## Overview:

I think this paper addresses an important issue for statistical modelling of extreme values and the characterization of coastal flooding hazard, which is to partial overcome the inevitable scarcity of data on extreme events and to account for outliers by incorporating historical information. Indeed, this paper shows how the use of historical information can improve the probabilistic and statistical modeling of extreme sea levels. This has been achieved by combining systematic (observed & artificial) ESLs with historical information using a bias reducing method based on a MC resampling method. I can see that this work could be used in many applications and bring improvement on the approaches already available for such applications.

On the whole, I felt able to follow the proposed method. I did, however, reach the end of subsection 3.2 and realize that I was uncertain about the used method of bias reduction (from the initial intermediate GPd to the last one) and how the first intermediate GPd can be used to continue the processes of bias reduction.

## General comments:

While a potentially useful method that could reduce uncertainties and in-crease the reliability of extremal estimates, there are, however, at least four major comments to which authors have to provide real and concrete answers, primarily to ensure that the usefulness of the authors' proposed method is conveyed at a standard that is consistent with that of the underlying concept:

1- I wonder if the proposed method provides a real improvement relative to the classic approach (the use of a threshold of perception method and the maximum of likelihood applied to both systematic and historical data, especially if we know that gathered hist information is exhaustive and represents the largest values during the historical period.

2- Regarding the first intermediate GPd, authors had to incorporate the historical information by developing a truncated GPd conditional to the largest historical data instead of simply substituting corresponding artificial events with known historical ESLs. Authors can refer to the censoring approaches proposed by Parent and Bernier (2003) (Parent, E., Bernier, J., 2003. Bayesian POT modeling for historical data. J. Hydrol. 274 (1–4), 95–108.)

3- The first "intermediate" GPd is not really intermediate. According to me, it is a distribution completely different from the initial one. A different object and cannot be used to continue the process of bias reduction used in this paper.

4- Several methods of bias reduction are presented in the literature. Authors are invited to give more details about the method of bias reduction used in this paper and give a quantitative evaluation of this bias.

## Specific comments:

### The issue regarding the duration of historical period:

Lines 135-138: “Prosdocimi (2018) notes that this issue is analogous to the common statistical problem of estimating the size of a population and compares several methods available in literature, including maximum likelihood, method of moments and maximum spacing”. The relationship between the maximum likelihood, method of moments and maximum spacing and the main issue regarding the duration of observation related to the incorporation of historical information with systematic data, is not clear for me. In addition, the authors sated in the following sentence (lines 138-139) that the same issue was considered by Engeland et al. (2018): “these methods as well as graphical and Bayesian concepts were also explored by Engeland et al. (2018) when considering flooding of Norwegian catchments”. This is not true because for some watersheds, Engeland et al. (2018) followed Prosdocimi (2017) and set the length of the historical period to be the time span from the fitst historical event to the end of the historical period plus the average time spacing between the historical events and another subjective method for the rest of watersheds.

### Systematic observations exceeding the threshold of perception

Lines 145-155 The authors stated this idea (including all the systematic observations higher than the threshold of perception in the historical information) was suggested in 2012 by the German Association for Water Management, Sewage and Waste (DWA). Please note that this is a very old recommendation of the United States Water Resources Council in their Bulletin 17B (USWRC, 1982) and was developed in the literature by many other scientists (Ouarda et al. 1998 among many others:

- T. B.M.J. Ouarda, P. F. Rasmussen, B. Bobée et J. Bernier (1998): Use of historical information in hydrologic frequency analysis Vol. 11, Journal of Water Science 41–49
- Ouarda, T., Hamdi, Y., Bobee, B., 2004. A general system for frequency estimation in hydrology (FRESH) with historical data. In: Benito, G., Thorndycraft, V.R. (Eds.), Systematic, Paleoflood and Historical Data for the Improvement of Flood Risk Estimation: Methodological Guidelines. CSIC, Madrid, pp. 55–70.
- Payrastre et al. (2011)…

### Methods for incorporating historical information (lines 145-173)

In line 145, authors talk about three methods suggested by the German Association for Water Management, Sewage and Waste (DWA) for the incorporation of historical

extremes with systematic observations (DWA, 2012). The two last methods (lines 155-173) were proposed to estimate the distribution functions parameters, are general and have nothing to do with the incorporation of historical information.

## Technical corrections (or clarifications):

- Journals names are missing in some references:
  Line 515 Haigh et al. (2014b)
  Line 515 Hamdi et al. (2015)
  Line 518 Hastings (1970)
  Line 523 Hinkel et al. (2015)
  Line 527 Jenson et al. (2008)
  And many others…

- Is there a need to detail the Gringorten plotting position formula?
- Authors don't need to detail the DWA methods. Don't need to present the equations since they are not used in the developed method.
- There is a repetition when talking about the benefits of hist information. Many things were repeated in Introduction/section 2.2.
- The choice of the POT threshold must be justified. Why the 98% extreme quantile and why the 1m?
- The choice of the 2 days period in the declustering method must be justified.
- Last paragraph of section 3.2: which type of bootstrap? parametric or non-parametric ? and the choice must be justified.
- Don't need to detail the BIC criteria (don't need to put the equation).