Authors' replies to Reviewer 2 comments for NHESS-2021-236

We would like to thank Dr. Philip Ward and the two anonymous Reviewers for their helpful comments and suggestions, which much improved our manuscript. We appreciate the constructive comments received, which are discussed in detail. In general, some of the comments in particular, dealing with the way comparisons between extreme value distributions are done, prompted a shift in the focus of the manuscript. As Reviewer 2 puts it, the manuscript provides a comparison between possible approaches to extreme water level analysis. We have thus proposed a more balanced title, which does not needlessly over-emphasize the most recently proposed method and more objectively reflects the findings and the newly introduced analyses.
Below we provide our discussion of Reviewer's comments (in blue italic font) and describe the changes addressing them.

Three approaches are compared in this research, i.e., GEV distribution on annual peak maxima (GEV-BM), the Metastatistical Extreme Value Distribution (MEVD), and GEV distribution on peaks over a higher threshold (GEV-POT). With respect to the latter approach, wouldn't it be better to rely on a Generalized Pareto Distribution (GPD) when threshold exceedances are considered? As far as I remember, GPD is a derivation of GEV for POT data; as such, is it conceptually correct to test a GEV distribution rather than a GPD on POT data? Please comment on this in the Methods section and/or extend the explanation in the Introduction (e.g. lines 37-39).
We now realize that the original description of the traditional extreme value analysis methods was not entirely clear. We have renamed what was originally called GEV-POT as POT-GPD, to avoid the confusion pointed out by this Reviewer. In particular, we do use a GPD distribution for the selected events over a high threshold in the POT approach (please also see discussion on differences between POT-GPD and the MEVD approach used here). The Materials and Methods section in the revised manuscript now contains a more detailed description of both GEV-BM and POT-GPD methods that clarifies the above points.

Lines 15-19 in the Introduction. As you speak of "active field" as for the modeling of extreme value probability of occurrence, you could reference more recent works.
Thank you for the suggestion. We agree and the Introduction section in the revised manuscript now contains more recent works, e.g. the following:

a. Miniussi, A., and Marra, F.: Estimation of extreme daily precipitation return levels at-site and in ungauged locations using the simplified MEV approach. Journal of Hydrology, 603(B), 126946, https://doi.org/10.1016/j.jhydrol.2021.126946, 2021.

b. Mekonnen, K., Melesse, A. M., Woldesenbet, T. A.: Effect of temporal sampling mismatches between satellite rainfall estimates and rain gauge observations on modelling extreme rainfall in the Upper Awash Basin, Ethiopia. Journal of Hydrology, 598, 126467, https://doi.org/10.1016/j.jhydrol.2021.126467, 2021.

c. Cancelliere, A.: Non Stationary Analysis of Extreme Events. Water Resources Management, 31, 3097-3110, https://doi.org/10.1007/s11269-017-1724-4, 2017.

d. Elvidge, S., and Angling, M.J.: Using extreme value theory for determining the probability of Carrington-like solar flares. Space Weather, 16, 417-421. https://doi.org/10.1002/2017SW001727, 2018.

e. Rypkema, D.C., Horvitz, C.C., and Tuljapurkar, S.: How climate affects extreme events and hence ecological population models, Ecology, 100, 6, https://doi.org/10.1002/ecy.2684, 2019.

f. Chan, S., Chu, J., Zhang, Y., and Nadarajah, S.: An extreme value analysis of the tail relationships between returns and volumes for high frequency cryptocurrencies, Research in International Business and Finance, 59, 101541, https://doi.org/10.1016/j.ribaf.2021.101541, 2022.

Line 29 in the Introduction. The list of reference is rather long; perhaps it would be enough to cite a few works and the "references therein".
Thanks for the suggestion.

Line 48 in the Introduction. You can also cite Solari et al. (2017).
Thanks, Solari et al. (2017) is now cited in the revised manuscript.

Page 4, Fig. 1. Please reduce the y-axis range for Marseille plot.
*A*greed. The y-axis range for Marseille was reduced as shown in the following revised Figure 1.

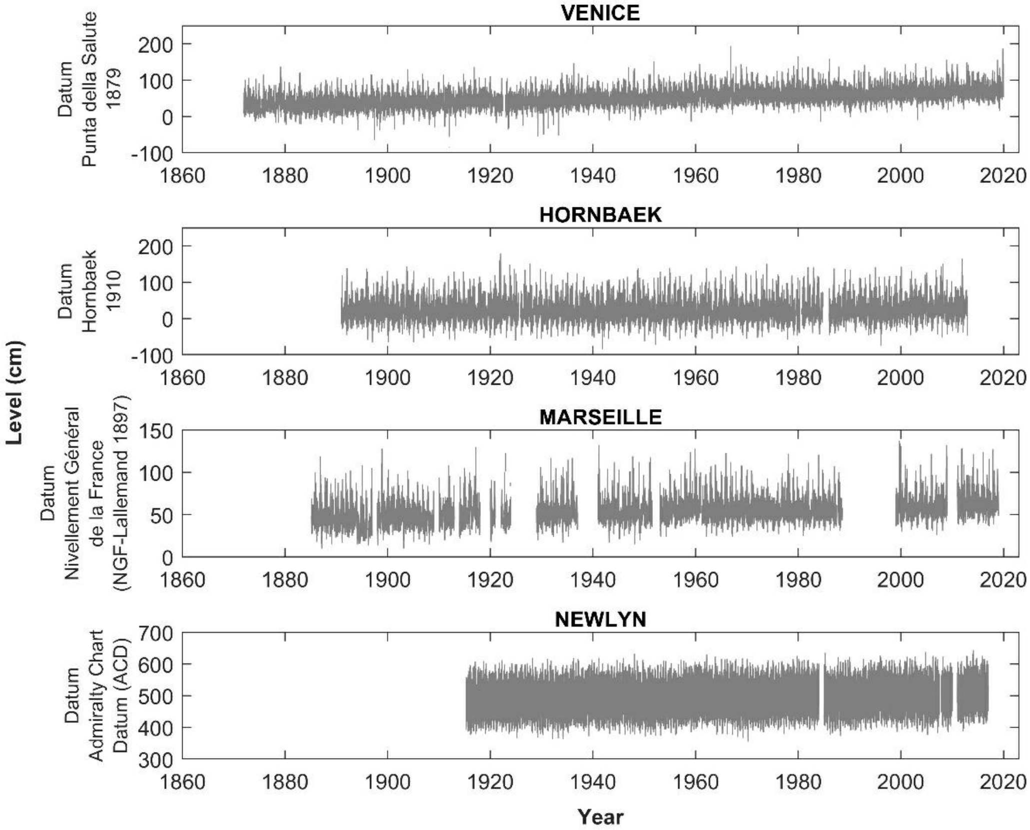

Figure 1. Revised Figure 1.

Page 5, line 107. If I understood correctly, "year" in the following line should be replaced with "block".
Thanks for this careful correction. It is correct, in the revised manuscript we have changed "year" with "block".

I would swap Section 2.2.1 and Section 2.2.2. First explain how you pre-processed the data, then the distribution used to model them.
Agreed.

Section 2.2.1. I think you should explain what are the cumulative distributions F you tested for the ordinary values, and which one did you choose.
In the revised manuscript we have clarified the potential distributions that were tested for sea level frequency analysis (i.e. Gamma, Weibull and Generalized Pareto distributions). Based on the comparative evaluation of the performance of these three probability distributions, described in the

revised supplementary materials, the Generalized Pareto Distribution emerged as the best model for the “ordinary” water level peaks (see revised Section 3.2, lines 264-268).

Section 2.2.2, line 134. The fact that you neglect the interactions between tides and surges means that gauges are placed in deep waters. Is that true? Please add the respective water depths in Table 1 if such info are available.

This statement may have generated some confusion and needs additional discussion. We now clarify that the tide-surge interaction is significant and needs to be taken into account when the surge and tide components are studied separately.

The revised manuscript at line 134 now explains that: “However, this effect is significant and needs to be taken into account when the surge and tide components are studied separately. Since here we do not attempt to separate these contributions but only analyze the sum given by the combination of the water level setup, induced by meteorological forcing, and the astronomical tide, hereafter we will neglect their non-linear interactions and will consider the observed sea level as the sum of additive components”.

Section 2.2.2. Please number the equations.

Thank you, agreed.

Section 2.2.3, lines 171-173. This paragraph is unclear. Indeed, looking at the correlograms (Fig. S1) it seems that independent events are achieved for no lags. This aspect is crucial so it should be better explained. Correlograms also reveal that tides are relevant (negligible) in Venice and Newly (Marseille and Hornbaek). Perhaps you could comment on this in the paper.

Thank you for this comment. In the Supporting Information, we have added Figure S2. The revised manuscript now discusses (lines 171-173) that: “The analysis of the correlograms of selected water level peaks shows that some correlation persists also for long time lags and also in the de-clustered time series. Even though the strength of this correlation is relatively small (the ACF is always less than 0.3) this periodic correlation should be considered in the interpretation of subsequent results, as it may impact the performance of statistical modelling. The de-clustering process does significantly decrease correlation as may be seen by comparing Figure S1 (ACF prior to de-clustering) and Figure S2(after de-clustering). Interestingly, it is seen that the tidal contribution (that generates periodicities in the ACF) is strongly visible in Venice and Newlyn, while it is quite small in Hornbaek and Marseille. However, after de-clustering, the residual correlation is associated with the tidal component also in Hornbæk and Marseille. The existing literature, which focuses on the storm-surge component only, uses shorter time lag values due to the shorter correlation of the surge component due to atmospheric drivers. For example, the independence…”.

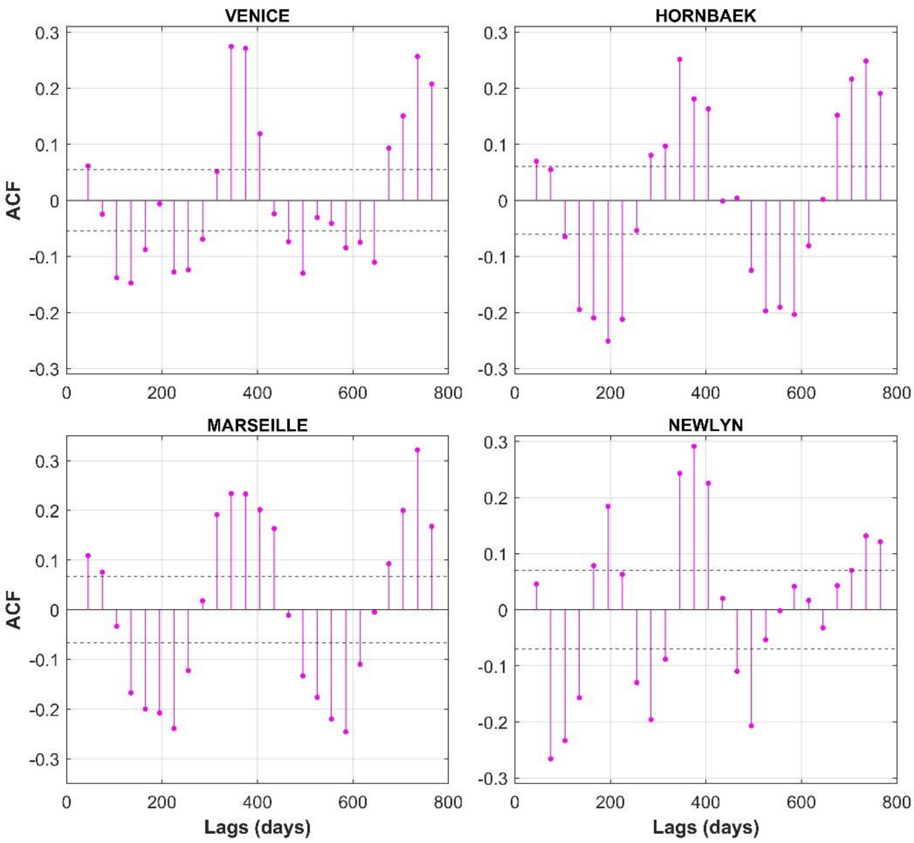

Figure S2. Correlograms for independent daily maxima water levels obtained by dividing the time series (of independent events) into 30-day bins.

Section 2.2.3, line 177. Please use consistent tenses throughout the paper when referencing other works. For instance, here you say "Bernardara et al. (2011) adopted", while previously you use the present tense (e.g. page 6, line 117 or later in the paper at page 8, line 212).

Thank you, we have revised the tenses throughout the manuscript.

Page 9, line 239. I do not understand why the return period is expressed for annual maxima (AM). Apologies but I am not familiar with the MEVD, however it is clear that it allows to select multiple events per year. Then, why Eq. (3) is defined with respect to AM data?

Exactly, the MEVD uses more observations to estimate the parameters of the distribution of "ordinary values". However, the MEVD cumulative distribution (Eq. (2)) is still the distribution of the annual maxima (line 100-103), estimated using a much greater sample than just yearly maxima.

It is not clear the purpose of Section 3.1, given that no non-stationary distributions are subsequently employed. However, if you want to keep it, I suggest adding the confidence intervals of the slopes fitted to the data (and perhaps comment them with respect to the p-values of the Mann-Kendall test).

Section 3.1 is a preliminary analysis of the extreme sea levels to understand whether long-term trends, unrelated to sea-level rise but to other factor (e.g., human-induced factors, morphological variations, etc.), are highlighted in the "cleaned up" signal (i.e. without mean sea level) of long time series of sea-level observation considered in this application. As suggested by this Reviewer, we have added confidence intervals in Figure 2.

Section 3.2, lines 277-279. What does it mean that thresholds are selected "based on local tidal ranges"? This is a pivotal step of the study, please extend the explanation (you could also add it to the Methods section).

Tidal range is defined as the sea level elevation between high and low water level over a tidal cycle. Differences in tidal range are often important, and are related to variations in coastal processes and

morphology. Tidal range varies between locations and time scales according to the hydrodynamic response of a particular bay or estuary to astronomical tidal forcing. Since the choice of the threshold is site-specific, we must consider this parameter in its selection.
As highlighted from the Reviewer 1, we have moved the lines 276-277 to line 121. The revised manuscript now discusses here that: “The threshold is set to be large enough to filter out water level peaks that are likely to be associated to conditions without any storm contribution and sufficiently low to maximize the amount of information used. In addition to the above, we choose the threshold value that produces the minimum estimation error under the MEVD framework”.

Figures S2-S6. Please use a 1:1 axis ratio. This would help to assess the quality of the fit.
Thank you, agreed.

Figure 5. Levels in the return period plots refer to z or h? I find the terminology rather confusing throughout the whole manuscript, e.g. sometimes you talk about storm surge, some other time about extreme sea levels. Please be consistent.
We apologize for the lack of clarity. It is z. To avoid confusion, in Figure 5 we have replaced “water level” with “total water level” (i.e. the variable z). In the revised manuscript, we have revised all the notation and terms used.

Conclusions, line 349. Please specify that MEVD outperforms the other distributions for long enough calibration periods.
Thank you for the suggestion. We have revised the sentence accordingly.