RESPONSE TO REFEREE 1 (Report 2)

Above all, I feel obliged to thank Referee 1 for his suggestions and continuous useful guidance.

I have read the answers of the author to my comments. Here are my further comments referring to the points listed in "nhess-2017-177-author_response-version2.pdf". I confirm that the paper is in my view interesting, the new structure has improved the readability of the article, but the new version remains affected by two shortcomings that I had already noticed in the first review:

a) the reliability of the hindcasts (particularly for waves and for intense/extreme events) is not convincingly documented

b) there insufficient effort to interpret the dependences that are found. On both these aspects I think that progresses are possible with a reasonable, but not huge effort.

As a general comment,
effort has been made to overcome both shortcomings. Details are given below.

Considering the specific answers of the author to my comment:

(1) Validation of hindcasts and its presentation

The goal of this paper is, indeed, not the validation of the hindcasts. However, the reliability of the analysis depends on the accuracy of the hindcasts, which is essential for this study. The minimal requirement is a documented validation of the two (storm surge and wave) hindcasts in previously published studies, which is what the author has partially added. However, the answer of the author is not fully clear on this respect, particularly as the published article validating the wave hindcast is concerned. I had formerly commented that "No sufficient attention is paid to assessing whether storm surge and wave simulations are capable of reproducing extremes values. Correlation and bias are not representative in this sense". I find in the revised version no new material answering to this request. The added information referring to Vousdoukas et al. (2016), in not relevant for intense and extreme events. Further absolute values are not representative of the actual importance of these errors. I think that values would be more informative if they are normalized with the values of the mean observation. Concerning waves, author refers to Phillips, B.T., Brown, J.M., Bidlot, J.-R. and Plater, A.J.: Role of Beach Morphology in Wave Overtopping Hazard Assessment. J. Mar. Sci. Eng. 5(1), 2017. I have checked this article, but I Have not found the validation of the wave hindcast that the author claim. I might have missed it. A specific indication by the author is required.

The description of the set of wave hindcasts used in the current study is contained in Page 5 (Line 8) of Phillips et al. (2017) as shown below:

"Here, we use data from a rerun of the European Centre Wave Model (ECWAM, [47]) Cycle 41R1 (28-km resolution) to provide a longer dataset. The model is forced by a six hourly ERA-interim wind field with no wave data assimilation. The effect of water level change and surface current due to tides and surge is negated. The output used is hourly H_s (significant wave height) and T_p (wave period) from 1979–2014 and was extracted from the offshore limit of the XBeach boundary. This global hindcast was produced in preparation of the ECMWF (European Centre for Medium Range Weather Forecasts) next reanalysis (ERA5)".

"The ECMWF wave hindcast exhibits an RMSE (root mean square error) of 0.31 m, a bias (the difference between the estimator's expected value and the true value of the parameter being estimated) of 0.06 m and a symmetric slope value of 1.11 for H_s when validated against the Liverpool Bay Wave Buoy" …

Reference ECWAM, [47]. "European Centre for Medium Range Weather Forecasts. IFS Documentation | ECMWF; European Centre for Medium Range Weather Forecasts: Reading, UK, 2016".

In synthesis, considering the new version I remain unconvinced that the statement "the overall performance of both surge and wave hindcasts is considered satisfactory" is documented at a convincing level.

---

*The set of storm surge hindcasts originated from Vousdoukas et al. (2016) have specifically used for projections of **extreme storm surge levels** along Europe and it seems as an appropriate dataset for the current paper. This is the main reason of utilizing such storm surge dataset in my current work. On the other hand, I have to admit that wave hindcasts based on the ERA5 significant wave reanalysis dataset used in my study have not been thoroughly tested as for the validity of their extreme values since ERA5 is still in production phase as it can be seen at*
https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production
https://www.ecmwf.int/en/about/media-centre/science-blog/2017/era5-new-reanalysis-weather-and-climate-data

Table 1: Details of the wave buoys used for the validation of extremes.

	ID	name	lat	lon	days	corr
MED	61217	Adriatic Sea	42.41	14.54	203	0.94
	61218	Adriatic Sea	43.83	13.72	1,588	0.87
	61280	Balearic Sea	40.69	1.48	2,764	0.87
BIS	62001	Bay of Biscay	45.20	-5.00	6,012	0.97
IRI	62091	Irish Sea	53.48	-5.43	4,991	0.93
	62094	Irish Sea	51.70	-6.70	3,727	0.94
	62301	Irish Sea	52.40	-4.70	5,339	0.93
	62303	Irish Sea	51.50	-5.10	7,426	0.94
NOS	62127	North Sea	54.00	0.70	1,158	0.92
	62142	North Sea	53.00	2.10	5,537	0.92
	62145	North Sea	53.10	2.80	5,796	0.92
	63115	North Sea	61.60	1.30	2,922	0.97
	HvH-LiG	North Sea	51.93	3.40	1,114	0.92

Due to this (limitation), an investigation was performed (as requested) over a set of 13 wave buoys along European coasts capable of ***providing enough hourly data*** *for such an analysis. The details of wave buoys used are shown in Table 1. Buoys over Mediterranean are denoted as MED, over Bay of Biscay as BIS, over Irish Sea as IRI and over North Sea as NOS. HvH-LiG refers to the wave buoy Lighteland Goeree stationed near the coastal area of Hook van Holland (NL).*

Based on these 13 wave buoys (Table 1), a dataset of 48,547 pairs of daily maxima was compiled comprising hindcast and observation wave values. The mean error (bias) of hindcasts was found to be equal to -0.29 m with a corresponding rmse of 0.56 m. From a closer investigation, it was found that in some cases hindcasts were not capturing the exact magnitude of extremes.

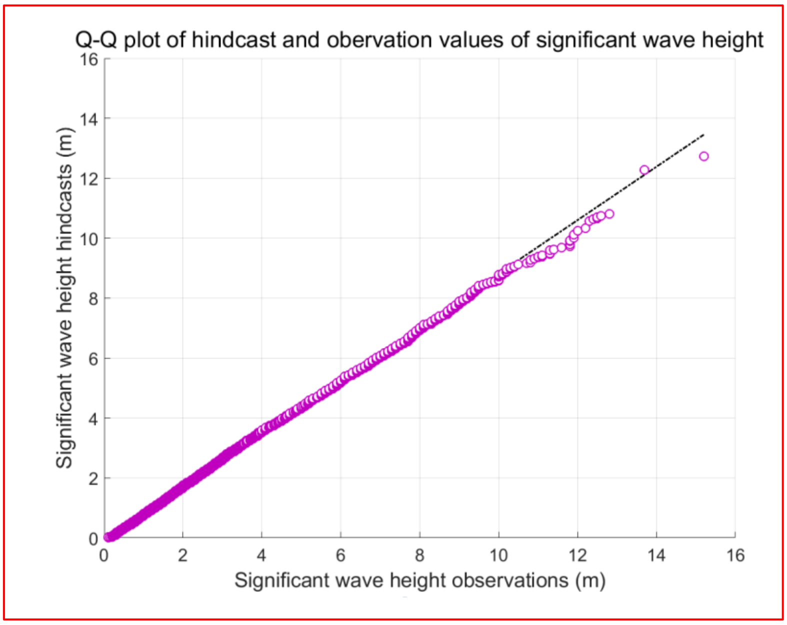

Figure 1. Q-Q (quantile-quantile) plot of hindcast and observation values of significant wave height

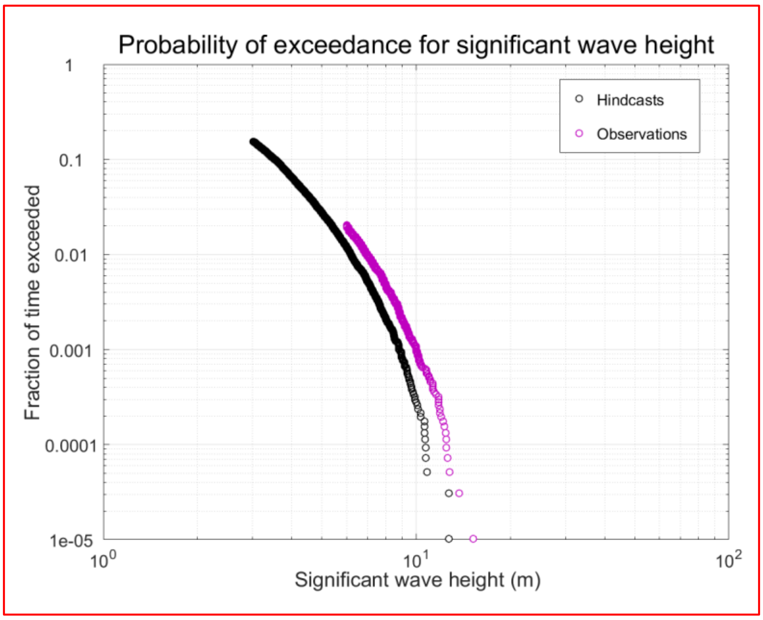

Figure 2. PoE (Probability of Exceedance) for wave observations (red color) and hindcasts (blue color).

In addition, both Figure 1 (Quantile-Quantile Plot) and Figure 2 (Probability of Exceedance Plot) seem to support this unavoidable limitation since due to the low resolution models used for reproducing time series of significant weather parameters (as in this case), extremes cannot be captured with their exact (high-impact) value but in most cases their "footprint" signal can be resolved as a spike of a lesser value.

A relevant example can be seen in Petroliagis and Pinson (2012) where the footprints of extreme wind speed values over Bremen airport are captured by ERA-Interim as footprint spikes although significantly underestimated (compared to observations) but still capable of resolving extremes as shown in Figure 3 (taken from Figure 7 of Petroliagis & Pinson).

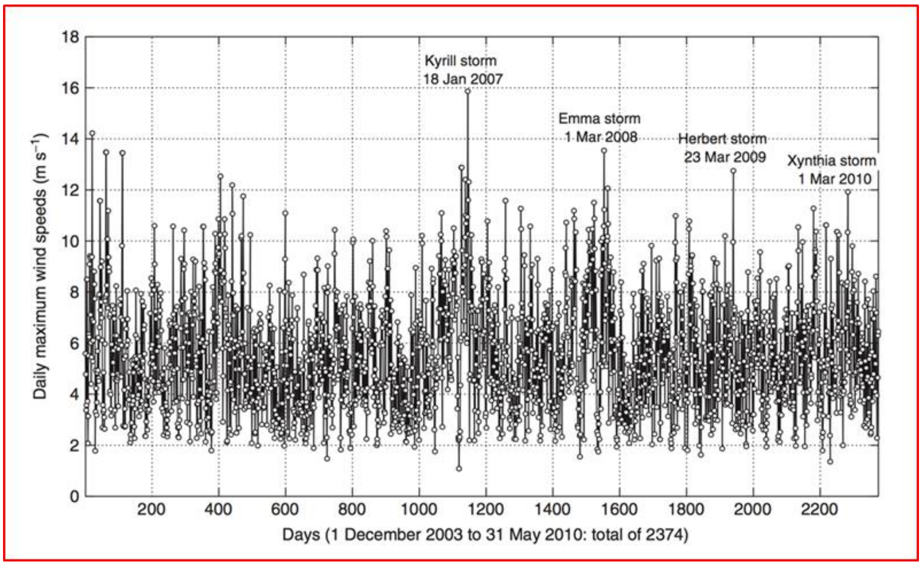

Figure 3. Time series of daily maximum wind speed values for Hannover over a period of 2374 days (from 1 December 2003 to 31 May 2010) in 'Reanalysis' mode. Peak values (significantly underestimated compared to observations) corresponding to Kyrill, Emma, Herbert and Xynthia storms are highlighted.

In a similar way, during the estimation of statistical dependence such footprints seem to be capable of determining the days of the ***most extreme wave daily maxima****. The main issue in estimating dependence is not the exact magnitude of extremes (i.e., how well are resolved by hindcasts) but rather if a spike (footprint of extreme) exists on a specific day denoting the exceedance over a critical percentile threshold of hindcasts. If such (correct) footprint is considered as a hit, Table 2 (current document) has been compiled containing the number of hits over a set of critical (hindcast & observation) wave thresholds in a POT (Peaks Over Threshold) environment.*

Taking into consideration that during the estimation of dependence (Table 6 and Table 7 of Technical Supplement), threshold (percentile) wave values ranging from 86.2 to 98.8% were used, this corresponds to 80 to 85% hits (correct footprint spikes of daily maxima denoting an extreme).

Lastly, in cases of ***compound (surge & wave)*** *footprints of extremes (resolved by hindcasts), Table 4 (Technical Supplement) has been compiled where the 98.5% percentile extremes of storm surge observations are compared to their corresponding hindcast values (falling in the same 98.5% category). Same way in Table 5 (Technical Supplement), the footprints of significant wave height observation extremes are compared to their corresponding hindcast (or lesser intensity) values.*

Table 2: Number of hits for various hindcast and observation thresholds (percentiles).

thrs (%)	***hind***	***obs***	***events***	***hits***	***score***
55	*1.51*	*1.80*	*6,617*	*6,129*	*93 %*
60	*1.81*	*2.10*	*5,673*	*5,135*	*91 %*
65	*1.97*	*2.30*	*5,041*	*4,586*	*91 %*
70	*2.17*	*2.50*	*4,506*	*4,023*	*89 %*
75	*2.41*	*2.80*	*3,737*	*3,329*	*89 %*
80	*2.69*	*3.05*	*3,281*	*2,821*	*86 %*
85	*3.05*	*3.40*	*2,527*	*2,147*	*85 %*
90	*3.53*	*4.00*	*1,699*	*1,439*	*85 %*
91	*3.65*	*4.10*	*1,586*	*1,314*	*83 %*
92	*3.77*	*4.30*	*1,394*	*1,159*	*83 %*
93	*3.91*	*4.50*	*1,243*	*1,033*	*83 %*
94	*4.08*	*4.70*	*1,070*	*881*	*83 %*
95	*4.29*	*4.90*	*948*	*768*	*81 %*
96	*4.56*	*5.20*	*775*	*620*	*80 %*
97	*4.89*	*5.60*	*601*	*479*	*80 %*
98	*5.38*	*6.20*	*417*	*337*	*81 %*
99	*6.18*	*7.20*	*233*	*186*	*80 %*

It becomes obvious that although hindcasts could not resolve the exact extremity of both surge and wave events, they were able to capture their footprints quite well. It is important to point out that hindcasts above all were ***capable of identifying and resolving all seven (7) compound events*** *that took place during the common time interval of 1,114 days.*

Based on above data and results, a new Section (Additional validation of wave hindcasts focusing on extremes) has been compiled and incorporated in the Technical Supplement as Sect. 2.

If the paper main goal is demonstration a methodology, this may be an acceptable limitation, but some rephrasing, admitting it and suggesting therefore, caution on the results that are found, is advisable.

The current work focuses on data preparation, parameter selection, methodology application and estimation of both correlation and statistical dependence between source variables. It also focuses on the prevailing (higher frequency) and dominant (higher intensity) low-level wind conditions over a set of preselected (top 80) extreme compound events. The critical time period during which such extremes take place is also analysed based on monthly frequency values of occurrence.

Nevertheless, taking into account the percentage of hits (% of correct spikes) presented in the previous section – ranging from 80 to 85% – caution (as requested) has been added in both Section 4 (Discussion) *and in Section 5 (Conclusions) just after the reference to the **compound validation over HvH** river ending point as shown below:*

> *"Since such "compound" validation is impossible to be repeated for all RIEN points, some caution with the exact levels of correlation and dependence should be bear in mind for the rest of the RIEN points".*

Other minor points in section 3

Section 3.1, line 6 depending on whether the hindcast includes the computation of nonlinear tide-surge interaction, the "residual" can include tide surge interaction as well.

Corrected (as suggested).

Section 3.1 Line 1 I think correct English expression would be "observation mode" … anyway note that I am not considering in this review typos.

Corrected (observation mode) throughout the text.

---

(2) Description of statistical methods

I am positive with the new version, which I think is more readable and useful to the reader. The section "statistical dependence" has now become very short, I suggest to merge it (as a separate subsection) with "Data and Methodology" in a single section.

This has been done. Statistical dependence has been incorporated into Data and Methodology that has become the new Joint Section 2 (Data and Methodology). Appropriate adaptations (to the text) have taken place also.

---

(3) Spatial variations of dependence and correlation, interpretation of results.

Obviously, the effect of atmospheric pressure (that in steady condition is described by the inverse barometer effect) is included in the model simulations. My comment was a suggestion that lack of dependence could be related to a large contribution of atmospheric pressure on storm surges, because waves are not affected by it. Large dependence could be related to the common action of wind on sea level and waves, depending on the characteristic of the fetch and on water depth along it. On this respect I think that the comment and *the added material does not answer to my request and the manuscript still would benefit from a discussion with an interpretation of results.*

The potential effect of atmospheric pressure in cases of low dependence has been added in both Section 4 (Discussion) and Section 5 (Conclusions). It seems that in cases of high (surge & wave) dependence, winds may have the leading (forcing) role whereas in cases of low dependence the contribution of atmospheric pressure tends to become more pronounced.

I accept that "a thorough understanding of all factors leading to a compound event is above the scope of this study ", but some discussion should be added, anyway. Any indication on under which conditions dependence would be high and when low?

As it has mentioned above, it seems that in cases of high dependence, winds (of an approaching storm) most probably have the leading role whereas in cases of low dependence the contribution of atmospheric pressure has the tendency of becoming more pronounced.

Based on this, it seems that compound events over northern sea areas are mostly driven (forced) by a common extreme (storm) wind event resulting in a high value of dependence between surge and wave, whereas a large contribution of atmospheric pressure affecting only storm surge might be one among other reasons for low dependence (of surge and wave) dependence over southern sea areas.

As already mentioned, the potential effect of atmospheric pressure in cases of low dependence has been added in both Section 4 (Discussion) and Section 5 (Conclusions) – see new updated version of manuscript.

I *suggest to split the final section "discussion and conclusion" in two sections: "discussion" and separated "conclusion". Reading it is not clear to me whether the goals of the paper is to demonstrate a methodology or to discuss dependences between storm surge and wave events at a representative set of European RIEN points.*

This (suggested) split has been taken place as Discussion (Section 4) and Conclusion (section 5). All necessary adaptations have also taken care of.

Comments on figures
In general figures need to be improved

As a general comment
almost all Figures have been updated (improved) as requested.

Names in figure 1 will likely not readable when reduced too journal size

---

Names in Figure 1 (River Ending Points) have been updated to be visible when reduced in journal size.

Quality of figure 2 is poor. I expect no geographical name will be readable and dots will be hardly visible when reduced to standard journal size

---

Figure 2 (both bias & rmse) has been updated. Quality has significantly improved.

Graphic quality of fig 3 is low. It contains a lot of useless details (e.g. roads and motorways labels) and unreadable (and useless in this context) names of towns and locations.

Figure 3 (main site positions) has been updated. Useless details have been removed.

In figs 4 and 6 annotations are too small. They will be unreadable in the printed journal article. This problem is present in most figures.

Figure 4 and 6 (using matlab) have been updated with new annotations significantly bigger so to be readable in a journal article.

Figures 1, 7, 8, 9 and 10 have also been updated.

In addition, after the split of Figure 2 a re-numbering of figures had to take place.

A renumbering of Sections took place also due to the merging of old Section 2 with Section 3.

References

Phillips, B.T., Brown, J.M., Bidlot, J.-R. and A.J. Plater, 2017. Role of Beach Morphology in Wave Overtopping Hazard Assessment. J. Mar. Sci. Eng. 5(1).

ECWAM, [47]. "European Centre for Medium Range Weather Forecasts. IFS Documentation | ECMWF; European Centre for Medium Range Weather Forecasts: Reading, UK, 2016".

Petroliagis, T.I. and P. Pinson, 2014. Early warnings of extreme winds using the ECMWF Extreme Forecast Index. Meteorological Applications, 21, 171–185. DOI: 10.1002/met.1339.