Proceedings of Proceedings of ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering
OMAE 2020
June 28-July 3, 2020, Fort Lauderdale, FL, USA

OMAE2020-19190

DRAFT FOR OMAE 2020

REVISED: SPATIAL FEATURES OF EXTREME WAVES IN GULF OF MEXICO

Ryota Wada
The University of Tokyo
Tokyo, Japan
Email: r_wada@k.u-tokyo.ac.jp

Philip Jonathan
Shell Research Ltd., London, UK
Lancaster University, UK

Takuji Waseda
The University of Tokyo
Tokyo, Japan

ABSTRACT

Extreme value analysis of significant wave height using data from a single location often incurs large uncertainty due to small sample size. Including wave data from nearby locations increases sample size at the risk of introducing dependency between extreme events and hence violating modelling assumptions. In this work, we consider extreme value analysis of spatial wave data from the 109-year GOMOS wave hindcast for the Gulf of Mexico, seeking to incorporate the effects of spatial dependence in a simple but effective manner. We demonstrate that, for estimation of return values at a given location, incorporation of data from a circular disk region with radius of approximately 5° *(long.-lat.), centred at the location of interest, provides an appropriate basis for extreme value analysis using the STM-E approach of Wada et al. (2018).*

NOMENCLATURE

STM Space-Time Maximum
E Exposure

INTRODUCTION

Estimation of extreme metocean conditions corresponding to return periods of 100 years and beyond is required for design of offshore production systems. However, as the duration of available wave records is typically shorter than the return period of interest, extrapolation using extreme value analysis is essential. Traditionally, only data from the location of interest is utilised for extreme wave estimation. Since the sample size for such data is insufficient for precise estimation, several approaches, such as spatial pooling [1], cyclone track-shifting [2] and explicit track modelling [3], have been proposed to increase sample size by making use of the data from nearby locations. Regional frequency analysis (RFA) gathers statistical information from different sites assuming the probability distributions of extreme values at different sites in the region are identical, except for a scale parameter ([4], [5]) and has been applied to extreme wave data analysis (e.g. [6]). Another approach is to model spatial dependency explicitly (e.g. [7], [8]).

There are several concerns when a sample from multiple locations is modelled wrongly assuming spatial independence, when the actual data is spatially dependent. The effective sample size is overestimated and hence uncertainty bands for return values are underestimated. Techniques such as block-bootstrapping can be used to inflate uncertainties to more realistic levels. Of course, the spatial scale of the extreme event will play a key role in the efficiency of the approach. The appropriateness of cyclone track-shifting or explicit track modeling depends on the veracity of the underpinning assumptions. Approaches based on explicit modelling of spatial dependence are complex: (a) temporally independent storm peak events must first be identified, and (b) transformed to unit Frechet scale following non-stationary marginal modelling involving covariates such as storm direction, season, longitude and latitude. Then (c) spatial dependence is characterised using an approximate composite likelihood estimation. The full procedure can only be used successfully when sample size is relatively large. STM-E [9] is a simple spatial statistical model for extreme value estimation of significant wave height under tropical cyclones. STM-E estimates extreme values for the (independent) space-time maxima of storm events. It also exploits the absolute spatial distribution of the maximum value observed per storm event, relative to the space-time maximum.

In this paper, we focus on extreme waves in the Gulf of Mexico. Here, strong winds caused by hurricanes drive the largest waves [[10], [11]]. These hurricanes are several hundred miles in diameter, and hurricane tracks are widespread throughout the region. In addition, a central region of Gulf of Mexico with high ocean temperatures, referred to as *hurricane alley*, exhibits hurricanes with increased intensity.

STM-E (see Methodology section) was previously applied to extreme wave analysis in the Gulf of Mexico [[12]], providing a spatially-smooth distribution of extreme waves. In the STM-E approach, the absolute spatial distribution of maximum observed significant wave height for a storm (referred to as the storm "exposure") is defined with respect to a spatial neighbourhood or "region". In the past, it was assumed that the extent of the exposure region corresponded to the whole spatial domain of interest. However, this need not be the case, especially when there is evidence for spatial non-stationarity of extreme events. In the current work therefore, we assess the sensitivity of return value estimates at a location using STM-E with respect to circular exposure regions of different size centred at that location. We seek to identify suitable radii of exposure region which (a) allow data for as many locations as possible to be combined for STM-E analysis, and (b) do not violate the assumptions (e.g. of spatial homogeneity within the exposure region) underlying the STM-E approach.

When the exposure radius is small, it is likely that storm peak significant wave height is homogeneous within the exposure region. However, only a small number of locations will be included in the STM-E analysis. As a result, STM-E analysis will be very similar to that of a single location. Conversely, when the exposure radius is very large, the whole ocean basin might be included in the exposure region. However, it might also be the case that storm peak significant wave height now shows a spatial trend within the exposure region, invalidating the modelling assumptions underlying STM-E. In other words, the exchangeability of STM, and of exposure, is violated. Over some range of exposure radii, it is likely that both criteria (a) and (b) can be satisfied simultaneously: this is the objective of the current study.

DATA and METHODOLOGY

GOMOS

Our study uses the Gulf of Mexico Oceanographic Study (GOMOS) data, a comprehensive metocean description for Gulf of Mexico hurricanes by Oceanweather Inc. Wave conditions during all significant hindcast events are simulated with a third generation wave model with 1/16th rectangular degree grid (7km) resolution. "GOMOS08" includes wave data from 1900 to 2008 with 379 hurricane occurrences.

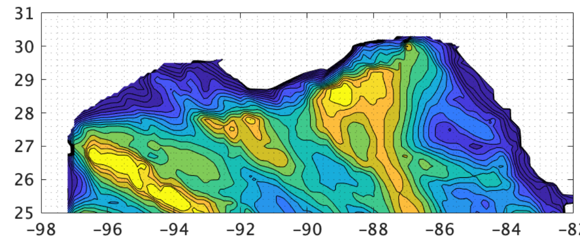

FIGURE 1. 100-YEAR RETURN VALUE FROM PER LOCATION APPROACH. COLOR SCALE AS IN FIGURE 2.

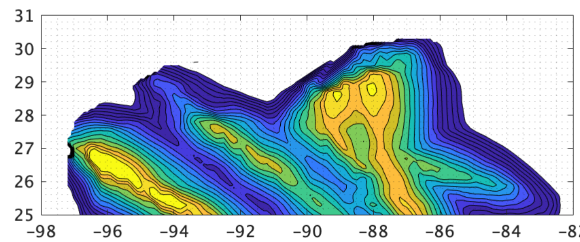

FIGURE 2. SPATIAL MAXIMUM OVER LARGEST 6 HURRICANES. COLOR SCALE AS IN FIGURE 1.

Figures 1 and 2 illustrate that an extreme wave analysis conducted per location using all 109 years of data produces a very similar spatial distribution of 100-year return value to the footprint of the maximum of the six strongest hurricanes in this area. This might suggest that the return value estimates from per-location analysis are overly influenced by the largest hurricanes.

STM-E

Here, we describe the STM-E approach briefly. We refer readers to [9] for further details of the method. The STM-E model was developed to characterize extreme waves offshore Japan, also dominated by tropical cyclones. The method relies on the estimation of two distributions from a sample of data, namely the distribution of spatio-temporal maximum (STM, S) and the exposure (E) within the exposure region. The STM-E estimate provides a parsimonious spatially-smooth distribution of extreme waves, with smaller uncertainties per location compared to estimates using data from a single location. Since S is a spatio-temporal maximum, all hurricane events within the exposure region contribute to the set $\{s_i\}$ of STM values, where index i indicates the hurricane number.. This increases the sample size for extreme value analysis compablack to single location analysis, which is impacted by only a subset of hurricanes that pass near by.

The STM is the largest significant wave height observed anywhere in the exposure region during the time period of the hurricane. A set of STM values $\{s_i\}_{i=1}^{n_S}$ for n_S hurricane events are extracted from the hindcast data. $\{s_i\}_{i=1}^{n_S}$ characterises the distribution of the STM random variable S. $F_{S|\psi_S}$, the conditional distribution of threshold (ψ_S) exceedances of S is estimated. The estimation of marginal distribution of STM can be achieved with parametric and non-parametric approaches. Here, we assume a generalised Pareto distribution and maximum-likelihood estimation.

Exposure is also characterised using hindcast data, as storm severity expressed as a fraction of STM at each location in the exposure region, therefore informing us about the maximum fractional influence of each of the hurricane events at each location in the region. The marginal distribution of exposure E_j at location r_j is defined empirically using the set $\{e_{ij}\}$, where

$$e_{ij} = \max_{t \in \mathscr{T}_i} \frac{h(r_j,t)}{s_i} \tag{1}$$

where $h(r_j,t)$ is the value of significant wave height at the location at time t within the time interval $\mathscr{T}_i$ of the storm, and s_i is the corresponding STM value. By combining the estimated distribution $F_{S|\psi_S}$ for STM, in terms of its density $f_{S|\psi_S}$, and the empirical distribution F_{E_j} for E_j at location j, we estimate the distribution F_{H_j} of storm severity H_j at location j in the exposure region corresponding to an arbitrary hurricane event, as follows. Since $H_j = E_j \times S$, we can write the cumulative distribution of H_j as

$$\begin{aligned} F_{H_j|\psi_S}(h) &= \mathbb{P}(H_j \leq h) \\ &= \int_s \mathbb{P}(E_j S \leq h | S=s) f_{S|\psi_S}(s) ds \\ &= \int_s \mathbb{P}(E_j \leq h/s) f_{S|\psi_S}(s) ds \\ &= \int_s F_{E_j}(h/s) f_{S|\psi_S}(s) ds \end{aligned} \tag{2}$$

where the final integral can be evaluated using numerical integration (e.g. [13]). In addition, the "N-year event" (or "N-year return value") $h_{j;N}$ is then defined as the fractile of the distribution of the annual maximum A_j with probability 1-1/N. That is

$$\begin{aligned} F_{A_j|\psi_S}(h_{j;N}) &= \exp\left[-\lambda\left(1 - F_{H_j|\psi_S}(h_{j;N})\right)\right] \\ &= 1 - \frac{1}{N} \end{aligned} \tag{3}$$

where λ is the expected number of threshold exceedances per annum.

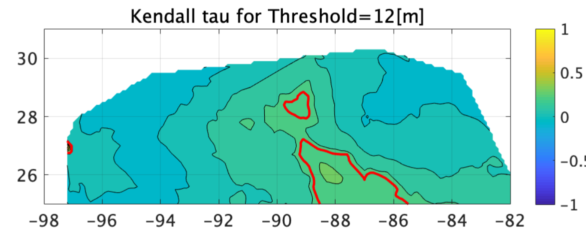

FIGURE 3. KENDALL'S TAU FOR RANK CORRELATION BETWEEN STM AND FULL-DOMAIN EXPOSURE FOR THRESHOLD 12m. THE AREA INDICATED BY RED INDICATES THE REGION WHERE THE TEST STATISTIC EXCEEDS ITS 95% CONFIDENCE INTERVAL UNDER THE NULL HYPOTHESIS.

Kendall's rank correlation

We assume here that the characteristics of E at a location are independent of STM, cyclone track, environmental covariates and time. These assumptions need to be justified. To achieve this we estimate Kendall's tau statistic for the rank correlation between E_j and S at each location. This rank correlation test is an easily implemented approach to check the independence of STM and exposure (see [9], Section 5.2). For each location, we consider the set of STM magnitudes and the set of exposure values (assuming that the exposure region corresponds to the full ocean basin), and estimate Kendall's tau τ statistic. If STM and exposure are independent, τ is approximately Gaussian-distributed with zero mean and variance $2(2n_S+5)/(9n_S(n_S-1))$. In our previous paper [14], the results of Kendall's tau test showed dependency in rank correlation when applying STM-E with extreme value threshold of 12m as depicted in Fig. 3; a positive correlation in the central region of the Gulf of Mexico was revealed. When a higher threshold of 13m is applied, the rank correlation does not reject the assumption of STM and exposure being independent, as depicted in Fig. 4. The figure also marks 18 locations considered in subsequent analysis using circular exposure regions of different radii.

Sensitivity to region selection

We next conducted a numerical experiment by changing the radius of an (assumed circular) exposure region, centred on the location of interest. Various radii are considered, ranging from zero (i.e. a single location approach) to 10° (i.e. a full ocean basin approach). The radius is defined in degrees longitude and latitude. The number of locations included in the exposure region for each radius is depicted in Fig 5-7. When a large enough radius is set, the analysis coincides with the full region analysis. This experiment was conducted for the 18 locations shown as the black circles in Fig 4. The numerical experiment was conducted taking all hurricane events into account.

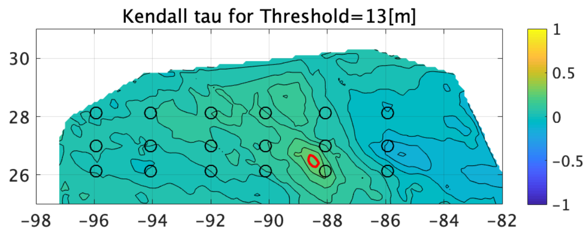

FIGURE 4. KENDALL'S TAU FOR RANK CORRELATION BETWEEN STM AND FULL-DOMAIN EXPOSURE FOR THRESHOLD 13m. THE AREA INDICATED BY RED INDICATES THE REGION WHERE THE TEST STATISTIC EXCEEDS ITS 95% CONFIDENCE INTERVAL UNDER THE NULL HYPOTHESIS. THE BLACK CIRCLES INDICATE LOCATIONS SELECTED FOR SUBSEQUENT ANALYSIS WITH CIRCULAR EXPOSURE REGIONS OF DIFFERENT RADII.

RESULTS AND DISCUSSION

Results for all 18 locations with exposure radius showed similar features. For all the figures, the x-axis shows the radius of exposure. Results for three arbitrary locations are summarised in Fig 5-7. In each figure, the top panel shows the number of locations included in the analysis as a function of exposure radius. The case of 0° radius coincides with the per-location approach, and the radius of 10° is equivalent to STM-E applied for the full region. The second panel shows the max, median and minimum exposure at the point of interest. The range of exposures widens with increasing radius. The third panel shows the estimated 100-year return value for the location and the STM. Red triangles indicate the 100-year return value of STM within the region, which increases with radius since larger values of STM are admitted for analysis as the region expands. Black crosses indicate the per-location 100-year return value (estimated using Equation(2)) incorporating both STM and exposure distributions. The bottom panel shows Kendall's rank correlation (between S and E_j) together with its 95% confidence interval under the null hypothesis that there is no rank correlation present. The statistic seems to deviate somewhat from zero as we increase the region size, but generally remains within its confidence interval. Location 3 was intentionally chosen to depict a case where the Kendall's tau fails for the full region analysis.

Selection of radius of exposure region

The analysis above reveals the trade-off between (a) increasing sample size for STM-E analysis and (b) introducing spatial dependence between STM and exposure, as a function of radius of exposure region. An ideal scenario would be to isolate the largest exposure region with does not violate the Kendall's tau test. For the Gulf of Mexico case, STM-E for radius of 5° shows a pragmatic balance between lack of dependence of STM and exposure for the selected points, and stability of 100-year return value estimate with respect to exposure radius. Nevertheless, computational effort increases with exposure radius. We are also interested in establishing the extent of differences in inferences for different choices of exposure radius, and whether it might be possible to perform STM-E analysis adequately with smaller exposure regions. We therefore next choose to compare 100-year return value estimates using four different exposure radii: i.e. per location analysis, STM-E with 2.5° radius, STM-E with 5° radius, and the full region STM-E analysis. The full region STM-E is applied for threshold of 13m, a condition that meets the assumption of independent STM and exposure. Here, the per location analysis is also revisited. In the per location approach for Fig 1, maximum likelihood approach based on high threshold (10m) is applied. The large variance occurs from the small sample size. In the per location approach in Fig 8, we use another approach, extracting 30 extreme samples to ensure enough sample size for maximum-likelihood estimation. The number 30 was derived by evaluating 100yr RP value stability against sample size for the selected 18 point. A review on choice of methodology for extreme value analysis is given in [15]. The result is depicted in Fig 8. Although results are sensitive to threshold choice, we observe the spatial distributions from STM-E become spatially smoother with increasing exposure radius. Note that the largest 100 year return values in the per-location approach are truncated. The figures suggests that STM-E with 5° radius is in good agreement with that obtained using the full US Gulf of Mexico as exposure region.

There are two main hyper-parameters in the application of STM-E, namely extreme value threshold choice and radius of exposure region (when the regions are assumed to be circular discs). Diagnostic tests can assist hyper-parameter selection to some extent, but in reality a variety of possible values for hyper-parameters may remain. In this case, a suitable practical approach would be to adopt an ensemble model including equal contributions from models corresponding to various sensible combinations of hyper-parameters.

Region size and extreme events

It appears rational that the shape and size of exposure region for STM-E analysis should be motivated by the spatial extents of phenomena driving extreme seas. Exposure region acts to select contributing data to STM and exposure, relative to the location of interest. The radius of maximum wind of hurricanes is around 80km, and the radius of outermost closed isobar is known to be several hundred kilometres [16]. The magnitude of hurricane scale therefore appears to coincide with the suggested scale for STM-E. The choice of exposure regions as circular discs is arbitrary, by also motivated by physical considerations; it is reasonable to assume that spatial dependence might be related to distance from the centre of the exposure region on average. Further, exposure region shapes reflecting hurricane track orientation, mainly from south to north, deserve attention.

CONCLUSION

The motivation for this work is to devise a simple statistical approach to return value estimation for significant wave height incorporating data from multiple dependent locations in a spatial neighbourhood, using straightforward methods of extreme value analysis. This is achieved using a local spatial version of the STM-E approach of [9], in which a circular exposure region of specified radius is adopted, such that modelling assumptions underlying STM-E are satisfied within the exposure region. A suitable exposure radius of 5°, i.e. around 500km, was estimated. For larger exposure regions, Kendall's rank correlation test suggests dependence between STM and exposure, invalidating the STM-E approach. As future work, the authors will undertake a thorough simulation study to demonstrate the performance of the local STM-E methodology, and to investigate the sensitivity of inferences to hyper-parameter choice.

ACKNOWLEDGMENT

We thank Oceanweather for use of GOMOS data, and Andrew Cox for useful comments on this work. One of the authors (R. Wada) would like to thank the Fundamental Research Developing Association for Shipbuilding and Offshore (REDAS) for financial support.

REFERENCES

[1] Heideman, J. C., and Mitchell, D. A., 2009. "Grid point pooling in extreme value analysis of hurricane hindcast data". *J. Waterway, Port, Coastal, Ocean Eng.,* ***135***, pp. 31–38.

[2] Vickery, P., Wadhera, D., and Stear, J., 2010. "A synthetic model for Gulf of Mexico hurricanes". *Offshore Technology Conference, Houston, OTC-20699*.

[3] Vickery, P. J., Skerlj, P. F., and Twisdale, L. A., 2000. "Simulation of hurricane risk in the U.S. using empirical track model". *J. Struct. Eng.,* ***126***, pp. 1222–1237.

[4] Darlymple, T., 1960. "Flood frequency analysis, us geological sur". *Water Supply Pap,* ***1543***.

[5] Hosking, J. R. M., and Wallis, J. R., 2005. *Regional frequency analysis: an approach based on L-moments*. Cambridge university press.

[6] Vanem, E., 2017. "A regional extreme value analysis of ocean waves in a changing climate". *Ocean Engineering,* ***144***, pp. 277–295.

[7] Davison, A. C., Padoan, S. A., and Ribatet, M., 2012. "Statistical modelling of spatial extremes". *Statist. Sci.,* ***27***, pp. 161–186.

[8] Ross, E., Kereszturi, M., van Nee, M., Randell, D., and Jonathan, P., 2017. "On the spatial dependence of extreme ocean storm seas". *Ocean Eng.,* ***145***, pp. 1–14.

[9] Wada, R., Waseda, T., and Jonathan, P., 2018. "A simple spatial model for extreme tropical cyclone seas". *Ocean Engineering,* ***169***, pp. 315–325.

[10] Jonathan, P., and Ewans, K. C., 2013. "Statistical modelling of extreme ocean environments with implications for marine design : a review". *Ocean Eng.,* ***62***, pp. 91–109.

[11] Wada, R., and Waseda, T., 2019. "Assessment of data-inherited uncertainty in extreme wave analysis". *Journal of Offshore Mechanics and Arctic Engineering*, pp. 1–13.

[12] Wada, R., and Waseda, T., 2018. "Benchmark for sources of uncertainty in extreme wave analysis". *Proc. 37rd Conf. Offshore Mech. Arct. Eng. (Accepted 2wa018)*.

[13] Ross, E., Randell, D., Ewans, K., Feld, G., and Jonathan, P., 2017. "Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference". *Ocean Eng.,* ***142***, pp. 315–328.

[14] Wada, R., Jonathan, P., Waseda, T., and Fan, S., 2019. "Estimating extreme waves in the gulf of mexico using a simple spatial extremes model". In ASME 2019 38th International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers Digital Collection.

[15] de Zea Bermudez, P., and Kotz, S., 2010. "Parameter estimation of the generalized pareto distribution—part i". *Journal of Statistical Planning and Inference,* ***140***(6), pp. 1353–1373.

[16] Liu, K., and Chan, J. C., 1999. "Size of tropical cyclones as inferred from ers-1 and ers-2 data". *Monthly Weather Review,* ***127***(12), pp. 2992–3001.

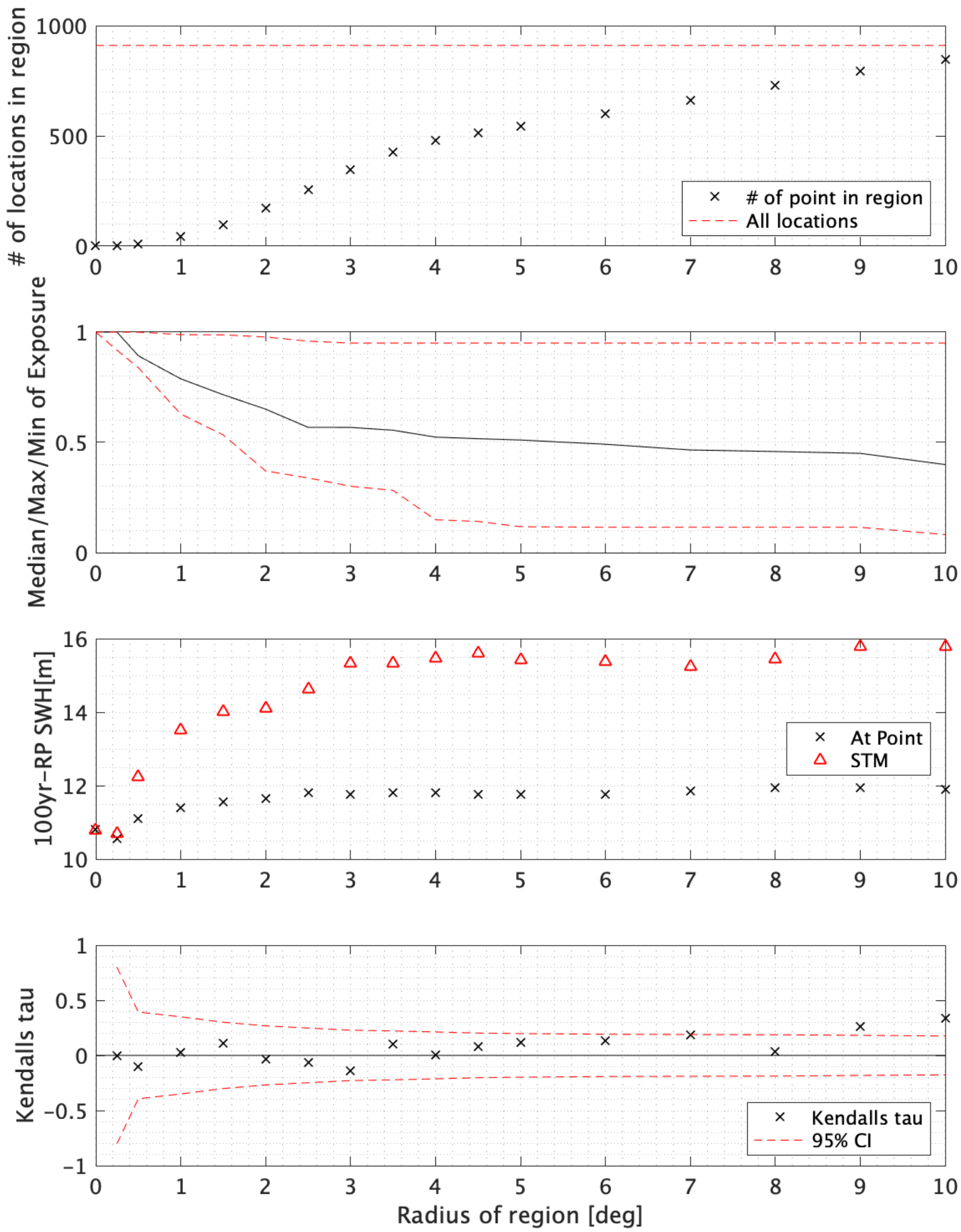

FIGURE 5. ARBITRARY LOCATION 1 : SENSITIVITY ANALYSIS FOR DIFFERENT EXPOSURE RADIUS. TOP: POINTS INCLUDED, 2ND: MAX, MEDIAN, MIN OF EXPOSURE, 3RD: 100-YEAR RETURN VALUE FOR STM AND POINT OF INTEREST, BOTTOM: KENDALL'S TAU STATISTICS WITH 95% CONFIDENCE INTERVAL

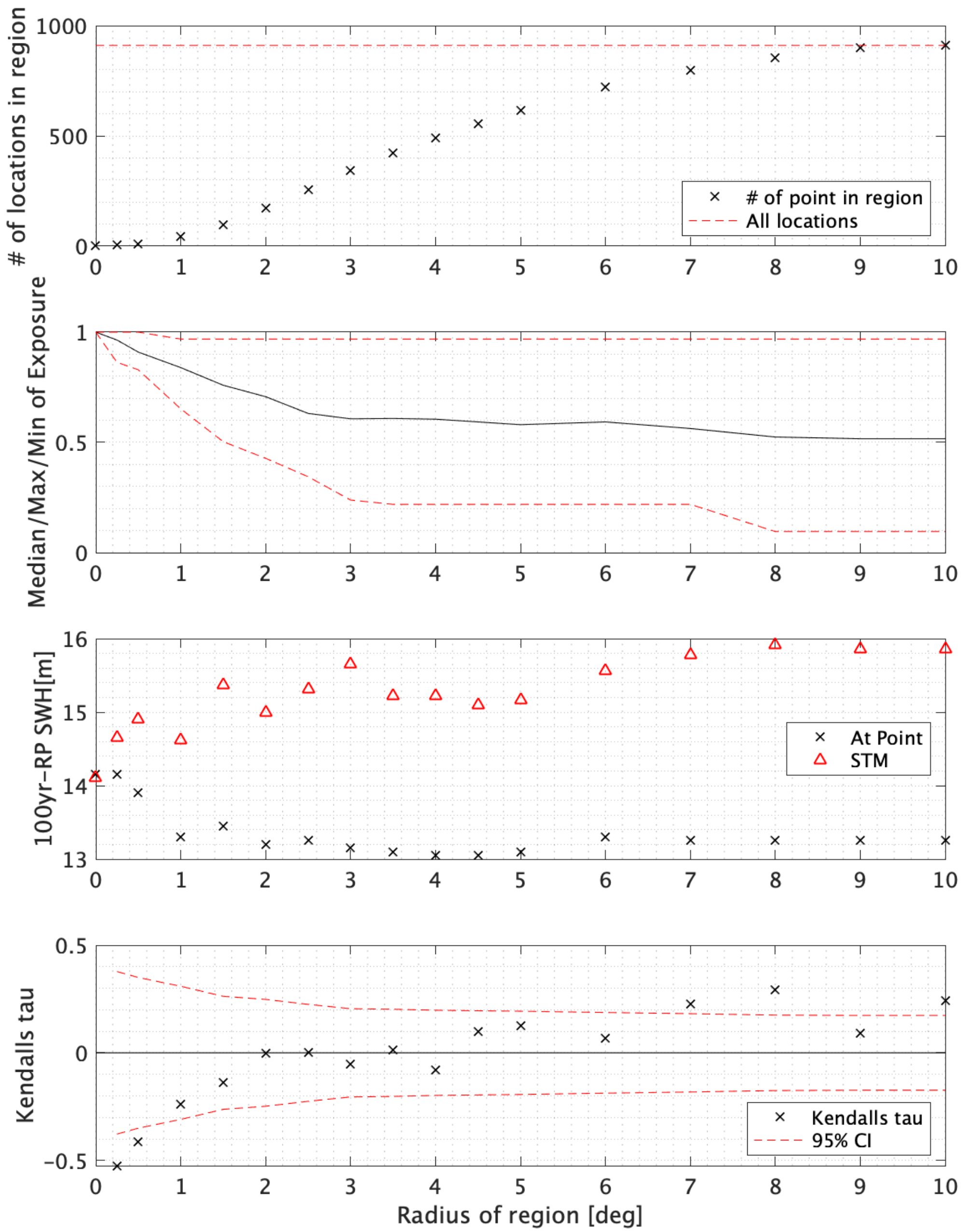

FIGURE 6. ARBITRARY LOCATION 2 : SENSITIVITY ANALYSIS FOR DIFFERENT EXPOSURE RADIUS. TOP: POINTS INCLUDED, 2ND: MAX, MEDIAN, MIN OF EXPOSURE, 3RD: 100-YEAR RETURN VALUE FOR STM AND POINT OF INTEREST, BOTTOM: KENDALL'S TAU STATISTICS WITH 95% CONFIDENCE INTERVAL

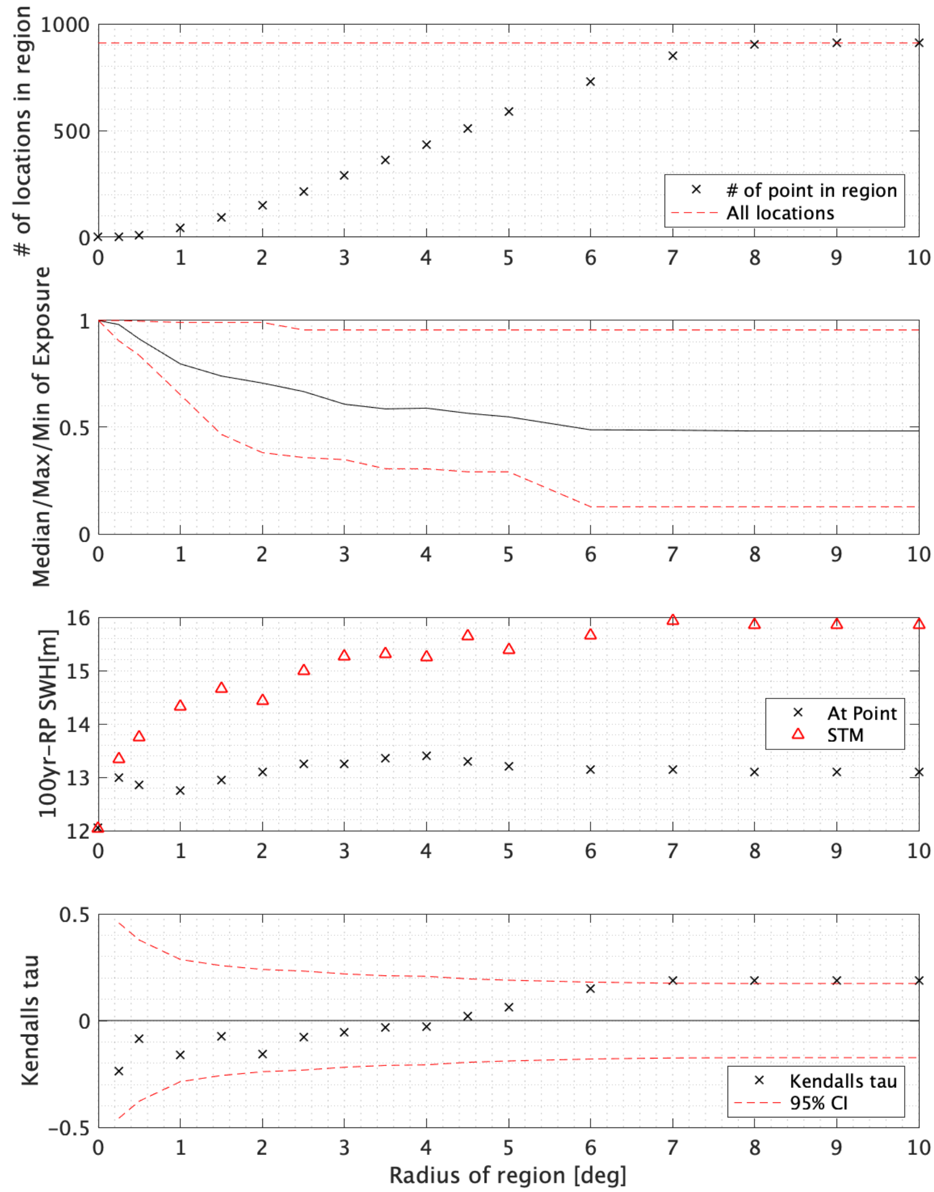

FIGURE 7. ARBITRARY LOCATION 3 : SENSITIVITY ANALYSIS FOR DIFFERENT EXPOSURE RADIUS. TOP: POINTS INCLUDED, 2ND: MAX, MEDIAN, MIN OF EXPOSURE, 3RD: 100-YEAR RETURN VALUE FOR STM AND POINT OF INTEREST, BOTTOM: KENDALL'S TAU STATISTICS WITH 95% CONFIDENCE INTERVAL

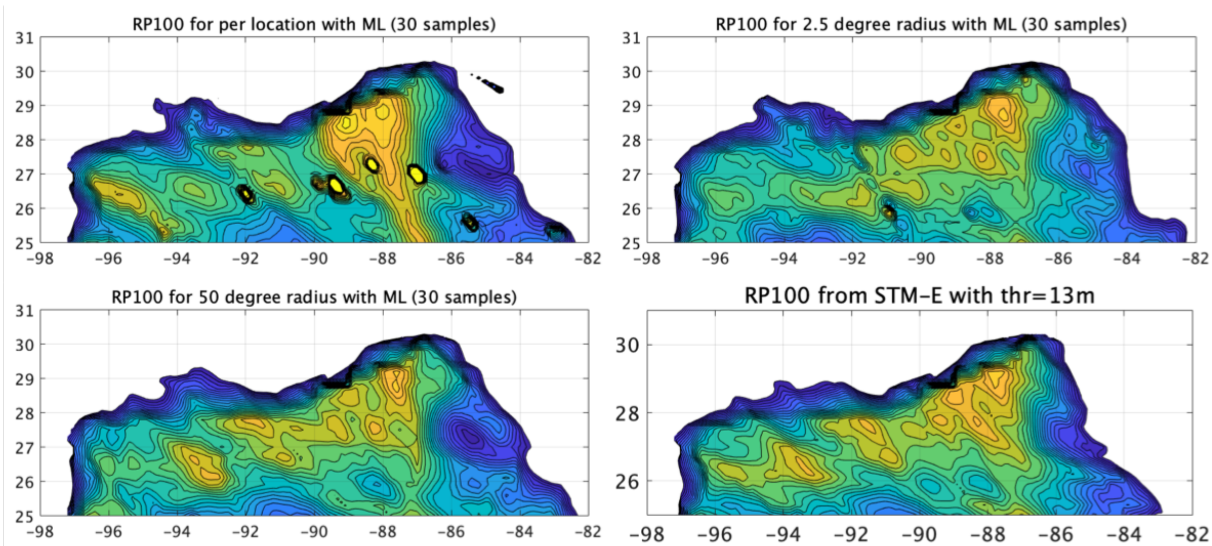

FIGURE 8. 100-YEAR RETURN VALUE FROM PER-LOCATION APPROACH, STM-E WITH 2.5° RADIUS, STM-E WITH 5° RADIUS, AND STM-E WITH FULL DOMAIN AS EXPOSURE REGION (WITH COMMON COLOR SCALE)