The work presents a method and a database for obtaining high-resolution boundary conditions for flood risk assessment worldwide. The method is an extension of a previous regional study (Sweet et al., 2020), with certain changes and innovations in some aspects. The method's steps and its application are well-structured and clearly described in the article. The results are analyzed and validated from different perspectives, combining different databases. The global-scale application is very interesting and useful for the scientific community. Despite the well-structured and well-written nature of the work, and the thorough analysis and verification of the results, certain points of the methodology and results need to be clarified and discussed.

Abstract

The second paragraph is focused on a general description of RFA method illustrated through the example of Cyclone Yasi, while it should be personalized to your study. I would describe the method based on your application globally and provide details about the results obtained. This expansion will emphasize the method's capabilities and the valuable resource for coastal risk assessments.

INTRODUCTION

I notice the absence of the reference to Calafat et al. () in the text, even though it is included in the references section but not cited in the body of the work. This reference is particularly relevant, representing one of the most significant contributions in recent years where the RFA method is applied.

Lines 125-128: “The principle of an RFA is founded on the basis that homogenous region can be identified, throughout which similar meteorological forcings and resultant storm surge or wave events could occur, even if the extreme events have not been seen in part of that region in the historical record (Hosking and Wallis, 1997).”

The key to the method lies in the similarity of meteorological forcings. In your global-scale application, you divide the entire globe into grid cells of 1 degree and then apply a 400-radius at the grid centroid. Could this definition of homogeneous regions significantly impact your results? Have you conducted a sensitivity analysis concerning meteorological forcings? This question is also related to the one about the Heterogeneity test.

In the grid cell example shown in Figure 2A, could it be possible to consider that extreme water levels along the east coast of Florida and the west coast of Florida could be generated by common storm track? Perhaps not the extra-tropical, but in the case the tropical cyclones? The way homogeneous regions are defined could potentially explain some of your “undesired results” (lines 556 – 560; lines 619-622: decreases in ESL exceedance probabilities compared to the single site analysis).

DATA

FES2014: higher resolution and performed better than those of GTSM (lines 262-262). Have you checked the differences between the tides between GTSM-ERA5 and FES2014? In the case of GTSM-ERA5, the non-linear interactions between tides and

surges are included because they have been simulated together. How might these effects influence sea levels (tide + surge)? What implications arise from the overestimation of the linear sum of these dynamics (Arns et al., 2020)?

You are using a global wave reanalysis at a resolution of 0.5. Have you considered the potential impact of this coarse resolution, particularly along the coast where the propagation process that modifies waves approaching the coast is not accounted for? How might this simplification influence your results? Have you contemplated applying a simplified method to address this limitation?

Regarding the definition of COAST-RP dataset, It is not very clear for me the sentence "in extra-tropical regions, a 38-year timeseries of ERA5 is used" (lines 226-227).

METHODS

I think Figure 3 helps to understand the methodology, but I would indicate and structure it based on the 5 key steps that you enumerate in lines 240-249. I would use the same structure for the following subsections: 3.1 Data processing, they might correspond to i) and ii), and 3.2 RFA (which corresponds to the rest of steps). This last subsection, I would divide it regarding the steps and scheme in Figure 1 (for example, the downscaling and the bias correction could be better in separated subsections)

I believe Figure 3 contributes to a better understanding of the methodology. However, I suggest indicating and structuring it based on the five key steps enumerated in lines 240-249. Similarly, I propose adopting the same structure for the subsequent subsections. For example, 3.1 Data Processing, which may correspond to i) and ii), and 3.2 RFA, aligning with the remaining steps, could be further divided. For example, 3.2 subsection could be divided based on the steps and the scheme presented in Figure 1. For instance, downscaling and bias correction could be more effectively presented in separate subsections for clarity and coherence.

Lines 270-271: "wave setup is interpolated to the nearest record location using a nearest-neighbour approach". I think a nearest-neighbour approach cannot be considered as an interpolation method, you are directly assigning the closest node

Lines 271-272: "to account for the lack of wave setup in sheltered areas…"
I would phrase it the other way around; it is the type of coastline that determines whether the wave setup is considered or not.

Lines 278-279: Could an overestimation of water levels result from adding the daily maximum wave setup to the daily highest water level? Have you checked if the "total water level," calculated as the sum of storm surge, astronomical tide, and wave setup at an hourly scale, and then selecting the extreme water levels, exhibits peaks that are very similar (in the same order) to the sum of the maximum daily values of surge-tide and maximum daily significant wave height?

Lines 282-284: Tide gauges are assumed to be located in sheltered regions such as bays and estuaries, and consequently, wave setup is not considered necessary. Why is it necessary to fit a copula between daily peak water levels and daily maximum significant wave heights when tide gauge records fall outside the temporal range of the ERA5 data for providing predictions of daily maximum significant wave heights (lines 279 – 281)? If this fitting is required, which database are you utilizing for surge + tide and significant wave heights? Is it GTSM-ERA5, FES2014, and ERA5?

Lines 306-307: why the final number of tide gauges are 836 while at the beginning the total number were 2223 (line 258)? I am not clear on how the discretization of regions leads to this reduction. (same comment as the other reviewer)

Figure 2. I suggest to plot the tide gauges selected (Figure 2A) in the same color as the time series in Figure B.

Lines 342-343: the index flood u defined as the 98^{th} percentile for all locations – same comment as the other reviewer

How does the Heterogeneity test operate in terms of assessing the homogeneity of a region? What characteristics are considered to deem time series within a region as homogeneous? If the test fails due to an anomalous record (lines 354), at what point could the anomalous record be considered an extreme event resulting from a tropical cyclone?

Lines 366:367: How could the empirical threshold of 0.35 for the shape parameter impact the results of high return periods? Are you aware of typical values for the shape parameter in the case of tropical cyclones (TCs)? While you mention "expert judgment" (line 368), could you provide a reference supporting this approach? As you mention in the results, section 4.3 (lines 555-557), some gauges show decreases in the return levels, and this could be driven by the shape parameter that maybe limit to much large water levels?

Lines 394-396: Regarding the last stage, which involves removing the bias in the high-frequency portion of the exceedance probability curves, the bias is quantified based on the divergence in the 1-in-1-year return period at each tide gauge/GTSM-ERA5 location. With this approach, a constant bias correction is applied at each location; however, the bias can be higher for higher return periods, as shown in Figures 6 and 8. How might this simplification impact the results?

Additionally, why do you use the 99th percentile of tidal elevations to interpolate the bias across coastal locations? Why not use tidal range, as in the interpolation of index flood u, or the 98th percentile, as the peak over threshold to select the extreme events?
Lines 421-424: description of COASTAL-RP dataset that should be included in section 2.

Lines 436-438: not completely clear, maybe 1000 grid cells which have between 3 and 10 GTSM-ERA5 record locations?

RESULTS

4.1 Global application of RFA

I wonder if in areas with tropical cyclones, a Generalized Extreme Value (GEV) distribution could be the best curve to represent the behavior of extreme water levels. What happens if there are two families of extremes due to extratropical storms and tropical storms with significantly different magnitudes? Perhaps considering the use of a mixed extreme value distribution, as suggested by O'Grady et al. (2022), might be beneficial.

4.4 Tropical Cyclone Yasi

I don't believe the case study of Cyclone Yasi illustrates how the RFA methodology enhances the representation of rare extreme events in the ESL exceedance probabilities. I think this section aims to demonstrate the regionalization process (as mentioned in lines 538-539). In some ways, these results suggest that the RFA method behaves like an interpolation of the shape parameter?

Line 495: Figure 5B.

Lines 503-507. It's not clear to me what is intended to be communicated with these sentences.

4.3 Comparison with GESLA

Areas where there is an improvement using RFA, should be also reflected in a better agreement with COAST-RP (section 4.4)?

4.4 Quantifying the improvements made by the RFA when compared to single site analysis

Why not plot the difference with COAST (GTSM with STORM) instead of GTSM-ERA? This would provide an overview of the magnitude of the differences and how your method approaches the COAST-RP results, not just in terms of the spatial pattern. On the other hand, the comparison with COAST-RP should only informs about the improvement in those areas prone to tropical cyclones.