REVIEW, "COASTAL EXTREME SEA LEVELS ..."

MARTÍN, AMORES, ORFILA, TOOMEY, MARCOS

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.

The following major comments can be considered in a revised manuscript:

- 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?

The Introduction would benefit from clear questions and hypotheses to motivate the study.

– 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.

– 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.

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

– 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.

- 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.
- 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?
- 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?
- 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.
- 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?

The following minor comments can also be considered:

- Line 57: "trough" should be 'through' for correct spelling.
- 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.
- 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.
- 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?
- Line 78: Should be 'data were' for subject-verb agreement.
- Line 97: "last" should likely be 'latest.' Please give the actual version numbers for SCHISM and WWM-III.
- Line 122: "where" should be 'were' for correct spelling.

– Line 275: “are” should be ‘area’ for correct spelling.

– 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.