Earth's Future

RESEARCH ARTICLE
10.1029/2025EF007072

Special Collection:
Forcing, response, and impacts of coastal storms in a changing climate

A Deep Learning Framework for Extreme Storm Surge Modeling Under Future Climate Scenarios

Emiliano Longo[1], Andrea Ficchì[1], Martin Verlaan[2,3], Sanne Muis[3,4], and Andrea Castelletti[1,5]

[1]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milano, Italy, [2]TU Delft, Delft, The Netherlands, [3]Deltares, Delft, The Netherlands, [4]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, [5]CMCC Foundation—Euro-Mediterranean Centre on Climate Change, Milano, Italy

Abstract Coastal regions are increasingly exposed to sea-level rise and intensifying storm surges, underscoring the urgent need for accurate long-term predictions of extreme water levels to support robust adaptation planning. Physics-based hydrodynamic storm surge models remain the gold standard for such projections, but are computationally demanding, limiting their feasibility for producing the large scenario ensembles needed under deep uncertainty. Artificial intelligence surrogate models have emerged as a promising alternative. Yet, current approaches often underrepresent rare extremes and lack validation under future climate conditions, constraining their application for long-term planning. Here, we develop a deep learning surrogate model trained on hydrodynamic simulations from the Global Tide and Surge Model (GTSM), with both historical reanalysis and high-resolution climate projections (CMIP6 HighResMIP). Using New York City, a highly vulnerable urban coastline with extensive surge records, as a testbed, we demonstrate the model's ability to represent extreme storm surges under both historical and mid-21st-century scenarios. To enhance performance on extremes, we propose a novel asymmetric loss function, combining quantile and expectile losses, which substantially improves predictions of rare storm surge events, while maintaining high overall performance. Fine-tuning with climate model outputs further aligns the surrogate's estimates with those of the hydrodynamic model across spatial and temporal scales. Under future climate forcing, projections obtained with the surrogate model closely reproduce the response of GTSM, capturing projected trends in extreme events. This open-data-based framework provides a computationally efficient and globally transferable approach for storm surge projection, enabling the large-scale scenario analyses required for climate-resilient coastal planning.

Plain Language Summary While rising mean sea levels are expected to be the dominant driver of future coastal flooding, storm surges remain a critical and highly uncertain contributor to coastal flood risk. In this study, we develop an Artificial Intelligence model that accurately emulates storm surge simulations for New York City, offering a much faster alternative to traditional physics-based models. Our model closely reproduces the storm surge levels from the Global Tide and Surge Model, including extreme events. We demonstrate that training the model using asymmetric loss functions, such as a new combination of quantile and expectile terms, significantly improves the accuracy of the surrogate model for rare but high-impact storm surge events, under both present and future climate scenarios. When applied to future climate scenarios, the model maintains high accuracy if it is fine-tuned using historical simulations from the same climate model. This framework offers a fast, flexible, and transferable tool for projecting storm surge extremes based on openly available global data sets, and can be applied to other coastal locations worldwide.

Key Points:

- Neural network models can efficiently and accurately emulate storm surge predictions from hydrodynamic models
- Using a custom asymmetric loss function significantly improves the prediction of extreme storm surge events
- Fine-tuning using historical climate model simulations enhances surrogate performance under future forcing scenarios

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Castelletti,
andrea.castelletti@polimi.it

Citation:

Longo, E., Ficchì, A., Verlaan, M., Muis, S., & Castelletti, A. (2026). A deep learning framework for extreme storm surge modeling under future climate scenarios. *Earth's Future*, *14*, e2025EF007072. https://doi.org/10.1029/2025EF007072

Received 6 AUG 2025
Accepted 4 MAR 2026

1. Introduction

Beyond the well-documented rise in mean sea levels, climate change is also reshaping the high-frequency dynamics of storm surges and waves through alterations in atmospheric and oceanic circulation patterns (Bernier et al., 2024; Intergovernmental Panel On Climate Change (Ipcc), 2022). While long-term regional variations in mean sea level are projected to remain the primary driver of extreme coastal water levels and flooding in the coming decades, particularly under high-emission scenarios (Tebaldi et al., 2021; Vousdoukas et al., 2018), episodic processes such as storm surges and wave action continue to play a critical role in coastal flood risk, potentially amplifying its impacts (Bernier et al., 2024; Intergovernmental Panel On Climate Change (Ipcc), 2023; Kirezci et al., 2020; Lashley et al., 2025). In many regions, the intensity of low-probability extreme

water level events is projected to increase, potentially exacerbating hazards in vulnerable coastal zones (Muis et al., 2020, 2023; Vousdoukas et al., 2018). A growing body of regional and global studies underscores the likelihood of more frequent and severe coastal hazards under climate change. Drivers include the accelerated sequence of storm and tropical cyclone (TC) events (Knutson et al., 2020; Xi et al., 2023), intensification of wind speeds in major TCs (Patricola & Wehner, 2018), and an increased probability of compound flooding from the concurrence of extreme precipitation and meteorological tides (Bevacqua et al., 2020). These projections reinforce the imperative to develop robust and adaptive strategies capable of navigating deep uncertainty in future coastal dynamics (Aziz et al., 2024; Bernier et al., 2024; Oddo et al., 2020). Contemporary frameworks for assessing future coastal risk typically fall into three broad categories: (a) hydro-dynamic models, which resolve the physical processes underpinning coastal hazards, (b) statistical approaches, which is based on a representation of the probability distributions derived from historical or synthetic data sets, and (c) data-driven approaches, including Artificial Intelligence (AI)-based models, which rely on empirical relationships between input and output variables. Each paradigm offers distinct advantages but is constrained by specific assumptions and simplifications that limit its utility in adaptation planning (Eilander et al., 2023; Feng et al., 2025; Garner & Keller, 2018; Herman et al., 2020; Mayo & Lin, 2022).

Hydro-dynamic models excel at representing nonlinear interactions among mean sea level, tides, surges, wave setup, and other drivers of compound flooding (Leijnse et al., 2021; Muis et al., 2020; Shimura et al., 2022). They accommodate deep uncertainty and ambiguity in system behavior (Kopp, Garner, et al., 2023; Muis et al., 2023), and probabilistic frameworks allow the propagation of uncertainties in climate inputs, model parameters, and structural configurations (Kopp, Oppenheimer, et al., 2023). Ensemble simulations further enable the exploration of plausible futures and their associated likelihoods (Herman et al., 2020; Lin et al., 2019). The integration of outputs from Global Climate Models (GCMs) into hydrodynamic simulations ensures coherent representations of hydro-meteorological drivers under future emission scenarios. Notably, the increasing fidelity of high-resolution GCM experiments (e.g., HighResMIP within CMIP6) has enhanced the capacity to simulate tropical cyclone behavior and associated hazards (Haarsma et al., 2016; Roberts et al., 2020).

Yet, despite these advantages, dynamic models remain highly computationally intensive. Accurately resolving multi-scale processes requires fine-resolution grids, making large-scale ensemble simulations prohibitively expensive in many cases (Muis et al., 2020; Pachev et al., 2023; Wang et al., 2023). Statistical approaches offer improved computational efficiency but often rely on simplifying assumptions about dependencies among sea level components (Benito et al., 2024; Lin et al., 2016; O'Grady et al., 2022; Vousdoukas et al., 2018). These approximations frequently omit critical dynamics, such as the timing and interaction of surge and tide, or fail to coherently represent meteorological drivers necessary for capturing compound flooding (Ragno et al., 2023; Sarhadi et al., 2025), thereby limiting their utility in comprehensive risk assessments and adaptation planning. Moreover, because statistical models require large sample sizes to derive reliable risk estimates, they often depend on synthetic scenarios generated by dynamic models, ultimately inheriting the same computational burdens and scope limitations they were designed to bypass (Krien et al., 2015; Lin et al., 2012; Mayo & Lin, 2022).

To overcome the computational constraints of dynamic models, recent studies have increasingly explored data-driven approaches, as surrogate models for storm surge dynamics. A broad body of data-driven methodologies for related coastal and flood-risk applications exist in the literature (Qin et al., 2023), including Gaussian process emulators (Betancourt et al., 2020; Liu & Guillas, 2016), Radial Basis Function (Rueda et al., 2019), Random Forests (Lecacheux et al., 2021), k-nearest neighbors (Tausía et al., 2023), physics-informed learning (Raissi et al., 2019), and neural networks (Harter et al., 2024; Hermans et al., 2025; Ishida et al., 2020). Within the broad landscape of data-driven approaches, neural networks represent a flexible class of models that leverage their universal approximation properties, making them particularly suitable for regression problems involving complex and non-linear input-output mapping. While such models lack explicit representations of underlying physical processes, they offer rapid inference capabilities for applications ranging from real-time forecasting and emergency response (Qin et al., 2023; Wang et al., 2023; Xie et al., 2023) to generating large ensembles for long-term risk projections (Harter et al., 2024; Ishida et al., 2020; Jiang et al., 2024; Kaufmann et al., 2024; Tiggeloven et al., 2021).

AI-based frameworks exhibit substantial diversity in their architectures and target applications. Some models incorporate endogenous states to predict system evolution sequentially (i.e., timestep by timestep), which is

particularly suited to forecasting (Qin et al., 2023; Wang et al., 2023; Xi et al., 2023), while others employ exogenous states to surrogate dynamic models, facilitating long-term scenario projections (Ayyad et al., 2023; Ishida et al., 2020). Differences in target variables (e.g., tide gauge observations vs. synthetic hydrodynamic outputs), spatio-temporal resolutions, and forcing data further distinguish these approaches (Harter et al., 2024; Qin et al., 2023; Xi et al., 2023). Neural networks have demonstrated promising regression performance in storm surge applications (Ayyad et al., 2023; Jiang et al., 2024; Kaufmann et al., 2024), including both cost sensitive learning approach (Hermans et al., 2025) and input features selection to enhance predictions of extremes (e.g., wind stress over wind speed (Harter et al., 2024), or average temperature signals for climate-driven variability (Ishida et al., 2020)). However, the vast majority of AI models developed to date are trained and validated using tide gauge data or historical storm surge reanalysis, which are limited in both spatial coverage and resolution (e.g., (Bruneau et al., 2020; Hermans et al., 2025)). Only a few studies assess the performance of neural networks for surge modeling across the global gauge network, highlighting the equatorial/tropical zone as the region where such models perform the worst (Tadesse et al., 2020; Tiggeloven et al., 2021).

Despite these advances, critical gaps remain in the development of AI models for storm surge projections (Bernier et al., 2024), particularly regarding their ability to capture extreme events (Camps-Valls et al., 2024). Neural networks frequently underestimate rare, high-impact surges, precisely the events most relevant for adaptation planning and early warning systems (Camps-Valls et al., 2024; Harter et al., 2024). Moreover, AI models have yet to be evaluated under non-stationary future conditions, a key requirement for extending beyond historical baselines and ensuring their robustness in a changing climate (Bernier et al., 2024). Most existing AI applications focus on direct prediction of surge levels using observed tide gauge data, while comparatively less attention has been given to surrogate modeling approaches that emulate hydrodynamic models. In particular, there is a lack of systematic evaluation of AI-based surrogates trained on outputs from physically based models, especially on storm surge projections. It remains unclear how transferable surrogate models are across different climate inputs (models or scenarios), or whether their strong performance holds under future scenarios without model-specific fine-tuning. Such evaluations are essential to support large-scale scenario generation under future climate ensembles, within a computationally efficient yet physically consistent framework.

At present, global storm surge projections remain limited to a few computationally intensive hydrodynamic modeling efforts using GCM forcings. These include ADCIRC coupled with MRI-AGCM meteorological fields (Shimura et al., 2022), as well as iterations of the Global Tide and Surge Model (GTSM) (Muis et al., 2020, 2023; Vousdoukas et al., 2018). The highest-resolution global projections to date have been generated with GTSM 3.0, driven by HighResMIP data sets and providing open-access outputs for five GCM simulations through 2050 (Copernicus Climate Change Service, 2022; Muis et al., 2023). This data set presents an unprecedented opportunity to train AI surrogates for storm surge projections, a strategy that remains largely unexplored.

Here, we develop a neural network surrogate model of GTSM 3.0 using these high-resolution global storm surge projections (Muis et al., 2023). The modeling framework is designed to be globally applicable because it can be trained in any location where GTSM outputs are available. In our case, we train our surrogate model using only meteorological forcings and without providing location-specific morphological information. This formulation allows the model to learn the local storm-surge response implicitly from the underlying hydrodynamic simulations used for training. As storm-surge dynamics depend on local bathymetry, coastline geometry, and regional atmospheric regimes, surrogate models remain location-specific unless retrained using all location-specific inputs. Our analysis focuses on the New York City coastline, a critical case study due to its population density, economic exposure, and extensive prior research on coastal flood risk (Sarhadi et al., 2024; Strauss et al., 2021).

Leveraging ERA5 and HighResMIP forcing data sets, we target surge and tide outputs to preserve the nonlinear interactions captured by GTSM. We assess model performance under both historical and projected future conditions, with particular attention to the use of quantile regression for improving the representation of extremes. Specifically, we (a) perform a quantitative evaluation of the surrogate model against GTSM outputs (both reanalysis and projection) using standard regression metrics and Extreme Value Analysis (EVA), and (b) compare the effectiveness of training the surrogate using either the Mean Squared Error (MSE) or a novel asymmetric loss function, combining quantile and expectile losses, in reducing the underestimation of storm surge peaks.

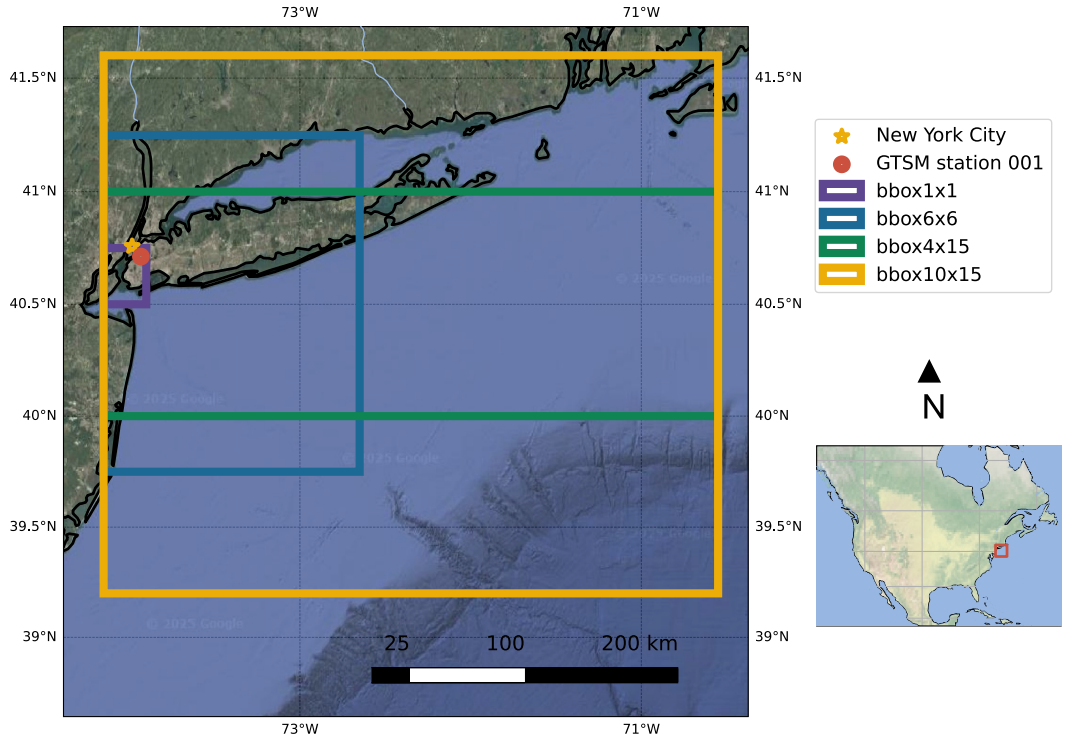

Figure 1. Map of the New York City coastline highlighting the GTSM station target and showing the bounding boxes of the atmospheric forcing fields.

2. Material and Methods

2.1. Case Study and Data

This study develops a surrogate model targeting GTSM outputs at a tide gauge near New York City (NYC). The closest GTSM node to Manhattan's southern tip lies at coordinates (40.71, −74.02) in the EPSG:4326 reference system (Figure 1). The NYC coastline is among the most studied urban coastal zones worldwide, due to its high exposure to coastal flood risk in low-lying metropolitan areas. This risk became particularly evident during Hurricane Sandy in October 2012, which caused 43 fatalities and over US$ 60 billion in damages (Marsooli & Wang, 2020; Strauss et al., 2021). Storm surges in this region are primarily driven by extratropical cyclones (ETCs) in the cool season (November–March) and tropical cyclones (TCs) during the hurricane season (June–November), both producing strong onshore winds and low atmospheric pressure that elevate coastal water levels (Orton et al., 2016). Notably, although the majority of storm surge events in NYC have been caused by ETCs, the most severe historical storm surges have been associated with TCs that produce extremely rare events (Catalano & Broccoli, 2018; Towey et al., 2022). In addition to storm surge, astronomical tides play a critical role in modulating coastal flood risk in NYC, exhibiting pronounced spatial and temporal variability along the complex coastline and within the NYC bay. This variability increases the likelihood of in-phase interactions between storm surge and tide, exacerbating local water level extremes (Colle et al., 2008; Georgas et al., 2014; Marsooli & Wang, 2020). Furthermore, the region's intricate bathymetry and coastal geometry, characterized by estuaries and tidal inlets, can amplify both tidal and surge dynamics at specific locations (Colle et al., 2008; Irish & Cañizares, 2009). The interplay of climate drivers, sea level rise, and local hydrodynamic processes makes projections of future storm tides and coastal flood risk in NYC highly uncertain (Garner et al., 2017; Lin et al., 2019), underscoring the need for fast and accurate surrogate models capable of supporting large ensemble projections of future sea levels.

The GTSM simulations span multiple periods: reanalysis (1979–present), historical (1950–2014), and future projections (2016–2050). The data set is provided at different resolutions: 0.1° along the global coastline and 0.25°–1° for ocean grid points (Copernicus Climate Change Service, 2022). GTSM dynamically simulates interactions among mean sea level, astronomical tide, and storm surge, capturing nonlinear surge-tide interactions.

The total water level relative to mean sea level is calculated as the sum of the astronomical tide and storm surge. GTSM v3.0 is a global, depth-averaged barotropic ocean model that solves the non-linear rotating shallow-water equations on an unstructured grid. Atmospheric forcing enters the model through surface wind stress and gradients of mean sea-level pressure. Mean sea level rise from climate models is used as a spatially varying, slowly evolving modification of the background water depth and reference sea surface. As a result, changes in mean sea level influence tidal propagation, dissipation and tide-surge interaction through changes in geometry rather than as a direct dynamical forcing. To capture both synoptic-scale (ETCs and TCs) and meso-scale features, the atmospheric forcings require hourly or sub-hourly resolution. GTSM operates with a 10-min time step and employs a variable-resolution mesh, achieving high fidelity along coasts while maintaining computational efficiency offshore (Muis et al., 2020).

To train the surrogate model, we use the same atmospheric data sets that force GTSM: ERA5 reanalysis (C3S, 2018) and GCM simulations from HighResMIP (Haarsma et al., 2016). ERA5 provides $0.25° \times 0.25°$ hourly data from 1979 onward; we focus on this modern satellite era to maximize consistency and data quality, excluding the pre-1979 back-extension available only recently. For projections, we adopt the CMCC-CM2-VHR4 model, one of only a few HighResMIP GCMs forced into GTSM, offering coupled atmosphere–ocean fields at 25×25 km resolution and 6-hourly time steps (Scoccimarro, Bellucci, & Peano, 2017), and capable of representing realistically TCs of category-5 on the Saffir-Simpson scale (Scoccimarro, Fogli, et al., 2017). CMCC-CM2-VHR4 atmospheric forcings are linearly interpolated to hourly resolution to match ERA5.

A comparison between forcing distributions over historical and future climate reveals a marked warming trend and an increase in the frequency of low-pressure events, while wind distributions appear stationary in isolation (see Figure S1 in Supporting Information S1). However, the effective non-stationarity of these drivers for storm surge requires joint analysis beyond the scope of this study.

2.2. Problem Formulation

To build a surrogate model of GTSM, we define a neural-network function F_θ whose behavior is determined by a set of trainable parameters θ, optimized via gradient-based learning, which maps the forcing values (χ) into the tide and surge signals estimated at time t, as Formalized in Equation 1. As we target a single GTSM output station, the output of the surrogate model is the estimate of tide and surge values at the selected location, that is, $\left[\hat{y}_t^{\text{tide}},\ \hat{y}_t^{\text{surge}}\right]$, in NYC. The tide and surge components are targeted separately to enable a clear and distinct assessment of model performance for each physical process. This decoupling allows us to identify under which forcing conditions and input configurations the neural network improves or degrades its skill for each component individually. Moreover, the modular formulation enables the development of two distinct surrogate modules, one for the astronomical tide and one for the storm surge. While treated as separate outputs, the storm surge component implicitly captures its non-linear dependence on the astronomical tide level and on mean sea level, consistent with known physical interactions represented in the hydrodynamic model. The surrogate model's function and its inputs can be defined as follows:

$$\left[\hat{y}_t^{\text{tide}},\ \hat{y}_t^{\text{surge}}\right] = F_\theta\left(\chi_t, \ldots, \chi_{t-T+1}\right) \tag{1}$$

where:

$$\chi_t = \left[\mathbf{X}_t^1 \quad \ldots \quad \mathbf{X}_t^K \quad ; \quad x_t^1 \quad \ldots \quad x_t^W\right] \tag{2}$$

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \ldots & x_{1,m} \\ \vdots & & \vdots \\ x_{n,1} & \ldots & x_{n,m} \end{bmatrix} \tag{3}$$

The surrogate model is forced with K spatially distributed variables and W spatially independent variables, both sampled for T time steps backward, including the current one. Specifically, we use four spatially distributed atmospheric variables—10 m zonal and meridional wind components, mean sea level pressure, and 2 m air

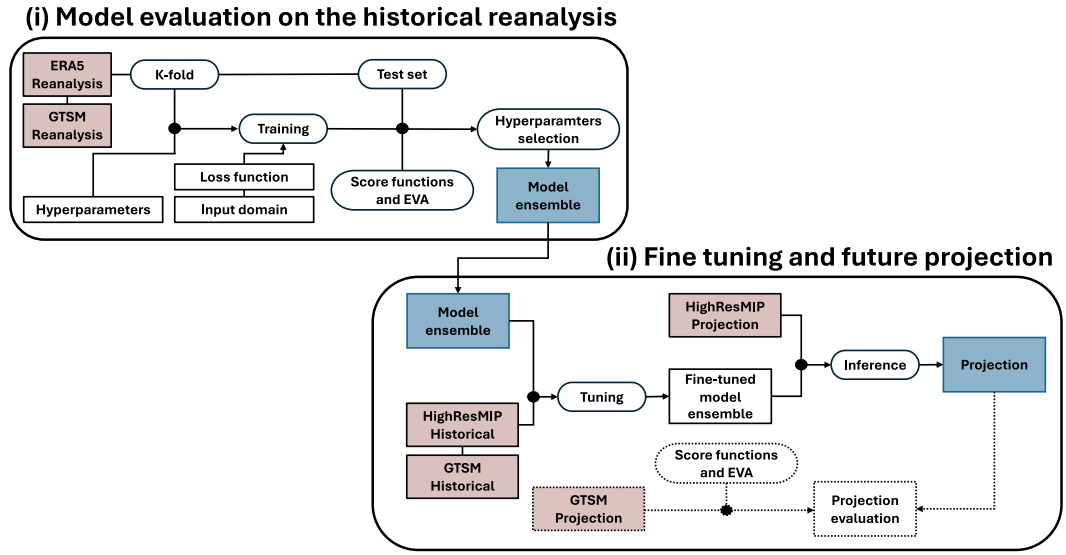

Figure 2. Two-stage framework for surrogate model development: (i) evaluation on the historical reanalysis data set (ERA5), exploring spatio-temporal domains and loss functions; (ii) fine-tuning and generation of future projections. Dashed links and boxes refer to the projection evaluation step. Red boxes indicate available input and target data, while blue boxes represent generated data and models.

temperature. The latter, though uncommon in storm surge models, provides information on large-scale atmospheric conditions under climate change, as demonstrated in Ishida et al. (2020). Additionally, five spatially independent predictors (moon and sun altitude and azimuth, and mean sea level) complement the forcing data set. Solar and lunar altitude and azimuth are used as scalar inputs for the surrogate model to learn the tidal signal and its interaction with slowly varying mean sea level directly from the data.

The number of time steps T (time window for inputs) and the spatial extent of atmospheric gridded variables, that is, the bounding box of (n,m) grid cells, are treated as the problem's hyperparameters to tune. We test $T = \{6-, 12-, 24-, 240-h\}$, and $(n,m) = \{(1,1),(6,6),(4,15),(10,15)\}$, representing four different bounding boxes of increasing spatial extent as displayed in Figure 1. The choice of the spatial extent of the atmospheric forcing domain represents an important trade-off between the physical representativeness of inputs and computational feasibility. Enlarging the atmospheric bounding box increases the dimensionality of the input tensors quadratically, leading to a rapid growth in memory requirements and training cost. While parallel computing enables fast training and inference, memory saturation becomes a limiting factor beyond a certain spatial extent, imposing constraints on batch size and model depth. The selected bounding box therefore reflects a compromise between atmospheric representativeness, model complexity, and numerical efficiency.

2.3. Model Training and Fine-Tuning

We build the function F_θ in Equation 1 using neural networks. We experiment with four architectures differing in their handling of spatio-temporal dependencies: a Feed Forward Neural Network (FFNN), two convolutional networks (2D and 3D), and a Long Short-Term Memory (LSTM) network. The optimization of parameters θ is performed using the Adam gradient descent algorithm (Kingma & Ba, 2014). The training framework consists of two stages (Figure 2):

1. The first phase is the comparative evaluation of the performance of different models trained on the historical reanalysis data set, by testing different configurations of the spatio-temporal domain of the inputs, that is, time window T and grid cells (n, m) in Equations 1–3, and two different loss functions (see next Section 2.3.1);
2. The second phase is the fine-tuning of the model (Weiss et al., 2016) and future projection generation; for the fine-tuning, the best-performing model selected over the reanalysis data set is retrained using the historical simulations of the GCM model (CMCC-CM2-VHR4), by applying a smaller learning rate, and the fine-tuned model is run in inference mode using the GCM projection over the future period; the future surrogate projection is finally validated against the GTSM output available using the same GCM future forcings.

The fine-tuning is performed to adjust the model to the different spatio-temporal structures of the ERA5 reanalysis compared to the CMCC-CM2-VHR4 projections. To select the best surrogate model, we evaluate the performance with overall regression scores and scores on estimated return levels based on Extreme Value Analysis (EVA). We train the model on the reanalysis data set using k-fold cross-validation (with k = 5) and a fixed test set of 5 years (1980, 1995, 2006, 2010, 2015), selected to capture a representative range of diverse atmospheric and storm-surge conditions, while not being affected by any large-scale climate regime or trend. As a result of the k-fold cross-validation, the training produces an ensemble of 5 models.

2.3.1. Loss Functions

The surrogate model is trained by minimizing a loss function over the N time steps of the training set using gradient-based optimization. To evaluate the performance of surrogate models across both average and extreme regimes, we initially examined several candidate loss functions, including the Mean Squared Error (MSE), Quantile Loss (QL), Expectile Loss (EL), and a linear combination of quantile and expectile losses. Based on empirical evidence of their ability to optimize and balance overall accuracy (i.e., over average conditions) and the representation of extremes (see Figure S2 in Supporting Information S1), we selected two loss functions for detailed comparison within the proposed framework: (a) the Mean Squared Error (MSE) as a baseline, as it optimizes overall accuracy but performs poorly on extremes, and (b) a linear combination of QL and EL, referred to as the Weighted (sum of) Quantile and Expectile (WQE) loss, which enhances the representation of extremes while preserving overall performance (comparable to MSE).

The MSE, widely adopted for regression tasks, targets the conditional mean of the output distribution. For our bivariate output—tide and surge components—the MSE is formalized in Equation 4 and serves as the baseline loss function, consistent with previous machine learning applications to storm surge modeling (Harter et al., 2024; Ishida et al., 2020; Jiang et al., 2024). By minimizing the MSE, the model targets the conditional expected value of the output variable. Given our bivariate output, including tide and surge components (Equation 1), the MSE can be defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\mathrm{tide}} - \hat{y}_i^{\mathrm{tide}}\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\mathrm{surge}} - \hat{y}_i^{\mathrm{surge}}\right)^2 \tag{4}$$

On the other hand, aiming to improve the representation of the right tail of the surge distributions while preserving overall performance, we adopt a novel custom tilted loss function defined as a weighted sum of quantile and expectile losses, termed the WQE loss. QL regression targets the conditional quantile of the output variable, by minimizing the QL function (Equation 7) such that $100 \cdot \tau\%$ of the observations lie below the predicted value $\hat{y}$, where τ_1 denotes the target quantile (Koenker & Hallock, 2001). Expectile Loss (EL), targets the conditional expectile defined by τ_2 through a smooth, squared, asymmetric loss (see Equation 8), which makes it more sensitive to large deviations and tail values (Bellini & Di Bernardino, 2017; Newey & Powell, 1987). Empirical results show that EL outperforms QL in representing extremes while maintaining lower overall error, as shown in Figure S2 of Supporting Information S1.

Our proposed WQE loss combines QL and EL into a single objective function by a linear combination of QL and EL (Equation 6), and remains tilted for $\tau_2 > 0.5$ with a relatively large coefficient w_2 applied to the expectile term, thereby penalizing underestimation of the target variable. Related approaches based on weighted combinations of quantile and expectile losses were recently proposed in the literature, albeit with different estimation target (Atanane et al., 2025). Although the WQE loss does not explicitly target either the conditional quantile or the conditional expectile, it empirically yields improved performance in representing extreme values, for example, 100-year return period levels (similarly to EL), while simultaneously improving MSE, outperforming both QL and EL on average conditions (Figure S2 in Supporting Information S1).

To specifically target the right tail of the distributions, the linear weights of the quantile (w_1) and expectile (w_2) components are set to $w_1 = \frac{1}{6}$ and $w_2 = \frac{5}{6}$, with $\tau_1 = 0.25$ and $\tau_2 = 0.82$. These parameters are selected through a trial-and-error procedure aimed at maximizing tail performance without degrading average accuracy (see Figure S3 in Supporting Information S1 for a sensitivity analysis with respect to τ_2). For large error magnitudes, the expectile component becomes dominant, leading to a stronger penalization of underestimation, while the penalty smoothly vanishes as the error approaches zero (see Equation 6 and the graphical illustration of alternative

asymmetric loss functions in Figure S4 of Supporting Information S1). Compared to standard MSE regression, WQE regression systematically reduces peak underestimation, a behavior that directly addresses a known limitation of MSE-based approaches (Gupta et al., 2009; Harter et al., 2024). The WQE loss function is defined as follows:

$$\mathrm{WQE} = \frac{1}{N}\sum_{i=1}^{N}\rho\left(y_i^{\mathrm{tide}} - \hat{y}_i^{\mathrm{tide}}\right) + \frac{1}{N}\sum_{i=1}^{N}\rho\left(y_i^{\mathrm{surge}} - \hat{y}_i^{\mathrm{surge}}\right) \tag{5}$$

where:

$$\rho(\varepsilon) = w_1\varphi_1(\varepsilon) + w_2\varphi_2(\varepsilon) \tag{6}$$

$$\varphi_1(\varepsilon) = \begin{cases} \tau_1\varepsilon & \varepsilon \geq 0 \\ (\tau_1 - 1)\varepsilon & \varepsilon < 0 \end{cases} \tag{7}$$

$$\varphi_2(\varepsilon) = \begin{cases} \tau_2\varepsilon^2 & \varepsilon \geq 0 \\ (1 - \tau_2)\varepsilon^2 & \varepsilon < 0 \end{cases} \tag{8}$$

2.3.2. Hyperparameter Tuning

We evaluate multiple hyperparameter combinations (see Table S1 in Supporting Information S1) to identify the optimal network architecture. We select the hyperparameters via trial and error, minimizing the loss function on both the validation and training set. When the performance difference between alternative configurations is negligible, we select the model with fewer parameters to favor the simplest and efficient models. In addition, to increase model generalization and limit overfitting, we select hyperparameters that minimize the difference between loss functions in training and test sets. The relevant MSE scores for the hyperparameter selection on training and validation sets are reported separately for each architecture in Supporting Information: FFNN (Table S2 in Supporting Information S1), Conv (Table S3 in Supporting Information S1), Conv3D (Table S4 in Supporting Information S1), and LSTM (Table S5 in Supporting Information S1).

2.4. Extreme Value Analysis

Extreme storm surge events (particularly positive extremes) represent valuable information to design effective adaptation strategies, alongside projections of mean sea level rise. To evaluate the goodness of the model in reproducing positive storm surge extreme values, we perform a Peaks Over Threshold (POT) analysis by fitting a Generalized Pareto Distribution (GPD) (Equation 9)—a flexible distribution used to model exceedances over a high threshold—and compare the right tail behavior of our surrogate model with that of GTSM outputs.

$$\mathrm{P}(Y \leq y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{\frac{-1}{\xi}} \tag{9}$$

To estimate the shape and scale parameters (ξ and σ) of the GPD, we use the Maximum Likelihood Estimation method. For the POT analysis, we set the threshold to the 99th percentile of the distribution (computed over the whole period of interest), as in Muis et al. (2023), and we apply a 24-hr window for de-clustering to isolate the highest peak from each extreme storm surge event. To estimate the confidence intervals of the fitted distribution, we apply bootstrap resampling with 1,000 iterations from the original storm surge extremes samples (Qi, 2008). To assess the goodness of the surrogate model to emulate extreme events, we compute the bias of return value estimates. The return value is the level associated with a given return period $\bar{T}(\cdot)$ (Equation 10). The return period for a given return value y can be defined as:

$$\bar{T}(y) = \frac{1}{\lambda(1 - \mathrm{P}(Y \leq y))} \tag{10}$$

To compute the yearly return period, we use λ equal to the annual frequency based on the number of independent extreme events from the POT analysis (Lin et al., 2016).

2.5. Model Evaluation

Model performance is evaluated using four complementary metrics: Mean Squared Error (MSE), Quantile Loss (QL) for the quantile 0.9, Root Mean Squared Error (RMSE), and additive bias. MSE and RMSE quantify the average performance of the regression. RMSE is included for comparability with prior studies on data-driven storm surge modeling, and references therein Harter et al. (2024); Hermans et al. (2025); Kaufmann et al. (2024), and to provide a physically interpretable error metric in meters. The Quantile Loss evaluates how closely the model estimates a higher output value specifically close to the 0.9 quantile of the conditional output variable (thus the tendency to overestimate values). The additive bias is included to assess overestimation or underestimation of the model providing a physically interpretable metric in meters.

For the astronomical tide component, considered deterministic and bounded for a given mean sea level, we restrict evaluation to MSE and RMSE metrics. For the storm surge component, we supplement regression scores with Extreme Value Analysis (EVA) metrics to quantify the surrogate's ability to reproduce high-return period events. We select the best-performing model based on a trade-off between overall regression accuracy and fidelity in estimating extreme return levels. The ERA5-based training uses 5-fold cross-validation, producing an ensemble of five surrogate models whose average performance serves as the benchmark. For future projections, we directly compare surrogate predictions against GTSM outputs, quantifying improvements due to fine-tuning and evaluating differences between MSE- and MQE-trained models. Finally, we assess the projected changes in return levels for selected periods and validate against GTSM projections.

3. Results and Discussion

We first investigate the influence of the spatio-temporal domain size—defined by the number of grid cells in the bounding box and the number of time steps—on the performance of the selected models across the four architectures (FFNN, Conv, Conv3D, LSTM). This analysis is conducted over the historical period using ERA5 reanalysis data (see Figure 2). Second, we focus on the best-performing model identified in the first phase, and evaluate its ability to project extreme values. We compare performance with and without fine-tuning, to evaluate the model transferability to climate projections. Overall, the Results and Discussion section is structured according to the two-step framework depicted in Figure 2.

3.1. Model Evaluation on the Historical Reanalysis

Since tide levels are driven solely by astronomical forcing and mean sea level (which remains constant across the spatial domain considered), the regression performance for tides is unaffected by the spatial extent and improves only as the temporal window increases. The FFNN architecture achieves the best performance for tide prediction, with a 1×1 bounding box and a time window of 240 hr, yielding the lowest MSE scores on the test set (Table S6 in Supporting Information S1). This finding is consistent with previous studies (Ishida et al., 2020). The resulting tide model achieves a test RMSE of 3.6 cm, which aligns closely with existing benchmarks for data-driven tide forecasts, where RMSE values typically range between 3.6 and 4.9 cm (Yang et al., 2020; Zhang et al., 2023).

In contrast to tide levels, the regression accuracy for storm surge is highly sensitive to the spatiotemporal domain (Figure 3), consistently with previous findings (Tiggeloven et al., 2021). Model performance improves as the size of the spatiotemporal input domain increases, with average scores on the test set showing substantial gains (Figure 3). All four architectures achieve their best results with a spatial domain of 10×15 grid cells and a temporal window of 24 hr, that is, the largest spatiotemporal extent considered for the three-dimensional input domain. Across all architectures, expanding the input domain from 6×6 grid cells and a 6-hr time window to 10×15 grid cells and 24 hr yields an average improvement of 58% in MSE and 31% in QL when using MSE loss and 52% in MSE and 34% in QL when using WQE loss.

All architectures perform comparably well on the test set for both MSE and QL scores, although the 2D and 3D convolutional networks exhibit slightly superior performance, achieving improvements of up to 10%. A comparison of models trained separately with the two loss functions (Figures 3a and 3b) reveals that MSE-trained

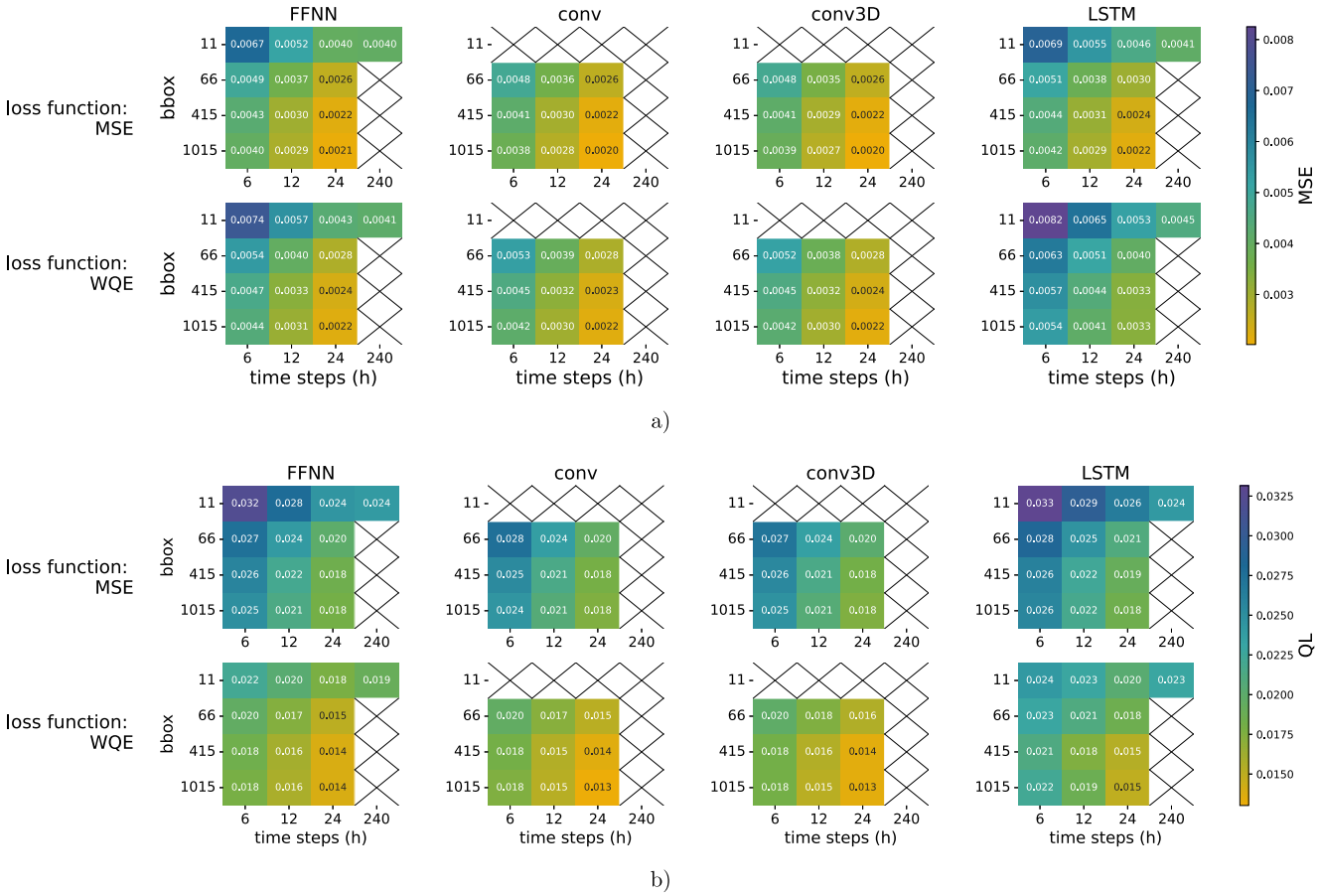

Figure 3. Comparative evaluation of surrogate model performance across different spatio-temporal input configurations with heatmaps showing the MSE (a) and QL (b) scores on the test set for different bounding box (bbox) sizes and time steps. The scores are computed for the FFNN, conv, conv3D and LSTM models ensembles. Both panels (a, b) show the scores for models trained with MSE loss (upper row) and WQE loss (lower row). The bounding box (bbox) sizes considered are: 1×1 (11), 6×6 (66), 4×15 (415), and 10×15 (1015) grid cells. The 1×1 spatial domain is used only with FFNN and LSTM models, and the 240 hr time step is used only with the 1×1 spatial domain.

models achieve 14% lower MSE scores relative to WQE-trained models, while WQE-trained models produce 25% lower QL scores than their MSE-trained counterparts.

Among the architectures, the 2D convolutional network emerges as the best-performing model for storm surge estimation with both metrics. Notably, the lowest RMSE of 4.5 cm is achieved by the 2D convolutional model trained with the MSE loss function.

To estimate the total water level (defined as the sum of tides and storm surges), different models can be selected to optimize the accuracy of each component, that is, using the FFNN for tides and the 2D convolutional network for storm surges—and their outputs combined (see Figure S5 in Supporting Information S1). This approach achieves the lowest average RMSE for total water level at 5.7 cm on the test set. State-of-the-art neural network models for surge and total water level regression typically target either gauge observations (Ishida et al., 2020; Kaufmann et al., 2024) or outputs from dynamic models (Ayyad et al., 2023; Xie et al., 2023). These studies often focus on either surrogate modeling for simulation tasks (Ayyad et al., 2023; Harter et al., 2024; Ishida et al., 2020; Kaufmann et al., 2024; Tadesse et al., 2020) or forecasting applications (Ramos-Valle et al., 2021; Xie et al., 2023). Reported RMSE values in state-of-the-art studies range from 1 to 8 cm globally, and from 3.5 to 8.5 cm specifically along the New York City coastline (Ayyad et al., 2023; Ramos-Valle et al., 2021). Despite the diversity of data-driven frameworks and case studies in the literature, our RMSE values for storm surge (4.5 cm) and total water level (5.7 cm) are comparable to existing benchmarks.

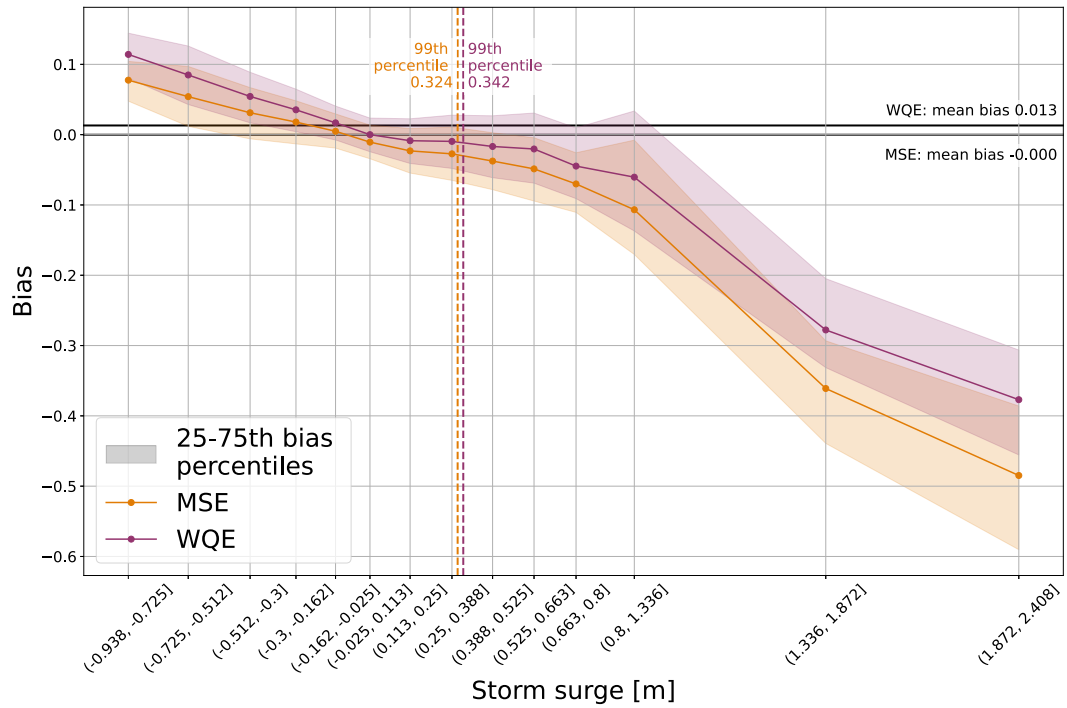

Figure 4. Bias for each interval along the storm surge distribution. Solid horizontal lines represent the score over the entire data set. Dashed vertical lines represent the 99th percentile of the distribution which is used as threshold for the POT extreme analysis.

Although the two loss functions produce broadly similar results and lead to consistent conclusions regarding optimal hyperparameters and spatiotemporal domain size, their performance diverges across the range of storm surge magnitudes (Figure 4). Overall, the models tend to overestimate negative surge values and underestimate positive ones, with the magnitude of both biases increasing proportionally to the absolute surge value. A similar bias pattern has been reported in other case studies (Harter et al., 2024). Compared to the MSE loss, the WQE loss tends to further overestimate negative surge values but mitigates the underestimation of positive ones. This difference in bias is roughly proportional to the absolute surge magnitude, reflecting the variability of the conditional distribution for larger absolute values. Consequently, the WQE-trained models substantially offer more accurate estimates of storm surge extremes, which is crucial for informing coastal adaptation planning.

The bias in the 10- and 100-year return period (RP) storm surge estimates is sensitive to the size of the spatio-temporal input domain. In general, the accuracy of 10- and 100-year RP estimates improves as the input domain expands, with the notable exception of the LSTM model, which shows no reduction in bias with increasing spatial extent (Figure 5). Despite substantial variability in EVA bias scores, the FFNN, 2D convolutional, and 3D convolutional models exhibit a consistent reduction in the underestimation of return values when trained on the largest spatio-temporal domains, although a saturation of performance is visible for the two largest spatial domains (from 4 × 15 to 10 × 15 grid cells, with 24 hr time window).

The MSE and WQE loss functions also produce distinct patterns in return value estimation. Models trained with the WQE loss demonstrate lower bias on extremes compared to MSE-trained models, in line with the improvements across the full storm surge distribution (Figure 4). Performance further varies by architecture, with WQE-trained convolutional models—particularly the 2D one—achieving superior accuracy in estimating 10- and 100-year RP extremes by reducing underestimation of return values. Taken together, regression and EVA analyses identify the 2D convolutional model as the best, achieving both the highest overall regression accuracy and the most reliable return value estimates.

The 2D convolutional model trained with the WQE loss function substantially reduces the underestimation of return values compared to the MSE loss, improving estimates by 59%–97% across return periods from 1 to 1,000 years. This best-performing model achieves the most accurate representation of extremes, with

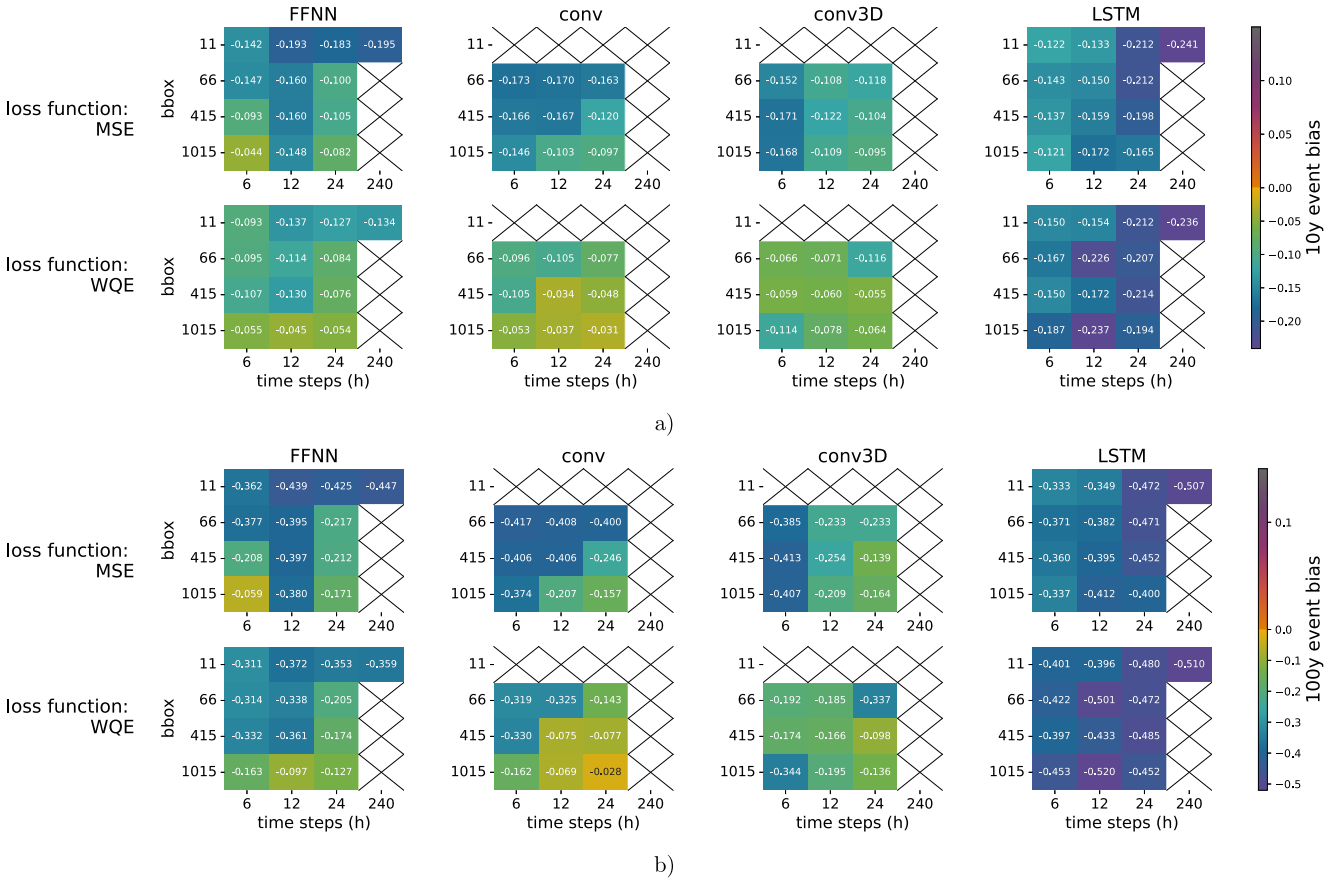

Figure 5. Comparative evaluation of surrogate model performance across different spatio-temporal input configurations with heatmaps showing the bias on 10-year Return Period (RP) events (a) and 100-year RP events (b) on the test set. The scores are computed for the FFNN, conv, conv3D and LSTM models ensembles. Both panels (a, b) show the scores for models trained with the MSE loss (upper row) and the WQE loss (lower row). The bounding box sizes considered are: 1×1 (11), 6×6 (66), 4×15 (415), and 10×15 (1015) grid cells. The 1×1 spatial domain is used only with FFNN and LSTM models, and the 240 hr time step is used only with the 1×1 spatial domain.

underestimations of return values of only 3.4%, 2.9%, 3.7%, and 0.2% for the 1-, 10-, 100-, and 1,000-year RPs, respectively, over the historical period (Figure 6). Furthermore, relative to the MSE-trained models, the WQE-trained model reduces the underestimation of the upper bound of the 95% confidence interval by 61%, 63%, 77%, and 99% for the same return periods. Overall, adopting the WQE loss function yields more conservative

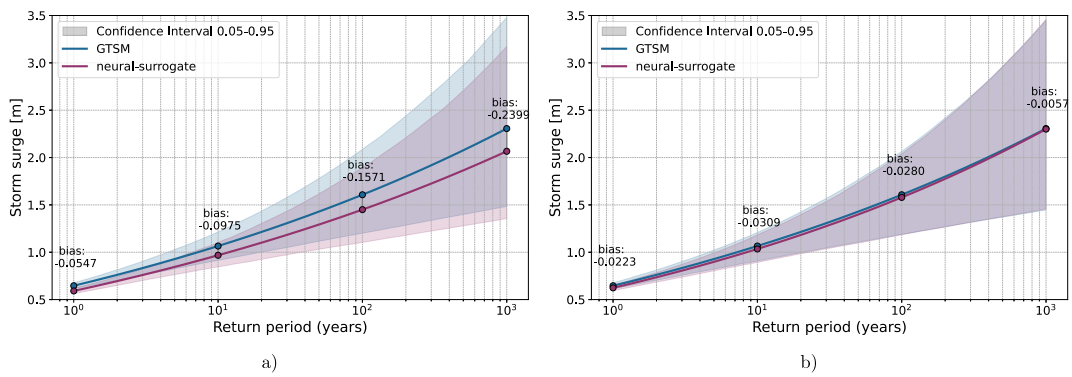

Figure 6. Comparison of the bias of extreme return values estimates from 1 to 1,000-year RP for the best model ensemble obtained with the MSE loss function (a) and WQE loss function (b) over the historical period. The EVA results here refer to the reanalysis data, with EVA performed on the validation and test set together.

Table 1
Comparison of MSE, QL, and RMSE Scores on the Projected Storm Surge Time Series for Surrogate Models and Fine-Tuned Surrogate Models Trained Using Either the MSE or WQE Loss Function

Model	MSE	QL	RMSE
Surrogate MSE	0.0048	0.0268	0.0694
Surrogate WQE	0.0047	0.0187	0.0685
Surrogate MSE fine-tuned	0.0040	0.0197	0.0634
Surrogate WQE fine-tuned	0.0045	0.0152	0.0673

estimates of surge peaks, mitigating the underestimation of the most severe low-probability events. The convolutional model trained with WQE on a 10 × 15 spatial grid and a 24-hr temporal domain is identified as the best-performing architecture when considering both regression scores and extreme value estimates. The return value bias scores for the 1-, 10-, 100-, and 1,000-year RPs are −2.2, −3.1, −2.8, and −0.6 cm, respectively. Notably, the 10-year RP bias of −3.1 cm is competitive in magnitude to that of GTSM, which shows an average bias of −10 cm for 10-year RP events when evaluated globally against the tide gauge network (Muis et al., 2020).

3.2. Fine Tuning and Future Projection

We assess the performance of models trained with both MSE and WQE loss functions under future scenarios to determine whether the advantages of the WQE-trained model observed on historical reanalysis data persist in projections. For this purpose, we use atmospheric forcing projections from the HighResMIP experiment with the CMCC-CM2-VHR4 global circulation model (GCM), consistently with the forcing used in surge projections with GTSM. We also compare the models derived from hyperparameter tuning on the reanalysis data set with their counterparts fine-tuned on the historical simulation of the GCM. Fine-tuning is performed to adjust for minor discrepancies between the spatio-temporal domains of the GCM projections and the reanalysis data. To ensure consistency in spatio-temporal resolution, the GCM projections are linearly interpolated to match the reanalysis grid and temporal resolution used in the training phase. The fine-tuning process improves performance in future scenarios for both MSE- and WQE-trained models (Table 1). The improvement from fine-tuning is more pronounced for the MSE-trained model than for the WQE-trained model. Specifically, performance gains for the MSE- and WQE-trained models are 16.7% and 4.3% in terms of MSE score, 26.5% and 18% for QL, and 8.6% and 1.8% for RMSE, respectively. Across all three metrics (MSE, QL, and RMSE), fine-tuning enhances predictive performance while preserving the relative differences between MSE- and WQE-trained models established during the initial hyperparameter tuning phase.

Figure 7 illustrates the four highest storm surge peaks in the future trajectory for the four models compared in Table 1. The trajectories highlight the differences in estimates across models, with the WQE-trained model showing a clear tendency to reduce the underestimation observed in the MSE-trained model, in some cases shifting toward a slight overestimation. This behavior aligns with the performance scores and EVA results, which

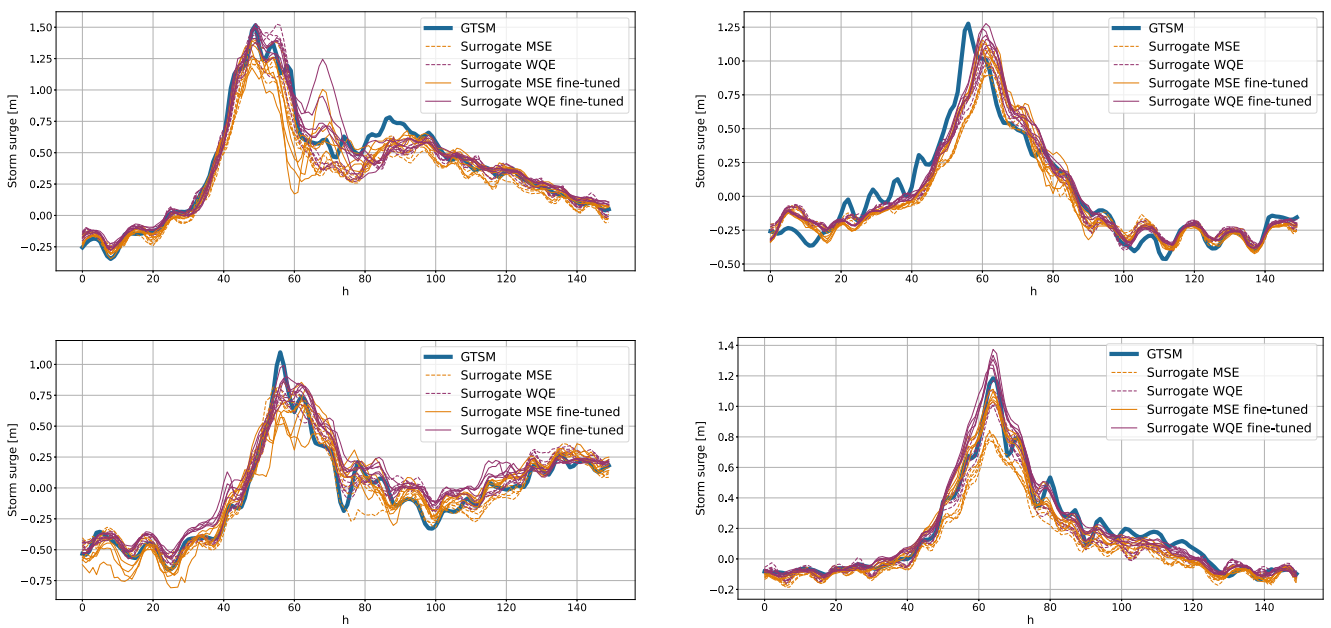

Figure 7. Storm surge peaks projections of the four highest peaks in the 2016–2050 horizon simulated with the CMCC-CM2-VHR2 GCM. Comparison between fine-tuned models (solid lines) and models obtained at the first step (dashed lines); models are trained either with the MSE loss (orange lines) or the WQE loss (purple lines).

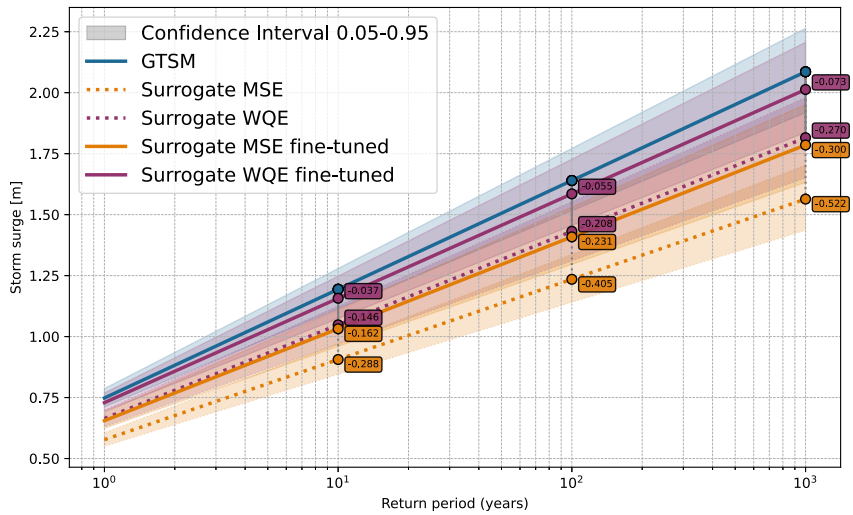

Figure 8. Comparison of bias of future return values estimates, with the fine-tuned models (solid lines) and the models obtained at the first step, without fine-tuning (dashed lines); the models are trained with either the MSE (orange lines) or the WQE loss function (purple lines).

quantify the improved handling of extremes by the WQE model. The divergence between the two loss functions becomes particularly evident at surge peaks where the conditional probability distribution exhibits higher variance, as the WQE result and expected value (MSE) diverge more substantially.

The model trained with the WQE loss function delivers more accurate estimates of return values in future projections as well (Figure 8). Consistent with the overall regression score results, fine-tuning both models on the historical GCM simulation also enhances their accuracy in estimating return values. Notably, the WQE fine-tuned model produces estimates that are closer to the actual return period values and associated confidence intervals from the GTSM reference projection. The WQE-trained model provides a better representation of extremes even without the fine-tuning step, outperforming the MSE-trained model after fine tuning. When fine tuning is applied, the WQE-trained model achieves a substantially larger improvement in extreme-value representation, with average bias reductions of 43% and 73.7% for the MSE- and WQE-trained models, respectively, across the 10- to 1,000-year return periods (Figure 8).

Finally, given the superior performance of the WQE-trained and fine-tuned model, particularly in projecting extreme events, this model is better suited than the MSE-trained and non-fine-tuned counterpart for assessing changes in extreme return values between the historical reanalysis and future projections (Figure 9).

Assuming time-invariant model bias—a reasonable assumption supported by the results in Figures 6 and 8—the surrogate model enables direct estimation of projected changes in return period events for the considered GCM scenario. For both storm surge and storm tide, the surrogate model reproduces trends closely aligned with GTSM results (Figures 9a and 9b). Specifically, the storm surge projections of GTSM and the surrogate models show changes of +15% (GTSM) and +16% (surrogate) for the 1-year RP, +11% and +11% for the 10-year RP, +1.7% and +0.01% for the 100-year RP, and −9.5% and −13% for the 1,000-year RP event, respectively. Similarly, for storm tide return values, the projected increases are +6% (GTSM) and +9% (surrogate) for the 1-year RP, +12% and +13% for the 10-year RP, +19% and +16% for the 100-year RP, and 28% and 18% for the 1,000-year RP event (Figure 9). Future storm surge projections show a moderate increase in event heights for return periods up to 100 years, while events with return periods longer than 100 years tend to decrease in magnitude. This difference can be due to the presence of an extremely rare event in the historical reanalysis (Hurricane Sandy, 2012), which influences the estimation of extreme return levels. On the other hand, for the projected storm surge, return levels with the same magnitude are extrapolated using the GPD distribution. For storm tide, the differences across return periods are higher. This behavior can mainly be attributed to changes in the astronomical tide height together with the random timing between tide and surge in the projected series, which likely leads to higher estimated extreme events. It is important to note that future projections obtained with the neural-network surrogate closely reproduce the response of the GTSM model under climate forcing, capturing projected trends and variability in extreme

10.1029/2025EF007072

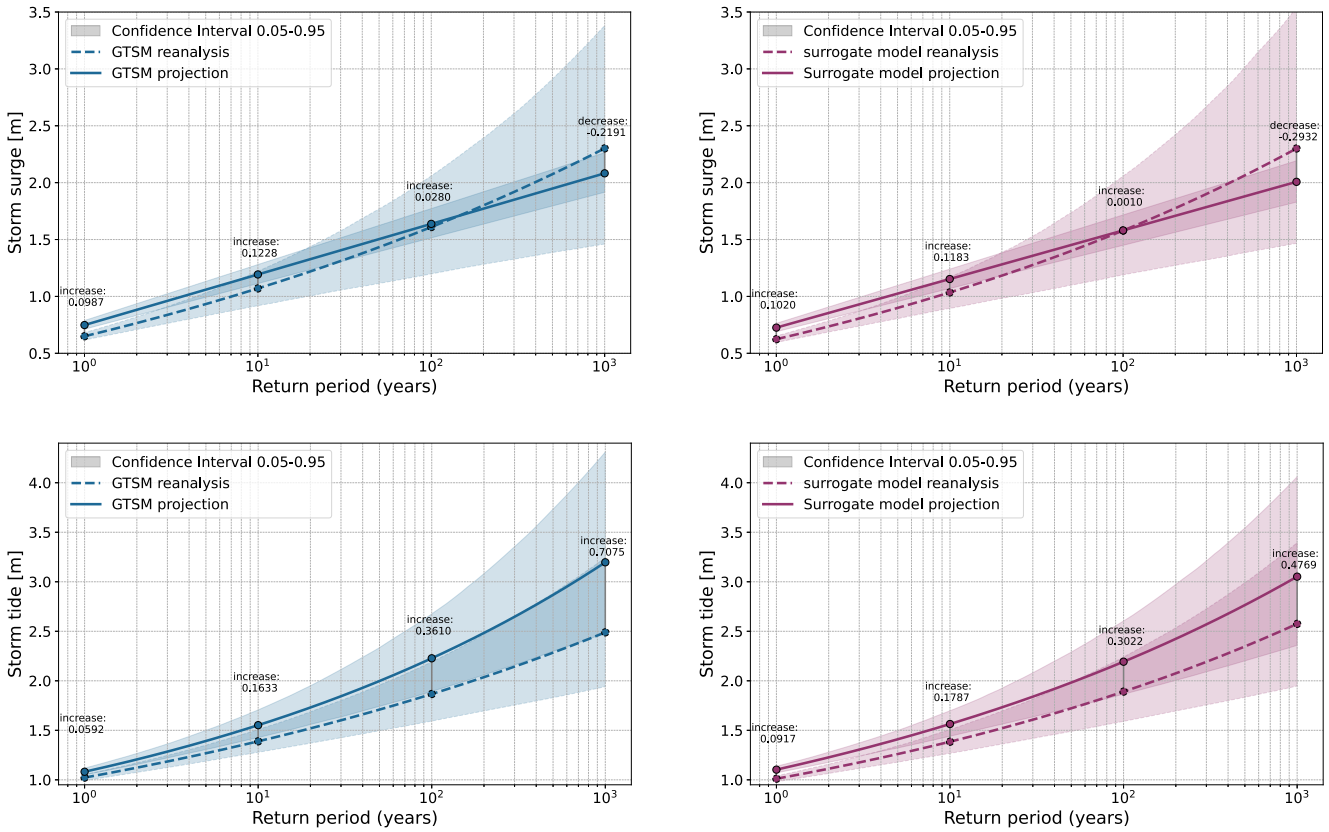

Figure 9. Changes in return values from historical reanalysis period (dashed line) to future projection period (solid line) for the GTSM simulation (a) and the surrogate model trained with the WQE loss (b), both forced by the same GCM projection. Results are shown for both storm surge (upper row) and storm tide (lower row). Shaded areas denote the 5%–95% confidence intervals. Text annotations indicate the absolute changes in return level (in meters) for selected return periods, with positive values indicating increases and negative values indicating decreases.

events (with comparable uncertainty ranges). Projected changes in extreme water levels therefore reflect the behavior of GTSM and are not intended to provide an independent physical interpretation derived from the surrogate itself. Detailed analyses of future projections, including results for the same case study, are provided in Muis et al. (2020), while further discussion of key findings and differences across HighResMIP GCM projections is presented in Muis et al. (2023).

This strong agreement highlights the consistency of the surrogate model in emulating hydrodynamic outputs, demonstrating its potential for efficiently estimating projected changes in storm surge risk under future scenarios. Once trained on historical simulations, the surrogate model offers two key advantages over physics-based models like GTSM: (a) a significant reduction of computational cost, and (b) a greater flexibility to focus on specific local targets. However, our results show that fine-tuning the surrogate on each climate model using corresponding hydrodynamic simulations is critical to maintain high accuracy under future forcings. This requirement limits model transferability and presents a barrier to scaling ensemble projections across multiple GCMs (especially when hydrodynamic simulations are not available)—unless surrogate models can be enhanced to generalize without climate model-specific retraining.

4. Conclusions

Storm surge modeling is computationally intensive, often requiring substantial time and computing resources that limit its feasibility for ensemble generation and large-scale applications. AI-based surrogate models offer a promising, time-efficient alternative to traditional physics-based dynamical models. However, existing AI

surrogates in the literature tend to underestimate storm surge extremes, and their accuracy under future climate scenarios remains largely unexplored.

In this work, we present a globally applicable framework for developing storm surge and tide surrogate models of the Global Tide and Surge Model (GTSM). For the first time, we leverage the global surge projections produced by GTSM, forced with both historical reanalysis data and HighResMIP climate scenarios, to train and evaluate AI models. We demonstrate the effectiveness of this framework through a case study in New York City (NYC), selected for its high exposure to coastal flood risk, complex tidal and surge dynamics, and its status as a well-studied benchmark for comparison with state-of-the-art models.

Using ERA5 reanalysis data, we assess the sensitivity of different surrogate model architectures to the spatio-temporal input domain, evaluating both regression accuracy and the ability to predict extreme events. Leveraging HighResMIP simulations from the CMCC-CM2-VHR4 global climate model, we project storm surge and storm tide trajectories and return levels through 2050, comparing surrogate predictions with GTSM projections. Two loss functions are tested during neural network training: Mean Squared Error (MSE) and the linear combination of QL and EL, referred to as the Weighted (sum of) Quantile and Expectile (WQE) loss. The WQE-trained model emphasizes performance on extremes, showing a reduction in the underestimation of positive extremes when compared with the MSE-trained model. We also evaluate surrogate performance in regression and extreme value estimation under future climate forcing using GTSM projections.

Our results reveal three key findings: (a) Larger spatiotemporal input domains improve both regression accuracy and extreme event estimation, with convolutional architectures outperforming FFNN and LSTM models, particularly for extreme levels. (b) The asymmetric WQE loss function effectively reduces the underestimation of extremes, outperforming MSE in return value predictions. (c) When fine-tuned on historical GCM simulations, the surrogate model maintains consistent performance under future scenarios, closely matching GTSM-projected changes in return levels.

In summary, this work demonstrates that AI-based surrogate models can deliver reliable storm surge projections, including extremes, when trained with an appropriate loss function and fine-tuned on climate model data. Although the present surrogate model relies on locally sampled atmospheric drivers, this setup does not explicitly capture large-scale dynamics, such as basin-scale pressure gradients. Future development could incorporate large-scale drivers at a coarse-resolution or circulation indices as predictors, to better represent large-scale dynamical drivers. The proposed framework, based on globally available data sets, is potentially transferable to other locations and can support practical applications in coastal adaptation planning. Under the current formulation, once trained, our surrogate model remains location-specific, as it is not trained by providing local morphological information, but implicitly learns the storm-surge dynamics of the training location. Further work could investigate the transferability of our surrogate or similar models when informed with location-specific, morphological information on bathymetry and coastal geometry. Nevertheless, our findings indicate that surrogate model performance is sensitive to GCM-specific characteristics of the forcing data, highlighting current limitations in cross-model transferability. While fine-tuning with historical simulations from the target GCM substantially improves performance under future scenarios, the fine-tuning requirement may limit the operational use of surrogates in multi-model ensemble frameworks, especially for GCMs lacking historical simulations with hydrodynamic models such as GTSM. Addressing these limitations will require dedicated research on generalization strategies aimed at reducing dependence on model-specific tuning. Future research should also extend this analysis to additional case studies and GCMs to develop regional and ensemble-based AI models for robust climate risk assessments.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

All GTSM data (Muis et al., 2023) are publicly available from the Climate Data Store (CDS) at: https://doi.org/10.24381/cds.a6d42d60 (Copernicus Climate Change Service, 2022) and can be retrieved using the CDS API. All the ERA5 data (Hersbach et al., 2020) are publicly available from the Climate Data Store (CDS) at: https://doi.org/10.

24381/cds.e2161bac (C3S, 2018). All the HighResMIP data (Haarsma et al., 2016) can be obtained from the Earth System Grid Federation (ESGF) data portal (https://esgf-metagrid.cloud.dkrz.de/search). The azimuth and altitude of the Moon and Sun were modelled in a Python environment using the Ephem package (https://pypi.org/project/ephem/). The software and analysis scripts (Python code) used to develop the surrogate model, and reproduce the results and figures are available at: https://doi.org/10.5281/zenodo.16748613 (Longo, 2026).

Acknowledgments
Emiliano Longo and Andrea Ficchì acknowledge support from the AXA Research Fund Fellowship on Coastal Livelihoods, under the PRINTFLOODS (Prediction Intelligence for Floods) project. Emiliano Longo is also funded by a PhD scholarship of Politecnico di Milano. Andrea Castelletti acknowledges funding by the EU Horizon 2020 project CLINT (Climate Intelligence: Extreme events detection, attribution and adaptation design using machine learning) under Grant Agreement 101003876. We would like to thank Joppe Vermeulen and Chiheb Ben Hammouda (Utrecht University) for their helpful comments on the loss functions. Open access publishing facilitated by Politecnico di Milano, as part of the Wiley - CRUI-CARE agreement.

References

Atanane, A., Mkhadri, A., & Oualkacha, K. (2025). An efficient hybrid approach of quantile and expectile regression. *Statistical Papers*, *66*(6), 144. https://doi.org/10.1007/s00362-025-01761-3

Ayyad, M., Hajj, M. R., & Marsooli, R. (2023). Climate change impact on hurricane storm surge hazards in New York/New Jersey coastlines using machine-learning. *npj Climate and Atmospheric Science*, *6*(1), 88. https://doi.org/10.1038/s41612-023-00420-4

Aziz, F., Wang, X., Mahmood, M. Q., Awais, M., & Trenouth, B. (2024). Coastal urban flood risk management: Challenges and opportunities A systematic review. *Journal of Hydrology*, *645*, 132271. https://doi.org/10.1016/j.jhydrol.2024.132271

Bellini, F., & Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance*, *23*(6), 487–506. https://doi.org/10.1080/1351847X.2015.1052150

Benito, I., Aerts, J. C. J. H., Eilander, D., Ward, P. J., & Muis, S. (2024). Stochastic coastal flood risk modelling for the east coast of Africa. *npj Natural Hazards*, *1*(1), 10. https://doi.org/10.1038/s44304-024-00010-1

Bernier, N. B., Hemer, M., Mori, N., Appendini, C. M., Breivik, O., De Camargo, R., et al. (2024). Storm surges and extreme sea levels: Review, establishment of model intercomparison and coordination of surge climate projection efforts (SurgeMIP). *Weather and Climate Extremes*, *45*, 100689. https://doi.org/10.1016/j.wace.2024.100689

Betancourt, J., Bachoc, F., Klein, T., Idier, D., Pedreros, R., & Rohmer, J. (2020). Gaussian process metamodeling of functional-input code for coastal flood hazard assessment. *Reliability Engineering & System Safety*, *198*, 106870. https://doi.org/10.1016/j.ress.2020.106870

Bevacqua, E., Vousdoukas, M. I., Zappa, G., Hodges, K., Shepherd, T. G., Maraun, D., et al. (2020). More meteorological events that drive compound coastal flooding are projected under climate change. *Communications Earth & Environment*, *1*(1), 47. https://doi.org/10.1038/s43247-020-00044-z

Bruneau, N., Polton, J., Williams, J., & Holt, J. (2020). Estimation of global coastal sea level extremes using neural networks. *Environmental Research Letters*, *15*(7), 074030. https://doi.org/10.1088/1748-9326/ab89d6

C3S. (2018). ERA5 hourly data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/CDS.ADBB2D47

Camps-Valls, G., Fernández-Torres, M.-, Cohrs, K.-H., Höhl, A., Castelletti, A., Pacal, A., et al. (2024). AI for extreme event modeling and understanding: Methodologies and challenges. *arXiv*. Retrieved from http://arxiv.org/abs/2406.20080(arXiv:2406.20080[cs])

Catalano, A. J., & Broccoli, A. J. (2018). Synoptic characteristics of surge-producing extratropical cyclones along the northeast coast of the United States. *Journal of Applied Meteorology and Climatology*, *57*(1), 171–184. https://doi.org/10.1175/JAMC-D-17-0123.1

Colle, B. A., Buonaiuto, F., Bowman, M. J., Wilson, R. E., Flood, R., Hunter, R., et al. (2008). New York city's vulnerability to coastal flooding: Storm surge modeling of past cyclones. *Bulletin of the American Meteorological Society*, *89*(6), 829–842. https://doi.org/10.1175/2007BAMS2401.1

Copernicus Climate Change Service. (2022). Global sea level change time series from 1950 to 2050 derived from reanalysis and high resolution CMIP6 climate projections [Dataset]. *ECMWF*. https://doi.org/10.24381/CDS.A6D42D60

Eilander, D., Couasnon, A., Sperna Weiland, F. C., Ligtvoet, W., Bouwman, A., Winsemius, H. C., & Ward, P. J. (2023). Modeling compound flood risk and risk reduction using a globally applicable framework: A pilot in the Sofala province of Mozambique. *Natural Hazards and Earth System Sciences*, *23*(6), 2251–2272. https://doi.org/10.5194/nhess-23-2251-2023

Feng, K., Lin, N., Kopp, R. E., Xian, S., & Oppenheimer, M. (2025). Reinforcement learning–based adaptive strategies for climate change adaptation: An application for coastal flood risk management. *Proceedings of the National Academy of Sciences*, *122*(12), e2402826122. https://doi.org/10.1073/pnas.2402826122

Garner, A. J., Mann, M. E., Emanuel, K. A., Kopp, R. E., Lin, N., Alley, R. B., et al. (2017). Impact of climate change on New York City's coastal flood hazard: Increasing flood heights from the preindustrial to 2300 CE. *Proceedings of the National Academy of Sciences*, *114*(45), 11861–11866. https://doi.org/10.1073/pnas.1703568114

Garner, G. G., & Keller, K. (2018). Using direct policy search to identify robust strategies in adapting to uncertain sea-level rise and storm surge. *Environmental Modelling & Software*, *107*, 96–104. https://doi.org/10.1016/j.envsoft.2018.05.006

Georgas, N., Orton, P., Blumberg, A., Cohen, L., Zarrilli, D., & Yin, L. (2014). The impact of tidal phase on Hurricane Sandy's flooding around New York City and long Island sound. *Journal of Extreme Events*, *1*(1), 1450006. https://doi.org/10.1142/S2345737614500067

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*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High resolution model intercomparison project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Harter, L., Pineau-Guillou, L., & Chapron, B. (2024). Underestimation of extremes in sea level surge reconstruction. *Scientific Reports*, *14*(1), 14875. https://doi.org/10.1038/s41598-024-65718-6

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., & Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty. *Water Resources Research*, *56*(2), e24389. https://doi.org/10.1029/2019WR025502

Hermans, T. H. J., Ben Hammouda, C., Treu, S., Tiggeloven, T., Couasnon, A., Busecke, J. J. M., & Van De Wal, R. S. W. (2025). Computing extreme storm surges in Europe using neural networks. Sea. *Ocean and Coastal Hazards*. https://doi.org/10.5194/egusphere-2025-196

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Intergovernmental Panel On Climate Change (Ipcc). (2022). *The ocean and cryosphere in a changing climate: Special report of the intergovernmental panel on climate change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157964

Intergovernmental Panel On Climate Change (Ipcc). (2023). *Climate change 2021 – The physical science basis: Working group I contribution to the sixth assessment report of the intergovernmental panel on climate change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Irish, J. L., & Cañizares, R. (2009). Storm-wave flow through tidal inlets and its influence on Bay flooding. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *135*(2), 52–60. https://doi.org/10.1061/(ASCE)0733-950X(2009)135:2(52)

Ishida, K., Tsujimoto, G., Ercan, A., Tu, T., Kiyama, M., & Amagasaki, M. (2020). Hourly-scale coastal sea level modeling in a changing climate using long short-term memory neural network. *Science of the Total Environment*, *720*, 137613. https://doi.org/10.1016/j.scitotenv.2020.137613

Jiang, W., Zhong, X., & Zhang, J. (2024). Surge-NF: Neural fields inspired peak storm surge surrogate modeling with multi-task learning and positional encoding. *Coastal Engineering*, *193*, 104573. https://doi.org/10.1016/j.coastaleng.2024.104573

Kaufmann, C. L. G., Gallo, M. N., & De Camargo, R. (2024). Predicting storm surge extremes on the southeast Brazilian Coast: Long-term projections with neural networks. *Regional Studies in Marine Science*, *79*, 103846. https://doi.org/10.1016/j.rsma.2024.103846

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. https://doi.org/10.48550/ARXIV.1412.6980

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

Knutson, T., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C.-H., Kossin, J., et al. (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, *101*(3), E303–E322. https://doi.org/10.1175/BAMS-D-18-0194.1

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *The Journal of Economic Perspectives*, *15*(4), 143–156. https://doi.org/10.1257/jep.15.4.143

Kopp, R. E., Garner, G. G., Hermans, T. H. J., Jha, S., Kumar, P., Reedy, A., et al. (2023). The framework for assessing changes to Sea-level (FACTS) v1.0: A platform for characterizing parametric and structural uncertainty in future global, relative, and extreme sea-level change. *Geoscientific Model Development*, *16*(24), 7461–7489. https://doi.org/10.5194/gmd-16-7461-2023

Kopp, R. E., Oppenheimer, M., O'Reilly, J. L., Drijfhout, S. S., Edwards, T. L., Fox-Kemper, B., et al. (2023). Communicating future sea-level rise uncertainty and ambiguity to assessment users. *Nature Climate Change*, *13*(7), 648–660. https://doi.org/10.1038/s41558-023-01691-8

Krien, Y., Dudon, B., Roger, J., & Zahibo, N. (2015). Probabilistic hurricane-induced storm surge hazard assessment in Guadeloupe, lesser Antilles. *Natural Hazards and Earth System Sciences*, *15*(8), 1711–1720. https://doi.org/10.5194/nhess-15-1711-2015

Lashley, C. H., Puleo, J., Shi, F., & Nederhoff, K. (2025). Role of waves in forecasting extreme coastal flooding under a warming climate: Insights from Norfolk, Virginia. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *151*(3), 05025001. https://doi.org/10.1061/JWPED5.WWENG-2169

Lecacheux, S., Rohmer, J., Paris, F., Pedreros, R., Quetelard, H., & Bonnardot, F. (2021). Toward the probabilistic forecasting of cyclone-induced marine flooding by overtopping at Reunion Island aided by a time-varying random-forest classification approach. *Natural Hazards*, *105*(1), 227–251. https://doi.org/10.1007/s11069-020-04307-y

Leijnse, T., Van Ormondt, M., Nederhoff, K., & Van Dongeren, A. (2021). Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind- and wave-driven processes. *Coastal Engineering*, *163*, 103796. https://doi.org/10.1016/j.coastaleng.2020.103796

Lin, N., Emanuel, K., Oppenheimer, M., & Vanmarcke, E. (2012). Physically based assessment of hurricane surge threat under climate change. *Nature Climate Change*, *2*(6), 462–467. https://doi.org/10.1038/nclimate1389

Lin, N., Kopp, R. E., Horton, B. P., & Donnelly, J. P. (2016). Hurricane Sandy's flood frequency increasing from year 1800 to 2100. *Proceedings of the National Academy of Sciences*, *113*(43), 12071–12075. https://doi.org/10.1073/pnas.1604386113

Lin, N., Marsooli, R., & Colle, B. A. (2019). Storm surge return levels induced by mid-to-late-twenty-first-century extratropical cyclones in the Northeastern United States. *Climatic Change*, *154*(1–2), 143–158. https://doi.org/10.1007/s10584-019-02431-8

Liu, X., & Guillas, S. (2016). Dimension reduction for Gaussian process emulation: An application to the influence of bathymetry on tsunami heights. *arXiv*. https://doi.org/10.48550/arXiv.1603.07888

Longo, E. (2026). A deep learning framework for extreme storm surge modelling under future climate scenarios [Software]. *Software Release v1.3. Zenodo*. https://doi.org/10.5281/ZENODO.16748613

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

Mayo, T. L., & Lin, N. (2022). Climate change impacts to the coastal flood hazard in the northeastern United States. *Weather and Climate Extremes*, *36*, 100453. https://doi.org/10.1016/j.wace.2022.100453

Muis, S., Aerts, J. C. J. H., Antolínez, J. A., Dullaart, J. C., Duong, T. M., Erikson, L., et al. (2023). Global projections of storm surges using high-resolution CMIP6 climate models. *Earth's Future*, *11*(9), e2023EF003479. https://doi.org/10.1029/2023EF003479

Muis, S., Apecechea, M. I., Dullaart, J., De Lima Rego, J., Madsen, K. S., Su, J., et al. (2020). A high-resolution global dataset of extreme Sea levels, tides, and storm surges, including future projections. *Frontiers in Marine Science*, *7*, 263. https://doi.org/10.3389/fmars.2020.00263

Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*(4), 819. https://doi.org/10.2307/1911031

Oddo, P. C., Lee, B. S., Garner, G. G., Srikrishnan, V., Reed, P. M., Forest, C. E., & Keller, K. (2020). Deep uncertainties in sea-level rise and storm surge projections: Implications for coastal flood risk management. *Risk Analysis*, *40*(1), 153–168. https://doi.org/10.1111/risa.12888

O'Grady, J. G., Stephenson, A. G., & McInnes, K. L. (2022). Gauging mixed climate extreme value distributions in tropical cyclone regions. *Scientific Reports*, *12*(1), 4626. https://doi.org/10.1038/s41598-022-08382-y

Orton, P., Hall, T., Talke, S., Blumberg, A., Georgas, N., & Vinogradov, S. (2016). A validated tropical-extratropical flood hazard assessment for New York Harbor: Flood assessment for New York Harbor. *Journal of Geophysical Research: Oceans*. https://doi.org/10.1002/2016JC011679

Pachev, B., Arora, P., del Castillo-Negrete, C., Valseth, E., & Dawson, C. (2023). A framework for flexible peak storm surge prediction. *Coastal Engineering*, *186*, 104406. https://doi.org/10.1016/j.coastaleng.2023.104406

Patricola, C. M., & Wehner, M. F. (2018). Anthropogenic influences on major tropical cyclone events. *Nature*, *563*(7731), 339–346. https://doi.org/10.1038/s41586-018-0673-2

Qi, Y. (2008). Bootstrap and empirical likelihood methods in extremes. *Extremes*, *11*(1), 81–97. https://doi.org/10.1007/s10687-007-0049-8

Qin, Y., Su, C., Chu, D., Zhang, J., & Song, J. (2023). A review of application of machine learning in storm surge problems. *Journal of Marine Science and Engineering*, *11*(9), 1729. https://doi.org/10.3390/jmse11091729

Ragno, E., Antonini, A., & Pasquali, D. (2023). Investigating extreme sea level components and their interactions in the Adriatic and Tyrrhenian Seas. *Weather and Climate Extremes*, *41*, 100590. https://doi.org/10.1016/j.wace.2023.100590

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

Ramos-Valle, A. N., Curchitser, E. N., Bruyère, C. L., & McOwen, S. (2021). Implementation of an artificial neural network for storm surge forecasting. *Journal of Geophysical Research: Atmospheres*, *126*(13), e2020JD033266. https://doi.org/10.1029/2020JD033266

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., et al. (2020). Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble. *Journal of Climate*, *33*(7), 2557–2583. https://doi.org/10.1175/JCLI-D-19-0639.1

Rueda, A., Cagigal, L., Pearson, S., Antolínez, J. A., Storlazzi, C., Van Dongeren, A., et al. (2019). HyCReWW: A hybrid coral reef wave and water level metamodel. *Computers & Geosciences*, *127*, 85–90. https://doi.org/10.1016/j.cageo.2019.03.004

Sarhadi, A., Rousseau-Rizzi, R., & Emanuel, K. (2025). Physics-based hazard assessment of compound flooding from tropical and extratropical cyclones in a warming climate. *Earth's Future*, *13*(1), e2024EF005078. https://doi.org/10.1029/2024EF005078

Sarhadi, A., Rousseau-Rizzi, R., Mandli, K., Neal, J., Wiper, M. P., Feldmann, M., & Emanuel, K. (2024). Climate change contributions to increasing compound flooding risk in New York City. *Bulletin of the American Meteorological Society*, *105*(2), E337–E356. https://doi.org/10.1175/BAMS-D-23-0177.1

Scoccimarro, E., Bellucci, A., & Peano, D. (2017). CMCC CMCC-CM2-VHR4 model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.1367

Scoccimarro, E., Fogli, P. G., Reed, K. A., Gualdi, S., Masina, S., & Navarra, A. (2017). Tropical cyclone interaction with the ocean: The role of high-frequency (Subdaily) coupled processes. *Journal of Climate*, *30*(1), 145–162. https://doi.org/10.1175/jcli-d-16-0292.1

Shimura, T., Pringle, W. J., Mori, N., Miyashita, T., & Yoshida, K. (2022). Seamless projections of global storm surge and ocean waves under a warming climate. *Geophysical Research Letters*, *49*(6), e2021GL097427. https://doi.org/10.1029/2021GL097427

Strauss, B. H., Orton, P. M., Bittermann, K., Buchanan, M. K., Gilford, D. M., Kopp, R. E., et al. (2021). Economic damages from Hurricane Sandy attributable to sea level rise caused by anthropogenic climate change. *Nature Communications*, *12*(1), 2720. https://doi.org/10.1038/s41467-021-22838-1

Tadesse, M., Wahl, T., & Cid, A. (2020). Data-driven modeling of global storm surges. *Frontiers in Marine Science*, *7*, 260. https://doi.org/10.3389/fmars.2020.00260

Tausía, J., Delaux, S., Camus, P., Rueda, A., Méndez, F., Bryan, K., et al. (2023). Rapid response data-driven reconstructions for storm surge around New Zealand. *Applied Ocean Research*, *133*, 103496. https://doi.org/10.1016/j.apor.2023.103496

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., et al. (2021). Climate model projections from the scenario model intercomparison project (ScenarioMIP) of CMIP6. *Earth System Dynamics*, *12*(1), 253–293. https://doi.org/10.5194/esd-12-253-2021

Tiggeloven, T., Couasnon, A., Van Straaten, C., Muis, S., & Ward, P. J. (2021). Exploring deep learning capabilities for surge predictions in coastal areas. *Scientific Reports*, *11*(1), 17224. https://doi.org/10.1038/s41598-021-96674-0

Towey, K. L., Booth, J. F., Rodriguez Enriquez, A., & Wahl, T. (2022). Tropical cyclone storm surge probabilities for the east coast of the United States: A cyclone-based perspective. *Natural Hazards and Earth System Sciences*, *22*(4), 1287–1300. https://doi.org/10.5194/nhess-22-1287-2022

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., & Feyen, L. (2018). Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nature Communications*, *9*(1), 2360. https://doi.org/10.1038/s41467-018-04692-w

Wang, T., Liu, T., & Lu, Y. (2023). A hybrid multi-step storm surge forecasting model using multiple feature selection, deep learning neural network and transfer learning. *Soft Computing*, *27*(2), 935–952. https://doi.org/10.1007/s00500-022-07508-8

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Xi, D., Lin, N., & Gori, A. (2023). Increasing sequential tropical cyclone hazards along the US East and Gulf coasts. *Nature Climate Change*, *13*(3), 258–265. https://doi.org/10.1038/s41558-023-01595-7

Xie, W., Xu, G., Zhang, H., & Dong, C. (2023). Developing a deep learning-based storm surge forecasting model. *Ocean Modelling*, *182*, 102179. https://doi.org/10.1016/j.ocemod.2023.102179

Yang, C.-H., Wu, C.-H., & Hsieh, C.-M. (2020). Long short-term memory recurrent neural network for tidal level forecasting. *IEEE Access*, *8*, 159389–159401. https://doi.org/10.1109/ACCESS.2020.3017089

Zhang, X., Wang, T., Wang, W., Shen, P., Cai, Z., & Cai, H. (2023). A multi-site tide level prediction model based on graph convolutional recurrent networks. *Ocean Engineering*, *269*, 113579. https://doi.org/10.1016/j.oceaneng.2022.113579