Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2021-55-RC1, 2021

Comment on nhess-2021-55

Anonymous Referee #1

Referee comment on "Correlation of wind waves and sea level variations on the coast of the seasonally ice-covered Gulf of Finland" by Milla M. Johansson et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-55-RC1, 2021

This manuscript uses observations of sea level, wind and waves to infer their relationships in the Gulf of Finland. The authors compute rank correlations and bivariate distributions (with and without dependence) between sea level and significant wave heights.

I have some concerns regarding the methods applied and the interpretation of the results, that I outline below, with some suggestions for improvement. Some of these suggestions may require big changes in the manuscript, though. Some moderate to minor comments follow.

First, when fitting the marginal distributions, in section 2.8, exponential functions are used for sea level and wave height. The exponential function has one parameter to fit, so I assume that the 2 parameters here refer to the two marginal CCDFs (one parameter each). I am not convinced that exponential functions are the most appropriate for extreme value analyses, which are the values the authors intend to extrapolate. Statistically, they may indicate better performance when using tests as AIC or BIC simply because the number of parameters to be fitted is smaller (one vs 2 or 3 in GEV, etc). But this does not mean that the fitting is better. In fact, extreme distributions do converge to these families of distributions (provided they fulfil the hypotheses). This choice is based on the results by Leijala (2018) which are in turn based on Särkkä (2017). The only reason provided there to use the exponential function was the number of parameters, but I do not think this is justified enough. Also, to me, one major issue here is why are the authors fitting the entire dataset to a distribution if the focus is on extremes. I think it would be better to select only those values considered as extremes in either one or another variable. That would make the rest of the computations easier too.

In the results on the correlations (section 3.1) I disagree with the statement that Hs and sea level are correlated. The correlation is not apparent from figure 3, since values of Hs exceeding 1 m may occur with nearly any value of sea level. The analysis provided later looking at correlations as a function of wind direction is more meaningful.

Actually, Figure 5 is very illustrative. However, I do not understand the results. When wind blows from the SW then there exists correlation between Hs and sea level at the buoy in Suomenlinna; but this buoy has some islets in its SW so it should be on their shadow. Am I missing something? or perhaps I am misinterpreting the angles. In any case it would be useful to specify the convention used for the directions in the text.

Still in section 3.1, I do not think that computing the correlations between spectra is the way to calculate how the correlation changes for different timescales in the time series. Cross-spectra provides that information partly. The alternative is to filter the original records using band-pass filters of the desired frequencies and then correlate each of them. Also, when correlations are stated, it is necessary to provide confidence intervals. I very much doubt that 0.2 is a significant correlation, for example. I am not sure that this part of the section is meaningful, at least not in its present form.

In section 3.2, when using all hourly values within the time series, any of the copulas seem to fit the observations satisfactorily, as shown in figure 7. Given the variability of the correlations with wind (and thus wave) direction, it would make sense to restrict the analyses to those periods of time when the two variables do show a coherent behaviour. This would surely improve the estimates.

In summary, I think the authors are working with ta data set for which it is worth exploring the dependences. In my opinion this should be done differently though. First, the authors should consider using only extreme for both variables. This reduces the number of data but the relationship is clearer. Second, extreme should be fitted with a suitable distribution for extremes. Third, events of either sea level or waves caused by different wind directions are very likely to belong to different families of distribution, since they probably arise from different atmospheric perturbations (e.g. they travel in distinct directions). This implies that the data should be treated and analysed separately. Copula functions should be also fitted for every subset in terms of direction and using the corresponding rank correlation. There are statistical tests to select the best copula fit, in case they show similar performance when compared to the observations. The comparison to the independent case is useful but cannot be taken as realistic if the Kendall correlation is high. Finally, I would suggest to remove the correlations of the spectra.

Other comments:

-Page 2, lines 17-19: there are many others: Wahl et al (2015) (https://doi.org/10.1038/NCLIMATE2736) for rain and storm surges; Arns et al (2017) (https://doi.org/10.1038/srep40171) for surges and waves; Marcos et al (2019) (10.1029/2019GL082599) for surges and waves too but globally. And references therein...

- equation 1 in page 7: this is instantaneous water level at a water depth of around 20 m, where waves are measured/modelled. Sea level maxima are generally defined over periods longer than just a few seconds, so the exact meaning of z must be clearly specified to avoid misinterpretations.

-p. 8, l. 20: please, provide the full name of the library used

-p.8, l. 22: units of the frequency are missing (I guess h^-1).

-p.10, l.3-4: normalization will not impact the Pearson correlations

-p.10, l.5: weaker->weaken?

-p. 15, l.8-9: the bias of the model in extreme waves is not discussed enough. This is an important shortcoming of this work

-p. 17,l.33: the observed sea levels correlate-> actually, this is true only for particular prevailing wind conditions.

-p.18, l. 2-3: "including hypothetical no-ice wave heights during the ice season did not markedly alter the correlation,". This is the expectation, right? why would this change if the relationship wind and waves remain the same?