Learning Storm Surge with Gradient Boosting

Benjamin Pachev\textsuperscript{a,*}, Eirik Valseth\textsuperscript{a}, Clint Dawson\textsuperscript{a,b}

\textsuperscript{a}Oden Institute for Computational Engineering and Sciences, UT Austin
Austin, 78712, Texas, USA
\textsuperscript{b}Department of Aerospace Engineering and Engineering Mechanics, UT Austin
Austin, 78712, Texas, USA

Abstract
Storm surge is a major natural hazard for 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 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. While these models can attain good accuracy and are highly efficient, they are often limited to a small geographical region and a fixed set of output locations.

We develop a novel surrogate model for peak storm surge prediction based on gradient boosting. Unlike most surrogate approaches, our model is not explicitly constrained to a fixed set of output locations or specific geographical region. The model is trained with a database of 446 synthetic storms that make landfall on the Texas coast and obtains a mean absolute error of 0.25 meters. We additionally present a test of the model on Hurricanes Ike (2008) and Harvey (2017).

Keywords: Storm Surge, ADCIRC, Machine Learning

1. Introduction

In the last four decades, tropical cyclones have caused over one trillion dollars of damage in the United States alone \cite{1}. Storm surge is directly responsible for much of the property damage from tropical cyclones \cite{2} and nearly half of the fatalities \cite{3}. 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 \cite{4}. Thus, predicting storm surge is crucial in order to assess long-term risk to coastal

\*Corresponding author
Email address: benjaminpachev@gmail.com (Benjamin Pachev)
infrastructure and property from tropical cyclones. While accurate physical models of storm surge such as the ADvanced CIRCulation (ADCIRC) model [5] 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, [6]). However, low-fidelity models neglect key physics and consequently can have high errors.

A different approach is to construct a data-driven or surrogate model from a set of high-fidelity simulations. This has the potential to attain high accuracy while still being much faster than the high-fidelity model, enabling accurate statistical studies. There have been a number of proposed surrogate models for storm surge in recent years [7]. These models attain good accuracy but are typically restricted to making predictions at a fixed set of points. This complications generalization of the model to other geographical regions or to different output points than those with which the model was trained. Furthermore, many approaches rely on black-box machine learning techniques, resulting in models which are difficult to interpret.

The goal of this study is to develop a surrogate model for storm surge that is easy to interpret and to generalize to different regions and output locations. The remainder of the paper is organized as follows: in Section 2 we lay out our modeling approach, in Section 3 we present our numerical results. Finally, we conclude with a discussion of our results and future work in Section 4.

2. Modeling Approach

Let \( \Omega \) denote the spatial domain, \( x \in \Omega \), and let \( t \in \mathbb{R} \) denote time. A storm \( s \) is defined by vector-valued meteorological forcing data \( f_s(x, t) \) (this could include wind, pressure, etc.). The corresponding storm surge is given by the scalar valued function \( \eta_s(x, t) \) and is presumed to functionally depend on the forcing \( f_s \). The time-dependent storm surge prediction problem computes the mapping from \( f_s \) to \( \eta_s \). While ADCIRC and other high-fidelity physical codes solve the time-dependant storm surge prediction problem, most surrogate models solve the simpler problem of predicting maximum storm surge, defined as \( \eta_{s, \text{max}}(x) = \max_t \eta_s(x, t) \). For many applications, such as coastal risk assessment, knowing \( \eta_{s, \text{max}} \) is sufficient. Hereafter we will drop the superscript and refer to the maximum storm surge as \( \eta(x) \).

2.1. Surrogate problem formulation

Constructing a surrogate model requires a set of sample inputs and outputs from a high-fidelity storm surge model, e.g. \( (f_s(x, t), \eta_s(x)) \) \( s=1 \). Because of computational constraints, \( n \) is typically small - a few hundred or thousand. The standard approach in the literature is to select a fixed set of locations \( (x_i)_{i=1}^N \subset \Omega \) and predict \( \eta_s(x_i) \) for \( i = 1 \ldots N \). This approach has the following issues: firstly \( N \), the number of output locations, is typically much larger than \( n \), the number of training examples. This is typically dealt with by applying dimensional reduction techniques such as principal component analysis [8, 9, 10].
A bigger issue is that any given storm might affect only a small fraction of the output locations. For large enough $\Omega$, this will be true for every storm. Computing predictions over the entire domain for every storm could be highly inefficient and potentially bias the model.

We can address these issues by considering a domain of relevance, $\Omega_s \subset \Omega$, for each storm $s$. Predictions are made only for locations inside of $\Omega_s$, and anything outside of $\Omega_s$ is ignored. Note that to be useful in prediction $\Omega_s$ must be determined \textit{a priori} from $f_s$ without knowledge of $\eta_s$. A simple approach is to let $\Omega_s$ be a neighborhood of the landfall location (or point of closest approach for bypassing storms). Introducing $\Omega_s$ means that the per-storm output is no longer of fixed dimension $N$. Most machine learning algorithms require a fixed output dimension, so we reformulate the problem. Instead of considering one storm $s$ as one training example, we consider each combination of storm and output location $(s, x_i)$ as a training example. While this results in many more training examples, it reduces the output dimension to 1.

2.2. 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. The flow of coastal and oceanic water is governed by the shallow water equations [11]:

\[ \frac{\partial \zeta}{\partial t} + \nabla \cdot (Hu) = 0, \quad \text{in } \Omega, \]
\[ \frac{\partial u}{\partial t} + \nabla \cdot (Hu) + g \nabla \zeta = F, \quad \text{in } \Omega, \]

(1)

where $\zeta$ denotes sea water surface elevation, $u$ the depth-averaged velocity field, $H = \zeta + h_b$ the total water column elevation, $h_b$ the bathymetry, $\Omega$ the computational domain (e.g., the coastal ocean), and $F$ the source term with components from tides, friction, and atmospheric forcing. To solve this set of transient non-linear 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 [5, 12]. 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. [13].

ADCIRC has been extensively validated for past hurricanes, including Ike [14], Gustav [15] (both 2008), Katrina, Rita (both 2005) [16], Harvey (2017) [17], as well as others. In our current study, we employ the ADCIRC mesh validated for the Texas coast in [14] which contains 3,352,598 nodes and 6,675,517 triangular elements. This mesh covers the entire North Atlantic 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 10m, see Figure 1 for a plot of the mesh in Galveston Bay. To en-

Figure 1: ADCIRC mesh from [14] in the Galveston Bay area.

sure 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 through a friction coefficient used in a Manning’s $n$ friction formula [18] to ascertain physically relevant friction forcing. Due to its extensive validation and capabilities, ADCIRC represents an ideal candidate to ascertain training data for our machine learning algorithm.

2.3. Training Data

The dataset used in this study consists of ADCIRC output for 446 synthetic storms that make landfall on the Texas coast [19]. 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 [7, 9, 20]. For the synthetic storms, we apply their atmospheric and tidal forcings to our ADCIRC model which uses an extensively validated finite element mesh, see [14] to ensure high fidelity simulation results on the Texas coast. 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 minute temporal resolution simulated on the same ADCIRC mesh. Time-series water elevation output is available at each node with a temporal resolution of two hours, in addition to the maximum attained surge for each node in the finite element mesh.

While best-track data was not directly part of the ADCIRC input, we include it in the training data to simplify computing the domain of relevance $\Omega_s$. Best-track data consists of the location of the storm’s eye, the central pressure, and a number of other storm parameters. In this study, we only make use of the eye location data - all other storm parameters are ignored.

2.4. Surrogate Model

To construct the surrogate model, we used gradient boosting, a powerful technique for function approximation that is implemented in many machine learning libraries. Here we provide a brief overview of gradient boosting. More detailed presentations can be found in the literature [21, 22].

Suppose we wish to approximate an unknown function $F : \mathbb{R}^m \rightarrow \mathbb{R}$ from its values at a set of points $y_i = F(x_i), i = 1 \cdots N$. We can formulate this as an optimization problem

$$\min_{f} \sum_{i=1}^{N} l(f(x_i), y_i),$$

where $l$ is some smooth loss function (e.g. mean squared error). Gradient boosting starts with an initial guess $F_0(x)$ and proceeds iteratively. Let $F_n$ be the approximation at iteration $n$. The next iteration is constructed in two stages:

1. The negative gradients of the loss $-\frac{\partial l(F_n(x_i), y_i)}{\partial F_n(x_i)}$ are computed, and a ‘weak learner’ $h_{n+1}(x_i)$ is fit to the gradients. This is done with a simple function approximation algorithm, such as a regression tree.
2. The approximation is updated as $F_{n+1}(x) = F_n(x) + \beta_{n+1} h_{n+1}(x)$, where $\beta_{n+1}$ is obtained through line search.

This is essentially gradient descent - except the descent occurs in function space as opposed to parameter space. Instead of updating parameter values at each step, a new weak model is fit to correct the current model predictions. The result is a weighted ensemble of simpler models, and in final form is similar to other approaches such as random forests [23].

We use the popular XGBoost [24] library for gradient boosting. XGBoost implements a number of improvements to the base gradient boosting algorithm, including regularization to prevent overfitting and using second-order derivative information to accelerate convergence. It is extremely scalable and easy to use. In addition, XGBoost has native support for computing feature importances. Feature importances indicate how much each input variable contributes to the model predictions, and makes the results more interpretable.

2.4.1. Feature Engineering

We used a combination of local or location-specific and global or storm-specific features. For each storm, we use the track data to determine the landfall time and location. This forms the basis for the majority of the features. A few, such as bathymetry or distance to the coast, depend only on the output location and don’t change from storm to storm. The local features include:

- Mean, Max, and min pressure/wind at the output location over a two-hour window before landfall.
- Distance to landfall location
- Latitude & Longitude
- Bathymetry
- Distance to the coast

The global storm features are computed as the mean, max, and min of the local wind and pressure features. While additional storm parameters such as forward speed, central pressure, etc. were provided in the best-track data, we elected not to use these to avoid overfitting. The wind and pressure data were provided on a regular grid which did not correspond to the ADCIRC mesh. Consequently, the local wind and pressure were determined by interpolation.

2.4.2. Domain of Relevance and Output Locations

We determined the domain of relevance $\Omega_x$ for each storm by taking a fixed radius $r$ about the landfall location. Choosing $r$ too small would neglect potentially useful data, and choosing $r$ too large would include a large number of locations with low surge - which could bias the model towards underprediction. Such bias is bad because the most extreme surges are typically the most important. We chose to set $r$ uniformly to 100 kilometers based on manual inspection.
of a few sample storms. The ADCIRC mesh used to generate the training data has over 3 million nodes. Over one-third are on the coastline and could be potential locations for predicting storm surge. To speed up training time, we randomly selected about 10,000 nodes to use as output locations. Increasing this number did not significantly change the final accuracy, but did significantly increase training time. For any given storm, about one-fourth of the output locations fell within the domain of relevance.

2.4.3. Training
We randomly selected 90% of the storms for training and cross-validation. The remaining 10% of the data was set aside for final model validation. All feature and parameter selections were based on cross-validation performance and finalized before the model was validated on the holdout set. We performed group-aware cross validation to ensure that training examples corresponding to the same storm are not split between the training and test folds. While XGBoost supports a lot of hyper parameters, the default settings tend to work well. Consequently, we left all hyper parameters at their default values except for the number of boosting rounds, which was set to 200.

The final training was done in parallel on 56 CPUs and took about 3 minutes in real time. The results are presented in Section 3.

3. Results
We first evaluated the model on the holdout set, which consisted of 10% of the synthetic storms. Overall, we see very good performance. The average absolute error over the holdout dataset was .25 meters. As indicated by the results shown in Figure 2, the model error was fairly consistent across storms. While average error is important, it is critical for applications to correctly predict the most severe surges. In Figure 3, we show the average prediction error as a function of the true storm surge (the shaded region indicates the variance). While the absolute error increases for higher surges, the relative error actually decreases.

3.1. Sample synthetic storms
In Figures 4 and 5, we visually compare of our model’s predictions to the ADCIRC output for two storms in test set. The first storm has fairly typical error while the second has one of the highest errors in the test set. The model slightly underpredicts the surge for the first storm and overpredicts the surge for the second. In both cases, however, the general location of the most extreme storm surges is correctly captured.

3.2. Feature Importance
After training the model, we used XGBoost’s native feature importances to rank the features. While there was slight variation from run to run, the most important features by far were the local wind statistics, which corresponds to
Figure 2: Distribution of mean error across storms.

Figure 3: Average error as a function of true storm surge.
Figure 4: Comparison of true surge values to predictions for a synthetic storm. Map data courtesy of OpenStreetMap [25].

Figure 5: Comparison of true surge values to predictions for a synthetic storm. Map data courtesy of OpenStreetMap [25].
physical intuition. Global storm statistics were second in importance, followed by storm-independent features such as latitude/longitude and bathymetry. Local pressure did not have a significant effect on the model.

3.3. Validation of Historical Storms

We also tested our model on two historical storms - Harvey (2017) and Ike (2008). The performance on real storms was not as good as on the validation set of synthetic storms. The average error was .94 meters for Ike and .66 meters for Harvey. As Figure 6 shows, the model systematically under-predicts the most extreme storm surges for Ike. We observe similar - albeit less extreme under-prediction for Harvey. This is probably not due to a bias towards lower predictions, because as shown in Figure 5 the model is capable of over-prediction. A more plausible explanation is that Ike and Harvey in some sense are outside of the distribution of synthetic storms used for training. We quantify this by plotting the storm mean wind magnitude against the mean surge level. In Figure 7 the relation between mean wind magnitude and mean surge for Ike as compared to the validation set of synthetic storms is shown. While the synthetic storms follow a clear trendline, Ike has a much higher surge level relative to wind magnitude. This explains the under-prediction for Ike - to the model, it 'looks like' a weaker storm.

To further test this hypothesis, we created 10 additional synthetic storms that closely resembled Ike but with perturbed tracks. We trained a new model with the additional 10 storms included the training set. Unsurprisingly, the mean prediction error for Ike improved from .94 meters to .79. Interestingly, the
error for Harvey improved as well - from .66 meters to .58. The improvement for Harvey is significant. It suggests that generating training data from historical storms could be better than relying on synthetic storms alone.

4. Conclusion

We have presented a novel surrogate model for predicting maximum storm surge. Our model achieves good accuracy with relatively little feature engineering or parameter tuning, and is easy to interpret. In addition, we have proposed a local formulation of the storm surge prediction problem that eliminates the need for hard-coding a set of output locations into the model. This allows for greater efficiency during training by letting the model focus only on the relevant regions for each storm. Another advantage is that a model trained with the local formulation can be directly evaluated for output locations not present in the training set. In the future, we hope to leverage this capability to develop models that cover a wider geographical region than what is present in the training data.

We have further shown that in some sense the synthetic training data fail to capture real hurricanes such as Ike and Harvey. As this dataset was intended for use in coastal risk assessment [19], this finding is somewhat concerning. However, we leave further investigation of this issue to future work, as it is not the focus of the current study.

Acknowledgements

The authors would like to gratefully acknowledge the use of the ADCIRC allocation at the Texas Advanced Computing Center at the University of Texas.
at Austin.

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.

References

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

[2] J. E. Neumann, K. Emanuel, S. Ravela, L. Ludwig, P. Kirshen, K. Bosma, J. Martinich, Joint effects of storm surge and sea-level rise on us coasts: new economic estimates of impacts, adaptation, and benefits of mitigation policy, Climatic Change 129 (1) (2015) 337–349.

[3] E. N. Rappaport, Fatalities in the united states from atlantic tropical cyclones: New data and interpretation, Bulletin of the American Meteorological Society 95 (3) (2014) 341–346.

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

[5] R. A. Luettich, J. J. Westerink, N. W. Scheffner, et al., Adcirc: an advanced three-dimensional circulation model for shelves, coasts, and estuaries. report 1, theory and methodology of adcirc-2ld1 and adcirc-3dl (1992).

[6] C. P. Jelesnianski, SLOSH: Sea, lake, and overland surges from hurricanes, Vol. 48, US Department of Commerce, National Oceanic and Atmospheric Administration . . . , 1992.

[7] J.-W. Lee, J. L. Irish, M. T. Bensi, D. C. Marcy, Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning, Coastal Engineering 170 (2021) 104024.

[8] G. Jia, A. A. Taflanidis, Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment, Computer Methods in Applied Mechanics and Engineering 261 (2013) 24–38.

[9] G. Jia, A. A. Taflanidis, N. C. Nadal-Caraballo, J. A. Melby, A. B. Kennedy, J. M. Smith, Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms, Natural Hazards 81 (2) (2016) 909–938.
[10] A. P. Kyprioti, A. A. Taflanidis, N. C. Nadal-Caraballo, M. Campbell, Storm hazard analysis over extended geospatial grids utilizing surrogate models, Coastal Engineering 168 (2021) 103855.

[11] W.-Y. Tan, Shallow water hydrodynamics: Mathematical theory and numerical solution for a two-dimensional system of shallow-water equations, Elsevier, 1992.

[12] W. J. Pringle, D. Wirasaet, K. J. Roberts, J. J. Westerink, Global storm tide modeling with adcirc v55: unstructured mesh design and performance, Geoscientific Model Development 14 (2) (2021) 1125–1145.

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

[14] M. E. Hope, J. J. Westerink, A. B. Kennedy, P. Kerr, J. C. Dietrich, C. Dawson, C. J. Bender, J. Smith, R. E. Jensen, M. Zijlema, et al., Hindcast and validation of hurricane ike (2008) waves, forerunner, and storm surge, Journal of Geophysical Research: Oceans 118 (9) (2013) 4424–4460.

[15] C. Forbes, R. A. Luetttich, C. A. Mattocks, J. J. Westerink, A retrospective evaluation of the storm surge produced by hurricane gustav (2008): Forecast and hindcast results, Weather and Forecasting 25 (6) (2010) 1577–1602.

[16] J. C. Dietrich, S. Tanaka, J. J. Westerink, C. N. Dawson, R. Luetttich, M. Zijlema, L. H. Holthuijsen, J. Smith, L. Westerink, H. Westerink, Performance of the unstructured-mesh, swan+ adcirc model in computing hurricane waves and surge, Journal of Scientific Computing 52 (2) (2012) 468–497.

[17] J. A. Goff, J. M. Swartz, S. P. Gulick, C. N. Dawson, A. R. de Alegría-Arzaburu, An outflow event on the left side of hurricane harvey: Erosion of barrier sand and seaward transport through aransas pass, texas, Geomorphology 334 (2019) 44–57.

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

[19] C. D.-C.-N. Dawson, A. P. C. Shukla, B. Kaiser, E. C. Kutanoglu, Adcirc simulation of synthetic storms in the gulf of mexico, Texas FEMA Hurricane Winds and Surge (2021).

[20] B. Bass, P. Bedient, Surrogate modeling of joint flood risk across coastal watersheds, Journal of Hydrology 558 (2018) 159–173.
[21] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.

[22] J. H. Friedman, Stochastic gradient boosting, Computational statistics & data analysis 38 (4) (2002) 367–378.

[23] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[24] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.

[25] OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org, https://www.openstreetmap.org (2022).