Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-190-AC1, 2020

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Wave height return periods from combined measurement–model data: A Baltic Sea case study" *by* Jan-Victor Björkqvist et al.

Jan-Victor Björkqvist et al.

jan-victor.bjorkqvist@fmi.fi

Received and published: 22 September 2020

R1: This is an interesting and very well written paper. I have no objection to it being published as submitted, but I would very much like to see the authors' responses to my comments.

Our response: Thank you. We appreciate you agreeing to review our paper. Please see our responses to your comments below.

R1: The most interesting and novel part of the paper is the treatment of sampling variability in extreme value estimation. Figure 2 is very informative. I have not seen a

Printer-friendly version

Discussion paper

spectral analysis of Hs history before. It is interesting that filtering removed information from time scales longer than three hours, despite the fact that the Gaussian filter half powerpoint was one hour. The moving averages of course had side lobes. It might be worthwhile to investigate more sophisticated digital filters. But the more interesting question is what time scales are really represented in hindcast data that is reported every hour but is based on three or six hour wind fields. Perhaps more work with filtering the continuous measurement time series would help answer that question.

Our response: We want to start by pointing out that the one hour value for the Gaussian filter was the standard deviation, meaning that it is expected to filter also longer time scales, since the filter also “reaches” them beyond the one standard deviation. Compared to moving averages or Fourier-filter, the Gaussian filter doesn’t have a equally sharp, well defined, cut-off. This is why we also compared it to the moving average, and ensured that the filter is functioning on time scales close to what Forristal et al. (1996) recommended.

We fully agree that the issue of what time scales are actually represented (both in modelled and measured data) is by no means obvious, nor solved in this paper. For modelled data the use of the identical forcing with different intervals (i.e. every 15 minutes, every hour etc.) should probably be used, since using different products introduces other sources of uncertainty. This would also give the opportunity to see which filters (if any) can consolidate wave data generated with the different wind forcings. Here a wide variety of more advance filters should be used (as you suggested).

The data used in our study can, unfortunately, not meet the needs of a more detailed study into this subject, because of other sources of uncertainty. We therefore had to limit ourselves to raising the subject up for discussion and making a first attempt at a reasonable solution.

R1: Different users are interested in different time scales. Ship designers often want to know the three hour sea state for use in model basins. Calculating extreme values of individual wave heights from shorter averaging times where Hs is not varying may be more accurate. Would calculations of individual wave heights from the hindcast data match those from thirty minute measurements (or chi-squared augmented hindcasts)?

Our response: Unless the lack of sampling variability is accounted for, then model and measurement data probably doesn't match for extreme values (as noted by Forristall et al. 1996 and our Fig. 3c). For mean values they should match. If sampling variability is properly accounted for, then also extreme statistics should match (assuming a perfect model etc.), but to account for this perfectly may not be possible (see also the following answer). The longer the averaging time, the less of an issue sampling variability becomes, but using an assumption of stationarity for three hours could probably be a large source of error, especially in sheltered areas and small basins. So you have to choose your poison in many cases.

R1: The difference in return periods between the chi-squared and filtered analyses deserves comment. If the chi-squared augmentation worked perfectly, wouldn't they be equal? Looking at Figure 3, it seems that the extreme wave height in the measurements has a larger deviation from the smooth curve than most of the artificial chi-squared data. That makes me think that the measurement is an outlier to the chi-squared distribution. Why don't you plot the variability of the measurements against a chi-squared distribution to check that?

Our response: If the chi-squared augmentation AND the filtering were perfect, then the results from the two data sets should agree. However, both are likely to be flawed, and this is the reason we decided to include both approaches even though they make the paper a bit more difficult to grasp. You are probably right that the maximum measured wave height is perhaps on the tail end of the distribution (which, to be fair, is not surprising for an unexpectedly high measurement). Of course, even if the chi-squared augmentation is perfect, it can only match the variability in an average sense (please also note, that panel a) and b) are different storms, since neither hindcast covered 2019).

As to plotting the variability of the measurements against a chi-squared distribution. This is a very attractive idea, but it would require us to know the "true" underlying significant wave height to relate the variation to that value (i.e. the measured values needs to be normalized by the "true" values, otherwise we are just plotting the distribution of the significant wave height, not the chi-squared distribution of the variation). One attempt to find the "true" value is the filtered time series, but we know this is not perfect. The other approach is to add variability to the model, but then we obviously assume that the variability follows a certain form. The Hs-spectra in Fig. 2 are essentially an attempt to visualize how well we are capturing the differences between the "true"/modelled and the measured/chi-squared-aumented values, even though we can never know both values in either pair.

R1: And finally, why do you think the hindcast of the recent storm was so bad?

Our response: We are not quite sure which storm this is referring to. If it refers to Fig 5, then in out opinion the hindcast was not that bad, with a quite accurate timing, although with a slight overestimation of the significant wave height. For the bias, the most obvious culprit is the wind forcing. The HARMONIE wind product is known to produce a positive bias in the modelled significant wave height. As to why this is, it is probably a part of the more general problem of wave model development, namely that we have to "tune to the mean" even though we are interested in the extremes. In other words, the physics might change in extremely high winds.

We also want to point out that several aspects of the wave field were simulated correctly, as the wave period and wave direction time series (Fib 5 b & c) show. Lastly, the accuracy of the ice product is normally also a possible source of error in this region, but this was not the case during this mild winter.

NB

Dr Jani Särkkä informed us, that while the ERA-Interim had a resolution of 3 hours, the downscaled product actually had a temporal resolution of 1 hours. Nonetheless, it is evident from the spectra in Fig. 2 that the WAM data doesn't capture the same temporal scales as WAM forced with a wind forcing with a native temporal resolution of 1 hour (compared to Fig. 5.1, page 39 in Björkqvist, 2020). We will correct this to the text and amend the discussion to reflect what we stated above.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-190, 2020.