Coastal Engineering 186 (2023) 104406

Contents lists available at ScienceDirect

Coastal Engineering

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

A framework for flexible peak storm surge prediction

Benjamin Pachev [a,*], Prateek Arora [b], Carlos del-Castillo-Negrete [a], Eirik Valseth [a,c,d], Clint Dawson [a]

[a] Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, 78712, TX, USA
[b] Department of Civil and Urban Engineering, New York University, Brooklyn, 11201, NY, USA
[c] The Department of Data Science, Norwegian University of Life Sciences, Elizabeth Stephansens vei 15, Ås, 1430, Norway
[d] Simula Research Laboratory, Kristian Augusts gate 23, Oslo, 0164, Norway

ARTICLE INFO

Keywords:
Storm surge
ADCIRC
Machine learning

ABSTRACT

Storm surge is a major natural hazard in coastal regions, responsible both for significant property damage and loss of life. Accurate, efficient models of storm surge are needed both to assess long-term risk and to guide emergency management decisions. While high-fidelity regional- and global-ocean circulation models such as the ADvanced CIRCulation (ADCIRC) model can accurately predict storm surge, they are very computationally expensive. Consequently, there have been a number of efforts in recent years to develop data-driven surrogate models for storm surge.

Here we develop a novel surrogate model for peak storm surge prediction based on a multi-stage approach. In the first stage, points are classified as inundated or not. In the second, the level of inundation is predicted for each point. Additionally, we propose a new formulation of the surrogate problem in which storm surge is predicted independently for each point. This new formulation has the potential to allow for predictions to be made directly for locations not present in the training data, and significantly reduces the number of required model parameters.

We demonstrate our modeling framework on two study areas: the Texas coast and the northern portion of the Alaskan coast. For Texas, the model is trained with a database of 446 synthetic hurricanes. The model is able to accurately match ADCIRC predictions on a test set of synthetic storms. We further present a test of the model on Hurricanes Ike (2008) and Harvey (2017). For both storms, we find that the model predictions have comparable accuracy to ADCIRC hindcasts when compared to actual observational data. For Alaska, the model is trained on a dataset of 109 historical surge events. We test the surrogate model on actual surge events including the recent Typhoon Merbok (2022) that take place after the events in the training data. As with the Texas dataset, the surrogate model achieves decent performance against observational data. In both cases, the surrogate models are many orders of magnitude faster than ADCIRC.

1. Introduction

In the last four decades, tropical cyclones have caused over one trillion dollars of damage in the United States alone (National Centers for Environmental Information (NCEI), 2022). Storm surge is directly responsible for much of the property damage from tropical cyclones (Neumann et al., 2015) and nearly half of the fatalities (Rappaport, 2014). Hurricane Katrina (2005) was the costliest hurricane on US record, with hundreds of billions of dollars in property damage and over 1200 deaths, most of which were caused by the extreme storm surge (Blake et al., 2011). Thus, predicting storm surges is crucial in order to assess the long term risk to coastal infrastructure and property from tropical cyclones. While high-fidelity physics-based models of storm surge such as the ADvanced CIRCulation (ADCIRC) model (Luettich et al., 1992; Pringle et al., 2021) have been developed, they require significant computational resources. This makes statistical risk studies infeasible or limited in scope. One workaround is to use a fast, low-fidelity physical model such as SLOSH (Sea, Lake, and Overland Surges from Hurricanes, (Jelesnianski, 1992)). However, low-fidelity models neglect key physics and consequently can have high errors.

An alternative to physics-based modeling is surrogate modeling, in which a parametric model is fit to data generated from observations or outputs of a high-fidelity physical model. An early example of this approach to modeling storm surge is the storm surge response function

* Corresponding author.
E-mail addresses: benjaminpachev@utexas.edu (B. Pachev), pa2178@nyu.edu (P. Arora), carlos.delcastillo@utexas.edu (C. del-Castillo-Negrete), eirik.valseth@nmbu.no, eirik@oden.utexas.edu (E. Valseth), clint@oden.utexas.edu (C. Dawson).

https://doi.org/10.1016/j.coastaleng.2023.104406
Received 31 March 2023; Received in revised form 30 August 2023; Accepted 1 October 2023
Available online 4 October 2023

by Irish et al. (2009), Resio et al. (2009) which models maximum storm surge as a function of landfall location, central pressure, and the hurricane pressure radius for a fixed forward speed and track angle. Recent work has focused on developing Machine Learning (ML) and Deep Learning (DL) based storm surge prediction models that can account for a wider variety of hurricane characteristics e.g., Lee (2006, 2008, 2009), Ian et al. (2022), Al Kajbaf and Bensi (2020), Chen et al. (2012), Kim et al. (2015), Hashemi et al. (2016), Bezuglov et al. (2016), Kim et al. (2018), Lee et al. (2021a). These models aim to mimic the accuracy of high fidelity physical storm surge models, but with a significant reduction in computational requirements. In Ian et al. (2022), Ian et al. compared the performance of several machine learning models to predict storm surge and observed that tree-based ensemble methods provided the best estimates of storm surge. Whereas in Al Kajbaf and Bensi (2020), Al Kajbaf and Bensi compared the performance of DL-based Artificial Neural Networks (ANN), ML-based Gaussian Process Regression (GPR), and Support Vector Machines (SVM) methods to predict storm surge levels. Finally, Al Kajbaf and Bensi predicted the storm surge with the least errors using ANN compared to GPR and SVM methods in Al Kajbaf and Bensi (2020). Lee et al. used ANN, including Back Propagation Neural Network (BPNN), to model the storm surge from typhoons in Taiwan in Lee (2006, 2008, 2009). These studies used wind velocity, wind direction and atmospheric pressure as inputs, in addition to a harmonic analysis of tidal levels. However, these data-driven storm surge models struggled to generalize to different tidal seasons due to limited training data. In Chen et al. (2012), Chen et al. compared BPNN to the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) algorithm to predict storm surge at a few tidal gauges on the east coast of Taiwan. Four input parameters were used: wind speed, wind direction, air pressure and simulated water level from ADCIRC. The ANFIS model outperformed BPNN in all cases.

More advanced ML techniques were developed by Kim et al. (2015) in which a larger set of six input parameters were used: latitude, longitude, the central pressure of the storm, moving speed of the storm, heading direction of the storm, and radius of exponential scale pressure. This study relied on a dataset of 446 synthetic storms developed using historical tropical cyclone activity in the Gulf of Mexico. In Hashemi et al. (2016), Hashemi et al. used 1050 synthetic storms simulated by the US Army Corp of Engineers for the North Atlantic Comprehensive Coastal Study to predict storm surge and compared the performance of DL based ANN, and ML based Support Vector Machine (SVM) algorithms. The developed prediction model was applied to weather stations within Rhode Island, USA, using four storm parameters: central pressure deficit, the radius of maximum winds, translation speed, and storm heading direction. The results reported in Hashemi et al. (2016) indicate that ANN performed better than SVM for storm surge predictions in the considered scenarios.

Overall, DL and ML based surrogate models have been capable of predicting storm surge with reasonable accuracy. However, the currently existing models, e.g. Chen et al. (2012), Kim et al. (2015), Hashemi et al. (2016), Kim et al. (2018), Lee et al. (2021a) ignore the complex bathymetry and topography of coastal regions which can introduce uncertainties in storm surge prediction, e.g., in Alaska, USA where storm–tidal interactions are significant (Zhang et al., 2021). To overcome these shortcomings, we developed a high-fidelity surrogate storm surge prediction model which is not limited to storm characteristics alone, and includes complex bathymetry and topography features of coastal regions. Improved ML driven predictions of storm surge can be used to determine the expected performance of existing and planned infrastructure as well as during preparations for incoming hazards (Arora and Ceferino, 2022).

In this paper, we present a novel flexible framework to develop regional surrogate models for peak storm surge prediction. The framework is applied to the Texas and northern Alaskan coastlines. The surrogate models rely on Gradient Boosting ML and ANN-based DL algorithms. A total of 135 and 172 input features are used for Texas and Alaska, respectively. The input features include bathymetric features, tidal harmonics, wind, pressure, as well as other storm characteristics. The trained models are validated on the historic events of Hurricane Harvey (2017), and Hurricane Ike (2008) in Texas and Typhoon Merbok (2022) in Alaska. The remainder of the paper is organized as follows: in Section 2 we discuss the study areas and creation of corresponding datasets. Section 3, we detail the construction of surrogate models. In Section 4 we present our numerical results for both study areas. Finally, we conclude with a discussion of our results and future work in Section 5.

2. Dataset generation

There are two potential sources of training data for a machine learning model of storm surge: observations, or output from a high resolution numerical model. Observational data is limited to only a few points in the spatial domain (e.g. NOAA or USGS tide gauges). Consequently, it is not feasible to train a model with full spatial output from observational data alone. This necessitates the use of a physics-based model to generate training data with a significantly larger spatial extent. However, it is worth noting that a data-driven model, while much computationally cheaper to evaluate, will inherit the same biases and errors present in the training data. Consequently, the model used for data generation should be as accurate as possible.

2.1. The ADCIRC model

Storm surge is a physical process that is a result of sea water flow induced by winds and is often further increased in magnitude by tides. In some cases, a distinction is made between storm tide and storm surge, with storm surge referring only to the excess water elevation above the natural tide. We do not make this distinction. By storm surge, we mean the total water elevation above the datum (e.g. what is sometimes called storm tide). The governing model we use for shallow water flow are the two dimensional shallow water equations (SWE) which consist of the depth averaged equations of mass conservation as well as x and y momentum conservation (Tan, 1992):

Find $(\zeta, \mathbf{u})$ such that:

$$\begin{aligned}
&\frac{\partial \zeta}{\partial t} + div(H\mathbf{u}) = 0, \text{ in } \Omega,\\
&\frac{\partial (Hu_x)}{\partial t} + div\left(Hu_x^2 + \tfrac{g}{2}(H^2 - h_b^2), Hu_xu_y\right) - g\zeta\frac{\partial h_b}{\partial x} = F_x, \text{ in } \Omega,\\
&\frac{\partial (Hu_y)}{\partial t} + div\left(Hu_xu_y, Hu_y^2 + \tfrac{g}{2}(H^2 - h_b^2)\right) - g\zeta\frac{\partial h_b}{\partial y} = F_y, \text{ in } \Omega,
\end{aligned} \tag{1}$$

where ζ is the free surface elevation above the geoid, h_b the bathymetry, $H = \zeta + h_b$ is the total water column, $\mathbf{u} = \{u_x, u_y\}^{\mathrm{T}}$ is the depth averaged velocity field, and the source terms F_x, F_y represent potential relevant sources which induce flow, including: Coriolis force, tidal potential forces, wind stresses, and wave radiation stresses. Ω is the computational domain, e.g., the coastal ocean. To solve this set of transient nonlinear partial differential equations for physically relevant shallow water flows in storm surge, a numerical method is required. In this study, we consider the well established numerical shallow water equation solver ADCIRC (Luettich et al., 1992; Pringle et al., 2021). ADCIRC was initially developed by Luettich, Westerink, and Scheffner to model shallow water flows in coastal regions, estuaries, and shelves.

ADCIRC uses a continuous Galerkin finite element method to spatially discretize the weak form corresponding to (1) and utilizes a finite difference technique to advance the solution in time. Over the last few decades, ADCIRC has been extensively developed and improved to include physics and other features that are critical to accurate modeling of shallow water flows during hurricane storm surge events including: waves, tides, bottom friction, levees and floodwalls, wetting and drying, and high resolution representation of bathymetry and topography.

ADCIRC distinguishes itself from many other coastal circulation models in the use of the finite element method for spatial discretization. This allows users to develop unstructured meshes of variable resolution throughout the computational domain. Hence, it is well suited to model phenomena such as hurricane storm surge which relies on both computational efficiency as well as variable resolution of a large domain, e.g., the Atlantic Ocean and the US East Coast. Furthermore, ADCIRC has been optimized for high-performance computing through MPI parallelization and is the backbone of several operational storm surge forecasting systems, see, e.g., Dresback et al. (2013). ADCIRC has been extensively validated for past hurricanes, including Ike (Hope et al., 2013), Gustav (Forbes et al., 2010) (both 2008), Katrina, Rita (both 2005) (Dietrich et al., 2012), Harvey (2017) Goff et al. (2019), as well as others. Due to its extensive validation and capabilities, ADCIRC represents an ideal candidate to generate training data for our surrogate models.

2.2. Texas

The first region we will consider is the coast of Texas. Fortunately, there is a preexisting publicly available dataset of ADCIRC output for 446 synthetic storms that make landfall on the Texas coast (Dawson et al., 2021). The synthetic storms were originally developed to assess flooding risk for computation of insurance premiums. The dataset is similar to the synthetic storm databases used in previous surrogate modeling studies (Lee et al., 2021b; Jia et al., 2016; Bass and Bedient, 2018). For each synthetic storm we have the best track data, as well as wind and pressure fields with a .05 degree spatial resolution and a 15 min temporal resolution. Time-series water elevation output is available for the entire spatial domain at a temporal resolution of two hours, in addition to the maximum attained surge.

The ADCIRC mesh used to generate the synthetic storms contains $3,352,598$ nodes and $6,675,517$ triangular elements. This mesh covers the entire North Atlantic ocean and the Gulf of Mexico to ensure necessary resolution of all far-field events physics that are critical to accurately resolve storm surge on the Texas coast. The mesh contains extraordinary high resolution on the Texas coast with element size on the order of 10 m, see Fig. 1, for a plot of the mesh in Galveston Bay.

To ensure physically relevant simulations of flow near the coast, in estuaries, and in floodplains, this mesh also contains a spatially variable classification of the sea bottom and land defined though a friction coefficient used in a Manning's n friction formula (Manning et al., 1890) to ascertain physically relevant friction forcing.

Each storm in the database is a tropical cyclone. Consequently, the database of storms includes best track data as well as meteorological forcing and water elevation output. Best track data consists of the location of the storm eye, the central pressure, and a number of other storm parameters. In this study, we will only make use of the eye location data.

2.3. Alaska

The second region considered in this study is Alaska. An often under-studied region, modeling storm surge adequately along the Alaskan coastline has become increasingly important as it has experienced increasingly frequent and intense Extra-Tropical Cylcones (ETCs) (Graham and Diaz, 2001; Terenzi et al., 2014). These events, such as the recent Typhon Merbok (2022), produced surges greater than $3m$ threatening the critical and unique natural, social, and economic system in Alaska (Henry and Heaps, 1976; Reimnitz and Maurer, 1979; Kowalik, 1984; Johnson and Kowalik, 1986; Kinsman and DeRaps, 2012; Wicks and Atkinson, 2017). While many prior studies have constructed surrogate surge models for the Texas coast, to our knowledge this work is the first to do so for Alaska.

Many challenges exist in modeling storm surge along the Alaskan coast. First, as it is an under-studied region, there is little available data in the sense that there are fewer gauges measuring water elevation values and few, if any, comprehensive data-sets on flooding events in the region (Buzard et al., 2021). Furthermore, since ETCs develop sometimes from not just closed low pressure systems that can be easily parameterized, best track data is generally not available for these storms and a good tracking algorithm that captures the full surge event is generally not available (Mesquita et al., 2009). Lastly, there is the additional challenge of modeling the effects of sea ice coverage in the region, as this has great variable and non-trivial impact on the momentum transfer from air to sea as well as propagation of surges (Kowalik, 1984; Johnson and Kowalik, 1986; Blier et al., 1997). This effect is often critical for accurate simulations and is just now being incorporated into models via parameterizations of the sea ice interaction based on the sea ice coverage (Joyce et al., 2019).

To build a training set for storm surge along the Alaska coastline we first had to identify storm surge events. This is done by applying a modified version of the identification algorithm used in Wicks and Atkinson (2017) to identify storm surge events across three stations that see the most frequent storm surge events: Nome, Red Dog Dock, and Unalakleet. Fig. 2 shows an example of an identified peak surge event associated with Typhoon Merbok at Nome Alaska. See Appendix B for more details on the storm surge identification algorithm.

Having identified date-ranges of relevant storm surge events, we compiled meteorological forcing data from the CFSv1 (Saha et al., 2006) and CFSv2 (Saha et al., 2014) data-sets, consisting of wind, air pressure, and sea ice concentration. The data is gridded and has a spatial resolution of 0.3122 degrees. A reduced resolution version of the ADCIRC mesh used in Joyce et al. (2019) consisting of 443,770 nodes and 850,855 triangular elements was used and ADCIRC was run to compute the true water levels across the whole domain for each storm surge event and compile the final raw training dataset. We note once again, the base training data-set consist of just the meteorological forcing data (wind, pressure, and sea ice coverage) and ADCIRC output water levels, with no best track data (unlike Texas). The base data-set of storm surge events and ADCIRC simulations has been published to the DesignSafe (Rathje et al., 2017) Data repository (Del-Castillo-Negrete et al., 2023).

3. Model construction

Let Ω denote the spatial computational domain with coordinates $\mathbf{x}$, and let t denote time. A storm s is defined by vector valued meteorological forcing data $\mathbf{f}_s(\mathbf{x},t)$ (e.g., wind and pressure). The corresponding storm surge is given by the scalar valued function $\eta_s(\mathbf{x},t)$ and is presumed to functionally depend on the forcing $\mathbf{f}_s$. The time dependent storm surge prediction problem computes the mapping from $\mathbf{f_s}$ to η_s. While ADCIRC and other high-fidelity numerical models solve the time dependent storm surge prediction problem, here we solve the simplified problem of predicting maximum storm surge, defined as $\eta_s^{max}(\mathbf{x}) = \max_t \eta_s(\mathbf{x},t)$. In many applications, such as coastal risk assessment, knowledge of η^{max} is sufficient. Thus, we will drop the superscript and refer to the maximum storm surge as $\eta(\mathbf{x})$. Hence, in the following surrogate model construction, we do not aim to predict the time series of water surface elevation, only the maximum surge at $\mathbf{x}$

3.1. Point-based formulation

Constructing a surrogate model requires a set of sample inputs and outputs from a high-fidelity storm surge model, e.g $(\mathbf{f}_s(\mathbf{x},t),\eta_s(\mathbf{x}))_{s=1}^n$. Because of computational constraints, n is typically small, a few hundred or thousand. However, replicating the high resolution outputs of storm surge model requires making predictions at a large number of points N. Directly predicting a high dimensional output vector with a model such as a neural network would require a large number of model parameters relative to the number of training examples. Such a model could be impossible to train due to data limitations (i.e the number of

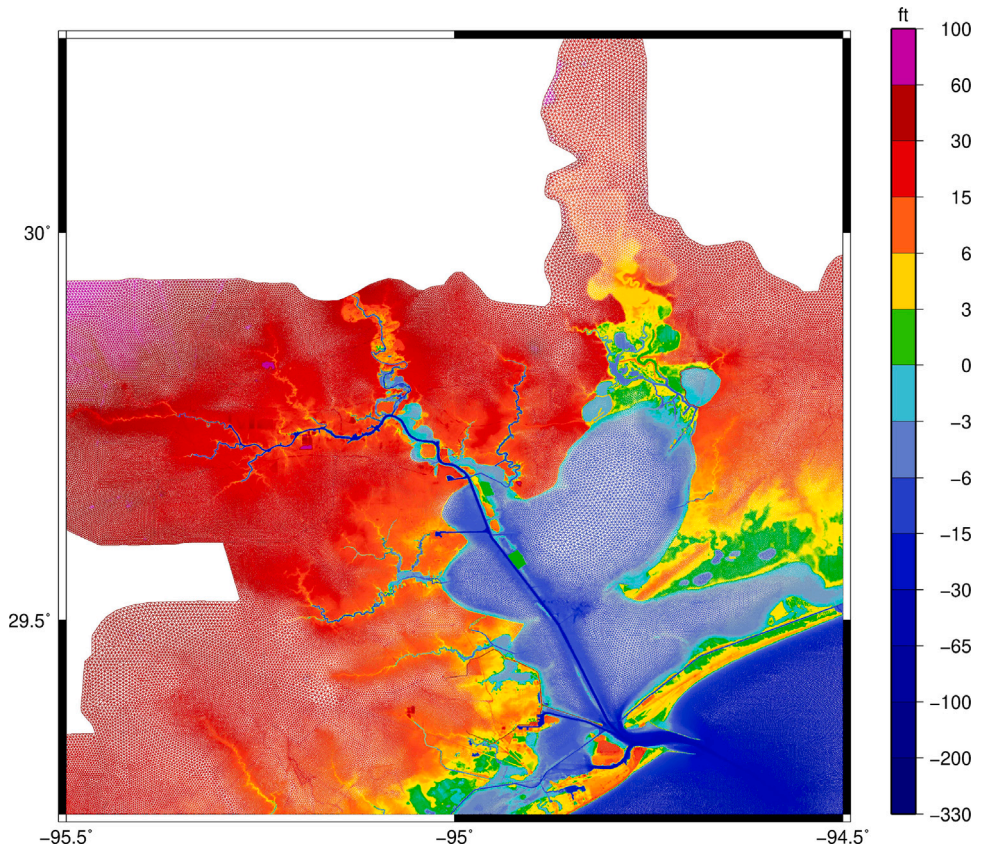


Fig. 1. ADCIRC mesh from Hope et al. (2013) in the Galveston Bay area.

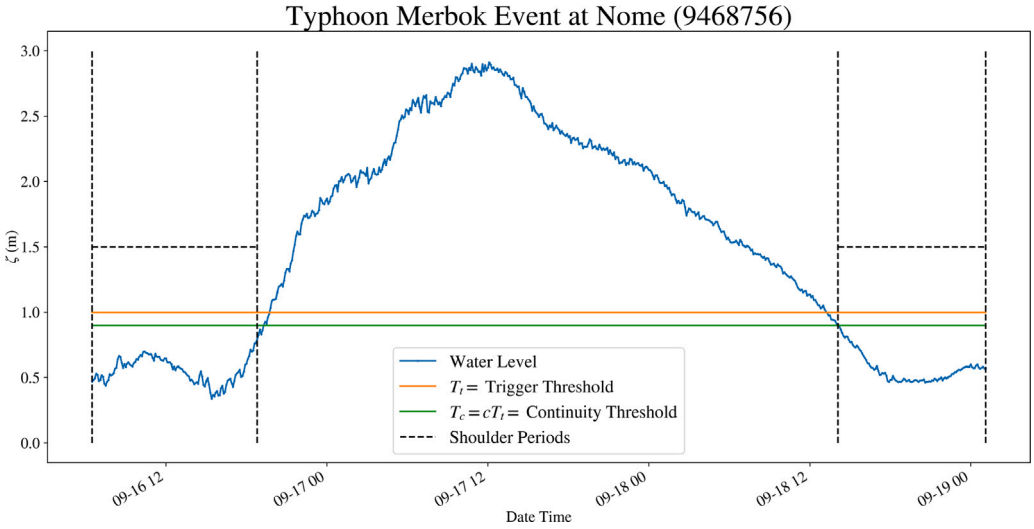


Fig. 2. Typhoon Merbok event as identified by Algorithm 1 at Nome, Alaska.

surge events in the training dataset would be much smaller than the output dimension). Consequently, a different approach is needed.

One solution is to predict the storm surge only at a small number of points, as is often done in the literature (Lee, 2006; Kim et al., 2015). However, this approach is problematic if high resolution output is desired, as interpolation from the sparse predictions can result in significant error. Another approach is to use dimensional reduction techniques such as principal component analysis (Jia and Taflanidis, 2013; Jia et al., 2016; Kyprioti et al., 2021). However, this explicitly constraints the model to a low dimensional manifold generated by the training data. In the case of principal component analysis, the manifold is linear. This approach could potentially struggle to capture the full range of physical outcomes, as there is no guarantee that the space of potential maximum surge profiles can be effectively projected into a low-dimensional linear space.

To circumvent the need for dimensional reduction while avoiding a large number of model parameters, we propose a novel *point-based* formulation of the surrogate problem. Instead of considering one storm s as one training example, we consider each combination of storm and output location $(s, \mathbf{x_i})$ as a training example. While this significantly increases the number of predictions that are needed, it reduces the model output dimension to one. This reduces the minimum number

of model parameters, while simultaneously increasing the number of training examples. Both of these changes make the estimation problem more computationally feasible. Another added benefit of this formulation is that the model has the possibility to directly make predictions for locations it has not seen in the training data, including points potentially outside of the study domain. Furthermore, the point-based formulation gives us the flexibility to make predictions only in a small area of interest within a much larger study domain.

3.2. Feature engineering

The preceding abstract formulation applies to both study areas. The next steps are to construct a good feature representation and select an appropriate machine-learning model. These steps will differ somewhat for each study area, as there are differences in the available data. However, the basic approach is the same.

Since the modeling problem is posed for a single point in the domain, the feature representation needs to encode only the information relevant to predicting storm surge for the given location. Although the surge at any given point can depend on the entirety of the bathymetry, wind, and pressure fields, in practice, not all of the data will be relevant. Our goal is to develop a compact feature representation that is still robust enough to enable accurate predictions. In addition, we want to discourage overfitting the data. For instance, we exclude the latitude and longitude from the training data. While storm surge does vary significantly by location, this variation is driven by bathymetry and coastline shape, not the actual values of the latitude and longitude. The input forcing data vary in time, however, the desired output of maximum surge does not. We remove the temporal component of the data by computing the mean, max, and min for the wind vector, wind magnitude, and pressure in time. The Alaska dataset additionally contains sea ice concentration data — which we process in the same manner as the wind and pressure. For the synthetic tropical cyclones in the Texas dataset, these statistics are computed over a fixed window from six hours before landfall to six hours after landfall. Selecting, shorter averaging time frames lead to a deterioration of the results, whereas longer times does not change the results in a measurable way. For the surge events in the Alaska dataset, these statistics are computed over the entire available time-series. We remark that before computing any statistics, the forcing data are first interpolated to the nodes of the ADCIRC mesh. Consequently, the processing of the forcing data is performed in a standardized manner for all surge events in both study areas.

The maximum surge at a given point can potentially depend on the forcing data and bathymetry over the entire spatial domain. However, we expect spatial proximity to play a role. Consequently, more weight should be given to local over global information. Nevertheless, it is important to at least include some global information to account for large-scale effects. To test these hypotheses, we compute spatial statistics for each of the temporally aggregated forcing variables over a variety of length scales. For convenience, the local neighborhoods used to compute these spatial statistics are all regular lat-lon boxes centered at the prediction point in question. The box sizes are 0.1, 0.2, and 0.4 degrees. Global statistics are also computed over the entire spatial domain for each storm. A similar procedure is applied to the bathymetry. Slightly different box sizes are used - 0.05, 0.1, 0.4, and 1.0 degrees (note that the spatial statistics are in addition to the pointwise bathymetry).

For storms that make landfall, we compute the distance to the landfall location. We also compute the distance to the coastline. Up to this point, the Texas and Alaska datasets can be handled with an essentially identical approach. The main difference arises when we consider the effect of tidal forcing. Unlike the Texas dataset, the Alaska dataset consists of real surge events, most of which are not hurricane strength intensity. Furthermore, Alaska experiences much stronger tidal variations than the Gulf Coast. The combination of weaker winds and stronger tides means that tidal forcing will significantly affect storm surges. To account for the effect of tides, we consider two approaches. The first approach is to account for tides in the training data. We do this by using ADCIRC to determine amplitudes and phases at each node in the Alaska mesh for eight major tidal constituents: M2, S2, N2, K2, O1, K1, P1, and Q1. The amplitudes of the constituents are then added as features. The second approach is to remove the tidal contribution from the maximum surge, and attempt to predict only the meteorological contribution. To determine the tidal contribution to maximum surge, we reran each event in the Alaska study area with meteorological forcing turned off. The maximum surge from the tidal run was then subtracted from the maximum surge from the run with both tides and meteorology to obtain the meteorological component of maximum surge. As the Texas dataset consists of synthetic events, there are no contributions from tides included in these results. While the tidal range in Texas is small compared to Alaska, several watersheds in Texas are sensitive to the effects of tides during storm surges, e.g., in the low-lying areas in vicinity of the City of Houston. However, the impact of tides, as noted in e.g., Irish et al. (2009), inclusion of tides can with reasonable accuracy be done through superposition.

In total, we generate three datasets that can be used to train surrogate models. These are summarized in Table 1. These three datasets are available on DesignSafe (see Appendix A) (Pachev et al., 2023-08-30).

3.3. Model architecture

Previously, researchers have developed neural network-based regression models to predict storm surges. However, one challenge for modeling storm surge is that for any given storm, a large number of points in the spatial domain may experience zero surges. Predicting which locations will remain dry is a classification problem, not a regression. Thus, we model each problem separately in two stages instead of just developing a simple regression model. On the other hand, predicting the level of surge for inundated areas is a regression problem. Consequently, we construct two models, one for classifying points as wet or dry, and one for predicting the inundation level at wet points.

We consider various model architectures for the classification and regression stages. Following previous work (Irish et al., 2009; Resio et al., 2009), we consider storm surge to be a process depending nonlinearly on winds and other forcings. Consequently, both the classification of points as inundated or not and the prediction of maximum inundation levels require a model with the ability to approximate nonlinear functions. Artificial neural networks (ANN) can capture highly nonlinear interactions and have been applied successfully across various disciplines, making them a popular choice for surrogate models (Kubat, 1999). In this work, we also consider gradient tree boosting, a decision tree-based, non-neural, machine learning method that has been widely applied to nonlinear classification and regression problems (Chen and Guestrin, 2016a; Rumelhart et al., 1986). Finally, we compare the performance of different architecture combinations with ANN and gradient tree boosting-based models.

ANN are inspired by biological neuron systems with a series of interconnected neurons Chen et al. (2012). In this work, we consider three different neural network architectures listed in Table 2. The largest neural network, ‘nn3’, is depicted in Fig. 3. Each network consists of a sequence of layers. The first layer is referred to as the *input layer*, while the last layer is called the *output layer*. Each layer of the network maps a vector of inputs to a vector of outputs according to the formula:

$$n = f(W \times I + b), \tag{2}$$

where n is the vector of outputs, I is the vector of inputs, W is the weight matrix, and b is the bias vector. The weight matrices and bias vectors are layer-dependent parameters that the network fits to the training data. The function f is known as the *activation function* and

Table 1
Training datasets for surrogate models.

Study area	Number of features	Prediction variable
Texas	135	Meteorological Component of Maximum Surge
Alaska	172	Total Maximum Surge
Alaska	164	Meteorological Component of Maximum Surge

Table 2
Architecture of neural networks.

Neural network	Number of hidden layers	Size of hidden layers
nn1	3	(256, 512, 256)
nn2	5	(256, 512, 1024, 512, 256)
nn3	7	(256, 512, 1024, 2048, 1024, 512, 256)

may vary between layers. It is generally nonlinear in order to enable the model to approximate nonlinear functions. Our networks use the rectified linear unit (ReLU) (Agarap, 2019) activation for all layers except the output layer. The ReLU activation is given by

$$f(x) = x,\ if\ x > 0\ else\ 0. \tag{3}$$

The choice of activation function for the output layer depends on the problem at hand. When predicting the maximum inundation level for inundated points, we use the ReLU activation function because it constrains the output to be nonnegative. However, when classifying a point as inundated or not, the output needs to be a probability between zero and one. In this case we use the sigmoid activation (Narayan, 1997), which is given by

$$g(x) = \frac{1}{1+e^{-x}}. \tag{4}$$

The weights and biases for each layer are determined by minimizing an appropriate loss function of the training data. Selection of the loss function depends on the problem. For the regression network, we minimize the mean squared error, given by

$$E = \frac{1}{N}\sum(\eta_i - F(x_i))^2. \tag{5}$$

Here E is the loss or error, N the total number of observations in the training data, η_i is the true maximum storm surge, and $F(x_i)$ is the predicted storm surge. For classification, we minimize the binary cross entropy loss, given by

$$L = -\frac{1}{N}\sum\left[y_i log(G(x_i)) + (1-y_i) log(1-G(x_i))\right]. \tag{6}$$

Here y_i is equal to one for inundated nodes, and is otherwise zero. The input data are indicated by x_i, and $G(x_i)$ is the inundation probability output by the classification network.

In addition to neural networks, we also considered gradient tree boosting. The gradient tree boosting method combines the concept of gradient descent and decision trees. The method aims to minimize the same loss functions in Eqs. (5) and (6). However, instead of fitting a predetermined set of parameters, the method iteratively constructs a sequence of decision trees. Each tree is trained to correct the errors of the previous trees. Predictions are generated by summing the outputs of each tree. For gradient boosting, we use the popular XGBoost library (Chen and Guestrin, 2016b). XGBoost can be used for both classification and regression problems, and is extremely scalable. We refer the interested reader to Breiman (2001) and Friedman (2002) for a more in-depth discussion of decision trees and gradient boosting.

3.4. Model training

For each dataset, we used 90% of the surge events for training, and the remainder for testing. The same training-test split was used across all experiments with a given dataset unless otherwise indicated. For the Texas data, the selection of surge events for training was random. On the other hand, the Alaska data consists of real historical surge events and has a natural chronological ordering. We consequently use a temporal split for the Alaska events, wherein the first 90% are used for training, and the most recent 10% for testing. This tests the ability of the models to predict future events from past events. It also ensures the test data include the recent Typhoon Merbok, an important test case for validation of the model.

3.4.1. Parameter settings

All neural network models were trained using a single NVIDIA Ampere A100 GPU on the Lonestar6 system at the Texas Advanced Computing Center. The XGBoost models were trained in parallel on 32 CPUs on Lonestar6. We used the Adam optimizer (Reddi et al., 2019) from the Keras library in Python to train all neural networks. The batch size was set to 2048. The number of epochs was set to 50 for the classification networks and 100 for the regression networks. This was done to balance training times between the two stages, as more data points are used in classification than regression (regression only uses inundated locations). The learning rate was set to 10^{-4} and adaptively decreased with a minimum of 10^{-6}. The number of boosting rounds used for the XGBoost model was 250.

3.4.2. Selection of prediction points

To focus model predictions on areas most affected by storm surge, we restrict predictions to points that are both close to the coastline and have a low enough elevation to potentially experience significant inundation. We first filtered points to be within ten kilometers of the coastline. Secondly, we restricted the maximum elevation above the geoid to be four meters. Predictions are also made for some points in the ocean — we filtered these to be only for points with a bathymetry of less than two meters. While somewhat arbitrary, this threshold based approach is easy to implement, and achieves the desired effect of constraining predictions to coastal regions with low elevation. We remark that a few surge events in the Texas dataset do exceed four meters of storm surge, which means that this filtering process may exclude some points that would experience inundation. However, another purpose of the filter is to limit the dataset size to facilitate model training.

Dataset size is particularly important for Texas. Even after the above restrictions, the Texas dataset is too large to fit in GPU memory, rendering training impractical. This is because the ADCIRC mesh used for the Texas study consists of over 3 million nodes and has a very high resolution near the coast — approximately half of the nodes fall within the coastal regions described above. Consequently, further reductions are applied. Firstly, we make predictions only at every tenth node (determined by the ordering of the nodes in the ADCIRC mesh). Secondly, predictions are only made for points that are within a radius of 150 km of the landfall location. This was done to focus the model predictions on regions that experience significant surge and mitigate bias towards underprediction of the most extreme surges. Even after all these reductions, the size of the Texas dataset is 20 GB.

By contrast, both Alaska datasets (with and without tides) are about one gigabyte with no extra filtering. This is due to the smaller size of the Alaska mesh (around 400 K nodes), its lower resolution near the coastline, and the number of surge events (109 for Alaska compared to 447 for Texas).

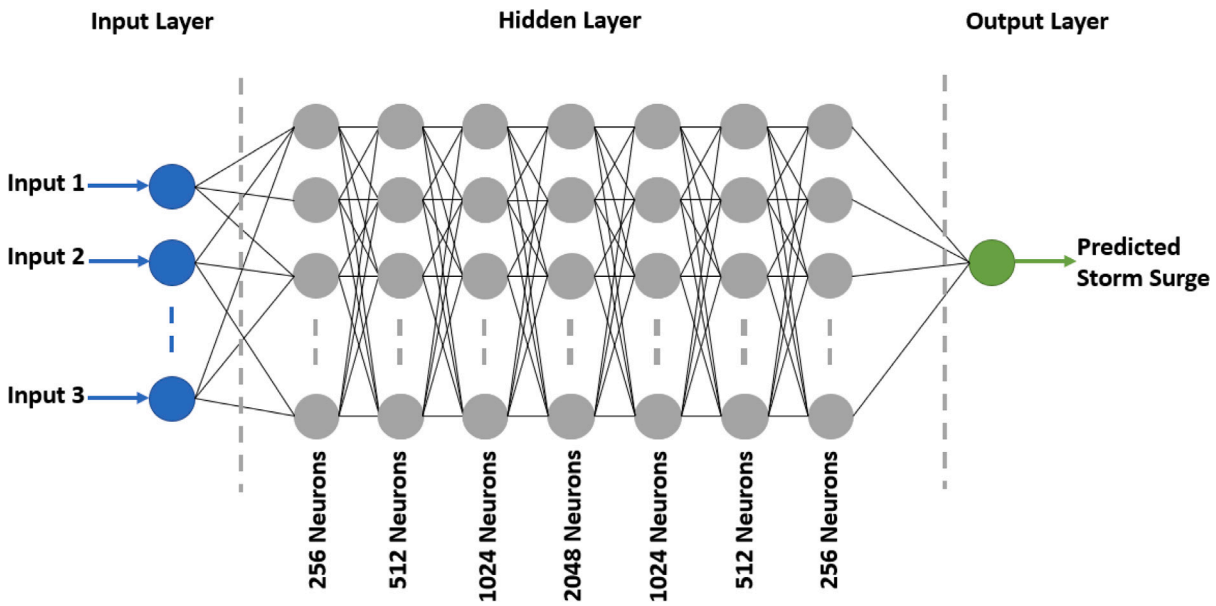


Fig. 3. Feed forward neural network architecture for storm surge prediction. For regression ANN, all linear layers are followed by the ReLU activation function. For classification ANN, all linear layers except the last one are followed by the ReLU activation function, and the sigmoid activation function follows the last layer.

4. Results

4.1. Model selection

Within the two stage framework, we can choose potentially different model architectures for the classification and regression stages. For each stage, we consider three neural network architectures, as well as XGBoost. This results in sixteen candidate two-stage models. For each dataset, we performed a grid search over all candidate model configurations. The best-performing models for each dataset are summarized in Table 3. Results for all sixteen model configurations are listed in Tables C.5, C.6, and C.7 in the appendix. All model configurations for the Texas area took under one hour to train. The Alaska models required about two minutes. Model performance was evaluated based on the overall RMSE, R^2 score, and the RMSE taken over the most extreme 5% of surges. These error metrics are computed across all points, including those whose inundation is incorrectly classified by the model. Consequently, they reflect the performance of both stages of the model.

For the Texas dataset, the best-performing model configuration uses the largest neural network model, 'nn3', for both classification and regression. In general, performance increases with the size of the network, and the neural networks do better than XGBoost. However, slightly better performance for extreme surges is obtained by using a smaller neural network as the classifier. We see different results for the Alaska datasets. The best models use XGBoost for regression and a neural network for classification. Furthermore, the neural network models have much worse performance compared to XGBoost when used for regression. This is likely because the Alaska data is a lot smaller than the Texas data, and is evidently insufficient to train the neural network models.

The Texas model does well at classifying inundated points, with error rates of less than 4% for both inundated and dry points. However, the Alaska data is heavily imbalanced towards inundation, which explains the high rates of misclassification of dry points as wet. This imbalance is driven by the Alaska mesh, which does not extend as far inland as the Texas mesh, resulting in a small number of dry points.

4.1.1. Removing correlated features

Many of the input variables, or features, are highly correlated. Consequently, it may be possible to eliminate some of them without significantly affecting performance. We use a threshold-based approach to reduce the number of features. First, a correlation threshold value τ is chosen between zero and one. Secondly, all pairs of input variables are ordered by the magnitude of their correlation coefficient, in descending order. Input variables are then iteratively eliminated until there are no remaining pairs with a correlation coefficient greater than τ in absolute value. Setting τ equal to one corresponds to no reduction in the number of features. Setting τ equal to zero would require all input variables to be completely uncorrelated. Fig. 4, illustrates the effect of the choice of the correlation threshold τ on the number of features and model performance.

Overall, increasing the number of features improves performance, but the gains are almost negligible when $\tau \geq 0.9$. Consequently, the choice of $\tau = 0.9$ represents a good trade-off between the number of input variables and model performance. Unless otherwise indicated, in the remaining experiments, we present results for models trained on reduced sets of features obtained by setting $\tau = 0.9$.

Finally, after setting $\tau = 0.9$, we selected 58 features (out of 135) and 79 features (out of 172) for the Texas and Alaska models, respectively, with 35 common features for both regions. The correlation among features obtained by computing spatial statistics may vary according to the regions' local characteristics (e.g., bathymetry and coastline shape). Thus, different features were retained for the Texas and Alaska models. However, correlations between pointwise variables (e.g., maximum wind at the prediction location) depend more on the overall storm characteristics. Thus, out of the 14 initially considered common pointwise features, 12 were retained for the Texas model, and all 14 were retained for the Alaska model. Some of the pointwise retained features were coastal distance, pointwise bathymetry, maximum wind speed, maximum wind in x and y directions, and maximum pressure for both Texas and Alaska regions. Additionally, landfall distance was also retained for Texas (it is not present in the Alaska data). Readers can refer to supplementary Table S1 for a complete list of features and a summary of retained features in supplementary Table S2 for the Texas and Alaska models.

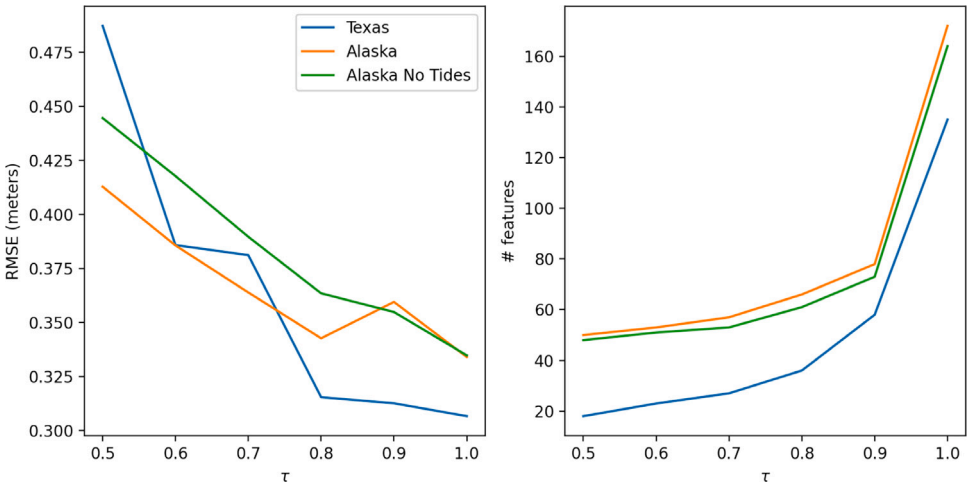


Fig. 4. Effect of correlation filter τ on model performance and the number of features.

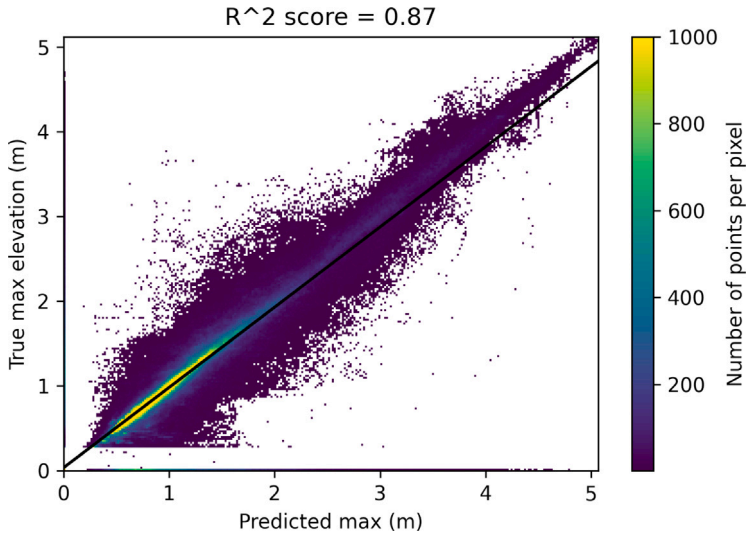


Fig. 5. Texas model fit.

Table 3
Best-performing models for each dataset. Top 5% RMSE is the RMSE taken over the most extreme 5% of surges. DryAsWet indicates the percentage of dry points classified as wet, while WetAsDry indicates the percentage of wet points classified as dry. Note that Alaska* denotes Alaska results without tides.

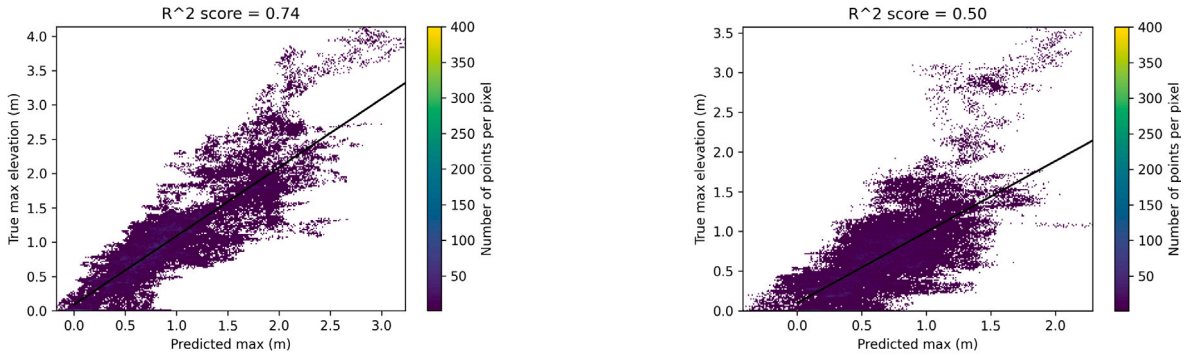
Dataset	Classifier	Regressor	R^2	RMSE (m)	Top 5% RMSE (m)	DryAsWet	WetAsDry
Texas	nn3	nn3	0.87	0.31	0.64	3.91%	3.20%
Alaska	nn1	xgb250	0.76	0.33	0.74	83.94%	0.00%
Alaska*	nn3	xgb250	0.54	0.33	0.84	78.20%	0.11%

4.2. Performance on test data

The Texas model achieves a good fit on the test synthetic storms, with an R^2 score of 0.87, as shown in Fig. 5. The error is greatest for the most extreme surges. This is to be expected because there are far fewer points with extreme surge, making it difficult to predict.

The Alaska model does not do as well as Texas, but still performs reasonably. However, excluding tides significantly impacts model performance, see Fig. 6. In Fig. 6, we observe forking of the results at the higher surge magnitudes. This is due to Typhoon Merbok, which had significantly higher surge than any other event in Alaska. We remark that while all three models (Texas, Alaska, Alaska without tides) have comparable RMSE, the R^2 score varies significantly between the three models. The comparatively worse model fits for the Alaska region are definitely affected by the much smaller dataset size, as well as the strong influence of tides relative to winds.

To get an idea of the spatial variation in model performance, we compute the root mean squared error at each prediction point across all storms. We show results for the Texas and Alaska models in Figs. 7 and 8. As illustrated in Fig. 7, the Texas model has consistent performance throughout the spatial domain. The model performs a little better on the western portion of the Texas coast, and a little worse near Galveston bay. One potential explanation for this difference is the change in coastline orientation from S-N to SW-NE. The model heavily relies on the



(a) Alaska model fit. (b) Alaska model fit without tides.

Fig. 6. Effect of including tides on Alaska model fit.

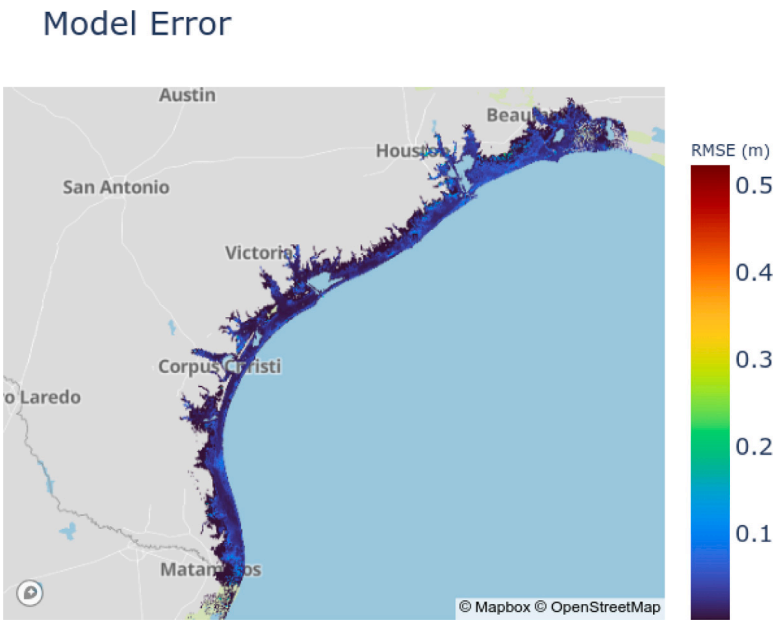


Fig. 7. Spatial variation in Texas model errors. Map data courtesy of OpenStreetMap (OpenStreetMap contributors, 2022).

x and y components of the wind vector for its feature representation. Consequently, it is not rotationally invariant, and likely is sensitive to coastline orientation. However, the inclusion of additional features such as bathymetric statistics and the wind magnitude likely allow the model to partially compensate for this sensitivity. This is borne out by the relative spatial consistency of the results. A second reason for slightly worse performance near Galveston bay is that the upper Texas coast is more prone to higher surge events due to its wider shelf. Higher surge correlates with higher model error.

Examining Fig. 8, we find more significant variations in model performance. The highest errors are found in the bay near the Nome and Unalakleet NOAA stations. We expect this for two reasons. First, as noted by Joyce et al. (2019), the mesh in these areas has not been fully resolved, and accurate resolution is critical to capturing the localized gradients in the tidal amplitudes due to dissipation as the tide propagates in these region. Furthermore, given that observations of high water levels from these stations were used to select the events in the Alaska dataset, it stands to reason that this area would experience the most extreme surges. Consequently, it is no surprise that it has the highest errors as well.

In addition to spatial variation, we consider how the model performance varies by storm. In Fig. 9 we plot the RMSE for each storm in the test datasets against the mean surge across all prediction points. The Texas data shows a clear positive correlation between RMSE and the storm intensity. The Alaska data also indicates higher RMSE for more intense storms. There is a single storm of low intensity and high RMSE — this is likely an outlier, but may indicate that the model has difficulty with low intensity storms.

4.2.1. *Model robustness*

Ideally, the model performance will be insensitive to the selection of training and test datasets. To verify this, for each model, we performed a ten-fold cross validation experiment. First, the data was subdivided into ten parts. For each part of the data, a model was trained using the remaining 90% of the data, and predictions were made on the selected 10%. For the Texas model, we performed an additional cross validation experiment in which the storms were partitioned spatially by the latitude of the landfall location instead of randomly. This forced the model to make predictions for storms whose landfall locations fall outside of the spatial region covered by the test data. The results are summarized in Table 4. We find that overall, the models are stable with respect to the choice of training data. When predicting outside the domain covered by the training dataset, the Texas model performs worse, but not by much.

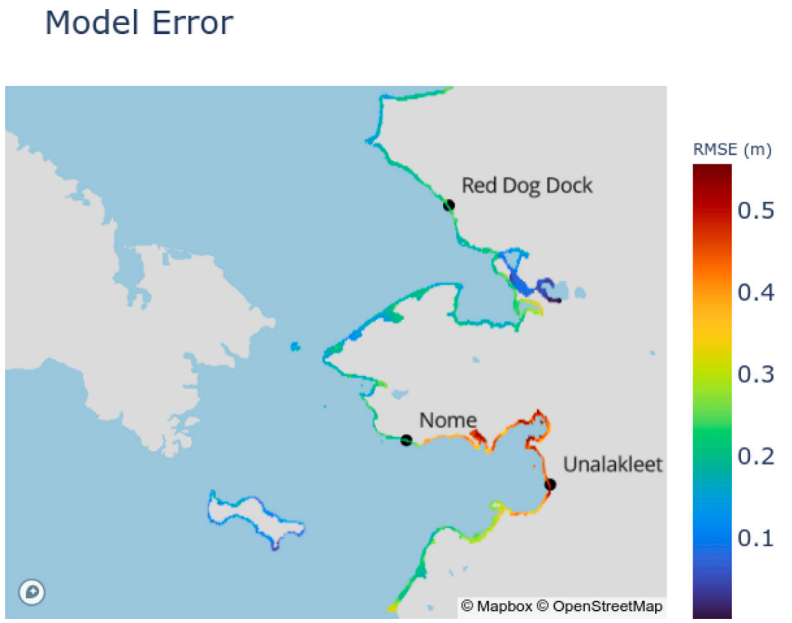
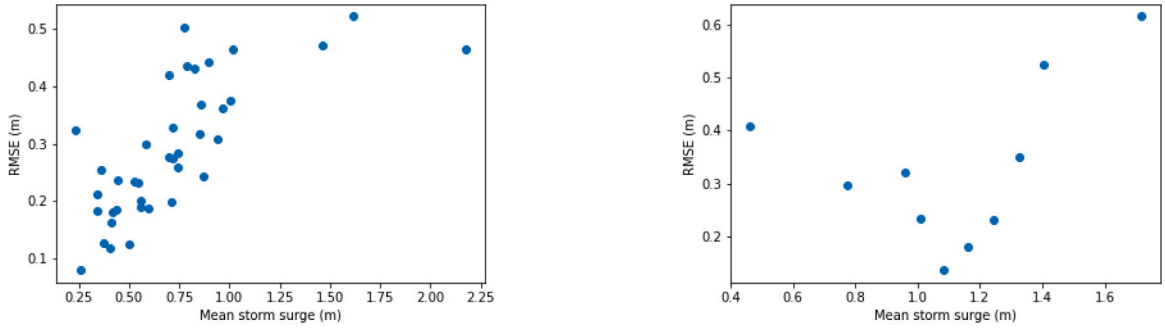


Fig. 8. Spatial variation in Alaska model errors. Map data courtesy of OpenStreetMap (OpenStreetMap contributors, 2022).



(a) Texas model performance by storm.

(b) Alaska model performance by storm.

Fig. 9. Model performance in Texas and Alaska.

Table 4
Statistics of model performance across a ten-fold cross-validation experiment.

Dataset	Mean RMSE (m)	Standard deviation (m)	Min RMSE (m)	Max RMSE (m)
Alaska	0.34	0.01	0.33	0.35
Alaska Without Tides	0.35	0.01	0.34	0.37
Texas	0.29	0.00	0.29	0.29
Texas (Spatial Split)	0.34	0.01	0.33	0.36

4.3. Validation against observational data

The purpose of a surrogate model is to act as a computationally inexpensive substitute for the underlying high-fidelity model — in this case ADCIRC. As demonstrated earlier, our models can replicate ADCIRC predictions to a certain degree of error. However, ADCIRC itself has biases and errors — which may be passed on to the surrogate model. By comparing the surrogate model predictions to actual observational data, we can get an idea as to what degree the model approximation error compounds with the ADCIRC model biases.

We tested the Texas model for Hurricanes Ike (2008) and Harvey (2017). These hurricanes were chosen based on their intensity, relevance to the study region, and availability of observational data. While Harvey generated extreme flooding in the Houston region due to a combination of rainfall and storm surge, we focus on the first landfall of Hurricane Harvey near Corpus Christi, which was primarily a storm surge event.

Although our ultimate goal is to validate model predictions against observational data, we first compare model predictions to ADCIRC. For each storm, we ran an ADCIRC simulation on the same mesh used to generate the synthetic data. Each simulation used 740 CPUs on Lonestar6 and was completed in around two hours. As when constructing the training data, we determined the landfall location for each hurricane and predicted the maximum surge for coastal points within a 150 km neighborhood of the landfall. However, we do not apply any further downsampling. Consequently, the model makes predictions for many locations not present in the training dataset. Because the initial selection of training locations was evenly distributed across the training dataset, these predictions fall within the regime of interpolation. One advantage of evaluating the model at these locations instead of using linear or nearest-neighbor interpolation is that the model can account for the local bathymetry, which would be otherwise ignored.

We find that the model accuracy for Ike and Harvey is a little lower than the test dataset but still high. For Ike, the model replicates ADCIRC

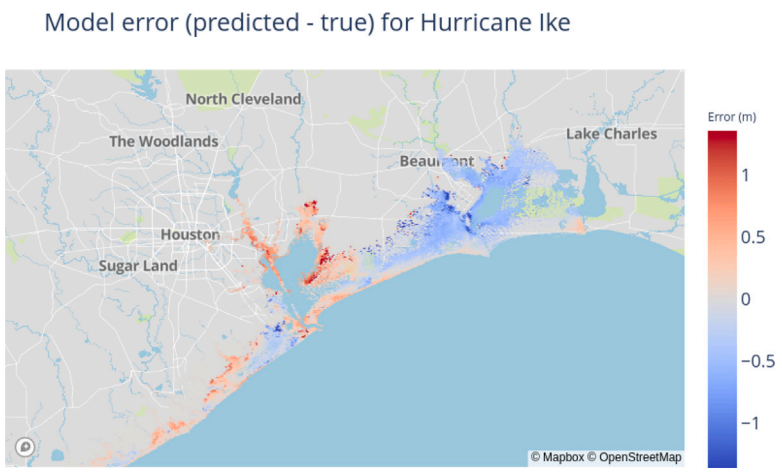


Fig. 10. Model error for Ike. Map data courtesy of OpenStreetMap (OpenStreetMap contributors, 2022).

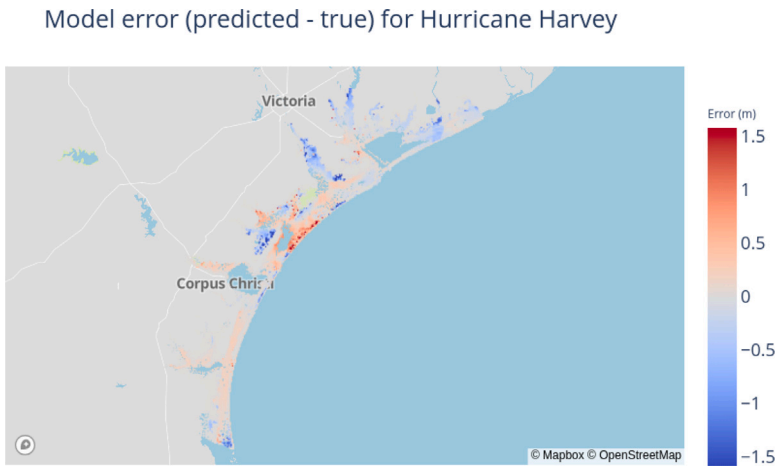


Fig. 11. Model error for Harvey. Map data courtesy of OpenStreetMap (OpenStreetMap contributors, 2022).

predictions with an R^2 score of 0.85 and an RMSE of 0.57 m. For Harvey, the model attains an R^2 score of 0.68 and an RMSE of 0.47 m. Plots of the errors for each storm are shown in Figs. 10 and 11. In these figures, the largest positive errors are clustered in areas where the hurricane surges were the largest, in eastern Galveston Bay and northeast of Corpus Christi for Hurricanes Ike and Harvey, respectively. Examining Fig. 9, the RMSE increases as mean storm surge increases. Hence, this distribution of errors follows a similar trend.

Next, we compare both the model predictions and ADCIRC predictions to observational data. For Ike we use 66 maximum elevation observations from a mixture of short-term USGS monitoring devices (East et al., 2008) and NOAA gauges. For Harvey we use 22 maximum elevation observations at NOAA gauges. The results are displayed in Fig. 12. From the results in this figure, the surrogate performs very similarly to ADCIRC when compared to real observations. Evaluation of our surrogate model takes seconds on a single CPU whereas running ADCIRC requires hours on hundreds of CPUs. Hence, this surrogate represents a speedup of several orders of magnitude. We remark that the ADCIRC performance for Ike is not indicative of the state-of-the-art. Better model performance was attained in Hope et al. (2013) by coupling ADCIRC to a short wave model and running an extensive tidal spinup beforehand. This reflects the fact that the model quality depends on the training data. Nevertheless, we have demonstrated our model is an effective surrogate for the ADCIRC model with no short wave coupling.

4.3.1. Alaska

One reason for the temporal partition of the Alaska data was to ensure that Typhoon Merbok would lie in the test dataset. Typhoon Merbok was significantly more intense than any other event in the dataset. Unsurprisingly, the model struggles with underprediction of the most extreme surges (see Fig. 13), with an RMSE of 0.61 m, and an R^2 score of 0.65. Accurate prediction of extreme surges for Merbok would likely require augmenting the training dataset with synthetic storms of higher intensity. We also here notice in Fig. 13 that the largest errors are in regions where the storm surge is largest (near Unalakleet and east of Nome).

As previously noted, observational data is more sparse for Alaska than for Texas. The only NOAA stations in the study domain are Unalakleet, Nome, and Red Dog Dock. These are the same stations used to select the surge events for the Alaska dataset. For each event in the test dataset, we compare the predictions of the surrogate XGBoost model and ADCIRC to observed NOAA data. At each station, we compute the RMSE over all events in the test data. These results are summarized in Fig. 14 where we observe that the ADCIRC model performs better than the surrogate model by about 0.3 to 0.5 m across stations.

5. Conclusions

We have presented a novel framework for developing surrogate models of storm surge. This framework is developed to ascertain the

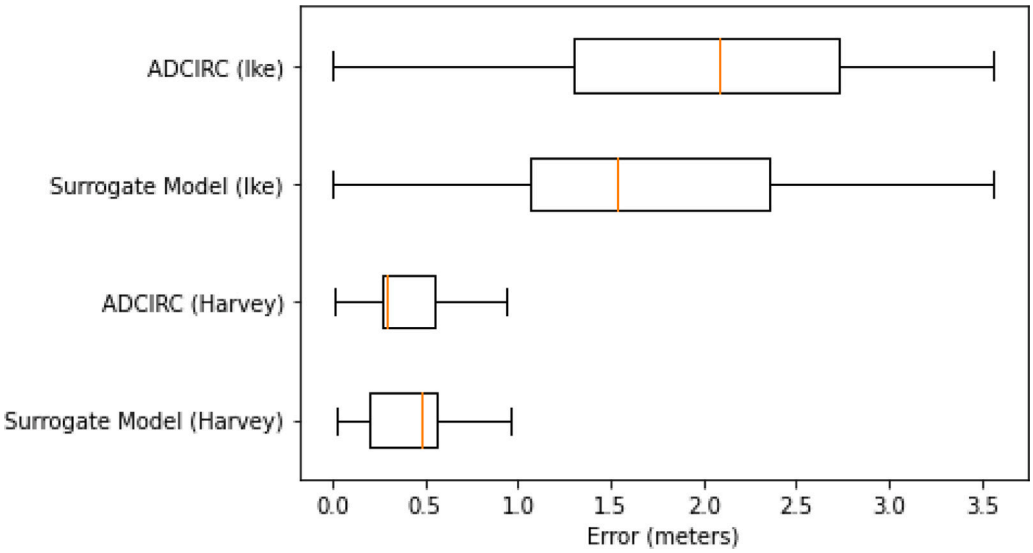


Fig. 12. Distribution of errors across observation points.

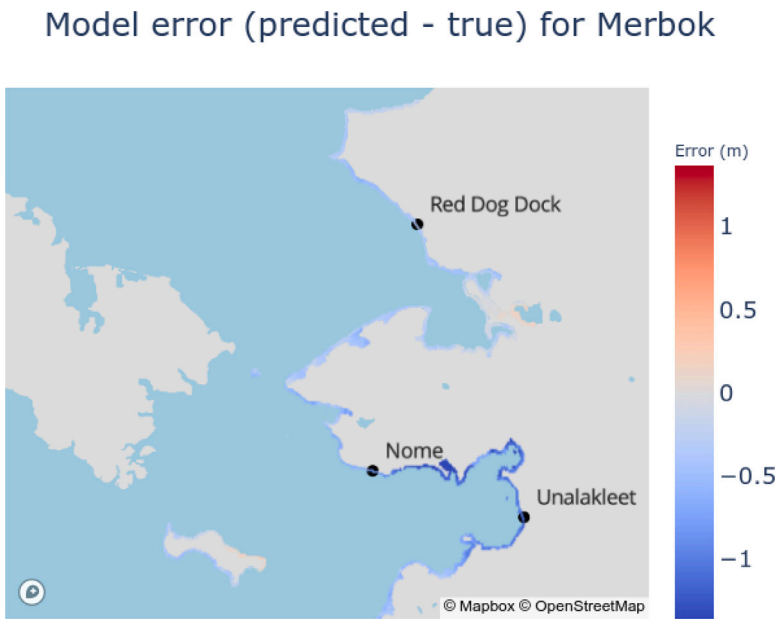


Fig. 13. Model error for Merbok. Map data courtesy of OpenStreetMap (OpenStreetMap contributors, 2022).

peak storm surge at selected points in the physical domain in response to physical forcing. Input forcings to the surrogates are the same physical drivers used in physics based models, e.g., winds and pressure, while the output is only the peak surge. We observe reasonable accuracy for two distinct regions with vastly different training data. The resulting surrogate models use a local formulation based on feature engineering, introduced in Section 3.2, that enables them to make predictions at locations not seen in the training data. Our framework is flexible enough to handle both landfalling tropical cyclones and general surge events. In the first case, we were able to accurately predict hurricane surges from synthetic training data. In the second, our model is able to use past events to make accurate predictions for future surge events. The same pipeline and code were used for these disparate applications with only minimal adjustments. Comparisons to observational data show that our surrogate models attain similar accuracy as the base ADCIRC model while being orders of magnitude faster. In Texas, our surrogate model is based on training data from synthetic hurricane events and does not contain tides. Hence, the surrogate may under or over-predict the peak surge in areas sensitive to tidal components of storm surge. However, the results presented in Section 4.3 indicates that the Texas model still performs well on the physical hurricanes where tides are included.

An interesting direction of future research is the development of larger scale storm surge models. The local formulation of our surrogate models means that the number of model parameters does not depend on the size of the study area or the desired output resolution. This is a significant advantage over past surrogate modeling approaches in terms of scalability.

A central question in the development of larger-scale models is how to more efficiently leverage training data. Generating training data is a major constraint for surrogate modeling, and models are frequently constrained by the availability of preexisting data. Often one region may be densely covered, while another region has little to no coverage.Consequently, we plan to test how well our model can use

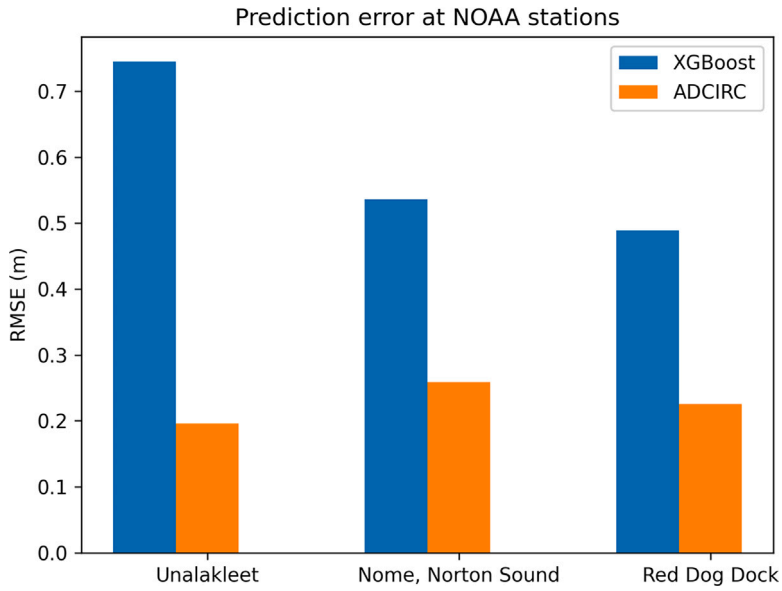


Fig. 14. Validation of surrogate model and ADCIRC predictions against NOAA observations. The RMSE is taken over events in the test data.

training data from densely covered parts of the domain to improve predictions in parts of the domain with more sparse coverage. Along these lines, we will investigate the prediction capabilities of points and areas that lie, potentially far away, from the training data set, as this new approach enables such predictions. In parallel, we plan to improve the feature representation used by the model to facilitate this kind of transfer prediction. Finally, we will consider additional classes of neural network architectures beyond the simple feed-forward networks studied in this work.

CRediT authorship contribution statement

Benjamin Pachev: Development of point-based surrogate model formulation, Generation of plots and figures, Use of supercomputing resources to run experiments, Editing and proofreading of the manuscript, Discussions on plots, experiments, and ideas to improve the model. **Prateek Arora:** Development of two-stage modeling approach, Survey of relevant literature, Development of neural network architecture and training methodology, Editing and proofreading of the manuscript, Discussions on plots, experiments, and ideas to improve the model. **Carlos del-Castillo-Negrete:** Generation and publication of Alaska dataset of storm surge events, Formatting the research code into a Python package for eventual publication, Editing and proofreading of the manuscript, Discussions on plots, experiments, and ideas to improve the model. **Eirik Valseth:** Description of the ADCIRC model and consultation on the physics of storm surge, Editing and proofreading of the manuscript, Discussions on plots, experiments, and ideas to improve the model. **Clint Dawson:** Oversight and procurement of funding, Editing and proofreading of the manuscript, Discussions on plots, experiments, and ideas to improve the model.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Links to the data and code are contained in the paper.

Acknowledgments

The authors would like to gratefully acknowledge the use of the "ADCIRC", "DMS23001". and "DMS21031" allocations at the Texas Advanced Computing Center at the University of Texas at Austin, as well the use of the DesignSafe platform for data storage.

Funding

This work was supported by the United States Department of Energy under grant DE-SC0022211 - MuSiKAL: Multiphysics Simulations and Knowledge discovery through AI/ML technologies. In addition, Benjamin Pachev was partially supported by a CSEM Fellowship from the Oden Institute for Computational Science and Engineering. The DesignSafe project is financially supported by the National Science Foundation, United States under grant CMMI-1520817.

Appendix A. Software and data contributions

We have made an effort to make all code and data contained in this study publicly available and open source, from raw data-set generation, to feature engineering and model training and validation. As mentioned before, the base storm datasets are available on the DesignSafe platform (Dawson et al., 2021; Del-Castillo-Negrete et al., 2023). These consist of the input files to run the full ADCIRC model for each storm used in training and validation, along with the full outputs for each run. While significant computational resources may be required to re-run the whole set of storms, code to run these ensembles via the DesignSafe platform using an HPC application is publicly available via the HPC application (Del-Castillo-Negrete, Carlos Emilio, 2022-02-15T18:44:55Z). We note a small base allocation of HPC resources to run these simulations is publicly available on DesignSafe to any natural hazards researcher, with the option for more allocations available upon request.

The code to generate features from the raw ADCIRC outputs and to train models is available in an open-source python package available on the PyPi repository (Pachev and Del-Castillo-Negrete, 2023), and can be easily installed running '`pip install adcirc-rom`'. Furthermore the python package contains a convenient command line interface (CLI) accessible via the '`arom`' command. After installing, run '`arom --help`' to see options for the two main entry-points to the CLI,

'`arom dataset`' and '`arom model`', with the first containing entry-points for feature engineering and the later containing entry-points for model training and validation. Note that while the library can run in serial environments, running on HPC architectures is recommended to speed up feature generation and model training. The final feature sets used for this study, along with the trained models and validation outputs are available as well in DesignSafe (Pachev et al., 2023-08-30). The engineered features can be used to train new models, while the pre-trained models can be used for direct prediction using new data.

Appendix B. Storm surge event identification algorithm

Algorithm 1 presents the modified version of the identification algorithm used in Wicks and Atkinson (2017) for identifying storm surge events. The identification algorithm operates by first finding intervals where the residual between predicted tides and observed water levels for a given tidal station gauge is above a given threshold, called the trigger threshold. The challenge becomes to cluster these events into discrete non-overlapping events. The algorithm clusters events by iterating between each gap between events. If the gap is shorter than the "lull" period or the surge does not go below a certain continuity threshold, expressed as a fraction of the trigger threshold, then the events surrounding that gap into one event. This process is repeated for each gap until all events are grouped into distinct ones.

Algorithm 1 Storm Surge Events Algorithm: Algorithm based off of Wicks and Atkinson (2017).
$R(t_i)$: The residual between the predicted tides and the observed water levels at times $\{t_i\}_{i=1}^{m}$
T : Trigger threshold (meters)
c : Continuity factor $0 < c <= 1$
L, S : Lull and Shoulder periods (hours)

GET-SURGE-EVENTS(R, T, c, L, S)
  $\{s^j\} \leftarrow$ Contiguous distinct sets $s^j = \{t_i : R(t_i) >= T\}$ ▷ Identify candidate surge events
  $\{g^j\} \leftarrow$ Contiguous distinct sets $g^j = \{t_i : R(t_i) < T\}$ ▷ Time gaps between events.
  for g^j **do**
    if If $R(t_i) >= cT$ for all t_i in g^j OR len$(g^j) < L$ **then**
      $s^{j-1} \leftarrow s^{j-1} + s^j$ ▷ Join events around g^j if cT not crossed or gap shorter than L
    else
      $s^j \leftarrow \{t_i : t_i > \min(s^j) - S\} \cup s^j \cup \{t_i : t_i < \max(s^j) + S\}$ ▷ Add S hours before/after event
      $E \leftarrow s^j$ ▷ Add to identified events set
    end if
  end for
  return E
end

We used the publicly available NOAA CO-OPs API (Anon, 2023) as our source for tidal gauge data, and focused on the three stations that see the most frequent storm surge events: Nome, Red Dog Dock, and Unalakleet. We ran Algorithm 1 for each stations data, across the whole range of data available for each stations, selected only the positive surge events, and then combined the date-ranges across each station, merging overlapping events. The dataset is fully available at Del-Castillo-Negrete et al. (2023).

Appendix C. Model selection results

See Tables C.5–C.7.

Table C.5
Comparison of models on the Texas test data.

Classifier	Regressor	R^2	RMSE (m)	Top 5% RMSE (m)
xgboost	xgboost	0.83	0.36	0.73
xgboost	nn1	0.85	0.33	0.69
xgboost	nn2	0.85	0.33	0.70
xgboost	nn3	0.85	0.33	0.73
nn1	xgboost	0.84	0.35	0.66
nn1	nn1	0.85	0.33	0.64
nn1	nn2	0.86	0.33	0.62
nn1	nn3	0.86	0.33	0.64
nn2	xgboost	0.84	0.34	0.66
nn2	nn1	0.87	0.31	0.61
nn2	nn2	0.86	0.32	0.58
nn2	nn3	0.87	0.31	0.58
nn3	xgboost	0.85	0.33	0.62
nn3	nn1	0.87	0.31	0.61
nn3	nn2	0.87	0.31	0.62
nn3	nn3	0.87	0.31	0.64

Table C.6
Comparison of models for the Alaska data with tides.

Classifier	Regressor	R^2	RMSE (m)	Top 5% RMSE (m)
xgboost	xgboost	0.73	0.36	0.87
xgboost	nn1	0.54	0.46	0.98
xgboost	nn2	0.57	0.45	1.15
xgboost	nn3	0.59	0.44	1.00
nn1	xgboost	0.76	0.33	0.74
nn1	nn1	0.54	0.46	1.06
nn1	nn2	0.51	0.48	1.15
nn1	nn3	0.43	0.52	1.01
nn2	xgboost	0.74	0.35	0.85
nn2	nn1	0.50	0.49	1.20
nn2	nn2	0.61	0.43	1.05
nn2	nn3	0.53	0.47	1.03
nn3	xgboost	0.71	0.37	0.92
nn3	nn1	0.55	0.46	0.99
nn3	nn2	0.41	0.53	1.27
nn3	nn3	0.56	0.45	1.08

Table C.7
Comparison of models for the Alaska data without tides.

Classifier	Regressor	R^2	RMSE (m)	Top 5% RMSE (m)
xgboost	xgboost	0.44	0.37	0.99
xgboost	nn1	0.15	0.46	1.18
xgboost	nn2	0.21	0.44	1.11
xgboost	nn3	−0.13	0.53	1.42
nn1	xgboost	0.45	0.37	1.04
nn1	nn1	0.05	0.48	1.02
nn1	nn2	0.08	0.47	1.20
nn1	nn3	−0.02	0.50	1.25
nn2	xgboost	0.50	0.35	0.92
nn2	nn1	0.02	0.49	1.17
nn2	nn2	0.05	0.48	1.37
nn2	nn3	−0.08	0.51	1.22
nn3	xgboost	0.54	0.33	0.84
nn3	nn1	−0.05	0.51	1.15
nn3	nn2	−0.05	0.51	1.20
nn3	nn3	0.00	0.49	1.21

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.coastaleng.2023.104406.

References

Agarap, A.F., 2019. Deep learning using rectified linear units (ReLU). arXiv:1803.08375.
Al Kajbaf, A., Bensi, M., 2020. Application of surrogate models in estimation of storm surge:A comparative assessment. Appl. Soft Comput. 91, 106184. http://dx.doi.org/10.1016/j.asoc.2020.106184, URL https://www.sciencedirect.com/science/article/pii/S1568494620301241.
Anon, 2023. NOAA Tides and Currents, https://tidesandcurrents.noaa.gov/.

Arora, P., Ceferino, L., 2022. Probabilistic and machine learning methods for uncertainty quantification in power outage prediction due to extreme events. EGUsphere 2022, 1–29. http://dx.doi.org/10.5194/egusphere-2022-975, URL https://egusphere.copernicus.org/preprints/egusphere-2022-975/.

Bass, B., Bedient, P., 2018. Surrogate modeling of joint flood risk across coastal watersheds. J. Hydrol. 558, 159–173. http://dx.doi.org/10.1016/j.jhydrol.2018.01.014, URL https://www.sciencedirect.com/science/article/pii/S0022169418300143.

Bezuglov, A., Blanton, B., Santiago, R., 2016. Multi-output artificial neural network for storm surge prediction in north carolina. CoRR abs/1609.07378, arXiv:1609.07378.

Blake, E.S., Landsea, C., Gibney, E.J., 2011. The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts).

Blier, W., Keefe, S., Shaffer, W.A., Kim, S.C., 1997. Storm surges in the region of Western Alaska. Mon. Weather Rev. 125 (12), 3094–3108. http://dx.doi.org/10.1175/1520-0493(1997)125<3094:SSITRO>2.0.CO;2.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Buzard, R.M., Christ, J., Overbeck, J.R., Endres, K.L., Plumb, E.W., 2021. Coastal Flood Impact Assessments for Alaska Communities. State of Alaska, Department of Natural Resources, Division of Geological.

Chen, T., Guestrin, C., 2016a. XGBoost: A scalable tree boosting system. CoRR abs/1603.02754, arXiv:1603.02754.

Chen, T., Guestrin, C., 2016b. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Chen, W.-B., Liu, W.-C., Hsu, M.-H., 2012. Predicting typhoon-induced storm surge tide with a two-dimensional hydrodynamic model and artificial neural network model. Nat. Hazards Earth Syst. Sci. 12 (12), 3799–3809. http://dx.doi.org/10.5194/nhess-12-3799-2012, URL https://nhess.copernicus.org/articles/12/3799/2012/.

Dawson, C., C., D.-C.-N., Shukla, A., B., P., C., K., Kutanoglu, E., 2021. ADCIRC simulation of synthetic storms in the gulf of Mexico. http://dx.doi.org/10.17603/ds2-68a9-0s64, Texas FEMA Hurricane Winds and Surge, DesignSafe-CI.

Del-Castillo-Negrete, C., Pachev, B., Arora, P., Valseth, E., Dawson, C., 2023. Alaska 1 m surge events - nome, red dog dock, unalakleet (1992 - 2022). Alaska Storm Surge Events http://dx.doi.org/10.17603/ds2-4rnb-j321.

Del-Castillo-Negrete, Carlos Emilio, 2022-02-15T18:44:55Z. Tapis Pylauncher. UT Austin Computational Hydraulics Group, URL https://github.com/UT-CHG/tapis-pylauncher.

Dietrich, J.C., Tanaka, S., Westerink, J.J., Dawson, C.N., Luettich, R., Zijlema, M., Holthuijsen, L.H., Smith, J., Westerink, L., Westerink, H., 2012. Performance of the unstructured-mesh, SWAN+ ADCIRC model in computing hurricane waves and surge. J. Sci. Comput. 52 (2), 468–497.

Dresback, K.M., Fleming, J.G., Blanton, B.O., Kaiser, C., Gourley, J.J., Tromble, E.M., Luettich Jr., R.A., Kolar, R.L., Hong, Y., Van Cooten, S., et al., 2013. Skill assessment of a real-time forecast system utilizing a coupled hydrologic and coastal hydrodynamic model during Hurricane Irene (2011). Cont. Shelf Res. 71, 78–94.

East, J.W., Turco, M.J., Mason Jr., R.R., 2008. Monitoring inland storm surge and flooding from hurricane Ike in Texas and Louisiana, September 2008. Surge 29, 95–20833.

Forbes, C., Luettich, R.A., Mattocks, C.A., Westerink, J.J., 2010. A retrospective evaluation of the storm surge produced by Hurricane Gustav (2008): Forecast and hindcast results. Weather Forecast. 25 (6), 1577–1602.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.

Goff, J.A., Swartz, J.M., Gulick, S.P., Dawson, C.N., de Alegria-Arzaburu, A.R., 2019. An outflow event on the left side of Hurricane Harvey: Erosion of barrier sand and seaward transport through Aransas Pass, Texas. Geomorphology 334, 44–57.

Graham, N.E., Diaz, H.F., 2001. Evidence for intensification of North Pacific winter cyclones since 1948. Bull. Am. Meteorol. Soc. 82 (9), 1869–1894. http://dx.doi.org/10.1175/1520-0477(2001)082<1869:EFIONP>2.3.CO;2.

Hashemi, M.R., Spaulding, M.L., Shaw, A., Farhadi, H., Lewis, M., 2016. An efficient artificial intelligence model for prediction of tropical storm surge. Nat. Hazards 82, 471–491.

Henry, R.F., Heaps, N.S., 1976. Storm surges in the southern Beaufort sea. J. Fish. Res. Board Can. 33 (10), 2362–2376. http://dx.doi.org/10.1139/f76-283.

Hope, M.E., Westerink, J.J., Kennedy, A.B., Kerr, P., Dietrich, J.C., Dawson, C., Bender, C.J., Smith, J., Jensen, R.E., Zijlema, M., et al., 2013. Hindcast and validation of Hurricane Ike (2008) waves, forerunner, and storm surge. J. Geophys. Res.: Oceans 118 (9), 4424–4460.

Ian, V.-K., Tse, R., Tang, S.-K., Pau, G., 2022. Performance analysis of machine learning algorithms in storm surge prediction. In: IoTBDS. pp. 297–303.

Irish, J.L., Resio, D.T., Cialone, M.A., 2009. A surge response function approach to coastal hazard assessment. Part 2: Quantification of spatial attributes of response functions. Natural Hazards 51, 183–205.

Jelesnianski, C.P., 1992. SLOSH: Sea, Lake, and Overland Surges from Hurricanes, Vol. 48. US Department of Commerce, National Oceanic and Atmospheric Administration.

Jia, G., Taflanidis, A.A., 2013. Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment. Comput. Methods Appl. Mech. Engrg. 261, 24–38.

Jia, G., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby, J.A., Kennedy, A.B., Smith, J.M., 2016. Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms. Nat. Hazards 81 (2), 909–938.

Johnson, W.R., Kowalik, Z., 1986. Modeling of storm surges in the Bering Sea and Norton Sound. J. Geophys. Res.: Oceans 91 (C4), 5119–5128. http://dx.doi.org/10.1029/JC091iC04p05119.

Joyce, B.R., Pringle, W.J., Wirasaet, D., Westerink, J.J., Van der Westhuysen, A.J., Grumbine, R., Feyen, J., 2019. High resolution modeling of western Alaskan tides and storm surge under varying sea ice conditions. Ocean Model. 141, 101421. http://dx.doi.org/10.1016/j.ocemod.2019.101421.

Kim, S.-W., Lee, A., Mun, J., 2018. A surrogate modeling for storm surge prediction using an artificial neural network. J. Coast. Res. (85 (10085)), 866–870.

Kim, S.-W., Melby, J.A., Nadal-Caraballo, N.C., Ratcliff, J., 2015. A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling. Nat. Hazards 76, 565–585.

Kinsman, N.E.M., DeRaps, M.R., 2012. Coastal Hazard Field Investigations in Response to the November 2011 Bering Sea Storm, Norton Sound, Alaska, Tech. Rep. RI 2012-2 v. 1.1, 1.1 ed. Alaska Division of Geological & Geophysical Surveys, pp. RI 2012–2 v. 1.1. http://dx.doi.org/10.14509/24484.

Kowalik, Z., 1984. Storm surges in the Beaufort and Chukchi Seas. J. Geophys. Res.: Oceans 89 (C6), 10570–10578. http://dx.doi.org/10.1029/JC089iC06p10570.

Kubat, M., 1999. Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7. Knowl. Eng. Rev. 13 (4), 409–412.

Kyprioti, A.P., Taflanidis, A.A., Nadal-Caraballo, N.C., Campbell, M., 2021. Storm hazard analysis over extended geospatial grids utilizing surrogate models. Coast. Eng. 168, 103855.

Lee, T.-L., 2006. Neural network prediction of a storm surge. Ocean Eng. 33 (3), 483–494. http://dx.doi.org/10.1016/j.oceaneng.2005.04.012, URL https://www.sciencedirect.com/science/article/pii/S002980180500140X.

Lee, T.-L., 2008. Back-propagation neural network for the prediction of the short-term storm surge in Taichung harbor, Taiwan. Eng. Appl. Artif. Intell. 21 (1), 63–72. http://dx.doi.org/10.1016/j.engappai.2007.03.002, URL https://www.sciencedirect.com/science/article/pii/S0952197607000425.

Lee, T.-L., 2009. Predictions of typhoon storm surge in Taiwan using artificial neural networks. Adv. Eng. Softw. 40 (11), 1200–1206. http://dx.doi.org/10.1016/j.advengsoft.2007.06.005, URL https://www.sciencedirect.com/science/article/pii/S0965997809001379.

Lee, J.-W., Irish, J.L., Bensi, M.T., Marcy, D.C., 2021a. Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning. Coast. Eng. 170, 104024. http://dx.doi.org/10.1016/j.coastaleng.2021.104024, URL https://www.sciencedirect.com/science/article/pii/S0378383921001691.

Lee, J.-W., Irish, J.L., Bensi, M.T., Marcy, D.C., 2021b. Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning. Coast. Eng. 170, 104024.

Luettich, R.A., Westerink, J.J., Scheffner, N.W., 1992. ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1, Theory and methodology of ADCIRC-2DD1 and ADCIRC-3DL. Coastal Engineering Research Center (US).

Manning, R., Griffith, J.P., Pigot, T., Vernon-Harcourt, L.F., 1890. On the flow of water in open channels and pipes.

Mesquita, M.d.S., Atkinson, D.E., Simmonds, I., Keay, K., Gottschalck, J., 2009. New perspectives on the synoptic development of the severe October 1992 Nome storm. Geophys. Res. Lett. 36 (13), http://dx.doi.org/10.1029/2009GL038824.

Narayan, S., 1997. The generalized sigmoid activation function: Competitive supervised learning. Inform. Sci. 99 (1), 69–82. http://dx.doi.org/10.1016/S0020-0255(96)00200-9, URL https://www.sciencedirect.com/science/article/pii/S0020025596002009.

National Centers for Environmental Information (NCEI), 2022. Billion-dollar weather and climate disasters. http://dx.doi.org/10.25921/stkw-7w73, URL https://www.ncdc.noaa.gov/billions/.

Neumann, J.E., Emanuel, K., Ravela, S., Ludwig, L., Kirshen, P., Bosma, K., Martinich, J., 2015. Joint effects of storm surge and sea-level rise on US Coasts: New economic estimates of impacts, adaptation, and benefits of mitigation policy. Clim. Change 129 (1), 337–349.

OpenStreetMap contributors, 2022. Planet dump. retrieved from https://planet.osm.org,.

Pachev, B., Del-Castillo-Negrete, C.E., 2023. Adcirc-rom. URL https://github.com/UT-CHG/adcirc-rom.

Pachev, B., Del-Castillo-Negrete, C.E., Arora, P., Clinton, D., Valseth, E., 2023-08-30. ADCIRC Reduced Order Modeling. DesignSafe-CI, URL https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published//PRJ-3870.

Pringle, W.J., Wirasaet, D., Roberts, K.J., Westerink, J.J., 2021. Global storm tide modeling with ADCIRC v55: Unstructured mesh design and performance. Geosci. Model Dev. 14 (2), 1125–1145.

Rappaport, E.N., 2014. Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. Bull. Am. Meteorol. Soc. 95 (3), 341–346.

Rathje, E.M., Dawson, C., Padgett, J.E., Pinelli, J.-P., Stanzione, D., Adair, A., Arduino, P., Brandenberg, S.J., Cockerill, T., Dey, C., et al., 2017. DesignSafe: New cyberinfrastructure for natural hazards engineering. Nat. Hazards Rev. 18 (3), 06017001.

Reddi, S.J., Kale, S., Kumar, S., 2019. On the convergence of adam and beyond. arXiv preprint arXiv:1904.09237.

Reimnitz, E., Maurer, D.K., 1979. Effects of storm surges on the Beaufort Sea Coast, Northern Alaska. ARCTIC 32 (4), 329–344. http://dx.doi.org/10.14430/arctic2631.

Resio, D.T., Irish, J., Cialone, M., 2009. A surge response function approach to coastal hazard assessment–part 1: Basic concepts. Natural Hazards 51, 163–182.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. nature 323 (6088), 533–536.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., Becker, E., 2014. The NCEP climate forecast system version 2. J. Clim. 27 (6), 2185–2208. http://dx.doi.org/10.1175/JCLI-D-12-00823.1.

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., den Dool, H.M.V., Pan, H.-L., Moorthi, S., Behringer, D., Stokes, D., Peña, M., Lord, S., White, G., Ebisuzaki, W., Peng, P., Xie, P., 2006. The NCEP climate forecast system. J. Clim. 19 (15), 3483–3517. http://dx.doi.org/10.1175/JCLI3812.1.

Tan, W.-Y., 1992. Shallow Water Hydrodynamics: Mathematical Theory and Numerical Solution for a Two-Dimensional System of Shallow-Water Equations. Elsevier.

Terenzi, J., Jorgenson, M.T., Ely, C.R., Giguère, N., 2014. Storm-surge flooding on the Yukon-Kuskokwim Delta, Alaska. Arctic 67 (3), 360–374, arXiv:24363780.

Wicks, A.J., Atkinson, D.E., 2017. Identification and classification of storm surge events at Red Dog Dock, Alaska, 2004–2014. Nat. Hazards 86 (2), 877–900. http://dx.doi.org/10.1007/s11069-016-2722-1.

Zhang, X., Chu, D., Zhang, J., 2021. Effects of nonlinear terms and topography in a storm surge model along the southeastern coast of China: A case study of Typhoon Chan-hom. Nat. Hazards 107, 551–574.