

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:

- 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.
- 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.
- 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!).
- 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. 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.