

Article

Probabilistic Storm Surge Estimation for Landfalling Hurricanes: Advancements in Computational Efficiency Using Quasi-Monte Carlo Techniques

Aikaterini P. Kyprioti ¹, Ehsan Adeli ¹, Alexandros A. Taflanidis ^{1,*} , Joannes J. Westerink ¹ and Hendrik L. Tolman ²

¹ Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN 46556, USA; akypriot@nd.edu (A.P.K.); eadeli@nd.edu (E.A.); jjw@nd.edu (J.J.W.)

² Office of Science and Technology Integration National Weather Service, NOAA/DOC, Silver Spring, MD 20910, USA; Hendrik.Tolman@NOAA.gov

* Correspondence: a.taflanidis@nd.edu



Citation: Kyprioti, A.P.; Adeli, E.; Taflanidis, A.A.; Westerink, J.J.; Tolman, H.L. Probabilistic Storm Surge Estimation for Landfalling Hurricanes: Advancements in Computational Efficiency Using Quasi-Monte Carlo Techniques. *J. Mar. Sci. Eng.* **2021**, *9*, 1322. <https://doi.org/10.3390/jmse9121322>

Academic Editor: Giuseppe Roberto Tomasicchio

Received: 27 October 2021

Accepted: 19 November 2021

Published: 23 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: During landfalling tropical storms, predictions of the expected storm surge are critical for guiding evacuation and emergency response/preparedness decisions, both at regional and national levels. Forecast errors related to storm track, intensity, and size impact these predictions and, thus, should be explicitly accounted for. The Probabilistic tropical storm Surge (P-Surge) model is the established approach from the National Weather Service (NWS) to achieve this objective. Historical forecast errors are utilized to specify probability distribution functions for different storm features, quantifying, ultimately, the uncertainty in the National Hurricane Center advisories. Surge statistics are estimated by using the predictions across a storm ensemble generated by sampling features from the aforementioned probability distribution functions. P-Surge relies, currently, on a full factorial sampling scheme to create this storm ensemble, combining representative values for each of the storm features. This work investigates an alternative formulation that can be viewed as a seamless extension to the current NHC framework, adopting a quasi-Monte Carlo (QMC) sampling implementation with ultimate goal to reduce the computational burden and provide surge predictions with the same degree of statistical reliability, while using a smaller number of sample storms. The definition of forecast errors adopted here directly follows published NWS practices, while different uncertainty levels are considered in the examined case studies, in order to offer a comprehensive validation. This validation, considering different historical storms, clearly demonstrates the advantages QMC can offer.

Keywords: landfalling storms; probabilistic storm surge estimation; forecast errors; quasi-Monte Carlo

1. Introduction

Prediction of storm-related impacts within prevention (pre-disaster), emergency management, and post-disaster settings has emerged as a key priority in natural hazard risk mitigation efforts [1,2]. The recent, increasingly active, and destructive hurricane seasons [3–6], as well as concerns related to the future effects of sea level rise, storm intensification, and increased hurricane occurrence rate projections on coastal areas [7–9], further stress the importance of research efforts that attempt to address this priority. Among the different impacts, storm surge—the wind-driven rise of water above astronomical tide levels during landfalling events—has been repeatedly proven to be an especially destructive force [10,11] and, therefore, its accurate estimation has been receiving increased attention. Relevant efforts towards this direction include the development of improved numerical models to predict surge for a given storm or an ensemble of storms [12–14], the establishment of online storm surge databases for easier access to relevant information for both researchers and the broader public [15,16], the creation of collaborative geospatial environments for

providing real-time surge estimates and hypothetical surge projections to planners and emergency managers [17], and the development of probabilistic tools to provide guidance during landfalling storms while explicitly considering all uncertainties related to the storm forecast [18]. The last is of particular importance and it is the focus of this paper. During landfalling events, predictions of the expected storm surge across the impacted regions are critical for guiding evacuation and emergency response/preparedness decisions, both at regional and national levels. At the same time, forecasts of the storm track, intensity, and size have an associated error [19] that impacts these surge predictions [20,21]. The uncertainties associated with the aforementioned forecast errors need to be quantified within a probabilistic setting [18] to support, ultimately, well-informed decisions [20].

To address this need, the National Weather Service (NWS) Meteorological Development Lab (MDL) created the Probabilistic tropical storm Surge (P-Surge) model [18,22]. P-Surge uses the National Hurricane Center's (NHC) official advisory for the storm current/forecasted features (such as track/intensity/size information), along with historical errors associated with the forecasts to provide probabilistic estimates for the anticipated storm surge. The historical errors are utilized to specify the probability distribution functions for four different storm features, quantifying, ultimately, the uncertainty in the NHC advisories [18]. These features correspond to (1) the cross-track variation, (2) the along-track variation (i.e., the storm forward translational speed), (3) the storm size, represented by the radius of maximum winds, and (4) the storm intensity, represented by the maximum velocity of sustained winds. For each geographic location of interest within the domain of the storm's impact, two different classes of statistical products are utilized to characterize the anticipated surge [18]: (1) the probability that the storm surge will exceed specific reference thresholds (for example, the probability that the surge will exceed 2 m) and (2) the storm surge value with specific probability of being exceeded (for example, the surge with 10% probability of being exceeded). These statistics are estimated by generating a storm ensemble that represents the aforementioned probability distribution functions, predicting the storm surge for each of the storms using the NWS's Sea, Lake, and Overland Surges from Hurricanes (SLOSH) model [12,14], and then appropriately combining SLOSH results over the storm ensemble. A full factorial sampling is currently utilized to create this storm ensemble [18]. For each of the four storm features a small number of representative values is considered, chosen based on the available intuition about the importance of the respective forecast errors. For example, for the storm intensity, strong, medium, and weak storms are considered, while for the storm size, large, medium, and small storms are examined. Each of these representative values is provided an associated probability mass that represents its relative likelihood based on the respective probability distribution functions. The storm ensemble is obtained by considering all possible combinations of these values (full factorial design), and the weight for each of the storms is obtained by the product of the probability masses for each of its respective features.

This paper investigates an alternative sampling for the storm ensemble, using a quasi-Monte Carlo (QMC) formulation [23,24], with ultimate goal to reduce the computational burden: Provide statistical estimates for the surge with the same degree of accuracy, while using a smaller number of sample storms. QMC utilizes deterministic, low-discrepancy sequences for numerical integration and has been proven highly efficient for the estimation of statistical measures like the ones considered in this application [23]. Its distinct advantage is that it promotes a space-filling sampling in the probabilistic assessment, which is shown to improve the estimation accuracy. The full factorial sampling on the other hand, relying on the combination of representative values to leverage the available physical intuition about the importance of the individual forecast errors, does not enjoy the same space-filling properties. Our objective here is to investigate the QMC sampling potential for improving the computational efficiency in probabilistic storm surge estimation during landfalling events. It is important to recognize that a variety of aspects impact the accuracy of the storm surge estimates in this context, like the quality of information for the spatio-temporal hurricane variation, the magnitude of the forecast errors, the predictive accuracy of the

numerical model utilized to calculate the expected surge, and the ways the interactions between tides, surge, and waves are addressed in that model. The focus of this paper is strictly on the uncertainty propagation component of this problem, i.e., the sample-based estimation of the storm surge statistics of interest, and not on the characterization of these uncertainties (forecast error description), or on the accuracy of the storm surge numerical predictions. All the discussions regarding the improvement of accuracy refer explicitly to this uncertainty propagation component. In fact, the accuracy should be interpreted in one of two equivalent ways: either as a reduction in the statistical errors for the same computational burden (same number of numerical simulations) or as an achievement of the same level of statistical accuracy with a reduced computational burden (smaller number of numerical simulations). The computational savings established in this context can promote a faster real-time forecasting process or can accommodate the reduction of statistical or even numerical errors. Such a reduction in the numerical errors can be established by leveraging the computational savings originating from a more efficient uncertainty propagation to accommodate the use of higher accuracy numerical models for the surge estimation.

For formulating the uncertainty propagation problem in our study, the definition of forecast errors follows directly the published NWS practices and different uncertainty levels are considered in the examined case studies to offer a comprehensive validation of the proposed methodology. For facilitating the storm surge predictions in these case studies, a surrogate model formulation is adopted, though it should be stressed that the proposed advances are independent of the selected storm surge numerical model. Surrogate models have been gaining increased popularity for providing accurate storm surge predictions in the past decade [25–28]. Here, they are preferred simply in order to facilitate the extensive validation studies and the fact that, as will be shown later, the reference solutions in these studies require very large storm ensembles to perform accurate comparisons. The use of a storm surge surrogate model is able to accommodate the described challenges for the validation case studies that are considered here. It should be stressed, though, that the authors are not suggesting there is a need to replace the currently used SLOSH model with a surrogate for the probabilistic NWS predictions. This is done only to facilitate the validation studies examined in this paper.

The remaining parts of the paper are organized as follows. In Section 2 the probabilistic storm surge estimation problem for landfalling events is reviewed, and in Section 3 the factorial sampling approach for the estimation of the surge statistics is discussed. Both these sections are based on the relevant published literature [18,22] but further establish some necessary mathematical formalism in the problem description to support the QMC advances. Section 4 introduces in detail the QMC implementation. Section 5 contains an extensive validation study, considering different historical storms and different advisories for each storm, as well as different definitions for the forecast errors. Section 6 offers the conclusions of the study.

2. Probabilistic Characterization of Surge for Landfalling Storms

2.1. Storm Characterization

During landfalling storms, the NHC provides advisories for the past and forecasted storm track, size, and intensity characteristics. Track is described by the latitude, s_{lat} , and longitude, s_{lon} , of the storm center, size by the radius of maximum winds, R_{mw} , and intensity by the maximum sustained wind speed, v_w , and/or the pressure difference between the center of the storm and the ambient pressure, DP . Examples of NHC advisories can be found at [16]. Following NHC standards, v_w is going to be taken herein to correspond to the 1-min sustained 10-m wind speed and will be typically presented in units of knots (kts). Part (a) of Figure 1 shows an example of the storm track (s_{lat}, s_{lon}), wind speed v_w , and pressure difference DP for advisory 18 of super storm Sandy, one of the historical storms that will be used in the case studies examined later in this paper. Both past and forecast information is shown. As also shown in part (a) of the figure, for future projections (forecasts), only the wind speed information is available for characterizing the storm

intensity in NHC advisories. Note that functional approximations exist [29] that relate R_{mw} , v_w , and DP using, additionally, the available storm track information. As such, one of these parameters can be determined if the other two characteristics (and the storm track) are known.

The NHC advisory information for the storm track, size, and intensity can be used by an appropriate parametric model to create wind and pressure fields that describe the storm evolution over time. These fields can be subsequently used as forcing for a storm surge numerical model to provide predictions for the anticipated surge. Since NHC forecasts do not provide information for the future evolution of R_{mw} , a common practice is to use the current estimate of R_{mw} as the future forecast for the storm size [18]. If R_{mw} is not readily available, then current information for v_w and DP can be leveraged to infer this value based on the aforementioned functional approximations that relate these three storm characteristics. This is the approach that is adopted in the case studies considered later in this paper.

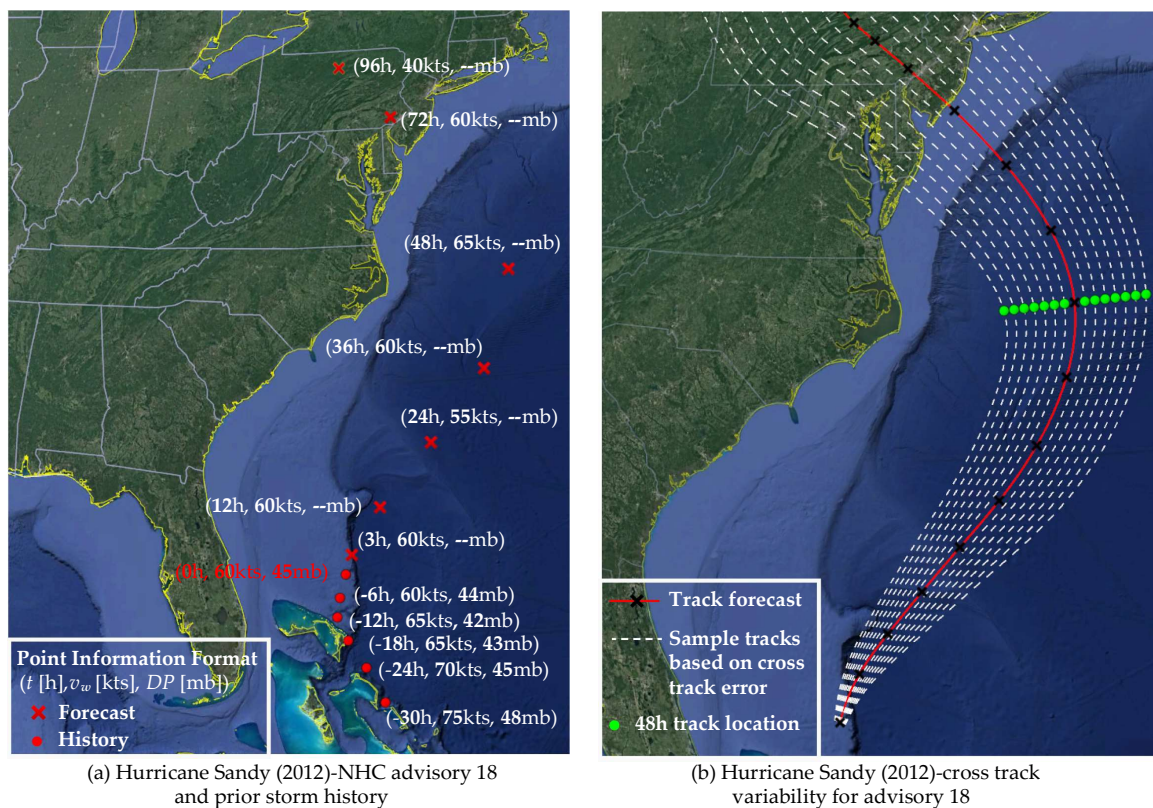


Figure 1. (a) NHC advisory 18 for super storm Sandy, combining past (indicated with $t < 0$) and forecast (indicated with $t > 0$) information for track and intensity (sustained wind speed v_w and pressure difference DP). (b) Storm tracks obtained considering the cross-track variability of the nominal track forecast shown in part (a). Green circles correspond to the storm track forecast at 48 h from the current position and represent the samples for the cross-track error for the current sampling process implemented in P-Surge.

To establish a mathematical formalism, we will denote by \mathbf{x} the four-dimensional vector of input parameters for characterizing each storm, referenced herein also as storm features:

$$\mathbf{x} = \begin{bmatrix} R_{mw} \\ s_{lat} \\ s_{lon} \\ v_w \end{bmatrix} \quad (1)$$

The temporal variation of these characteristics will be described using notation $\mathbf{x}(t)$, where t corresponds to time, with $t < 0$ characterizing the prior history of the storm (historic track) and $t > 0$ the future forecast (advisory track). The NHC forecast provides the nominal (median) predictions for the storm features, $\{\tilde{\mathbf{x}}(t); t > 0\}$, with tilde \sim over \mathbf{x} used herein to denote the nominal storm characteristics. The NHC advisory, originally is provided for variable time intervals, but it can be easily interpolated into hourly intervals. Such an interpolation is independently performed for each of the four storm features.

2.2. Forecast Errors and Uncertainty Description

As discussed in the introduction, NHC forecasts are associated with certain errors. These errors can be described as variations in the size, the intensity, and the along and cross position of the storm center relative to the nominal track [18]. The along-track variability impacts the translational velocity of the storm, while the cross-track variability impacts the position of the storm track itself, including storm bearing. Part (b) of Figure 1 shows the different storm tracks that originate from the cross-track variability of the nominal track advisory shown in part (a) of Figure 1. The different storm tracks shown in this figure are obtained through the sampling process detailed later in Section 3. Let $\Delta R_{mw}(t)$ and $\Delta v_w(t)$ denote, respectively, the variability in the size and intensity of the storm parameters, compared to their nominal values, so that $R_{mw}(t) = \tilde{R}_{mw}(t) + \Delta R_{mw}(t)$ and $v_w(t) = \tilde{v}_w(t) + \Delta v_w(t)$. Note that, based on the $R_{mw}(t)$ and $v_w(t)$ values, the $DP(t)$ value can be obtained as discussed in the previous section; therefore, if that storm characteristic is required by the numerical model used to predict the storm surge, as is typically the case, then it can be inferred based on the values of $R_{mw}(t)$ and $v_w(t)$ [18]. Let also $\Delta s_{along}(t)$ and $\Delta s_{cross}(t)$ denote, respectively, the along- and cross-track variability of the storm. For the track variability, both $\Delta s_{along}(t)$ and $\Delta s_{cross}(t)$ jointly influence the updated storm track coordinates $s_{lat}(t)$ and $s_{lon}(t)$, determining, respectively the variation of the storm perpendicular to the nominal track and the variation along this updated track, impacting, ultimately, the translational storm speed. The process of getting the updated track based on the nominal storm track $\{\tilde{s}_{lat}(t), \tilde{s}_{lon}(t)\}$ and the $\Delta s_{along}(t)$ and $\Delta s_{cross}(t)$ variation is the following. First, each point of the nominal track is varied by $\Delta s_{cross}(t)$ perpendicular to its bearing (at each t) to obtain a cross-path modified track. This process ultimately alters also the bearing of the storm track. Then, each point of that modified track is varied along the track by $\Delta s_{along}(t)$ to obtain the final track. The latter variation maintains the bearing of the cross-modified track.

The vector characterizing the variability in the storm features is finally defined as:

$$\Delta \mathbf{x}(t) = \begin{bmatrix} \Delta R_{mw}(t) \\ \Delta s_{cross}(t) \\ \Delta s_{along}(t) \\ \Delta v_w(t) \end{bmatrix} \quad (2)$$

Each of the four components of this vector will be denoted using subscript i , Δx_i , with $i = 1, \dots, 4$ corresponding, respectively, to characteristics related to ΔR_{mw} , Δs_{cross} , Δs_{along} , and Δv_w . The forecast errors for the NHC advisory are typically defined for 12-h intervals [0 12 ... 120] h and are used to characterize the probability distribution for $\Delta \mathbf{x}(t)$ [18]. The errors can be interpolated into hourly intervals to facilitate an easier combination with the interpolated nominal track forecast. For the probability distribution of $\Delta \mathbf{x}(t)$, independence is assumed between the four storm features, since they represent fundamentally different storm characteristics, while for each storm feature, perfect correlation between different times is assumed. This perfect correlation means that if, for example, the $v_w(t)$ value is larger by 1 standard deviation than its nominal value at $t = 24$ h, it will still be larger by 1 standard deviation than its median value at $t = 48$ h. Of course, the latter standard deviation, describing the magnitude of the variability, will change for different t and it will

become larger (larger forecast error) as t increases. This ultimately leads to the following probabilistic description:

$$\Delta x_i(t) = \sigma_i(t)\varepsilon_i; i = 1, \dots, 4 \quad (3)$$

where ε_i are independent random variables quantifying the uncertainty in the NHC forecast, and $\sigma_i(t)$ are scaling parameters that dictate the size of the forecast errors at different times. The vector with components ε_i will be denoted herein as ε .

For $\Delta s_{along}(t)$, $\Delta s_{cross}(t)$, and $\Delta v_w(t)$ a Gaussian probability distribution is assumed for the forecast errors [18]. This means that in the representation of Equation (3) ε_i are independent, identically distributed standard Gaussian variables and $\sigma_i(t)$ are the standard deviations associated with the forecast prediction errors. The latter are selected proportionally to the forecast error statistics:

$$\sigma_i(t) = e_i(t)/a; i = 2, \dots, 4 \quad (4)$$

where $e_i(t)$ is the 5-year mean absolute error statistics for forecasts, which started with at least hurricane strength, and a is scaling parameter that relates the standard deviation of the normal distribution to the absolute error, taken currently as 0.7979 [22]. Parts (a)–(c) of Figure 2 show examples for $e_i(t)$ based on the 2005, 2011, 2012, and 2020 NHC forecast errors for hurricanes. These years correspond to the events that will be utilized later in the case studies. It is evident that as t increases, the errors associated with the NHC forecast become larger and that the accuracy of the forecasts has improved over the years since the average errors in 2020 are substantially smaller than the average errors of 2005.

For ΔR_{mw} , there is no underlying assumed probability distribution; it is treated as a discrete random variable with three possible values representing small-, medium-, and large-sized storms, taken to correspond to the 15th, 50th, and 85th percentiles of the storm size error [18]. As also shown in part (d) of Figure 2, based on the nominal storm size $\tilde{R}_{mw}(t)$, five different error distributions are distinguished. These ultimately define possible values for the storm size, along with the associated probability masses (probability weights) as:

$$\begin{aligned} &\{\Delta R_{mw}^{(k_1)}(t) | \tilde{R}_{mw}(t); k_1 = -1, 0, 1\} \\ &P_1^{(k_1)} = 0.4 \text{ if } k_1 = 0 \\ &P_1^{(k_1)} = 0.3 \text{ if } k_1 = -1, 1 \end{aligned} \quad (5)$$

where $k_1 = -1$ corresponds to the 15th percentile sample, $k_1 = 0$ to the 50th percentile sample, and $k_1 = 1$ to the 85th percentile sample, while “|” represents the conditional value; samples are denoted herein with superscripts in parenthesis. This means that in the representation of Equation (3) the product $\sigma_i(t)\varepsilon_i$ may be taken to correspond directly to $\Delta R_{mw}^{(k_1)}(t)$, and that for the different examined t values, the same percentile index k_1 is always used. To unify the representation across the different storm features, notation ε_1 will still be used to describe the samples for R_{mw} , in this case representing which of the three different percentiles to adopt across the different t values. The ordering utilized in this study for the ΔR_{mw} samples, as also shown in part (d) of Figure 2, is such that the $k_1 = -1$ leads to the smallest-sized storms.

This uncertainty description leads ultimately to the probability distribution $p(\varepsilon)$ for ε , corresponding to a standard Gaussian probability density function for $\{\varepsilon_i; i = 2, 3, 4\}$ and to the probability masses shown in Equation (5) for ε_1 . It is important to acknowledge that this uncertainty characterization related to the forecast errors evolves over time. The information presented here, and what will be also presented in Section 3, is based on version 2.7 of P-Surge and on the published literature around it [18,22]. Though such an evolution over time will change the $p(\varepsilon)$ definition, the computational approaches for the efficient propagation of the uncertainty stemming from $p(\varepsilon)$, which is where the emphasis of this paper is placed, are not impacted. Only if there is a modification in the assumption of perfect correlation of the forecast errors between different time instances there will be an impact, since that can drastically modify the dimensionality of vector ε .

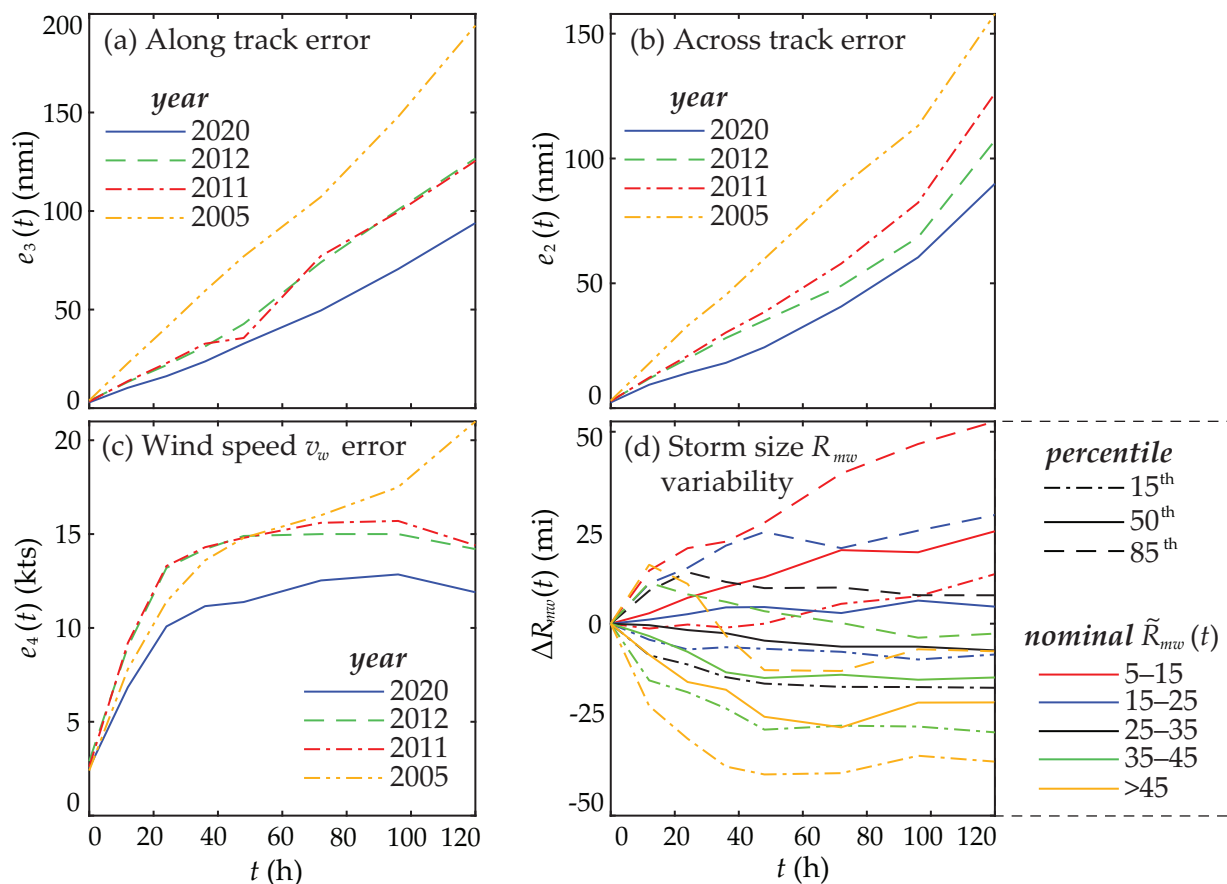


Figure 2. Uncertainty description as a function of time for the storm characteristics: (a–c) 5-year mean absolute error for the variation of the along-track, cross-track, and the wind speed, respectively, for four different error descriptions [2005, 2011, 2012, and 2020 updates of NHC forecast errors]; (d) the 15th, 50th, and 85th percentiles for the radius of maximum winds for different nominal radius values.

2.3. Probabilistic Surge Estimation

Let z_n denote the peak storm surge at location n of the geographic domain of interest. This surge can be estimated through any appropriate numerical model utilizing the storm characteristics $\mathbf{x}(t)$. For a storm characterized through deviation ϵ from the nominal storm forecast $\tilde{\mathbf{x}}$, notation $z_n(\epsilon | \tilde{\mathbf{x}})$ will be used herein for the surge description. For explicitly considering the NHC forecast errors in the predictions, different statistics of the storm surge can be estimated. Within the mathematical formalism established in Section 2.2, this is equivalent to the propagation of the uncertainty stemming from $p(\epsilon)$. Statistics of interest include the probability that the surge will exceed different thresholds of interest, or the threshold with specific probability of being exceeded [20,22]. Using the total probability theorem, the probability that the surge will exceed a specific threshold b is given by:

$$P_n(b) = P(z_n > b) = \int P(z_n > b | \epsilon) \cdot p(\epsilon) d\epsilon = \int I[z_n(\epsilon | \tilde{\mathbf{x}}) > b] \cdot p(\epsilon) d\epsilon \quad (6)$$

where, according to the notation established for the surge, $z_n(\epsilon | \tilde{\mathbf{x}})$, $P(z_n > b | \epsilon)$ is equal to $I[z_n(\epsilon | \tilde{\mathbf{x}}) > b]$, with $I[\cdot]$ corresponding to the indicator function, which is 1 if the quantity inside the brackets is satisfied, else it is 0. Note that the integral over $p(\epsilon_1)$ corresponds ultimately to a summation over its three discrete values. The threshold corresponding to exceedance probability p is given by:

$$b_n^p \text{ such that } P(z_n > b_n^p) = p \quad (7)$$

Current statistical products provided by the NWS focus on probability values of $p = 50\%$ (median) and $p = 10\%$ and surge threshold values of $b = 1.52$ m (5 ft) and 3.05 m (10 ft).

The estimation of $P_n(b)$ or b_n^p requires the calculation of the four-dimensional probabilistic integral of Equation (6). One approach to perform this calculation is to use Monte Carlo integration [30]. This approach leads to estimates with randomized characteristics, meaning that the result changes each time the estimation is performed, and to an accuracy that reduces as a function of $1/\sqrt{N}$ where N is the total number of simulations (samples) used, representing the computational burden necessary for the calculation of the desired statistics. Alternatively, deterministic approaches such as the use of factorial sampling, as currently performed in P-Surge [22], or the use of QMC sampling can be employed. These alternatives are discussed in the next two sections.

3. Estimation of Surge Statistics Using Full Factorial Sampling

One approach for estimating the surge statistics is to use full factorial sampling for the four storm features, combining discretized representative samples, along with probability weights for each of them. In the current P-Surge implementation [18,22], the discretization of the probability distribution is different for ΔR_{mw} , Δv_w , and Δs_{along} , adopting a discretization that results in roughly equal probability masses for each of the samples, as shown in part (a) of Figure 3 for Δv_w , and different for Δs_{cross} , adopting a discretization based on equally spaced samples, as shown in part (b) of Figure 3. For Δs_{cross} , the physical description of the samples (equal distance) is given a larger priority than the probabilistic description (equal probability masses), since the cross-track variability is expected to have a significant impact on the storm surge [18]. The discretization process for each of the storm features is detailed next.

3.1. Sampling for Each of the Storm Features

For Δv_w , $n_w = 3$ values are utilized for the corresponding random variable, ε_4 , representing the 15th ($k_4 = -1$), 50th ($k_4 = 0$), and 85th ($k_4 = 1$) percentiles, or weak, medium, and strong intensity storms, respectively. As shown in part (a) of Figure 3, this equivalently corresponds to a partitioning of the probability space to 30%–40%–30% subsets and to the selection of the point at the middle of the subset as the representative sample. This leads to pair of samples and respective probability weights $\{\varepsilon_4^{(k_4)}, P_4^{(k_4)}; k_4 = -1, \dots, 1\}$, given by:

$$\begin{aligned} \{\varepsilon_4^{(\pm 1)} = \pm \Phi^{-1}(0.15) = \pm 1.036, P_4^{(\pm 1)} = 0.3\} \\ \{\varepsilon_4^{(0)} = 0, P_4^{(0)} = 0.4\} \end{aligned} \quad (8)$$

with $\Phi(\cdot)$ denoting the standard Gaussian cumulative distribution function. For R_{mw} , $n_{mw} = 3$ values are utilized, as described by Equation (5), representing small-, medium-, and large-sized storms, respectively. Note that this discretization is identical to the one used for Δv_w , with the only difference that the underlying probability distribution function is not Gaussian. Similarly, for Δs_{along} a total of n_a partitions are utilized with the suggested values for n_a ranging between 3 [18] to 7 [22]. The increase in the value of n_a originated from the incorporation of tides in P-Surge and the need to better capture the coupling between the generated storm surge and the tidal variations when calculating the resultant water elevation. The along-track variation influences, as discussed earlier, the storm forward speed, which impacts the temporal variation of the storm surge and ultimately affects the total water elevation when it is combined with the evolving tides, even if the part of the storm surge itself would not be affected by such along-track variations. In the case studies that will be presented later, $n_a = 3$ was used, adopting earlier recommendations [18], since, as will be discussed later, the impact of tides will be ignored here. This ultimately leads to a definition similar to Equation (8), with samples representing the 15th ($k_3 = -1$), 50th

($k_3 = 0$), and 85th ($k_3 = 1$) percentiles. The pair of samples and respective probability weights in this case corresponds to $\{\varepsilon_3^{(k_3)}, P_3^{(k_3)}; k_3 = -1, \dots, 1\}$, where:

$$\begin{aligned} \{\varepsilon_3^{(\pm 1)} = \pm \Phi^{-1}(0.15) = \pm 1.036, P_3^{(\pm 1)} = 0.3\} \\ \{\varepsilon_3^{(0)} = 0, P_3^{(0)} = 0.4\} \end{aligned} \quad (9)$$

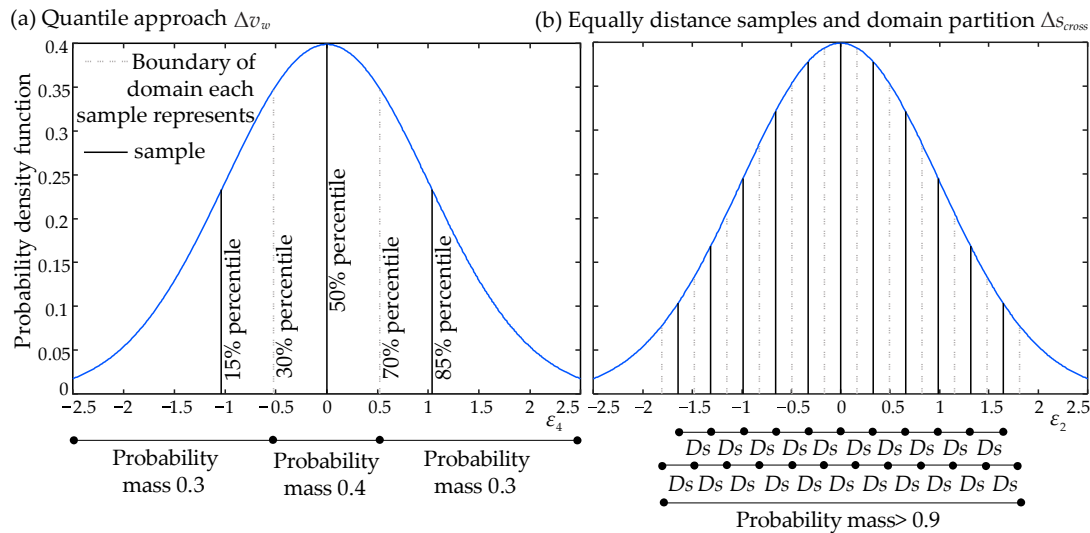


Figure 3. Illustration of the sampling process from a standard normal distribution using: (a) probability mass approach (shown for v_w) or (b) equally spaced samples and domain partition for Δs_{cross} .

For Δs_{cross} , the number of samples n_c is not pre-determined, but it is decided based on the $R_{mw}(t)$ value for $t = 48$ h [18]. The samples cover at least 90% of the normal distribution and they are sampled in such a way that the distance between storm tracks at the 48-h forecast is equal to the 48-h forecast for R_{mw} . The sampling for Δs_{cross} is illustrated in part (b) of Figure 3, while the resultant storm tracks have been presented earlier, in part (b) of Figure 1. The sampling can be conditional on the actual R_{mw} sample, so different for each of the three $\Delta R_{mw}^{(k_1)}(t)$, or on the smallest-sized storm, so uniform for all R_{mw} samples and based on the smallest $\Delta R_{mw}^{(k_1)}(t)$ sample. The step for the Δs_{cross} samples is:

$$Ds = (\tilde{R}_{mw}(48) + \Delta R_{mw}^{(1)}(48)) / \sigma_2(48) \quad (10)$$

where $\sigma_2(48)$ represents, as discussed earlier, the standard deviation of the forecast prediction error for the cross-track at $t = 48$ h. To cover the desired 90% of the standard normal distribution with a step of Ds , the total number of samples considered is $n_c = (2[1.65/Ds]^{ceil} + 1)$, where $[]^{ceil}$ is the ceiling function and 1.65 is the domain between the 5% and 50% percentiles (or between 50% to 95% percentiles) of the standard Gaussian distribution. The pair of samples and respective probability weights are $\{\varepsilon_2^{(k_2)}, P_2^{(k_2)}; k_2 = -(n_c - 1)/2, \dots, (n_c - 1)/2\}$, where:

$$\{\varepsilon_2^{(\pm k_2)} = \pm k_2 Ds, P_2^{(\pm k_2)} = \Phi(0.5 + k_2 Ds + Ds/2) - \Phi(0.5 + k_2 Ds - Ds/2)\} \quad (11)$$

$$P_2^{(k_2)} = \frac{P_2^{(k_2)}}{\sum_{k=-(n_c-1)/2}^{(n_c-1)/2} P_2^{(k)}} \quad (12)$$

with P_2 denoting the unnormalized probability. The normalization in the denominator of Equation (12) is implemented to guarantee that the sum of probability weights is equal to 1. As discussed above, contrary to the other three storm features, the selection of samples for

Δs_{cross} is not based on percentiles, but rather on an equal distance between the samples, and that the probability domain each sample represents is not centered on the sample with respect to probability mass distribution (probability mass to the left and right of the sample is not equal within the domain).

3.2. Full Factorial Combination and Estimation of Surge Statistics

The full factorial sampling is accomplished by an exhaustive combination of the discrete, representative samples for the four storm features, leading to a total of $N = n_a n_{mw} n_c n_w$ samples and associated probability weights:

$$\boldsymbol{\varepsilon}^{(k)} = \begin{bmatrix} \varepsilon_1^{(k_1)} \\ \varepsilon_2^{(k_2)} \\ \varepsilon_3^{(k_3)} \\ \varepsilon_4^{(k_4)} \end{bmatrix} \quad \begin{matrix} k_1 = -1, 0, 1 \\ k_2 = -(n_c - 1)/2, \dots, (n_c - 1)/2 \\ k_3 = -(n_a - 1)/2, \dots, (n_a - 1)/2 \\ k_4 = -1, 0, 1 \end{matrix} \quad (13)$$

and $P^{(k)} = P_1^{(k_1)} P_2^{(k_2)} P_3^{(k_3)} P_4^{(k_4)}$

This sample set, which is the one corresponding to the current P-Surge formulation, will be denoted as X_{ps} herein, and is composed of the pairs $\{\boldsymbol{\varepsilon}^{(k)}, P^{(k)}; k = 1, \dots, N\}$ given by Equation (13). Based on each sample $\boldsymbol{\varepsilon}^{(k)}$, the sample for $\Delta \mathbf{x}$ can be obtained from Equation (3), which, combined with the nominal storm characteristics, yields the complete parametric storm description \mathbf{x} . The latter parametrization can be ultimately used for the numerical surge prediction. As pointed out earlier, if the numerical surge prediction requires additional knowledge of DP , the $DP(t)$ values for $t > 0$ can be obtained based on the values for $R_{mw}(t)$ and $v_w(t)$ [18].

Using the above defined sample set X_{ps} , the desired surge statistics described by Equations (6) and (7) are obtained, respectively, as:

$$P_n(b) = \sum_{k=1}^N I[z_n(\boldsymbol{\varepsilon}^{(k)}|\tilde{\mathbf{x}}) > b] P^{(k)} \quad (14)$$

$$b_n^p \text{ such that } \sum_{k=1}^N I[z_n(\boldsymbol{\varepsilon}^{(k)}|\tilde{\mathbf{x}}) > b_n^p] P^{(k)} = p \quad (15)$$

4. Quasi-Monte Carlo Implementation

The full factorial sampling discussed in the previous section accommodates a physically intuitive definition of the storm ensemble, as it is based on an independent definition of representative samples for each of the storm features. This independence further accommodates a seamless exploration of the impact of each of these features on the surge statistics. It has, though, an inherent drawback: It does not have space-filling properties, shown to provide distinct advantages for the estimation of probabilistic integrals such as the ones examined here [23,24]. For this reason, a Quasi-Monte Carlo (QMC) sampling scheme is examined in this study as an alternative approach.

For QMC, it is convenient to utilize the same probability space for all storm features, so that the sampling process can be unified. The standard is to transform each of the Δx_i into a uniform in $[0, 1]$ probability model. Denoting by ε_i the transformed random variables with uniform probability in $[0, 1]$, the following relationship can be used to transform them to the original ε_i , with each one having the probability distribution discussed in Section 2.2:

$$\varepsilon_1 = \begin{cases} \varepsilon_1^{(-1)} & \text{if } \varepsilon_1 < 0.3 \\ \varepsilon_1^{(0)} & \text{if } 0.3 \leq \varepsilon_1 \leq 0.7 \\ \varepsilon_1^{(1)} & \text{if } \varepsilon_1 > 0.7 \end{cases} \quad (16)$$

$\varepsilon_i = \Phi^{-1}(\varepsilon_i); i = 2, 3, 4$

The sampling of ε_i in order to estimate the statistics of interest through QMC is performed using low-discrepancy sequences [23,24]. A low-discrepancy sequence with N elements is a sequence with the property that, for all intermediate values of $N_m < N$, its subsequence (with elements $1, \dots, N_m$) has a low discrepancy, leading to an equal partitioning of the (uniform) probability space for any chosen subsequence. They are frequently referenced as quasi-random sequences, due to their common use as a replacement of uniformly distributed random numbers for Monte Carlo (MC) sampling. It should be noted, though, that, in fact, they are not random; rather, they are completely deterministic sequences. Different types of low-discrepancy sequences exist [23] with common representatives including the Halton, van der Corput, and Sobol sequences, which, despite their subtle differences, offer similar accuracy improvements in an MC setting. In this study, Halton low-discrepancy sequences are utilized.

The reason why a QMC formulation is recommended for the problem examined here is two-fold. First, the uncertain space is low dimensional (four variables only) and the statistical quantities of interest, corresponding to relatively large occurrence probabilities (typically not lower than 5% [22]), do not constitute rare events. QMC has been shown to be quite robust for this type of implementation. Second, QMC avoids any randomness in the MC results, something that is very well aligned with the current P-Surge formulation. An alternative space-filling approach would have been to use Latin-hypercube sampling (LHS). Though LHS shares some of the variance-reduction advantages of QMC and a similar underlying foundation [23,30] of trying to evenly populate the entire probability space, it, unavoidably, introduces randomness in the surge predictions. The surge statistics will be different every time the algorithm is implemented. For this reason, QMC is the recommended approach in this application.

The QMC sampling and estimation of surge statistics is implemented through the following steps.

Step 1: Samples $\{\varepsilon^{(k)}; k = 1, \dots, N_r\}$ are obtained for ε based on the chosen low-discrepancy sequence.

Step 2: Each of the components for $\varepsilon^{(k)}$ is transformed to the respective value of $\varepsilon_i^{(k)}$ utilizing Equation (16). These ultimately replace the samples of Equation (13), while the weights of Equation (13) are transformed to equal probability weights of $P^{(k)} = 1/N_r$.

The samples created through this formulation will be denoted herein as X_{qmc} . Samples for $\Delta x_i(t)$ are then obtained utilizing the $\varepsilon_i^{(k)}$ samples of Step 2 according to Equation (3) and are combined with the nominal storm characteristics to yield the complete parametric storm description. The latter parametrization can be ultimately used for the numerical surge prediction. Finally, using the sample set X_{qmc} , the desired surge statistics described by Equations (6) and (7) are obtained, respectively, as:

$$P_n(b) = \frac{1}{N_r} \sum_{k=1}^{N_r} I[z_n(\varepsilon^{(k)} | \tilde{\mathbf{x}}) > b] \quad (17)$$

$$b_n^p \text{ such that } \frac{1}{N_r} \sum_{k=1}^{N_r} I[z_n(\varepsilon^{(k)} | \tilde{\mathbf{x}}) > b_n^p] = p \quad (18)$$

The overall proposed formulation is identical to the current P-Surge workflow with the only difference being the replacement of the samples in Equation (13) with the samples obtained through the two-step QMC sampling process that was discussed above. This makes the implementation of this alternative formulation within the current NWS workflow straightforward.

5. Illustrative Case Studies and Validation

5.1. Details for the Case Studies

The validation of the proposed methodology will consider three different case study storms: hurricane Katrina (2005), hurricane Irene (2011), and super storm Sandy (2012). These cases cover both the Gulf coast (Katrina) and the North Atlantic (Irene, Sandy) and, as shown in Figure 4, include both landfalling storms (Katrina, Sandy), as well as storms with bypassing characteristics (Irene). For each storm, two different NHC advisories will be utilized. (1) For Katrina, advisories 23 and 18 are used, corresponding, respectively, roughly to 24 h and 48 h before the storm made landfall, (2) for Sandy, advisories 26 and 18 are used, corresponding, respectively, roughly to 36 h and 72 h before landfall, and (3) for Irene, advisories 30 and 26 are used, corresponding, respectively, roughly to 24 h and 48 h before the storm bypassed New York and started gradually losing strength. The selection of two advisories per storm, instead of a single one, is made to avoid any dependence of the results on the explicit characteristics of the chosen advisory. The specific advisories are selected so that the storm would have a considerable impact on the study region and the advisories are sufficiently different (sufficient time difference between them). To simplify the presentation, the earlier advisory (smaller advisory number), representing the time further from landfall, will be denoted by A_1 and the later advisory (larger advisory number), representing the time closer to landfall, will be denoted by A_2 . Since the storm impact is typically larger in the geographic domain close to its landfall and peak surge in this domain occurs temporally close to landfall [18,31,32], advisory A_1 ultimately corresponds to a case study with larger influence of the forecast error uncertainties; the storm is further away from landfall and so, according to Figure 2 errors, the variability of the storm track, size, and intensity anticipated characteristics by the time the storm reaches landfall, when a bigger impact is anticipated, is larger. This is also clearly demonstrated by the cross-track variability shown in Figure 4.

To investigate more clearly the influence of the forecast uncertainties, for each storm two different scenarios are examined: (1) the storm occurring on its actual year, denoted as U_1 , and (2) the storm occurring in 2020, denoted as U_2 . In each case, the NWS forecast error of that specific year is utilized. Since the forecast errors keep decreasing as the years progress, which is evident from the results of Figure 2, substantially smaller uncertainties are ultimately considered for scenario U_2 . The scaling parameter a is taken to be 0.7979 for all cases apart from Katrina in U_1 scenario, for which the value used at that year, 0.6745, is adopted. Figure 4 shows, for each storm and advisory, the nominal NHC forecast track, as well as the tracks that correspond to 1 standard deviation from the nominal track based on the identified cross track variability. Evidently scenario U_1 leads to a larger variability for the storm track, as expected.

To faithfully mimic the typical P-Surge implementation, advisories are assumed to provide information for the storm track, $DP(t)$ and $v_w(t)$ for the past hurricane history ($t < 0$), while, for the storm forecast ($t > 0$), only the storm track and $v_w(t)$ are available. Based on the provided DP and v_w , the storm size R_{mw} is estimated. The current estimate for $R_{mw}(0)$ is kept constant for future predictions ($t > 0$), representing the nominal storm size forecast for $t > 0$. For the relationship between v_w , DP , and R_{mw} , updated results from study [33] are utilized. The same relationships are used to provide the $DP(t)$ predictions for $t > 0$, based on the values of $v_w(t)$ and $R_{mw}(t)$ for the respective time t . The $DP(t)$ predictions are ultimately necessary for performing the storm surge numerical estimation.

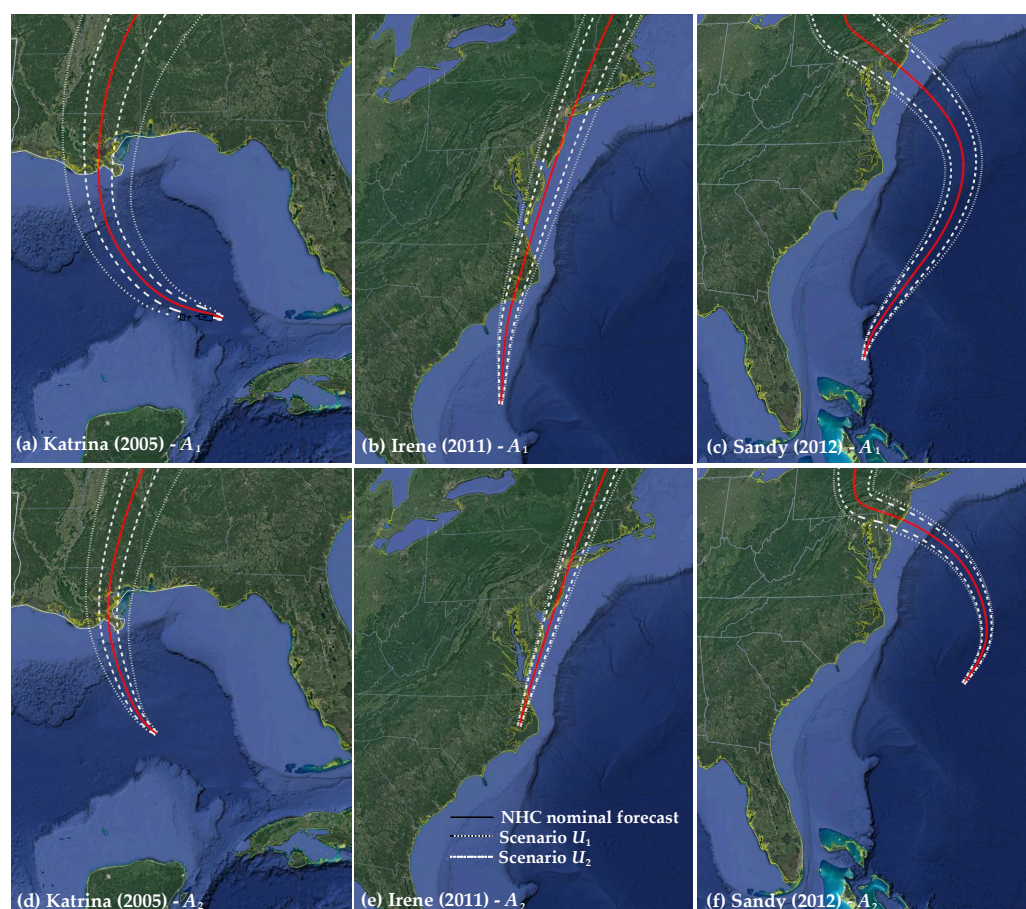


Figure 4. Storm tracks for hurricanes Katrina (2005), Irene (2011), and Sandy (2012), considering two advisories in each case: NHC nominal forecast (red solid line), as well as the track corresponding to 1 standard deviation for the cross-track variability based on the forecast errors of the year of the storm (scenario U_1 in dotted line), or the 2020 forecast errors (scenario U_2 in dash-dotted line). Top row corresponds to advisory A_1 and bottom to advisory A_2 .

5.2. Storm Surge Prediction and Validation Measures

As discussed in the introduction, and in order to support the extensive case studies, the predictions of the storm surge are performed using a storm-surge surrogate model. Details for the development of the surrogate models for the North Atlantic, used for hurricanes Irene and Sandy, and the Louisiana region, used for hurricane Katrina, are discussed in references [34] and [25], respectively. The surrogate models were developed using databases of ADCIRC simulations provided by the U.S Army Corps of Engineers through their Coastal Hazards System [15]. The accuracy of the established surrogate models is very high, with correlation coefficient with respect to the original database over 98.5%, providing a very high degree of confidence for their use within the case studies considered here. Each surrogate model offers surge predictions for close to 1 million computational nodes over the geographic domain of impact for each storm. The input to the surrogate model is identical to the one currently used for SLOSH within the P-Surge implementation: the storm track and information for storm size (R_{mw}) and intensity (DP).

The output z considered in all comparisons corresponds to the peak storm surge above ground without considering the tidal effects. This is the quantity that can be directly predicted by the established surrogate models (the surge above the datum provided by these models is simply converted to surge above ground). As discussed in Section 3.1, although tides are a deterministic process, that is not directly influenced by the forecast errors, there is a secondary effect related to the timing of the peak storm surge and how this couples with the tidal phases to produce the resultant peak water elevation (tides + storm surge). There-

fore, the influence of the forecast errors, especially those that are related to the along-track variation, are expected to be impacted by the additional consideration of tides. For this reason, the value n_a used in this study for the factorial sampling is 3, corresponding to an earlier recommendation when the impact of tides was not included in the P-Surge framework [18]. Even though the fact that the tides are not considered in the case studies is posing a limitation for the presented work (which is dictated by the need to use surrogate models to accommodate the extensive validation studies), the comparison between the QMC and the factorial sampling schemes still remains consistent. The settings examined by the two sampling schemes are appropriate for the adopted assumption related to the exclusion of tidal effects. Therefore, no impact of this assumption is expected on the comparisons of the two sampling schemes established across the case studies.

The validation is performed by comparing the output predicted through the storm sampling approach to a reference value. Reference results are denoted herein using the $\hat{\cdot}$ notation and are obtained by considering a Monte Carlo implementation based on Latin Hypercube Sampling (LHS) with a very large number of samples (100,000). This large number of samples is guaranteed to provide estimates with a high degree of confidence (very small coefficient of variation) and, thus, it can serve as the reference (benchmark) values for establishing the necessary comparisons. The computational workflow for the estimation of the surge statistics using LHS Monte Carlo is identical to the QMC implementation with the only difference that the samples $\xi^{(k)}$ originate from LHS sampling in $[0, 1]$ instead of a low-discrepancy sequence.

As validation measures, the mean error and the normalized mean error are utilized. For the probability of exceeding threshold b , the validation statistics are, respectively:

$$ME^b = \frac{1}{N_S} \sum_{n \in S} |P_n(b) - \hat{P}_n(b)| \text{ and } NME^b = \frac{\sum_{n \in S} |P_n(b) - \hat{P}_n(b)|}{\sum_{n \in S} \hat{P}_n(b)} \quad (19)$$

where S denotes a specific subset of nodes that are of interest (defined next) and N_S is the total number of such nodes within the subset. For the threshold corresponding to probability of exceedance p , the validation statistics for the mean error and the normalized mean error are:

$$ME^p = \frac{1}{N_S} \sum_{n \in S} |b_n^p - \hat{b}_n^p| \text{ and } NME^p = \frac{\sum_{n \in S} |b_n^p - \hat{b}_n^p|}{\sum_{n \in S} \hat{b}_n^p} \quad (20)$$

Instead of using all nodes within the domain of interest to estimate these statistics, only nodes corresponding to higher surge values are utilized. For this reason, the subset S is defined to correspond to the nodes for which the median surge estimate exceeded 1 m. This is separately established for each storm advisory and uncertainty scenario examined and guarantees that the comparisons are constrained to locations with considerable surge over the normal tides, avoiding nodes with smaller surge values that can lead to an erroneous confidence of the statistical accuracy (easier to accurately estimate these statistics).

For each type of output statistic, three different cases are examined. For the probability of exceeding a specific threshold, regarding the definition of that threshold b , values of 1.52 m (5 ft), 2.28 m (5.7 ft), and 3.05 m (10 ft) are utilized, extending to the common range of interest for probabilistic surge estimation [18]. For the threshold with specific probability of exceedance, probabilities p equal to 50%, 10%, and 5% are utilized. The first probability corresponds to the median estimates for the expected surge, while the latter two represent worse surge conditions that are still quite plausible. In particular, the threshold with 10% exceedance probability is a key forecasting product provided by the NWS [18].

5.3. Results and Discussion

Two different storm ensemble sampling implementations are considered: the QMC, utilizing the already defined set X_{qmc} , and the full factorial sampling adopted in P-Surge, utilizing the set X_{ps} . For brevity, the latter will be referenced herein as factorial sampling. For QMC, different N_r values will be examined. For the factorial implementation, the guidelines for P-Surge are adopted, utilizing, as discussed in Section 3.1, three representative values for the forecast errors related to size, intensity, and along-track variability, and n_c representative values for the forecast error related to the cross-track variability. This brought the total number of simulations for the factorial implementation to $N = 27n_c$. Recall that the selection of the three representative values for the along-track variability is made in order to provide consistency with the P-Surge recommendation when the impact of tides is not considered, whereas the n_c value is dependent on the step Ds given by Equation (10). As such, the n_c and, therefore, the N values are different for each advisory, impacting the nominal \tilde{R}_{mw} estimate, as well as for each uncertainty scenario, impacting the target cross-track variability at 48 h and the uncertainty in the R_{mw} estimate at 48 h. Note that, even though it would have been interesting to examine the convergence characteristics also for the factorial sampling (and not only for the QMC), such comparisons fall outside the scope of the present study. The investigation of other possible combinations that vary, for example, the number of representative values per forecast error or choose different types of discretization to select these representative values, would fundamentally shift the focus of the study. More importantly, though, the recommendations reviewed in Section 3 and adopted here have been established by the NWS in order to accommodate the best average accuracy, across multiple years of refinement and advances. In other words, for a given N , the accuracy that is established by the adopted factorial sampling implementation should be considered as close to the optimal one, at least for a typical storm.

Figures 5–7 show the normalized accuracy metrics for the threshold with probability of exceedance p , NME^p , and for the probability of exceeding threshold b , NME^b , for hurricanes Katrina, Irene, and Sandy, respectively. Figure 8 shows accuracy metrics (without normalization) for the same quantities of interest for hurricane Irene. The top row in each figure corresponds to the threshold with probability of exceedance p , and the bottom to the probability of exceeding threshold b . The first two columns correspond to advisory A_1 and uncertainty scenarios U_1 and U_2 , respectively, while the last two columns to advisory A_2 and have the same order for the uncertainty scenarios as in advisory A_1 (so U_1 and U_2 , respectively). In each subplot, the QMC results for varying N_r values are shown as curves and the factorial sampling implementation as single points, since they correspond to a unique sample size. To distinguish more easily the factorial sampling implementation, a vertical, dashed, gray line is added in all cases. Figures 9 and 10 present different hazard maps, depicting the spatial distribution of surge statistics over the geographic domain of interest, for advisory A_2 for super storm Sandy looking at scenario U_1 , for three different sampling approaches: the reference case (left column), the factorial P-Surge sampling (middle column), and the QMC sampling for a N_r value similar to the one used for the factorial P-surge sampling (right column). Figure 9 shows results for thresholds corresponding to exceedance probabilities of 50% and 10%, while Figure 10 shows results for the probabilities of exceeding thresholds 1.52 m (5 ft) and 3.05 m (10 ft). All hazard maps are focused on the geographic domain where the storm has a strong impact, corresponding to non-zero values for the thresholds and the probabilities examined.

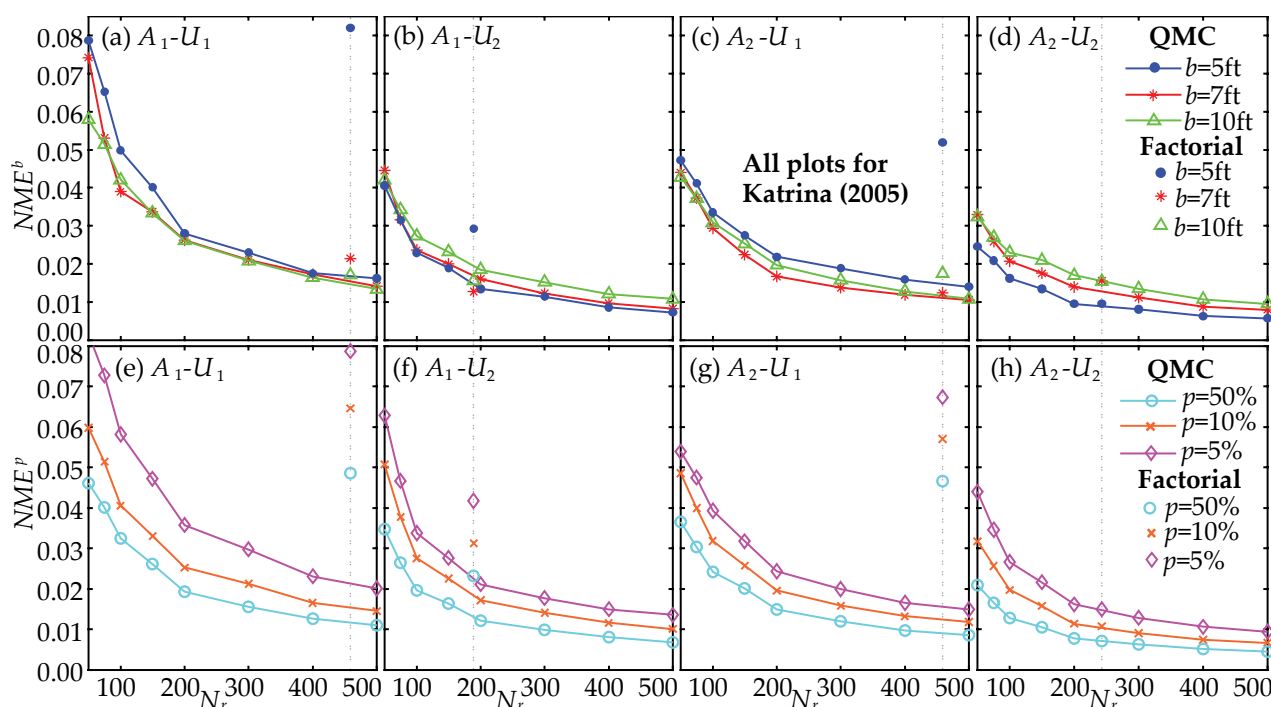


Figure 5. Normalized accuracy metrics NME^b (top row) and NME^p (bottom row) as functions of the number of samples N_r used for hurricane Katrina (2005) for two different advisories and two uncertainty scenarios. First two columns correspond to advisory A_1 and uncertainty scenarios U_1 (first column) and U_2 (second column), while the two last columns correspond to advisory A_2 and uncertainty scenarios U_1 (third column) and U_2 (fourth column), respectively. Results for both the QMC (curves as function of samples) and the factorial sampling (for a specific sample size) are shown. A gray, dashed, vertical line is added to distinguish more easily the factorial sampling.

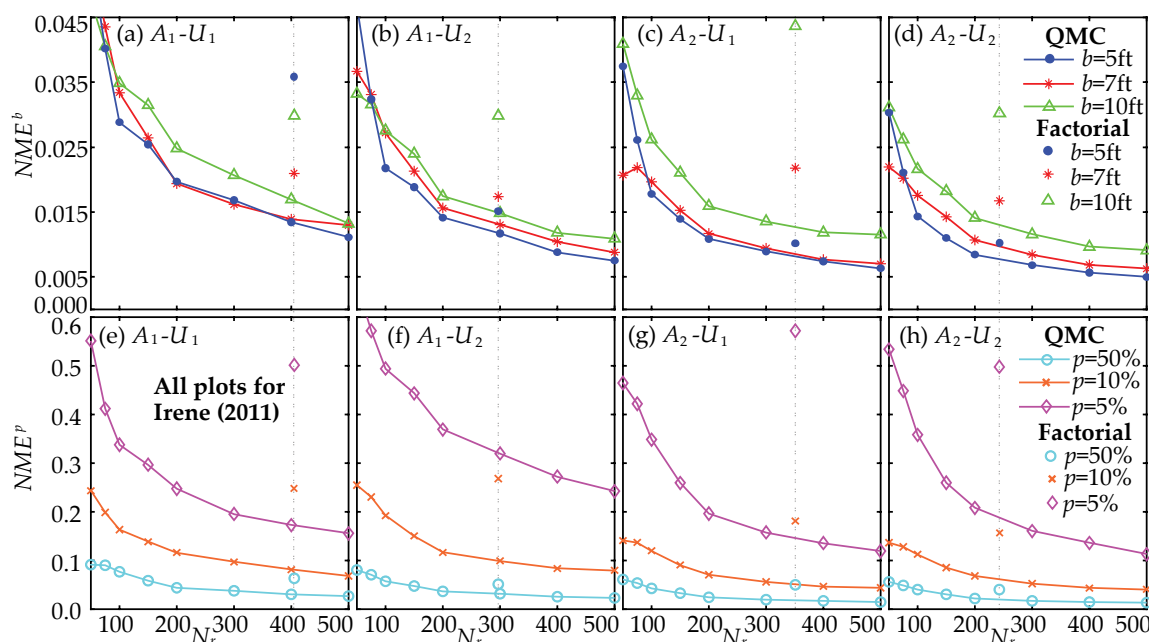


Figure 6. Normalized accuracy metrics NME^b (top row) and NME^p (bottom row) as functions of the number of samples N_r used for hurricane Irene (2011) for two different advisories and two uncertainty scenarios. First two columns correspond to advisory A_1 and uncertainty scenarios U_1 (first column) and U_2 (second column), while the two last columns correspond to advisory A_2 and uncertainty scenarios U_1 (third column) and U_2 (fourth column), respectively. Results for both the QMC (curves as function of samples) and the factorial sampling (for a specific sample size) are shown. A gray, dashed, vertical line is added to distinguish more easily the factorial sampling.

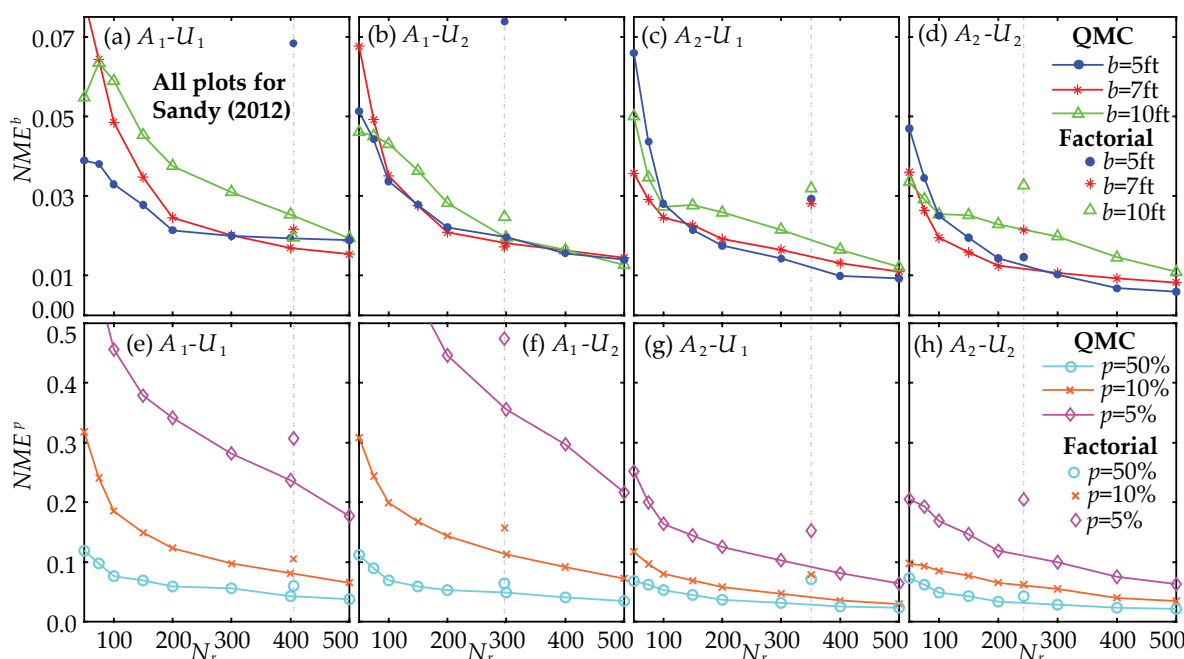


Figure 7. Normalized accuracy metrics NME^b (top row) and NME^p (bottom row) as functions of the number of samples N_r used for hurricane Sandy (2012) for two different advisories and two uncertainty scenarios. First two columns correspond to advisory A_1 and uncertainty scenarios U_1 (first column) and U_2 (second column), while the two last columns correspond to advisory A_2 and uncertainty scenarios U_1 (third column) and U_2 (fourth column), respectively. Results for both the QMC (curves as function of samples) and the factorial sampling (for a specific sample size) are shown. A gray, dashed, vertical line is added to distinguish more easily the factorial sampling.

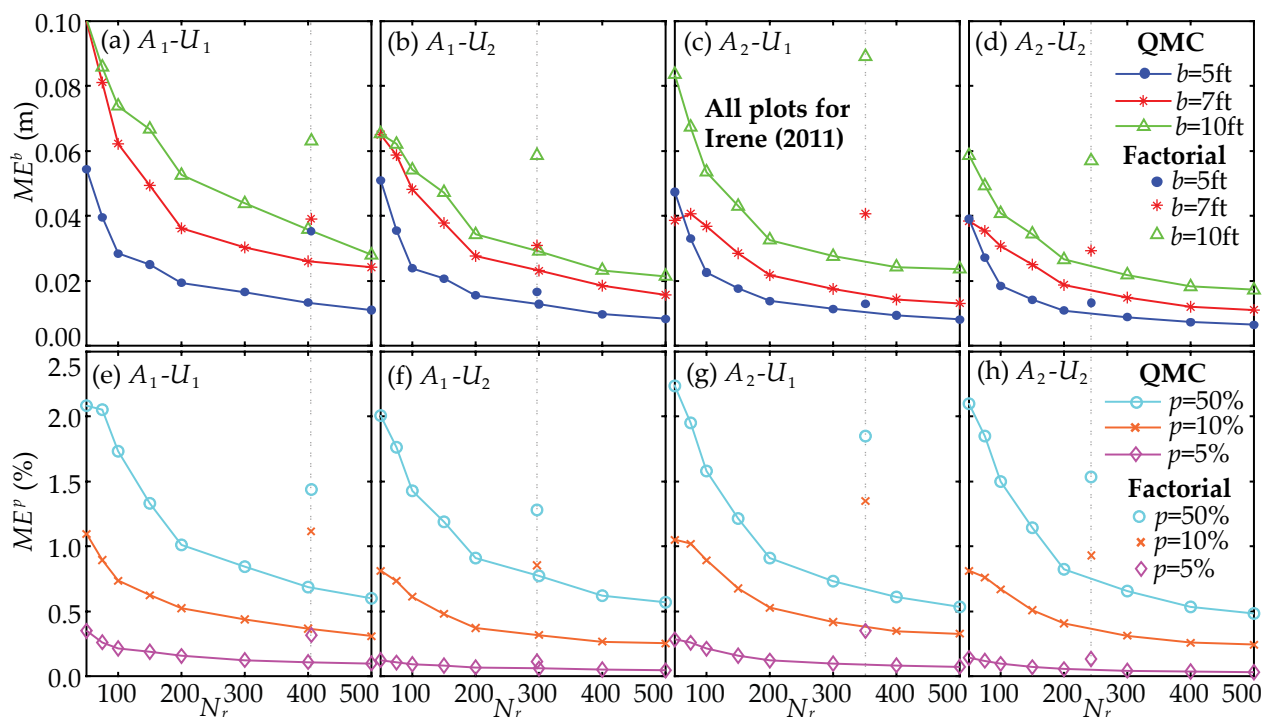


Figure 8. Accuracy metrics ME^b (top row) and ME^p (bottom row) for hurricane Irene (2011) for two different advisories and two uncertainty scenarios. First two columns correspond to advisory A_1 and uncertainty scenarios U_1 (first column) and U_2 (second column), while the two last columns correspond to advisory A_2 and uncertainty scenarios U_1 (third column) and U_2 (fourth column), respectively. Results for both the QMC (curves as function of samples) and the factorial sampling (for a specific sample size) are shown. A gray, dashed, vertical line is added to distinguish more easily the factorial sampling.

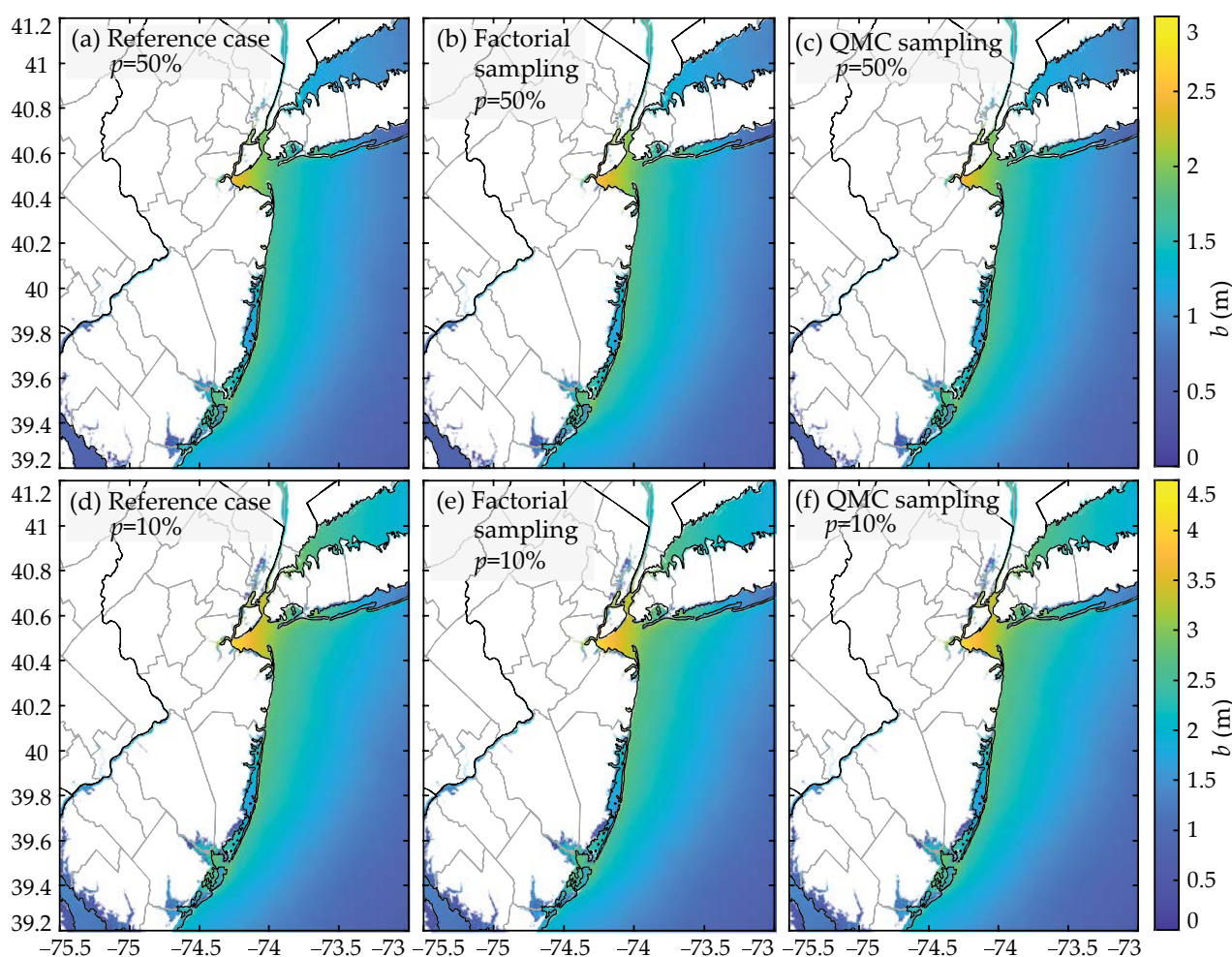


Figure 9. Distribution over the spatial domain of interest of the thresholds corresponding to exceedance probabilities 50% (top row) and 10% (bottom row) for advisory A_1 and scenario U_1 for super storm Sandy. Three different sampling approaches are presented for estimating the desired statistics: The reference case (left column) is presented to establish the comparisons between the factorial P-Surge sampling (middle column) and the QMC sampling (right column) for a N_r value similar to the one used for the factorial P-surge sampling.

A variety of interesting remarks can be made based on the results presented. The key one is that the QMC sampling approach offers computational advantages over the factorial sampling. It is evident in Figures 5–8 that, as the number of samples N_r increases, the accuracy of QMC consistently increases, as expected, while for the same size of storm ensembles (same number of samples), its accuracy is the same or higher compared to the accuracy established by the factorial sampling implementation. As discussed in the introduction, this improvement in accuracy can be equivalently interpreted as an ability to accommodate considerable computational savings. Though trends change from storm to storm and across the different statistical quantities examined, for the majority of the cases a consistent 2- up to 3-fold reduction of the ensemble size (computational savings) using the proposed QMC implementation can be established, while offering the same statistical accuracy as the current factorial sampling formulation. The reduction in some of the cases is substantially larger than these numbers, especially when examining probabilistic quantities that represent less frequent events, for example, thresholds with smaller exceedance probabilities p , or probabilities of exceeding larger surge values b . Since these statistical products are important for making well-informed decisions during landfalling events, the recorded improvement in the associated computational burden (required number of storm simulations) for obtaining the same level of accuracy estimates in the surge statistics should be treated as an important achievement. As it was already

mentioned earlier, results vary from case to case, depending on the storm features and on the uncertainty description, especially since the number of samples for the P-Surge factorial implementation is dependent on the R_{mw} value that dictates the number of cross-track samples utilized. Nevertheless, consistent benefits can be observed in all results. This behavior is anticipated and it agrees with the well-known advantages of quasi-Monte Carlo estimation when compared to alternative methods for estimating statistics in problems similar to the one examined here [23,24]. Moreover, the QMC implementation shows a consistent performance across all statistical quantities examined, independently of the storm, advisory, or forecast error size (U_1 and U_2). The same remark does not apply for the factorial sampling, whose behavior can change from advisory to advisory across the different quantities examined. This robustness in accuracy that the QMC implementation enjoys further stresses its advantages; it offers a reliable performance that is independent of the storm characteristics and the size of the forecast error. Especially the robustness, with respect to the latter quantity, is a highly desirable property, since it means that the implementation does not need to be finely tuned if, in the future, the size of the forecast errors and, consequently, the uncertainty description for the storm features reduces.

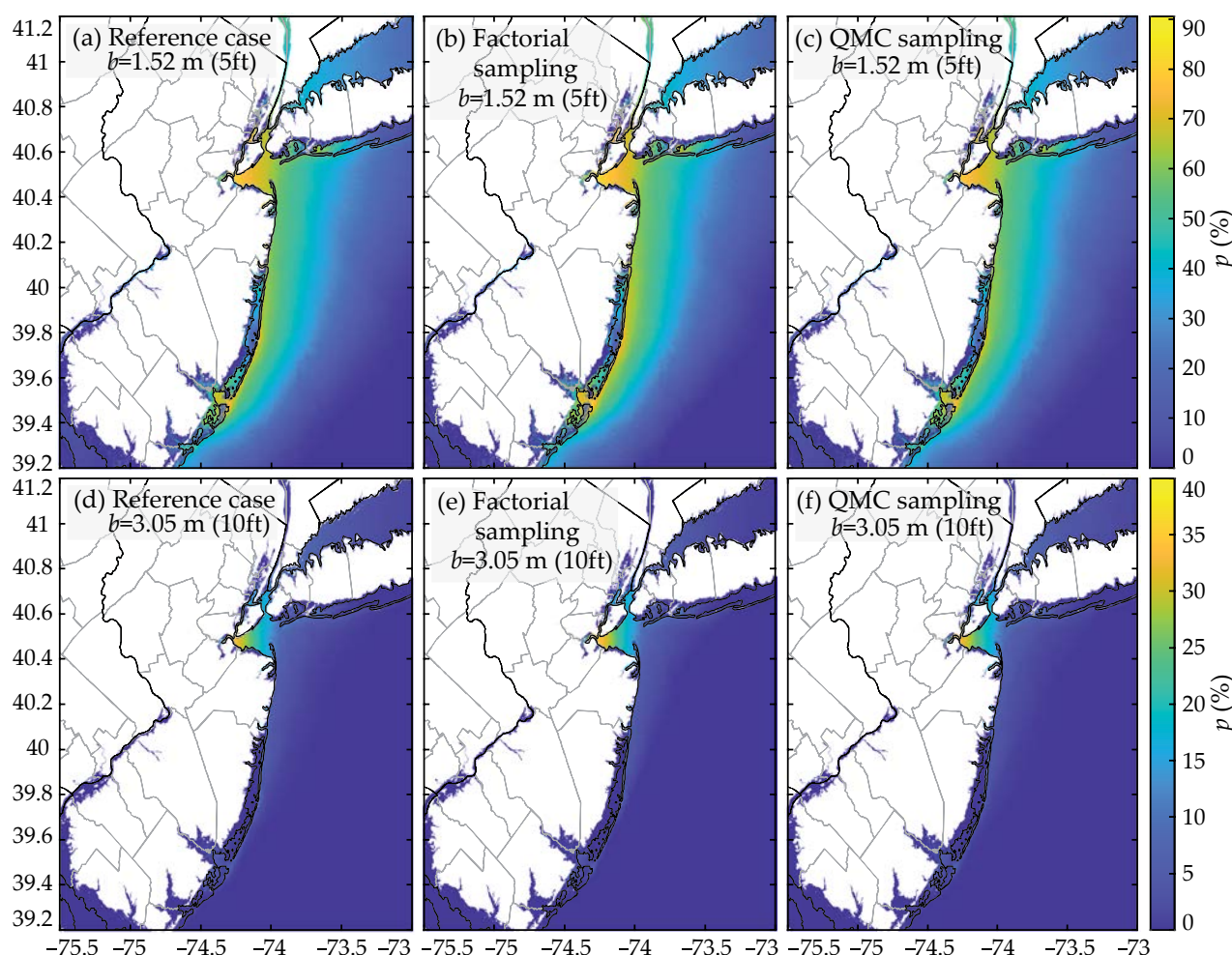


Figure 10. Distribution over the spatial domain of interest of the probabilities of exceeding thresholds 1.52 m (5 ft) (top row) and 3.05 m (10 ft) (bottom row) for advisory A_1 and scenario U_1 for super storm Sandy. Three different sampling approaches are presented for estimating the desired statistics: The reference case (left column) is presented to establish the comparisons between the factorial P-Surge sampling (middle column), and the QMC sampling (right column) for a N_r value similar to the one used for the factorial P-surge sampling.

It is important to note, though, that the accuracy for estimating the surge statistics is high across all the methods for all the examined storms/advisories. This can be easily observed in the non-normalized statistics reported in Figure 8, but more importantly in the hazard maps presented in Figures 9 and 10. The error sizes in Figure 8 are very small, while no differences are discernible for the hazard maps between the reference, the QMC, and the factorial P-Surge implementations. Trends are identical for all other cases (hurricanes, advisories, uncertainty scenarios) and, for this reason, only selective results are reported in the aforementioned three figures. These trends show that, even though QMC does offer advantages with respect to the computational efficiency in the estimation of surge statistics, it is important to acknowledge that the current P-Surge factorial sampling implementation provides highly accurate estimates for the products the NWS currently provides to emergency response managers, just at a higher computational cost. The aforementioned potential advantages, though, provide a clear argument that the alternative QMC formulation should be promoted over the existing factorial sampling.

A comparison between the hazard maps in the first and second row of Figures 9 and 10 makes it evident that the impact of the uncertainties associated with the forecast errors is significant; the values of the surge for different exceedance probabilities in Figure 9 and the probability of exceeding different surge values in Figure 10 are drastically different. To better observe the differences in the figures, note that different scales are used in each row. Of course, the differences will change depending on the size of the considered uncertainties, but the comparison here clearly demonstrates an important variability for the estimated surge even when the storm is close to landfall, corresponding, as discussed earlier, to a case that the impact of the forecast error uncertainties is expected to be smaller. This observation further stresses the importance of explicitly considering these uncertainties within a formal uncertainty propagation step, as NWS currently does, and the advantages offered by the QMC formulation, since it allows for a more accurate propagation of these uncertainties to estimate the various surge statistics.

Going back to the error metrics in Figures 5–8, and comparing across the different subplots, it is evident that the larger the forecast uncertainties, for example, comparing U_1 against U_2 or A_1 against A_2 , the smaller the accuracy (the larger the errors) for estimating the surge statistics. This is, of course, anticipated, since larger uncertainties are increasing the surge variability, posing a greater challenge for the accurate estimation of its statistics. Trends are more evident in the comparison between U_1 and U_2 , since, between advisories A_1 and A_2 , the nominal storm features also changed, something that can fundamentally alter the storm surge manifestation itself. Similarly, a comparison across different storms should be cautiously made, since the size and strength of each storm impacts the results, especially the statistics that relate to probability of exceeding specific surge values. For stronger storms, with a larger generated surge, the probabilities are larger and, so, the accuracy in the estimation of these probabilities is higher, since fewer challenges exist in the calculation of more frequent statistics (see also the discussion in the next paragraph). This is the main reason why errors for Katrina for NME^b are substantially smaller (Figure 5). The same does not apply for NME^p since the compared statistics in this case are adaptively chosen to correspond to the same percentile for the surge, creating a more consistent comparison across storms and advisories. Independently of the size of the forecast uncertainties, the QMC formulation appears to offer evident advantages. In fact, these advantages are typically higher in cases where these uncertainties are larger, further stressing the potential that this formulation has to offer; it can improve computational efficiency for implementations for which accounting for the forecast uncertainties is more critical (bigger influence of the uncertainties).

A final important comparison needs to be established with respect to the different statistical quantities. Statistics that represent less frequent events, for example, thresholds for smaller exceedance probabilities p or probabilities of exceeding larger surge values b , are associated, in general, with smaller accuracy for the QMC formulation. Even though this trend does not hold for all the cases examined for the NME^p statistic, as it is affected

by a range of factors, including the normalization adopted and the selection of the nodes for which the estimation of the error metrics is performed, it is predominantly true, while it always holds for NME^b . As discussed earlier, the QMC implementation has the same good performance regardless of the statistical quantity examined, even for quantities that correspond to infrequent events, and, more importantly, its behavior shows no outliers (abnormal errors for a specific statistical quantity examined), something that cannot be always said for the factorial sampling. Both these are critical features: QMC can be reliably used to estimate even statistics for infrequent surge values, and, more importantly, it is shown to provide robust estimates across all cases. It is interesting to note that the factorial sampling provides in some instances lower normalized accuracy for frequent statistical quantities, for example, for the median surge values, represented by NME^p for $p = 50\%$. This is, of course, possible and does not represent an alarming trend, especially if one considers, as discussed earlier, that errors are, in general, small. The consistency, in contrast, of the QMC results further stresses the benefits that the latter can offer.

Overall, the extensive validation across different storms, advisories (per storm), levels of uncertainty (per advisory), and statistical quantities of interest (per advisory) demonstrate the advantages of the QMC formulation in a broad setting. This formulation can reliably facilitate a significant reduction of the computational burden, requiring a smaller number of simulations (smaller storm ensemble size) for estimating the current surge products offered by the NWS or even accommodating the accurate estimation of less frequent events compared to the ones currently considered, for example, thresholds corresponding to low (1% or lower) exceedance probabilities.

6. Conclusions

A quasi-Monte Carlo (QMC) formulation was examined for the propagation of the errors associated with the NHC (National Hurricane Center) forecasts during landfalling storms. The existing framework for the uncertainty quantification associated with these errors and for the estimation of the different surge statistics of interest was first reviewed. The implementation of this framework was facilitated through the P-Surge tool of the NWS (National Weather Service) and it was based on the generation of a storm ensemble that represented the underlying probability distributions for the storm features (as prescribed by the forecast errors) and on the subsequent simulation of the anticipated surge for each of these storms. A deterministic, full factorial sampling scheme is currently implemented within P-Surge for the selection of the storm ensemble selection. A QMC formulation for this selection was then proposed in this paper as a potential alternative. This formulation still promotes a deterministic sampling, maintaining this desirable feature of P-Surge, but replaces the factorial, combinatorial sampling with space-filling, low-discrepancy sequences. Overall, it represents a minimal deviation from the existing P-Surge workflow, which can be seamlessly integrated within it.

The validation of this newly introduced sampling formulation considered three storms with fundamentally different characteristics, two different advisories per storm and two different scenarios for the forecast error definitions. It also examined different surge statistics as products of interest, extending to both the surge with specific probability of exceedance p , as well as the probability of exceeding a specific surge threshold b . Different values were examined for both b and p so that the achieved accuracy could be examined for both frequent and infrequent events. The uncertainty quantification followed the setup of version 2.7 of P-Surge. For accommodating the extensive validation study, the numerical surge estimation was performed using a surrogate model instead of the SLOSH numerical model currently employed within the P-Surge workflow. This use of surrogate models did not allow the explicit consideration of the tidal influence in the analyses. Although this exclusion of tides in the case studies is an undeniable limitation, the comparison of the examined sampling methods was performed under equal terms, since, for the factorial sampling, the number of representative samples for the along-track variability was selected in an appropriate way to reflect the absence of the tidal effect. Results showed that although

both the existing factorial P-Surge sampling and the proposed QMC sampling provide estimates with high accuracy, the QMC formulation offers considerable advantages. In fact, in the majority of the advisories and for the statistical quantities examined, the same level of accuracy can be obtained by QMC utilizing at least 2–3 times smaller storm ensembles, accommodating considerable computational savings, something that allows the delivery of products within a smaller time window, facilitating an even more efficient real-time forecasting process. QMC was also shown to have remarkable robustness in accuracy performance, with no outliers in the accuracy trends across all the (extensive) case studies examined. Results clearly demonstrate the benefits that QMC can offer for the estimation of surge statistics during landfalling storms.

It is important to acknowledge that the overall accuracy of the storm surge estimates is impacted by a number of factors, as detailed in the introduction, but the focus of this paper was strictly on the uncertainty propagation component of this problem. The potential improvement in statistical accuracy offered by the QMC implementation for this component can be leveraged, though, to establish advances across the remaining components, for example, towards the improvement of the grid discretization to reduce numerical errors in the surge estimates.

Author Contributions: Conceptualization, A.A.T., J.J.W., H.L.T., and A.P.K.; methodology, A.P.K., E.A., and A.A.T.; software, E.A.; validation, A.P.K. and A.A.T.; writing—original draft preparation, A.A.T. and A.P.K.; writing—review and editing, A.A.T., A.P.K., H.L.T., J.J.W., and E.A.; project administration, A.A.T.; funding acquisition, A.A.T. and J.J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Oceanographic and Atmospheric Administration (NOAA), grant number NA19OAR0220089. The views and opinions expressed in this paper are those of the authors and do not represent NOAA.

Data Availability Statement: Data utilized for the creation of the surrogate models were provided by the Army Corps of Engineers through the Coastal Hazards System <https://chs.ercd.dren.mil/>. Data for storm advisories can be found in <https://rammb-data.cira.colostate.edu>. All other data used are reported in the manuscript.

Acknowledgments: Authors would like to thank Arthur Taylor for offering guidance with respect to the sampling implemented in version 2.7 of P-Surge, for providing access to the NHC forecast errors statistics and advisories used in this study, and for feedback related to the importance of the along-track variability of storms when tides are incorporated into the analysis. They would also like to thank the Army Corp of Engineers, Coastal Hydraulics Laboratory of the Engineering Research and Development Center for providing access to the databases that were used to develop the surrogate models for the storm surge predictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kennedy, A.; Cox, D.; Irish, J.; Kaihatu, J.; Lynett, P.; Tomiczek, T. Envisioning the Future Coast: Coastal Engineering Research in the Coming Decades. In Proceedings of the A Report from the Coastal Engineering Research Framework Workshop, Arlington, VA, USA, 13–14 November 2018. [\[CrossRef\]](#)
2. Geological Society of America. Managing U.S. Coastal Hazards. Available online: <https://www.geosociety.org/gsa/positions/position22.aspx> (accessed on 20 November 2021).
3. Pasch, R.J.; Roberts, D.P.; Blake, E.S. The 2019 Atlantic Hurricane Season: An Active and Destructive Year. *Weatherwise* **2020**, *73*, 32–39. [\[CrossRef\]](#)
4. Klotzbach, P.J.; Schreck III, C.J.; Collins, J.M.; Bell, M.M.; Blake, E.S.; Roache, D. The extremely active 2017 North Atlantic hurricane season. *MWR* **2018**, *146*, 3425–3443. [\[CrossRef\]](#)
5. Blake, E.S. The 2017 Atlantic hurricane season: Catastrophic losses and costs. *Weatherwise* **2018**, *71*, 28–37. [\[CrossRef\]](#)
6. Wood, K.M.; Klotzbach, P.J.; Collins, J.M.; Schreck, C.J. The record-setting 2018 eastern North Pacific hurricane season. *Geophys. Res. Lett.* **2019**, *46*, 10072–10081. [\[CrossRef\]](#)
7. Dangendorf, S.; Marcos, M.; Wöppelmann, G.; Conrad, C.P.; Frederikse, T.; Riva, R. Reassessment of 20th century global mean sea level rise. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5946–5951. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Javeline, D.; Kijewski-Correa, T. Coastal homeowners in a changing climate. *Clim. Chang.* **2019**, *152*, 259–274. [\[CrossRef\]](#)

9. Lim, Y.-K.; Schubert, S.D.; Kovach, R.; Molod, A.M.; Pawson, S. The roles of climate change and climate variability in the 2017 Atlantic hurricane season. *Sci. Rep.* **2018**, *8*, 1–10.
10. Robertson, I.N.; Riggs, H.R.; Yim, S.C.; Young, Y.L. Lessons from Hurricane Katrina storm surge on bridges and buildings. *J. Waterw. Port Coast. Ocean. Eng.* **2007**, *133*, 463–483. [[CrossRef](#)]
11. Marshall, J.S.; Daniel, A.; Lyda, A.; Roueche, D.; Davis, B.; Djima, W.; Heo, Y.; Kijewski-Correa, T.; Moravej, M.; Rittelmeyer, B.; et al. StEER-Hurricane Dorian: Field Assessment Structural Team (FAST-1) Early Access Reconnaissance Report (EARR). *Des. Safe* **2019**, *3*, 10. [[CrossRef](#)]
12. Jelesnianski, C.P.; Chen, J.; Shaffer, W.A. *SLOSH: Sea, Lake, and Overland Surges from Hurricanes*; NOAA Technical Report, NWS 48; US Department of Commerce: Silver Spring, MD, USA, 1992.
13. Luettich, R.A., Jr.; Westerink, J.J.; Scheffner, N.W. *ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries*; Report 1. Theory and methodology of ADCIRC-2DDI and ADCIRC-3DL.; Dredging Research Program Technical Report DRP-92-6; U.S Army Engineers Waterways Experiment Station: Vicksburg, MS, USA, 1992.
14. Glahn, B.; Taylor, A.; Kurkowski, N.; Shaffer, W.A. The role of the SLOSH model in National Weather Service storm surge forecasting. *Natl. Weather. Dig.* **2009**, *33*, 3–14.
15. Nadal-Caraballo, N.C.; Campbell, M.O.; Gonzalez, V.M.; Torres, M.J.; Melby, J.A.; Taflanidis, A.A. Coastal Hazards System: A Probabilistic Coastal Hazard Analysis Framework. *J. Coast. Res.* **2020**, *95*, 1211–1216. [[CrossRef](#)]
16. Regional and Mesoscale Meteorology Branch. Available online: <https://rammb-data.cira.colostate.edu/> (accessed on 20 November 2021).
17. Kijewski-Correa, T.; Taflanidis, A.; Vardeman, C.; Sweet, J.; Zhang, J.; Snaiki, R.; Wu, T.; Silver, Z.; Kennedy, A. Geospatial environments for hurricane risk assessment: Applications to situational awareness and resilience planning in New Jersey. *Front. Built Environ.* **2020**, *6*, 549106. [[CrossRef](#)]
18. Taylor, A.A.; Glahn, B. Probabilistic guidance for hurricane storm surge. In Proceedings of the 19th Conference on Probability and Statistics, Bethlehem, PA, USA, 17–19 March 2022.
19. Chen, J.H.; Lin, S.J.; Magnusson, L.; Bender, M.; Chen, X.; Zhou, L.; Xiang, B.; Rees, S.; Morin, M.; Harris, L. Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophys. Res. Lett.* **2019**, *46*, 4495–4501. [[CrossRef](#)]
20. Hamill, T.M.; Brennan, M.J.; Brown, B.; DeMaria, M.; Rappaport, E.N.; Toth, Z. NOAA's future ensemble-based hurricane forecast products. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 209–220. [[CrossRef](#)]
21. Resio, D.T.; Irish, J.L.; Westerink, J.J.; Powell, N.J. The effect of uncertainty on estimates of hurricane surge hazards. *Nat. Hazards* **2012**, *66*, 1443–1459. [[CrossRef](#)]
22. Gonzalez, T.; Taylor, A. Development of the NWS Probabilistic Tropical Storm Surge Model. In Proceedings of the 33rd Conference on Hurricanes and Tropical Meteorology, Ponte Vedra, FL, USA, 11 May 2021.
23. Lemieux, C. *Monte Carlo and Quasi-Monte Carlo Sampling*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
24. Caflisch, R.E. Monte carlo and quasi-Monte Carlo methods. *AcNum* **1998**, *1998*, 1–49. [[CrossRef](#)]
25. Jia, G.; Taflanidis, A.A.; Nadal-Caraballo, N.C.; Melby, J.; Kennedy, A.; Smith, J. Surrogate modeling for peak and time dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms. *Nat. Hazards* **2015**, *81*, 909–938. [[CrossRef](#)]
26. Kyprioti, A.P.; Taflanidis, A.A.; Nadal-Caraballo, N.C.; Campbell, M.O. Incorporation of sea level rise in storm surge surrogate modeling. *Nat. Hazards* **2020**, *105*, 531–563. [[CrossRef](#)]
27. Bass, B.; Bedient, P. Surrogate modeling of joint flood risk across coastal watersheds. *JHyd* **2018**, *558*, 159–173. [[CrossRef](#)]
28. Al Kajbaf, A.; Bensi, M. Application of surrogate models in estimation of storm surge: A comparative assessment. *Appl. Soft Comput.* **2020**, *91*, 106184. [[CrossRef](#)]
29. Jelesnianski, C.P.; Taylor, A. A preliminary view of storm surges before and after storm modifications. In *NOAA Technical Memorandum*; Environmental Research Laboratories; Weather Modification Program Office: Washington, DC, USA, 1973.
30. Robert, C.P.; Casella, G. *Monte Carlo Statistical Methods*, 2nd ed.; Springer: New York, NY, USA, 2004.
31. Forbes, C.; Rhome, J.; Mattocks, C.; Taylor, A. Predicting the storm surge threat of Hurricane Sandy with the National Weather Service SLOSH model. *J. Mar. Sci. Eng.* **2014**, *2*, 437–476. [[CrossRef](#)]
32. Sebastian, A.; Proft, J.; Dietrich, J.C.; Du, W.; Bedient, P.B.; Dawson, C.N. Characterizing hurricane storm surge behavior in Galveston Bay using the SWAN+ ADCIRC model. *Coast. Eng.* **2014**, *88*, 171–181. [[CrossRef](#)]
33. Knaff, J.A.; Zehr, R.M. Reexamination of tropical cyclone wind–pressure relationships. *Weather. Forecast.* **2007**, *22*, 71–88. [[CrossRef](#)]
34. Kyprioti, A.P.; Taflanidis, A.A.; Nadal-Caraballo, N.C.; Campbell, M. Storm hazard analysis over extended geospatial grids utilizing surrogate models. *Coast. Eng.* **2021**, *29*, 103855. [[CrossRef](#)]