Response to Reviewer #2 of our manuscript entitled **Coastal extreme sea levels in the Caribbean Sea induced by tropical cyclones** submitted to *Natural Hazards and Earth System Sciences.*

Ariadna Martín, Angel Amores, Alejandro Orfila, Tim Toomey, Marta Marcos

October 7, 2022

In their manuscript "Coastal extreme sea levels in the Caribbean Sea induced by tropical cyclones," the authors Martín et al. use a new database of synthetic tropical cyclones as forcing to simulations of wind waves and storm surges in the Caribbean Sea. It is shown that the wind waves and storm surges vary significantly at coasts around the basin, due to differences in storm evolution, local bathymetry, and other characteristics. The manuscript is generally well-written.

It is my recommendation that this manuscript can be **rejected** for publication.

To this reviewer, there are two critical flaws in this study. First, the validation is inconsistent with the rest of the study and is inconclusive about the quality of the model predictions. The two validation storms are represented by data-assimilated atmospheric products, whereas the synthetic storms are constructed from a simple parametric model. And the two validation storms are (apparently) not described by sufficient observations of wind waves and coastal sea levels, so it is not possible to validate the model predictions for these storms. Without a validation, it is not possible to trust the later predictions for the synthetic storms – this is recognized by the authors, who note an insufficient nearshore resolution and possible boundary effects as reasons for poor predictions of wind waves and storm surge, respectively. The validation should be expanded to investigate these and other potential problems, either by validating against all available observations for these storms, or by selecting other storms with useful observations.

Second, the findings are not necessarily novel, and it is not clear what is the contribution to our scientific understanding. Previous studies have investigated storm-induced hazards in the Caribbean Sea and elsewhere, and they have characterized the wind waves and storm surge that is possible along their coastlines, as well as the relative roles of atmospheric pressures, winds, and waves as drivers. What are the gaps in those previous studies, not in number of storms considered, but in understanding of coastal processes? How will this study help to fill those gaps? As-is, the study is impressive in the amount of computations that have been performed, but it is lacking in connecting those computations to a novel contribution. Because these two flaws will require a substantial refocusing of the manuscript, likely with additional computations, this reviewer recommends the present manuscript to be rejected for publication.

> A — We thank the reviewer for the time taken to evaluate our work and for the assessment provided. We respond to all concerns below in detail. In particular, we have extended our validation by 1) including consistent comparisons using both outputs from the atmospheric reanalysis and the parametric model, 2) adding more TCs and 3) accounting for coastal sea levels. We have also highlighted the strong points in our work.

Main comments:

(1) — Lines 25–37: It is not clear (at least to this reviewer) what will be learned by reading this manuscript. It is stated that "we focus on the ocean hazards generated by TCs in terms of wind-waves and storm surges." Have these hazards not been characterized previously either basin-wide or at specific locations in the Caribbean Sea? Do we expect this 1000- storm study to provide new insights into the magnitudes of waves and surges in this region? If so, then why?

It is also stated that "[w]e analyse in detail the outputs of the numerical simulations to quantify the role of the different forcing factors." Have these roles not been quantified previously? This

reviewer is aware of several studies that quantified the relative roles of atmospheric pressures, winds, and waves on the magnitude of storm surge. Do we expect this 1000-storm study to provide new insights? If so, then why?

> A — We have modified this paragraph in the introduction to highlight the objectives and the main novelties of this study. First, we perform a basin-scale study, in contrast to earlier assessments that are limited to local cases. Second, we tackle both coastal sea levels and wind-waves, using a coupled model, which is also a new perspective. Third, we investigate the main drivers of coastal hazards, by separating the different sources (pressure, winds and waves). Finally, we exploit a new, large and comprehensive synthetic dataset over the Caribbean. We have referenced the current literature of which we are aware, but we will be keen to cite any other work that the reviewer provides in relation to the topic and that might be missing.)

(2) — Lines 49–63: Can the authors provide more justification for how the storms were selected? This reviewer is interested in two aspects. First, why use the maximum wind speed (presumably, see minor comment below) as a proxy for the storm intensity? Why not use the minimum central pressure, or maximum radius, or an integrated quantity like the total kinetic energy? The STORM database includes several parameters for each storm, not just the maximum wind speed, but it seems like the current methodology is ignoring them.

Second, the random selection of subsamples seems suboptimal. Why not use a maximum dissimilarity algorithm to identify the top 1000 storms that best span the parameter space? Or surely there are other methods that could be considered? The authors' method appears to work okay, given the convergent errors shown in Figure 1, but it would be nice to see a brief discussion of why this method was selected over other options.

> A — To select the subset of TCs from the entire dataset we used maximum wind speed because this is the variable that is better represented in STORM. It shows the highest correlation between the North Atlantic basin in the STORM dataset compared to the IBTrACS dataset [Bloemendaal et al., 2020].
>
> Regarding the methodology, there are certainly other algorithms that serve the same purpose. Our method is simply one of them and we believe is does work properly. We demonstrate so when we compare the distributions of the total dataset and the subsets. We determine objectively the required size of the subsample, and we show that the distributions of the entire dataset and the subsample are consistent temporally and spatially. To put it into context, we have now cited other methods, too.

(3) — Lines 79–81 and 170–173: Not sure what this means. There were 46 sea-level records in the region – they must have observed something useful about the water levels. The authors refer to a lack of "footprint" – does this mean that the observation records do not include the effects of the storms? The lack of water-level validation is a critical flaw in this study. Somehow the water-level predictions need to be validated, either for these or other storms. Without a validation of the water levels, the rest of the results cannot be trusted.

> A — We agree. This was a weak point also raised by other reviewer, who even provided an additional reference, so we have extended our validation to include measurements from both waves and coastal tide gauges. Please, refer to the new section 3.2.

(4) — Lines 92–94: References are needed for these methods. Why use a distance of 300 km? Why use a reduction of 20 percent?

> A — We have included the corresponding references. With respect to the 300 km distance, this is justified because the transfer of energy between the atmosphere and the ocean at such a long distance from the TC eye is negligible, according to Holland [1980] and Holland et al. [2010]. As for the 20 % parameter, we have included a reference in the paper [Willoughby and Black, 1996]. We would like to note here that this reduction will not alter the results as we do not account for winds over land.

(5) — Sections 2.1 and 2.2: There is a mismatch between the atmospheric forcing used for the validation and the rest of the study. The validation storms use a data-assimilated product, which should be accurate (although this reviewer is not convinced that ERA5 can resolve the full dynamics of a tropical cyclone), whereas the synthetic storms are developed with a parametric model. The validation storms have different resolution (1 hr, 30 km) than the synthetic storms (3 hr, 5 km). Can the authors comment on how these differences may affect the validation?

More importantly, why not generate the atmospheric forcing for Wilma and Tomas in the same way as the synthetic storms? The authors could use the historical track information for these two storms, push it through the parametric model, and then be able to compare apples-to-apples.

A — We agree with the reviewer. We have therefore performed tests and we have changed all the format of validation section in order to add synthetic fields and also more events (related to next question).

(6) — Figure 4: Why use only two buoys per storm? Why not do a comprehensive validation by using all available observations? As-is, the reader can assume that these buoys were cherry-picked to show the best results.

A — It was never our intention to show biased results. The validation of waves with buoys is generally very good, as the winds far from the TC eye are well reproduced. In the new validation section 3.2 we have now included all available data for buoys and tide gauges.

(7) — Lines 184–185: "but the generated waves are less intense due to the smaller fetch area." The Caribbean Sea is a large basin, with a minimum width of 600 km at its narrowest. Why would any waves be fetch-limited in this basin?

A — We meant smaller fetch in comparison to the Atlantic counterpart, where TC have much more space to develop. This has been clarified in the text.

(8) — Lines 204–205 and 294–295: This claim should be explored, ideally via a more-comprehensive validation with the full set of available observations. But more importantly, why do the authors think the wave set-up is under-estimated? Should it be more than 5 percent of the total contribution?

A — For wave setup effects to be well modelled in the nearshore, a high spatial resolution is required. The separation between nearby grid points needs to be able to represent the nearshore slope and the wave breaking. This is not the case with a coastal resolution of $\sim$ 2 km. Therefore, it is not possible to validate wave setup when it is not well reproduced by the model. We are aware that it is underestimated in most cases. The work by Amores et al. [2020] demonstrated that, under conditions of strong sustained storms, wave setup may contribute up to 40-50% to total coastal water levels. We believe that, in regions with wide continental shelves and under strong TC, at least part of the wave setup is captured by the model.

(9) — Lines 305–306: The repository should include more than just the return periods. The selected 1000 synthetic storms should be included, both in their parameters from the STORM dataset and their pressure/wind fields from the parametric model. The SCHISM and WWM-III input files should also be included. This will allow for reproducibility of the study results.

A — We initially decided to limit the product to what we consider the most useful (in terms of impact analysis) product of the work. But we agree with the reviewer that reproducibility is very important, so we are now sharing all the outputs in the same data repository 10.5281/zenodo.7069110

(10) — Section 4: An overarching comment is that it is not clear (at least to this reviewer) what is novel about the study findings. It should be expected that the windward islands are affected by Atlantic storms, whereas the west side of the Caribbean is affected by storms from that basin. It should also be expected that regions with narrow shelfs and deep offshore bathymetry will have smaller storm surges that are forced mainly by the storm's pressure deficit, whereas regions with wide shelfs and shallow offshore bathymetry will have larger storm surges that are forced mainly by the storm's winds. Can the authors do more to contextualize their findings and motivate their novelty?

A — Please, see our response above to the main comment (1). We have included some additional text in the introduction to state the major novelties and strengths of the study. We are well aware and agree with the generalities described by the reviewer, but our work quantifies the coastal hazards from storm surges and wind-waves along all the coastlines in the Caribbean induced by TCs and this is a result that, to our knowledge, has not been published before.

Other specific comments:

(1) — Line 57: "trough" should be 'through' for correct spelling.

A — Done

(2) — Figure 1: It would be helpful to describe what is meant by "speed." This reviewer assumes it is the maximum wind speed at any location/time during the storm. But it could be something else, e.g. the forward speed of the storm.

A — The reviewer is correct, is Maximum wind speed and we have changed this in the Figure.

(3) — Line 65: Again, it is assumed that these speeds (e.g. 111 km/h) refer to the maximum wind speeds, but this should be clarified in the text.

A — Done

(4) — Figure 2: This reviewer struggled to see the tracks and labels in this figure, as they were depicted in black on a mostly blue background. Not sure what to suggest to make these features to be more legible. What if the track and labels were in white?

A — This figure has undergone some modifications due to changes in the validation

(5) — Line 78: Should be 'data were' for subject-verb agreement.

A — Changed

(6) — Line 97: "last" should likely be 'latest.' Please give the actual version numbers for SCHISM and WWM-III.

A — Done.

(7) — Line 122: "where" should be 'were' for correct spelling.

A — Done

(8) — Line 275: "are" should be 'area' for correct spelling.

A — Done

(9) — Line 280: When the letter 'm' is shown in italic font, this reviewer assumes it is a variable, e.g. 25 times m. If it is meant to be a unit (meters), then it should not be in italic font

A — Changed

References

Nadia Bloemendaal, Ivan D Haigh, Hans de Moel, Sanne Muis, Reindert J Haarsma, and Jeroen CJH Aerts. Generation of a global synthetic tropical cyclone hazard dataset using storm. *Scientific data*, 7(1):1–12, 2020. doi: 10.1038/s41597-020-0381-2.

Greg J Holland. An analytic model of the wind and pressure profiles in hurricanes. *Monthly weather review*, 108(8):1212–1218, 1980.

Greg J Holland, James I Belanger, and Angela Fritz. A revised model for radial profiles of hurricane winds. *Monthly weather review*, 138(12):4393–4401, 2010. doi: https://doi.org/10.1175/2010MWR3317.1.

HE Willoughby and PG Black. Hurricane andrew in florida: Dynamics of a disaster. *Bulletin of the American Meteorological Society*, 77(3):543–550, 1996. doi: https://doi.org/10.1175/1520-0477(1996)077⟨0543:HAIFDO⟩2.0.CO;2.

A. Amores, M. Marcos, D. S. Carrió, and L. Gómez-Pujol. Coastal impacts of storm gloria (january 2020) over the north-western mediterranean. *Natural Hazards and Earth System Sciences*, 20(7):1955–1968, 2020. doi: 10.5194/nhess-20-1955-2020.