Efficient probabilistic storm surge estimation through adaptive importance sampling across storm advisories

WoongHee Jung[1], Alexandros A. Taflanidis[1], Aikaterini P. Kyprioti[1], Ehsan Adeli[1], Joannes J. Westerink[1] and Hendrik Tolman[2]

[1]University of Notre Dame, Department of Civil and Environmental Engineering and Earth Sciences

[2] National Weather Service, NOAA / DOC

Abstract

During landfalling tropical storms, probabilistic predictions of the storm surge constitute important products for guiding emergency response/preparedness decisions. The probabilistic formulation for these predictions is established by considering forecast errors for the intensity, size, cross- and along-track variability of the National Hurricane Center (NHC) advisories. These errors define ultimately the vector of uncertain storm features, with corresponding probability distributions chosen based on historical data. Propagation of the uncertainty in these features, serving as input to a numerical model for predicting storm surge, provides the desired statistical products, for example, the water level corresponding to a specific exceedance probability. This estimation is repeated whenever the NHC updates the storm advisory. Monte Carlo (MC) simulation is considered here as the numerical tool for facilitating this uncertainty propagation. Specifically, the implementation of adaptive importance sampling (IS) across the storm advisories is examined to improve MC computational efficiency, with objectives to attain better accuracy estimates or accommodate predictions using a smaller number of storm surge simulations. IS achieves this objective by introducing a proposal density (IS density) to choose a storm ensemble with higher contribution to the examined probabilistic product(s). In the proposed implementation, storm surge simulation results from the current advisory are leveraged to select the IS density to use for the next advisory, establishing an IS workflow that involves minimal additional computational burden, since readily available information is utilized. The requirement to estimate the storm surge across a large geographic domain, leading to the definition of a large number of quantities of interests (QoIs), poses a significant challenge in this case, since these quantities typically represent competing IS choices. Different schemes are discussed to establish a compromising solution, with emphasis on the use of principal component analysis to reduce the computational burden for the IS density selection. Alternative IS formulations are examined, and an adaptive selection of the IS characteristics is discussed, utilizing a novel global sensitivity analysis scheme and an efficient estimation of the anticipated IS accuracy. In order to guarantee robustness, since the IS density is established using limited information (i.e, limited number of simulations) involving a large number of QoIs, and, more importantly, it is chosen based on the information of the current advisory but is

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

implemented in the next one, a defensive IS scheme is introduced. Validation considering different historical storms demonstrates robustness and improvements in accuracy, accomplished by the proposed sharing of information across advisories through an adaptive IS formulation. Extensions to accommodate a quasi-Monte Carlo implementation are also discussed.

Keywords: landfalling storms; storm forecast errors; probabilistic surge estimation; Monte Carlo simulation; adaptive importance sampling; high-dimensional output.

1 Introduction

Evacuation and emergency preparedness decisions at regional and national levels during landfalling tropical storms rely on estimates of the expected storm surge impact (Taylor and Glahn 2008; Smith et al. 2011; Chen et al. 2019; Kijewski-Correa et al. 2020). Such estimates can be obtained by using the National Hurricane Center's (NHC) official advisory for the storm current/forecasted track, size and intensity characteristics. These characteristics provide the input to an appropriate hydrodynamic numerical model, such as the Sea, Lake, and Overland Surges from Hurricanes (SLOSH) model (Jelesnianski et al. 1992; Glahn et al. 2009) [current selection in the NHC workflow] or the ADvanced CIRrculation Model for Shelves, Coastal Seas, and Estuaries (ADCIRC) model (Luettich et al. 1992), to predict the anticipated surge for the corresponding storm scenario. To establish well-informed decisions in this setting, uncertainties associated with the NHC forecast advisory need to be explicitly accounted for (Hamill et al. 2012; Resio et al. 2012). Such uncertainties can be quantified through forecast errors for four different storm features (Taylor and Glahn 2008): (a) the cross-track variation; (b) the along-track variation (i.e. the storm forward translational speed); (c) the storm size, represented by the radius of maximum winds; and (d) the storm intensity, represented by the maximum velocity of sustained winds. Through analysis of historical data (errors from past forecasts), probability distributions for these forecasted storm features can be defined (Gonzalez and Taylor 2018). This ultimately leads to a formal definition of the probabilistic storm surge predictions for landfalling storms (Taylor and Glahn 2008; Kyprioti et al. 2021a), established by combining the official NHC advisory, representing the nominal track, size, and intensity, with the uncertainty description corresponding to the aforementioned four storm features, representing the variability of the track, size, and intensity. The predictions of interest correspond to statistical products like the storm surge with specific probability of being exceeded, or the probability that the storm surge will exceed specific reference thresholds. Mathematically, these can be described through probabilistic integrals across the uncertain storm features for different quantities of interest (QoIs) (Kyprioti et al. 2021a), with the latter corresponding to the storm surge for different locations within the geographic domain of storm impact. Numerically, these products can be obtained by creating an ensemble of storm scenarios based on the official NHC advisory and the forecast error probabilistic description, then predicting the storm surge

for each of them through a hydrodynamic model as discussed earlier, and finally estimating the desired statistics across the ensemble. This estimation is repeated whenever a new advisory becomes available, typically every 6 hr, to provide updated statistical products for emergency response/preparedness decisions.

The National Weather Service (NWS) has created the Probabilistic tropical storm Surge (P-Surge) model (Taylor and Glahn 2008; Gonzalez and Taylor 2018) to support the aforementioned uncertainty quantification and propagation. The traditional P-Surge formulation uses a factorial sampling for establishing statistical predictions, defining a small number of representative values for each storm feature, and considering all possible combinations of these values (factorial design) to generate the required storm ensemble. The relative weight of each of these storms is obtained based on the relative likelihood of the corresponding representative values, utilizing the underlying probability distribution of the forecast errors. Recently (Kyprioti et al. 2021a), an alternative implementation was examined for the uncertainty propagation in this setting, adopting quasi-Monte Carlo (QMC) through the use of low discrepancy sequences (Lemieux 2009) for the numerical integration of the associated probabilistic integrals. It was shown in (Kyprioti et al. 2021a) that QMC can improve computational efficiency and provide statistical estimates for the surge with the same degree of accuracy as factorial sampling, while using a smaller number of storm simulations. Both these approaches though, factorial-sampling or QMC, focus strictly on the uncertainty in the input representation, trying to create more appropriate storm ensemble scenarios, either through the selection of representative values or through the use of low discrepancy sequences. They do not utilize any information regarding the surge output from the different (readily available) simulations to improve the uncertainty-propagation efficiency.

By contrast, this paper establishes a framework for improving the computational efficiency of probabilistic surge estimates by explicitly leveraging information about the numerically predicted surge. Specifically, it examines a Monte Carlo (MC) simulation for the uncertainty propagation and leverages the fact that probabilistic predictions need to be repeated across the different NHC advisories. The objective is to use the results from the MC implementation at the current advisory to improve the computational efficiency of the MC estimates for the next advisory. This can facilitate adaptive, intelligent decisions for the uncertainty propagation with no additional computational burden, established simply by using the readily available information for the numerical simulations at each advisory, and sharing information across the advisories. The computational statistics tool investigated to establish the desired improvement in the computational efficiency is importance sampling (IS) (Robert and Casella 2004; Kroese et al. 2011). IS is formulated by introducing a proposal density that focuses the MC sampling on input regions that have larger contributions to the integrand of the probabilistic integral under consideration. Through the appropriate selection of this proposal density, the variability of the MC estimator can be drastically reduced, creating smaller statistical errors in the uncertainty propagation, and therefore delivering the desired computational

efficiency (Robert and Casella 2004; Kroese et al. 2011). This selection requires knowledge of the simulation output, though, since it relies on knowledge about the entire integrand and not only about the probabilistic distribution of the uncertain input. This requirement can be seamlessly accommodated within the probabilistic surge estimation by using results from the current advisory to select the IS proposal density for the next advisory.

The novel contributions of this manuscript are: (i) the introduction of the IS formulation across storm advisories; and (ii) a number of advances to support an adaptive IS implementation within this setting, to maximize the computational benefits. The biggest challenge for this IS implementation is the need to establish a single proposal density across all QoIs, which for a typical application (large geographic domain of impact, with many locations of interest) corresponds to a high-dimensional output vector. Although IS applications have examined in detail implementations to high-dimensional inputs (Au and Beck 2003; Ehre et al. 2021), their focus is typically on a single output. Studies that have examined applications to high-dimensional output problems are very rare. The challenge in such cases originates from the fact that for each of these outputs the optimal proposal density will be different (Hesterberg 1988) and, as it will be shown later, in probabilistic storm surge estimation these densities are conflicting. For establishing a proposal density that offers a balanced compromise across them, different schemes are examined here. Emphasis is placed on reducing the numerical burden for this selection even for applications with very high-dimensional output (millions of locations of interest where surge needs to be estimated), and an approach using principal component analysis (PCA) (Jolliffe 2002) as a dimensionality reduction tool is discussed. Moreover, adaptivity is introduced in the IS implementation. Past studies (Au and Beck 1999; Medina and Taflanidis 2014) have demonstrated the importance of adaptive IS selection, established by examining different candidate IS choices and promoting the one with the best anticipated efficiency. A similar scheme is advocated here, leveraging a highly efficient global sensitivity analysis (GSA) that was recently established for probabilistic surge estimation applications (Jung et al. 2022), and an efficient prediction of the anticipated accuracy of the alternative IS implementations. For the proposal density, formulations relying on both marginal and joint distributions across the inputs are discussed as alternative options, whereas GSA is used to prioritize the inputs, so that the limited available information can be better utilized to choose densities only for the more influential inputs (Schuëller et al. 2004). To establish robustness, to accommodate the implementation across the large number of QoIs using limited information (limited number of simulations) and, more importantly, the fact that the IS density is chosen based on the current advisory but is implemented in the next one, a defensive IS scheme (Hesterberg 1995) is introduced. Finally, extensions to couch the IS formulation within a QMC framework are also discussed.

The remainder of the paper is organized as follows. Section 2 reviews the probabilistic surge estimation problem, while Section 3 discusses the basics of the MC and IS formulations. Section 4 presents all details

of the IS formulation tailored to the probabilistic surge estimation, including all necessary advances to support the high dimensionality of the output and the implementation across advisories. It also presents the proposed computational workflow. Section 5 discusses the details for the validation case studies, considering different historical storms to demonstrate robustness. Section 6 presents the validation results examining three different setups: (i) adaptive IS selection for a single advisory, (ii) IS implementation across advisories, and (iii) IS integrated with a QMC scheme.

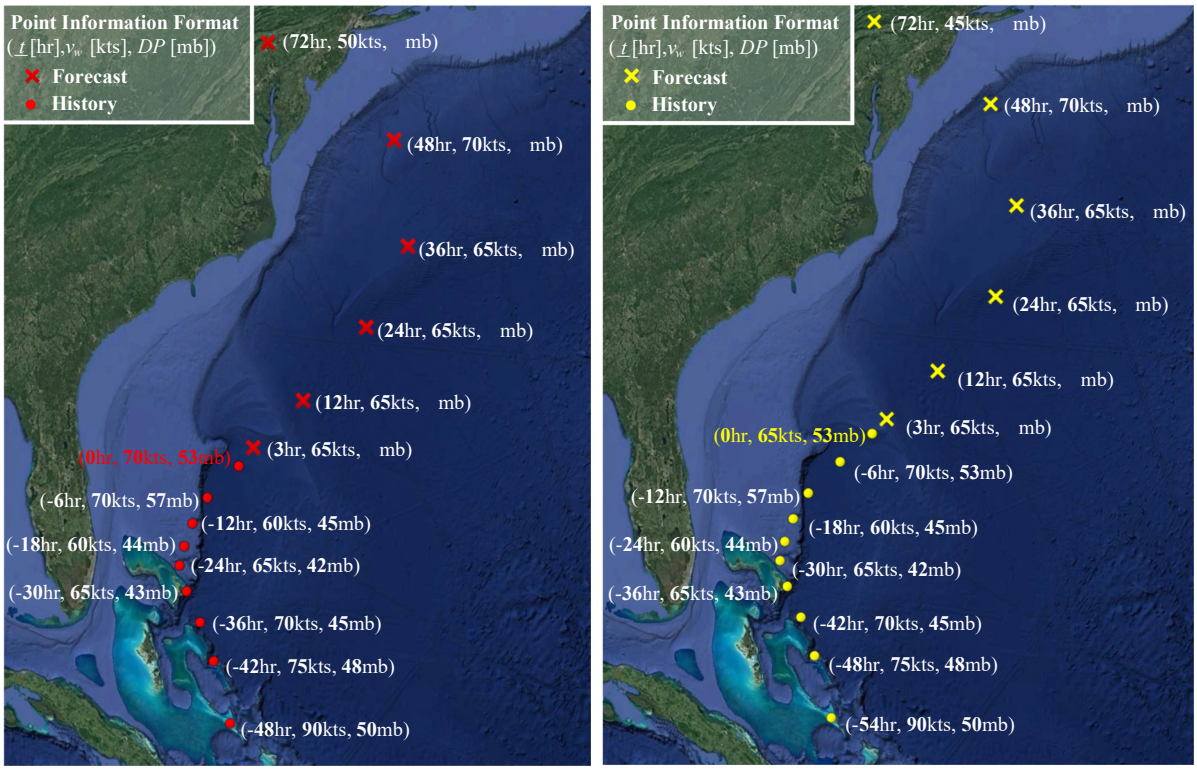
2 Formulation of probabilistic surge description for landfalling storms

This section reviews the NWS uncertainty quantification framework (Taylor and Glahn 2008; Gonzalez and Taylor 2018) for the formulation of probabilistic surge estimates. A detailed mathematical description of the framework was recently discussed in (Kyprioti et al. 2021a). Here this description is revisited, placing emphasis on establishing formalisms to accommodate the estimation and sharing of information across storm advisories, an issue not addressed in the aforementioned past studies.

2.1 Storm characterization

During landfalling storms, NHC provides advisories for the past and the forecasted track, size, and intensity characteristics. The track is described by the latitude, s_{lat}, and longitude, s_{lon}, of the storm center, the storm size by the radius of maximum winds, R_{mw}, and the storm intensity by the maximum sustained wind speed, v_w, and/or by the pressure difference between the center of the storm and the ambient pressure, DP. Information for R_{mw} might not be readily available, and in such cases it is customary to infer it based on the remaining characteristics, using functional approximations (Jelesnianski and Taylor 1973) that relate R_{mw}, v_w, and DP, using, additionally, the storm track information. This is the reason for including also information about DP in the advisories, to accommodate such an inference for R_{mw}, if necessary. The NHC advisories include information for both the history (prior evolution) of the storm characteristics as well as the forecasted storm evolution for the next few days and are updated typically in regular intervals (every 6 hr). Figure 1 shows an example of two subsequent advisories from superstorm Sandy, one of the historical storms that will be used in the case studies examined later in the paper. As also seen in this figure, NHC forecasts do not provide information for the future evolution of R_{mw} or DP, and the standard practice (Taylor and Glahn 2008) is to use the current estimate of R_{mw} for the future forecast definition. The completed NHC advisory information for the evolution of the storm track, size, and intensity can be then used by a parametric model to provide the evolution of wind and pressure fields over time. These fields can be subsequently used as forcing for a storm surge numerical model, such as the SLOSH (currently used by NHC) or ADCIRC (alternative option) models discussed in the introduction, to provide predictions for the anticipated surge.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)



(a) Hurricane Sandy (2012)-NHC advisory 22 and prior storm history

(b) Hurricane Sandy (2012)-NHC advisory 23 and prior storm history

Figure 1. Data from NHC advisories 22 [part (a)] and 23 [part (b)] for superstorm Sandy, showing past (indicated with $\underline{t} < 0$) and forecast (indicated with $\underline{t} \geq 0$) information for track and intensity.

To mathematically formalize this setup, let **q** denote the four-dimensional vector of features used to describe the size, track, and intensity of each storm:

$$\mathbf{q} = \begin{bmatrix} R_{mw} \\ s_{lat} \\ s_{lon} \\ v_w \end{bmatrix} \tag{1}$$

The variation with time of these characteristics will be described using notation $\mathbf{q}(t)$ where t denotes time. A superscript in parenthesis $.^{(k)}$ will be used hereinafter to explicitly denote, when needed, the kth storm advisory, provided at time $t^{(k)}$. Let also $\underline{t}^{(k)} = t - t^{(k)}$ denote the centered time for the kth advisory with $\underline{t}^{(k)} < 0$ corresponding to history and $\underline{t}^{(k)} \geq 0$ to forecasts. When appropriate, the superscript will be removed from $\underline{t}^{(k)}$, using $\underline{t}$ to denote the temporal aspects for the future storm evolution characteristics. The NHC advisory combines information for the history $\{\mathbf{q}(\underline{t}^{(k)}); \underline{t}^{(k)} < 0\}$ of the storm features as well as nominal (median) predictions $\{\tilde{\mathbf{q}}(\underline{t}^{(k)}); \underline{t}^{(k)} \geq 0\}$ for their future evolution (forecast) with notation $\tilde{}$ used above the vector **q** to denote the nominal predictions. To simplify discussions we will use notation $\{\tilde{\mathbf{q}}(\underline{t}^{(k)})\}$

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

to describe the *k*th NHC advisory, with the understanding that for $\underline{t}^{(k)} < 0$ the information corresponds to the actual recorded storm evolution (with a deterministic definition of all the associated storm features), and not to nominal predictions. Also shorthanded notations $\underline{t}^{(k)+}$ and $\underline{t}^{(k)-}$ are used to describe forecast $\underline{t}^{(k)} \geq 0$ and history $\underline{t}^{(k)} < 0$ estimates, respectively, for the *k*th advisory.

2.2 Uncertainty quantification for storm features

The uncertainty quantification for calculating the probabilistic surge estimates is established by accounting for forecast errors in the nominal advisory forecast $\{\tilde{\mathbf{q}}(\underline{t}^{(k)+})\}$, considering variations in the size, the intensity, and the along and cross position of the storm center relative to the nominal track (Taylor and Glahn 2008). The vector characterizing the variability of the storm features for $\underline{t} \geq 0$ is defined as:

$$\Delta\mathbf{q}(\underline{t}) = \begin{bmatrix} \Delta R_{mw}(\underline{t}) \\ \Delta s_{cross}(\underline{t}) \\ \Delta s_{along}(\underline{t}) \\ \Delta v_{w}(\underline{t}) \end{bmatrix} \tag{2}$$

where $\Delta R_{mw}(\underline{t})$ and $\Delta v_{w}(\underline{t})$ denote, respectively, the variability (error) in the size and intensity storm parameters, while $\Delta s_{along}(\underline{t})$ and $\Delta s_{cross}(\underline{t})$ denote, respectively, the along- and cross-track variability of the storm. Note that the definition of the variability $\Delta\mathbf{q}(.)$ is independent of the advisory, which is the reason that is presented as function of $\underline{t}$ and not $\underline{t}^{(k)}$. Of course, for combining with the nominal advisory, eventually $\Delta\mathbf{q}(.)$ is evaluated for the $\underline{t}^{(k)}$ values. The combination of $\{\tilde{\mathbf{q}}(\underline{t}^{(k)+})\}$ and $\Delta\mathbf{q}(\underline{t}^{(k)+})$ leads to the definition of $\{\mathbf{q}(\underline{t}^{(k)+})\}$ for the storm having forecast error $\Delta\mathbf{q}(\underline{t}^{(k)+})$ with respect to the nominal advisory $\{\tilde{\mathbf{q}}(\underline{t}^{(k)+})\}$. Combining this information with the history $\{\mathbf{q}(\underline{t}^{(k)-})\}$ provides the complete information for the storm sample scenario $\{\mathbf{q}(\underline{t}^{(k)})\}$. For deriving $\{\mathbf{q}(\underline{t}^{(k)+})\}$, size and intensity are defined directly through their respective variations, $R_{mw}(\underline{t}^{(k)+}) = \tilde{R}_{mw}(\underline{t}^{(k)+}) + \Delta R_{mw}(\underline{t}^{(k)+})$, $v_{w}(\underline{t}^{(k)+}) = \tilde{v}_{w}(\underline{t}^{(k)+}) + \Delta v_{w}(\underline{t}^{(k)+})$, while for the storm track, the combination of $\Delta s_{along}(\underline{t}^{(k)+})$ and $\Delta s_{cross}(\underline{t}^{(k)+})$ jointly influences the updated storm track $s_{lat}(t)$ and $s_{lon}(t)$ in the following way: the nominal track $\{\tilde{s}_{lat}(\underline{t}^{(k)+}), \tilde{s}_{lon}(\underline{t}^{(k)+})\}$ is varied by $\Delta s_{cross}(\underline{t}^{(k)+})$ perpendicular to its bearing to obtain a cross-path modified track; then each point of that modified track is varied along the track by $\Delta s_{along}(\underline{t}^{(k)+})$ to obtain the final track. The latter variation maintains the bearing of the cross-modified track but changes the translational speed. Figure 2 demonstrates an example, showing different storm tracks that originate from the cross-track variability of the nominal track advisory shown in part (a) of Figure 1.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

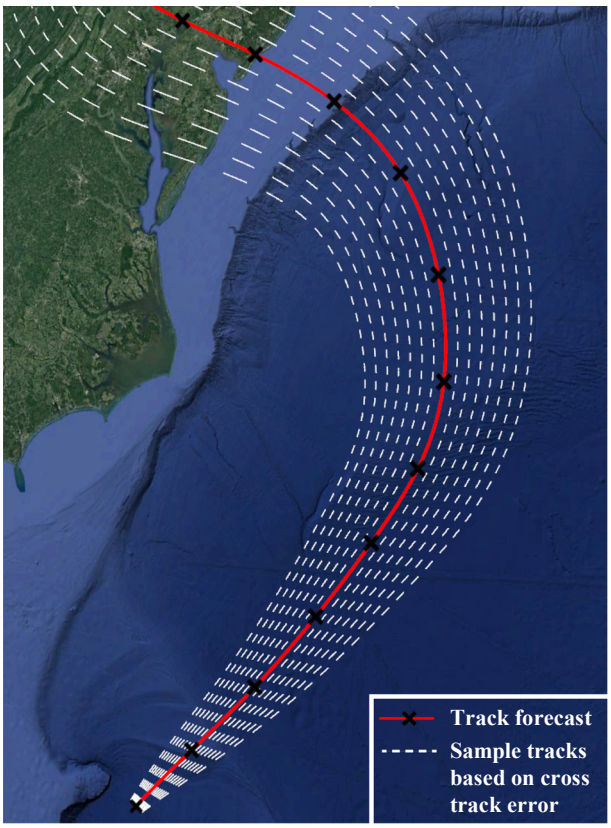


Figure 2. Storm tracks obtained considering cross-track variability of the nominal track forecast shown in part (a) of Figure 1.

The probabilistic description for $\Delta\mathbf{q}(t)$ is based on a statistical analysis of past forecast errors. The current NWS formulation is to assume: (i) independence between the four forecast errors for storm features, since they represent fundamentally different storm characteristics; and (ii) perfect correlation across times (Taylor and Glahn 2008). Note that the perfect correlation in time means that if, for example, the $v_w(\underline{t})$ value is larger by one standard deviation than its nominal value at $\underline{t} = 24$ hr, it will still be larger by one standard deviation than its median value at $\underline{t} = 48$ hr. Of course, the latter standard deviation, describing the magnitude of the variability, will change at different times, and become larger (larger forecast error) as $\underline{t}$ increases. These assumptions simplify the uncertainty description, requiring a four-dimensional vector to represent the random variables. This vector is denoted by $\mathbf{x} \in \mathbb{R}^{n_x}$ (n_x=4) herein and the probabilistic description of its independent components, $\{x_i; i = 1,...,4\}$ is discussed next.

Following (Taylor and Glahn 2008) a Gaussian probability distribution is assumed for $\Delta s_{along}(\underline{t})$, $\Delta s_{cross}(\underline{t})$ and $\Delta v_w(\underline{t})$ leading to (Kyprioti et al. 2021a):

$$\Delta q_i(\underline{t}) = \sigma_i(\underline{t}) x_i;\ i=2,...,4 \tag{3}$$

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

where x_i's are independent standard Gaussian random variables with $\sigma_i(\underline{t})$ corresponding to their standard deviation. This standard deviation represents the scaling parameter that dictates the size of the forecast error at different times and is selected based on the 5-year mean absolute error statistics, $e_i(\underline{t})$, as:

$$\sigma_i(\underline{t}) = e_i(\underline{t}) / a;\ i=2,...,4 \tag{4}$$

with a taken as 0.7979 (Gonzalez and Taylor 2018). Part (b) of Figure 3 shows examples for $e_i(\underline{t})$ based on the 2012 NHC forecast errors for hurricanes, corresponding to one of the years examined in the case studies later. It is evident that as $\underline{t}$ increases, the errors associated with the NHC forecast become larger.

For the size parameter $\Delta R_{mw}(\underline{t})$, the established probabilistic description (Taylor and Glahn 2008) corresponds to a discrete random variable representation, with three possible values representing small, medium, and large size storms, assigned to the 15th, 50th, and 85th percentiles for the storm size error conditional on the nominal storm size $\tilde{R}_{mw}(\underline{t})$. Equivalently (Kyprioti et al. 2021a), these define three possible values $\{\Delta R_{mw}^{(r)}(\underline{t}) \,|\, \tilde{R}_{mw}(\underline{t});\ r=-1,0,1\}$, with probability masses 0.3 (for r=-1), 0.4 (for r=0) and 0.3 (for r=1) linked to the 15th percentile, 50th percentile and 85th percentile storm size errors, respectively. This uncertainty description is shown in part (a) of Figure 3. To unify the uncertainty characterization across all storm features, a transformation to the standard Gaussian space is also adopted for the probabilistic description of ΔR_{mw}. This leads ultimately to the relationship:

$$\Delta R_{mw}(\underline{t}) = \begin{cases} \Delta R_{mw}^{(-1)}(\underline{t}) \,|\, \tilde{R}_{mw}(\underline{t}) \text{ if } x_1 < \Phi^{-1}(0.3) \\ \Delta R_{mw}^{(0)}(\underline{t}) \,|\, \tilde{R}_{mw}(\underline{t}) \text{ if } \Phi^{-1}(0.3) \le x_1 \le \Phi^{-1}(0.7) \\ \Delta R_{mw}^{(1)}(\underline{t}) \,|\, \tilde{R}_{mw}(\underline{t}) \text{ if } x_1 > \Phi^{-1}(0.7) \end{cases} \tag{5}$$

where x_1 is a standard Gaussian variable and $\Phi(.)$ denotes the standard Gaussian cumulative distribution function.

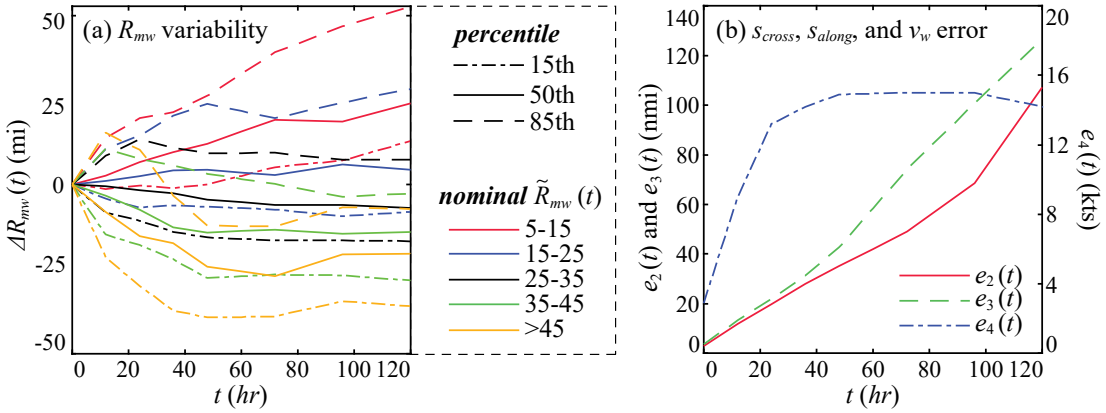


Figure 3. Uncertainty description as a function of time for the storm characteristics: (a) the 15th, 50th and 85th percentiles for the radius of maximum winds for different nominal radius values; (b) 5-year mean absolute error for the cross-, along-track variation and the wind speed, for the 2012 NHC forecast error description.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

2.3 Probabilistic surge estimates

The uncertainty description is ultimately facilitated through the four-dimensional random variable vector $\mathbf{x}$, with components $\{x_i; i=1,...,4\}$ following (independent) standard Gaussian distributions. Knowledge of $\mathbf{x}$ provides $\Delta\mathbf{q}(\underline{t}^{(k)+})$ through Eqs. (3) and (5) which, as explained in Sections 2.1 and 2.2, can be combined with the nominal forecasts $\{\tilde{\mathbf{q}}(\underline{t}^{(k)+})\}$ and the storm history $\{\mathbf{q}(\underline{t}^{(k)-})\}$ to define the input, $\{\mathbf{q}(t)\}$, to a hydrodynamic numerical model to predict the surge across the geographic domain the storm impacts. We are interested in the statistics of the peak surge across n_z discretized locations within this domain, with n_z corresponding to a large number in typical applications. Notation z_j will be used to denote this surge at location j and notation $z_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})$ to explicitly denote its dependence on the deviation $\mathbf{x}$ from the nominal storm predictions for the kth advisory $\tilde{\mathbf{q}}^{(k)} = \{\tilde{\mathbf{q}}(\underline{t}^{(k)})\}$. Mathematical formalism "|" is used herein to denote conditional relationships. Note that in an effort to simplify the notation, the temporal dependence has been removed in the $\tilde{\mathbf{q}}^{(k)}$ description.

To support emergency response decisions, different statistics of interest for the peak storm surge can be estimated within this uncertainty quantification setting (Hamill et al. 2012; Gonzalez and Taylor 2018), with the most popular one being the probability that the surge will exceed a specific threshold b. Using the total probability theorem, this probability for the kth advisory and jth location is expressed as:

$$P_j^{(k)}(b) = \int I[z_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)}) > b]\,p(\mathbf{x})\,d\mathbf{x} \tag{6}$$

where $I[.]$ corresponds to the indicator function, which is one if the quantity inside the brackets is satisfied, else it is zero, and $p(\mathbf{x})$ corresponds to the probability distribution model of $\mathbf{x}$ (standard Gaussian as detailed earlier). The inverse problem can provide another interesting statistical quantity, the surge corresponding to a specific probability p_t of exceedance, which for the kth advisory is given by:

$$b_j^{(k)p_t} \text{ such that } P_j^{(k)}(b_j^{(k)p_t}) = p_t \tag{7}$$

To generalize discussions so that different probabilistic surge estimates can be easily accommodated, let $h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})$ denote the consequence measure associated with the statistics of interest, and

$$H_j^{(k)} = E_p[h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})] = \int h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})\,p(\mathbf{x})\,d\mathbf{x} \tag{8}$$

the probabilistic integral defining these statistics, where $E_p[.]$ represents expectation under probability model $p(\mathbf{x})$. For example, for the probability of exceedance given by Eq. (6) $h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)}) = I[z_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)}) > b]$, whereas for the expected (mean) surge $h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)}) = z_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})$. The objective of the uncertainty propagation, which is the focus of the remainder of this paper, is the efficient estimation of the four-

dimensional integral of Eq. (8). This will be accomplished using MC simulation tools. Readers interested in the estimation of this integral through factorial sampling or QMC techniques are referred to (Kyprioti et al. 2021a).

3 Monte Carlo and importance sampling fundamentals

The MC estimator for the integral of Eq. (8) is formulated by considering a set $\{\mathbf{x}^l : l = 1,...,N\}$ of N independent identically distributed (i.i.d) samples for $\mathbf{x}$ from a proposal density $f^{(k)}(\mathbf{x})$, with $\mathbf{x}^l$ denoting the lth sample. In the context of the probabilistic surge estimation, the set $\{\mathbf{x}^l : l = 1,...,N\}$ represents the ensemble of N storms utilized to estimate the desired statistical products. Note that a dependence of the proposal density on (k) is used herein, since that density can change across advisories. Using the N i.i.d samples, $\mathbf{x}^l \sim f^{(k)}(\mathbf{x})$, with $\sim$ representing distribution proportional to, the MC estimator, for the kth advisory and jth QoI, denoted $\hat{H}_j^{(k)}$, is obtained as (Robert and Casella 2004; Kroese et al. 2011):

$$\hat{H}_j^{(k)}(f^{(k)}) = \frac{1}{N}\sum_{l=1}^{N} h_j(\mathbf{x}^l \mid \tilde{\mathbf{q}}^{(k)}) \frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)} \tag{9}$$

where the dependence of the estimator on the proposal density selection is explicitly denoted using notation $\hat{H}_j^{(k)}(f^{(k)})$. Provided that the expression $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)}) p(\mathbf{x}) / f^{(k)}(\mathbf{x})$ is bounded, meaning that the support of $f^{(k)}(\mathbf{x})$ is greater than the support of the integrand $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)}) p(\mathbf{x})$, the estimator $\hat{H}_j^{(k)}(f)$ is unbiased, with coefficient of variation, $\delta_j^{(k)}$, representing its statistical accuracy as an approximation of $H_j^{(k)}$, given by (Robert and Casella 2004; Kroese et al. 2011):

$$\delta_j^{(k)}(f) = \frac{1}{\sqrt{N}} \frac{\sqrt{Var_{f^{(k)}}[h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)}) p(\mathbf{x}) / f^{(k)}(\mathbf{x})]}}{H_j^{(k)}} \tag{10}$$

where $Var_{f^{(k)}}[.]$ denotes variance under the probability model $f^{(k)}(\mathbf{x})$.

The simplest choice for the proposal density is to use $f^{(k)}(\mathbf{x}) = p(\mathbf{x})$, which corresponds to a direct MC estimation. In MC with IS, the density $f^{(k)}(\mathbf{x})$, termed IS density in this case, is chosen so that the MC sampling concentrates in regions of the input $\mathbf{x}$ that correspond to larger values (i.e. with higher importance) for the integrand $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)}) p(\mathbf{x})$ that is associated with $H_j^{(k)}$ (Robert and Casella 2004; Kroese et al. 2011). The IS objective is to reduce the variance appearing in the numerator of Eq. (10), and therefore improve the accuracy of the estimator $\hat{H}_j^{(k)}(f^{(k)})$. As also evident from Eq. (10) the variance of the estimator $\hat{H}_j^{(k)}(f^{(k)})$ is $Var_{f^{(k)}}[h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)}) p(\mathbf{x}) / f^{(k)}(\mathbf{x})] / N$, implying that the improvement in accuracy can be

achieved either by an increase in the number of simulations N, which entails an increase in the computational effort, or by a decrease of the variance $Var_{f^{(k)}}[h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})p(\mathbf{x})/f^{(k)}(\mathbf{x})]$ through an appropriate selection of $f^{(k)}(\mathbf{x})$. IS attempts to achieve this improvement through the latter venue, selecting $f^{(k)}(\mathbf{x})$ to minimize $Var_{f^{(k)}}[h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})p(\mathbf{x})/f^{(k)}(\mathbf{x})]$. The optimal IS density, establishing this minimization, is (Kroese et al. 2011):

$$f_j^*(\mathbf{x}|\tilde{\mathbf{q}}^{(k)}) = \frac{|h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})|\,p(\mathbf{x})}{\int |h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})|\,p(\mathbf{x})d\mathbf{x}} \propto |h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})|\,p(\mathbf{x}) \tag{11}$$

where $\propto$ denotes proportionality, and the chosen notation for the optimal density is intended to stress that it is a function of both the nominal NHC advisory [dependence on $\tilde{\mathbf{q}}^{(k)}$] and the location [dependence on j]. Note that the benefits from the IS implementation will be greater when the statistics examined correspond to infrequent events, leading to larger differences between $p(\mathbf{x})$ and $f_j^*(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})$ (Robert and Casella 2004; Kroese et al. 2011).

Using directly the optimal density of Eq. (11) within the IS scheme is impractical since it requires the knowledge of the entire integrand. Sample-based approximations of this optimal density will be examined instead. Such approximations have been shown to be highly efficient for selecting IS densities in other types of applications (Au and Beck 1999; Medina and Taflanidis 2014). The sample-based IS is formulated, as detailed in Section 4.1, by fitting a probability density function to samples distributed proportionally to the optimal density of Eq. (11).

A key component in establishing improved accuracy when utilizing such sample-based approximations for the IS density is some form of adaptive selection, established by examining different candidate choices $f^c(\mathbf{x})$ for the IS density and promoting the most appropriate one (Medina and Taflanidis 2014). This requires an estimation of the anticipated MC accuracy if $f^c(\mathbf{x})$ were used as proposal density. To achieve this objective using the readily available simulations, which were performed using $f^{(k)}(\mathbf{x})$ as the sampling density, the approximated variance is first expressed as:

$$Var_{f^c}\left[h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x})}{f^c(\mathbf{x})}\right] = E_{f^c}\left[\left(h_j(\mathbf{x}|\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x})}{f^c(\mathbf{x})}\right)^2\right] - \left(H_j^{(k)}\right)^2 \tag{12}$$

Then, the second moment appearing in Eq. (12), the only component that depends on $f^c(\mathbf{x})$, can be calculated using MC with the sample set $\{\mathbf{x}^l : l = 1,...,N\} \sim f^{(k)}(\mathbf{x})$ as (Medina and Taflanidis 2014):

$$E_{f^c}\left[\left(h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x})}{f^c(\mathbf{x})}\right)^2\right]=\int\left(h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x})}{f^c(\mathbf{x})}\right)^2 f^c(\mathbf{x})d\mathbf{x}$$
$$=\int\left(h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})\right)^2\frac{p(\mathbf{x})}{f^c(\mathbf{x})}\frac{p(\mathbf{x})}{f^{(k)}(\mathbf{x})}f^{(k)}(\mathbf{x})d\mathbf{x} \qquad (13)$$
$$\approx\frac{1}{N}\sum_{l=1}^{N}\left(h_j(\mathbf{x}^l\,|\,\tilde{\mathbf{q}}^{(k)})\right)^2\frac{p(\mathbf{x}^l)}{f^c(\mathbf{x}^l)}\frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)}$$

The minimization of the variance $Var_{f^c}[h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})p(\mathbf{x})/f^c(\mathbf{x})]$, which is the IS objective, corresponds ultimately to the minimization of the second moment $E_{f^c}[(h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})p(\mathbf{x})/f^c(\mathbf{x}))^2]$ that can be approximated through Eq. (13). This will be leveraged in the developments discussed in the next section.

Beyond the mathematical description it is important to stress that based on the coefficient of variation of Eq. (10), the reduction of the IS estimator variance of Eq. (12) leads to a proportional reduction of computational effort to establish the same statistical error. As such this variance should be interpreted as being directly proportional to the MC-based computational burden, or, equivalently, to the achieved convergence rate as N increases. This feature is essential for assessing the improvements established through the proposed adaptive IS implementation.

4 Adaptive IS formulation for probabilistic surge estimation

The implementation of the sample-based IS formulation discussed in Section 3 couched within the probabilistic surge estimation has the following three challenges: (i) the necessity to choose the sample-based proposal densities using limited information since the number of simulations N is typically small (Kyprioti et al. 2021a) due to the large computational burden of the numerical models used to predict the storm surge; (ii) the need to establish a compromise across conflicting outputs with each of them leading to a different optimal density [dependence of optimal density of Eq. (11) on j]; (iii) the requirement to share information across the storm advisories, basing the selection of the proposal density for the next advisory on the responses available for the current advisory, which correspond to different nominal track characteristics [dependence of optimal density of Eq. (11) on (k)]. In Sections 4.1-4.3, each of these challenges is separately discussed, with the overall computational framework presented in Section 4.4.

The framework is established by choosing IS densities utilizing the storm surge simulations from the current (k) advisory, with the intension to use these simulations for the next (k+1) advisory. This means that simulations $\{h_j(\mathbf{x}^l\,|\,\tilde{\mathbf{q}}^{(k)}):l=1,...,N\}$ for the current advisory obtained using sample set $\{\mathbf{x}^l:l=1,...,N\}\sim f^{(k)}(\mathbf{x})$ are readily available to support the IS selection. Note that for the first advisory $f^{(1)}(\mathbf{x})=p(\mathbf{x})$ since no adaptive IS selection has been yet established. We will denote as $\varphi(\mathbf{x})=p(\mathbf{x})/f^{(k)}(\mathbf{x})$ the ratio of $p(\mathbf{x})$ to the proposal density used at the current advisory. Let also

$\mathbf{H}^{(k)} \in \mathbb{R}^{N \times n_z}$ and $\mathbf{X} \in \mathbb{R}^{N \times n_x}$ denote the observation and input matrices defined through this information, with the *l*th row of each matrix corresponding, respectively, to the output (across all surge locations) and input for the *l*th simulation. Also let $\{r_j : j = 1,...,n_z\}$ denote user-defined priority weights assigned to each QoI, dictating the relative importance of establishing higher statistical accuracy for each QoI. These weights will be utilized when aggregating information across the QoIs. The vector of priority weights will be denoted as $\mathbf{r} \in \mathbb{R}^{n_z}$.

4.1 Sample-based approximation of the optimal density

Let $\pi(\mathbf{x})$ denote the target density we want to approximate. This density may correspond directly to the optimal IS density $f_j^*(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)})$ given by Eq. (11), though different options will be discussed in Section 4.2. The readily available samples $\{\mathbf{x}^l : l = 1,...,N\} \sim f^{(k)}(\mathbf{x})$, provide information for $\pi(\mathbf{x})$ if a weight $w(\mathbf{x}) = \pi(\mathbf{x}) / f^{(k)}(\mathbf{x})$ is attached to each of them (Kroese et al. 2011). Let $\{w^l : l = 1,...,N\}$ denote the weight-set across the sample set, with $w^l = w(\mathbf{x}^l)$. Utilizing the sample/weight pair $\{(\mathbf{x}^l, w^l) : l = 1,...,N\}$ the sample-based approximation for $\pi(\mathbf{x})$ can be established by fitting some distribution. Two different approaches can be employed for this task: (i) resampling first the weighed set to obtain samples for $\pi(\mathbf{x})$ with no weights, and then establishing a distribution fit for these samples or; (ii) using directly the weighted samples for the fit, with the weights representing the relative likelihood of each sample. In all examples considered in this manuscript, the latter implementation is adopted, since it was found to accommodate higher robustness.

With respect to the type of distribution, either non-parametric or parametric approximations can be considered in this setting (Silverman 1998). Non-parametric densities, such as kernel approximations, will face challenges since the number *N* of available samples will be generally small. For this purpose, parametric densities are promoted here. Specifically, a Gaussian Mixture Model (GMM) formulation (McNicholas and Murphy 2008) is chosen, an approach shown in previous studies (Kurtz and Song 2013; Geyer et al. 2019) to be highly appropriate for sample-based IS approximations. The GMM fit is established using maximum likelihood estimation, implemented through the Expectation-Maximization (EM) algorithm (Moon 1996; McNicholas and Murphy 2008), whereas the weights for the samples $\{w^l : l = 1,...,N\}$ are accommodated using the likelihood adjustments outlined in (Geyer et al. 2019). Interested readers can find numerical implementation details for the EM algorithm in the latter study.

To improve the robustness of the distribution fit, both marginal and joint density formulations are examined, and, additionally, the consideration of IS densities only for the most influential input components is considered. The motivation for examining these alternative options is the same: address the fact that limited information is available (small number of samples *N*) for the IS density selection. It has been shown

in a number of studies (Schuëller et al. 2004; Medina and Taflanidis 2014; Jia et al. 2017), that the formulation of sample-based proposal densities has the potential to lead to overfitting in such settings, when the number of available samples to describe the densities is small. The prioritization of input components or the use of marginal densities promotes the extraction of a smaller amount of information from the available sample set, something that can accommodate greater robustness in the IS selection. At the same time, such choices may reduce the IS efficiency. For this reason, different candidate IS formulations are established, then the efficiency for each of them is approximated using Eq. (13), and the most appropriate one is chosen. For the input prioritization, an efficient, recently proposed GSA (Jung et al. 2022) is leveraged. Utilizing matrices $\mathbf{H}^{(k)}$ and $\mathbf{X}$ along with vector $\mathbf{r}$ and set $\{\varphi(\mathbf{x}^l) = p(\mathbf{x}^l) / f^{(k)}(\mathbf{x}^l) : l = 1,...,N\}$ this GSA implementation provides aggregated importance indicators for the first-order sensitivity indices for input $\mathbf{x}$, quantifying the importance of each x_i towards the observed surge variability. Note that these indicators are established across all QoIs, and ultimately represent the proportion of the total variance explained by the variations in each x_i. The IS formulation can then focus on only the important components. If $\mathbf{x}_s$ denotes the important input components and $\mathbf{x}_{\sim s}$ the remaining ones, then the candidate IS density is $f^c(\mathbf{x}) = f^c(\mathbf{x}_s)p(\mathbf{x}_{\sim s})$ and entails a choice of only $f^c(\mathbf{x}_s)$ for the lower dimensional $\mathbf{x}_s$, since for the remaining input components the original distribution $p(\mathbf{x}_{\sim s})$ is utilized [no IS formulation]. The selection of $f^c(\mathbf{x}_s)$ can be established for the entire vector $\mathbf{x}_s$ (joint formulation) or independently for each of its components (marginal formulation). The terms IS_M (marginal) and IS_T (joint) will be used herein to distinguish these two implementations. The algorithmic implementation of these concepts will be presented in detail in Section 4.4, after additional details for the IS formulation are discussed.

4.2 Selection of IS densities across conflicting outputs

As mentioned earlier, each of the examined outputs (QoIs) has a different optimal density, $f_j^*(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)})$, and these densities can be conflicting with one another. Figures 4 and 5 illustrate this challenge, showing the mean for $f_j^*(x_i \mid \tilde{\mathbf{q}}^{(k)})$ for each of the forecast errors looking at two different advisories of superstorm Sandy (Figure 4) and hurricane Gustav (Figure 5). Should be noted that this mean is of great importance for the IS selection as it dictates the shift of the proposal density towards the peak of the probabilistic integrand. The probabilistic estimation in both these figures pertains to the 10% exceedance probability $P_j^{(k)}(b_j^{(k)10\%})$. It is evident from the results that even with respect to simple statistics (mean) of $f_j^*(x_i \mid \tilde{\mathbf{q}}^{(k)})$, let alone the distributions themselves, great variability exists within the geographic domain of storm impact, indicating that the different QoIs represent conflicting decisions. For example, note that some QoIs move the original zero mean Gaussian distribution towards positive values while others towards negative values. These conflicts are expected to reduce, unfortunately, the overall IS efficiency, since a

compromising IS density will need to be implemented, which can be quite different from the actual optimal density for each output. Moreover, the selection of this density that balances the conflicting objectives needs to be carefully executed, while also accommodating applications with a large number of outputs. Two different formulations are examined to satisfy these requirements.

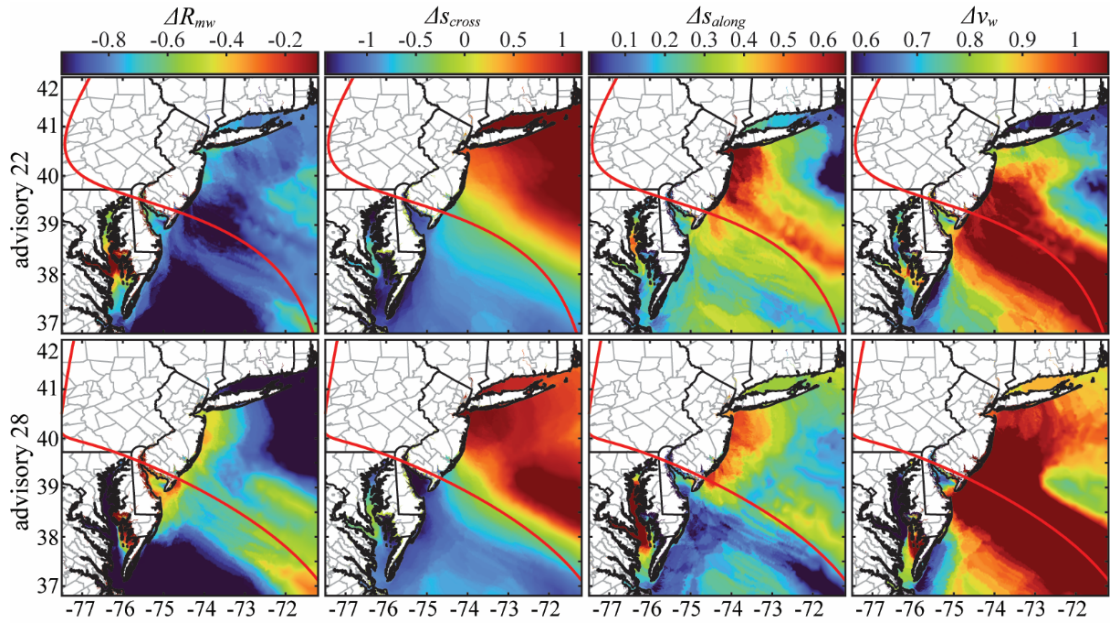


Figure 4. Distribution of mean of the forecast error optimal IS density corresponding to 10% exceedance probability statistics for advisories 22 and 28 of superstorm Sandy.

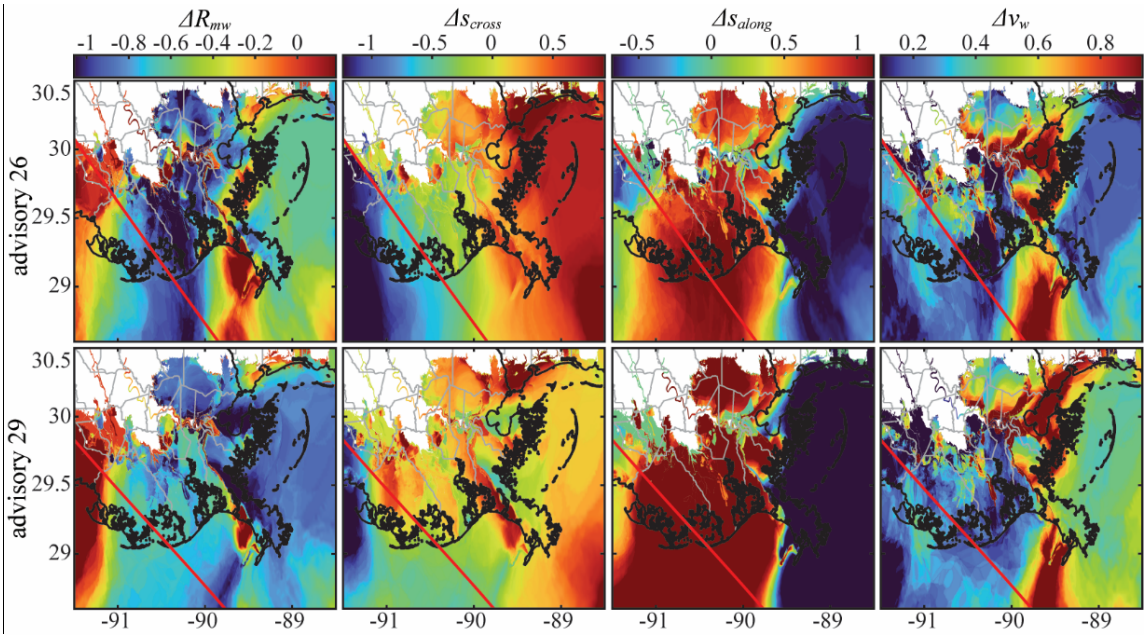


Figure 5. Distribution of mean of the forecast error optimal IS density corresponding to 10% exceedance probability statistics for advisories 26 and 29 of hurricane Gustav.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

The first formulation utilizes PCA as a dimensionality reduction technique. PCA identifies a low-dimensional vector of latent outputs, also called principal component vector, that best explains the variability of the original data $\mathbf{H}^{(k)}$ (Jolliffe 2002). Each principal component $\{y_v; v = 1,\ldots,n_v\}$ has an associated importance λ_v, representing the variance of the data explained by the component. To incorporate the priority weights r_j for each output and the fact that each observation is associated with weight $\{\varphi(\mathbf{x}^l); l = 1,...,N\}$, a weighted (generalized) PCA is implemented (Greenacre 1984; Jolliffe 2002), with details reviewed in Appendix A. Note that the incorporation of the observation weights $\{\varphi(\mathbf{x}^l); l = 1,...,N\}$ leads ultimately to a PCA that examines the variance of the observations under probability model $p(\mathbf{x})$, instead under the proposal density $f^{(k)}(\mathbf{x})$ [if these weights were not used], maintaining a focus on the original integrand. As detailed in Appendix A, this incorporation leads to adjustment of observations by $\sqrt{\varphi(\mathbf{x}^l)}$. To accommodate the dimensionality reduction, only the more important components are retained, with the exact number chosen so that these retained components explain a significant portion of the original output variance (say over 99%). This process leads to values of n_v in the range of 20-50 for the application of interest, as will be discussed in the case study examples. The output matrix for the principal components for the N available simulations is denoted by $\mathbf{Y}^{(k)} \in \mathbb{R}^{N \times n_v}$ and can be obtained through the implementation discussed in Appendix A. For the y_v component, the responses across the simulations $\{y_v(\mathbf{x}^l); l = 1,\ldots,N\}$ are given by the vth column of matrix $\mathbf{Y}^{(k)}$. The optimal IS density for each component is taken to correspond to the integrand representing its expected value under probability model $p(\mathbf{x})$, leading to:

$$f_v^*(\mathbf{x}^l \mid \mathbf{H}^{(k)}) \propto \frac{\left|y_v(\mathbf{x}^l)\right|}{\sqrt{\varphi(\mathbf{x}^l)}} p(\mathbf{x}^l) \tag{14}$$

where the division by $\sqrt{\varphi(\mathbf{x}^l)}$ incorporates the fact that the generalized PCA implementation introduced, equivalently, such weights for each observation, and that these weights need now to be balanced out to obtain the proportionality to the intended integrand. Note that the dependence of the IS density on $\mathbf{H}^{(k)}$ [impacting the entire PCA implementation] is explicitly noted. Each of the n_v densities given by Eq. (14) represents a different target density $\pi_v(\mathbf{x})$ that can be approximated through the sample-based approach discussed in Section 4.1 (algorithmic details discussed later in Section 4.4). Subscript $_v$ is used herein to distinguish the target density defined for each principal component. Since the number of such densities is small (value of n_v is small), a separate fitting to each of these densities can be examined, leading to an approximation $\hat{f}_v^*(\mathbf{x} \mid \mathbf{H}^{(k)})$ for each. The final candidate IS density is obtained by a weighted sum of the individualized IS densities for each principal component (Hesterberg 1988), with weights corresponding to λ_v. Since the latter weights represent the variance of the principal components, and the IS objective is to

reduce the MC variability which is connected to the output variance, selecting them as weights is the recommended choice. Principal components with larger variance should have proportionally increased importance in guiding the IS selection. Note that the priority weights, r_j for each output have been incorporated in the PCA, and therefore influence the eigenvalues λ_v. This leads to the following candidate density definition:

$$f^c(\mathbf{x}\,|\,\mathbf{H}^{(k)}) = \frac{\sum_{v=1}^{n_v} \lambda_v \hat{f}_v^*(\mathbf{x}\,|\,\mathbf{H}^{(k)})}{\sum_{v=1}^{n_v} \lambda_v} \tag{15}$$

An alternative approach can be accommodated if the averaging across the principal components densities is established based on the actual optimal densities, and not their approximations. This leads to target IS density:

$$f_{PCA}^*(\mathbf{x}\,|\,\mathbf{H}^{(k)}) = \frac{\sum_{v=1}^{n_v} \lambda_v f_v^*(\mathbf{x}\,|\,\mathbf{H}^{(k)})}{\sum_{v=1}^{n_v} \lambda_v} \propto \sum_{v=1}^{n_v} \lambda_v \frac{\left|y_v(\mathbf{x}^l)\right|}{\sqrt{\varphi(\mathbf{x}^l)}}\, p(\mathbf{x}) \tag{16}$$

representing the target density $\pi(\mathbf{x})$. This density can be then directly approximated through the sample-based implementation discussed in Section 4.1. The two alternative PCA-based, formulations will be distinguished using notation *PCA$_{in}$-IS* for the approach using sample-based approximation to the IS density of each individual component, having candidate IS density given by Eq. (15), and notation *PCA-IS* for the approach using sample-based approximation to the weighted density across the components, with candidate density given by the sample-based approximation for the target density in Eq. (16). *PCA-IS* entails a smaller computational burden, since it involves only one GMM-fit, though *PCA$_{in}$-IS* represents a selection with greater versatility (since the approximation is implemented for individual components), and therefore greater potential IS efficiency.

The second formulation for accommodating the selection of proposal densities across the conflicting outputs does not include any dimensionality reduction step. In this case, establishing IS densities for each of the outputs and then combining them, an approach similar to the *PCA$_{in}$-IS* formulation, is impractical, since the number n_z of such densities will be prohibitively large. Instead, the individual densities given by Eq. (11) are first combined to define the target IS density:

$$f^*(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)}) = \frac{\sum_{j=1}^{n_z} r_j \gamma_j f_j^*(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})}{\sum_{j=1}^{n_z} r_j \gamma_j} \propto \sum_{j=1}^{n_z} r_j \gamma_j \left|h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})\right| p(\mathbf{x}) \tag{17}$$

where γ_j corresponds to the additional weights provided for each density, beyond the priority weights associated with each QoI. Similar to the PCA formulation, the weights γ_j are chosen equal to the variance of each of the outputs, since the objective of the IS is the minimization of the MC variability, which is

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

connected to the output variance as explained earlier. This indicates that outputs with larger variance should have proportionally increased importance in guiding the IS selection, which is what the formulation of Eq. (17) with weights corresponding to the output variance accommodates. The priority weights r_j incorporate, additionally the relative importance of reducing the estimator variance for each QoI. The definition of the weights γ_j, as well as their estimation using the readily available samples, is:

$$\begin{aligned}\gamma_j &= Var_p[h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})] \\ &\approx \frac{1}{N}\sum_{l=1}^{N}\left(h_j(\mathbf{x}^l\,|\,\tilde{\mathbf{q}}^{(k)})\right)^2\frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)} - \left(\frac{1}{N}\sum_{l=1}^{N}h_j(\mathbf{x}^l\,|\,\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)}\right)^2\end{aligned} \tag{18}$$

The use of Eq. (18) in Eq. (17) completely defines the IS target density $\pi(\mathbf{x})$ which can be then approximated through the sample-based implementation discussed in Section 4.1. This formulation denoted as *OrO-IS* (original output IS) resembles the *PCA-IS* implementation, though in this case there is no requirement to transform to the latent output space to establish the combination of densities. In this context, the distinction is that Eq. (17) [*OrO-IS*] establishes the combination with respect to the original output of interest, whereas Eq. (16) [*PCA-IS*] with respect to the latent output.

One final topic that needs to be examined is the assessment of the proposal density efficiency across the different outputs. For a specific output, this efficiency is characterized by the variance of Eq. (12) which can be approximated using the result of Eq. (13). For multiple outputs, the weighted average should be used instead. Utilizing a normalization with respect to the weighted average MC estimator variability (without IS) the following efficiency measure is defined for the *k*th advisory:

$$F^{(k)}[f^c(\mathbf{x})] = 1 - \frac{\frac{1}{n_z}\sum_{j=1}^{n_z} r_j Var_{f^c}\left[h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x})}{f^c(\mathbf{x})}\right]}{\frac{1}{n_z}\sum_{j=1}^{n_z} r_j Var_p[h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})]} \tag{19}$$

This efficiency represents the reduction of computational effort accommodated by the proposal density $f^c(\mathbf{x})$ for establishing the same average statistical accuracy (same average variance) across the outputs as the direct MC estimation. For example, value 0.1 means that computational effort can be reduced by 10% while maintaining the same statistical accuracy. Positive values represent reduction of computational cost (improvement), while negative values represent increase of computational cost.

For accommodating the desired adaptive IS implementation, an MC-estimate of this measure using the readily available samples $\{\mathbf{x}^l : l = 1,...,N\} \sim f^{(k)}(\mathbf{x})$, can be established. This estimate, termed herein as *efficiency approximation*, is:

$$\hat{F}^{(k)}[f^c(\mathbf{x})] = 1 - \frac{\frac{1}{n_z}\sum_{j=1}^{n_z} r_j \left[\frac{1}{N}\sum_{l=1}^{N} \left(h_j(\mathbf{x}^l \mid \tilde{\mathbf{q}}^{(k)}) \right)^2 \frac{p(\mathbf{x}^l)}{f^c(\mathbf{x}^l)} \frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)} - \left(\frac{1}{N}\sum_{l=1}^{N} h_j(\mathbf{x}^l \mid \tilde{\mathbf{q}}^{(k)}) \frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)} \right)^2 \right]}{\frac{1}{n_z}\sum_{j=1}^{n_z} r_j \left[\frac{1}{N}\sum_{l=1}^{N} \left(h_j(\mathbf{x}^l \mid \tilde{\mathbf{q}}^{(k)}) \right)^2 \frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)} - \left(\frac{1}{N}\sum_{l=1}^{N} h_j(\mathbf{x}^l \mid \tilde{\mathbf{q}}^{(k)}) \frac{p(\mathbf{x}^l)}{f^{(k)}(\mathbf{x}^l)} \right)^2 \right]} \quad (20)$$

Note that both the efficiency definition of Eq. (19) and its approximation given by Eq. (20) can be provided for any desired proposal density selection $f^c(\mathbf{x})$, something explicitly denoted in their functional dependence notation.

4.3 Robustness of IS densities across advisories

The last issue that needs to be addressed is the fact that the selection of proposal densities is based on the current, *k*th advisory, as evident by the dependence on $\tilde{\mathbf{q}}^{(k)}$ or $\mathbf{H}^{(k)}$ of the target densities presented in Section 4.2, but is employed in the next (*k*+1)th advisory which might correspond to different nominal track characteristics $\tilde{\mathbf{q}}^{(k+1)}$. Loss of efficiency, but most importantly loss of robustness, may arise if the promoted IS density does not sufficiently cover the important regions (for efficiency) or support (for robustness) of the updated integrand $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k+1)}) p(\mathbf{x})$. Robustness is a bigger concern here, since if the support of $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k+1)}) p(\mathbf{x})$ is not a subset of the support of the promoted IS density, the IS estimator will be biased (Robert and Casella 2004; Kroese et al. 2011). A defensive IS formulation (Hesterberg 1995) is adopted to improve robustness. If $f_{IS}^c(\mathbf{x})$ denotes the IS density approximation formulated based on any of the approaches outlined in Section 4.2, then the defensive IS, $f_{DIS}^c(\mathbf{x} \mid \alpha)$, is obtained as a combination of $f_{IS}^c(\mathbf{x})$ and $p(\mathbf{x})$:

$$f_{DIS}^c(\mathbf{x} \mid \alpha) = \alpha f_{IS}^c(\mathbf{x}) + (1-\alpha) p(\mathbf{x}) \quad (21)$$

with $0 \le \alpha \le 1$ representing the defensive IS weight. The selection of Eq. (21) as IS density guarantees that the support of $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k+1)}) p(\mathbf{x})$ is a subset of the support of $f_{DIS}^c(\mathbf{x} \mid \alpha)$ [or, equivalently, that the quotient $h_j(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k+1)}) p(\mathbf{x}) / f_{DIS}^c(\mathbf{x})$ is bounded], and therefore guarantees unbiased MC predictions.

The choice of α needs to be established with care, since small values of α can lead to a significant reduction in IS efficiency, with the defensive IS density $f_{DIS}^c(\mathbf{x})$ deviating significantly from the promoted IS density $f_{IS}^c(\mathbf{x})$. An adaptive selection is suggested here for α by comparing the (reduced) efficiency $\hat{F}^{(k)}(f_{DIS}^c)$ of $f_{DIS}^c(\mathbf{x})$ to the (optimal) efficiency $\hat{F}^{(k)}(f_{IS}^c)$ of $f_{IS}^c(\mathbf{x})$ [both approximated through Eq. (20)], and selecting an appropriate defensive weight α so that the resulting reduction for $\hat{F}^{(k)}(f_{DIS}^c)$ when

compared to $\hat{F}^{(k)}(f_{IS}^{c})$ is less than some threshold ρ. This choice corresponds to the greater robustness that can be infused based on the tolerance for reduction of IS efficiency.

Note that this defensive IS scheme enhances robustness not only for the sharing information across the storm advisories, but additionally for accommodating decisions using small sample sets across a large number of QoIs. This provides an additional safeguard beyond the ones discussed in Sections 4.1 and 4.2 to support the IS formulations when using sample-based approximations with a limited number of samples while trying to balance across conflicting QoIs. Due to the aforementioned characteristics the defensive IS, $f_{DIS}^{c}(\mathbf{x}\,|\,\alpha)$, given by Eq. (21) might outperform the original IS $f_{IS}^{c}(\mathbf{x})$ for some α values. For this reason, a further improvement of the optimal IS is first established, selecting α to maximize the *efficiency approximation* given by Eq. (20):

$$\alpha^{*} = \arg\max_{\alpha\in[0,1]} \hat{F}^{(k)}[f_{DIS}^{c}(\mathbf{x}\,|\,\alpha)] \tag{22}$$

Identification through Eq. (22) of α^{*}=1 means that $f_{IS}^{c}(\mathbf{x})$ outperforms any defensive IS scheme, whereas identification of α^{*}<1 means that there is a defensive density that outperforms, on average, the sample-based approximation of the optimal IS density identified earlier. The $f_{DIS}^{c}(\mathbf{x}\,|\,\alpha^{*})$ is ultimately the promoted optimal IS, and then a further decrease in α is considered to provide further robustness within the defensive IS scheme, as discussed in the previous paragraph.

4.4 Workflow for the IS density formulation

Combining the concepts discussed in Sections 4.1-4.3, the IS workflow for the probabilistic storm surge estimation during landfalling storms is established. Note that different variants exist for this workflow, depending on: (i) whether sample-based approximations are established for the marginal (IS_M) or joint (IS_T) distributions; and (ii) what target density selection (PCA_{in}-*IS,* *PCA-IS*, or *OrO-IS*) is made for accommodating the high-dimensional output. Below the detailed workflow for the PCA_{in}-*IS* considering both IS_M and IS_T formulations is presented first, with the differences in certain steps for accommodating the other variants discussed later on. At the *k*th advisory, the steps for the adaptive IS selection are the following.

Step 1 [priority ranking]: Using the matrices $\mathbf{H}^{(k)}$ and $\mathbf{X}$ along with the priority vector $\mathbf{r}$ and the set $\{\varphi(\mathbf{x}^{l}) = p(\mathbf{x}^{l}) / f^{(k)}(\mathbf{x}^{l}) : l = 1,...,N\}$, estimate the aggregated first-order importance indicators for $\mathbf{x}$ following the implementation described in detail in (Jung et al. 2022). This provides a ranking of the four input components x_i based on their importance. This step can be skipped, and a generic importance

ranking can be used instead, since as shown in (Jung et al. 2022), x_2 has always the highest importance, with x_1 and x_4 following with similar importance, and x_3 having typically the lowest importance.

Step 2 [dimensionality reduction]: Perform PCA for matrix $\mathbf{H}^{(k)}$ with priority output weights $\mathbf{r}$ and sample weights $\{\varphi(\mathbf{x}^l) = p(\mathbf{x}^l) / f^{(k)}(\mathbf{x}^l) : l = 1,...,N\}$, and retain n_v components so that the explained variance of Eq. (A.1) is greater than a threshold ψ_o. Obtain the matrix of latent responses $\mathbf{Y}^{(k)}$ (details in Appendix A) and eigenvalues λ_v.

Step 3 [target density definition]: Define the target density for each principal component $\pi_v(\mathbf{x})$ to correspond to $f_v^*(\mathbf{x} \mid \mathbf{H}^{(k)})$ given by Eq. (14). Estimate the weights $w_v(\mathbf{x}) = \pi_v(\mathbf{x}) / f^{(k)}(\mathbf{x})$ for the sample set $\{\mathbf{x}^l : l = 1,...,N\}$, providing the weight set $\{w_v^l : l = 1,...,N\}$. Repeat this for each principal component.

Step 4 [sample-based approximation of target density]: Utilizing the sample/weight set $\{(\mathbf{x}^l, w_v^l) : l = 1,...,N\}$ from Step 3, perform a GMM fit to obtain a sample-based approximation of $f_v^*(\mathbf{x} \mid \mathbf{H}^{(k)})$, denoted as $\hat{f}_v^*(\mathbf{x} \mid \mathbf{H}^{(k)})$. Define the final target density $\hat{f}^*(\mathbf{x} \mid \mathbf{H}^{(k)})$ by Eq. (15).

Set counter n=1.

Step 5 [IS input definition]: Define $\mathbf{x}_s$ to correspond to the n most important components of $\mathbf{x}$ based on the priority ranking of Step 1.

Step 6 [candidate density definition]: Define the candidate proposal density for the joint (IS_T) formulation as $f_{T_n}^c(\mathbf{x}) = p(\mathbf{x}_{\sim s})\hat{f}^*(\mathbf{x}_s \mid \mathbf{H}^{(k)})$ and for the marginal (IS_M) formulation as $f_{M_n}^c(\mathbf{x}) = p(\mathbf{x}_{\sim s})\prod_{s_i=1}^{n}\hat{f}^*(x_{s_i} \mid \mathbf{H}^{(k)})$ where x_{s_i} denotes the s_i element of vector $\mathbf{x}_s$. For n=1, only $f_{M_n}^c(\mathbf{x})$ needs to be considered (the two densities are identical).

Step 7 [candidate density refinement and efficiency approximation]: Consider the defensive IS density given by Eq. (21), with $f_{IS}^c(\mathbf{x})$ corresponding either to $f_{M_n}^c(\mathbf{x})$ or $f_{T_n}^c(\mathbf{x})$, and investigate whether an additional refinement can be established by identifying an optimal defensive weight α based on Eq. (22). This process provides the final optimal densities (i) $f_{DM_n}^c(\mathbf{x} \mid \alpha^*)$, denoting the refined marginal defensive density starting with $f_{M_n}^c(\mathbf{x})$, and (ii) $f_{DT_n}^c(\mathbf{x} \mid \alpha^*)$, denoting the refined joint defensive density starting with $f_{T_n}^c(\mathbf{x})$, as well as their respective approximated efficiencies, given by Eq. (20).

Step 8 [repeat across input definitions]: if $n<n_x$, update $n=n+1$ and repeat Steps 5-8.

Step 9 [selection of best density]. Considering the efficiency of refined densities $\{f_{DM_n}^c(\mathbf{x} \mid \alpha^*); n = 1,...,n_x\}$ and $\{f_{DT_n}^c(\mathbf{x} \mid \alpha^*); n = 2,...,n_x\}$ across all repeated Steps 7, choose the final IS

density $f_{IS}^{c}(\mathbf{x})$ as the one with the best approximated efficiency. This density will be denoted herein as *promoted-IS*.

Step 10 [robustness enhancement through defensive IS scheme]. Consider again the defensive IS density $f_{DIS}^{c}(\mathbf{x}\,|\,\alpha)$ given by Eq. (21), implemented for the density $f_{IS}^{c}(\mathbf{x})$ identified in Step 9, and select defensive weight α^c so that the efficiency reduction of $f_{DIS}^{c}(\mathbf{x}\,|\,\alpha)$ compared to $f_{IS}^{c}(\mathbf{x})$ is ρ. Note that if $f_{IS}^{c}(\mathbf{x})$ corresponds to $\alpha^*<1$, then the final $f_{DIS}^{c}(\mathbf{x}\,|\,\alpha^c)$ is merely formulated as a further decrease of α. The corresponding density $f_{DIS}^{c}(\mathbf{x}\,|\,\alpha^c)$ will be denoted herein as *robust-IS*, and represents the IS proposal density for the next advisory $f^{(k+1)}(\mathbf{x})$.

This ten-step algorithm will be denoted as AIS-SA (adaptive importance sampling across storm advisories) herein. It is important to note that the computational cost of AIS-SA primarily originates from the GMM fit and the dimensionality reduction, and this cost is very small when compared to the cost for the numerical simulations to estimate the storm surge. By sharing information across the advisories, AIS-SA requires no additional such simulations, making the overall adaptive implementation very efficient. More importantly, in real-time operational setting, AIS-SA can be implemented once statistical products have been delivered for a specific advisory, while waiting the next advisory, imposing, therefore, no additional operational cost, and allowing all available computational resources to be used for the storm surge simulations (to provide the desired statistical products).

An alternative implementation for the sample-based approximation can be established if the fit is performed for each of the densities appearing in the $f_{T_n}^{c}(\mathbf{x})$ and $f_{M_n}^{c}(\mathbf{x})$ expressions in Step 6 separately. To accomplish this, the order of Steps 4 and 5 needs to be flipped, and the GMM fit in the new Step 5 (previous Step 4) is established considering either the sample/weight set $\{(\mathbf{x}_s^l, w_v^l): l=1,...,N\}$ to obtain approximation $\hat{f}_v^*(\mathbf{x}_s\,|\,\mathbf{H}^{(k)})$ or the sample/weight set $\{(x_{s_i}^l, w_v^l): l=1,...,N\}$ to obtain approximation $\hat{f}_v^*(x_{s_i}\,|\,\mathbf{H}^{(k)})$ for each component of $\mathbf{x}_s$. This alternative formulation involves a larger computational burden as it requires a larger number of GMM distribution fits compared to the single fit required in the original formulation.

For the other two IS variants, some adjustments are needed in some of the algorithmic steps. For the *PCA-IS* implementation an update is needed for Steps 3 and 4 as follows:

Step 3 [target density definition]: Define the target density across the principal components to correspond to $f_{PCA}^{*}(\mathbf{x}\,|\,\mathbf{H}^{(k)})$ given by Eq. (16). Estimate the weights $w(\mathbf{x})=f_{PCA}^{*}(\mathbf{x}\,|\,\mathbf{H}^{(k)})/f^{(k)}(\mathbf{x})$ for the sample set $\{\mathbf{x}^l: l=1,...,N\}$, providing weight set $\{w^l: l=1,...,N\}$.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

Step 4 [sample-based approximation of target density]: Utilizing the sample/weight set $\{(\mathbf{x}^l, w^l) : l = 1,..., N\}$ from Step 3, perform a GMM fit to obtain the sample-based approximation of $\hat{f}^*(\mathbf{x} \mid \mathbf{H}^{(k)})$.

For the *OrO-IS*, Step 2 is removed, all conditionings on $\mathbf{H}^{(k)}$ are replaced by a conditioning on $\tilde{\mathbf{q}}^{(k)}$, whereas in Steps 3 and 4, the target density across the principal components $f^*_{PCA}(\mathbf{x} \mid \mathbf{H}^{(k)})$ is replaced by the target density across outputs $f^*(\mathbf{x} \mid \tilde{\mathbf{q}}^{(k)})$ given by Eq. (17).

5 Illustrative case studies overview

5.1 Case study characteristics

Three different historical storms are examined as illustrative case studies, corresponding to: hurricane Gustav (2008), hurricane Irene (2011), and superstorm Sandy (2012). These storms are chosen to offer a comprehensive demonstration examining both landfalling (Gustav and Sandy) and bypassing (Irene) storms, across different geographic regions, that include both the Gulf of Mexico (Gustav) and the North Atlantic (Irene and Sandy). For each storm, different series of NHC advisories will be utilized: (i) for Gustav, advisories 26-29 (in total 4 advisories) are used, corresponding to a period range roughly from 48 hr to 24 hr before the storm makes landfall, (ii) for Sandy advisories 22-28 (in total 7 advisories) are used, corresponding to a period range roughly from 72 hr to 24 hr before landfall, while (iii) for Irene advisories 26-30 (in total 5 advisories) are used, corresponding to a period range roughly from 48 hr to 24 hr before the storm bypasses New York and gradually starts losing its strength. The earliest advisory (with the smallest indexing number), corresponding to the furthest time from landfall, will be denoted by $A^{(1)}$ and each subsequent advisory will be presented with an increasing superscript number. The tracks for some of the advisories for all these storms are shown in Figure 6.

According to the current NWS probabilistic framework (Gonzalez and Taylor 2018), for each advisory, the available information corresponds to: the storm track, $DP(\underline{t})$ and $v_w(\underline{t})$ for the past hurricane history $\underline{t}^{(k)} < 0$ and the storm track and $v_w(\underline{t})$ for future forecasts $\underline{t}^{(k)} \geq 0$. Based on the provided *DP* and v_w, the storm size R_{mw} is estimated for the hurricane history. The current estimate for $\underline{t}^{(k)} = 0$, $R_{mw}(0)$, is kept constant for future predictions ($\underline{t}^{(k)} \geq 0$), representing the nominal storm size forecast. For the relationship among v_w, *DP* and R_{mw}, results from the study (Knaff and Zehr 2007) are utilized.

As detailed earlier, the proposed advances are independent of the underlying physical model that is used for the storm surge simulations (for the peak surge predictions). For the case studies examined here, a surrogate model of ADCIRC is chosen for this purpose. The use of a surrogate model instead of ADCIRC is necessitated by the extensive case studies considered and the need to estimate reference solutions for the

estimated surge, to validate the IS formulation. Details for the development of the surrogate models for the North Atlantic (used for hurricanes Irene and Sandy) and for the Louisiana region (used for hurricane Gustav) can be found in (Kyprioti et al. 2021b) and (Jia et al. 2016), respectively. The surrogate models were developed using databases of ADCIRC simulations for the two aforementioned regions of interest and are available by the U.S Army Corps of Engineers through their Coastal Hazards System (Nadal-Caraballo et al. 2020). The accuracy of the established surrogate models is very high, having a correlation coefficient with respect to the original database of over 98.5%, providing a very high degree of confidence for their use within the case studies considered here. The part of the ADCIRC grid within the domain of impact for these storms includes n_z=1,860,021 nodes for the North Atlantic case studies and n_z=1,552,341 for the New Orleans case studies. These numbers showcase the large dimensionality of the examined output, stressing the IS formulation challenges.

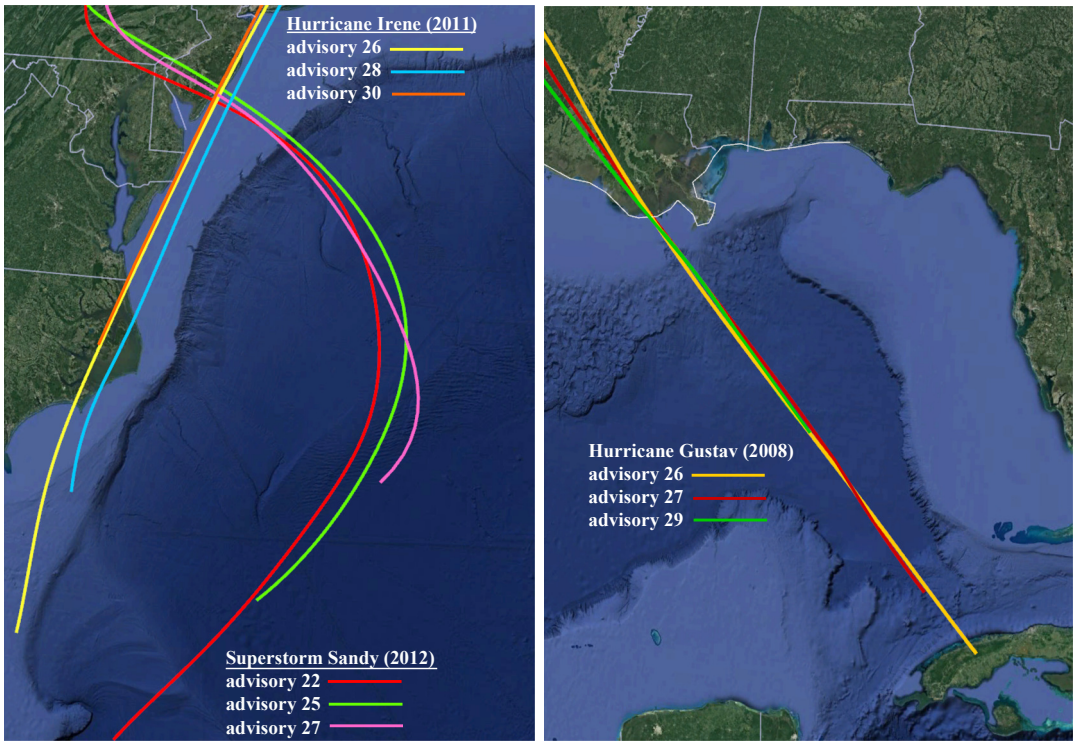


Figure 6. Storm tracks for the case study storms for hurricanes Irene (2011), Sandy (2012) [left column], and Gustav (2008) [right column].

5.2 Case study computational details

Unless otherwise defined, the number of samples for the direct MC and IS implementations is chosen to be N=500. The efficiency reduction for the *robust-IS* selection is set to ρ=0.1. Five different statistics are considered to quantify the probabilistic performance, encompassing the typical statistics used in probabilistic surge forecasting (Kyprioti et al. 2021a) corresponding to: (i) the surge threshold corresponding to probability of exceedance $P_j^{(k)}(b_j^{(k)p_t})$ for four different values of p_t=[0.01, 0.05, 0.1, 0.2],

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

denoted herein as S_{p_t} [for example $S_{0.01}$ corresponds to $P_j^{(k)}(b_j^{(k)p_t})$ for p_t=0.01]; (ii) the mean surge $E_p[z_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})]$, denoted herein as S_m. As discussed in section 2.3 these statistics lead, ultimately, to different definitions of the consequence measure $h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})$. The statistics corresponding to lower p_t values represent events with less frequent characteristics, whereas S_m and statistics corresponding to higher p_t values focus on the mean response behavior. Also even though these definitions can be combined to derive a formulation of a single IS density that could optimally balance across all of them, an independent implementation (adaptive IS for each of them separately) will be considered here. Unless otherwise specified, no preference is provided for any nodes, with priority weights, r_j, set equal to 1 for all nodes.

For facilitating the validation of the results, reference estimates are obtained using a large number of quasi-random samples (i.e. N_r=5000) by QMC. Let $\{\mathbf{x}^r : r = 1,...,N_r\} \sim p(\mathbf{x})$ denote this sample set. The reference estimation pertains to both the probabilistic integral of Eq. (8) as well as the efficiency of the IS density of Eq. (19), given, respectively, by:

$$H_j^{(k)} = \frac{1}{N_r}\sum_{l=1}^{N_r} h_j(\mathbf{x}^r\,|\,\tilde{\mathbf{q}}^{(k)}) \tag{23}$$

$$F^{(k)}[f^c(\mathbf{x})] = 1 - \frac{\frac{1}{n_z}\sum_{j=1}^{n_z} r_j\left[\frac{1}{N_r}\sum_{l=1}^{N_r} h_j{}^2(\mathbf{x}^r\,|\,\tilde{\mathbf{q}}^{(k)})\frac{p(\mathbf{x}^r)}{f^c(\mathbf{x}^r)} - \left(H_j^{(k)}\right)^2\right]}{\frac{1}{n_z}\sum_{j=1}^{n_z} r_j\left[\frac{1}{N_r}\sum_{l=1}^{N_r} h_j{}^2(\mathbf{x}^r\,|\,\tilde{\mathbf{q}}^{(k)}) - \left(H_j^{(k)}\right)^2\right]} \tag{24}$$

Note that both of these equations correspond to QMC-based estimates of the actual probabilistic quantities, but since the value of N_r is very large, they are considered herein as the reference results. When needed to draw distinctions, the estimate of the efficiency using Eq. (20) and the N readily available simulation samples will be referenced, as mentioned earlier, as *efficiency approximation* whereas the estimate using Eq. (24) and the large number of N_r QMC samples will be referenced as *actual efficiency*. It should be stressed that the quantities in Eqs. (23) and (24) are calculated here simply to accommodate the intended validation, and are not available in practical applications. This is the reason that the adaptive IS selection in AIS-SA is based on the *efficiency approximation*.

For the *PCA$_{in}$-IS* and *PCA-IS* formulations, threshold ψ_o is set to 99%, leading to the identification of number of principal components in the range of 20-50, depending on the statistic examined.

6 Illustrative case studies: results and discussion

To establish a comprehensive assessment of the IS implementation, three different setups are examined, with increasing levels of complexity, starting with the adaptive IS selection for a single advisory, and then extending to the implementation across advisories.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

6.1 Adaptive IS selection for a single advisory

In this section, the IS performance within each advisory is examined. This setup accommodates comprehensive comparisons for a variety of topics related to the IS implementation for probabilistic surge estimation, circumventing challenges that may arise by the fact that differences in the nominal storm characteristics may occur between advisories. The implementation and performance across advisories will be examined in the next section. Specifically, the focus here is on the potential IS benefits, the efficiency comparisons across the different IS variants, as well as the challenges associated with the implementation across different locations (different QoIs) or for different statistics of interest.

Results are presented for advisory $A^{(1)}$ for all three storms as well as advisory $A^{(4)}$ for hurricane Gustav. All three IS variants are considered here, *PCA_{in}-IS* (IS density of individual principal components), *PCA-IS* (IS density across principal components), or *OrO-IS* (IS density across original outputs), whereas results are initially presented for IS densities considering different number of input components, n=1,…,4, according to the AIS-SA implementation presented in Section 4.4 [repetition of Steps 5-8], for both the marginal (independent IS densities for each input component) and joint IS (single IS density across all input components) formulations. Even though a specific n value and formulation are promoted in Step 9 of the algorithm, the discussion here focuses on the intermediate results. For notational simplicity, each of the densities will be denoted herein as M_n [for the marginal IS_M formulation] or T_n [for the joint IS_T formulation], where subscript $._n$ denotes the number of input components considered in the IS formulation. Results for T_1 are not presented, since this is identical to the M_1 case. Initially, the discussion focuses on the *promoted-IS* density in Eq. (21) for the α value found in Eq. (22).

The IS implementation is initially examined for two different statistics, the surge thresholds corresponding to 1% exceedance probability $S_{0.01}$ and 10% exceedance probability $S_{0.1}$ with results presented, respectively in Figures 7 and 8. Each figure presents the approximated and actual IS efficiency in Eqs. (20) and (24), respectively, with different marker styles and colors. Results are presented for the cases discussed in the previous paragraph: for each of the three IS variants (columns of figures) and for different formulations of the IS density (discrete cases in the x-axis). In each figure, results are presented for the different storms and advisories (rows of figures).

Focusing on the overall IS efficiency first, we can observe significant improvements in accuracy, and therefore proportional reduction in computational burden (in most instances reduction of computational burden by 1.5 to 2 times can be achieved). The benefits are greater for statistics corresponding to less frequent events; this is showcased here by a bigger improvement for $S_{0.01}$ compared to $S_{0.1}$ [compare same variant implementation between Figures 7 and 8]. This is expected. As discussed earlier, since IS densities will have greater differences than $p(\mathbf{x})$ for rarer events, contributing to greater potential benefits offered by the IS implementation. The overall degree of improvement shows significant variability across storms and

advisories. This degree depends on a range of attributes, for example on how well the actual optimal IS density is approximated by the proposed GMM formulation, and on how the provided information allows for the accurate identification of this optimal IS, with the most important being, the required degree of compromise across the different surge locations for selecting the IS density. This will be, evidently, case dependent.

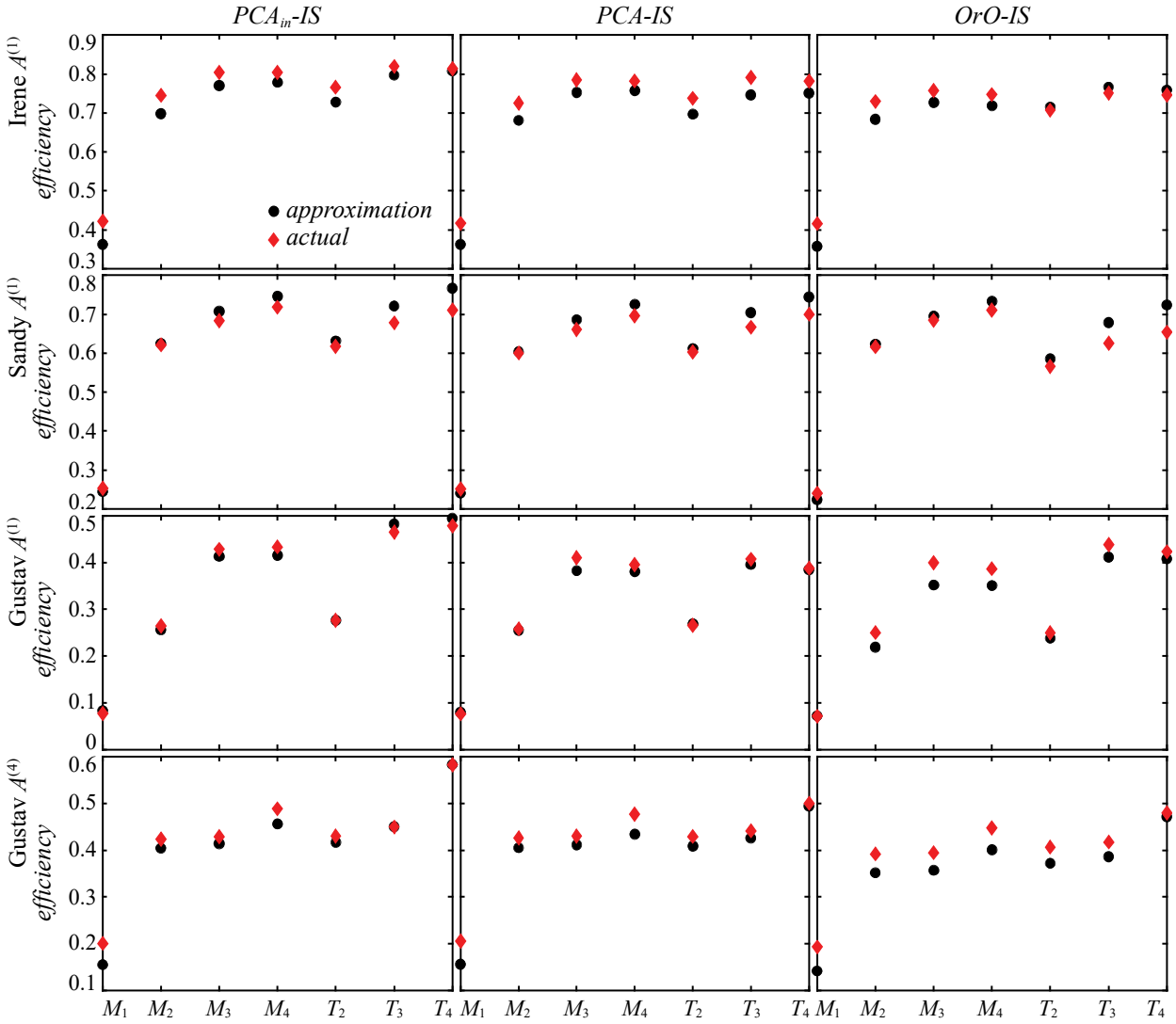


Figure 7. Approximated and actual IS efficiency for a specific advisory for different IS variants (columns of figure) and different formulations for the IS density. Results are presented for different storms and advisories (rows of figure) and correspond to the $S_{0.01}$ statistic. Robust IS formulation is not considered in these results.

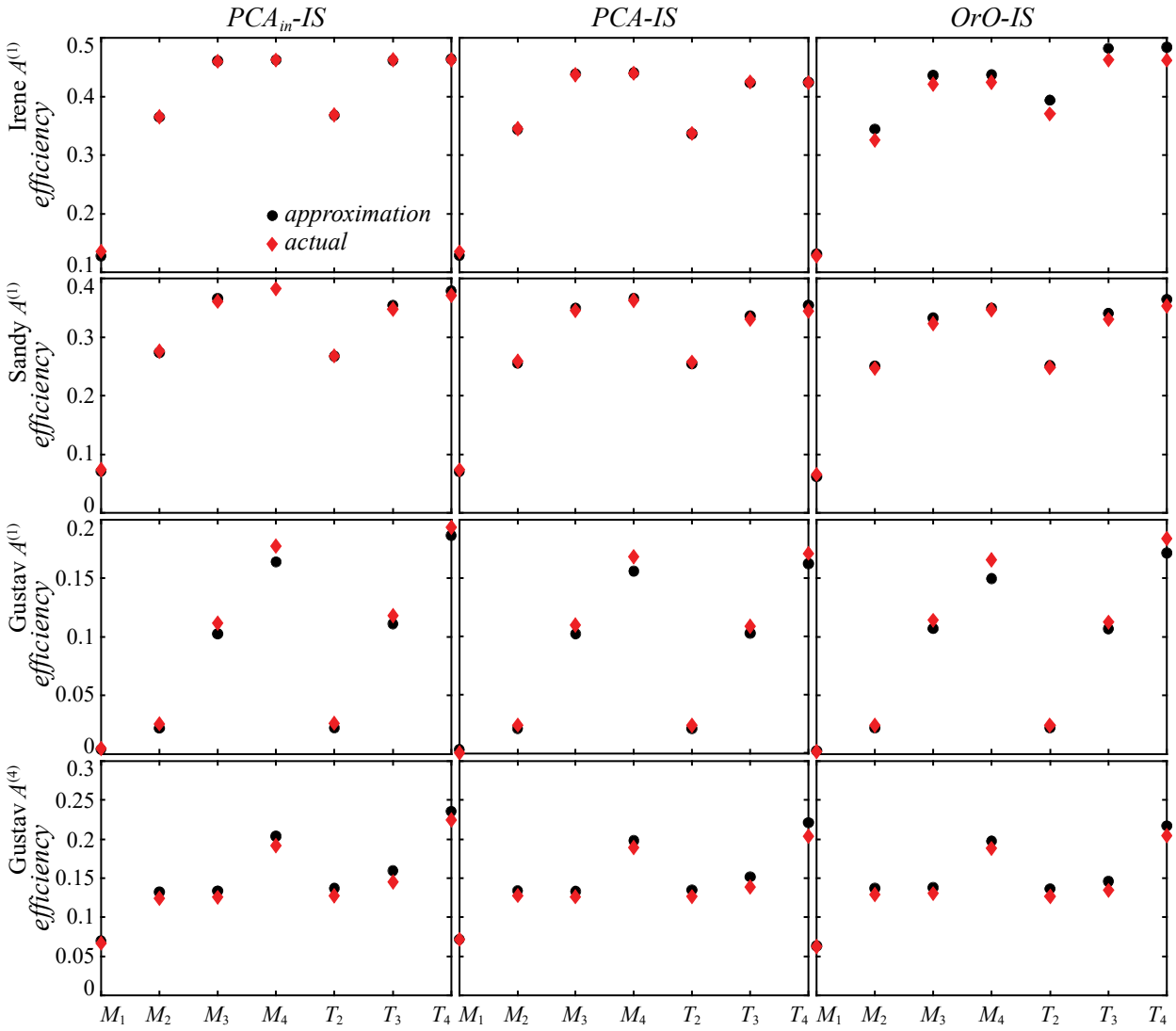


Figure 8. Approximated and actual IS efficiency for a specific advisory for different IS variants (columns of figure) and different formulations for the IS density. Results are presented for different storms and advisories (rows of figure) and correspond to the $S_{0.1}$ statistic. Robust IS formulation is not considered in these results.

Examining, next, the differences in the efficiency as the number of IS input variables increases (for both the marginal M_n or joint formulations T_n), results show variability across the different advisories, with n=4 emerging consistently, as the preferable implementation. To better frame the influence of the number of IS input variables, the GSA results (performed in the first step of the adaptive IS formulation) need to be considered. The GSA for these advisories identified consistently Δs_{cross} as the most influential input, with Δv_w and ΔR_{mw} following with similar values of the first order Sobol' indices, and Δs_{along}, having smaller importance. For hurricane Irene that has bypassing characteristics, Δs_{along} is identified to have very small importance, agreeing with the trends reported in (Jung et al. 2022). It is therefore no surprise that no difference, and perhaps some efficiency loss, is reported when moving from n=3 to n=4 for Irene (including Δs_{along} in the IS formulation). Another interesting trend is the larger relative efficiency improvement when

moving from n=1 to n=2 (including Δv_w or ΔR_{mw} in the IS formulation), as opposed to when moving from n=0 (no IS implementation) to n=1 (including only Δs_{cross} in the IS formulation). Despite Δs_{cross} being the dominant input with respect to global sensitivity characteristics, it is not the most advantageous to include in the IS formulation. The reason for it is easily understood if one looks at the results in Figures 4 and 5 earlier. For Δs_{cross} the optimal IS characteristics exhibit significant higher variability within the geographic domain, demonstrating clear competing preferences (some nodes promote values substantially less than 0 and some other nodes values substantially larger than 0). Even though Δs_{cross} is more influential as individual input (higher GSA aggregated importance indicators), the competing optimal IS characteristics for it and the need, as discussed earlier, to establish a compromise across them, reduce its relative influence within the IS formulation, allowing input variables with smaller variability with respect to the IS characteristics (see Figures 4 and 5) to emerge as more advantageous options. If a single QoI was examined, then the advantages associated with the IS formulation for different input variables would be in complete agreement with the GSA results. These trends can be, additionally, leveraged to understand the smaller efficiency improvement offered by IS for hurricane Gustav: the optimal IS characteristics for this hurricane for the influential parameters ΔR_{mw} and Δs_{along} exhibit significant larger variability within the domain of interest, indicating IS with greater degree of competing preferences, something that ultimately reduced the efficiency for the promoted, compromising, IS solution.

Comparing across the IS variants (different columns in the figures), results show that all yield similar performance with the PCA-based (*PCA_{in}-IS, PCA-IS*) outperforming the original output based *OrO-IS*, and the IS density approximation considering individual principal components *PCA_{in}-IS* outperforming IS density across principal components *PCA-IS*. This shows that the proposed dimensionality reduction (through PCA) accommodates improved efficiency, allowing IS to better leverage the available limited information for the adaptive selection of the IS proposal densities. The superiority of *PCA_{in}-IS* over *PCA-IS* is also expected, as discussed in Section 4.2, and the improvements margins, which for some cases (meaning advisories or storms) are not insignificant, justify its use despite the higher computational complexity (as also detailed in Section 4.2). Overall, this discussion reveals a preference in terms of both efficiency and robustness for *PCA_{in}-IS*, whereas discussion in this and the previous paragraph demonstrate the benefits of the proposed adaptive implementation, examining different choices and promoting the one with the highest projected efficiency.

To examine some additional trends, Figures 9 and 10 replicate Figure 8, but each one investigates a different feature of the IS formulation. Figure 9 presents results using a smaller number of N=100 simulations, with the objective to examine the impact of the amount of available information on the adaptive IS selection. Figure 10 introduces priority weights, r_j, in the implementation, to examine the robustness of the AIS-SA when certain QoIs appear to be more influential. In this case, the priority weights are chosen

equal to the surge variance for each location (approximated through the MC results), providing higher priority to locations that exhibit higher response variability. For the results in Figure 10, the default value of N=500 is used.

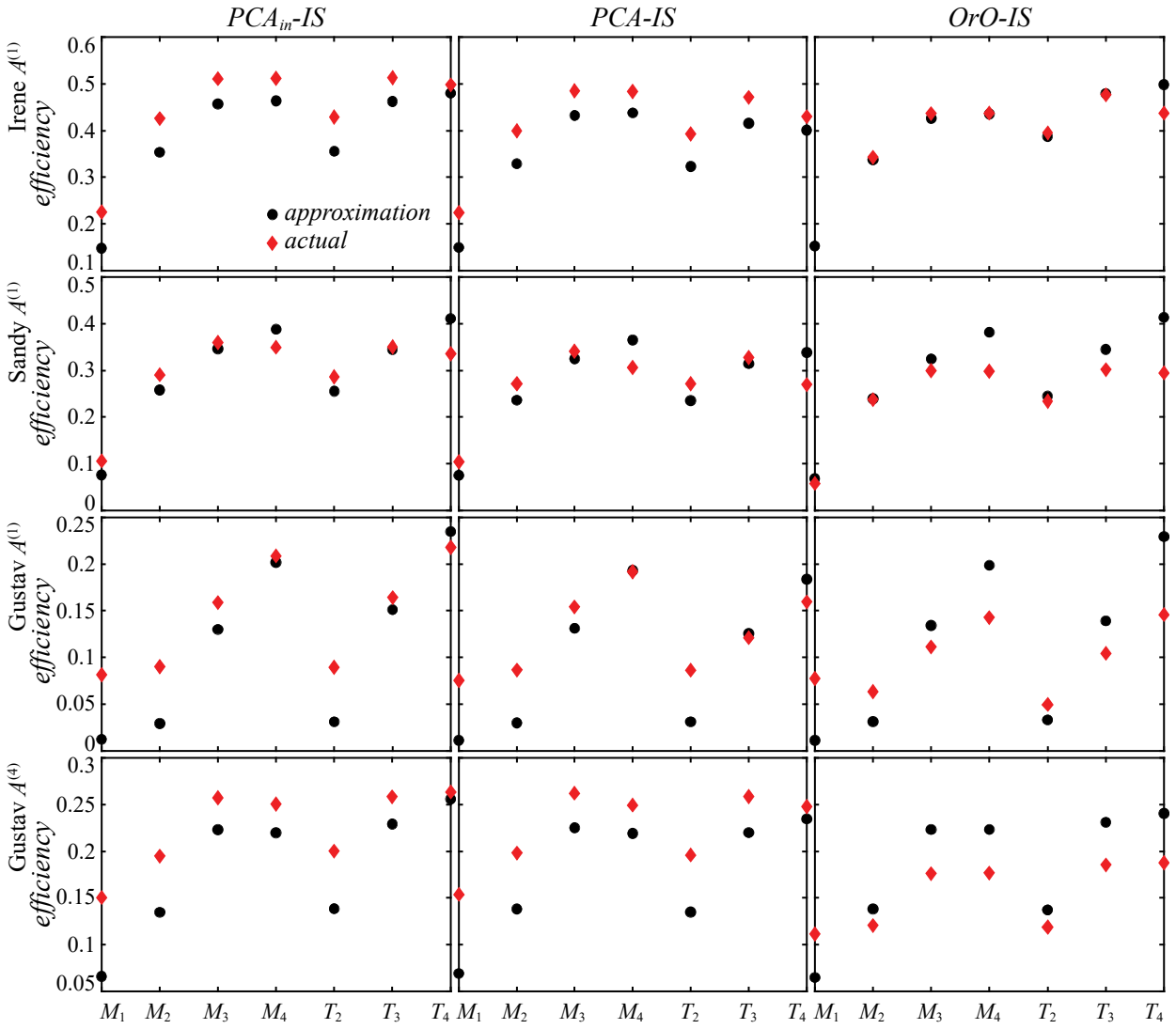


Figure 9. Approximated and actual IS efficiency for a specific advisory for different IS variants (columns of figure) and different formulations for the IS density. Results are presented for different storms and advisories (rows of figure) and correspond to the $S_{0.1}$ statistic, similar to Figure 8, but for N=100 number of simulations.

Examining the behavior of the marginal IS_M and joint IS_T density formulations across Figures 7, 8 and 9 one immediately observes similar performance, with the marginal IS_M implementation exhibiting some slightly better average performance, even though in some instances IS_T might yield higher efficiency. The better performance for the IS_M family of densities is more evident when examining the *actual efficiency* [given by Eq. (24)] (more on discrepancies to *efficiency approximation* [given by Eq. (20)] later) for the implementation that has a smaller number of simulations (Figure 9), something that translates to a limited

amount of information for selecting the adaptive IS. This is expected, and is the motivation for introducing the IS_M formulation, since challenges are anticipated when increasing the number of features of the density calibrated while using such limited information. This is manifested as some (slight) loss of robustness for the joint family of densities, especially when the number of input variables n considered in the IS implementation is larger. With respect to the influence of this number, as also discussed earlier, results show variability across the different examined advisories, with n=4 emerging consistently, as the preferable implementation, but with a relative reduction of efficiency when using smaller number of simulations N=100, occurring for the reasons explained just above.

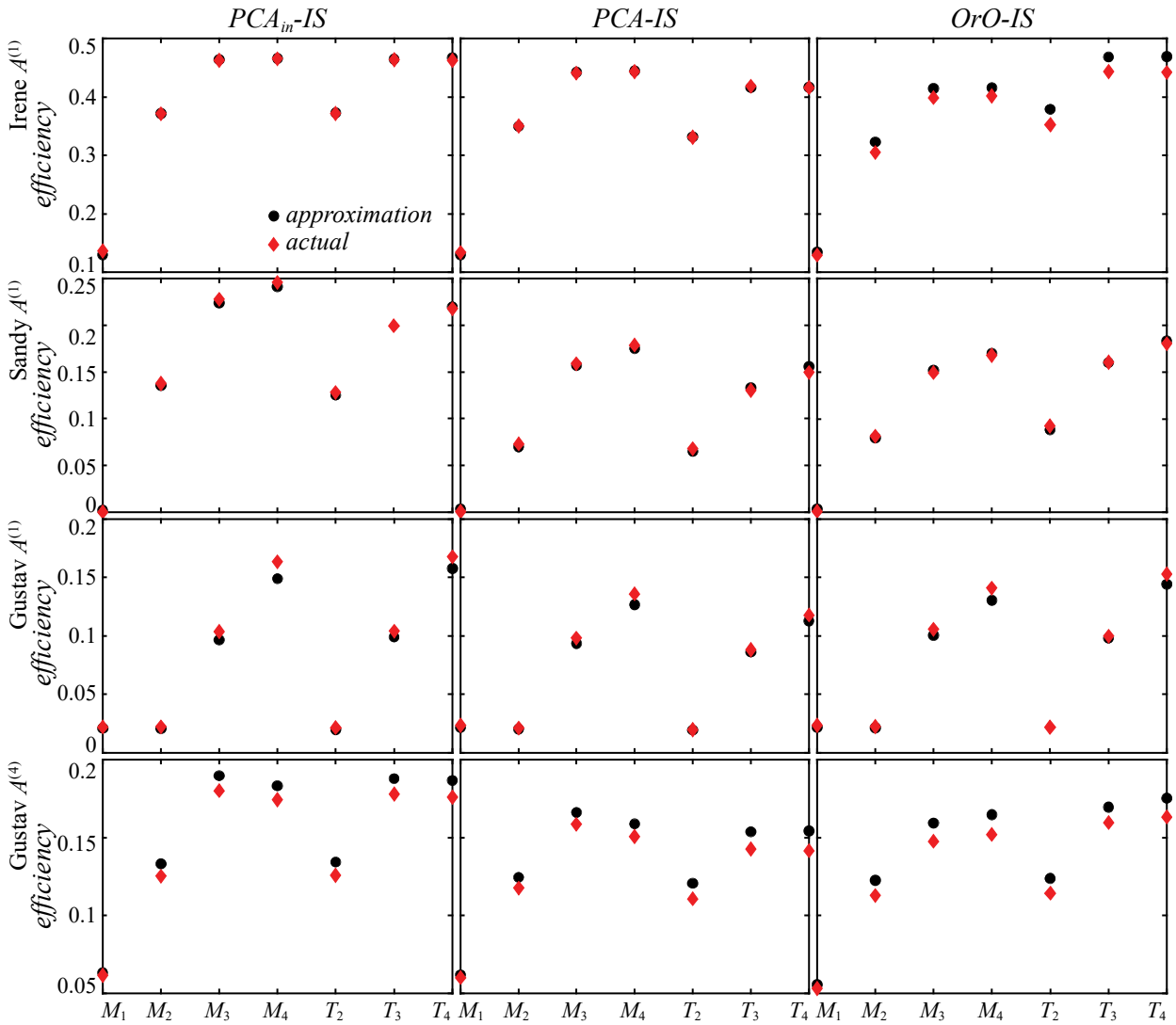


Figure 10. Approximated and actual IS efficiency for a specific advisory for different IS variants (columns of figure) and different formulations for the IS density. Results are presented for different storms and advisories (rows of figure) and correspond to the $S_{0.1}$ statistic, similar to Figure 8, but for priority weights r_j set equal to the variance of each location.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

Examining the trends in Figures 8 and 10 demonstrates that the overall degree of efficiency improvement also depends on the exact selection of the priority weights as well as the required degree of compromise across the different surge locations for selecting the IS density; for example, if the chosen r_j prioritize nodes with similar optimal IS density characteristics (for example same mean for the IS density), then the average IS benefits will increase. This discussion offers preliminary guidance that IS can provide important benefits within the application examined here, though the degree of improvement in computational efficiency will depend on details of the storm surge variability within the geographic domain of interest, and the statistics examined. Of course the exact IS benefits can be evaluated only when an implementation across the advisories is examined, something discussed in the next section.

Comparing, finally, the results for the actual and approximated efficiency [given by Eqs. (24) and (20) respectively], good overall agreement is reported. This agreement is better for the *PCA$_{in}$-IS* and *PCA-IS* variants, demonstrating, again, a preference for IS formulation utilizing a dimensionality reduction approach. The discrepancies are significantly bigger, as expected, for (i) the case corresponding to less frequent event [compare the same variant implementation between Figures 7 and 8], since a smaller amount of information is extracted from the same amount of resources (same number of available simulations), and (ii) for the case with smaller N [compare the same variant implementation between Figures 8 and 9], owing to the limited information utilized in this case to obtain the *efficiency approximation*. Nevertheless, the adaptive IS selection (comparing across all M_n and T_n cases) based on the *efficiency approximation* yields consistent results with the choices that would be made if the *actual efficiency* was known. This is the most important feature, as the *efficiency approximation* is directly leveraged within AIS-SA to select the best IS candidates.

Next, the IS efficiency across the different statistics is examined. In this case, the IS formulation is based on one statistic [for example surge thresholds corresponding to 10% exceedance probability $S_{0.1}$], while the efficiency is evaluated using a different statistic [for example surge thresholds corresponding to 1% exceedance probability $S_{0.01}$]. This means that different consequence measures $h_j(\mathbf{x}\,|\,\tilde{\mathbf{q}}^{(k)})$ are used for: (i) the AIS-SA algorithm in Section 4.4.; and (ii) the efficiency estimation according to Eq. (24). Comparison against the case that the same consequence measure is used across these two steps quantifies the efficiency loss established by using a different statistic in the IS selection. Figures 11 and 12 show the *actual efficiency* calculated using Eq. (24) with N_r=5000 quasi-random samples for each of the examined statistics (curves in the figures) for the IS density chosen based on every other one (x-axis discrete cases). Results correspond to the optimal density promoted through the AIS-SA algorithm and for the preferred variant, *PCA$_{in}$-IS*, instead of the comprehensive cases reported in the earlier figures in this section. Figure 11 presents results for the *promoted-IS* density (Step 9 of the algorithm) and Figure 12 for the *robust-IS*

density (Step 10 of the algorithm). In these figures, for each different curve, the highest efficiency is anticipated for IS selected based on the same statistic (same x-axis case), whereas the departure from this efficiency for the other implementations quantifies the loss when IS is chosen based on different criteria. Note that since decisions made within AIS-SA are based on the *efficiency approximation* but all validations utilize the *actual efficiency* some small deviations from the trend of having the highest efficiency corresponding to the same x-axis case are reasonable.

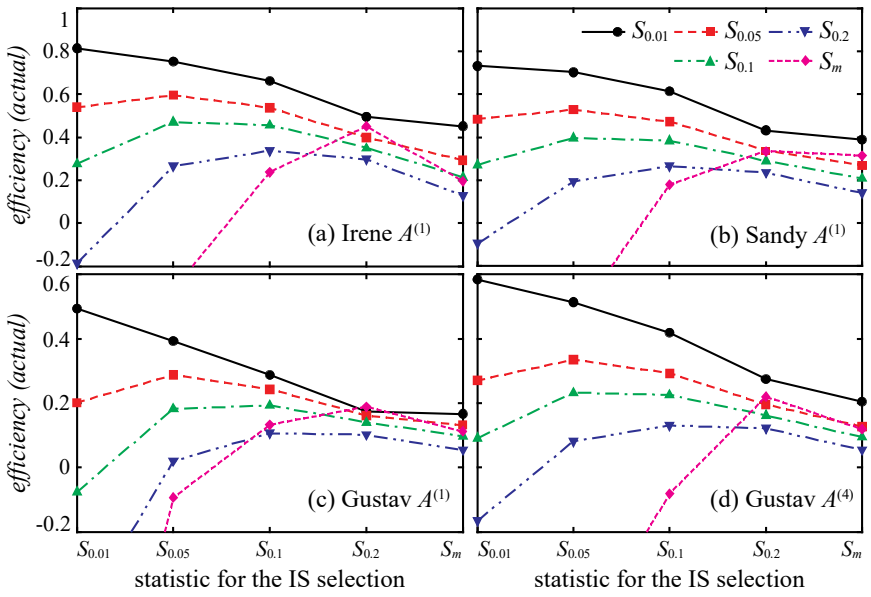


Figure 11. Efficiency for various statistics of interest (different curves) for different IS density selections (represented in the x-axis in each plot). Results correspond to the *promoted-IS* density and the PCA_{in}*-IS* variant.

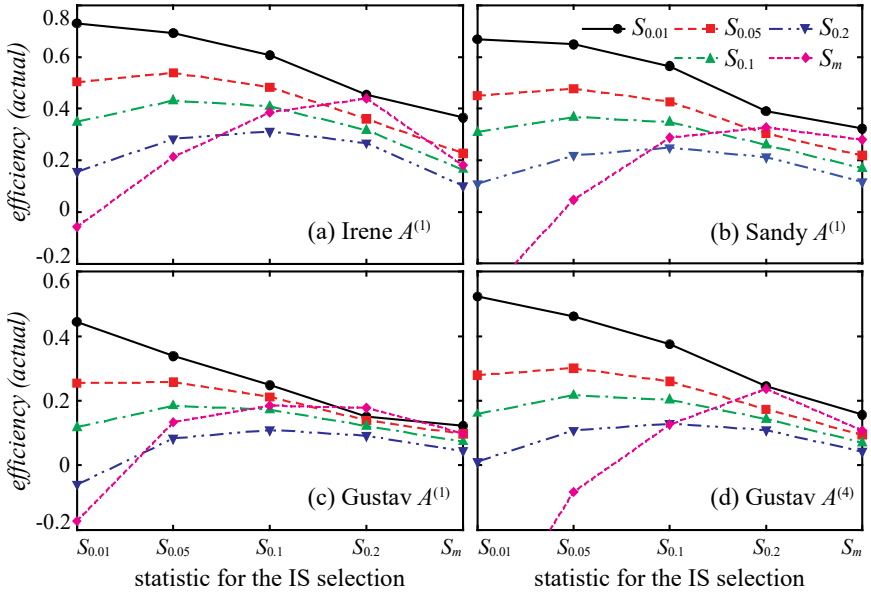


Figure 12. Efficiency for various statistics of interest (different curves) for different IS density selections (represented in the x-axis in each plot). Results correspond to the *robust-IS* density and the PCA_{in}*-IS* variant.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

Results across all these figures showcase significant dependence of the IS efficiency on the chosen statistic (consequence measure). This is manifested as a significant dispersion of the efficiency across the different statistics of interest for a specific IS selection (specific case in the x-axis of Figures 11 and 12). Statistics that correspond to less frequent events, exhibit greater benefits (higher potential efficiencies) from the IS formulation, further validating trends identified in Figures 8 and 9 and discussions in Section 3, but this comes with a stronger performance dependence on the statistic used for the IS selection (bigger variability across the curve). The opposite is also true with respect to the effect across other statistics of the IS density chosen based on less frequent statistics; the spread is larger when the IS is chosen based on less frequent events [$S_{0.01}$ case in the x-axis], leading for some other statistics even to loss of efficiency compared to the direct MC implementation [efficiency values less than 0]. Overall the IS selection using the intermediate statistic $S_{0.1}$ seems to perform well across all the statistic examined here, which, recall, encompass the typical statistics of interest in probabilistic surge forecasting applications. Moreover, the implementation of the *robust-IS* seems to greatly improve the performance across the entire range of statistics examined here: comparing Figure 12 to 11 one can observe a significant reduction in the dispersion of the efficiency (translating to an improvement in robustness) with a moderate only (ρ=0.1 in this case) reduction of efficiency for the targeted statistic. These discussions show that when different statistics are of interest, careful consideration of the one to base the IS selection is needed. The use of some intermediate statistical quantity, like the $S_{0.1}$ examined here, should be preferred, while the defensive IS formulation does provide additional robustness even in this context. The latter has an evident dependence on ρ, but this topic will be discussed further in the next section. Of course, as discussed earlier, the different statistics can be augmented in a single QoI definition, across not only multiple locations but also across different statistics definitions to better inform the AIS-SA formulation. In this case careful selection of the priority weights might be needed to better balance decisions across the different statistics of interest.

6.2 IS formulation across advisories

With the IS characteristics for a single advisory examined in detail, the validation moves on now to the proposed implementation across advisories. In this case, the identified as preferred variant *PCA_{in}-IS* is only considered, whereas results are reported both for the *promoted-IS* (Step 9 of the AIS-SA algorithm) and the *robust-IS* densities (Step 10 of the AIS-SA algorithm). To evaluate the efficiency and robustness of the AIS-SA algorithm with respect to the information sharing across advisories, another implementation is examined, corresponding to the optimal IS selection for the current advisory. This allows the efficiency accomplished by choosing IS densities across subsequent advisories to be compared to the best efficiency that would be achieved if one could explicitly choose and implement the IS density for the present advisory. For the *k*th advisory the first type of efficiency corresponds to $F^{(k)}[f^{(k)}(\mathbf{x})]$, i.e. using as proposal density

the one coming from the previous advisory, whereas the second type of efficiency corresponds to $F^{(k)}[f^{(k+1)}(\mathbf{x})]$, i.e. using as proposal density the optimal density identified at the present advisory. These two cases will be distinguished in the figures using terminology *previous* and *current*, with *previous* corresponding to the AIS-SA formal implementation (across advisories) and *current* to the best efficiency. Results in Figures 13 and 14 are reported for each of the three different storms for both the *promoted-IS* density and the *robust-IS* density. Figure 13 presents results for the $S_{0.01}$ statistic and Figure 14 for the $S_{0.1}$ statistic. Figure 15 examines a similar implementation as Figure 14 but in this case for N=100, to investigate the impact of the number of simulations on the adaptive IS formulation across advisories. Finally, Figure 16 presents results for the performance with respect to the $S_{0.01}$ statistic (similar to Figure 13) with IS density chosen based on the $S_{0.1}$ statistic (i.e. the exact densities utilized in Figure 14). This ultimately replicates the scenarios evaluated in Figures 11 and 12, but in this case with respect to the implementation across advisories.

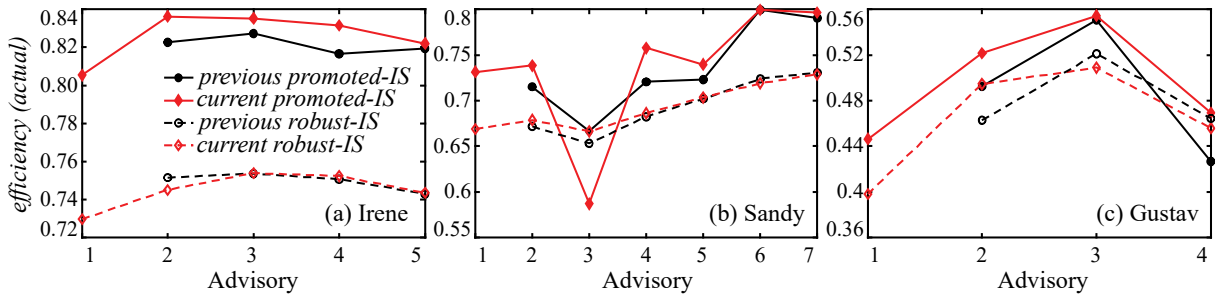


Figure 13. Efficiency of AIS-SA across advisories for the $S_{0.01}$ statistic for the three case study storms (columns of figure) for both the *promoted-IS* density and the *robust-IS* density. Results correspond to the *PCA$_{in}$-IS* variant. In each case the efficiency of the density chosen based on the current advisory is also shown.

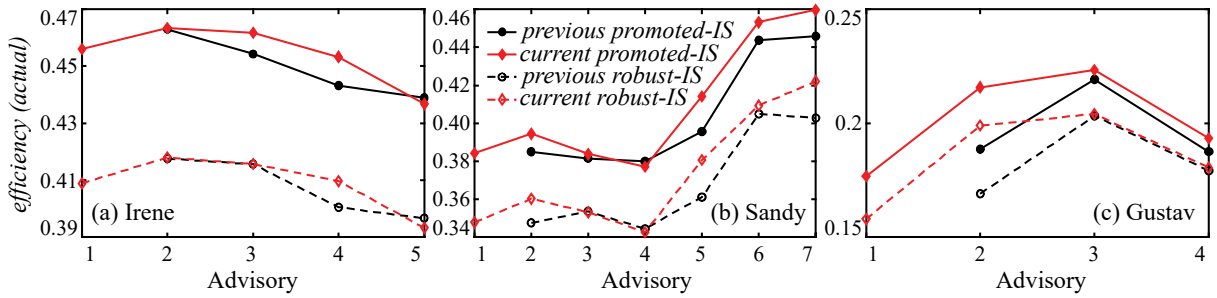


Figure 14. Efficiency of AIS-SA across advisories for the $S_{0.1}$ statistic for the three case study storms (columns of figure) for both the *promoted-IS* density and the *robust-IS* density. Results correspond to the *PCA$_{in}$-IS* variant. In each case the efficiency of the density chosen based on the current advisory is also shown.

Results clearly show that the proposed implementation, sharing information across advisories, works well, since the gap in performance between the *previous* and *current* IS implementations is small: the density chosen based on the results from the previous advisory as well as the density that would have been chosen if the results from the current advisory were known. Note that the fact that in some (very few)

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

instances the *previous IS* implementation outperforms the *current IS* implementation, is not something that should be anticipated (the formulation using current information should improve over the one using past information) but it can be explained: the selection of the IS densities is based on the *efficiency approximation* [calculated with Eq. (20) using readily available small number of simulations], but the validation here is performed with respect to the *actual efficiency* [calculated with Eq. (24) using a large number of simulations]. As also discussed for the trends reported in Figures 11 and 12, there can be instances that the *efficiency approximation* leads to (slightly) sub-optimal decisions when evaluated against the *actual efficiency*.

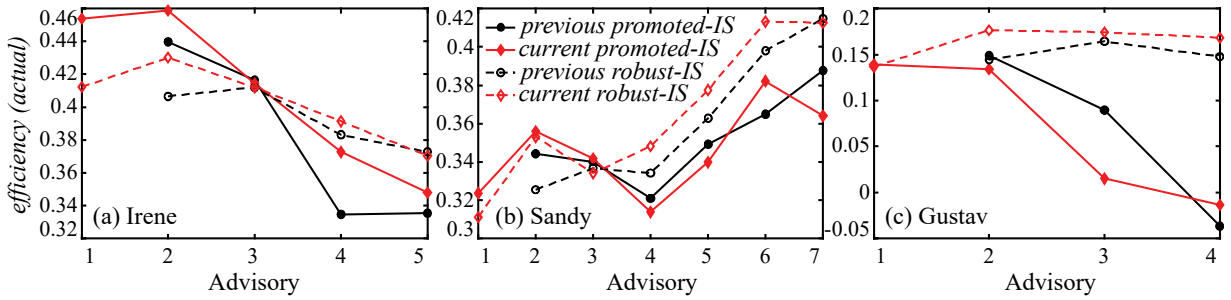


Figure 15. Efficiency of AIS-SA across advisories for the $S_{0.1}$ statistic similar to Figure 14, but for N=100 number of simulations, for the three case study storms (columns of figure) for both the *promoted-IS* density and the *robust-IS* density. Results correspond to the PCA_{in}*-IS* variant. In each case the efficiency of the density chosen based on the current advisory is also shown.

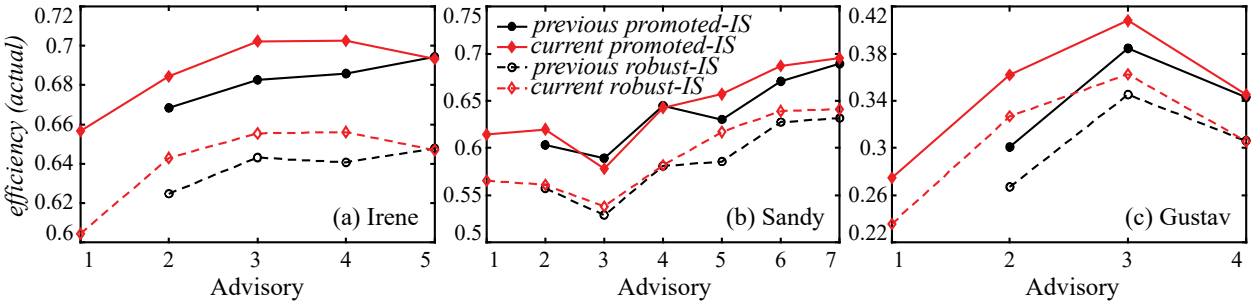


Figure 16. Efficiency of AIS-SA across advisories for the $S_{0.01}$ statistic similar to Figure 13, but for IS density chosen based on the $S_{0.1}$ statistic, for the three case study storms (columns of figure) for both the *promoted-IS* density and the *robust-IS* density. Results correspond to the PCA_{in}*-IS* variant. In each case the efficiency (for the $S_{0.01}$ statistic) of the density chosen based on the current advisory (for the $S_{0.1}$ statistic) is also shown.

Investigating further the trends in Figures 13-16, the implementation of the defensive IS strategy (*robust-IS* case) does not seem to be warranted from the perspective of sharing information across advisories, since as discussed above, the formulation based on the previous advisory information is very close to the *optimal* IS selection, indicating no loss of robustness by the fact that advisories change. It is warranted though to provide robustness against making selections using limited information for deciding on the IS densities (N=100), as the results in Figure 15 clearly indicate, especially for hurricane Gustav. The reason why this is more clearly evident for hurricane Gustav is the smaller relative efficiency of the IS

implementation for this storm (see also discussion in previous section), creating greater impact on any efficiency reduction. It is important to additionally note that here the effects from the use of smaller N are greater compared to the identical case investigated in Section 6.1 (Figure 9). These differences originate from the vulnerabilities associated with the use of the *efficiency approximation* in the AIS-SA algorithm and the fact that, in the implementation examined here, this approximation is estimated by using the IS density from the previous advisory. As already discussed in Section 6.1 (and illustrated in Figure 9), the differences between the approximated and actual efficiency are larger when the number of simulations utilized to calculate the former is small. When this estimation is based on proposal densities formulated on past information (previous advisory) these differences can become larger and create vulnerabilities for the decisions made according to them, as the results in Figure 15 clearly indicate. The *robust-IS* case addresses all these vulnerabilities, indicating the importance of using the *robust-IS* formulation when the number of simulations utilized for the proposal density selection is small. Even if N is large, since maintaining a level of robustness in the AIS-SA formulation is important, one can argue that trends here indicate that simply a smaller value of ρ should be utilized. This will reduce the gap between the *robust-IS* and the *promoted-IS* formulations, improving efficiency while still infusing a sufficient level of robustness, since the previous discussions and comparisons clearly demonstrated that as long as N is large, the sharing of information across advisories does not create significant vulnerabilities in the adaptive IS selection (indicating that. smaller robustness level is sufficient). Unfortunately, an adaptive way to choose ρ cannot be established since decisions are made based on the *efficiency approximation* (no other information available for AIS-SA) and the current advisory. The detailed exploration of an appropriate value of ρ requires estimation of the *actual efficiency* for the next advisory.

The results also verify the trends examined in more detail in section 6.1, with efficiency being greater for the $S_{0.01}$ statistic or the *robust* implementation providing benefits when the IS is based on a different statistic (compare the results in Figures 16 and Figures 11/12).

6.3 Integration within QMC implementation

It was shown in (Kyprioti et al. 2021a) that QMC offers an attractive formulation for probabilistic surge estimation, establishing deterministic solutions for a given N quasi-random samples, which is a preferred feature based on the existing NWS computational workflow (Taylor and Glahn 2008; Gonzalez and Taylor 2018), while improving the numerical accuracy. The AIS-SA can be integrated within such a QMC formulation if the IS selection is restricted to marginal distributions (IS_M). Note that the latter restriction is not necessary; the class of joint IS densities IS_T can in principle be considered, but in this case, some transformation to independent distributions will be needed to facilitate the QMC sampling (Lemieux 2009). The use of the marginalized IS_M family of densities promotes a seamless extension while, equally importantly, the IS_M formulation is comparable to or even slightly preferred over the IS_T formulation choice

as the results in Section 6.1 indicate. The extension is accommodated by using QMC sampling instead of random MC sampling to obtain the samples from $f^{(k)}(\mathbf{x})$. All other aspects of the algorithm remain the same. Note that a formal adoption of IS within a QMC setting may be also considered (Spanier and Maize 1994), though the formulation requires significant modifications of the existing computational workflow. For this reason, the simpler direct integration of AIS-SA within QMC is promoted here.

The validation is performed by examining the bias of the AIS-SA (utilizing QMC) or direct QMC predictions, given by Eq. (9) (note that for direct QMC $f^{(k)}(\mathbf{x}) = p(\mathbf{x})$ in this equation), to the reference results given by Eq. (23). The number of simulations for the AIS-SA implementation is greatly reduced to N=50, to test how the proposed integration of two accuracy improvement techniques (IS and QMC) can help in further reducing the computational burden. For the QMC sampling, Sobol' low discrepancy sequences are used. AIS-SA with QMC is compared to QMC using N, 1.5N, or 2N samples. Similar to (Kyprioti et al. 2021a) the average accuracy is expressed using the normalized mean error, given by:

$$NME^{(k)} = \frac{\frac{1}{n_z}\sum_{j=1}^{n_z}\left|H_j^{(k)} - \hat{H}_j^{(k)}\right|}{\frac{1}{n_z}\sum_{j=1}^{n_z}\left|H_j^{(k)}\right|} \tag{25}$$

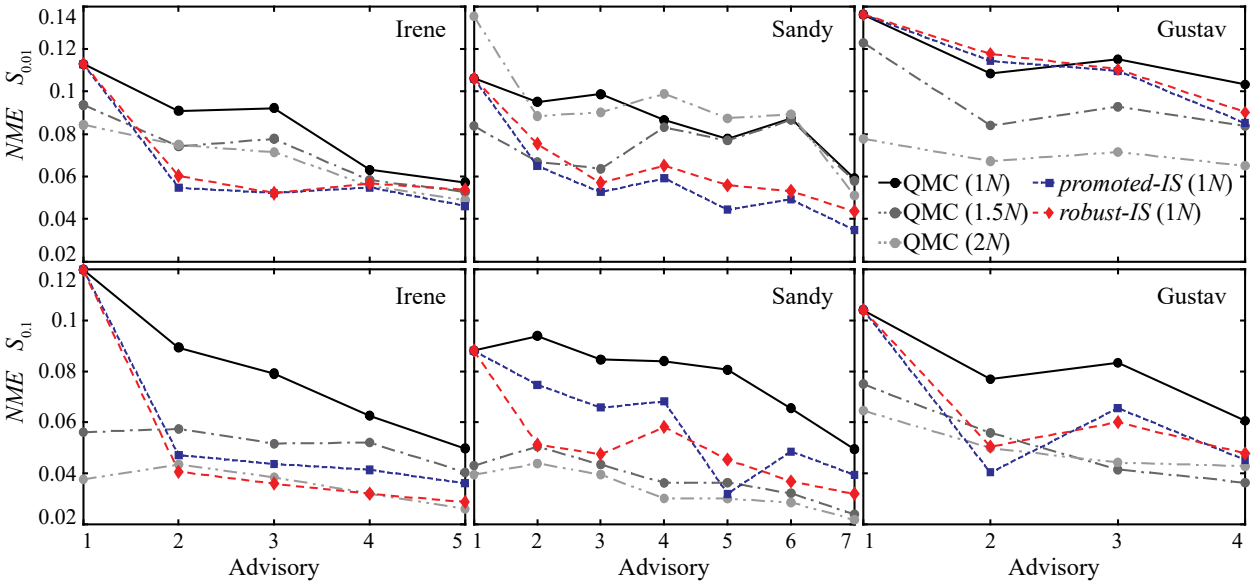


Figure 17. Accuracy (*NME*) of AIS-SA with QMC across advisories for the $S_{0.01}$ statistic (top row) and the $S_{0.1}$ statistic (bottom row) for the three case study storms (columns of figure) for both the *promoted-IS* density and the *robust-IS* density. Results correspond to the *PCA*$_{in}$*-IS* variant and IS density chosen based on the $S_{0.1}$ statistic. In each case the accuracy of direct QMC using N, 1.5N or 2N samples is also reported.

Results are presented in Figure 17 for both the *promoted-IS* density and the *robust-IS* density for performance assessed using either the $S_{0.1}$ statistic or the $S_{0.01}$ statistic, but with IS density chosen (in both instances) according to the $S_{0.1}$ statistic. This setting is chosen to better evaluate the implementation that

was promoted earlier for practical applications, using $S_{0.1}$ to decide on the IS characteristics even when considering an implementation to different statistics. Results show that AIS-SA can be seamlessly couched within a QMC formulation and provide additional computational savings that vary from storm to storm, but may reach a two-fold reduction: AIS-SA in all instances outperforms direct QMC and accomplishes similar accuracy to QMC using 1.5 or even 2 times larger storm ensembles. Considering that QMC was shown (Kyprioti et al. 2021a) to provide consistently at least a threefold reduction compared to the current practice of using factorial sampling (Gonzalez and Taylor 2018), the advantages of combining QMC and AIS-SA are clearly evident. Moreover, *robust-IS* and *promoted-IS* implementations show consistently similar accuracy, with a much smaller gap when compared to the comparisons in section 6.2. This indicates that within a QMC implementation, the *robust-IS* formulation may infuse robustness with only a small efficiency reduction.

7 Conclusions

This paper examined the implementation of importance sampling (IS) for the improvement of computational efficiency in real-time probabilistic surge estimation. The IS is established by sharing information across the different advisories, using the readily available simulation results from the current advisory to adaptively select a proposal density within a MC estimation setting, to use in the next advisory. The objective is to reduce the number of simulations needed in the next advisory to offer statistical estimates with the same accuracy. A sample-based, adaptive IS formulation was established in this paper, termed AIS-SA, to achieve this objective. The sample-based approach adopted for AIS-SA can be applied even when the number of available simulations, and consequently the information for the proposal density selection, is limited, as this is the setting which is expected in typical practical applications. A parametric density implementation was proposed based on Gaussian Mixture Models (GMMs), whereas a workflow that considers different candidate proposal densities (with different number of inputs and corresponding to either the joint or marginal distributions across these inputs), and selects the most favorable one according to its anticipated efficiency was established. The fact that the probabilistic estimation pertains to the storm surge for multiple locations within the geographic domain of storm impact poses a significant challenge, since a single proposal density needs to be promoted across all these quantities of interest (QoIs). Three different variants were examined to address this challenge, the first two considering a dimensionality reduction through principal component analysis (PCA) and considering a separate density for each principal component (*PCA$_{in}$-IS*) or a single density across all components (*PCA-IS*), and the third one selecting a single density across all original outputs (*OrO-IS*). Finally, for establishing robustness to accommodate the implementation across the large number of QoIs using limited information (limited number of simulations)

and, more importantly, the fact that the IS density is chosen based on the current advisory but is implemented in the next one, a defensive IS scheme (termed *robust* IS) was introduced.

The efficacy of the proposed computational workflow was established by considering three different storms with diverse characteristics. Main conclusions from the case study are the following:

- The most important conclusion is that information sharing across the storm advisories can accommodate computational benefits (improvement in accuracy with same computational budget) in real-time probabilistic surge estimation. The differences between subsequent advisories are, typically, not significant enough to lead to erroneous decisions. The adaptive IS implementation examined in this paper is one approach to leverage such information. Independent of the adopted approach, the results in the case studies provide tangible proof that such information should be utilized and can improve the efficiency of existing computational workflows (Taylor and Glahn 2008; Gonzalez and Taylor 2018).
- AIS-SA implementation yields satisfactory improvements in computational efficiency (in most instances reduction of computational burden by 1.5 to 2 times can be achieved), though it faces challenges in the fact that the different QoIs within the geographic domain of storm impact yield optimal choices for IS densities with competing features. Having to choose a proposal density that balances across all of them reduces the AIS-SA efficiency.
- The promoted IS variant is PCA_{in}-*IS*, establishing better efficiency and robustness by leveraging the dimensionality reduction (for robustness) and versatility of selecting separate IS densities for each principal component (for efficiency).
- The implementation across advisories has only a small impact on the AIS-SA efficiency, with IS densities based on the previous advisory having only a small deviation from the performance of the IS densities that could be chosen if simulation results from the current advisory were known.
- In the case studies examined here, the additional robustness established through the defensive-IS scheme is critical when the available information for formulating the IS densities across advisories is limited (small number of simulations). Even when adequate information is available, the recommendation is to keep utilizing a defensive strategy, though with a reduced threshold ρ for the efficiency reduction compared to the ρ=0.1 used in this study.
- The benefits from the IS implementation are larger for statistics that correspond to infrequent events (surge thresholds corresponding to 1% rather than 10% exceedance probability).
- When different statistics are calculated within the probabilistic surge estimation, the IS should be based on statistics with intermediate characteristics in terms of frequency of occurrence. The thresholds corresponding to 10% exceedance probabilities were shown in this study to provide good performance across a range of statistical products of interest.

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

- Comparison between marginal and joint candidate IS densities showed small differences. The former implementation can more easily support applications with reduced amount of information (limited simulations), and also can be more easily integrated within a Quasi-Monte Carlo (QMC) setting, and thus it is the recommended formulation.
- The integration of AIS-SA within a QMC implementation provides additional computational savings that vary from storm to storm, but may reach a two-fold reduction.

8 Statements and Declarations

Funding

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

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Conceptualization, A.A.Taflanidis, W.Jung and A.P.Kyprioti; methodology, W.Jung, and A.A. Taflanidis; software, W. Jung, and E.Adeli; validation, W.Jung, and A.A.Taflanidis; data curation, W.Jung and A.P.Kyprioti; writing—original draft preparation, A.A.Taflanidis, W.Jung and A.P.Kyprioti; writing—review and editing, A.A.Taflanidis, W.Jung, A.P.Kyprioti, E.Adeli, J.J. Westerink and H.Tolman; funding acquisition, A.A.Taflanidis and J.J. Westerink. All authors have read and agreed to the published version of 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, and for providing access to the NHC forecast error statistics and advisories used in this study. They would also like to thank the Army Corps 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.

Data Availability

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

Appendix A: Review of PCA implementation

Define the observation matrix $\bar{\mathbf{H}}^{(k)}$, with *l*th column corresponding to the central observations, obtained by subtracting from the elements of the *l*th column $\mathbf{H}^{(k)}$ of their mean value, given by Eq. (9). Then, define the weighted observations $\bar{\mathbf{H}}_r^{(k)} = \text{diag}(\sqrt{\boldsymbol{\varphi}})\bar{\mathbf{H}}^{(k)}\text{diag}(\sqrt{\mathbf{r}})$ (Greenacre 1984), where $\text{diag}(\sqrt{\boldsymbol{\varphi}})$ corresponds to diagonal matrix with elements $\{\sqrt{\varphi(\mathbf{x}^l)}; l = 1,..., N\}$ across its diagonal, and $\text{diag}(\sqrt{\mathbf{r}})$ to a diagonal matrix with $\sqrt{\mathbf{r}}$ across its diagonal. Consider, now, the eigenvalue problem for the weighted covariance matrix $[\bar{\mathbf{H}}_r^{(k)}]^T \bar{\mathbf{H}}_r^{(k)}$. The solution of this eigenvalue problem provides the vector of latent outputs (principal components), with the *v*th latent output denoted as y_v. The corresponding eigenvalue λ_v represents the portion of the total variance of the weighted original data $\bar{\mathbf{H}}_r^{(k)}$ that can be explained by y_v, while the eigenvector $\mathbf{P}_v \in \mathbb{R}^{n_z}$ facilitates the mapping to the latent space. Note that the weighting by $\sqrt{\mathbf{r}}$ ultimately provides priority r_j to the variance of each output when constructing the weighted covariance $[\bar{\mathbf{H}}_r^{(k)}]^T \bar{\mathbf{H}}_r^{(k)}$ for the PCA, whereas the weighting by $\sqrt{\boldsymbol{\varphi}}$ adjusts this covariance to refer to the original probability model $p(\mathbf{x})$ (Greenacre 1984; Jolliffe 2002). To accommodate a larger dimensionality reduction only the principal components corresponding to the n_v largest eigenvalues are retained, with n_v chosen such that the ratio

$$\psi = \frac{\sum_{n=1}^{n_v} \lambda_v}{\sum_{n=1}^{\min(n_z, N-1)} \lambda_v} \tag{A.1}$$

is greater than some threshold ψ_o (for example 99%). PCA ultimately maximizes the variance in the original data that has been preserved when considering only n_v components. The observation matrix for the latent outputs for the N available simulations can be obtained as $\mathbf{Y}^{(k)} = \bar{\mathbf{H}}_r^{(k)}\mathbf{P}$, where the projection matrix $\mathbf{P} \in \mathbb{R}^{n_z \times n_v}$ is the matrix with *v*th column corresponding to eigenvector $\mathbf{P}_v$.

References

Au SK, Beck JL (1999) A new adaptive importance sampling scheme. Structural Safety 21 (2):135-158

Au SK, Beck JL (2003) Importance sampling in high dimensions. Structural Safety 25 (2):139-163

Chen JH, Lin SJ, Magnusson L, Bender M, Chen X, Zhou L, Xiang B, Rees S, Morin M, Harris L (2019) Advancements in hurricane prediction with NOAA's next-generation forecast system. Geophysical Research Letters 46 (8):4495-4501

Ehre M, Papaioannou I, Willcox KE, Straub D (2021) Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse. Computer Methods in Applied Mechanics and Engineering 381:113826

Geyer S, Papaioannou I, Straub D (2019) Cross entropy-based importance sampling using Gaussian densities revisited. Structural Safety 76:15-27

Glahn B, Taylor A, Kurkowski N, Shaffer WA (2009) The role of the SLOSH model in National Weather Service storm surge forecasting. National Weather Digest 33 (1):3-14
Gonzalez T, Taylor A (2018) Development of the NWS'Probabilistic Tropical Storm Surge Model. Paper presented at the 33rd Conference on Hurricanes and Tropical Meteorology, Ponte Vedra, FL,
Greenacre MJ (1984) Theory and applications of correspondence analysis.
Hamill TM, Brennan MJ, Brown B, DeMaria M, Rappaport EN, Toth Z (2012) NOAA's future ensemble-based hurricane forecast products. Bulletin of the American Meteorological Society 93 (2):209-220
Hesterberg T (1995) Weighted average importance sampling and defensive mixture distributions. Technometrics 37 (2):185-194
Hesterberg TC (1988) Advances in importance sampling. Stanford University,
Jelesnianski CP, Chen J, Shaffer WA (1992) SLOSH: Sea, lake, and overland surges from hurricanes. NOAA Technical Report, NWS 48. US Department of Commerce, National Oceanic and Atmospheric Administration,
Jelesnianski CP, Taylor A (1973) A preliminary view of storm surges before and after storm modifications. NOAA technical memorandum. vol 3. Environmental Research Laboratories, Weather Modification Program Office, Washington, DC
Jia G, Taflanidis AA, Beck JL (2017) A new adaptive rejection sampling method using kernel density approximations and its application to subset simulation. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3 (2):D4015001
Jia G, Taflanidis AA, Nadal-Caraballo NC, Melby J, Kennedy A, Smith J (2016) Surrogate modeling for peak and time dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms. Natural Hazards 81 (2):909-938
Jolliffe IT (2002) Principal component analysis. Springer series in statistics, 2nd edn. Springer, New York
Jung W, Kyprioti AP, Adeli E, Taflanidis AA (2022) Exploring the sensitivity of probabilistic surge estimates to forecast errors. Natural Hazards in press
Kijewski-Correa T, Taflanidis A, Vardeman C, Sweet J, Zhang J, Snaiki R, Wu T, Silver Z, Kennedy A (2020) Geospatial environments for hurricane risk assessment: applications to situational awareness and resilience planning in New Jersey. Frontiers in Built Environment 6:549106
Knaff JA, Zehr RM (2007) Reexamination of tropical cyclone wind–pressure relationships. Weather and Forecasting 22 (1):71-88
Kroese DP, Taimre T, Botev ZI (2011) Handbook of Monte Carlo methods. Wiley Series in Probability and Statistics. John Wiley and Sons, New York
Kurtz N, Song J (2013) Cross-entropy-based adaptive importance sampling using Gaussian mixture. Structural Safety 42:35-44
Kyprioti AP, Adeli E, Taflanidis AA, Westerink JJ, Tolman HL (2021a) Probabilistic Storm Surge Estimation for Landfalling Hurricanes: Advancements in Computational Efficiency Using Quasi-Monte Carlo Techniques. Journal of Marine Science and Engineering 9 (12):1322
Kyprioti AP, Taflanidis AA, Nadal-Caraballo NC, Campbell M (2021b) Storm hazard analysis over extended geospatial grids utilizing surrogate models. Coastal Engineering:103855
Lemieux C (2009) Monte carlo and quasi-monte carlo sampling. Springer Science & Business Media,
Luettich RA, Jr. , Westerink JJ, Scheffner NW (1992) 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
McNicholas PD, Murphy TB (2008) Parsimonious Gaussian mixture models. Statistics and Computing 18 (3):285-296
Medina JC, Taflanidis A (2014) Adaptive importance sampling for optimization under uncertainty problems. Computer Methods in Applied Mechanics and Engineering 279:133–162. doi:10.1016/j.cma.2014.06.025
Moon TK (1996) The expectation-maximization algorithm. IEEE Signal Processing Magazine 13 (6):47-60
Nadal-Caraballo NC, Campbell MO, Gonzalez VM, Torres MJ, Melby JA, Taflanidis AA (2020) Coastal Hazards System: A Probabilistic Coastal Hazard Analysis Framework. Journal of Coastal Research 95 (sp1):1211-1216
Resio DT, Irish JL, Westering JJ, Powell NJ (2012) The effect of uncertainty on estimates of hurricane surge hazards. Natural Hazards 66 (3):1443-1459
Robert CP, Casella G (2004) Monte Carlo Statistical Methods. 2nd edn. Springer, New York, NY
Schuëller GI, Pradlwater HJ, Koutsourelakis PS (2004) A critical appraisal of reliability estimation procedures for high dimensions. Probabilistic Engineering Mechanics 19 (4):463-474

Author Personal Copy (DOI: 10.1016/j.coastaleng.2023.104287)

Silverman BW (1998) Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC. Boca Raton, Florida
Smith JM, Westerink JJ, Kennedy AB, Taflanidis AA, Smith TD (2011) SWIMS Hawaii hurricane wave, surge, and runup inundation fast forecasting tool. In: 2011 Solutions to Coastal Disasters Conference, Anchorage, Alaska, June 26-29.
Spanier J, Maize EH (1994) Quasi-random methods for estimating integrals using relatively small samples. SIAM Review 36 (1):18-44
Taylor AA, Glahn B (2008) Probabilistic guidance for hurricane storm surge. In: 19th Conference on probability and statistics.