Review of A probabilistic view of extreme sea level events in the Baltic Sea by Kuosmanen and Hieronymus

General comments

In their article "A probabilistic view of extreme sea level events in the Baltic Sea" the authors describe two hierarchical spatial extreme value models in which the parameters of the Generalized extreme value (GEV) distribution are modeled using two variants of latent Gaussian process descriptions within the Bayesian modeling framework. They compare the results against a baseline model with standard maximum likelihood estimation and two simpler Bayesian GEV models over a set of tide gauges on the coastline of the Baltic Sea. Their results show that the spatial models outperform the simpler modeling approaches, model robustness is improved and uncertainty reduced both in the GEV parameter and return level estimates.

This is to my knowledge the first time that such hierarchical spatial GEV models have been applied to observed annual sea level maxima in the Baltic Sea region. The article is within the scope NHESS, albeit being mathematically rather heavy, and in principle does present some technical developments over previous studies. The topic of spatial extreme value modeling is in my opinion timely, and I appreciate the effort authors have put to this work, as is evident also from the rendered notebooks available as supplementary material.

Unfortunately, the manuscript is a bit unfinished and not mature enough for publication. It was difficult for me in many places to follow the line of thought of the authors, because of the poorly structured text and quality of figures. For example, the description of methodology contains repetition and is unnecessarily long for communicating the main points to the reader. I also have some methodological concerns, which I hope the authors will address in their reply. Lastly, I was not able to access the code or the supplementary material linked to the Code and data availability and Supplement -sections and therefore was not able to review them. Overall, the manuscript needs a major revision before it can be recommended for publication in NHESS.

In the following, I will first list my major comments regarding the manuscript. I hope they will be helpful for the authors when improving the manuscript. I will also provide a short list of more specific comments.

Major comments

Manuscript content

Firstly, please provide sufficient background information on earlier work. There are plethora of both observation- and physics-based extreme value analysis work done previously in the Baltic sea region. For example, extreme value analysis of sea level extremes based on hydrodynamical model simulations has not been mentioned at all

(e.g. Lorenz and Gräwe, 2023), although it provides a complementary approach to spatially infer statistics of sea level extremes.

Secondly, the description of methods is long and contains heavy mathematical jargon. I encourage the authors to concentrate on delivering the core message of theoretical aspects in the main text and provide technical details in the supplementary material or as an appendix, if needed. For example, the theoretical background of the block-maxima approach in page 3 is long-winded ($D(u_n)$ condition completely undefined!) and should be shortened. Also, lines 110-120 are an overly complex description of GEV distribution support and tail behavior and could be compressed to a couple of sentences. Yet another location for simplification is Sect. 3.3.3, where the description of Matérn covariance function properties could be mostly moved (at least Eqs. 39-41) to an appendix and only a brief description of its properties with the length scale values kept in the main text. Sections 3.1.1-3.1.3 could be combined and shortened by removing unnecessary detailed descriptions of GESLA-3 data (e.g. lines 234-236) and including only the description of the final training data set.

Regarding the Results section, I encourage the authors to expand Sects. 4.2.1 and 4.2.2 . Currently, they only contain illustrative examples for some subsets of tide gauges without justification why these subsets were chosen in the first hand. Furthermore, these individual examples do not provide a full model validation within the study domain. Therefore,

1. To better show the overall (miss)match and reduction in uncertainty between Baseline and the two spatial models in Sect. 2.4.1, I suggest that the authors provide only one illustrative example, drop the rest and quantitatively show the reduction in uncertainty (e.g. fractional uncertainty) with respect to Baseline on all tide gauges either as a table or a map for some (e.g. 50-year) return level. A similar domain-wide analysis on the GEV parameter accuracy with respect to Baseline could be included here as well.
2. Similarly in Sect. 4.2.2, I suggest that leave-one-out validation is extended to cover all tide gauges and the change in the predicted return levels is compared against the full models in terms of suitable statistics.

Figures are too small and require reformatting. Please, increase the figure size throughout the manuscript, remove both redundant texts (e.g., orange texts for Baseline in the posterior density plots) and duplicate legends. Also, please mask out the Norwegian Sea from the spatial maps, as this region does not belong to the model domain. One way to gain space for the figures would be to keep panels only for one spatial model in Sects. 4.2.1 and 4.2.2, as the results are very similar for both Hilbert and Latent. The other model could then be moved to the supplementary material.

Finally, to put the results into a wider context, I suggest that the authors compare them against previous similar studies in the study region, where possible. For example, Lorenz and Gräwe (2023) provides a useful point of reference for the results obtained in this study.

The structure of the text

I encourage the authors to make a complete overhaul of the manuscript in order to improve its readability:

1. Remove orphaned and one-sentence paragraphs (page 3 being a primary example of this) and be consistent throughout the text in this matter.
2. Put all figure descriptions to the figure captions and move any explanations of results to the main text (particularly Figs. 5-10).
3. Check and remove unnecessary repetition from the manuscript (e.g. lines 69-74, 114-120, 164-169, 192-198 and 245-250)
4. Check the text for fragmented and erroneous sentence structures such as lines 28-30 and for the incorrect and repetitive use of words such as "respectively".
5. Please, check for any other typographical errors (e.g. missing spaces after Eq., Sect. and Fig.) throughout the text.

Methodological concerns

It is unclear from the manuscript, how Latent is formulated. The model description is currently basically a copy-paste of lines 320-327. Please, explain what is the difference between Hilbert and Latent in Sect. 3.3.4. Also, please elaborate in the manuscript what is the motivation for including the Hilbert space approximation Gaussian process model. For example, does it bring computational benefit over Latent?

The authors model the mean function of Gaussian process (at least for Hilbert) as a linear function of longitude and latitude. I am a bit concerned that this might be a slight oversimplification, particularly outside data rich regions in the Baltic Sea, as the Geometry of the Baltic Sea is quite complex. For example, it is well known that sea level extremes tend to increase towards the end of Gulf of Finland, the Bothnian Bay and Gulf of Riga. Have the authors thought of using other explanatory covariates in their models?

Detailed comments

Title: This is too general in my opinion. Perhaps "A probabilistic view of spatial extreme sea level events in the Baltic Sea" would better reflect the content of the manuscript?

Line 76: There are also other earlier studies acknowledging this issue (e.g. Suursaar et al., 2002; Soomere et al., 2018)

Lines 317-320: "The coordinate transformation is due to the lack of information pertaining the coordinate systems for the different tide gauge stations" This sentence in unclear to me, could you elaborate?

Lines 338-339: What are the units of the length scales and how were the values selected?

Eq. 42: Should $Y(s_i)$ be $Z(s_i)$?

Table 2: Please, reduce the number of decimals

Line 554: the Zenodo link points to the supplementary material by Räty et al. (2023)!

References

Suursaar, Ü., Kullas, T., and Otsmann, M.: A model study of the sea level variations in the Gulf of Riga and the Väinameri Sea, Cont. Shelf Res., 22, 2001–2019, https://doi.org/10.1016/S0278-
4343(02)00046-8, 2002.

Soomere, T., Eelsalu, M., and Pindsoo, K.: Variations in parameters of extreme value distributions of water level along the eastern Baltic Sea coast, Estuar. Coast Shelf S., 215, 59–68, 2018

Lorenz, M. and Gräwe, U.: Uncertainties and discrepancies in the representation of recent storm surges in a non-tidal semi-enclosed basin: a hindcast ensemble for the Baltic Sea, Ocean Sci., 19, 1753–1771, https://doi.org/10.5194/os-19-1753-2023, 2023