14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

HAZARD-CONSISTENT SELECTION OF STORM SCENARIOS FOR LONG-TERM STORM SURGE HAZARD ESTIMATION ACROSS LARGE GEOGRAPHIC REGIONS

WOONGHEE JUNG[1], ALEXANDROS A. TAFLANIDIS[1], NORBERTO C. NADAL-CARABALLO[2], MADISON C. YAWN[2], AND LUKE A. AUCOIN[2]

[1]Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame
Notre Dame, IN 46556
wjung2@nd.edu, a.taflanidis@nd.edu

[2]U.S. Army Corps of Engineers, Engineer Research and Development Center, Coastal and Hydraulics Laboratory, Coastal Hazards Group
Vicksburg, MS 39180
Norberto.C.Nadal-Caraballo@erdc.dren.mil, Madison.C.Yawn@erdc.dren.mil, Luke.A.Aucoin@erdc.dren.mil

Key words: Coastal hazard estimation, flood hazard study, tropical cyclones, hazard-consistent probabilistic scenario optimization, response-proxy.

Abstract. The recent destructive hurricane seasons have increased the importance of storm surge hazard estimation for discussing coastal community resilience. The hazard is generally represented by surge inundation probabilities for each location (hazard curves) or surge levels corresponding to specific exceedance probabilities across all locations (hazard maps) within the coastal domain of interest. It is assessed through a Monte Carlo simulation (MCS), utilizing an ensemble of storm scenarios and their occurrence rates that represent the regional climatology. The impact prediction of each storm scenario within the MCS is performed by a hydrodynamic-based numerical model with a high computational cost. To address the challenge, reducing the number of storm scenarios considered for the hazard estimation is desirable. This is especially true in pre-disaster planning settings when investigating the impact of adaptation measures on coastal resilience. Since the entire process needs to be repeated for each measure, the number of storm scenarios needs to be kept moderately low. This study investigates the selection of fewer storms from the full ensemble within this setting, which can produce consistent hazard products with some chosen ones obtained from the full ensemble. Beyond the identification, the occurrence rates of selected storms are also updated. The final subset can be subsequently utilized as a representative ensemble for the regional hazard estimation. The storm selection is established as a two-stage optimization: the inner-loop adjusts the occurrence rates for a candidate subset; the outer-loop searches for the best subset with the desired number of storms. This work extends the implementation to extended coastal regions with many locations of interest. In this case, it is computationally intractable to consider all the locations in the domain, and for this reason, a subset of representative locations is chosen through cluster analysis and considered for storm selection. Additionally, the correlation in the hazard description is better integrated into the storm selection by modifying the objective function for the outer-loop optimization.

1 INTRODUCTION

The importance of the accurate assessment of the long-term regional storm surge hazard has increased, given its critical role in guiding informed pre-disaster planning and ultimately enhancing the resilience of coastal communities[1]. The hazard is quantified by considering an ensemble of storm scenarios representing the regional climatology, predicting the storm surge levels for each storm, and combining these predictions with the occurrence rates of storms to deliver the desired hazard products[2], such as curves, providing the annual probabilities of surge exceeding different threshold values for specific locations, or maps, delineating the surge corresponding to specified annual exceedance probabilities (AEPs) for all locations across the coastal region of interest. To attain highly accurate hazard quantification, the current standard employs a high-fidelity numerical model, like ADCIRC[3], for predicting surge levels. However, the computational cost of the model is significant, requiring a reduction in the number of storm scenarios. This challenge is even critical for planning coastal adaptation measures, as it needs the repetition (updating) of the hazard estimation for each measure to assess their impact.

Different approaches have been proposed to support efficient regional hazard quantification, including the joint probability method[4], surrogate model-based techniques[5, 6], and hazard-consistent probabilistic scenario optimization (HPSO) approaches[7]. Particularly, HPSO has been shown to accommodate great reductions in the ensemble size without compromising the statistical accuracy of the hazard estimation, and is especially attractive when updating regional flood studies, for example, to examine the impact of flood intervention strategies as discussed earlier. The foundational principle for HPSO is that surge predictions for a response-proxy are available for a large number of storm scenarios. The selection of a reduced subset of representative scenarios is formulated in HPSO as an optimization, identifying the storms that yield hazard estimates with the smallest deviation from the target hazard (i.e., achieving hazard-consistency) for the specific response-proxy. The occurrence rates (weights) of the identified storms are also adjusted in this process. Then, numerical simulations for the actual response to support the desired hazard estimation can be performed only for the reduced-size ensemble. The response-proxy may correspond to surge predictions from a previous flood study when examining its future update with only a limited number of carefully selected scenarios.

Formally, HPSO is performed through a two-stage optimization[7]: an inner-loop adjusts the weights for a given storm subset through linear programming, while the outer-loop identifies the best subset with respect to an appropriate measure for the hazard-consistency to a predefined target. All previous HPSO studies have focused on applications with only a small number of locations (up to hundreds). This work extends these efforts by examining an implementation across an extended coastal region, with up to tens of thousands of locations of interest. This high-dimensional output imposes a significant computational burden on the inner-loop. To accommodate computational tractability, dimensionality reduction is examined by choosing a representative subset of locations to guide the storm ensemble selection. These locations are chosen using cluster analysis, leveraging spatial and surge response correlations. Additionally, a modification of the objective function for the outer-loop is examined to better integrate the spatial correlation of the hazard description. This is accomplished by using the correlation across the locations for each hazard map as the objective function. The performance of the extended HPSO will be examined in the case study, where the updating of the flood study is considered, assisted by HPSO using the previous flood study results as the response-proxy. Also, the impact of the response-proxy on the HPSO performance will be investigated in detail.

2 REVIEW OF REGIONAL COASTAL HAZARD ESTIMATION AND HPSO

2.1 Regional hazard quantification

Coastal hazard is quantified through an ensemble of N_t storm scenarios, with the sth storm having the weight λ^s. This weight may correspond to the annual occurrence rate based on the regional climatology or to the adjusted rate through HPSO as discussed later. The notation **S** will be used to succinctly describe the storm ensemble along with the weights. To support the hazard estimation, the surge is predicted for each storm through an appropriate numerical model, with the surge level for the sth storm at the ith location denoted as z_i^s and n_z representing the number of locations where the hazard are estimated. The annual exceedance probability (AEP) for the surge at the ith location exceeding a threshold b is estimated with **S** as[2]:

$$P_i(b\,|\,\mathbf{S}) = 1 - \exp\left(-\sum_{s=1}^{N_t} \lambda^s I[z_i^s > b]\right) \approx \sum_{s=1}^{N_t} \lambda^s I[z_i^s > b] \tag{1}$$

where $I[.]$ is the indicator function and the notation "|" denotes the dependence of the estimation on **S**. Based on Eq. (1), the surge level b_i^p corresponding to a specific AEP of interest, p, can be obtained by solving the inverse problem of $P_i(b_i^p\,|\,\mathbf{S}) = p$. The regional hazard is represented by the hazard-curve for a specific location with b_i^p across different p values, or the hazard-map for the distribution of the b_i^p across the coastal domain of interest for a specific p value.

2.2 HPSO implementation

The objective of HPSO is to leverage the response-proxy database $\{\tilde{z}_i^s : i = 1,...,\tilde{n}_z, s = 1,...N_t\}$ over the storm suite **S** to identify a reduced subset of n_r (n_r<N_t) representative scenarios and adjust their weights so that hazard-consistency to a desired target is achieved. The notation "~" represents quantities associated with the response-proxy and will be used hereinafter. Note that locations for which information is available in the response-proxy do not need to coincide with the locations for which the hazard needs to be estimated. The hope is that the hazard-consistency with respect to the response-proxy is translated to the actual response (updated flood study). The resultant storm ensemble will be denoted herein as $\mathbf{S}_r$. Hazard-consistency in this setting is quantified by a match to thresholds $\tilde{b}_i^{p_k}$ for n_p different chosen AEP values $\mathbf{p}$={p_k: k=1,…,n_p} across all $\tilde{n}_z$ locations. It should be noted that most practical applications utilize the same target AEP levels for all locations, though different targets can be easily accommodated as discussed later. For simplicity of discussions, the standard formulation considering the same AEP targets will be used herein. The targets $\{\tilde{b}_i^{p_k} : k = 1,...,n_p\}$ are obtained by utilizing hazard estimates $\tilde{P}_i(\tilde{b}\,|\,\mathbf{S})$ obtained through Eq. (1) with **S**. This is demonstrated in part (a) of Figure 1, with the shaded region representing the AEP range of interest and the horizontal dotted lines the specific AEP levels for the hazard definition. The approximate hazard obtained through the reduced-size storm ensemble $\mathbf{S}_r$ is also illustrated in the same plot by a blue dash-dotted line. HPSO identifies the optimal $\mathbf{S}_r$ that provides the best match (the best hazard-consistency) between the approximate and target hazard curves in this figure. With respect to the target hazard definition, for nearshore and onshore nodes that have remained dry in some of the storms, this hazard curve might not cover the entire AEP range of interest, a situation also illustrated in part (b) of Figure 1. For such locations, additional AEP levels, the so-called anchor levels, are introduced to the HPSO formulation, corresponding to the extremes of the hazard curve within the AEP range of interest, represented by the red triangles in Figure 1 and to different AEP levels that can be accommodated within the HPSO implementation.

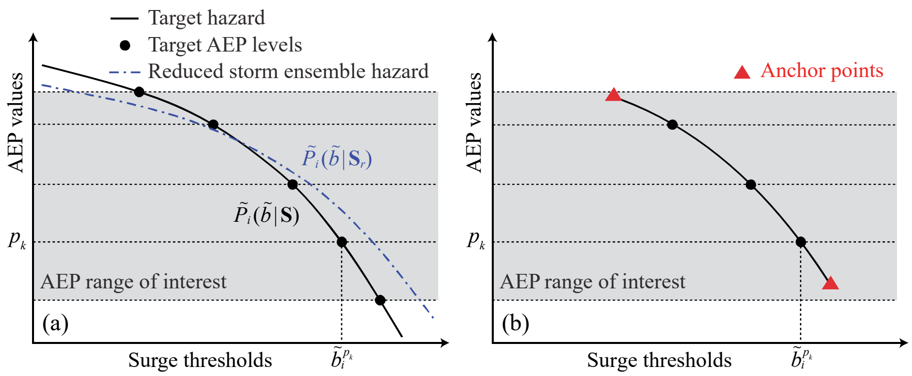

Figure 1: The target hazard (obtained using storm ensemble **S**) at the ith location quantified by the pair of target AEP levels and corresponding surge thresholds $\{(p_k, \tilde{b}_i^{p_k}) : k = 1,...,n_p\}$ [part (a)] and definition of anchor points [part (b)] when the hazard curve does not cover the entire AEP range of interest (shaded area). In part (a), the estimated hazard obtained using the reduced-size storm ensemble is also shown.

Computationally, HPSO is solved through double-loop optimization, as discussed in the introduction. The inner-loop optimizes weights for a specific storm ensemble using the sum of absolute deviations from the target as the objective function and a linear programming (LP) formulation. The outer-loop identifies the best storm ensemble using the same objective and a genetic algorithm (GA) based search. To formulate HPSO, let $\mathbf{x}$ denote the N_t-dimensional vector whose jth element indicates whether the jth storm belongs (x_j=1) or not (x_j=0) in the subset. Also, let $\boldsymbol{\lambda}(\mathbf{x}) = \{\lambda^r(\mathbf{x}) : r = 1,...,n_r\}$ represent the weights for the selected storms. Note that only n_r weights for the selected storms are considered in $\boldsymbol{\lambda}(\mathbf{x})$. Relative weights w_k and γ_i, respectively, for the kth AEP value and ith location can also be introduced for a relative prioritization. Define, now, the objective function across the target hazard as:

$$f(\mathbf{x},\boldsymbol{\lambda}(\mathbf{x})) = \sum\nolimits_{i=1}^{\tilde{n}_z} \gamma_i \sum\nolimits_{k=1}^{n_p} (w_k / p_k) \cdot \tilde{u}_i^k(\mathbf{S}_r) \Big/ \sum\nolimits_{i=1}^{\tilde{n}_z} \gamma_i \sum\nolimits_{k=1}^{n_p} w_k \tag{2}$$

where the absolute deviation for each AEP level and location is:

$$\tilde{u}_i^k(\mathbf{S}_r) = | p_k - \tilde{P}_i(\tilde{b}_i^{p_k} | \mathbf{S}_r) | \tag{3}$$

with $\tilde{P}_i(\tilde{b}_i^{p_k} | \mathbf{S}_r)$ obtained through Eq. (1) for the subset $\mathbf{S}_r$ defined through the pair $\{\mathbf{x},\boldsymbol{\lambda}(\mathbf{x})\}$.

The two-stage optimization in HPSO is formulated as follows. The inner-loop identifies the optimal n_r-dimensional weight vector $\boldsymbol{\lambda}(\hat{\mathbf{x}})$ for the storms in any subset $\hat{\mathbf{x}}$ though optimization:

$$\begin{aligned} &\boldsymbol{\lambda}^*(\hat{\mathbf{x}}) = \arg\min_{\boldsymbol{\lambda}(\hat{\mathbf{x}})} f(\hat{\mathbf{x}},\boldsymbol{\lambda}(\hat{\mathbf{x}})) \\ &\text{subject to } \quad \lambda^r(\hat{\mathbf{x}}) > 0 : r = 1,...,n_r \end{aligned} \tag{4}$$

The use of $f(\mathbf{x},\boldsymbol{\lambda}(\mathbf{x}))$ allows for transforming the inner-loop into LP and solving it efficiently when $\tilde{n}_z$ is small[7]. The outer-loop searches for the optimal storm subset with the weights identified through the inner-loop and corresponds to an integer optimization:

$$\begin{aligned} &\mathbf{x}^* = \arg\min_{\mathbf{x}} h(\mathbf{x},\boldsymbol{\lambda}^*(\mathbf{x})) \\ &\text{subject to } \sum\nolimits_{j=1}^{Nt} x_j = n_r \quad \text{with} \quad x_j \in \{0,1\} \ (j = 1,...,N_t) \end{aligned} \tag{5}$$

where $\boldsymbol{\lambda}^*(\mathbf{x})$ corresponds to the optimal weights from Eq. (4) for the storm subset $\mathbf{x}$ ($\hat{\mathbf{x}} = \mathbf{x}$) and $h(\mathbf{x},\boldsymbol{\lambda}^*(\mathbf{x}))$ represents an appropriate objective which may correspond to $f(\mathbf{x},\boldsymbol{\lambda}^*(\mathbf{x}))$ in Eq. (2) in the original HPSO or a new objective that will be introduced in Section 3.1.

To perform the double-loop optimization, at each iteration of the outer-loop GA optimization, the optimal storm weights are decided based on the inner-loop LP optimization. The convergence of the overall optimization scheme is decided by that of the outer-loop. Upon convergence, the optimal subset $\mathbf{S}_r^*$ corresponds to the pair $\{\mathbf{x}^*,\boldsymbol{\lambda}^*(\mathbf{x}^*)\}$. The repetition of the inner-loop optimization within this scheme requires a highly efficient implementation. For large geographic domains, corresponding to large $\tilde{n}_z$, the computational cost of the LP dramatically increases, creating an insurmountable computational burden for HPSO.

3 HPSO EXTENSIONS FOR EXTENDED REGIONS AND NOVEL OBJECTIVE FUNCTION

3.1 Implementation for extended regions

To accommodate a computationally efficient implementation for extended coastal regions, with a large value for $\tilde{n}_z$, the decrease in the number of locations involved in the optimization to a smaller number n_c ($n_c \ll \tilde{n}_z$) is considered. If the n_c locations are properly chosen, the computational complexity of the inner-loop optimization can be substantially reduced, leading to a significant improvement in the computational efficiency of the overall HPSO. Let subscript $\cdot_q$ distinguish the n_c representative locations, with $u_q^k(\mathbf{S}_r)$ corresponding to $u_i^k(\mathbf{S}_r)$ for each of these locations, leading to the modification of the objective function in Eq. (2) as:

$$f_c(\mathbf{x},\boldsymbol{\lambda}(\mathbf{x})) = \sum_{q=1}^{n_c} \underline{\gamma}_q \sum_{k=1}^{n_p} (w_k / p_k) \cdot \tilde{u}_q^k(\mathbf{S}_r) \Big/ \sum_{i=1}^{\tilde{n}_z} \underline{\gamma}_q \sum_{k=1}^{n_p} w_k \tag{6}$$

where $\underline{\gamma}_q$ denotes the adjusted weights per representative location. The proper definition of $\underline{\gamma}_q$ depends on how the representative locations are chosen, discussed later in this section. The computational cost of the LP is then reduced to $O([(n_p \cdot n_c)+n_r]^{2.5})$ from $O([(n_p \cdot \tilde{n}_z)+n_r]^{2.5})$[8], which in typical applications ($\underline{n_r} \ll \tilde{n}_z$), brings significant computational savings. Similarly, the objective function for the outer-loop optimization can be simplified. Once the storms are chosen by HPSO, considering the representative locations only, the rates of the chosen storms can be adjusted by solving the original problem of Eq. (4) using all locations. Though its computational cost is significant, as it is performed only once, the additional burden is not prohibitive.

The selection of the representative locations is accomplished by clustering all locations into n_c clusters using *k*-means[9] and selecting the center of each cluster as the representative location. This clustering utilizes a vector of features for each location and partitions all locations into clusters so that the distance of each location to the cluster center is minimized, where the distance is defined with respect to the feature vector. This selection leads to the adjusted weights, established by summing the weights of all the locations belonging to the *q*th cluster.

The selection of the feature vector is critical to ensure that locations in each cluster are similar enough and that the chosen locations establish a good representation. To achieve these, using both geospatial and response/target information is considered. Additional detailed discussions on the feature selection are included in [10]. Since neighboring locations are expected to exhibit similar surge responses by a storm, considering spatial locations is intuitive. The features in this case corresponds to the latitude and longitude of the *i*th location, denoted as $[e_i^1 \; e_i^2]$. However, spatial proximity does not always guarantee the response similarity due to complex coastal morphologies. Moreover, hazard targets could be dissimilar for neighboring locations. This motivates the inclusion of response and target hazard information. The latter $\{\tilde{b}_i^{p_k}: k=1,\ldots,n_p; i=1,\ldots,\tilde{n}_z\}$ can be directly used. But, the direct use of response $\{\tilde{z}_i^s : s=1,...,N_t\}$ is impractical if N_t is large, as it substantially increases the feature vector dimension and the computational cost for k-means. Following [11], the eigenvectors of the first few principal components obtained by principal component analysis are used, denoted as $\mathbf{\Psi} \in \Re^{n_z \times n_e}$ with n_e denoting the number of retained components. The feature vector for the ith location is finally defined as:

$$v_i^h = \begin{cases} e_i^h / Var(e_i^h) & \text{if } h=1,2 \\ \tilde{b}_i^{p_{h-2}} / [n_p Var(\tilde{b}_i^{p_{h-2}})] & \text{if } h=3,...,n_p+2 \\ \underline{\omega}^{(h-n_p-2)} [\mathbf{\Psi}]_{i(h-n_p-2)} / Var(v^h) & \text{if } h = n_p+3,...,n_p+n_e+2 \end{cases} \tag{7}$$

where $[.]_{ij}$ represents the element of the ith row and jth column, and $\underline{\omega}^j = \omega^j / \sum_{j=1}^{n_{ret}} \omega^j$ with ω^j corresponding to the eigenvalue for the jth principal component. Note that each component of feature vector is standardized, and equal weights are assigned to the three information sources.

3.2 Incorporation of spatial correlation over hazard maps in storm selection

Spatial correlation information for the targeted hazard can be readily incorporated within the outer-loop formulation of Eq. (5) by introducing a new objective $g(\mathbf{x}, \boldsymbol{\lambda}^*(\mathbf{x}))$. Note that this leads to a deviation from the original HPSO formulation. Although it has been shown that spatial correlation can be captured to some extent, even with objective $f(\mathbf{x}, \boldsymbol{\lambda}^*(\mathbf{x}))$ [12], this correlation may still exhibit (depending on the case) larger deviations from the target[13]. This motivates the modification of the objective that can explicitly consider spatial correlation information. The proposed objective considers the correlation coefficient between the resultant (for the chosen storm ensemble) and the original hazard maps for each AEP level, and is formulated by averaging the correlation coefficients across the levels:

$$g(\mathbf{x}, \boldsymbol{\lambda}^*(\mathbf{x})) = -\sum_{k=1}^{n_p} w_k \cdot \rho_k(\mathbf{S}_r) \Big/ \sum_{k=1}^{n_p} w_k \tag{8}$$

where the correlation coefficient for each AEP level is given by:

$$\rho_k(\mathbf{S}_r) = \frac{\sum_{i=1}^{\tilde{n}_z} \gamma_i [\hat{\tilde{b}}_i^{p_k}(\mathbf{S}_r) - \frac{1}{\tilde{n}_z}\sum_{i=1}^{\tilde{n}_z} \hat{\tilde{b}}_i^{p_k}(\mathbf{S}_r)][\tilde{b}_i^{p_k} - \frac{1}{\tilde{n}_z}\sum_{i=1}^{\tilde{n}_z} \tilde{b}_i^{p_k}(\mathbf{S}_r)]}{\sqrt{\frac{1}{\tilde{n}_z}\sum_{i=1}^{\tilde{n}_z} \gamma_i [\hat{\tilde{b}}_i^{p_k}(\mathbf{S}_r) - \frac{1}{\tilde{n}_z}\sum_{i=1}^{\tilde{n}_z} \hat{\tilde{b}}_i^{p_k}(\mathbf{S}_r)]^2} \sqrt{\frac{1}{\tilde{n}_z}\sum_{i=1}^{\tilde{n}_z} \gamma_i [\tilde{b}_i^{p_k} - \frac{1}{\tilde{n}_z}\sum_{i=1}^{\tilde{n}_z} \tilde{b}_i^{p_k}(\mathbf{S}_r)]^2}} \tag{9}$$

with $\hat{\tilde{b}}_i^{p_k}(\mathbf{S}_r)$ represents the threshold corresponding to $\tilde{P}_i(\tilde{b}_i \mid \mathbf{S}_r) = p_k$ for the subset $\mathbf{S}_r$.

The implementation over extended coastal regions discussed in Section 3.1 can be accommodated by modifying the correlation coefficient in Eq. (9) by replacing the summation over all locations with the summation over the representative locations, as done for Eq. (6).

4 ILLUSTRATIVE CASE STUDY

4.1 Coastal region of interest and database

The case study considers updating the storm surge hazard estimates for a specific St. Bernard Parish neighborhood in the Greater New Orleans eastern coastal region between the Mississippi River and Lake Borgne, as shown in part (a) of Figure 2. The original flood study, whose surge simulation results are utilized as a response-proxy for the update, is part of the U.S. Army Corps of Engineers (USACE) Coastal Hazards System[14]. The database consists of N_t=645 synthetic storms, created based on the regional climatology, with the annual occurrence rate for each storm, and their tracks are illustrated in part (b) of Figure 2. The surge simulations are performed using ADCIRC. In the domain of interest, the original grid consists of $\tilde{n}_z$=7,663 locations, while the updated grid consists of n_z=31,091 locations. The updated grid has better resolution, incorporating more detailed and updated flood protection measures.

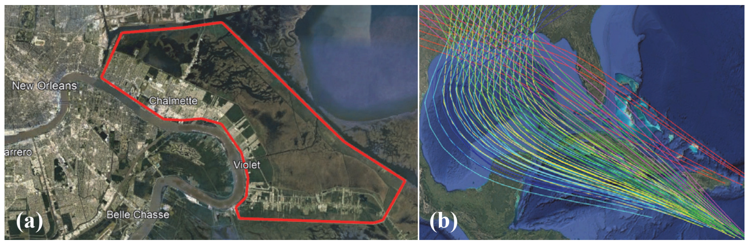

Figure 2: The coastal domain of interest (with red boundary) [part (a)] and storm tracks in the database [part (b)].

4.2 HPSO implementation details

Hazard-consistency is defined for AEPs in the range [0.001, 0.1] and specifically, at target levels, $\mathbf{p}$=[0.001 0.002 0.005 0.01 0.02 0.05 0.1]^T. To explore the HPSO implementation under different settings for a comprehensive validation, two values are examined for the size of the reduced storm subset, n_r=20 or 40, and four values for the number of representative locations, n_c=20, 50, 150, or 300. In practice, a single value would be used for either of these variables based on the available computational resources for updating the flood hazard (related to n_r) or performing HPSO (related to n_c). For the objective functions of the outer-loop, both $f(\mathbf{x},\boldsymbol{\lambda}^*(\mathbf{x}))$ in Eq. (2) and $g(\mathbf{x},\boldsymbol{\lambda}^*(\mathbf{x}))$ in Eq. (8) are considered. The respective implementations will be distinguished as HPSO$_{dev}$ and HPSO$_{cor}$, respectively. Their final objective function values will be denoted as $dev-O$ (deviation objective) and $cor-O$ (correlation objective), respectively, and used for evaluating the achieved hazard-consistency with respect to the response-proxy. Finally, uniform weights across all locations and target AEP levels are used (i.e., γ_i=1 and w_k=1).

The selection of storms and their weights by HPSO is established utilizing the response-proxy $\tilde{z}_i^s$. Their practical utility needs to be examined using the actual response z_i^s. For this, z_i^s is predicted for the entire N_t=645 storms, and the reference hazard (surge thresholds $b_i^{p_k}(\mathbf{S})$ corresponding to p_k AEP) is obtained using Eq. (1) with λ^s. The use of the reduced ensemble identified through HPSO $\mathbf{S}_r^*$ provides corresponding approximate estimates $b_i^{p_k}(\mathbf{S}_r^*)$. As validation measures, the final objective function values of HPSO are utilized again, evaluated with respect to the actual response by replacing $\tilde{n}_z$ with n_z in Eqs. (2) and (9), $\tilde{P}_i(\tilde{b}_i^{p_k}\,|\,\mathbf{S}_r)$ with $P_i(b_i^{p_k}(\mathbf{S})\,|\,\mathbf{S}_r^*)$ in Eq. (3), and $\hat{\tilde{b}}_i^{p_k}(\mathbf{S}_r)$ with $b_i^{p_k}(\mathbf{S}_r^*)$ and $\tilde{b}_i^{p_k}$ with $b_i^{p_k}(\mathbf{S})$ in Eq. (9). Notations $dev-P$ (deviation performance) and $cor-P$ (correlation performance) will be utilized to denote the updated objectives for the actual response. For a more detailed validation, the performance for individual target AEP levels will also be examined. The correlation coefficient between the approximate and reference hazard maps is utilized for this purpose, defined as:

$$\overline{CC}_k = \frac{\frac{1}{n_z}\sum_{i=1}^{n_z}\gamma_i[b_i^{p_k}(\mathbf{S})-\frac{1}{n_z}\sum_{i=1}^{n_z}b_i^{p_k}(\mathbf{S})][b_i^{p_k}(\mathbf{S}_r^*)-\frac{1}{n_z}\sum_{i=1}^{n_z}b_i^{p_k}(\mathbf{S}_r^*)]}{\sqrt{\frac{1}{n_z}\sum_{i=1}^{n_z}\gamma_i[b_i^{p_k}(\mathbf{S})-\frac{1}{n_z}\sum_{i=1}^{n_z}b_i^{p_k}(\mathbf{S})]^2}\sqrt{\frac{1}{n_z}\sum_{i=1}^{n_z}\gamma_i[b_i^{p_k}(\mathbf{S}_r^*)-\frac{1}{n_z}\sum_{i=1}^{n_z}b_i^{p_k}(\mathbf{S}_r^*)]^2}} \tag{10}$$

To examine the impact of the use of the response-proxy within the HPSO, an idealized HPSO implementation is also considered, utilizing the actual response instead of the response-proxy in all optimizations. Note that this implementation is examined here solely for establishing a comprehensive validation. It is accommodated by replacing $\tilde{z}_i^s$ with z_i^s within HPSO and adjusting the different variables accordingly. It is important to mention that it necessitates the use of representative locations, as n_z makes the original HPSO implementation computationally intractable. The terminology *ideal* HPSO is introduced to describe this implementation, while *practical* HPSO will be utilized for the actual HPSO formulation (using the response-proxy).

5 RESULTS AND DISCUSSION

5.1 Optimization implementation with respect to response-proxy

Results for the optimization aspects of HPSO are examined in Figure 3 for both HPSO$_{dev}$ and HPSO$_{cor}$ implementations for varying numbers of representative locations (n_c) [x-axis] and selected storms (n_r) [different markers]. For each implementation, the respective objective is presented, $cor-O$ for HPSO$_{dev}$ [part (a)] and $dev-O$ for HPSO$_{cor}$. In this subsection, only the *practical* HPSO is examined. Focusing on the use of the representative locations to facilitate the HPSO implementation to extended geographic domains, for both storm selection variants, the performance improves when more representative locations are considered, but the rate of improvement quickly decreases, reaching a plateau. This continuous improvement is expected, as a larger n_c allows for a larger amount of information in the selection of the optimal ensemble. The decrease in the rate of improvement ending up with a plateau is important, as they confirm the efficacy of utilizing a smaller number of representative locations, instead of all the locations. Therefore, the proposed extension of HPSO, utilizing representative locations, can improve computational efficiency substantially without significant loss of accuracy as long as a sufficient number of representative locations is utilized.

Beyond the number of representative locations, the size of the selected storm subset also contributes to performance improvement. This is again expected, as a larger n_r offers greater flexibility in the optimization for establishing hazard-consistency. These trends showcase that for the response-proxy, HPSO identifies a storm ensemble supporting good hazard-consistency.

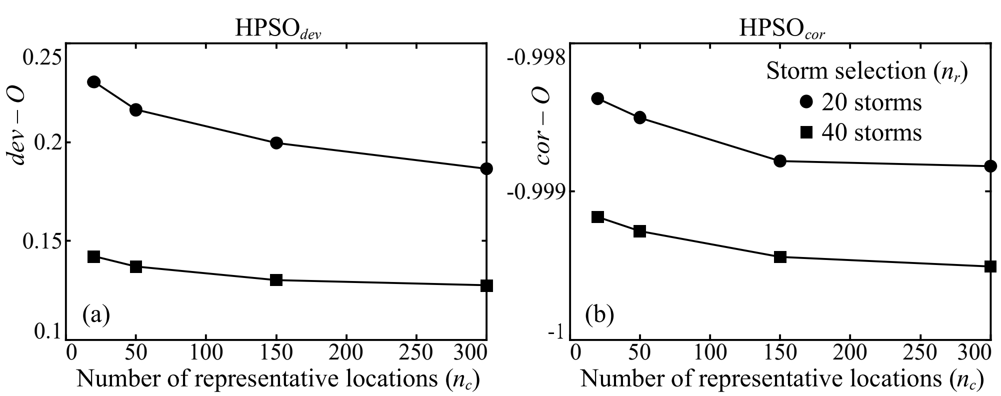

Figure 3: The objective functions (response-proxy) of HPSO$_{dev}$ and HPSO$_{cor}$ corresponding, respectively, to $dev-O$ [part (a)] and $cor-O$ [part (b)] for varying numbers of representative locations (n_c) [x-axis] and storm subset sizes (n_r) [different markers].

5.2 HPSO performance with respect to the actual response

The discussion moves next to the examination of the performance with respect to the actual response. Hereinafter, both the *practical* and *ideal* HPSO will be examined and compared, as this comparison ultimately quantifies the impact of the response-proxy in HPSO. Figure 4 examines the overall performance, replicating Figure 3 with respect to the updated objectives $dev-P$ for HPSO$_{dev}$ [part (a)] and $cor-P$ for HPSO$_{cor}$ [part (b)], for varying numbers of representative locations (n_c) [*x*-axis] and selected storms (n_r) [different markers]. Figure 5 examines the performance for individual AEP levels, presenting results of HPSO$_{dev}$ [part (a)] and HPSO$_{cor}$ [part (b)] using n_c=300 for the AEP levels [*x*-axis] for varying n_r [different line types].

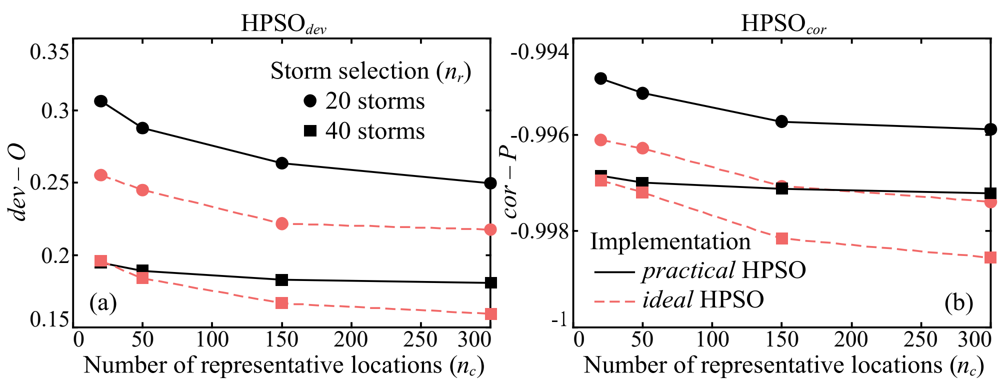

Figure 4: The performance objectives (actual response) of HPSO$_{dev}$ and HPSO$_{cor}$ corresponding respectively to $dev-P$ [part (a)] and $cor-P$ [part (b)] for varying numbers of representative locations (n_c) [*x*-axis] and storm subset sizes (n_r) [different markers]. Results for *practical* and *ideal* HPSO [different color and line types] are shown.

Focusing first on the *practical* HPSO [examining black lines and markers only], the overall trends with respect to the influence of n_c or n_r are identical to those observed in Figure 3. The similarity with respect to n_c is of particular importance, demonstrating that the extended HPSO is valid for its practical application and the use of the response-proxy does not promote different behavior by the representative locations. Comparing Figures 3 and 4, there is an evident performance decrease for the *practical* HPSO, as the updated objective values in Figure 4 are slightly higher than the ones reported in Figure 3. This shows an influence, as expected, of the response-proxy to guide the ensemble selection. This influence is better evaluated when comparing the *practical* and *ideal* HPSO implementations in Figure 4 [examining black and red lines and markers]. The gap between the curves is not significant, and it decreases when n_r increases, showcasing that the use of the response-proxy does not yield significantly sub-optimal solutions, especially when the size of the reduced storm ensemble is not very small. Unsurprisingly, the differences between *practical* and *ideal* HPSO are smaller for smaller n_c, as HPSO is performed with less amount of information, ultimately providing smaller comparative benefits when the actual response can be utilized. The overall optimization accuracy as well as the performance trends are similar when using either the response-proxy or the actual response. As this information is leveraged within HPSO to make decisions, for example, choosing n_r for the best hazard approximation, this similarity is significant. Additionally, even in the ideal scenario where the actual response could be used, the benefits would not be enhanced substantially.

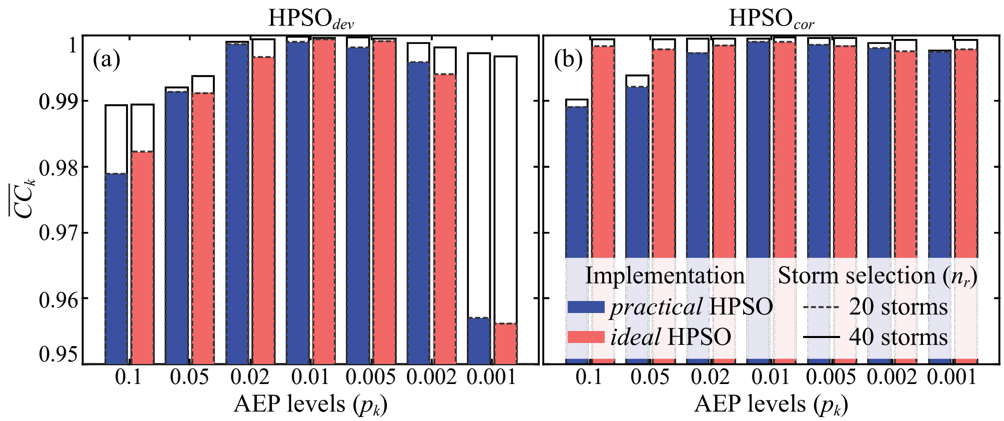

Figure 5: The correlation coefficient between the approximate and reference hazard maps for all target AEP levels [x-axis] for HPSO$_{dev}$ [part (a)] and HPSO$_{cor}$ [part (b)] with varying storm subset sizes (n_r) [different line types]. Results for the *practical* and *ideal* HPSO [different colors] with n_c=300 are shown.

The discussion shifts next to the performance per AEP level (Figure 5). Focusing first on the correlation coefficients for HPSO$_{cor}$ [part (b)], it is evident that the predicted hazard maps by both the *practical* and *ideal* HPSO are highly correlated to reference maps across all AEP levels. This verifies that the desired high correlation in hazard description is achieved through the new objective function. This is further supported by the fact that the performance of both HPSO implementations consistently improves across all AEP levels as n_r increases. Next, looking at the correlation coefficients for HPSO$_{dev}$ [part (a)], they also show high correlation but exhibit slightly different trends, the correlation of the *practical* and *ideal* HPSO is relatively low for AEP levels close to the boundary of the target range, especially when n_r is small. This decrease in correlation is expected, as the similarity of the hazard maps is not directly aligned with the objectives of HPSO$_{dev}$, and justifies the proposed HPSO$_{cor}$ in this work. However, the fact that both the *practical* and *ideal* HPSO implementations suffer from the same performance degradation indicates that it is not associated with the use of a response-proxy. This may be attributed to other factors, such as the characteristics of surge response database (particularly dry instances for specific nodes) or boundary effects (impacting the optimization) associated with the extremes of the AEP range of interest, which are discussed in detail in [10]. The AEP levels, for which the accuracy is relatively low, benefit more from the increase in n_r, achieving more considerable improvements. These observations indicate that HPSO$_{cor}$ establishes a better balance (consistent and robust) across all target AEP levels and provides a relative preference.

5.3 Illustration of achieved consistency for hazard curves and maps

The HPSO performance for the predicted hazard is finally evaluated for hazard curves for three different locations in Figure 6 and the hazard map for the AEP level p_k=0.01 in Figure 7. Results correspond to the optimal ensemble with n_r=40 storms selected through HPSO$_{cor}$ and n_c=300 representative locations. Since results for HPSO$_{dev}$ are practically identical, they are not presented here. In all instances, the hazard products obtained through both *ideal* and *practical* HPSO implementations as well as the reference hazard products are shown. In Figure 7, the errors of HPSO-based predictions from the reference are also shown in parts (d) and (e) (respectively, for *ideal* and *practical* HPSO implementations) for easier comparisons.

Results in both figures clearly demonstrate the high accuracy in the hazard predictions that can be achieved through HPSO, proving, once more, the validity of the proposed HPSO extension and its practical applicability using a response-proxy. In Figure 6, the two hazard curves, established through the different HPSO implementations, are very close to each other, and both are highly accurate when compared to the reference hazard curve. Some decrease in the accuracy at higher AEP levels is observed in parts (b) and (c) of the figure, but this occurs for both *ideal* and *practical* HPSO implementations, indicating that this behavior is not associated with the use of a response-proxy. Similarly, in Figure 7, the hazard maps established by the two HPSO implementations demonstrate high accuracy across the domain, with the small subdomains of lower accuracy being the same for both the *ideal* and *practical* HPSO implementations. These observations showcase that the use of the response-proxy does not create any vulnerabilities in HPSO, associated with lower accuracy for the predicted hazard due to the use of a reduced storm ensemble (i.e., the underlying objective of HPSO).

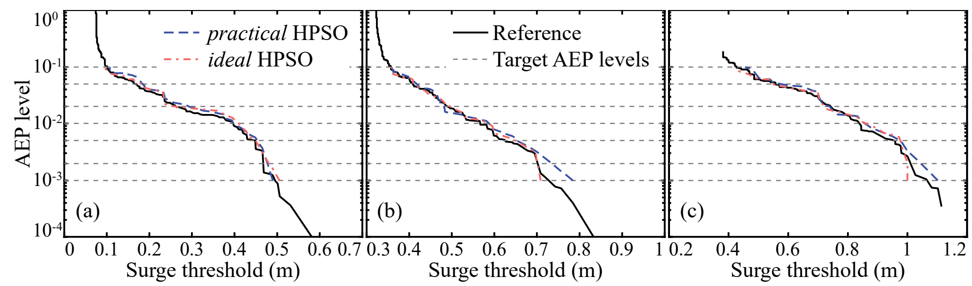

Figure 6: Comparison between the reference and approximated hazard curves through the *practical* and *ideal* HPSO implementations [different colors and line types] using the optimal storm ensemble of size n_r=40 identified with HPSO$_{cor}$ and n_c=300 representative locations.

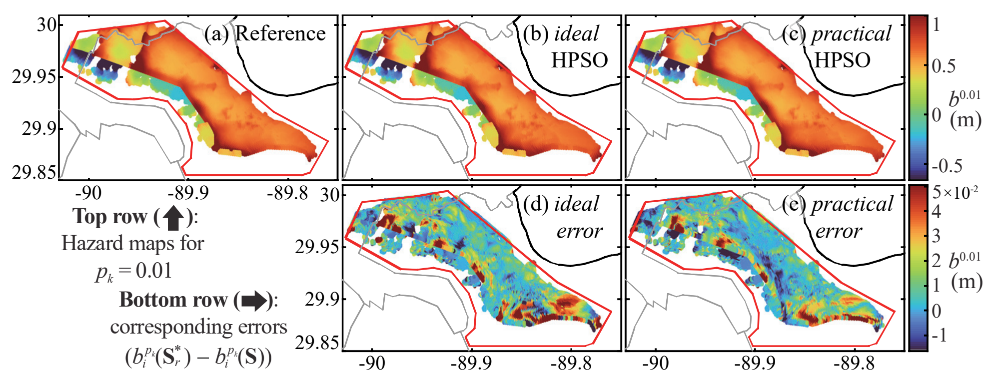

Figure 7: Hazard maps [top row] and errors [bottom row] corresponding to AEP level 0.01 established by: (a) Reference, (b) *ideal* HPSO, (c) *practical* HPSO with HPSO$_{cor}$, n_r=40, and n_c=300.

5 Conclusions

This paper discussed the extension of hazard-consistent probabilistic scenario optimization (HPSO) to enable its implementation over extended coastal domains with a large number of locations where the hazard is estimated, and its performance in a practical application. The extension included selecting a reduced number of representative locations through cluster analysis to ensure computational tractability, as well as modifying the objective function to promote high correlation in hazard description. The practical application, which considered the updating of the storm surge hazard assessment using surge simulations from the previous hazard estimation as a response-proxy, demonstrated that the proposed extension effectively supports the desired storm selection and is minimally affected by the use of the response-proxy.

REFERENCES

[1] Frazier TG, Wood N, Yarnal B, Bauer DH. Influence of potential sea level rise on societal vulnerability to hurricane storm-surge hazards, Sarasota County, Florida. Applied Geography. 2010;30:490-505.

[2] Resio DT, Boc SJ, Borgman L, Cardone V, Cox A, Dally WR et al. White paper on estimating hurricane inundation probabilities. Consulting Report prepared by USACE for FEMA2007.

[3] Luettich RA, Jr. , Westerink JJ, Scheffner NW. ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1. Theory and methodology of ADCIRC-2DDI and ADCIRC-3DL. Vicksburg,MS: Dredging Research Program Technical Report DRP-92-6, U.S Army Engineers Waterways Experiment Station; 1992.

[4] Toro GR, Resio DT, Divoky D, Niedoroda A, Reed C. Efficient joint-probability methods for hurricane surge frequency analysis. Ocean Engineering. 2010;37:125-34.

[5] Jung W, Taflanidis AA, Nadal-Caraballo NC, Yawn MC, Aucoin LA. Regional storm surge hazard quantification using Gaussian process metamodeling techniques. Natural Hazards. 2023:1-29.

[6] Resio D, Irish J, Cialone M. A surge response function approach to coastal hazard assessment – part 1: basic concepts. Natural Hazards. 2009;51:163-82.

[7] Legg MR, Nozick LK, Davidson RA. Optimizing the selection of hazard-consistent probabilistic scenarios for long-term regional hurricane loss estimation. Structural Safety. 2010;32:90-100.

[8] Cohen MB, Lee YT, Song Z. Solving linear programs in the current matrix multiplication time. Journal of the ACM (JACM). 2021;68:1-39.

[9] Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society Series C (Applied Statistics). 1979;28:100-8.

[10] Jung W, Taflanidis AA, Nadal-Caraballo NC, Yawn MC, Aucoin LA. Hazard-consistent scenario selection for long-term storm surge risk assessment over extended coastal regions. Coastal Engineering. 2024:104620.

[11] Kyprioti AP, Taflanidis AA, Nadal-Caraballo NC, Campbell M. Storm hazard analysis over extended geospatial grids utilizing surrogate models. Coastal Engineering. 2021:103855.

[12] Han Y, Davidson RA. Probabilistic seismic hazard analysis for spatially distributed infrastructure. Earthquake Engineering & Structural Dynamics. 2012;41:2141-58.

[13] Christou V, Bocchini P, Miranda MJ, Karamlou A. Effective sampling of spatially correlated intensity maps using hazard quantization: Application to seismic events. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. 2018;4:04017035.

[14] Nadal-Caraballo NC, Campbell MO, Gonzalez VM, Torres MJ, Melby JA, Taflanidis AA. Coastal Hazards System: A Probabilistic Coastal Hazard Analysis Framework. Journal of Coastal Research. 2020;95:1211-6.