<u>General comments</u>

I would like to thank the author for the careful attention to all three reviewers' comments. This manuscript is greatly improved, and I think will be fit for publication after minor revisions. In particular, the paper now nicely articulates its purpose as a "proof of concept" and appropriately describes the advances achieved by the TMAX method. The author added detail to the methods so that they are clear and easy to follow, and they now provide a robust error analysis.

The following are three significant revisions I suggest prior to publication are. Note that all are described in further detail in the "line-by-line" comments.
- Reassessing the assumption that the selection of one event per tidal cycle is sufficient for declustering
- Providing background on UK water levels (tidal range, storm surge climatology)
- Extending the comparison of TMAX and SSJPM globally using results from Enriquez et al. (2022) (note: I do not think this is necessary for publication, but would strengthen the paper)

<u>Line by line comments</u>

+ *Introduction:* The final paragraph of the introduction nicely frames the purpose of this study as to deliver a "generally applicable," rather than maximally accurate, method and "to examine the proof of concept for the TMAX method, rather than being an attempt to redefine UK." However, the beginning of the introduction primarily describes the value of EWL distributions in terms of providing accurate design heights. What is the of generally applicable but not maximally accurate design water levels? For example, Reviewer 3 suggests "broad regional screening." I would also suggest finding examples of papers that use tide gauge-based EWL distributions to compare characteristics of flood hazard / drivers of flooding across seasons or large geographic areas.

Major comment
+*L174-176, 196-201:* The text says, "Furthermore, assuming the storm surge duration is less than that of the tide, it also prevents the selection of multiple peaks from within the same storm event. Therefore, in addition to providing a determination of relevant peak value, it is also a form of declustering (see below)." Is this a good assumption? If it is, please provide references or an analysis showing that typical surge durations are less than 1 tidal cycle. Batstone et al. (2013) applies a declustering algorithm (the extremal index), and other ETC-impacted regions I'm more familiar with experience multi-day surge durations.

+*L192-196:* There may be something I'm not considering here, but I don't think it's true that threshold selection in a POT approach "involves tidal range and datum." In practice, the threshold is often selected using a percentile that targets a certain average number of events per year and is then adjusted via sensitivity testing (e.g. Batstone et al., 2013; Baranes et al., 2020; Enriquez et al., 2022). I recommend removing this part of the text.

+L279: Can you please provide detail on how were gauge drifts and slips were assessed?

+General comment: The paper would benefit from text describing tidal range and storm surge climatology of the UK to give the reader a sense for 1) the magnitude of the errors relative to total surge and 2) an idea of whether the 1 tide cycle independence criteria is reasonable.

+Figure 3: For a broad audience, I do not think there is value in showing 100 year ESLs relative to ODN. For example, it's unclear whether spatial variation is driven by variations in tidal range, storm climatology, or geography. I would instead recommend showing TMAX2009-EA2011 in this figure, and I would use filled markers with a color scale in addition to writing the values on the map. You could also turn this into a multi-panel figure where additional panels in the same format show tidal range or the 100-year EWL **relative to MHHW** (rather than ODN).

+General comment: Why do you only assess through 2018 for the extended time series analysis? Given that this now excludes the most recent 7 years of data, it is worth a brief explanation.

+General comment: This is not, in my opinion, necessary for publication, but I believe the manuscript would be stronger if it could demonstrate broad geographic applicability via comparison to results in regions beyond the UK. The SSJPM results from Enriquez et al. (2022) would be a good option (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JC018157). They use the qn-SSJPM (Baranes et al., 2020) to provide annual exceedance curves, but they also calculate the full-record exceedance curve. They use the GESLA-3 dataset, but I do not believe they provide the actual EWLs for each location. If you would like to strengthen the paper with this comparison, perhaps the authors could be contacted to request their results.