Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC1
https://doi.org/10.5194/nhess-2021-55-AC1, 2021

Reply on RC1

Milla M. Johansson et al.

Author 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-AC1, 2021

We thank Reviewer #1 for the constructive comments which helped us to improve the manuscript. Our detailed responses are given below.

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.

We added the formula of the two-parameter exponential function to the manuscript: $F = 1 - \exp(-\lambda(x-x0))$.

Our purpose in this study was to study the correlation between simultaneous sea level and wave height in general, not just the correlation of the extreme events (which may differ). Thus, we chose to use the entire data set of hourly values. The frequency distribution of these is not an extreme value distribution, and clearly does not qualify for fitting GEV, for instance. It is not well defined which form the distribution of hourly sea levels follows, and speculating with the best-fitting distribution is not in the core focus of this study (although it is an interesting topic for further studies). Thus, we chose a simple two-parameter function which fits the data.

So that our results would better highlight the fact that we are not focusing on extremes only, we chose to show the probability levels 1/10 and 1/100 h yr^{-1} and leave the 1/1000 h yr^{-1} values out.

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.

This is true, the correlation in the entire data set is not apparent. We removed the mention of this and changed the text in 3.1 to: "The simultaneous hourly sea levels and significant wave heights (Fig. 3) indicate some positive tail dependence -- high sea levels co-occurring with high waves -- both in the observed and simulated data. Though high waves also occur with moderate or even low sea levels, and high sea levels are not necessarily accompanied by high waves either."

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.

It is true that there are some small islands shadowing the wave buoy location in the south-west. There are also larger islands to the southeast. These surrounding islands certainly provide shelter, but there still a significant amount of energy that are propagated around them, especially the longer waves. This is evident form e.g. Figure 4 in Björkqvist et al. (2020).

As for the correlation, it is more driven by the sea-level dynamics than the waves. Strong winds from both southwest and east can generate waves, but only westerly winds will simultaneously drive up the sea level. The easterly winds actually empties the Gulf of Finland, hence the positive correlation for southwesterly winds and negative correlation for easterly winds.

References: Björkqvist, J.-V., Vähä-Piikkiö, O., Alari, V., Kuznetsova, A., and Tuomi, L., 2020: WAM, SWAN and WAVEWATCH III in the Finnish archipelago – the effect of spectral performance on bulk wave parameters, J. Oper. Oceanogr., 13, 55–70, DOI: 10.1080/1755876X.2019.1633236

We added the convention to the figure caption: “The direction given is the direction from where the wind blows (nautical convention), …”

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.

We decided to remove this part of Section 3.1 from the manuscript.

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.

We repeated the copula analysis separately for the data corresponding to westerly-wind situations, and easterly-wind situations. It turns out that the westerly-wind results are qualitatively similar, although with stronger correlation, than the results for the entire data set. The very low or non-existent correlation in the easterly-wind data leads to the copula method being rather useless.

We added to the text: “We tested the effect of wind direction on the copula results by doing the copula analysis separately for two subsets of the data: situations with westerly wind (160°--340°, Fig. 7b) and easterly wind (340° --160°, not shown). The results for the former subset are qualitatively similar to those including all data: copulas are divided into two groups. The stronger correlation (tau = 0.269) results in somewhat higher probabilities of high total water levels in the distributions. In the easterly-wind case, on the other hand, the non-existent correlation (tau = -0.012) leads to the copula analysis being somewhat pointless, as all the copulas end up close to the independent case.”

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.

The proposed analysis of extreme values is indeed interesting, but the aim of this study was to investigate the two processes more broadly. Also see our response above.

We repeated the copula analysis for two different data sets for different wind directions, see response above.

We added a goodness-of-fit analysis for the copulas by calculating the Cramér-von Mises statistics.

We removed the correlations of the spectra as suggested.

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...

We added these references to the Introduction.

- 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.

We clarified the text to read: "We defined the total water level, z_{max}, as the hourly maximum level to which the continuous water mass reaches as a combined effect of the still water level, z_{still}, and the wave action. To do this, we first assumed that z_{still} does not change during an hour. Neglecting site specific coastal effects, z_{max} in such case is determined by the highest individual wave crest, η_{max}. (Note that wave crest denotes the height above z_{still} and is thus half of the actual height of the highest wave.) At Suomenlinna the highest single wave crest during 30 minutes is approximately 92 % of the significant wave height (Björkqvist et al., 2019), which we rounded up to $\eta_{max} = H_s$ for hourly values. With these assumptions, the maximum water level elevation is…"

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

We added the full name and proper references.

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

We changed the units of all frequencies to $h\ yr^{-1}$.

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

This part of the manuscript was removed based on a previous comment.

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

This part of the manuscript was removed based on a previous comment.

-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

The overestimation of the highest values by the model is a shortcoming, but this is, nonetheless, not a study of the behaviour of extreme values. Our work studies the correlation of water level variations and surface waves more generally, and since the extreme values only account for a small fraction of the data, the main results are not dependent on the model performance for the highest wave heights. Indeed, Figure 5 shows that the correlation found in the observations is well reproduced using the model data. Also, the results from the Copulas are based on observations, and are therefore not tainted by possible model errors.

We have added a paragraph to the discussion that directly addresses this aspect of the model performance:

"The validation suggests that the highest wave heights are overestimated in the simulations. The attenuation of longer waves through bottom processes is indirectly accounted for in the model through the calibration against measurements. Nonetheless, this calibration might not hold for even harsher weather conditions and even longer waves. While the simulations offer a good tool for assessing the general dependence between the wave height and water level variations, they should not be used as is for analyzing extreme values. If the wave simulations of the parametric model are to be used for extreme conditions, the wave-bottom interactions need to be accounted for in a more explicit manner, and such improved model needs to be re-validated."

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

We changed the text to: "The sea level variations and significant wave heights show a positive correlation in general (tau = 0.20). The correlation depends on wind direction: southwesterly winds lead to strongest positive correlation (up to tau = 0.5), while northeasterly winds lead to zero or negative correlation."

-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?

We agree that this is not a surprising result, but also not immediately obvious because of two reasons:

1) The winds are somewhat stronger during the ice-time, which means that the hypothetical waves simulated for the ice time (but without including ice) might be higher than the waves during the ice-free time.

2) The sea level variations that are measured during the ice time are damped because of the ice, and are probably therefore smaller than during the ice-free time.

Points 1) and 2) together means that the correlation of the two variables during the hypothetical ice-free time is not trivially the same as for the ice-free time. The ice-time is short enough that this doesn't have a large impact, but we still feel it is worth to try to quantify the effect, especially since we expect that the seasonal ice-cover might change in the future.