

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC2  
<https://doi.org/10.5194/nhess-2021-55-AC2>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC2

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-AC2>, 2021

---

We thank Reviewer #2 for the constructive comments which helped us to improve our manuscript. Our detailed responses are given below.

*This paper assesses the role of sea level and wind waves in generating total water level events in the Gulf of Finland, which is covered by ice during certain times of the year, making it a more challenging but also very interesting analysis. I think the content of the manuscript is novel and deserves publication with NHESS. While most of the analysis is technically sound and well-presented there are some aspects which I think require a bit more work before the paper can be recommended for publication. I summarize these below split into one major general comment and several (mostly minor) specific comments.*

### *General comment:*

*I find the copula analysis part of the paper pretty weak. The authors decide to only use three copulas, without providing convincing arguments for that selection other than pointing to previous literature. In the past especially people using Matlab ended up focusing on the Archimedean copulas as those were implemented and easy to use. However, the development of the MvCAT copula toolbox by Sadegh et al. (2017) has made it much easier to draw from a larger set of copulas. In R there are also many more copulas readily available to use. More important than using a larger set of copulas would be to show whether or not the copulas that are used at the moment are actually capable of capturing the dependence structure of the observations. There are many different goodness-of-fit tests available. Without any such tests blindly applying a random set of copulas does not provide relevant insight. I strongly encourage the authors to invest a little bit more time in strengthening this part of the analysis, as I find it to be an important component (otherwise it could just be left out but that puts a hole into the analysis).*

We extended our analysis to include nine different copulas. We also calculated a goodness-of-fit test based on the Cramér-von Mises statistic.

### *Specific comments*

*P1, l6 make clear in the abstract that  $H_s$  is used to represent waves*

We added "significant wave height" to the abstract.

*P1, l11 the one hour per 100 years sounds a bit strange; I understand the reasoning but isn't it just a 100-year event in the end?*

We changed the sentence to a more general form: "Compared to this, a probability distribution of the sum calculated by assuming that the two variables are independent underestimates the probabilities of high total water levels."

*P1, l18 It would be good to already mention here when defining total water levels that tides are negligible, not everyone will know that and wonder about the definition*

We added this to the text: "On the Baltic Sea coasts, the still water level variations mainly consist of storm surges and longer-term variations, the amplitude of tides being small."

*P1, l21 "in many locations"*

Corrected.

*P2, l25 why not say "decimeters"?*

Corrected.

*P3, l29 consider replacing "patched" by "inferred"*

We changed this to "estimated".

*Fig. 1 mention in caption what the contours are*

We added this to the caption: "The contours show the water depth in metres."

*P4, l5 is the trend also statistically insignificant?*

The trend is statistically significant due to large number of data points. Our intention was

to say that its magnitude (-2 cm) is negligible compared to the amplitude of the variations. Statistical significance plays no role in this. We changed the paragraph to read:

"The average sea level in 1992–2019 at Helsinki was +0.21 m, variations ranging from -0.73 to +1.69 m (Table 1). The linear trend in the time series was -2 cm in 28 years – the land uplift and sea level rise practically canceling each other. These long-term phenomena thus have a negligible contribution to the sea level variations studied here."

*Sect. 2.2 mention somewhere early on why temperature is needed (it's not a typical variable one would use in most areas; recalling that it is used to determine ice-free periods would be helpful)*

We added the explanation that the temperatures were used as a forcing for the parametric wave model. The ice-free periods were determined from the ice charts.

*P6, l2 not sure if "long waves" is a good term here as it is basically reserved for long period waves (not travel distance)*

You are right, and we are actually referring to long period waves in the sense that they are longer than what can be generated by the local wind and the nearest fetch. This mix of different wave systems is typical in the archipelago (Björkqvist et al. 2019). We now realize the ambiguity and will change the text in the manuscript to the following:

"We then determined the fraction of the GoF wave energy that arrived to Suomenlinna; these waves have longer periods than what can be generated by the local wind and the nearest fetch, and will henceforth be called "long waves"."

*P7, l16 I had a hard time following this definition, if  $H_s$  is the average of the 33% highest waves how is the highest single wave lower than that?*

We clarified the text to explain that wave crest (above the still water level) is only half of the wave height:

"Neglecting site specific coastal effects,  $z_{\max}$  in such case is determined by the highest individual wave crest,  $\eta_{\max}$ . (Note that wave crest denotes the height above  $z_{\text{still}}$ , thus being 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), ..."

*Eq. 3: using theta which is special to the Archimedean (and one parameter) copulas makes that an alteration of Sklar's theorem*

This is true. We removed theta from the equation.

*P8, l19 first spell out and then introduce the abbreviation CCDFs*

Corrected.

*Fig. 3 when describing the results it would be helpful to recall that sim I has wave heights set to zero*

We added this to figure caption.

*P10, I5 "sea level variations are weaker"*

This part (analysis of the spectra) was removed from the manuscript based on a comment from Reviewer #1.

*Fig. 5 caption: the values for tau from the entire data are not in table 3 (the one for observations of 0.2 is included but just not rounded, but the 0.16 is not)*

The tau values (0.20 for observations 2016-2019, and 0.16 for simulations 1992-2019, set F) are given in Table 2. We added a reference to Table 2 in the figure caption.

*Fig. 7 & 8: make clear that it y-axis shows exceedance probability per hour (or translate values to exceedance probability per year)*

We translated the values in the y axes of the figures to " $\text{h yr}^{-1}$ ", and corrected the text accordingly.

*Fig. 8 It would be good to show the convolution results from figure 7c as well for direct comparison (only the one for 1992 to 2019); also see general comment above about testing which (or any) of the copulas are actually a good choice*

In Fig. 8, the convolution results from Fig. 7 (for the observations 2016-2019) are also shown. We clarified this in the caption: "The observation-based CCDF of total water level in 2016–2019, the CCDF calculated assuming the variables independent, as well as CCDFs calculated by taking  $10^8$  random samples based on the nine different copulas..."

We added a goodness-of-fit test for the copulas based on the Cramér-von Mises statistic.

*P15, I8 "is the highest"*

We changed this to "for the highest".

*P16, I6 "quite" is a better term then "peaceful"*

We changed this to "calm".

*P16, l6-8 that sentence wasn't clear to me please explain better what it's supposed to tell the reader and why it's relevant*

We tried to tell the reader that the 4-year period of observations seems to be on the calmer side when considering all possible 4-year periods of the study. This then reinforces that it is not representative in itself, and longer time series are needed for robust conclusions. We have modified this section by removing a lot of the numbers and only keeping the core message. We hope that this improves the readability. It now reads:

"The observational period of this study (2016--2019) was relatively calm: the 4-year maximum sea level (1.11 m) was lowest among the seven consecutive 4-year sets, while the maximum simulated significant wave height (1.89 m) ranked fourth. To obtain reliable estimates for probabilities of extreme sea levels on the Finnish coast, at least 30 years of data are generally used [...]"

*P17, l22-25 In the discussion further up the authors correctly point to the fact that the observed data is way too short to infer information for longer return periods (I would have commented on that if I hadn't seen that remark); why not focus (or at least add) results for a more reasonable return period, such as 10 years or so? At least we know the results would be more robust.*

We changed the manuscript so that we only show results for return periods of 10 and 100 years, and left 1000 years out.