Review comments to the revised version of “Variability of distributions of wave set-up heights along a shoreline with complicated geometry” (os-2019-25) by Tarmo Soomere, Katri Pindsoo, and Maris Eelsalu

I reviewed the original version of this manuscript, and concluded that it was worth publishing the results after major revisions of the manuscript. I have now read the authors’ response to my comments and the revised manuscript. While the authors have cleared up several issues raised by the reviewers, I still have a few reasons to feel less than confident with the results and conclusions to be able to recommend that this version of the manuscript is ready for publication. As some of my comments might change the results, I have to label the possible revisions as major. This is not to say that progress hasn’t been made with respect to the original version. Please find my comments below:

Main comment #1 (to the original manuscript): The authors have fully cleared up why the results are now different than in the previous submission to Earth System Dynamics. This matter has been fully addressed, and I leave it up to the editor how this should be reflected (if at all) in the final published manuscript.

Main comment #2 (to the original manuscript): My second main concern was the accuracy of the data. I recommended that the model would be run for a shorter time with a “full set-up” to validate the results. The authors chose to use measurements to validate their data, and while they are only available at one point, I think that this is a valid alternative to making additional model runs with a wind forcing from an atmospheric model. I accept this general approach taken by the authors to address my concerns. I also agree with the authors that a bias of 0.05 m is about as good as it gets. The revised manuscript states that the Tallinnamadal wave data are buoy data (although I have been under the impression that it is based on a pressure sensor). Is there a reference for these data?

So while I applaud the authors for using measurements to validate their model, I’m left feeling a bit uneasy about the results of the validation. The authors claim that “The appearance of the relevant empirical probability distributions of the occurrence of different wave heights is similar for both data sets (Fig. 3)”. Still, I can’t really see those two distributions being similar. The most important distinction is that a fit to measured data over 1 m would be reasonable well represented with a straight line (i.e. exponential fit), while a fit to the modelled data would be concave upwards. Doing the analysis on the model data would probably lead to a positive quadratic exponential, while a similar analysis on the measured data would result in a quadratic exponential close to zero. I understand that this is wave height and not wave set-up, but the former is a “forcing” for the latter. How can we trust the results of a quadratic exponential in the modelled wave set-up, if the shape of the distribution is not estimated correctly in the incoming wave height data?

Main comment #3 (to the original manuscript): This comment was concerning the fitting procedures and how they were presented. First, the authors have expanded on the description of the function they are fitting and what they are doing. This is now perfectly clear.

I also now understand why the authors want to use the gaps as a cut of the fitting: you want to avoid fitting the distribution to the “tail” of the data, which is not that statistically stable. I, however, still disagree that the gaps would have some kind of relevant meaning here. I would suggest that the upper limit of the fit would be a certain multiple of the mean wave set-up, or a certain amount of standard deviations above the mean. This would be a more objective and robust measure for the upper point than the gaps, which depend on the resolution of the binning, and also have a larger statistical variability (especially if this analysis should ever be reproduced using measurement data).

I also still disagree with the use of probability distributions instead of cumulative distribution. The authors said that the cumulative distribution can smooth out differences. Another way to see it is that using the probability distribution exaggerates differences. Reviewer #1 also suggested using cumulative distributions, and while he mentioned extreme value analysis, I want to point out that the fitting of cumulative distributions are in no way restricted to performing extreme value analysis (which I know that you are specifically not doing). Still, since I haven't seen this type of fitting before, I can't say how big of a deal this really is. I would strongly suggest fitting to the cumulative distribution (as this is standard practice), but if you are set on fitting to the probability distribution, then, ultimately, I won't stop you.

I also disagree with the authors claim that the highest values would only be a part of some "extreme value distribution" and therefore outside the basic distribution (page 13, lines 16-17). This is not the case. When calculating e.g. block maxima the points in the resulting extreme value distribution are still members of the original underlying distribution. For example, sufficiently long block maxima of an exponentially distributed variable should follow a Gumbel distribution, but all of the block maxima were still points (but vary far up the tail) in the original exponential distribution. If you want to exclude outliers, don't use the previously mentioned claim to do it, since it is fundamentally incorrect.

In summary, I do believe that the basic idea adopted by the authors is correct, namely that the last (highest) points are infamously unstable, and excluding them from the fit might be warranted. My disagreement is with the motivation (citing extreme value distributions) and the fitting technique (probability distributions and use of gaps) adopted to implement this idea.

Main comment #4 (to the revised manuscript): This comment is based on information that was not available in the original manuscript. You use two simplifications in calculating the refraction and shoaling. Assumption of shallow water and Snell's law. You mention that the water depth can be between 4 and 27 m. This would mean a (deep water) wavelength of about 260 m and 1700 m meter to satisfy the shallow water condition (wave period up to 33 s!). This is not reasonable, especially in the Gulf of Finland. Calculating the true phase and group speed using iteration from the full dispersion relation is trivial, and I see no reason not to use it.

The second problem is Snell's law, since it assumes that the isolines in the bathymetry are straight and parallel. This might often be a good approximation, but looking at Fig. 10, this is a questionable assumption in Tallinn Bay. The proper refraction can be calculated using the full equations (as done in e.g. WAM). As one of your contributions in this study is to investigate the complex geometry and it's effect on the wave set-up, I find it surprising that you have neglected effects that seems to be important in this geographical area.

Specific comment #1: To really see the effect of the fitting range, please continue the blue line as (for example) a dashed blue line outside the fitting range to illustrate how well it captures the data that were not used for fitting.

Specific comment #2: On page, lines you write: "In other words, in these locations the leading term a of the quadratic polynomial az+ bz+ c is positive at a 95% significance level."

This isn't really true, since you didn't do the analysis for only one point. If you do the analysis for all points then, on average, around 5% of the points should give a "false positive" when using a 95% significance (if the null hypothesis was true). For you 25% >> 5%, so the effect is real, but the confidence in a single point is no longer positive at a 95% confidence level, and treating it as such seems to be some unintentional version of p-hacking. (I understand that there was no malicious intent here even though I used the phrase p-hacking. It was used only to make the point clear. This comment was of a pure technical nature.)

Specific comment #3: Page 14, lines 16-17 “For large values of z this function behaves similarly to the probability density function of a Gaussian distribution.”

Did you mean for small values of z?

Specific comment #4: Figure 8. I find it hard to believe that it wouldn’t be possible to get a better Weibull fit than that presented in the figure, since the exponential distribution is contained in the Weibull family. Have you calculated the empirical cumulative distribution and plotted that in Weibull coordinates? You are claiming that Weibull is insufficient, so we need to make sure that this is right. As some other reviewer commented, it takes a lot to discard a three parameter Weibull.

Specific comment #5: As a suggestion, perhaps amend the “convex upwards” and “concave upwards” with (light tailed) and (heavy tailed) remarks. This might be especially useful the first time they are mentioned in the text, and perhaps in the discussion and conclusions, which some might read without going through the entire text and all the figures.

So e.g. on page 15, lines 26-27 the text would then read “The appearance of the distribution of modelled wave heights in the offshore (Fig. 3) is convex upwards (thin tailed) in the range of relatively frequent wave heights of 0.5–1.7 m.”