npj Natural Hazards
Supplementary Information for

Neglecting Spatiotemporal Rainfall Variability Underestimates Flood Hazard and Risk

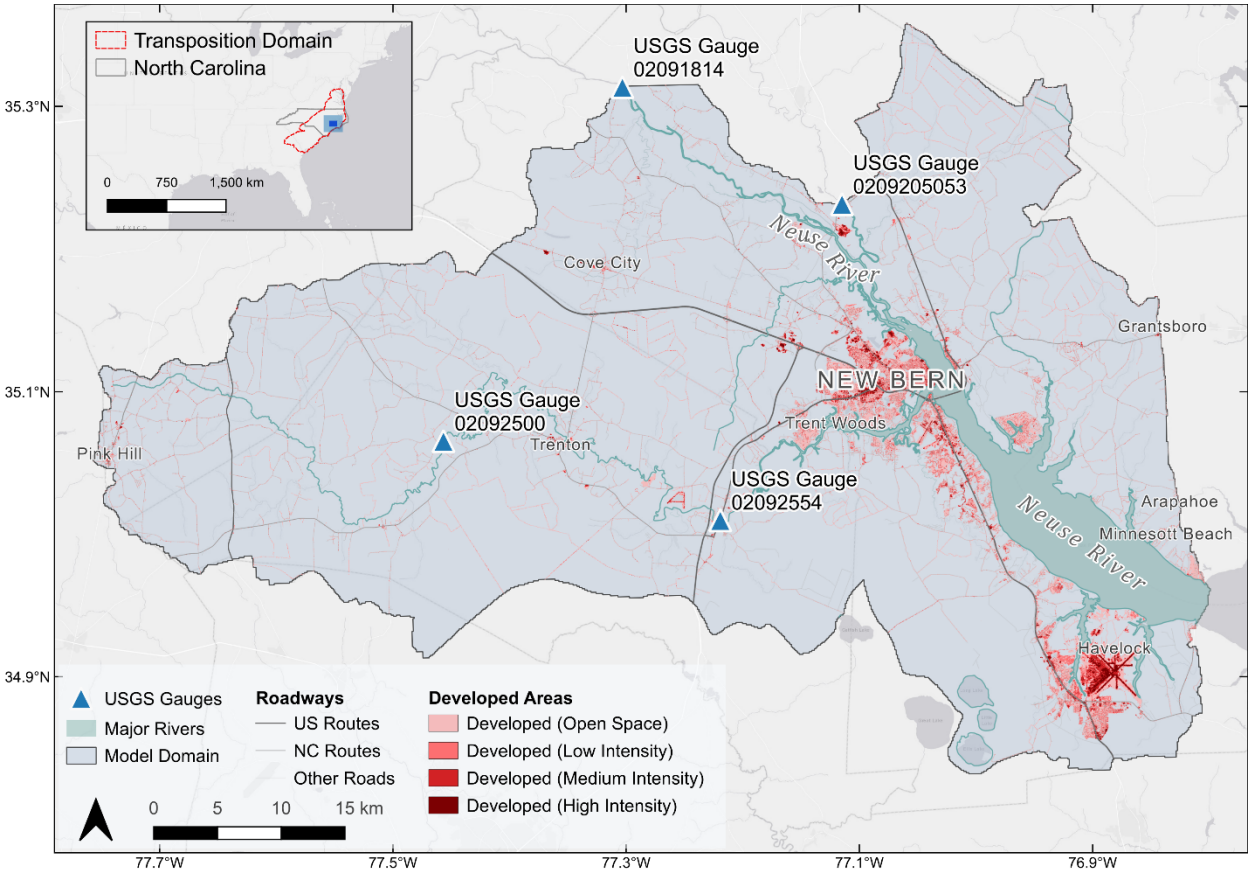
John A. Baer[1], Antonia Sebastian*[1,2], Lauren E. Grimley[1], James Doss-Gollin[3,4], Daniel B. Wright[5], Mohammad Ashar Hussain[5], Marissa Webber[1]

[1]Department of Earth, Marine, and Environmental Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]Environment, Energy and Ecology Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA
[4]Ken Kennedy Institute, Rice University, Houston, TX, USA
[5]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA

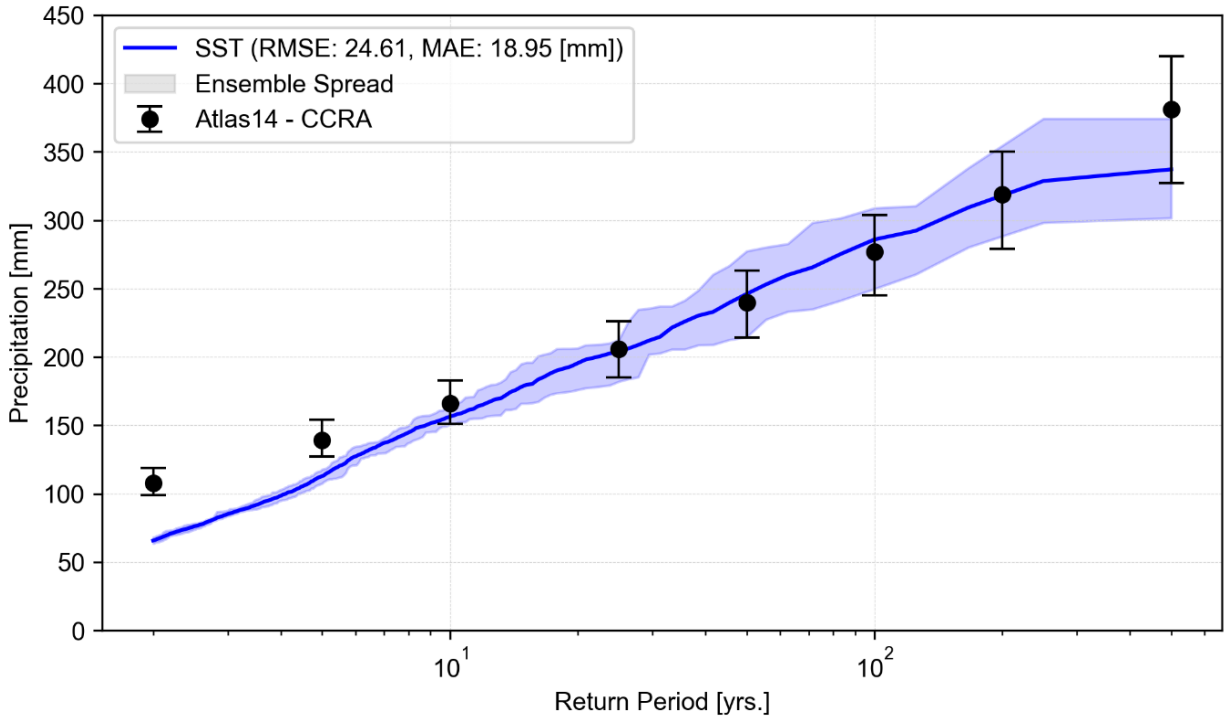
**Correspondence to:* Antonia Sebastian (asebastian@unc.edu)

Contents

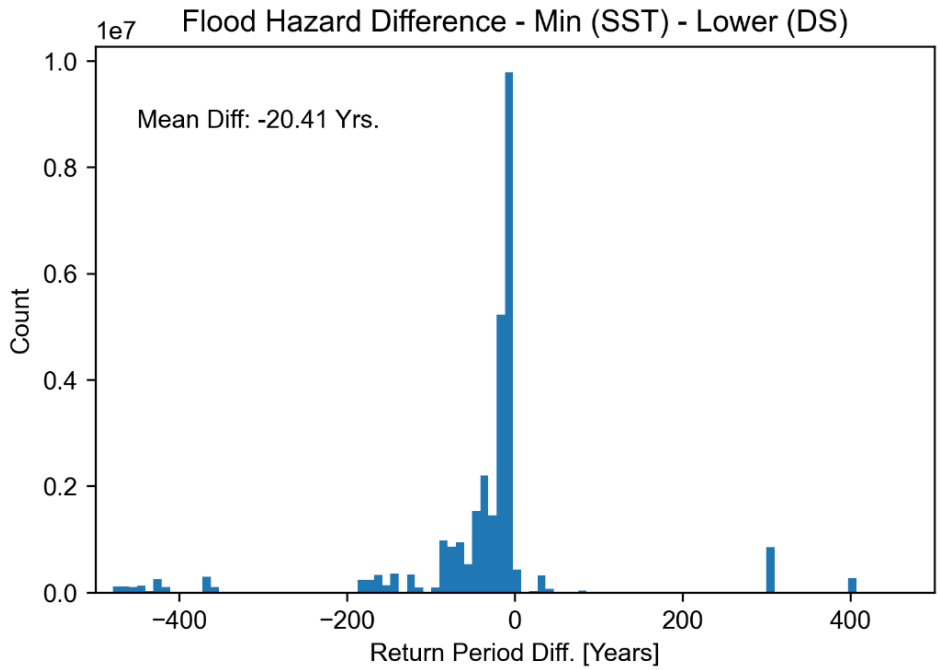
Figures S1-9



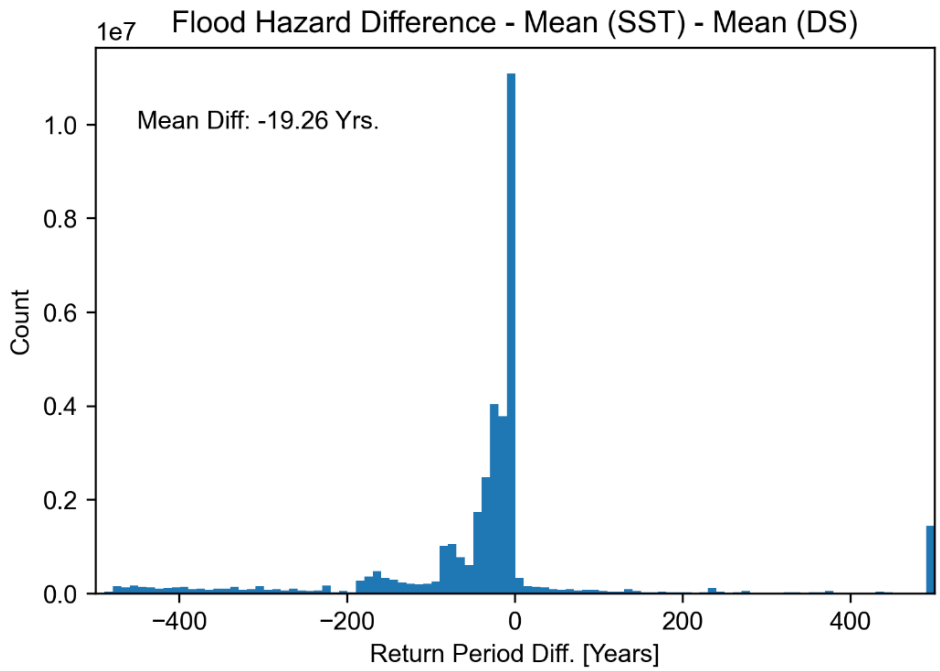
Supplementary Figure S1. A map of the study area and watershed encompassing New Bern, North Carolina. In the inset box, the red outline shows the transposition domain used for SST, while the blue shaded region shows the extent of the main map.



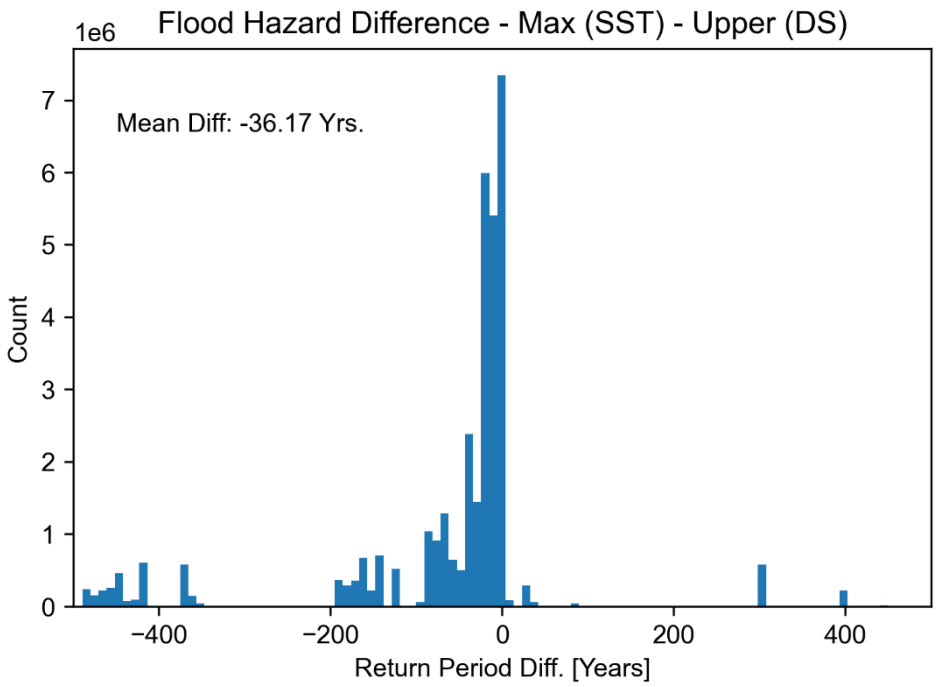
Supplementary Figure S2. A comparison of rainfall frequency curves produced by SST (blue) and Atlas 14 (black). The 90 percent confidence interval of the Atlas 14 IDF curve is shown as black error bars, while the ensemble spreads of the SST frequency curve is shown as a semi-transparent shaded region. The SST curve and ensemble spread closely match the Atlas 14 IDF curve for most return periods: treating the Atlas 14 IDF curve as the "true" value, we find a mean absolute error of 18.95 mm and an overall bias of -12.55 mm, suggesting that the SST rainfall volumes are similar to - but overall slightly less than – the design storm rainfall.



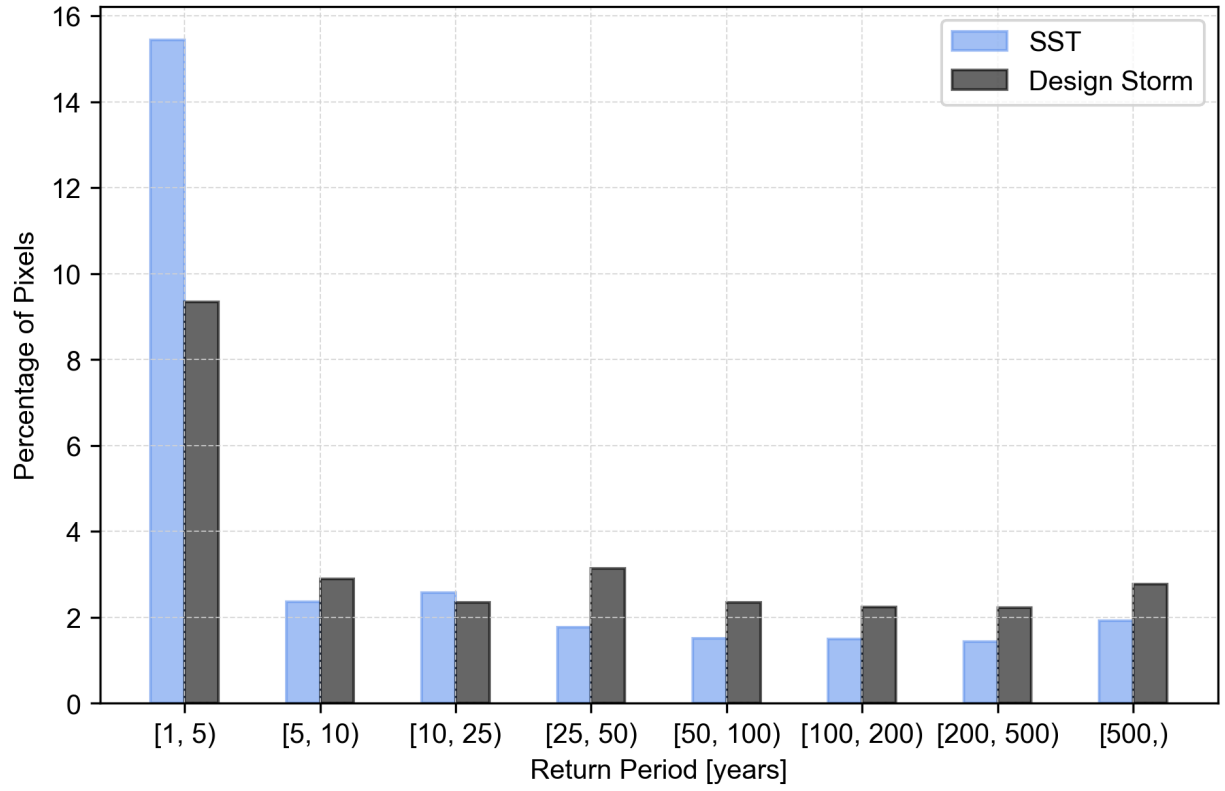
Supplementary Figure S3. A histogram of the differences in flood hazard between the five-meter hazard map produced by the SST ensemble minimum and the lower bound design storms. Negative values represent more frequent flooding from SST, while positive values represent more frequent flooding from the design storm approach.



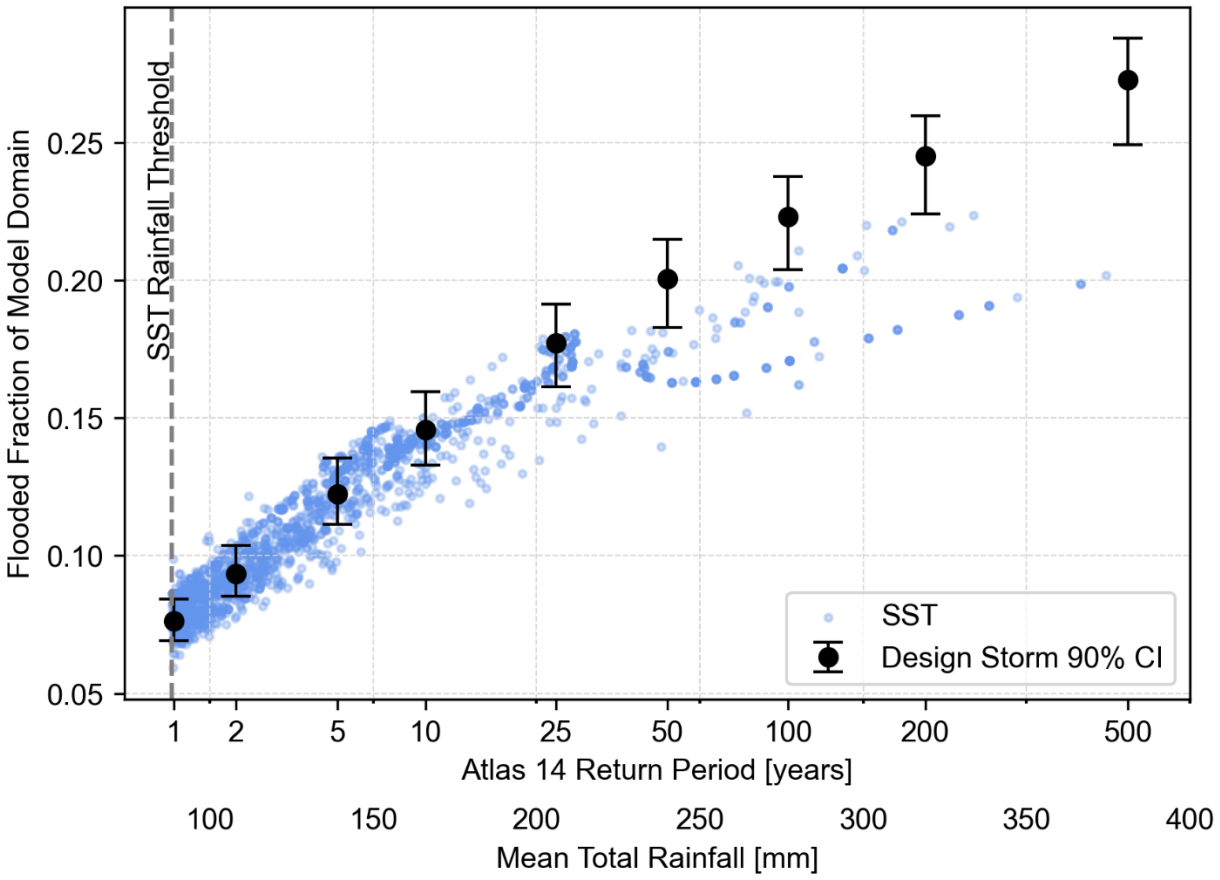
Supplementary Figure S4. A histogram of the differences in flood hazard between the five-meter hazard map produced by the SST ensemble mean and the mean design storms. Negative values represent more frequent flooding from SST, while positive values represent more frequent flooding from the design storm approach.



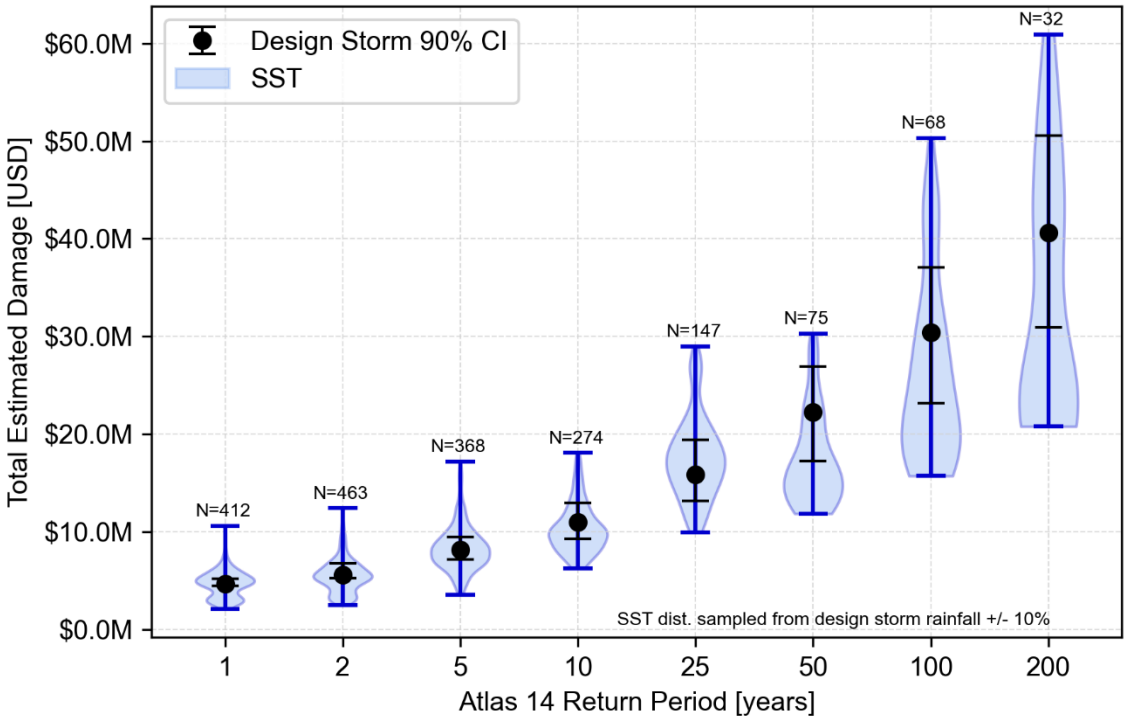
Supplementary Figure S5. A histogram of the differences in flood hazard between the five-meter hazard map produced by the SST ensemble maximum and the upper bound design storms. Negative values represent more frequent flooding from SST, while positive values represent more frequent flooding from the design storm approach.



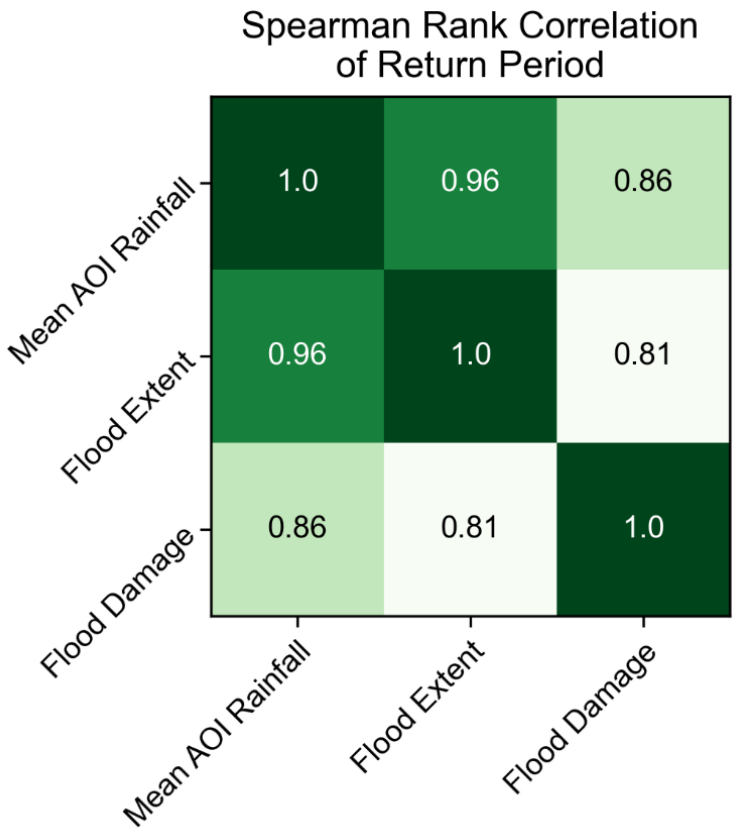
Supplementary Figure S6. Percentage of pixels in the mean downscaled flood hazard rasters that correspond to different return period bins (half-open intervals). Relative to the design storm approach, SST substantially increases the number of pixels that are in the five-year floodplain, and slightly increases the number of pixels in the 25-year floodplain.



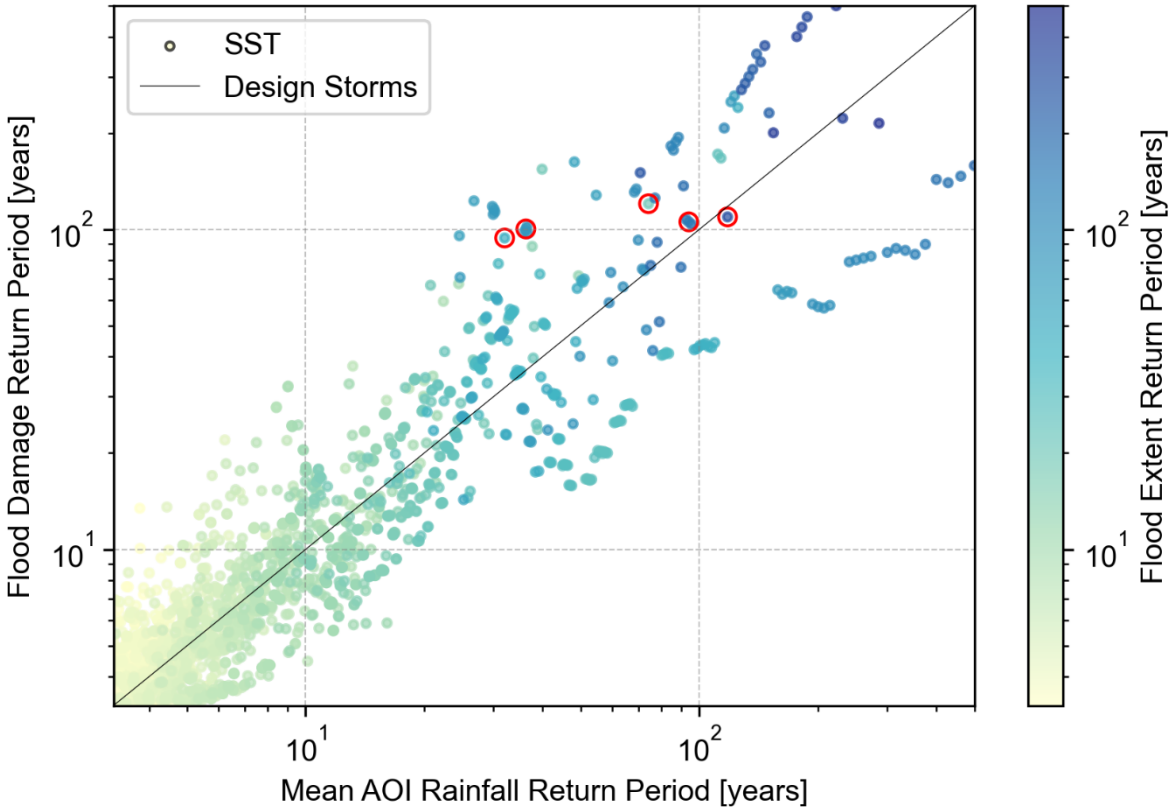
Supplementary Figure S7. Scatterplot of the relationship between mean total rainfall (in millimeters) and flood extent for SST storms and design storms. The 90 percent confidence interval of the design storms is represented as black error bars. The estimated frequency for each storm is also shown, based on IDF curves from the CCRA rain gauge. Generally, design storms produce floods with extents as large as or larger than the equivalent SST storm.



Supplementary Figure S8. Violin plot of the relationship between mean total rainfall and flood damages for SST and the design storm approach. The 90 percent confidence interval of the design storms is represented as black error bars, while the "N=" annotations at the top of each violin correspond to the number of SST storms included in that distribution. For each return period, the SST distributions were constructed by selecting all storms with mean total rainfall within 10 percent of the design storm of the same return period. Across all return periods, the distribution of risk estimates produced by SST have relatively thick tails that extend beyond the upper-bound design storm estimate.



Supplementary Figure S9. A map of the Spearman rank correlation for the return periods of each combination of rainfall, flood extent, and flood losses. More positive values and darker colors correspond to more positive correlation between two variables.



Supplementary Figure S10. A comparison of the frequency relationships between rainfall, flood extent (including permanent water bodies), and flood damage for model runs with SST storms. Design storms are represented by the black line, which shows the relationship that would result from assuming that these relationships are one-to-one. The points highlighted in red are shown on Figure 4 in the main text. While the relationships between storm, flood, and damage frequency are correlated, they are not one-to-one (as suggested by design storm approaches), even for small return period storms.

Tables S1-S2

Design Storm Return Period [Years]	Mean SST Return Period [Years]
1	1
2	1
5	2
10	3
25	7
50	20
100	69
200	185
500	389

Supplementary Table S1. A comparison of the hazard estimates produced by the SST ensemble mean and the mean design storms. For each return period, the comparison was performed by identifying all pixels in the design storm hazard raster equal to that return period, then calculating the mean return period of all corresponding pixels in the SST hazard raster. SST produces equal or smaller return periods for all design storm floodplains, particularly for areas with 10-year, 25-year, 50-year, and 100-year return periods.

Storm ID	Return Period [Years]		
	Basin-Averaged Rainfall	Flood Extent	Flood Damage
5-179-18	40	20	154
5-329-18	48	58	163
1-197-18	59	116	59
6-183-11	69	96	131
0-97-18	71	286	150
4-403-11	74	20	120
3-430-18	75	200	77
2-472-14	85	118	194
5-398-18	90	182	76
5-303-18	118	316	109
9-281-17	172	73	63
1-495-18	231	400	223
5-214-16	334	134	86

Supplementary Table S2. Examples of highly nonlinear frequency relationships of rainfall, flood extent, and flood damage for 13 SST storms. Storm ID numbers are constructed as follows: (Ensemble Member)-(Synthetic Year)-(Storm Arrival in Year).

Appendix A. Transposition Domain

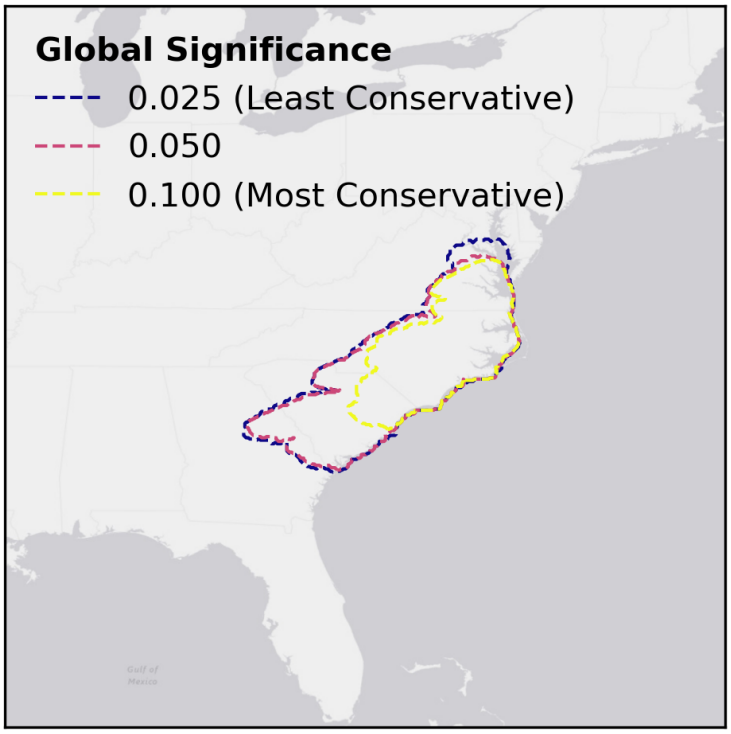


Figure A1. The three SLAM transposition domains considered. Each domain varies only in the GSL applied during application of the SLAM methodology, with a GSL of 0.025 representing the least conservative domain.

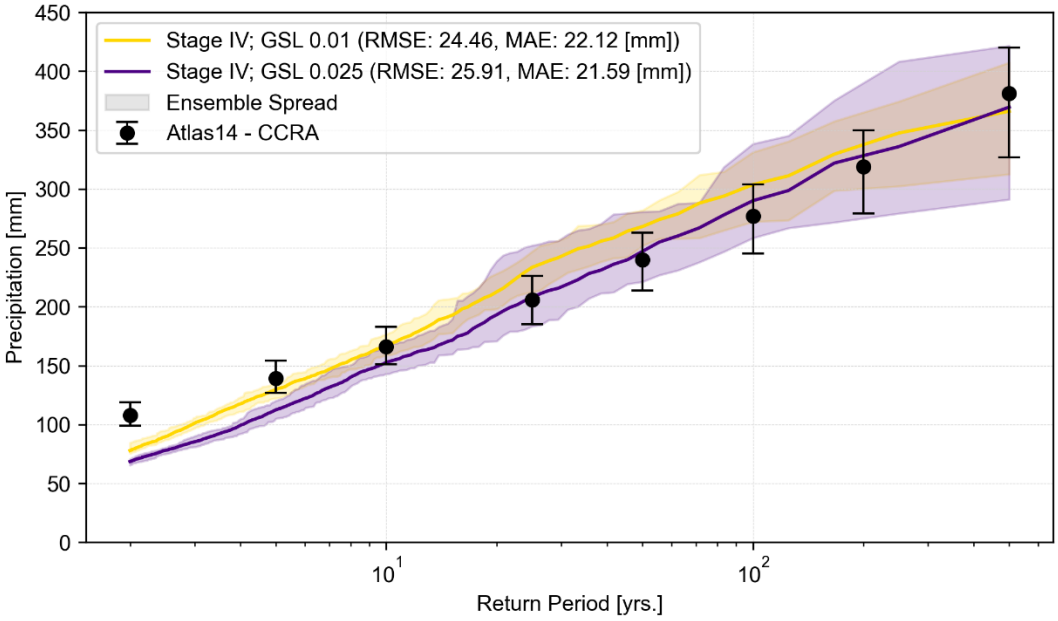


Figure A2. A comparison of the CCRA rainfall frequency curve from Atlas 14 against those produced by SST with the most conservative (GSL = 0.01) and least conservative (GSL = 0.025) transposition domains. The RMSEs and MAEs are shown for each curve relative to Atlas 14. Note that these rainfall frequency curves appear different than those in Fig. S2 because these were generated with RainyDay's "DURATIONCORRECTION" option, which enables more accurate rainfall frequency estimates (particularly for small return periods) but cannot be used in conjunction with hydrologic modeling.

Appendix B. Stochastic Storm Transposition

Tropical Cyclone Postprocessor

Coastal settings pose two major challenges for SST. First, rainfall patterns can change rapidly with increasing distance from the coast due to oceanographic effects [1,2], complicating the delineation of a transposition domain[3]. Second, tropical cyclone (TC) rainfall intensity declines as the storm makes landfall [4], such that transposing these storms randomly throughout the transposition domain – as is typically done in SST [5–7] – might result in unrealistically high rainfall distributions. To address these issues, we developed a postprocessor for RainyDay that can apply alternate transposition schemes based on user-defined storm types, a process we refer to as modified SST (M-SST).

To create the postprocessor, we modified RainyDay such that it saved relevant data from each transposed storm realization in a tabular format that could be read into our postprocessor. The postprocessor then identifies the storm type of each realization based on user-supplied classifications of the parent storm types. Parent storms can either be classified as "TC", in which case they are subject to cyclone-specific transposition methods, or "Other", in which case they are transposed exactly as they would have been by RainyDay. TCs are then transposed by randomly sampling a location from a user-specified coastline and shifting the track such that their landfall timing or minimum distance to land (if the parent storm did not make landfall) is preserved. The new track is validated within a user-specified tolerance by ensuring that points over land or over water in the observed track remain so in the transposed track. If the transposed track violates the validation criteria, the track is iteratively transposed until these criteria are satisfied. Other type-specific transposition schemes could theoretically be added in the future.

To apply the postprocessor to our study, we manually classified our parent storms as TCs using NOAA's Historical Hurricane Tracks database [8]. Landfall timing (or minimum distance to land) was calculated using storm tracks from the International Best Track Archive for Climate Stewardship (IBTrACS) database [9] matched to each parent storm. A simplified shapefile of the coastline, modified to remove estuaries and interpolated to add thousands of uniformly-spaced vertices along its length, was used to sample a random landfall location. For our synthetic storm ensemble, we selected a tolerance of zero, meaning that all points along the transposed track remained over land or over water as in the observed track (Fig. B1).

Validation of Storm Realizations

The M-SST storm realizations were validated by generating a rainfall frequency curve and comparing it against similar curves produced via traditional SST and Atlas 14 at the Coastal Carolina Regional Airport (CCRA) rain gauge (Fig. B2). M-SST produced a 24-hour rainfall frequency curve that more closely matched the Atlas 14 curve at CCRA than one produced by traditional SST, reducing the root-mean-square error (RMSE) from 31.03 to 24.61 mm and the mean absolute error (MAE) from 26.06 to 18.95 mm. Additionally, M-SST produced an ensemble spread similar to those of traditional SST and Atlas 14, suggesting it reasonably represents the distribution and frequency of extreme rainfall events in the study area. However, similar to traditional SST, M-SST underpredicted rainfall volumes for storms with return periods less than 10 years relative to Atlas 14. It also underpredicted rainfall volumes associated with the 500-year return period, potentially due to limitations induced by Stage IV's relatively short period of record, or by the tight tolerance we selected for TC transpositions (Fig. B3).

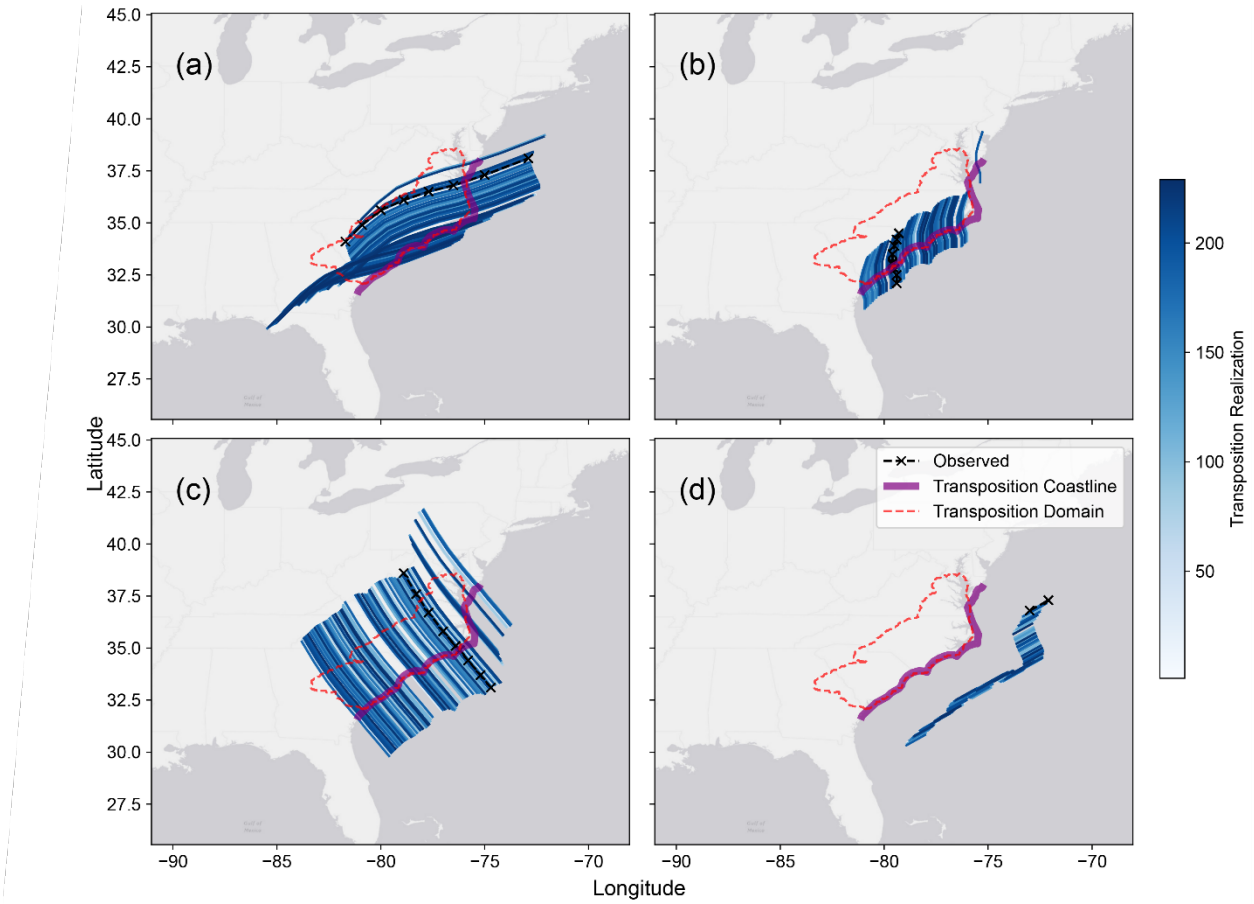


Figure B1. Examples of transposed TC tracks produced by M-SST for Hurricanes (a) Michael, (b) Gaston, (c) Isabel, and (d) Gustav. Transposed tracks remain over land and water at the same timesteps as their parent tracks, and only the portion of the TC track that falls within the 24-hour time frame of the parent storm is considered during validation. Any TC track can theoretically be transposed through this method, regardless of whether the storm approaches the transposition domain from the landward (a) or seaward (b, c) side or never makes landfall at all (d).

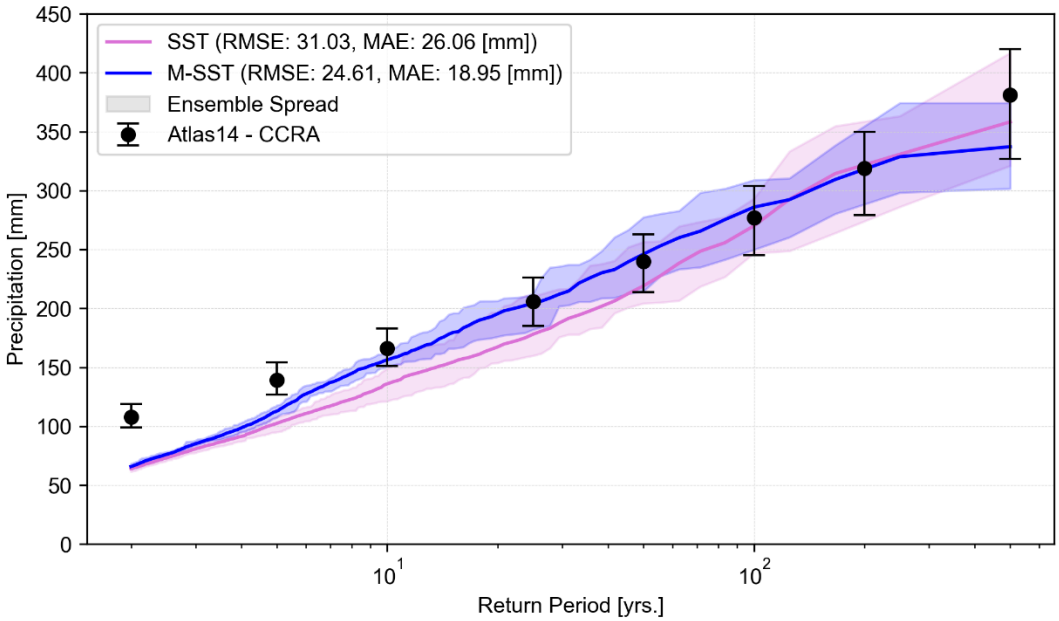


Figure B2. Comparison of the rainfall frequency curves produced by traditional SST (pink line), M-SST (blue line), and Atlas 14 (black dots). The 90 percent confidence interval of the Atlas 14 IDF curves are shown as black error bars, and the ensemble spreads of the SST and M-SST frequency curves are shown as semi-transparent shaded regions. The M-SST curve and ensemble spread more closely matches the Atlas 14 data for all except the 500-year return period, producing a lower average RMSE and MAE relative to the mean and upper and lower bounds of the Atlas 14 IDF curve than traditional SST.

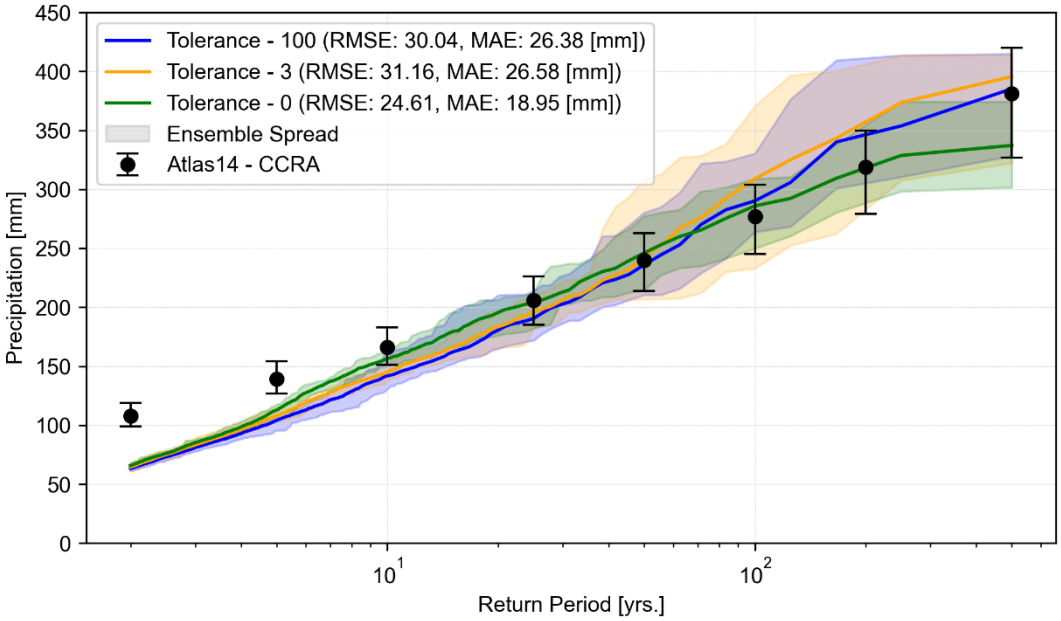


Figure B3. Comparison between the CCRA rainfall frequency curve from Atlas 14 and those produced by M-SST with a TC transposition tolerance of 0, 3, and 100. The RMSEs and MAEs are shown for each curve relative to Atlas 14.

Appendix C. Hydrodynamic Model

Development

The SFINCS model used in this study was adapted from an existing model of the Carolinas [10]. Channel bathymetry was adjusted along the Trent River, using a rectangular channel that preserved the effective bankfull cross-sectional area of the channel as measured in field cross-sections from the North Carolina Floodplain Mapping Program. The invert elevation of the Trent River was raised by 2 m upstream of United States Geological Survey gauge 02092554 based on analysis of simulated versus observed hydrographs. Typical ranges of Manning's n were defined for each land use type in the National Land Cover Dataset based on values from Arcement & Schneider (1989)[11], Chow et al. (1998)[12], and Savage et al. (2016)[13]. The initial model used the average of these ranges, but during calibration, Manning's n was increased to the upper end of these ranges (or slightly beyond) for undeveloped and natural land use categories, particularly for woody wetlands and small upland creeks, which helped better reflect the dense growth characteristics of the coastal swamp forests present in our model domain and resulted in improved model performance (Table C1).

Validation

The model was validated against five storm events of varying intensity and duration (Table C1), including Hurricanes Matthew and Florence. For Hurricane Matthew, we used wind field data from Oceanweather, Inc. (OWI) and for Hurricane Florence we use a modified OWI product described in Ratcliff (2022) [14]. For both TCs, we use time-series water level boundary conditions from an ADvanced CIRCulation (ADCIRC) model at the downstream boundary. For the other three validation events, wind stress was neglected, and time-series water level data was obtained from North Carolina Emergency Management (NCEM). For the 2009 and 2010 validation events, no time-series data was available, and so the historical average of the observed water levels at NCEM gages was used.

Validation metrics included bias, mean absolute error (MAE), and Nash-Sutcliffe (NSE) and Kling-Gupta (KGE) efficiencies between observed and modeled values of runoff volume, maximum water surface elevations, depths, and timing at locations where data was available. Simulated hydrographs for each validation event were compared against observed hydrographs from United States Geological Survey (USGS) gauges 02092554 and 02092500 located along the Trent River. For Hurricanes Matthew and Florence, hydrographs were also compared at USGS rapid deployment gauge 02092576 located at the confluence of the Trent and Neuse Rivers near downtown New Bern. Forty-six HWMs from the USGS were used to validate maximum water surface elevations from Florence, and six HWMs from NCEM were used to validate maximum water surface elevations from Matthew. No HWMs were available for the other three events.

SFINCS simulated hydrographs that matched observations for all five validation events (Fig. C1), resulting in an average KGE of 0.72 and an average peak elevation bias and RMSE of 0.38 and 1.14 meters, respectively (Fig. C2). The model predicted peak water levels at the upstream Trent River gauge (02092500) within half a meter, but overpredicted them at the downstream gauge (02092554) for all events except Hurricane Matthew. The model showed less bias in terms of total runoff volume, with a mean area-under-hydrograph error of just 4% (Table C2). These validation results suggest that hydrologic processes like precipitation, infiltration, and runoff are

captured by the model. We expect that large uncertainties in the peak water levels may be a result of incomplete river bathymetry information.

For Hurricanes Florence and Matthew, the model closely recreated observed HWMs in the densely developed areas around New Bern (Fig. C3), with an overall RMSE of 0.43 meters and a bias of just 0.02 meters (Fig. C4). These results represent a slight improvement in the domain-wide performance of the model from Grimley et al. (2025)[10] which covers eastern North and South Carolina and was validated for Hurricanes Florence and Matthew only. We validated the model for three additional events and made local refinements – including a finer grid resolution, refined channel bathymetry, and calibrated n – that improved the model skill in reproducing water levels. Overall, our model performs similarly to other hydrodynamic models used for hazard and risk modeling and demonstrates high skill for storms of varying magnitude and frequency (Table C3).

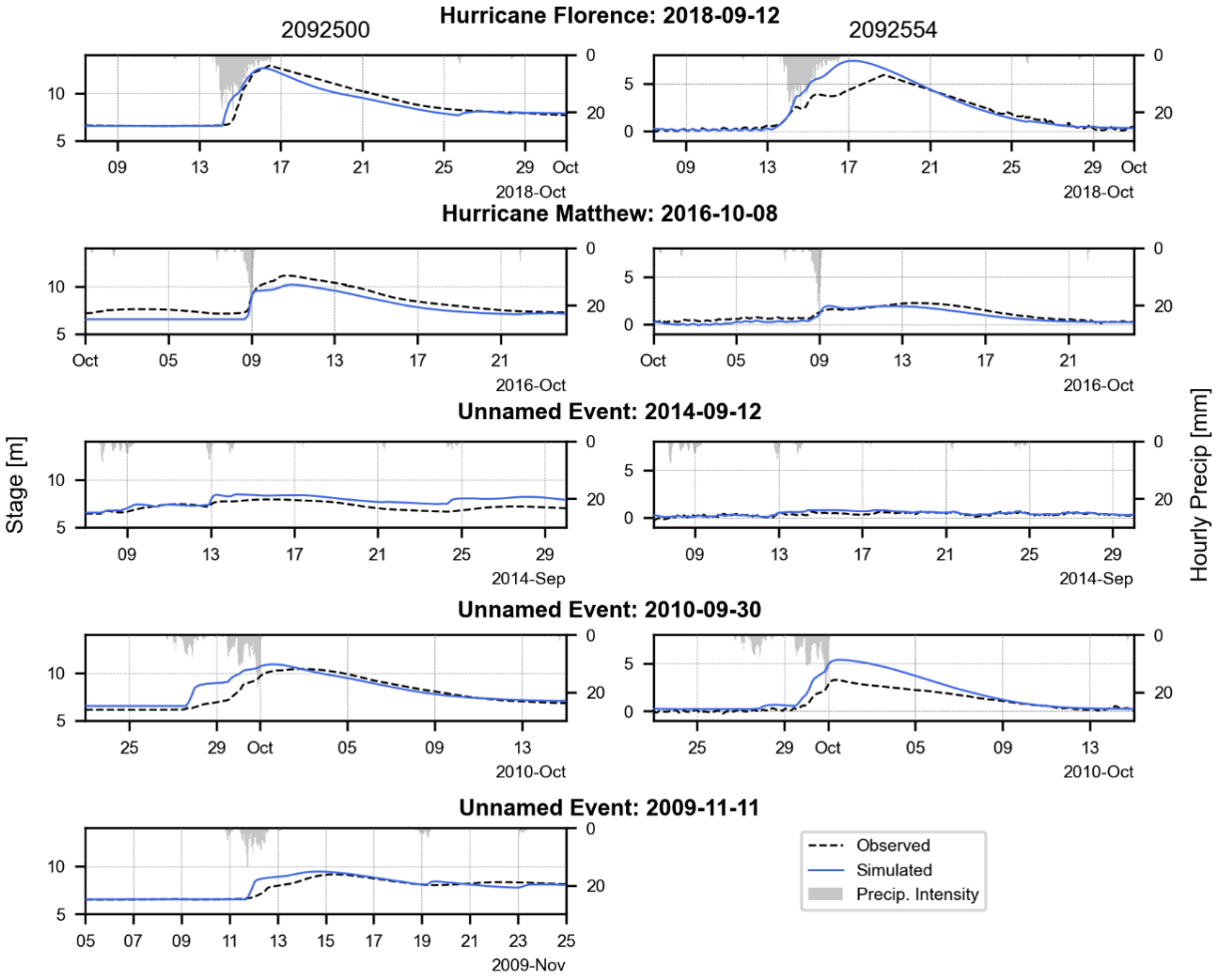


Figure C1. Comparison of observed hydrographs with those simulated by the hydrodynamic model for each of the five validation events. No data was available for gauge 2092554 during the 2009-11-11 event. Our model reproduces the shape of observed hydrographs – including peak height and timing - for all five storms, which vary in type and intensity.

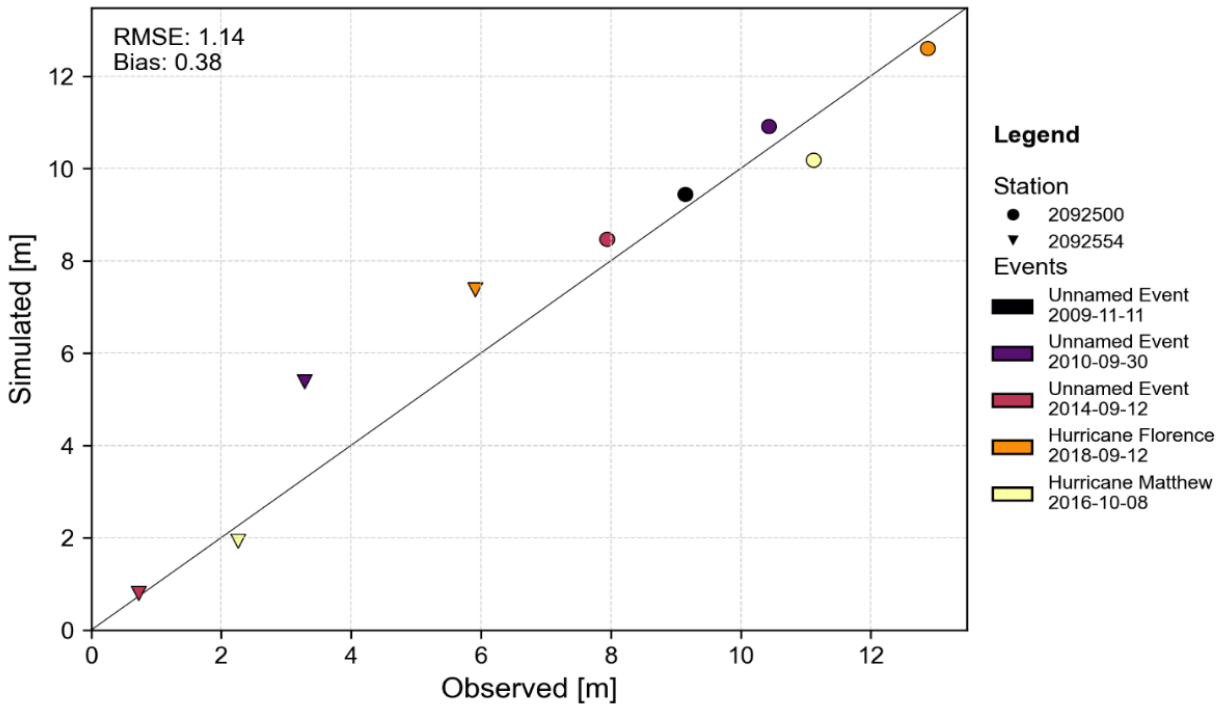


Figure C2. Scatterplot of simulated versus observed maximum water levels at USGS gauging stations. Our model can reproduce observed peak stage elevations but overpredict water levels at the downstream gauge (02092554) during the largest events.

5

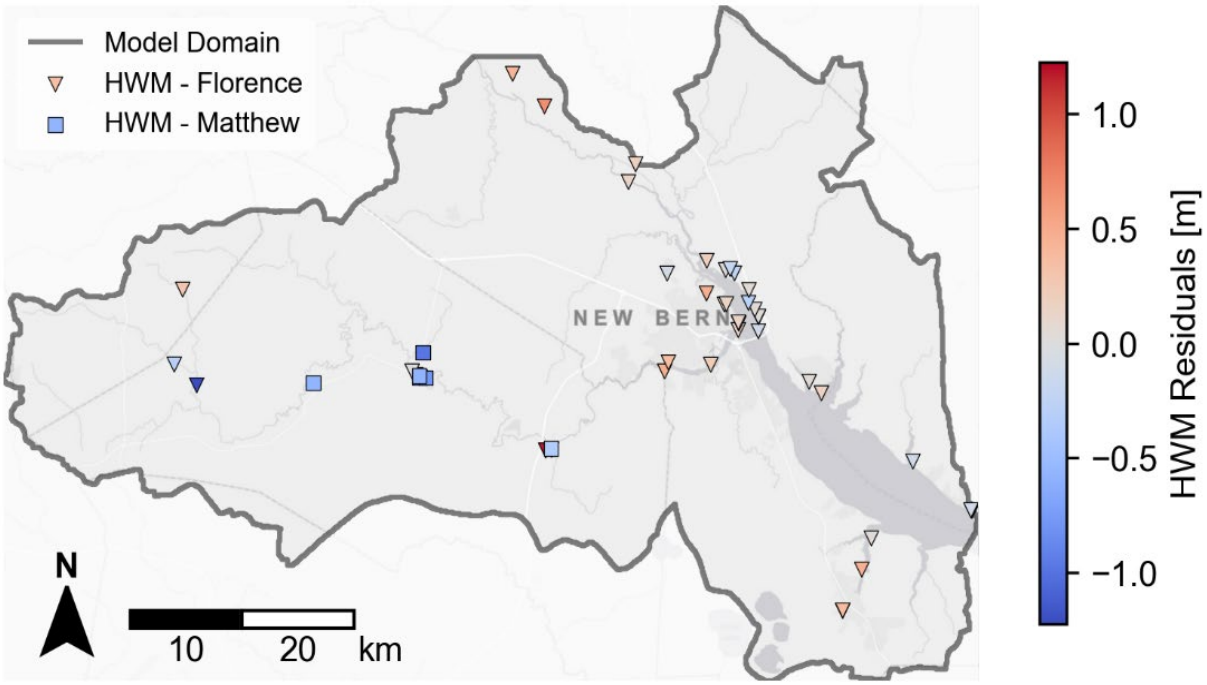


Figure C3. Map of model HWM error for Hurricanes Matthew and Florence. HWM error is calculated as the difference between the simulated maximum water elevation and the surveyed elevation of an HWM indicator, such as mud, seed, or foam lines, at the same coordinates. Locations where the model overpredicted maximum flood elevations are shown in red, while areas where the model underpredicted them are shown in blue. Our model tended to underpredict maximum flood elevations in the rural, western portion of the domain, but closely reproduced observed elevations in the developed areas near New Bern and Havelock.

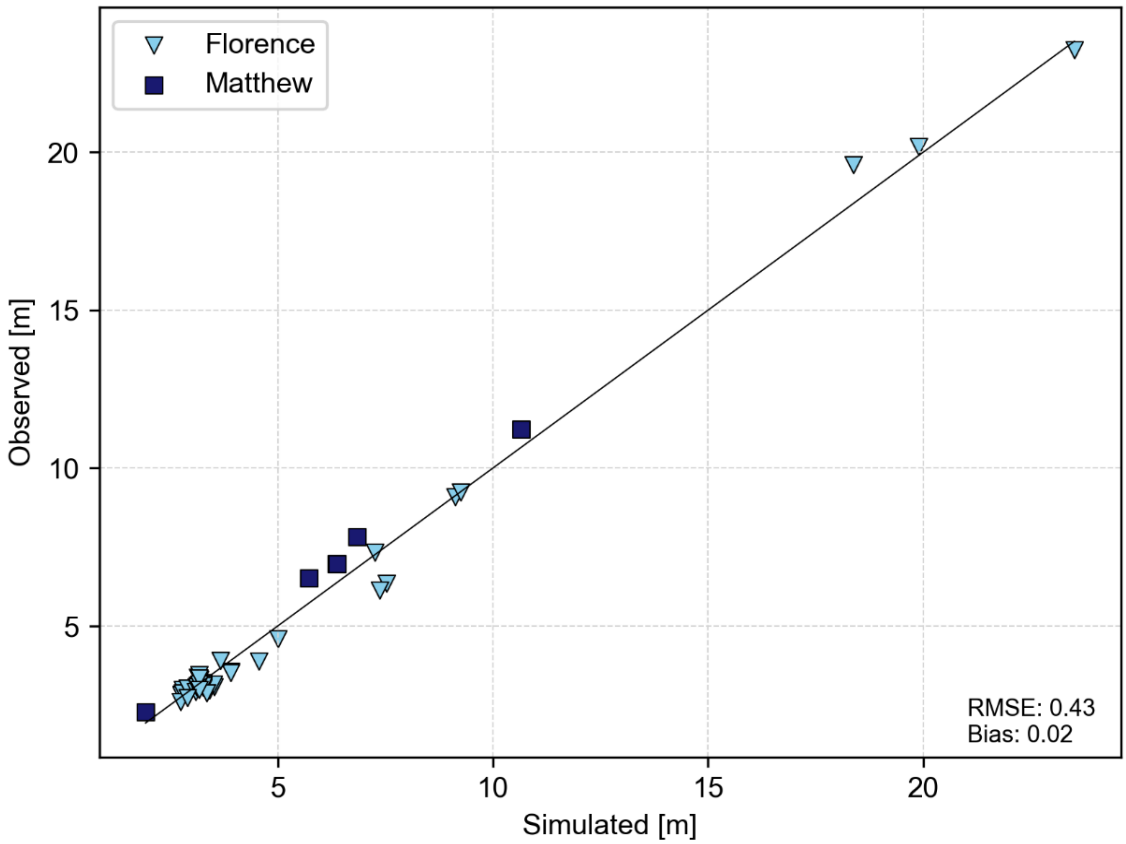


Figure C4. Scatterplot of simulated versus observed HWMs for Hurricanes Matthew and Florence. The black line represents perfect agreement between simulated and observed maximum flood elevations. Our model slightly underestimated HWMs for Matthew, but closely reproduced HWMs for Hurricane Florence, with a near-zero bias across these two events.

ID	NLCD Name	Ranges from Arcement & Schneider (1989), Chow et al. (1998), and Savage et al. (2016)			Final Model
		Low	Average (Initial Model)	High	
11	Open Water	0.02	0.023	0.025	0.025
21	Developed, Open	0.03	0.04	0.05	0.05
22	Developed, Low Intensity	0.08	0.1	0.12	0.12
23	Developed, Medium Intensity	0.06	0.1	0.14	0.15
24	Developed, High Intensity	0.12	0.16	0.2	0.2
31	Barren Land	0.023	0.027	0.03	0.03
41	Deciduous Forest	0.1	0.13	0.16	0.2
42	Evergreen Forest	0.1	0.13	0.16	0.18
43	Mixed Forest	0.1	0.13	0.16	0.2
52	Shrub/Scrub	0.07	0.115	0.16	0.16
71	Herbaceous	0.025	0.038	0.05	0.05
81	Hay/Pasture	0.025	0.038	0.05	0.05
82	Cultivated Crops	0.025	0.038	0.05	0.05
90	Woody Wetlands	0.07	0.115	0.16	0.35
95	Emergent Herbaceous Wetlands	0.07	0.115	0.16	0.085
--	Large Rivers	0.025	0.03	0.045	0.045
--	Smaller Rivers/Streams	0.035	0.045	0.055	0.1

Table C1. Comparison of the final Manning's roughness coefficients selected for each land use type against the typical ranges used in literature. The average of these ranges was used in the initial model, and were adjusted upward for the final model, particularly for small streams and woody wetlands, to reflect the characteristics of the coastal swamp forests in our study area.

Event Name	TC	Simulation Period	Total Rainfall [mm]	Max. 24-hr. Rainfall [mm]	Max. 72-hr. Rainfall [mm]	Return Period, Years (24-hr.)	Return Period, Years (72-hr.)
Hurricane Florence	Yes	9/7/2018 - 10/01/2018	565.7	335	515	269	974
Hurricane Matthew	Yes	10/01/2016 - 10/31/2016	210.5	151	180	7	7
Unnamed	No	9/07/2014 - 9/30/2014	175.6	48	74	<1	<1
Unnamed	No	9/23/2010 - 10/15/2010	407.0	213	302	29	67
Unnamed	No	11/5/2009 - 11/25/2009	213.2	140	173	5	5

Table C1. Observed storm events used to validate the hydrodynamic model. “Total rainfall” refers to the average total rainfall over the watershed during the simulation period. The maximum 24-/72-hour rainfall reflects the maximum amount of rainfall that fell in any 24-/72-hour window, averaged across the watershed. Rainfall return periods are listed relative to Atlas 14 IDF curves for the CCRA gauge and show that storms spanning a range of magnitudes and frequencies were used to validate our hydrodynamic model.

Event	Date	USGS Gauging Station	Area Under Hydrograph, Percent Error
Hurricane Florence	2018-09-12	02092500	-10%
		02092554	13%
Hurricane Matthew	2016-10-08	02092500	-6%
		02092554	-13%
Unnamed Event	2014-09-12	02092500	58%
		02092554	-34%
Unnamed Event	2010-09-30	02092500	-5%
		02092554	21%
Unnamed Event	2009-11-11	02092500	7%

Table C2. Hydrodynamic model area-under-hydrograph (AUH) error for the five validation events, where "area-under-hydrograph" reflects the total area under the hydrograph above the lowest stage level recorded during the simulation period, in arbitrary units. This metric serves as a proxy for runoff volume.

Event	Date	USGS Station	NSE	RMSE [m]	MAE [m]	KGE
Hurricane Florence	2018-09-12	02092500	0.91	0.55	0.36	0.89
		02092554	0.84	0.99	0.56	0.71
Hurricane Matthew	2016-10-08	02092500	0.65	0.72	0.67	0.90
		02092554	0.75	0.35	0.31	0.66
Unnamed Event	2014-09-12	02092500	-0.76	0.67	0.58	0.67
		02092554	0.33	0.17	0.13	0.63
Unnamed Event	2010-09-30	02092500	0.71	0.76	0.52	0.87
		02092554	0.66	1.04	0.70	0.42
Unnamed Event	2009-11-11	02092500	0.83	0.42	0.25	0.87

Table C3. Hydrodynamic model validation statistics for the five validation events. Statistics include the NSE, RMSE, MAE, and KGE. We find high KGE and NSE for our model, suggesting that the model can reproduce watershed responses to a variety of storm types and magnitudes.

Supplemental References

1. Curtis, S. Means and Long-Term Trends of Global Coastal Zone Precipitation. *Sci Rep* **9**, 5401 (2019).
2. Marra, F., Armon, M. & Morin, E. Coastal and orographic effects on extreme precipitation revealed by weather radar observations. *Hydrology and Earth System Sciences* **26**, 1439–1458 (2022).
3. Wright, D. B., Mantilla, R. & Peters-Lidard, C. D. A Remote Sensing-Based Tool for Assessing Rainfall-Driven Hazards. *Environ Model Softw* **90**, 34–54 (2017).
4. Zhai, G., Xu, W., Su, P., Qin, L. & Liao, X. Characteristics of precipitation changes during tropical cyclone processes in China from 1980 to 2019. *Sci Rep* **14**, 13654 (2024).
5. Wright, D. B., Smith, J. A., Villarini, G. & Baeck, M. L. Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. *Journal of Hydrology* **488**, 150–165 (2013).
6. Wright, D. B., Smith, J. A. & Baeck, M. L. Flood frequency analysis using radar rainfall fields and stochastic storm transposition. *Water Resources Research* **50**, 1592–1615 (2014).
7. Zhu, Z., Wright, D. B. & Yu, G. The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis. *Water Resources Research* **54**, 8983–8998 (2018).
8. National Oceanic and Atmospheric Administration. Historical Hurricane Tracks. https://coast.noaa.gov/hurricanes/ (2024).
9. Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. *Bull. Amer. Meteor. Soc.* **91**, 363–376 (2010).
10. Grimley, L. E. *et al.* Determining the Relative Contributions of Runoff, Coastal, and Compound Processes to Flood Exposure Across the Carolinas During Hurricane Florence. *Water Resources Research* **61**, e2023WR036727 (2025).
11. Arcement, G. J. & Schneider, V. R. *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains*. https://pubs.usgs.gov/publication/wsp2339 (1989) doi:10.3133/wsp2339.
12. Chow, V. T., Maidment, D. R. & Mays, L. W. *Applied Hydrology*. (McGraw-Hill Book Company, 1998).
13. Savage, J. T. S., Pianosi, F., Bates, P., Freer, J. & Wagener, T. Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research* **52**, 9146–9163 (2016).
14. Ratcliff, J. D. Analysis of Wind and Storm Surge of Hurricane Florence. (The University of North Carolina at Chapel Hill, United States -- North Carolina, 2022).