Review of *Importance of non-stationary analysis for assessing extreme sea levels under sea level rise*

- Overview:

The manuscript compares the return levels obtained from four different extreme value analyses accounting for non-stationarity in the Punta Della Salute tide gauge, Venice. The extreme value distributions include: (i) generalized extreme value (GEV) applied to a block maxima sampling (BM); (ii) generalized Pareto distribution (GPD) and (iii) a point process (PP) method, both applied to peak over a threshold (POT); and (iv) a joint probability method (JPM). In addition, the authors tested the implications of using three different detrending techniques, including (i) removing the annual mean sea level from the time series (MSL); (ii) removing the last 19 years' mean sea level (MSL_L); and (iii) not detrending the data before fitting the distributions.

I find the topic important for coastal flood risk assessment as traditional designs have been based on analyses using direct methods that ignore the non-stationarities, which can lead to an underestimation of the risk, as found in previous studies. In addition, having a comprehensive analysis of the different non-stationary extreme value methodologies would help on the way to obtaining a more standardized analysis, facilitating the comparison of the results between the studies. Thus, the results of the manuscript are relevant for the scientific community and coastal risk stakeholders after improving the work in some aspects, mainly related to

- the level of comprehensiveness of the analysis: it will be interesting for the scientific community to include one of the most utilized indirect methods: the revised joint probability method. In doing so, the authors will also reduce the existing overlapping with previous studies,
- the level of applicability to other study areas by including more tide gauge records in the analysis, and
- the level of replicability (some information relevant for reproducibility is missed).

- General comments:

  1. The evaluation of return levels obtained from this set of methods in particular (GEV, GPD, PP, and JPM) has not been performed before, to the best of my knowledge. However, previous studies compared the implications

of using different extreme value distributions when accounting for non-stationarity (e.g., Razmi et al 2017; Menendez et al 2010). Thus, the paper will benefit from the analysis of a wider set of methodologies, particularly from including the revised joint probability method (RJPM), which is one of the most widely applied indirect methods for estimating non-stationary return levels.

In this line, some overlap also exists in the analysis of the effects of different detrending techniques in the estimation of return levels (Haigh et al 2010).

Finally, the uncertainties in estimating the parameters of the different extreme value distributions could be included in the analysis, as done in previous studies (e.g., Cheng et al 2014) to offer a more exhaustive analysis.

By doing all of the above, the paper will provide a richer and more novel analysis of the different non-stationary extreme value analyses, as well as a more complete evaluation of the uncertainties derived from each method.

2. As is, the manuscript conclusions are appropriate for the study case (Punta Della Salute) only. In order to provide a quasi-standardized method for non-stationary analysis, the paper will benefit from applying the analysis to a larger set of tide gauge records, so the authors will be able to assess whether the conclusions can be extrapolated to other areas of study. Likewise, readers will be able to decide which method should be used according to the conditions of each case: tide range, relative relevance of surge vs tide in extremes – as in Dixon et al 1999-, location, record length, etc. In this sense, Haigh et al 2010 showed that differences between return levels estimated using different methods highly depend on the record length and on the presence of outliers (when using direct methods).

3. In addition to the above comments, some concerns about the reproducibility of the work arise, including (i) the use of phrases such as "*We used our own script to generate the empirical sea level distribution function*" (line 151); (ii) the authors do not provide information on the additional tide gauges used to calculate mean sea level prior to 1924; (iii) they do not provide information on

how they calculated the uncertainties shown for the stationary analysis in 30-year time windows (Figure 3 and Figure 4).

- Specific questions

Introduction

- Line 58: How do you define suitability? What do you use to compare the different distributions and decide which of them performs best?

Methods

- Tide gauge: Is the sea level record at this particular tide gauge affected by the activation of the MOSE? Can that influence the measured extreme sea levels?

- Tide gauges used to calculate the mean sea level prior to 1924: (i) in order to ensure reproducibility, the authors should provide the name and location of those tide gauges. (ii) Are these extra tide gauges affected by the same subsidence rate as Punta Della Salute? As a double-check, did you compare the mean sea level of the tide gauges and the Punta Della Salute one during the overlapping period?

- Joint probability method: instead of using JPM, the authors might consider the revised JPM (RJPM) or the skew surge revised JPM (SSRJPM). These two methods allow the fitting of extreme value distribution (thus obtaining extreme probabilities) and apply an extreme index that accounts for dependence between extreme events.
Line 151: to ensure reproducibility you might want to clarify which kind of empirical distribution you have used instead of stating sentences such as "*we use our own script to generate the empirical sea level distribution*".

- Comparing different model configurations: Could you please expand the explanation of the meaning of M_0 and M_1?

Results

- Figure 3: I don't think you explained how you calculated the gray envelope in the text. The analysis will benefit from the comparison of uncertainties resulting from the different model configurations. Also, why isn't the scale parameter (yellow line) shown for GEV, GPD (MSL), and PP? doesn't the scale parameter vary with time (equation 7)? Same for the location parameter (green line) for PP (NDT, MSL, MSL_L)?

- References

Haigh I D, Nicholls R, Wells N (2010). *A comparison of the main methods for estimating probabilities of extreme still water levels*. Coastal Engineering, vol 57 (9).

Dixon M J, Tawn J A (1999). *The effect of non-stationarity on extreme sea-level estimation*. Appl Statist (48).

Cheng L, AghaKouchak A, Gilleland E, Katz R W (2014). *Non-stationary extreme value analysis in a changing climate.* Climate Change.

Menendez M, Woodworth P L (2010). *Changes in extreme high water levels based on a quasi-global tide-gauge data set.* Journal of Geophysical Research – Oceans.

Razmi A, Golian A, Zahmatkesh Z (2017). *Non-Stationary Frequency Analysis of Extreme Water Level: Application of Annual Maximum Series and Peak-over Threshold Approaches.* Water Resources Management (31).