The paper “Bayesian extreme value analysis of extreme sea levels along the German Baltic coast using historical information” discusses the inclusion of historic extreme events, often disregarded as outliers, into extreme value analysis and adapts an established approach to the German Baltic coast. This new methodology is able to reduce uncertainties at some of the locations of the study and to shows that current design approaches might underestimate extreme sea levels significantly due to limited availability of long tide gauge records. The paper further discusses large scale variability found in the Travemünde record, which remains undefined but might be the reason for the underestimation of extreme sea levels in other records.

The authors present their research in a precise and easy to understand manner. Their work highlights the next step in possible improvements to the coastal protection in their study area and given a well-founded recommendation on the treatment of outliers in EVA.

I therefor recommend the publication after some minor revisions, mainly focused on some clarifications and extended discussions of its important findings.

General comments:

Data: The length of most of the hourly records does not seem to match the length of the GELSA3 records. Please provide the data sources used to extend the records.

Data preparation and filtering: A few short comments on the methods of data preparation and their potential influence on the results seem helpful for the reader. Specifically, these questions arise:

- While certainly the easiest methods of detrending, extending the linear trend into the first half of the 19th century (line 168) seems prone to overestimate sea level rise in extended timespan due to possible acceleration between the 19th and 20th century. This requires at least a short discussion of the underlying assumptions and their uncertainties and whether other option to fill this gap (e.g. a quadratic trend) were tested.
- The authors explain sufficiently, why an additional threshold for the selection of the historic events is necessary, but find that no systematic method was able to derive an appropriate threshold for all tide gauges. They therefor resort to a manual review of data. To understand the process of this selection, a description of the (possibly subjective) criteria for this review would be helpful to the reader.
- Furthermore, the consequences of this selection remain unclear until much later in the paper, where it is briefly mentioned that only the 1872 event remains at four stations after applying this threshold (line 355-357). This is important information for the reader and should be mention alongside the definition of the filter.
- In this context, it also is unclear which step of data preparation Fig. 2 depicts. Some parts of the data preparation (the extension of the AMAX timeseries) seems to have already been conducted, others (applying the threshold for historical data) not.

Results (ESL estimates): The results are presented in an overall clear and easy to understand fashion. Still, an addition to Fig. 4 would highlight the results of the study even more. I suggest to include the AMAX-GEV (Sys.+HI) estimate of Travemünde as well to show that the AMAX estimate benefits from the inclusion as well (even though the interpretation is limited by the availability of only one example). I further suggest adding a figure similar to Fig. 4, but depicting the HW1000 estimates, as supplementary material. These values are already given in Table 2, but I found Fig. 4 to enable a much easier comprehension of the findings. Therefor, the finding discussed in the text should also be appear in Fig. 4.

Results (ESL variability): The long-range dependence and importance of the 1872 event in sampling ESL data is already demonstrated sufficiently. While the assumption seems obvious to transfer this finding to ESL at the other tide gauges, the authors should comment on this assumption directly for example by determining the Hurst exponent for the Wismar and Warnemünde records as well to prove regional similarities or by referencing other sources discussing the variability.

Discussion: Even though EVA not considering historical data in not the focus of the paper, the discussion would benefit from an evaluation of both methods, whenever possible. At least for the Flensburg tide gauge it would be interesting to see, how an EVA of AMAX data, constructed from the hourly record, performs in comparison to the POT-GP method. This would enable more thorough comments on some aspects already mentioned in the paper:

- In the introduction it is mentioned, that the POT-GP method is generally preferred. Can this notion be confirmed for the German Baltic coast, if the length of the record is the same?
- More importantly, the importance of the inclusion of the 1872 in the AMAX record is shown during the analysis to ESL variability at Travemünde. An example of a tide gauge, where systematic records of this event are not available would further highlight this aspect.
- The inclusion of historic data is show to also improve the AMAX-GEV analysis at Travemünde, and improvements to the ESL estimate are found, due to the reduction of uncertainties. Analysis of a second tide gauge would give some indication, whether these improvements are similar at the other tide gauges, whether historic data improves POT-GP and AMAX-GEV analysis in a similar way at each tide gauge, and of the potential improvements, if additional historic data becomes available at the other tide gauges.

Discussion (1872 event): The authors briefly discuss the inclusion of the 1872 event and conclude, that this extreme event should be included in the analysis. This conclusion would strongly benefit from further evidence and discussion, for example by showing the estimated return frequency of the event with and without the inclusion of historical data. Further information should be given, which criteria were previously used to call the event an outlier. Are these criteria still valid or does the inclusion of historical data enable a new way of defining an outlier?

Specific comments:

Line 30-32: The mentioned paper of Weiss et al. (2014) uses the concept of homogenous regions for extreme wave height statistics in more offshore areas. Does literature on the feasibility of this method for estimating extreme still water levels exist? I suspect additional challenges posted by the strong influence of coastal bathymetry.

Line 40: Grammatical error and I suggest a more direct phrasing:

Despite this, <u>several statistical methods exist</u> to …

Line 62: Missing bracket after the reference to Fig. 1.

Line 72: … (<u>hereafter</u> referred as HW200) …

Line 73: … of past observations, ~~and~~ whose accuracy …

Lines 74-75: At least for Mecklenburg-Western Pomerania additional information is available (MLUV, 2012), which mentions the Gumbel distribution to generally result in the best fit. However, the reassessment of the described methodology conducted in 2021 does not yet seem to be published. Nonetheless, I suggest, extending the description of current design practices. Additionally, the

question arises, whether usage of the Gumbel distribution would be a logical addition to the analysis of this paper to enable the most direct comparison to current design practices.

Figure 2: Please provide a point of reference and a definition for the event's magnitude. I presume magnitude refers to the maximum height above MSL.

Figure 5: The description should include the name of the tide gauge to clarify its context.

Line 279: … series of AMAX ESLs at Travemünde …

Line 289-290: A question for clarity: does the historical data in this case only refer to historical data from literature or does it also include "historical" values from the AMAX record outside the 70-year window? As a small bump around 1943 is also visible in the "Sys+Hist" plot as well, I would suspect relatively large influence of the inclusion/exclusion of the 1872 event as well.

Line 387: … Wismar: 63~~y~~ years …

Line 388: I suggest using the abbreviation HW200 for consistency:

… estimates for HW200 increase …

References:

MLUV: Ministerium für Landwirtschaft, Umwelt und Verbraucherschutz Mecklenburg-Vorpommern: Regelwerk Küstenschutz Mecklenburg-Vorpommern - Bemessungswasserstand und Referenzhochwasserstand, Nr. 2-5, https://www.stalu-mv.de/serviceassistent/download?id=1634742, 2012