consists of bidirectional LSTM (Bi-LSTM) models that are garnering significant interest within the domain of WL prediction (Bai and Xu 2021; Fang et al. 2021; Zhang et al. 2022). Unlike traditional LSTM models that only rely on previous timesteps, the advantage of Bi-LSTM models is that input sequences are processed in both forward and backward directions (Equations 1 to 13 in the Supplementary material). The reader is referred to the study of Ahmed et al. (2022) for a more detailed explanation of Bi-LSTM. This dual viewpoint helps identify and learn features that may not be apparent when the sequence is analyzed in only one direction.

2.4.1 Bidirectional LSTM network models

The model architecture consists of three LSTM units that are set to vary in intervals of 32 since higher units tend to increase computational complexity (Figure 2a). We consider a "L2" regularization method to prevent overfitting by penalizing large weights. Also, we include dropout rates in the model architecture to prevent overfitting by randomly disabling a subset of neurons during the training process, thereby allowing the LSTM network to develop a more generalized understanding of the data and improve its performance on new and/or unseen data. The LSTM units use the "tanh" and hard "sigmoid" recurrent activation functions. In addition, we add a standard dense (fully connected) layer to output the final prediction. The loss functions consist of both mean absolute error (MAE) and mean squared error (MSE) for different variants of the models with the "Adam" optimizer as suggested in similar WL prediction studies (Huang et al. 2020). We reserve 30% of the training data for validation of the model's learning ability during the training process. An early stopping callback is also employed to monitor the validation loss, stop training if no improvement is observed for five consecutive epochs, and ultimately prevent overfitting and/or unnecessary computations.

We consider two training strategies in the LSTM algorithm: (i) train-test (TT) split, and (ii) time-series cross-validation (CV) split. The first method involves a single split into training (80%) and testing sets (20%) whereas the second method includes multiple training and testing sets created sequentially for a more comprehensive evaluation of the model's performance across different periods (Figure 2b and 2c). Unlike traditional CV strategies, here CV fold does not shuffle the data and therefore keeps the time sequence invariant (Kingphai and Moshfeghi 2022). We consider 10-fold CV to check the model's performance and potentially improve the prediction accuracy. Moreover, the LSTM model is trained using the training set for that specific split for each iteration of the loop. As the loop progresses, the size of the training set increases whereas the validation set consists of data points that come after the training set in time. As a result, the training and validation process involves learning from past data and validating the model's performance on unseen future data, respectively. Once all splits are processed, the final model is trained using the entire dataset.

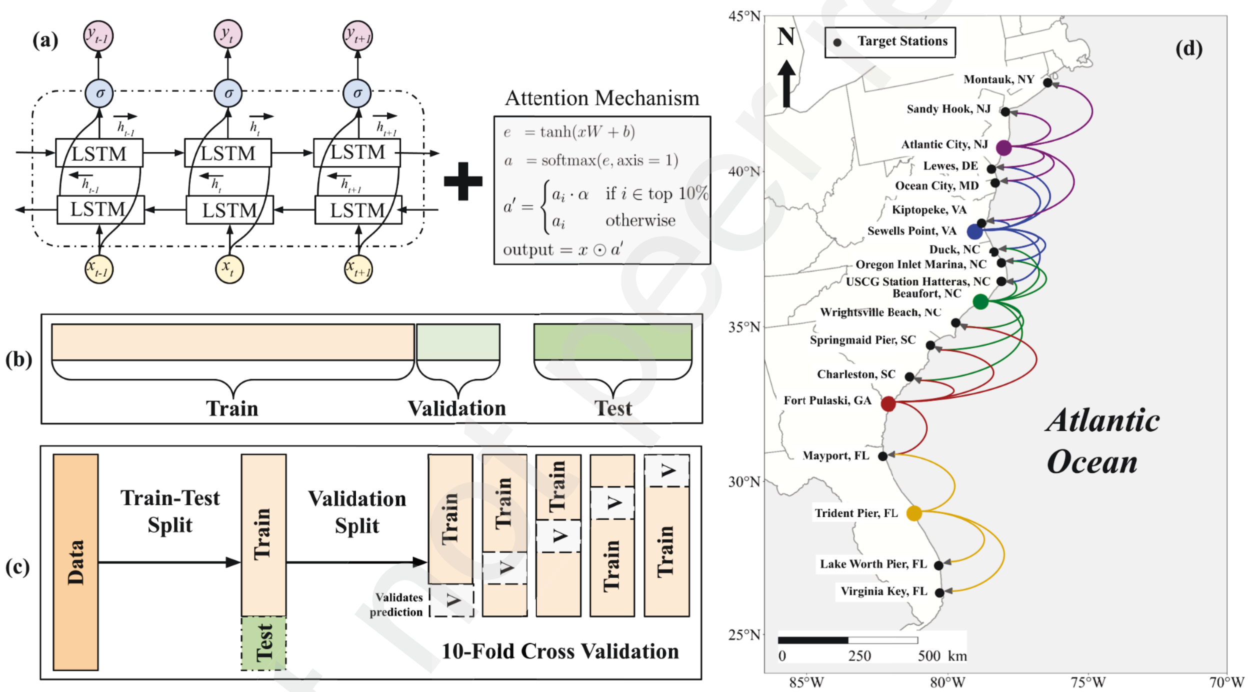

Figure 2. Schematic of model architecture, training, validation, and transfer learning approach. (a) Bidirectional LSTM network model with an attention layer mechanism to improve pattern recognition. The model employs two data training approaches: (b) train-test split, and (c) time-series cross-validation for model development. (d) Transfer learning approach to predict extreme water level evolution at target stations (black circles) using models developed in the closest training stations (colored circles).

2.4.2 Hyperparameter tuning

We conduct hyperparameter tuning to identify optimal values of LSTM units, dropout rate, and learning rate within specified ranges to train the models (Table 1). We set the tuner search to a maximum of 300 trials, after which the best hyperparameters are used to train the models. The model architecture relies on a Bayesian optimization technique for hyperparameter tuning that inherently functions in a sequential manner and leverages data from previous evaluations to inform subsequent runs (Wang et al. 2023). Such technique efficiently balances the exploration of new areas in the hyperparameter space with an emphasis on known suitable regions. This is particularly useful when each training iteration is computationally intensive since the aforementioned optimization technique can identify optimal hyperparameters with less time than methods like grid or random search (Marco et al. 2022). Additionally, its capacity to handle high-dimensional hyperparameter spaces and integrate prior knowledge about potential hyperparameters makes it a versatile choice (Bischl et al. 2023). Its proven success in real-world applications and its efficiency in finding robust hyperparameters with limited evaluations position it as a top choice for many practitioners (Wang et al. 2023).

Hyperparameters, identified through a rigorous tuning (or calibration) process on site-specific training data, tend to yield models that perform optimally within particular training domains. However, these models may not necessarily exhibit the same level of effectiveness across other target domains, even if both domains share similar morphological and hydrodynamic characteristics. We address this limitation by focusing on the range of values considered for hyperparameters. First, we define the values for specific parameters such as batch sizes (32, 64, 128, and 256), look-back times (6 and 24), loss functions (MAE and MSE), and data training strategies (TT and CV), while allowing other parameters to be determined through hyperparameter tuning. We then ensure that combinations of options from the aforementioned parameters occur precisely once, which in turn facilitates the creation of distinctive models with a unique set of hyperparameters across the five training stations (Figure 2).

Hence, this tuning process produces a set of suitable Bi-LSTM models with comparable performance for a given training station. We evaluate model's performance using several metrics that are recommended for models predicting WL dynamics (Abbaszadeh et al. 2020; Lee et al. 2021; Muñoz et al. 2022a). Those include the coefficient of determination (R^2), Mean Bias Error (MBE), Root Mean Square Error (RMSE), Kling-Gupta Efficiency (KGE) (Gupta et al., 2009), and Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). Based on the evaluation metrics, there might be a set of "transferable" LSTM models among the suitable ones from the training station for which inherent pattern recognition capabilities would be adequate for the target stations (Section 2.4.4). However, an increase in tuning trials poses a risk of overfitting, where models become excessively tailored to the training data and lose their predictive ability on new datasets. Here, we specify an appropriate number of tuning trials while employing different combinations of hyperparameters to generate a spectrum of suitable models for a training domain (Table 1). Specifically, we develop a total of 32 models at each training.

Table 1: Range of values considered for hyperparameter tuning.

Hyperparameter (3 LSTM units)	**Range of Tested Values**
LSTM Units	32 - 512 (step of 32)
Dropout Rate	0.10- 0.50 (step of 0.1)
Activation	Tanh, Sigmoid
L2 Regularization	1e-6 - 1e-3 (log sampling)
Learning Rate (Adam Optimizer)	1e-4 - 1e-2 (log sampling)
Maximum number of trials	300

Batch Size (b)	32 - 256 (step of 32)
Loss Function	MAE, MSE
Look-back time (h)	6, 24
Epochs	500 (with early stopping)
Validation Split	30% of training data

2.4.3 Attention layer mechanism

Attention layers are used to address inherent limitations of conventional RNNs (and LSTM) such as the tendency to lose information from earlier segments of extended sequences and difficulties to train models in areas of sharp and extreme changes (Rithani et al. 2023). These layers scan through the data, identify key features, and increase their influence in the training process. Here, we incorporate an attention layer in the model architecture for better model generalization during the training process. Specifically, the attention layer computes attention scores for each time step using a weight matrix (Glorot) and bias (zeros) initialization (Equations 14 to 15 in the Appendix). In addition, we customize this layer using a factor that amplifies the top 10% of the attention scores (Equation 16). This factor allows the model to focus more on crucial parts of the sequence, which could be abrupt changes of high or low levels in the data.

The choice of the Glorot initializer for the weight matrix in the attention layer is appropriate due to the use of tanh and sigmoid activation functions in the LSTM units (Glorot and Bengio 2010; Evangelista and Giusti 2021). The initializer keeps the scale of the gradients approximately the same in all layers of the LSTM network. Starting with zero biases ensures that all neurons in a layer initially produce outputs of roughly the same magnitude, which can be a good starting point for symmetric activation functions like tanh.

2.4.4 Transfer learning

We conduct TL to predict the evolution of EWLs at selected target stations by leveraging "gained knowledge" from the closest training stations (Figure 2d). Such knowledge includes hidden sequential patterns and nonlinear associations among input and target data features that are stored as model weights (Muñoz et al. 2021; Zhao et al. 2021). Note that most of the target stations are located in-between two training stations; except Montauk, NY and Sandy Hook, NJ as well as Lake Worth Pier, FL and Virginia Key, FL that are close to a single training station. Here, the TL approach consists in training, tuning, and validating Bi-LSTM models at each training station and subsequently saving the corresponding model weights. We leverage all available Bi-LSTM models (e.g., 32 models) and transfer them to the target stations in order to predict the evolution of EWLs within the predefined 7-day window (Section 2.3). Among these models, we identify "transferable" Bi-LSTM models based on the criteria that both KGE and NSE are above a threshold value of 0.70 at the target stations. This threshold ensures that each transferable model adequately accounts for the magnitude and timing of EWLs while also keeping its inherent pattern recognition capabilities on new input data that has not yet been observed at the target stations (e.g., those associated with future extreme events). Lastly, we assess the performance of transferable models at the target stations with identical evaluation metrics including R^2, MBE, RMSE, KGE, and NSE.

3. Results

3.1 Assessment of bidirectional LSTM network models

We first assess the performance of Bi-LSTM models at the training stations with and without the attention layer mechanism incorporated in the model architecture. For this, we consider the models' ability to capture: (i) peak and timing of EWLs, and (ii) evolution of EWLs for selected historic TCs (hurricanes) and extra-TCs (Nor'easter winter storms).

3.1.1 Extreme water levels

For convenience, we will focus on model predictions obtained from Fort Pulaski (FP), GA since this training station has been reporting complete consecutive hourly data from 1975 to present (Figure 3). Predictions of EWLs and associated RMSE, average peak error, and KGE metrics suggest that Bi-LSTM models with the attention layer (e.g., hereinafter referred to as Bi-LSTM-ATT) can capture the magnitude and timing of peak WLs with a higher accuracy than those without this layer (Figure 3b). Note that this includes the two most EWL events contained in the testing set (e.g., Hurricane Matthew (2016) and Hurricane Irma (2017)). More than half of Bi-LSTM models achieve a median RMSE, absolute peak error, and mean bias of 0.18 m, 0.68 m, and -0.11 m, respectively (Figure 3c to 3f and Table S1). Also, these models achieve very low to moderate performances with a median KGE and NSE of 0.50 and 0.15, respectively (Table S1). In contrast, the models' performance substantially improves after integrating the attention layer mechanism in the model architecture. In that regard, half of Bi-LSTM-ATT models show a reduction in the median RMSE, absolute peak error, and mean bias by 27% (0.13 m), 36% (0.44 m), and 55% (-0.05 m) with respect to the Bi-LSTM models only (Figure 3g to 3j and Table S2). Also, these models achieve a median KGE and NSE of 0.67 and 0.56, respectively (Table S2).

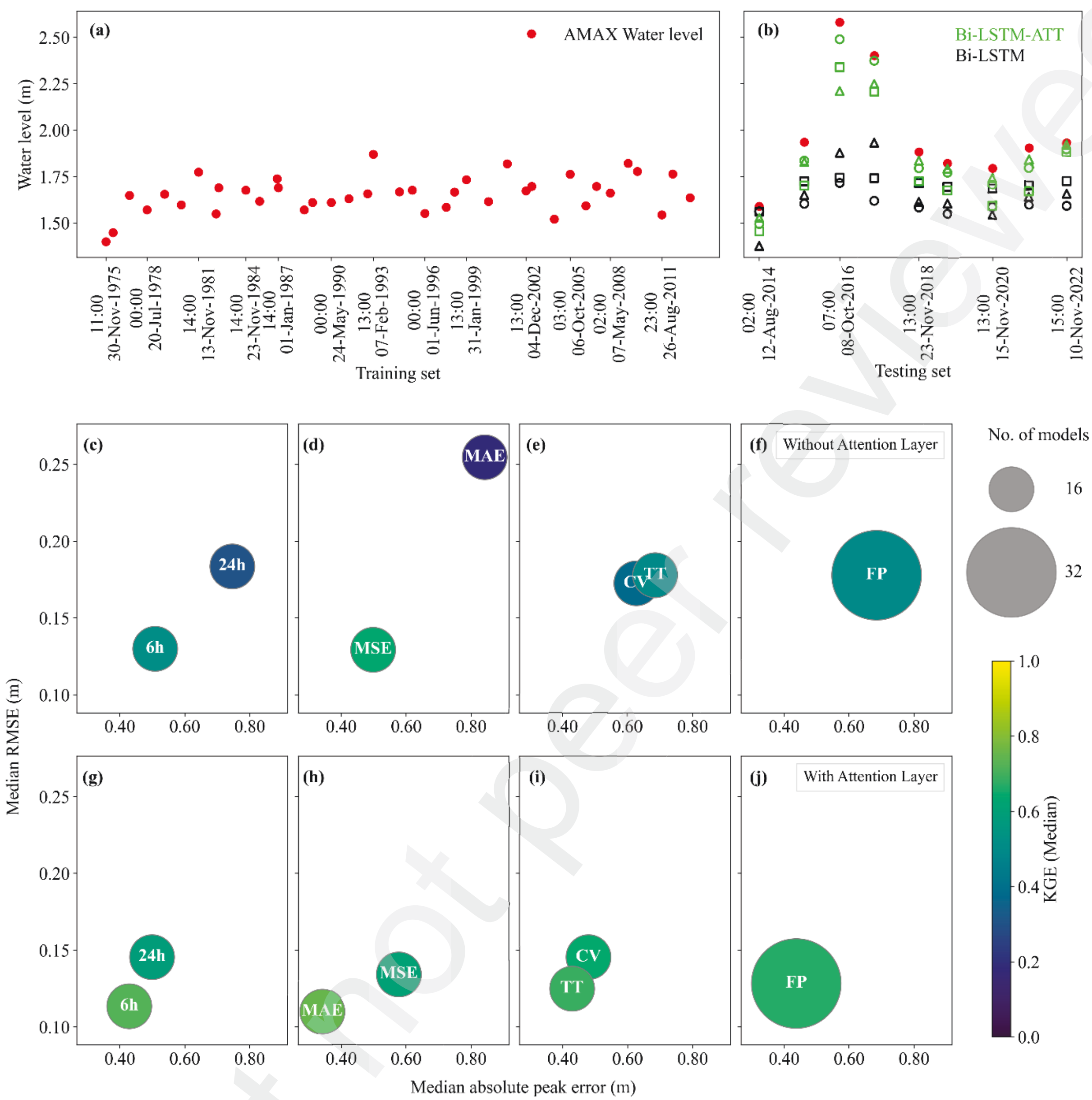

Figure 3. Assessment of models' performance in the testing set at Fort Pulaski (FP), GA. (a, b) Annual maximum water levels in the training set (80%), and testing set (20%) in addition to predictions with/out the attention layer mechanism. The Bi-LSTM and Bi-LSTM-ATT models are categorized into (c, g) look-back times of 6 h and 24 h, (d, h) loss functions focused on Maximum Absolute Error (MAE) and Mean Square Error, (MSE) (e, i) Train-Test (TT) and Cross Validation (CV) fold strategies, and (f, j) a collective unit to compare their overall performance.

At the remaining training stations, we evaluate predictions of EWLs derived from Bi-LSTM-ATT models only (Table S2). The models at Atlantic City (AC), NJ station achieve satisfactory model performance with a median KGE of 0.84 despite the relatively low median NSE of 0.36. The median RMSE and mean bias in this training station are 0.07 and 0 m, respectively. Training station Sewells Point (SW), VA has the best performing models with a relatively high median KGE and NSE of 0.92 and 0.80, respectively. These models achieve a median RMSE and mean bias of 0.05 m and -0.02 m, respectively. Similarly, results at Beaufort (BF), NC station perform satisfactorily with median KGE and NSE of 0.86 and 0.69, respectively. Also, this training station shows a median RMSE and mean bias of 0.05 m and -0.01 m, respectively. Lastly, the models of training station Trident Pier (TD), FL achieve a moderate to satisfactory performance with median KGE and NSE of 0.78 and 0.51, respectively. The models show a median RMSE and mean bias of -0.07 m and -0.03 m, respectively.

3.1.2 Evolution and peak of extreme water levels

Next, we assess the performance of Bi-LSTM-ATT models to predict the evolution of EWLs including the magnitude and timing of the peaks (Figure 4). The top-two Bi-LSTM-ATT models of AC predict the evolution of EWLs associated with a Nor'easter storm with relatively high accuracy. These models achieve KGE of 0.89 and 0.82 and NSE of 0.96 and 0.94 (Figure 4a). In contrast, the top-two models of SW capture the evolution of EWLs and almost perfectly match the magnitude of the peak associated with Hurricane Joaquin (2015) (Figure 4b). These models achieve KGE of 0.98 and 0.98 and NSE of 0.99 to 0.99. Similarly, the top-two models of BF achieve satisfactory KGE of 0.96 and 0.95 and NSE of 0.98 and 0.98. These models capture the evolution and almost perfectly match the magnitude of the peak associated with Hurricane Dorian (2019) (Figure 4c). The top-two models of FP can predict the evolution of EWLs but underpredict the peak associated with Hurricane Matthew (2016) (Figure 4d and 4f). These models achieve KGE of 0.87 and 0.85 and NSE of 0.91 and 0.88.

Lastly, the top-two models of TD capture the evolution of EWLs even though the second model overpredicts the magnitude of the peak associated with Hurricane Nicole (2022) (Figure 4e). These models achieve KGE of 0.72 and 0.70 and NSE of 0.84 and 0.93. Regarding the peak time difference (Figure 4f), all Bi-LSTM-ATT models developed for AC and BF stations show a 1-h lead difference and a perfect match with respect to the observed peak, respectively. Six models of SW station show a 1-h lag difference whereas two models of TP station show a 1-h lead difference with respect to the observed peak. Thirteen and ten models developed for FP stations show a 1-h lag difference and 1-h lead difference with respect to the observed peak, respectively. Overall, the time difference between observed and predicted peak WL is ±1 h for all trained models.

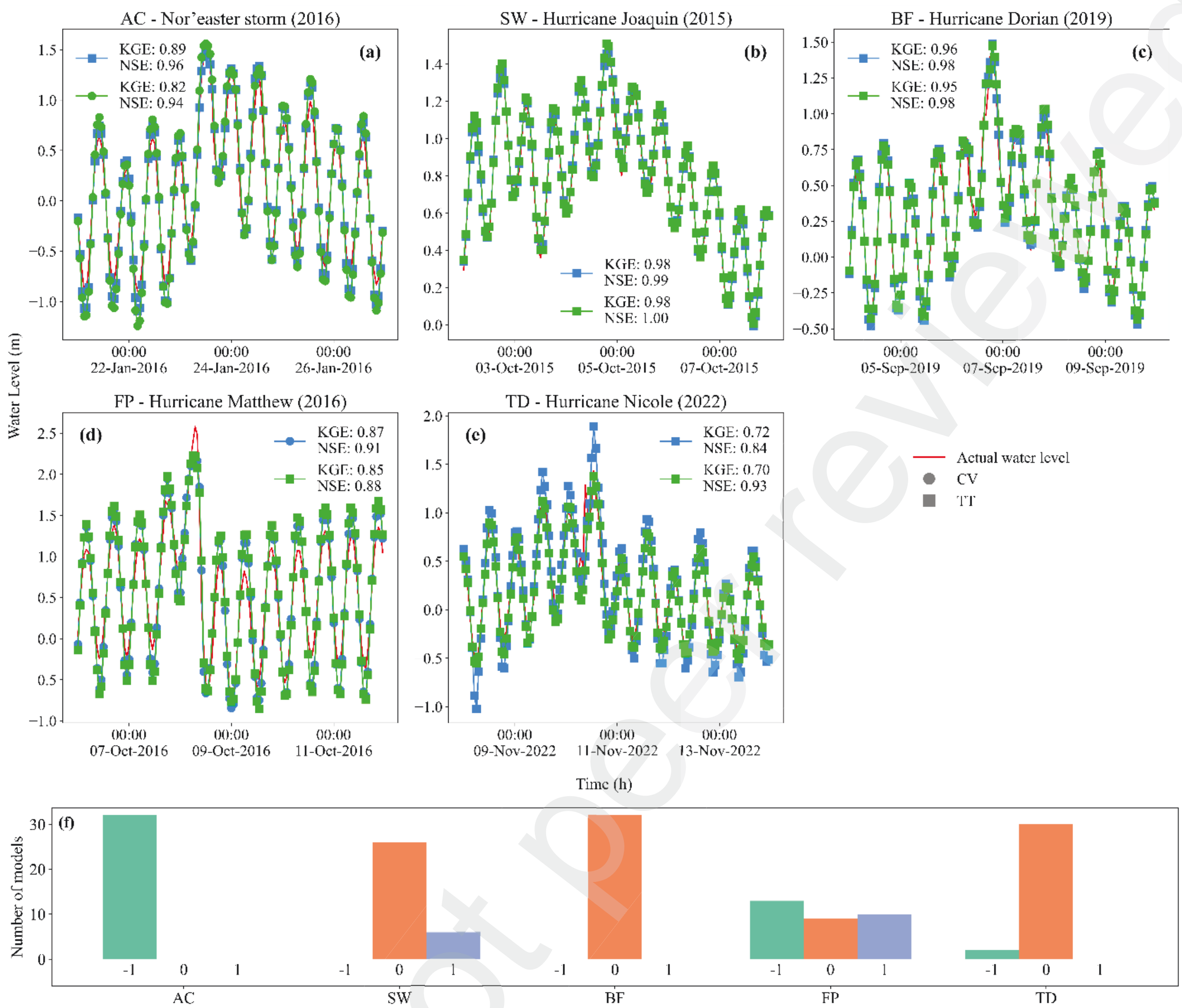

Figure 4. Assessment of models' performance to predict the evolution and peak of extreme water levels in the testing set. Predictions of the top-two Bi-LSTM-ATT models from the Train Test (TT) and/or Cross Validation (CV) split strategies at (a) Atlantic City, NJ (AC), (b) Sewells Point, VA (SW), (c) Beaufort, NC (BF), (d) Fort Pulaski, GA (FP), and (e) Trident Pier, FL (TD). (f) Observed and predicted peak time differences among the five training stations and 32 Bi-LSTM-ATT models.

3.2 Assessment of transfer learning approach

After evaluating the models' performance, we proceed with the assessment of transferable models from the closest training stations to target stations. For example, we transfer the pretrained Bi-LSTM and Bi-LSTM-ATT models and their associated model weights from Sewells Point, VA to the target station at Duck, NC (Figure 2d). Since most of the extreme events at Sewells Point, VA, including Hurricane Isabel (2003), Irene (2011) and Sandy (2012), are in the training set, the performance of all Bi-LSTM models are satisfactory, with metrics comparable to Bi-LSTM-ATT models (Table S3). Here, the goal is to predict the evolution of EWLs for relevant extreme events such as Hurricane Isabel (2003) and Dorian (2019) (Figure 5). Based on the threshold value of 0.70 (Section 2.4.4), there are no transferable Bi-LSTM models that can predict the evolution of both storm events. For Hurricane Isabel, the two best models achieve low KGE of -0.98 and -1.03 and NSE of -0.50 and -0.54 (Figure 5a). Although these two models show a better performance for Hurricane Dorian, they still achieve moderate KGE of 0.55 and 0.54 and NSE of 0.43 and 0.41 (Figure 5d). In contrast, the top-two transferable Bi-LSTM-ATT models achieve high KGE of 0.94 and 0.93 and NSE of 0.97 and 0.97 when predicting EWLs triggered by Hurricane Isabel (Figure 5b). For Hurricane Dorian, the top-two transferable Bi-LSTM-ATT models achieve a relatively high KGE of 0.72 and 0.74 as well as NSE of 0.93 and 0.93 (Figure 5e).

Furthermore, we assess the models' performance in terms of R^2 and RMSE for both extreme events and compare model predictions from the best transferable Bi-LSTM and Bi-LSTM-ATT models using a one-to-one plot (Figure 5c and 5f). Bi-LSTM models have poor generalization of EWLs with low predictive accuracy ($R^2 < 0.50$) and high error (RMSE > 0.30). In contrast, the Bi-LSTM-ATT models can predict the evolution of EWLs with a high predictive accuracy ($R^2 > 0.95$) and low error within an acceptable range (RMSE < 0.15 m). In general, RMSEs below 0.20 m are desirable for hurricane storm surge modeling (Muis et al. 2016). Following this analysis, we hereafter present the results derived from Bi-LSTM-ATT models only.

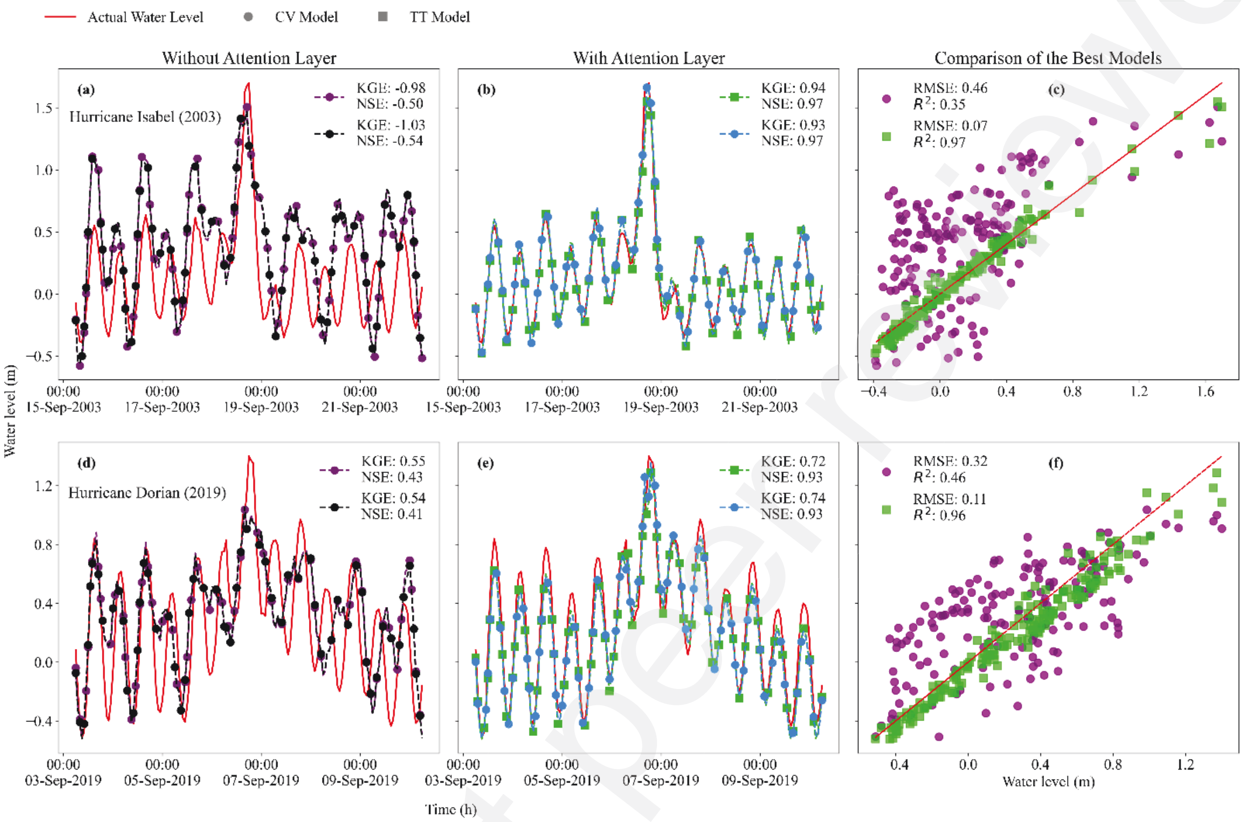

Figure 5. Assessment of transfer learning approach from Sewells Point, VA (training station) to Duck, NC (target station). Prediction of extreme water level evolution using the top-two (a, d) Bi-LSTM, and (b, e) Bi-LSTM-ATT models for Hurricane Isabel (2003) and Dorian (2019). (c, f) One-to-one comparison of top-two model predictive capabilities based on the aforementioned hurricane events.

3.3 LSTM-SAM framework

Once the Bi-LSTM-ATT models and TL approach have been assessed at selected training and target stations (Section 3.1 and 3.2), we introduce the LSTM-SAM framework to accurately predict the evolution of EWLs at target stations in the U.S. Atlantic Coast (Figure 2d). As mentioned before, we set a time window of 7-day centered around the peak to characterize the evolution of EWLs at the target stations and leverage the 32 Bi-LSTM-ATT models developed at each training station. Note that instead of considering the top-two transferable Bi-LSTM models at the target stations (Figure 5), the LSTM-SAM framework identifies a set of transferable models for which KGE and NSE are above 0.70 when evaluated with respect to TCs or extra-TCs events (Figure 6). For practical flood prediction purposes and decision-making support, Bi-LSTM-ATT models achieving the smallest peak deviation among the transferable models are considered as the optimal ones at each target station (e.g., models with the closest prediction to the peak WL).

There are 3 transferable models from training station AC to the target station in Montauk, NY that accurately predict EWL evolution of The Perfect Storm (1991) and Hurricane Sandy (2012) (Figure 6a and 6b). These models achieve satisfactory performances such as an average KGE and NSE above 0.75. Likewise, there are 21 transferable models from AC and SW that predict EWL evolution of Hurricane Isabel (2003) and Sandy (2012) at the target station in Kiptopeke, VA (Figure 6c and 6d). These models achieve an average KGE and NSE above 0.80. At the target station located in Duck, NC, LSTM-SAM identifies 14 transferable models from SW and BF that predict EWL evolution of Hurricane Isabel (2003) and Dorian (2019) (Figure 6e and 6f). These models show satisfactory performances with average KGE and NSE above 0.75. Similarly, the framework identifies 14 transferable models from TD to the target station in Lake Worth Pier, FL that accurately predict EWL evolution of both Hurricane Dorian (2019) and Nicole (2022) (Figure 6g and 6h). The models achieve average KGE and NSE above 0.85.

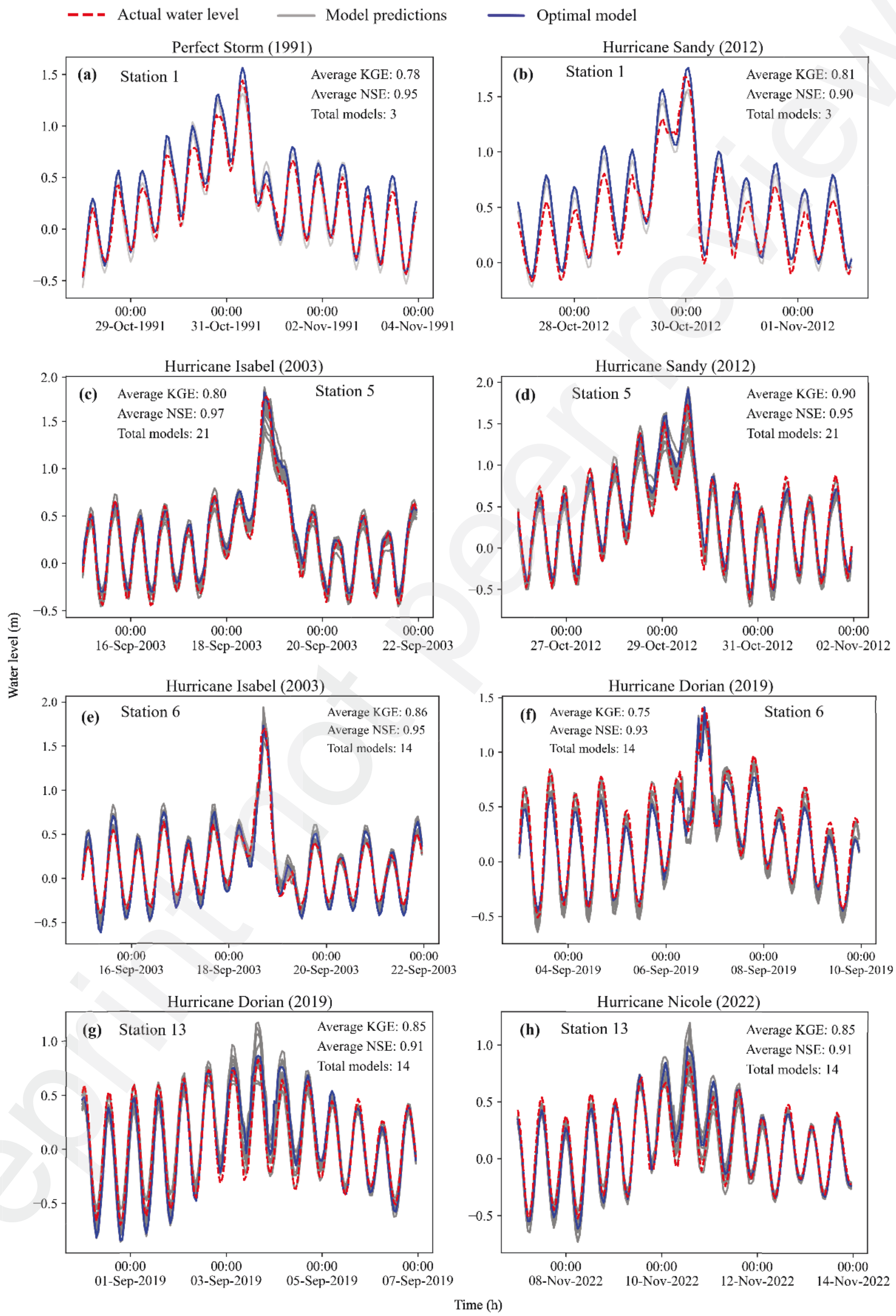

Figure 6. Extreme water level prediction for relevant hurricanes and Nor'easter winter storms in the U.S. Atlantic Coast. Each row panel shows two extreme events at the target stations, the number of transferable models, and their average performance in terms of KGE and NSE. These stations are (a, b) Montauk, NY, (c, d) Kiptopeke, VA, (e, f) Duck, NC, and (g, h) Lake Worth Pier, FL. The dashed red, blue, and gray lines represent observed water levels, optimal model, and water level predictions from all transferable models.

Results of the remaining target stations show average KGE and NSE ranging from 0.70 to 0.99 (Figure S2, Supplementary material). It is worth noting that we leverage WL data from the training station AC to predict the complete evolution of Hurricane Sandy (2012) at the target station in Sandy Hook, NJ (Figure S2b). This demonstrates the transfer model capability to predict EWL evolution even when tide-gauges fail or become inoperative. However, there are three target stations for which no Bi-LSTM-ATT models are completely transferable given the criteria that both NSE and KGE should be above 0.70 within the 7-day window (Section 2.4.4). Specifically, the LSTM-SAM framework does not identify any transferable models from SW and BF stations to the target station in Oregon Inlet, NC that can accurately capture the evolution of EWLs of Hurricane Floyd (1999) and Irene (2011) (Figure 7a and 7b). Similarly, the framework does not identify transferable models from SW and BF stations to USCG Station Hatteras, NC for Hurricane Matthew (2016) and Dorian (2019) (Figure 7c and 7d). Lastly, there are no transferable models from TD station to Virginia Key, Florida for Hurricane Irma (2017) and Nicole (2022) (Figure 7e and 7f). However, note that some Bi-LSTM-ATT models can effectively capture the peak WL within a shorter time window centered around the peak (e.g., ±1 day); hence, the relatively low KGE and NSE at the target stations are explained by an overprediction of WLs occurring before and after the peak WL. Therefore, the LSTM-SAM framework considers the model with the highest KGE and NSE as the optimal models for those three target stations (Figure S2, Supplementary material).

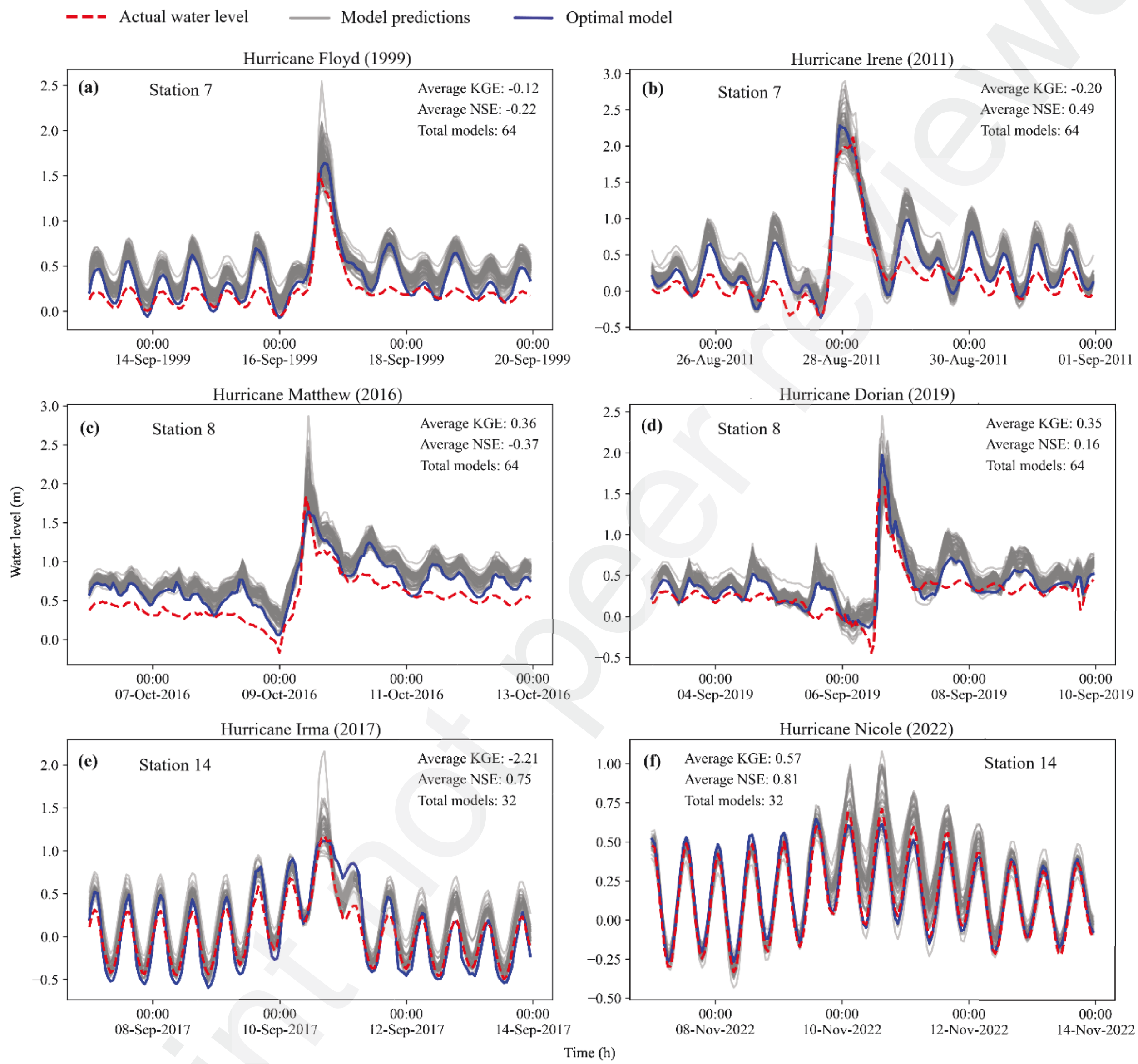

Figure 7. Extreme water level prediction for relevant hurricane events in the U.S. Atlantic Coast. Each row panel shows two extreme events at the target stations, total number of models, and their average performance in terms of KGE and NSE. These stations are (a, b) Oregon Inlet Marina, NC, (c, d) USCG Station Hatteras, and (e, f) Virginia Key, FL. The dashed red, blue, and gray

lines represent observed water levels, optimal model, and water level predictions from all available Bi-LSTM-ATT models developed at the corresponding training stations.

4. Discussion

Physically-based models can accurately predict EWLs; however, they are site-specific and not transferable to other domains, even with similar characteristics, due to their need for detailed topographic and bathymetric (topobathy) data (Santiago-Collazo et al. 2019; Bates 2022b). A feasible alternative to overcome this limitation consists in leveraging *state-of-the-art* deep learning models, such as LSTM networks, given their effectiveness for learning dynamic and/or sequential data including nonlinear interactions and hidden patterns from hydrometeorological input features (Tedesco et al. 2023). Conveniently, LSTM networks enable time-series prediction even in absence of geographical information or catchment characteristics that may remain quasi-invariant for a relatively long time (e.g., average slope, length, width, catchment size, among other input features). Although model transferability is still challenging (Zhao et al. 2021; Kratzert et al. 2024), adequate feature engineering procedures (Merizalde et al. 2023) and improvement of LSTM models' architecture (e.g., attention layer mechanisms) can increase the effectiveness of TL approaches over untrained (target) sites having similar hydrodynamic and morphologic features to the corresponding training areas. This in turn will help advance flood prediction efforts in large scale domains with high accuracy and less computational time (Ding et al. 2020; Li et al. 2021; Nearing et al. 2024). Here we present one of the first applications of TL to EWL prediction in coastal domains, and to our knowledge, the first outside the hydrological context.

4.1 Application of the LSTM-SAM framework

The proposed LSTM-SAM framework is trained on sequential WL data from tide-gage stations and follows a time-series decomposition in order to obtain hidden patterns and nonlinear associations such as seasonality, trend, harmonic tides, and NTR components (Section 2.3). Moreover, these input features are complemented with hydrometeorological data from reanalysis datasets. As a result, the framework is capable of applying learned knowledge to selected target stations and predicting the evolution of EWLs (Section 2.4.1). We ensure that Bi-LSTM networks are robustly calibrated through hyperparameter tuning, which resulted in 32 distinctive models across five training stations and a unique set of hyperparameters per model (Section 2.4.2). We argue that this method preserves inherent pattern recognition capabilities for each model (e.g., model weights) and increases the chances for identifying effective transferable models to target stations.

Conventional Bi-LSTM models show poor performance to predict EWLs at the training station particularly when there exist more extreme events in the test set (Figure 3a and 3b). This is partly due to a sequential training (80%) and test (20%) split given that the frequency and magnitude of EWLs is expected to increase by the end of the century (Santiago-Collazo et al. 2019; Bloemendaal et al. 2022; Boumis et al. 2023). In contrast, leveraging randomly selected subsamples in each batch during the calibration process facilitates a quicker model convergence (De la Fuente et al. 2024) and prevents anomalies in the training data. Nevertheless, training Bi-LSTM models on sequential data ensures that temporal relationships are fully considered in the learning process. Since these models are in general most effective in capturing changing trends of cyclic patterns (Wang et al. 2022), they do ignore some nuanced information of rare and abrupt changes during the training process. Therefore, there is a higher chance of incorrect estimation of equally rare but more extreme data values in testing sets which limits the model's ability to accurately predict EWLs.

Following this, we introduce a custom attention layer in the Bi-LSTM architecture (Section 2.4.3), which significantly improves the models' ability to capture TCs and extra-TC events (Figure 3b). Since hyperparameter tuning is predominantly inclined toward identifying the optimal set of model parameters that result in the most accurate EWL prediction in the training set, there is no guarantee that this optimal set leads to improved performance on unseen (test) sets (Tran et al. 2020). Therefore, by amplifying the top 10% of the attention scores, the model's ability to internally focus on the most relevant time steps is substantially improved and leads to more accurate EWL predictions. Moreover, the attention layer mechanism allows the models to perform identical operations consistently beyond the training set and therefore generalizing unseen data with similar characteristics in the test set. Since Bi-LSTM-ATT outperforms conventional Bi-LSTM models (Figure 3f to 3h), the proposed LSTM-SAM framework demonstrates the ability to effectively predict the evolution of EWLs even when higher WLs attributed to more frequent TCs and extra-TCs are expected.

We also observe comparable performance across all metrics for all Bi-LSTM-ATT models developed on the specified options of lookbacks, batch sizes, loss functions, and data training strategy (Table 2). It is worth noting that a higher batch size of 128 sample points shows the highest performance of all models in terms of KGE and NSE, suggesting that they capture both peak and timing of EWLs with higher accuracy. However, more than half of the models developed on a smaller batch size of 32 sample points have very low NSE scores, which tend toward inaccurate timing of EWLs' predictions. In addition, lower performance for models developed with 24-hour lookbacks compared to 6-hour lookbacks suggests that considering more previous time steps does not necessarily improve the model's predictive accuracy. Furthermore, the identical performance for both data training strategies as well as the loss functions suggests that either of these options are suitable to train the models.

Although most of the trained Bi-LSTM-ATT models can predict EWLs at the nearby target station, we observe that models with KGE and NSE above 0.70 demonstrate robust generalization capabilities of EWL evolution from onset, peak, to dissipation (Figure 4). NSE show a higher accuracy for predictions with correct timing despite the over- or underprediction of WLs (Figure 7e and 7f). In addition, an extended window size renders minor discrepancies in the timing of peak WLs such as 1-h lead or 1-h lag negligible in the overall prediction performance. On the other hand, KGE improve when the magnitude of predictions closely aligns with actual events, even if the prediction timing is inconsistent with actual WL observations (Figure 7c and 7d).

The low performance at target stations located at Oregon Inlet Marina, NC (NOAA ID: 8652587) and USCG Station Hatteras, NC (NOAA ID: 8654467) might be attributed to the geographic location of the tide-gages (Figure 2d). Unlike target stations that are directly exposed to the Atlantic Ocean, these stations are located behind Bodie's and Cape Hatteras' islands of the Outer Banks barrier island chain. Coastal areas surrounding the target stations experience about 1 m of mean tidal range on the ocean side and 0.3 m behind the island (Velasquez-Montoya et al. 2020). In addition, these areas benefit from vast coastal wetlands and protection infrastructure such as the Herbert C. Bonner bridge that alters tidal dynamics and attenuates storm surges and waves (Velasquez-Montoya et al. 2021, 2022). Consequently, these site-specific conditions significantly differ from those of nearby training stations including Sewells Point, VA (NOAA ID: 8638610) and Beaufort, NC (NOAA ID: 8656483).

Nevertheless, Bi-LSTM-ATT models have the potential to capture the peak WL which is crucial for supporting flood emergency response and decision-making. In fact, the optimal models for those target stations correctly capture the peaks of Hurricane Floyd (1999), Irene (2011), Matthew (2016), and Dorian (2019) (Figure 7a to 7d). Similarly, the evolution of EWLs at Virginia Key, Florida (NOAA ID: 8723214) is overestimated (Figure 3e and 3f). Nevertheless, the models can effectively capture the peaks of Hurricane Irma (2017) and Nicole (2022). The tide-gage is located at the entrance of Biscayne Bay close to the Bear Cut bridge, which is most likely responsible for WL being less representative of extreme events. The relatively lower water surface elevation of this station during Hurricane Irma compared to Trident Pier, FL (NOAA ID: 8721604) has also been noted in another study (Alarcon et al. 2022).

4.2 Limitations and future work

The waves and atmospheric variables obtained from ERA5 have a spatial resolution of approximately 31 km. This level of resolution may not correctly account for local variability; hence higher resolution data might improve the performance of LSTM-SAM framework. There are instances where wave components for target stations like Lewes, DE, and Charleston, SC needed to be complemented with those derived from the WIS portal data. As a result, some errors could have been introduced in the input features, reducing the accuracy of EWL predictions. Interestingly, these target stations have the lowest number of transferable models from the training stations (Figure S2). More advanced deep learning models like Transformers have attention mechanisms built into their design (Boussioux et al. 2022) and could be a worth-exploring alternative to the proposed Bi-LSTM-ATT models for predicting EWL evolution in coastal areas. We plan to extend the LSTM-SAM framework to inland target stations by taking into account the contribution of river discharge for accurate prediction of total WLs in coastal to inland transition zones (Serafin et al. 2017b; Bilskie and Hagen 2018; Muñoz et al. 2022b; Yin et al. 2022). Future work should focus on predicting spatiotemporal WL variability and flood inundation extent by combining Bi-LSTM-ATT and Convolutional Neural Networks (Gavahi et al. 2021).

5. Conclusion

In the present study, we characterize the evolution of extreme water levels (EWLs) at tide-gage stations distributed along the U.S. Atlantic Coast. To achieve this, we identify 5 training stations that were hit by historic hurricane events and contain complete consecutive hourly data spanning at least 40 years. Then, we leverage available WL and hydrometeorological time-series data to train bidirectional Long Short-Term Memory (Bi-LSTM) network models for each training station. Furthermore, we incorporate an attention layer mechanism in the model architecture and a transfer learning (TL) approach with the goal effectively predicting the evolution of EWLs at target (tide-gage) stations. The models highlight the significance of a longer batch size in enhancing model performance, while challenging the assumed benefits of longer look-back periods. The collection of models with the attention layer mechanism and TL approach is referred to as the LSTM-Station Approximated Models (LSTM-SAM) framework and is effectively applied to 14 target stations.

The LSTM-SAM framework predicts the onset, peak, and dissipation of multiple EWL events emerging from tropical cyclones (hurricanes) and extratropical cyclones (Nor'easter storms) with high accuracy. For this, the framework identifies "transferable" models based on KGE and NSE above 0.70 in order to ensure an accurate generalization of EWLs. Under these criteria, the LSTM-SAM framework demonstrates satisfactory performance with transferable models achieving average KGE, NSE, and RMSE ranging from 0.78 to 0.92, 0.90 to 0.97, and 0.09 to 0.18 at the target stations, respectively. Following these results, we conclude that the LSTM-SAM framework can accurately predict not only EWLs but also their evolution over time, which could assist in large-scale operational flood predictions like the National Water Model (NWM) or Coastal Emergency Risk Assessment (CERA). Future work should focus on predicting spatiotemporal WL variability and flood inundation extent by combining Bi-LSTM-ATT and Convolutional Neural Networks.

Acknowledgment

Partial financial support for this study is provided by the National Science Foundation, CAS-Climate Program (Award # 480948).

References

Abbaszadeh P, Gavahi K, Moradkhani H (2020) Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. Advances in Water Resources 145:103721. https://doi.org/10.1016/j.advwatres.2020.103721

Ahmed S, Saif AF, Hanif MI, et al (2022) Att-BiL-SL: Attention-Based Bi-LSTM and Sequential LSTM for Describing Video in the Textual Formation. Applied Sciences 12:. https://doi.org/10.3390/app12010317

Alarcon VJ, Linhoss AC, Kelble CR, et al (2022) Coastal inundation under concurrent mean and extreme sea-level rise in Coral Gables, Florida, USA. Natural Hazards 1–30

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

Almar R, Ranasinghe R, Bergsma EWJ, et al (2021) A global analysis of extreme coastal water levels with implications for potential coastal overtopping. Nat Commun 12:3775. https://doi.org/10.1038/s41467-021-24008-9

Altunkaynak A, Kartal E (2021) Transfer sea level learning in the Bosphorus Strait by wavelet based machine learning methods. Ocean Engineering 233:109116. https://doi.org/10.1016/j.oceaneng.2021.109116

Anderson DL, Ruggiero P, Mendez FJ, et al (2021) Projecting Climate Dependent Coastal Flood Risk With a Hybrid Statistical Dynamical Model. Earth's Future 9:e2021EF002285. https://doi.org/10.1029/2021EF002285

Arns A, Wahl T, Wolff C, et al (2020) Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts. Nature Communications 11:1–9. https://doi.org/10.1038/s41467-020-15752-5

Bai L-H, Xu H (2021) Accurate estimation of tidal level using bidirectional long short-term memory recurrent neural network. Ocean Engineering 235:108765

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 (2022a) Uneven burden of urban flooding. Nature Sustainability. https://doi.org/10.1038/s41893-022-01000-9

Bates PD (2022b) Flood Inundation Prediction. Annu Rev Fluid Mech 54:287–315. https://doi.org/10.1146/annurev-fluid-030121-113138

Bentivoglio R, Isufi E, Jonkman SN, Taormina R (2022) Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrology and Earth System Sciences 26:4345–4378

Bevacqua, Maraun D, Vousdoukas MI, et al (2019) Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. Science Advances 5:eaaw5531. https://doi.org/10.1126/sciadv.aaw5531

Bian G-F, Nie G-Z, Qiu X (2021) How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? Atmospheric Research 249:105339

Bilskie MV, Hagen SC (2018) Defining Flood Zone Transitions in Low-Gradient Coastal Regions. Geophysical Research Letters 45:2761–2770. https://doi.org/10.1002/2018GL077524

Bilskie MV, Zhao H, Resio D, et al (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

Bischl B, Binder M, Lang M, et al (2023) Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 13:e1484

Bloemendaal N, de Moel H, Martinez AB, et al (2022) A globally consistent local-scale assessment of future tropical cyclone risk. Science Advances 8:eabm8438. https://doi.org/10.1126/sciadv.abm8438

Boumis G, Moftakhari HR, Moradkhani H (2023) Coevolution of Extreme Sea Levels and Sea-Level Rise Under Global Warming. Earth's Future 11:e2023EF003649. https://doi.org/10.1029/2023EF003649

Boussioux L, Zeng C, Guénais T, Bertsimas D (2022) Hurricane Forecasting: A Novel Multimodal Machine Learning Framework. Weather and Forecasting 37:817–831. https://doi.org/10.1175/WAF-D-21-0091.1

Brakenridge GR (2021) Global active archive of large flood events, 1985-present. Available from DFO Flood Observatory, University of Colorado(Accessed on October, 2021 through personal communication)

Bruneau N, Polton J, Williams J, Holt J (2020) Estimation of global coastal sea level extremes using neural networks. Environmental Research Letters 15:074030

C. Huang, J. Zhang, L. Cao, et al (2020) Robust Forecasting of River-Flow Based on Convolutional Neural Network. IEEE Transactions on Sustainable Computing 5:594–600. https://doi.org/10.1109/TSUSC.2020.2983097

Chen D, Zhang J, Jiang S (2020) Forecasting the Short-Term Metro Ridership With Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks. IEEE Access 8:91181–91187. https://doi.org/10.1109/ACCESS.2020.2995044

Cleveland RB, Cleveland WS, McRae JE, Terpenning I (1990) STL: A seasonal-trend decomposition. J Off Stat 6:3–73

Codiga DL (2011) Unified tidal analysis and prediction using the UTide Matlab functions

de Amorim LBV, Cavalcanti GDC, Cruz RMO (2023) The choice of scaling technique matters for classification performance. Applied Soft Computing 133:109924. https://doi.org/10.1016/j.asoc.2022.109924

De la Fuente LA, Ehsani MR, Gupta HV, Condon LE (2024) Toward interpretable LSTM-based modeling of hydrological systems. Hydrol Earth Syst Sci 28:945–971. https://doi.org/10.5194/hess-28-945-2024

Ding Y, Zhu Y, Feng J, et al (2020) Interpretable spatio-temporal attention LSTM model for flood forecasting. Neurocomputing 403:348–359. https://doi.org/10.1016/j.neucom.2020.04.110

Douris J, Kim G (2021) The Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970-2019)

Evangelista LGC, Giusti R (2021) Short-term effects of weight initialization functions in Deep

NeuroEvolution. Evo 21

F. Zhuang, Z. Qi, K. Duan, et al (2021) A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE 109:43–76. https://doi.org/10.1109/JPROC.2020.3004555

Fang Z, Wang Y, Peng L, Hong H (2021) Predicting flood susceptibility using LSTM neural networks. Journal of Hydrology 594:125734. https://doi.org/10.1016/j.jhydrol.2020.125734

Fraehr N, Wang QJ, Wu W, Nathan R (2022) Upskilling Low-Fidelity Hydrodynamic Models of Flood Inundation Through Spatial Analysis and Gaussian Process Learning. Water Resources Research 58:e2022WR032248. https://doi.org/10.1029/2022WR032248

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

Ghanbari M, Arabi M, Kao S-C, et al (2021) Climate Change and Changes in Compound Coastal-Riverine Flooding Hazard Along the U.S. Coasts. Earth's Future 9:e2021EF002055. https://doi.org/10.1029/2021EF002055

Glavovic B, Dawson R, Chow WT, et al (2022) Cities and settlements by the sea

Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. JMLR Workshop and Conference Proceedings, pp 249–256

Gul S, Khan GM, Yousaf S (2022) Multi-step short-term $$PM_{2.5}$$ forecasting for enactment of proactive environmental regulation strategies. Environmental Monitoring and Assessment 194:386. https://doi.org/10.1007/s10661-022-10029-4

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.

Journal of hydrology 377, 80–91.

Hamidi E, Peter BG, Muñoz DF, et al (2023) Fast Flood Extent Monitoring With SAR Change Detection Using Google Earth Engine. IEEE Transactions on Geoscience and Remote Sensing 61:1–19. https://doi.org/10.1109/TGRS.2023.3240097

Hamitouche M, Molina J-L (2022) A review of ai methods for the prediction of high-flow extremal hydrology. Water Resources Management 36:3859–3876

Hewamalage H, Bergmeir C, Bandara K (2021) Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting 37:388–427

Hino M, Nance E (2021) Five ways to ensure flood-risk research helps the most vulnerable. Nature 595:27–29

Jafarzadegan K, Alipour A, Gavahi K, et al (2021) Toward improved river boundary conditioning for simulation of extreme floods. Advances in Water Resources 158:104059. https://doi.org/10.1016/j.advwatres.2021.104059

Kardhana H, Valerian JR, Rohmat FI, Kusuma MS (2022) Improving Jakarta's Katulampa Barrage Extreme Water Level Prediction Using Satellite-Based Long Short-Term Memory (LSTM) Neural Networks. Water 14:. https://doi.org/10.3390/w14091469

Khojasteh D, Glamore W, Heimhuber V, Felder S (2021) Sea level rise impacts on estuarine dynamics: A review. Science of The Total Environment 780:146470. https://doi.org/10.1016/j.scitotenv.2021.146470

Kingphai K, Moshfeghi Y (2022) On time series cross-validation for deep learning classification model of mental workload levels based on EEG signals. Springer, pp 402–416

Kratzert F, Gauch M, Klotz D, Nearing G (2024) HESS Opinions: Never train an LSTM on a single basin. Hydrol Earth Syst Sci Discuss 2024:1–19. https://doi.org/10.5194/hess-2023-275

Lee J-W, Irish JL, Bensi MT, Marcy DC (2021) Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning. Coastal Engineering 170:104024. https://doi.org/10.1016/j.coastaleng.2021.104024

Li W, Kiaghadi A, Dawson C (2021) Exploring the best sequence LSTM modeling architecture for flood prediction. Neural Computing and Applications 33:5571–5580. https://doi.org/10.1007/s00521-020-05334-3

Lindemann B, Müller T, Vietz H, et al (2021) A survey on long short-term memory networks for time series prediction. Procedia CIRP 99:650–655. https://doi.org/10.1016/j.procir.2021.03.088

Liu Y, Yang Y, Chin RJ, et al (2023) Long Short-Term Memory (LSTM) Based Model for Flood Forecasting in Xiangjiang River. KSCE Journal of Civil Engineering 27:5030–5040. https://doi.org/10.1007/s12205-023-2469-7

Mahdianpari M, Salehi B, Rezaee M, et al (2018) Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. Remote Sensing 10:1119. https://doi.org/10.3390/rs10071119

Marco R, Ahmad SSS, Ahmad S (2022) Bayesian hyperparameter optimization and Ensemble Learning for Machine Learning Models on software effort estimation. International Journal of Advanced Computer Science and Applications 13:

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. Frontiers in Built Environment 6:149. https://doi.org/10.3389/fbuil.2020.00149

Merizalde MJ, Muñoz P, Corzo G, et al (2023) Integrating geographic data and the SCS-CN method with LSTM networks for enhanced runoff forecasting in a complex mountain basin. Frontiers in Water 5:1233899

Moftakhari H, Muñoz DF, Akbari Asanjan A, et al (2024) Nonlinear Interactions of Sea-Level Rise and Storm Tide Alter Extreme Coastal Water Levels: How and Why? AGU Advances 5:e2023AV000996. https://doi.org/10.1029/2023AV000996

Moftakhari, Schubert JE, AghaKouchak A, et al (2019) Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries. Advances in Water Resources 128:28–38. https://doi.org/10.1016/j.advwatres.2019.04.009

Muis S, Lin N, Verlaan M, et al (2019) Spatiotemporal patterns of extreme sea levels along the western North-Atlantic coasts. Scientific Reports 9:3391. https://doi.org/10.1038/s41598-019-40157-w

Muis S, Verlaan M, Winsemius HC, et al (2016) A global reanalysis of storm surges and extreme sea levels. Nature communications 7:11969

Muñoz DF, Abbaszadeh P, Moftakhari H, Moradkhani H (2022a) Accounting for uncertainties in compound flood hazard assessment: The value of data assimilation. Coastal Engineering 171:104057. https://doi.org/10.1016/j.coastaleng.2021.104057

Muñoz DF, Moftakhari H, Moradkhani H (2020) Compound Effects of Flood Drivers and Wetland Elevation Correction on Coastal Flood Hazard Assessment. Water Resources Research 56:e2020WR027544. https://doi.org/10.1029/2020WR027544

Muñoz DF, Muñoz P, Moftakhari H, Moradkhani H (2021) From local to regional compound flood mapping with deep learning and data fusion techniques. Science of The Total Environment 782:146927. https://doi.org/10.1016/j.scitotenv.2021.146927

Muñoz DF, Yin D, Bakhtyar R, et al (2022b) Inter-Model Comparison of Delft3D-FM and 2D HEC-RAS for Total Water Level Prediction in Coastal to Inland Transition Zones. JAWRA Journal of the American Water Resources Association 58:34–49. https://doi.org/10.1111/1752-1688.12952

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology 10, 282–290.

Nearing G, Cohen D, Dube V, et al (2024) Global prediction of extreme floods in ungauged watersheds. Nature 627:559–563. https://doi.org/10.1038/s41586-024-07145-1

NOAA-NCEI (2023) National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters. https://doi.org/10.25921/stkw-7w73

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 Engineering 266:113048. https://doi.org/10.1016/j.oceaneng.2022.113048

Office for Coastal Management, 2024. Economics and Demographics [WWW Document]. URL https://coast.noaa.gov/states/fast-facts/economics-and-demographics.html (accessed 4.1.24).

Parker K, Erikson L, Thomas J, et al (2023) Relative contributions of water-level components to extreme water levels along the US Southeast Atlantic Coast from a regional-scale water-level hindcast. Natural Hazards. https://doi.org/10.1007/s11069-023-05939-6

Pörtner H-O, Roberts DC, Masson-Delmotte V, et al (2019) IPCC special report on the ocean and cryosphere in a changing climate. IPCC Intergovernmental Panel on Climate Change: Geneva, Switzerland 1:1–755

Rainey JL, Brody SD, Galloway GE, Highfield WE (2021) Assessment of the growing threat of urban flooding: A case study of a national survey. Urban Water Journal 18:375–381

Rashid MM, Moftakhari H, Moradkhani H (2024) Stochastic simulation of storm surge extremes along the contiguous United States coastlines using the max-stable process. Communications Earth & Environment 5:39. https://doi.org/10.1038/s43247-024-01206-z

Rithani M, Kumar RP, Doss S (2023) A review on big data based on deep neural network approaches. Artificial Intelligence Review 56:14765–14801. https://doi.org/10.1007/s10462-023-10512-5

Rueda A, Camus P, Tomás A, et al (2016) A multivariate extreme wave and storm surge climate emulator based on weather patterns. Ocean Modelling 104:242–251. https://doi.org/10.1016/j.ocemod.2016.06.008

Saksena S, Merwade V (2015) Incorporating the effect of DEM resolution and accuracy for improved flood inundation mapping. Journal of Hydrology 530:180–194. https://doi.org/10.1016/j.jhydrol.2015.09.069

Sanders BF, Schubert JE, Kahl DT, et al (2022) Large and inequitable flood risks in Los Angeles, California. Nature Sustainability. https://doi.org/10.1038/s41893-022-00977-7

Santiago-Collazo FL, Bilskie MV, Hagen SC (2019) A comprehensive review of compound inundation models in low-gradient coastal watersheds. Environmental Modelling & Software 119:166–181. https://doi.org/10.1016/j.envsoft.2019.06.002

Seleem O, Ayzel G, Bronstert A, Heistermann M (2023) Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany. Natural Hazards and Earth System Sciences 23:809–822

Serafin KA, Ruggiero P (2014) Simulating extreme total water levels using a time-dependent, extreme value approach. Journal of Geophysical Research: Oceans 119:6305–6329.

https://doi.org/10.1002/2014JC010093

Serafin KA, Ruggiero P, Stockdon HF (2017a) The relative contribution of waves, tides, and nontidal residuals to extreme total water levels on U.S. West Coast sandy beaches. Geophysical Research Letters 44:1839–1847. https://doi.org/10.1002/2016GL071020

Serafin KA, Ruggiero P, Stockdon HF (2017b) The relative contribution of waves, tides, and nontidal residuals to extreme total water levels on U.S. West Coast sandy beaches. Geophysical Research Letters 44:1839–1847. https://doi.org/10.1002/2016GL071020

Shen C (2018) A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resources Research 54:8558–8593. https://doi.org/10.1029/2018WR022643

Shen C, Appling AP, Gentine P, et al (2023) Differentiable modelling to unify machine learning and physical models for geosciences. Nature Reviews Earth & Environment 4:552–567. https://doi.org/10.1038/s43017-023-00450-9

Tan C, Sun F, Kong T, et al (2018) A Survey on Deep Transfer Learning. In: Kůrková V, Manolopoulos Y, Hammer B, et al. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. Springer International Publishing, Cham, pp 270–279

Tate E, Rahman MA, Emrich CT, Sampson CC (2021) Flood exposure and social vulnerability in the United States. Natural Hazards 106:435–457

Tedesco P, Rabault J, Sætra ML, et al (2023) Bias Correction of Operational Storm Surge Forecasts Using Neural Networks. arXiv preprint arXiv:230100892

Thomas A, Dietrich J, Asher T, et al (2019) Influence of storm timing and forward speed on tides and storm surge during Hurricane Matthew. Ocean Modelling 137:1–19. https://doi.org/10.1016/j.ocemod.2019.03.004

Tiggeloven T, Couasnon A, van Straaten C, et al (2021) Exploring deep learning capabilities for surge predictions in coastal areas. Scientific reports 11:17224

Velasquez-Montoya L, Overton MF, Sciaudone EJ (2020) Natural and anthropogenic-induced changes in a tidal inlet: Morphological evolution of Oregon Inlet. Geomorphology 350:106871

Velasquez-Montoya L, Sciaudone EJ, Smyre E, Overton MF (2021) Vulnerability indicators for coastal roadways based on barrier island morphology and shoreline change predictions. Natural Hazards Review 22:04021003

Velasquez-Montoya L, Wargula A, Nangle J, et al (2022) Hydrodynamics of a tidal inlet under gray to green coastal protection interventions. Frontiers in Earth Science 10:991667

Wahl, Jain S, Bender J, et al (2015) Increasing risk of compound flooding from storm surge and rainfall for major US cities. Nature Climate Change 5:1093–1097. https://doi.org/10.1038/nclimate2736

Wahl T, Haigh ID, Nicholls RJ, et al (2017) Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis. Nature Communications 8:16075. https://doi.org/10.1038/ncomms16075

Wang X, Jin Y, Schmitt S, Olhofer M (2023) Recent advances in Bayesian optimization. ACM Computing Surveys 55:1–36

Wing OE, Lehman W, Bates PD, et al (2022) Inequitable patterns of US flood risk in the Anthropocene. Nature Climate Change 12:156–162

Yin D, Muñoz DF, Bakhtyar R, et al (2022) Extreme Water Level Simulation and Component Analysis in Delaware Estuary during Hurricane Isabel. JAWRA Journal of the American Water Resources Association 58:19–33. https://doi.org/10.1111/1752-1688.12947

Zhang Y, Ragettli S, Molnar P, et al (2022) Generalization of an Encoder-Decoder LSTM model for flood prediction in ungauged catchments. Journal of Hydrology 614:128577. https://doi.org/10.1016/j.jhydrol.2022.128577

Zhao G, Pang B, Xu Z, et al (2021) Improving urban flood susceptibility mapping using transfer learning. Journal of Hydrology 602:126777. https://doi.org/10.1016/j.jhydrol.2021.126777

Zscheischler J, Martius O, Westra S, et al (2020) A typology of compound weather and climate events. Nature Reviews Earth & Environment 1:333–347. https://doi.org/10.1038/s43017-020-0060-z