7C.2 A METHOD FOR PROBABILISTIC SURGE FORECASTING WITH HIGH-FIDELITY MODELS

Taylor G. Asher UNC Chapel Hill, NC *
Matthew Plumlee Northwestern University, IL
Won Chang University of Cincinnati, OH
Matthew V. Bilskie Louisiana State University, LA
Rick A. Luettich Institute of Marine Sciences, UNC Chapel Hill, NC

* Corresponding author, UNC Department of Marine Sciences, email: taylor.asher.res at gmail

1 BACKGROUND

Full details of our work will be published in our forthcoming paper, Plumlee et al. 2020; the full citation is in the References.

There is substantial uncertainty in tropical cyclone surge forecasts, even at lead times of just 24 hours, due to the uncertainty in tropical cyclone meteorological forecasts. This has been a major obstacle to forecasting with high-fidelity hydrodynamic models due to the resulting computational cost of large ensembles. Current surge forecasting from the National Weather Service uses P-Surge (Forbes & Rhome 2011, Taylor & Glahn 2008) for probabilistic forecasts and peak surge atlases for scenario-based guidance. Both of these are built on the SLOSH model (Jelesnianski et al. 1992), a very fast and reliable, but also less accurate surge model (Kerr et al. 2013). Demand for more advanced models has led to alternate forecast products (e.g. Blanton et al. 2012) that rely on running a small number of simulations of a very high fidelity model, whose results can be provided as-is or interpreted by experts to provide specific guidance. These methods lead to fundamentally different guidance products and each has its limitations: The higher, unknown levels of error from a coarser model necessitate greater caution in guidance products. Conversely, hand-selected high-fidelity simulations do not come with a likelihood and require expert judgment to interpolate/extrapolate results to other scenarios.

2 METHODS

We have developed and applied an approach that attempts to deal with all of these concerns by combining a high-fidelity surge model with modern statistical tools. At the core of our approach is the goal of constructing an emulator—a very computationally cheap surrogate for the hydrodynamic model that is trained on simulation data—tailored for a particular hurricane. This allows arbitrary scenario-based surge estimates to be calculated in just a few seconds, while also supporting probabilistic forecasts. For an impending hurricane, a handful of surge model simulations are run to train the surrogate. These simulations use slightly different versions of the forecast hurricane, and are selected in a manner that minimizes error in the surrogate, given the forecast and its uncertainty. This approach takes into account both the fact that the forecast changes every advisory and that simulations were carried out in previous forecast cycles. Thus, our approach reuses previous simulations to maximize use of computational resources. Brief explanations of each of these elements are provided in the following subsections.

2.1 Surrogate

Peak storm surge response to a hurricane over a region is represented as a Gaussian Process emulator (Gu & Berger 2016). Gaussian process-based emulation of coastal water levels have been employed successfully in several studies (e.g. Zhang et al. 2019, Parker et al. 2019). We use a six-dimensional characterization of the storms at landfall as inputs to the surrogate: latitude (LAT), longitude (LONG), heading (H), forward speed (FS), maximum wind speed (MWS), and radius of 34kt isotach (R34). This characterization is unique to this work, and was chosen to conform to NHC forecast data while providing a good summary of hurricane properties relevant to storm surge.

2.2 Selecting Simulations

The goal is to select the storms to-be-simulated which will produce the most accurate estimate of the probabilistic surge hazard, given the forecast distribution, the emulator error distribution, and what simulations have already been run. We used a modified version of the integrated mean squared error (IMSE) criterion of Sacks et al. (1989) to define the minimization problem. To select the actual storm parameter combinations to be run, we used a randomized search algorithm (e.g. Gramacy & Lee 2009). Importantly, the effect on the error structure of new samples can be assessed *a priori* because the Gaussian process emulator's error structure is defined intrinsically.

The forecast parameters' covariance structure and error distribution were determined based on recent NHC forecast data (Cangialosi 2018), but without any knowledge of the current hurricane's future states (since this is unknowable).

2.3 Forecast Revision

Once a set of alternate landfall characteristics have been selected, the NHC forecast must be modified to create a full storm track for each revised storm. To ensure the surge forecasts are most relevant to the particular structure of the storm at hand, we utilize information on the full storm by perturbing the forecasted storm's track and characteristics (using the NHC forecast as our baseline) with a mixed physical-statistical model. The six parameters used in emulation are determined with a Gaussian process statistical model based on historical forecast data that takes as input the storm's current state, the forecast, and the desired landfalling state. We then use a generalization of the Holland (1980) relation (see Dietrich et al. 2018) to perturb additional parameters in the NHC forecast that are not in our surrogate, namely the individual isotachs. As an illustration of this process, a sequence of randomly drawn versions of Hurricane Michael, the storm used in our case study, is shown in Figure 1; more details on this are in the next section.

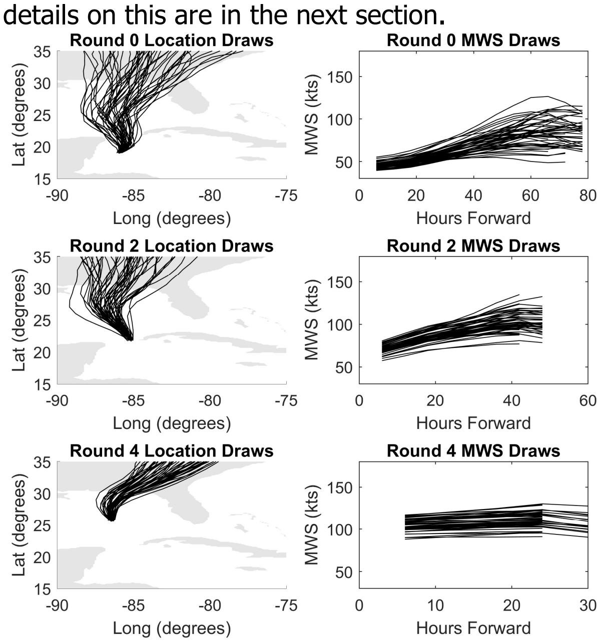

Figure 1: Randomly drawn tracks and intensities from forecast distributions at rounds 0, 2, and 4.

3 CASE STUDY & RESULTS

We carried out a pilot study of this method using 2018's Hurricane Michael. Michael represents an interesting test case in that the storm underwent rapid intensification for nearly all of its short 3-day lifespan before making landfall in the Florida Panhandle, which was consistently under-predicted by the NHC, leading to large, low-biased intensity errors (Beven et al. 2019). Conversely, track errors were unusually small, with landfall well-predicted 30 hours before landfall.

We began with a set of 30 simulations (termed "Round 0") covering a very wide range of storm states across the region of interest, the Gulf of Mexico from Mississippi through the Florida Panhandle. This provides a baseline dataset that could easily be produced across larger regions. We then performed forecasts of the surge at every-other NHC advisory, i.e. every 12 hours, over Michael's brief lifespan. This corresponds to five rounds, with forecasts at 60, 48, 36, 24, and 12 hours prior to landfall; we chose 12-hour intervals to reduce the human effort for this pilot study. We simulated 10, 20, or 30 hurricanes each 12-hour round. The ADCIRC hydrodynamic model (Luettich et al. 1992, Westerink et al. 2008) was used with a high-resolution mesh (Figure 2) designed for operational forecasting covering the northeastern Gulf of Mexico (Bilskie et al. 2020). The mesh has just over 2 million nodes, and has resolution in the tens of meters in nearshore and overland areas. Runs were done without tides nor wave coupling, discussed further later.

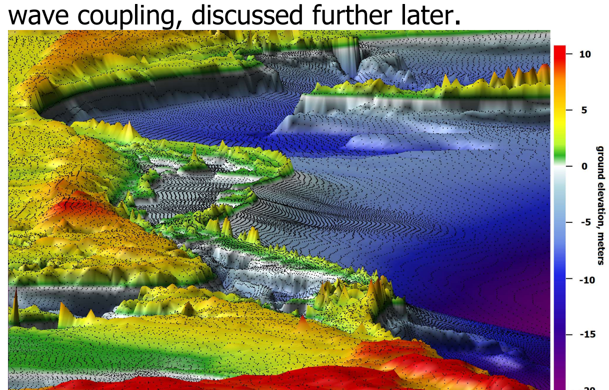

Figure 2: ADCIRC mesh across the Florida Panhandle. View is looking eastward from Panama City, black dots are mesh nodes.

We began with thirty simulations in round 1, then twenty runs in rounds 2 and 3, and ten in rounds 4 and 5. The evolution of sampled hurricane parameters is shown in Figure 3. The sampled (i.e. "design") storms can be seen to move with the shifting distribution, while filling in gaps left by prior rounds. The median predicted surge is shown against the hindcast surge in Figure 4. Here the predictions grow toward the hindcast surge as the rounds progress, though this is driven by Michael's increasing intensity forecasts.

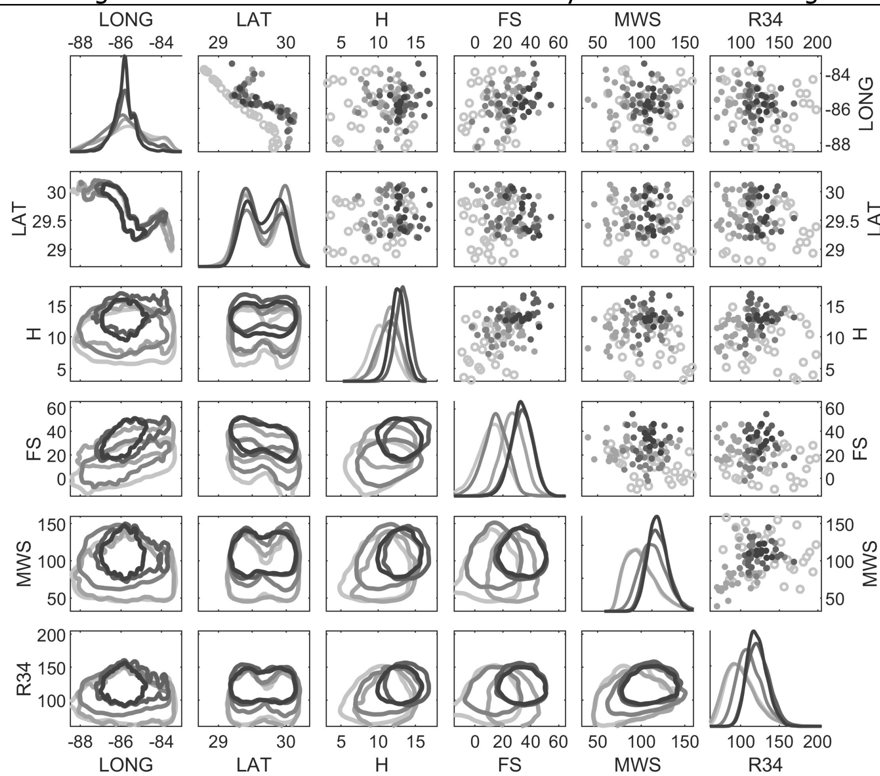

Figure 3: Parameter forecast distributions and samples; all data are for rounds 1-5 and darker colors imply later rounds. Bottom-left panels show the 90% forecast PDF contours. Diagonal panels show marginal forecast densities for each landfall characteristic. Top-right panels show the selected design storms' parameters; open circles are the round 0 parameters.

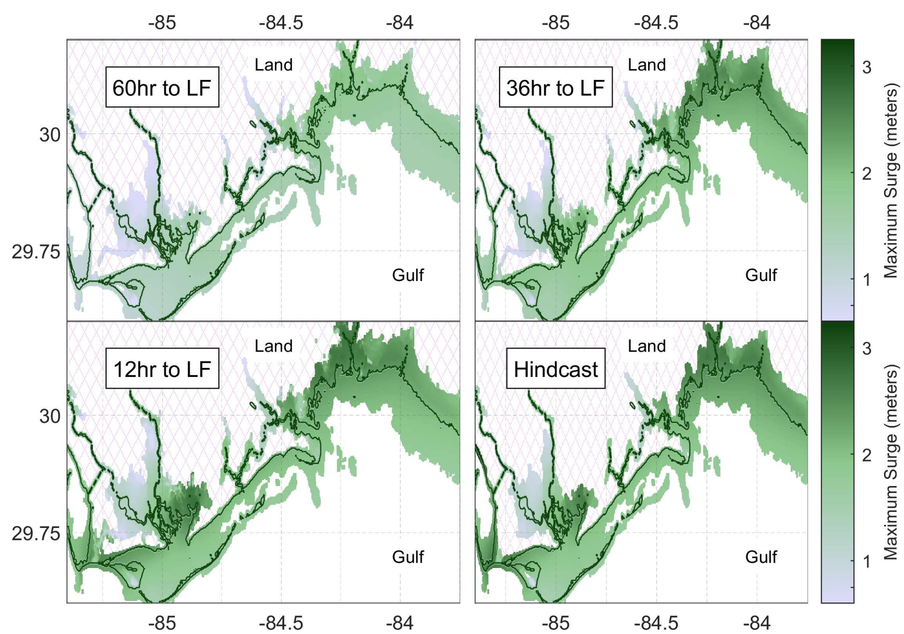

Figure 4: Hindcast peak surge (bottom right panel) and predictive median peak surge (other three panels) using the predictive distribution in the Florida Panhandle. Color is shown only for nodes shallower than 4 meters depth that wetted. Black line is the coastline.

To evaluate the forecast model's performance, several error metrics were calculated, including root mean squared error (RMSE), mean absolute error (MAE), the Dawid-Sebastiani score, and the misclassification rate. The Dawid-Sebastiani score evaluates the adequacy of both the forecast variance and the mean (see, e.g. Gneiting & Raftery 2007). The misclassification rate indicates the percentage of nodes predicted as being wet when they were actually dry or vice-versa. Performance was only evaluated for overland points within the study area that were wetted by simulations in order to not over-estimate performance by including easy-to-predict points with little surge. Results are shown in Figure 5, This figure also shows performance when the storm's landfall state is known. In other words, with forecast uncertainty removed. This allows a rough comparison of uncertainty due to the surrogate vs. that due to forecast uncertainty, although the two are intrinsically linked by the procedure employed: the surrogate depends on the simulations performed, which are chosen based on the forecast uncertainty. As can be seen, forecast uncertainty quickly dominates, suggesting that the surrogate model works reasonably well and is on par with errors in high quality surge model hindcasts of 20-30 cm. The oscillations from round to round when landfall is known are surprising, and may be attributable to the fact that this has been done for a single storm, meaning sample size is one.

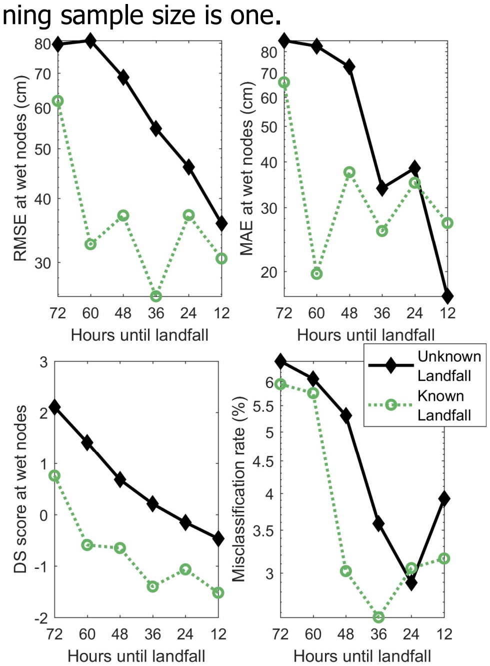

Figure 5: Surge forecast performance with known and unknown landfalling storm states.

Emboldened by the promising results thus far and wishing to ground-truth the work, we also carried out a comparison to real measured surge data and NWS's P-Surge probabilistic model. Results are shown in Figure 6. Note that this means there is additional uncertainty due to errors in the surge models and errors in parameterized meteorological forcing. Our system generally performs comparably to P-Surge. Note that one should not necessarily expect the median to correspond to the actual surge due to the effects of uncertainty on the response, and so focus should be placed on quantile coverage.

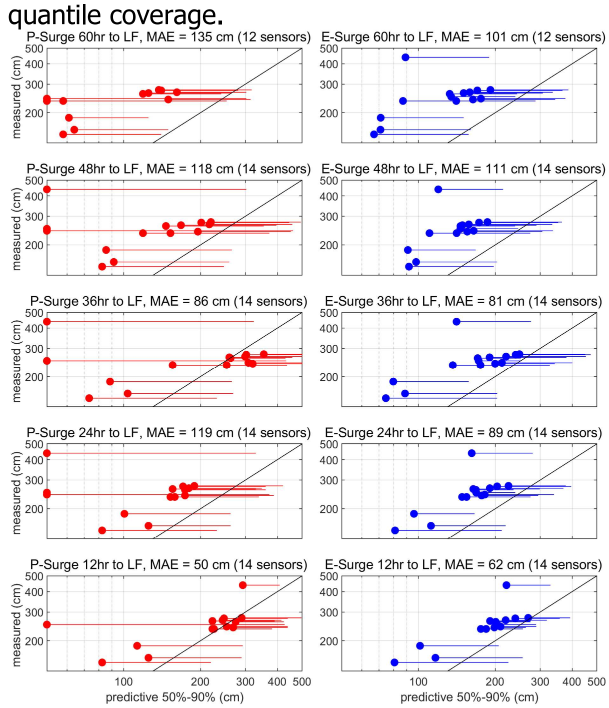

Figure 6: Predictive accuracy of P-Surge (left) and our proposed method (right) vs. observed surge data for rounds 1 (top) through 5 (bottom). The dot represents the forecast median and the line to the right extends to the 90% quantile of the forecast distribution. If no dot is present for P-Surge, it means there was no prediction for that round. If the 90% quantile was present for P-Surge but the 50% was not, then the 50% value is set to 50 cm for plotting purposes.

Some of the higher variance in the P-Surge predictions is likely attributable to the inclusion of tides, which were excluded for our simulations since tide range is small and Michael made landfall at mid-tide, meaning the effect of neglecting it should be small. We performed an additional hindcast simulation (not shown) including both tides and wave coupling to evaluate the effect of excluding these. Changes in peak surge were generally less than 10 to 30 cm, though the effect was larger, around 50 cm, in the area immediately around peak surge, which we attribute to wave setup. This is notable since the highest observation in Figure 6 coincidentally corresponds to the location of highest modeled surge in the storm, at Mexico Beach, FL, where model under-prediction is largest. This points to the importance of more advanced modeling, though it is difficult to read too much into a single data point.

4 CONCLUSION

We have developed a method to optimize use of computational resources for calculating both probabilistic and deterministic surges from impending hurricanes that is capable of being fully automated and run with presently available resources. The method leverages several statistical tools to optimally select which simulations to run through the costly hydrodynamic model in order to construct a cheap Gaussian process-based surrogate for the hydrodynamic model. The method is designed to function optimally in forecasting where the target is always moving as the storm changes. Since it does not require any fixed design, the method is robust against data losses, changes in available resources, and other operational challenges. The statistical methods were developed with speed in mind, and all steps can be run in under 20 minutes on a laptop. Prediction of surge from a new storm with the surrogate model takes seconds. In a pilot study with Hurricane Michael, the method was shown to be effective with, on average, 10 simulations per 6-hour forecast cycle. The method not only converges quickly, but also performed on par with NWS's P-Surge, though we emphasize it is difficult to generalize performance based on only one storm. Results are extremely promising, and far exceeded our expectations, especially given the unusual characteristics of Michael. For instance, a hurricane with more than 3 days of lead time would have permitted several earlier rounds of simulations, meaning errors at a given time-before-landfall should be improved.

Considerable improvements can be made to our method. Most notably, results suggest the hurricane forecast uncertainty is a driving component of total surge forecast uncertainty, and so better characterizing the hurricane forecast distribution is key to producing better surge estimates. Our current forecast distribution is independent of the particular storm under consideration, even though uncertainty in forecasts clearly varies storm-to-storm and across different forecast cycles. Many more advanced surrogate modeling strategies exist, and work on this will continue. We also anticipate needing to move beyond a purely "at-landfall" characterization in order to handle storms whose exact time of landfall is ill-defined due to irregular coastal geometry; Hurricane Irma's (2017) path toward the Florida peninsula illustrates the potential importance of this.

5 ACKNOWLEDGEMENTS

This material was based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute.
National Science Foundation under Grant Award Number ACI-1339723 and by the U.S. Department

6 REFERENCES

Beven, J.L., Berg, R., Hagen, A., 2019. Hurricane Michael (Tropical Cyclone Report No. AL142018). National Hurricane Center.

Bilskie, M. V., Hagen, S. C. and Medeiros, S. C., 2020. Unstructured finite element mesh decimation for real-time hurricane storm surge forecasting. Coastal Engineering. **Under Review**.

Blanton, B., McGee, J., Fleming, J., Kaiser, C., Kaiser, H., Lander, H., Luettich, R., Dresback, K., Kolar, R., 2012. Urgent Computing of Storm Surge for North Carolina's Coast. Procedia Computer Science, Proceedings of the International Conference on Computational Science, ICCS 2012 9, 1677–1686. https://doi.org/10.1016/j.procs.2012.04.185

Cangialosi, J.P., 2018. National Hurricane Center Forecast Verification Report: 2017 Hurricane Season. NOAA/NWS/NHC.

Cangialosi, J.P., 2019. National Hurricane Center Forecast Verification Report: 2018 Hurricane Season. NOAA/NWS/NHC.

Dietrich J. C., Muhammad A., Curcic M., Fathi A., Dawson C. N., Chen S. S., Luettich R. A., 2018. Sensitivity of Storm Surge Predictions to Atmospheric Forcing during Hurricane Isaac. Journal of Waterway, Port, Coastal, and Ocean Engineering 144. https://doi.org/10.1061/(ASCE)WW.1943-5460.0000419

Forbes, C., Rhome, J., 2011. An Automated Operational Storm Surge Prediction System for the National Hurricane Center, in: Estuarine and Coastal Modeling (2011), Proceedings. Presented at the Estuarine and Coastal Modeling, pp. 213–229. https://doi.org/10.1061/97807844124110.0001 3

Gneiting, T. and Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 359–378.

Gramacy, R. B. and Lee, H. K., 2009. Adaptive design and analysis of supercomputer experiments. Technometrics 51 130–145.

Gu, M. and Berger, J. O.(2016). Parallel partial Gaussian process emulation for computer models with massive output. The Annals of Applied Statistics 10 1317–1347.

Holland, G.J., 1980. An Analytic Model of the Wind and Pressure Profiles in Hurricanes. Mon. Wea. Rev. 108, 1212–1218. https://doi.org/10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;2

Jelesnianski, C.P., Chen, J., Shaffer, W.A., 1992. SLOSH: Sea, lake, and overland surges from hurricanes (Technical Report No. 48), NOAA Technical Report NWS 48. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, MD.

Kerr, P.C., Donahue, A.S., Westerink, J.J., Luettich, R.A., Zheng, L.Y., Weisberg, R.H., Huang, Y., Wang, H.V., Teng, Y., Forrest, D.R., Roland, A., Haase, A.T., Kramer, A.W., Taylor, A.A., Rhome, J.R., Feyen, J.C., Signell, R.P., Hanson, J.L., Hope, M.E., Estes, R.M., Dominguez, R.A., Dunbar, R.P., Semeraro, L.N., Westerink, H.J., Kennedy, A.B., Smith, J.M., Powell, M.D., Cardone, V.J., Cox, A.T., 2013. U.S. IOOS coastal and ocean modeling testbed: Inter-model evaluation of tides, waves, and hurricane surge in the Gulf of Mexico. J. Geophys. Res. Oceans 118, 5129–5172. https://doi.org/10.1002/jgrc.20376

Luettich, R.A., Westerink, J.J., Scheffner, N.W., 1992. ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1. Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL. (No. CERC-TR-DRP-92-6). Coastal Engineering Research Center, Vicksburg MS.

Parker, K., Ruggiero, P., Serafin, K.A., Hill, D.F., 2019. Emulation as an approach for rapid estuarine modeling. Coastal Engineering 150, 79–93. https://doi.org/10.1016/j.coastaleng.2019.03.004

Plumlee, M., Asher, T.G., Chang, W., Bilskie, M.V., 2020. High-Fidelity Hurricane Surge Forecasting Using Emulation and Sequential Experiments. Annals of Applied Statistics. **Under review**.

Sacks, J., Welch, W. J., Mitchell, T. J. Wynn, H. P., 1989. Design and analysis of computer experiments. Statistical Science 409–423.

Taylor, A.A., Glahn, B., 2008. Probabilistic guidance for hurricane storm surge, in: 19th Conference on Probability and Statistics. p. 8.

Westerink, J.J., Luettich, R.A., Feyen, J.C., Atkinson, J.H., Dawson, C., Roberts, H.J., Powell, M.D., Dunion, J.P., Kubatko, E.J., Pourtaheri, H., 2008. A Basin- to Channel-Scale Unstructured Grid Hurricane Storm Surge Model Applied to Southern Louisiana. Mon. Wea. Rev. 136, 833–864. https://doi.org/10.1175/2007MWR1946.1

Zhang, J., Taflanidis, A.A., Nadal-Caraballo, N.C., Melby, J.A., Diop, F., 2018. Advances in surrogate modeling for storm surge prediction: storm selection and addressing characteristics related to climate change. Nat Hazards. https://doi.org/10.1007/s11069-018-3470-1