Response to reviewer

Date: 15 July 2026

Manuscript title: **Global dataset of storm surges and extreme sea levels for 1950–2024 based on the ERA5 climate reanalysis**

We thank the reviewer for thoroughly considering the adjustments made to the manuscript in the first review round, and for suggesting further improvements towards the final publication. Below we address each comment of the reviewer (in blue) by providing a response and citing the changes to the manuscript text.

The author's have addressed all concerns raised during the revision, resulting in a more clear and streamlined paper, with only minor technical corrections remaining:

1. Resolution of all figures still need improving, particularly spatial plots (e.g. Figs 4,5) as I it became very blurry when zooming in to observe some global regions more closely.

   High-resolution figures are submitted to ESSD together with the final version of the manuscript.
2. I still find it a shame that more validation focusing on storm surges was not addressed, as the tidal signal (and its error) obscures some of the results in relation to the objectives of the paper. While I understand processing tide gauges to extract storm surges is time laborious, it had been done in previous validations of GTSM global reanalyses. In any case, I think a minimum effort could be done to add some storm surge validation, e.g. for the validated historical events, which really shouldn't require much effort. See my comments below.

   Figure 3: I would suggest removing the tidal signal and adding the observed surge signal instead, therefore validating waterlevels and storm-surges against observations for these selected events, and commenting it on the corresponding paragraph. This would require extracting the surge only for those locations and years.

   Thank you for suggesting this, I have implemented this comment by adjusting the Figure 3 and adding storm surge signal comparison, where the storm surge height was extracted from the tide gauge records (GESLA) using harmonic analysis. I have also included a comparison of peak surge height values.

   The following text has been added:

L224: “The storm surge signal was obtained from the tide gauge records by removing the tidal component using *hatyan* software (Veenstra, 2026).”

L233: “When comparing the modelled and measurement-derived surge timeseries, the peak surge height is underestimated for the North Sea storm (-9%) and typhoon Marie (-14%), but is overestimated (+14%) for Hurricane Carla.”

3. Lines 241-249: Here I would highlight both the general stability of the statistics (which gives confidence on the quality of the extended period), but notably also the differences we see. It seems you now mostly highlight the stability. Differences point to indeed some temporal variability in extremes, as you already mention, but I would more strongly highlight it as a topic of interest for the community and that your extended dataset offers a unique opportunity to explore (strong added value!).

   I agree with this comment, and have adjusted the text to emphasize the differences stronger. Next to climate variability, I now also mention the probable variation in the ERA5 extension quality in terms of representing the extreme storm events (this is also discussed in more detail in the Conclusions). The text was adjusted as follows:

   L246: “There are, however, several areas where larger differences are visible. In general, there are negative differences in high-latitude regions, such as Antarctica and Northern Europe, and positive differences difference for most of the rest of the globe. It is not clear what drives these differences in surge statistics, and the extended GTSM-ERA5 dataset provides an opportunity for further research to explore these differences across the full 75 year period. It is possible that some of these differences are caused by climate variability (Lobeto and Menendez, 2024). In regions with limited historical weather records, the differences can also be influenced by varying ability of ERA5 backward extension to represent historical extreme storm events. ”

4. Figure 7: I appreciate the author’s effort to include the observations. Unfortunately, high return period extremes (e.g. 100-year) for the plotted locations at which long term records are available don’t seem very sensitive to the original vs extended period (at least in terms of point estimates), so these results don’t illustrate the added value of the dataset very strongly. For me, the added value is a mix of 1) better estimation of point estimates, and return periods of contemporary extreme events ;2) reduction of confidence intervals; 3)Being able to do this globally as opposed to observations that are either

unavailable or too short. Here are some ideas that could help: maybe there are other locations among the 107 long-record stations that better showcase the added value for the computation of the point estimates?E.g. look for tide-gauge locations where the 100-yr return level is sensitive to the period used (from observed records), and show that the extended dataset indeed helps in these locations. And/Or maybe mix such plots for gauged locations with plots for ungauged locations where we also see a substantial impact of the extended period based on GTSM (e.g. as done in the first version of this plot, e.g. Buenos Aires), therefore highlighting the added value of the dataset in also enabling better extremes in non-observed regions.

I agree that some of the initially selected tide gauges in Figure 7 are not very informative for the reader. In order to address this comment, I have adjusted Figure 7 by 1) replacing the New York tide gauge by Montauk (USA, geographically close to New York, but with a more visible impact of the dataset extension on the point estimates for RP100); 2) replacing the Sydney tide gauge with the Cuxhaven (in Germany, with larger surges and also more visible change in the RP 100 estimate).

In the Conclusions section, I have now highlighted the added value of the dataset for EVA, while cautioning the dataset users that the simplified EVA presented in this manuscript is only intended for first-order assessments: L355: “The extended dataset can be used for a more robust estimation of extreme water levels, where the long record length can contribute to reducing confidence intervals and improving the accuracy of high return period point estimates. The extreme value statistics of total water levels calculated in this study are, this data is intended for first-order global overview of extremes.”

5. As a minimum, the authors could consider a somewhat more **elaborated description of the current plots in relation to observations (currently lines 299-306).** In some location we see the blue curve getting closer to the grey curve (Bergen, Hosojima, Freemantle). Or, in relation to the comment “the observed frequency of the individual extremes matches with the GTSM-ERA5305 datasets, which can be seen by comparing the blue and black dots in Figure 7”, we can see indeed that the extended dataset helps better identify return periods of specific events relative to the shorter dataset (e.g. in NY, the highest event on record for the shorter dataset -~2.8m – was being given a ~40yr RP, while for the longer period its ~70yr RP for both modelled and observed) – here the problem is that the **single-distribution fit is too stiff and is not being affected to this sensitivity in these very high events, but other kind of distributions would**. It

also showcases that the sensitivity and hence added value of the longer record on Figure 5b (impact on the 100-yr return level) might not be fully representative (and maybe underestimated), as it is limited by the **chosen extreme value distribution**. Your extended dataset enables other users to look at these kind of aspects more thoroughly.

Thank you for this comment, highlighting more clearly the challenges with comparing the EVA based on model results and observations. I have included the following changes to the manuscript to elaborate on the performance of the EVA fits and the comparison of event-specific return periods based on model and observations:

L296: "It must be noted that the quality of the extreme value analysis fits based on the peak-over-threshold method with fixed threshold selection varies from location to location. The chosen single-distribution fit is relatively stiff, and is only partially affected by the highest extremes in the data. In the examples illustrated in **Figure 7** this is particularly visible for locations affected by tropical cyclones, such as Montauk and Galveston."

L308: "the observed frequency of the individual extremes matches with the GTSM-ERA5-E dataset, which can be seen by comparing the blue and black dots in **Figure 7**. The extended dataset helps to better identify the return periods of specific events, compared to the original shorter dataset. For example, at the Esbjerg tide gauge, the most extreme event in the GTSM-ERA5 was initially given a return period of 40 years, while based on the longer record length it is assigned a return period of 70 years, both based on GTSM-ERA5-E and observations."

L365: "In future research, the use of more advanced and flexible extreme value analysis techniques could be applied to the entire global dataset in order utilize the longer record length towards obtaining more accurate estimates globally."