Combined Point by Point Response

22 December 2025

S E Taylor

Please note that a list of relevant changes would be too complex as 75% pf the paper has been revised. However, both mark-up and non-markup versions of the paper have been uploaded.

It seems the upload will accept only one file for the point by point response. Accordingly, please find below my detailed responses to Reviewer 1, 2 and 3 merged into this single document.

I very much hope you find this acceptable.

Kind Regards

S E Taylor

Reviewer #1

Comment: Many thanks to the reviewer #1 for taking the time and effort to prepare their comments.

General comments

This study presents a new technique for extreme value analysis of coastal tide gauge records called the TMAX method. The TMAX method primarily improves upon the AMAX method of Gumbel & Lieblein (1954) by using a subset of the highest total water level values of each tidal day, rather than annual maxima. The author applies the TMAX method to the same 35 UK tide gauge records used by Batstone et al. (2013) for development of the skew surge joint probability method, and results are compared to extreme water level estimates derived from the SSJPM and the AMAX method of Gumbel & Lieblein (1954).

The author describes several valid advantages of this method, including: 1) the method is simpler than the SSJPM because it does not require harmonic analysis or fitting a probability distribution; and 2) using the highest water level of each tidal data, rather than annual maxima, enables more elegant treatment of incomplete time series.

However, I disagree with one of the author's motivations for developing the TMAX method. They write: "SSJPM is complex, often requiring manual intervention." It is true that the SSJPM is complex, but papers such as Batstone et al. (2013) and Baranes et al. (2020) clearly demonstrate that application of joint probability methods in regions where tides are large relative to surge (and thus large surge events may not be included in the highest recorded total water levels) provides more precise and stable return level estimates with a narrowed uncertainty range. It would be more appropriate to apply the TMAX method in an area with smaller tides – or, perhaps, to apply the TMAX method to calculating skew surge statistics, and then convolving the TMAX probabilities with the tide probability distribution. In terms of the manual intervention, I interpreted that part of Batstone et al. (2013) as finding that manual intervention with the SSJPM was necessary in two distinct geographic regions (the Severn and James estuaries). To me, this points to a geographically linked phenomenon as a potential challenge (such as nonlinear tide-surge interaction or river influence), rather than the statistical method itself.

Comment: Yes, JPM's use the data more effectively in this respect, responding to surges occurring away from high tide. But they suffer because residual and astro-tide are not uncorrelated. So, it does provide stable return level estimates, but only after very skilled intervention.

Comment: The revised paper shows that the TMAX results for Avonmouth on the Severn Estuary, with very large tidal range are adequate with differences of around 0.05m

I interpret the primary conclusions of this paper as the TMAX method 1) giving "a significantly better internal fit and reduced variance" compared to the AMAX method, and 2) being "at least as accurate as the AMAX and SSJPM methods." In this paper's current form, I do not think that these conclusions are sufficiently supported.

Comment: I have taken this on-board and removed the " exaggerated" claims from the paper.

- In general, there is little analysis provided of the results. In particular, it is important to discuss how geography, record length, and tidal range affect the results.
- I think "accuracy" is determined by consistency with the SSJPM, yet it seems that for 9 of the 35 locations, TMAX and SSJPM results are significantly different, even at the 20-year return period level. It is difficult to interpret this result without further discussion of factors such as record length and geography.

Comment: The revised paper contains more detailed statistical analysis.

Comment: Accuracy is not necessarily determined by consistency with Batstone's results. However the revised text shows the standard deviation and applies the t-test and discusses this point under 4.4 Error Analysis.

- It's a bit unclear how "internal fit" is compared, is it a comparison of the analyses that extend through 2009 to the ones that extend through 2018? If so…

  o On average, the TMAX-derived return water levels increase when the analysis is extended from 2009 to 2018 (Table 4). This is likely due to detrending the entire dataset using a constant assumed rate of RSLR that likely increased between 2009 and 2018. To actually assess something like stability (what I am interpreting "internal fit" to mean), it would make more sense to use something like a moving 365-day window to remove sea level rise and variability, then compare the results over the two time periods. Baranes et al. (2020) showed that application of a joint probability method significantly improved stability compared to fitting a probability distribution to total water level using a Monte Carlo validation method and by comparing statistical fits over two time periods. A similar sort of analysis could be used here to support the author's conclusion.

  o If I am interpreting "internal fit" correctly, the only comparison of the AMAX and TMAX internal fits are in Table 3, which only show mean values.

Additional comments:

1. Fitting a constant sea level rise rate for the "AMAX to 2009" and "TMAX to 2009" analysis may facilitate comparison with the results in Batstone et al. (2013), but it will yield less robust statistics compared to using something like a moving 365-day mean, as rates of RSLR are variable, and inter-annual sea level variability may significantly affect the results. This should be discussed in the manuscript.

2. The use of a tidal day, rather than 1 year, is described as a novel approach. However, once the tidal days are reduced to a subset of the largest tidal days, I interpret this method as becoming equivalent to what Batstone et al. (2013) does in applying a peaks-over-threshold approach to skew surge (although this paper uses total water level). In other words, selecting a number of peak tidal day water levels that is equivalent to 5 times the record length is roughly equivalent to using the top 0.7%. This is thus equivalent to the innovation of the peaks-over-threshold approach, and the novelty here is overstated.

3. In the specific comments below, I highlight several parts of the introduction with questionable descriptions of published studies, the methods description is confusing, or the results are inconsistent.

Comment: The phrase "internal fit" is not used in the revised manuscript.

Comment: The revised paper no longer discusses sea level rise, since thee topic of the application of ESL analysis to non stationary series is a separate and distracts from the main point to the paper, (which is to show that TMAX is an adequate method of analysis). SLR has therefore been removed, except insofar as detrending is necessary for comparison.

Comment: The revised paper now shows std deviation and in addition to mean differences.

Comment: Since the study is no longer an assessment of SLR this point can be the topic of another study.

Comment: The revised paper includes a comparison of AMAX and POT in Section 3.3.

Specific comments

Lines 24-26: While tide gauge-based extreme value analysis is valuable for many reasons, the spatially varying nature flood return levels that you mention (lines 24-25), along with the fact that tide gauges are generally purposefully installed in wave-sheltered locations, make it somewhat rare that they are the only tool used to determine design elevations for coastal defence structures. Perhaps you could modify the text to describe alternative applications, such as for determining boundary conditions and/or validating numerical models used for coastal planning.

Lines 29-31: This definition of HAT could use some clarifying. Do you mean that HAT assumes average conditions for the *meteorological component* of tidal height (or perhaps of total water level)? This might be clearer than calling the meteorological component of water level "noise."

Lines 42-43: I'm curious why this is being highlighted as a particular weakness of the JPM when none of the extreme value analysis methods you discuss in the introduction provide flood duration information.

Lines 45-46: This is not quite right. The timing of the actual astronomical tide is shifted compared to the predicted astronomical tide (and the predicted tide is what's used to calculate the non-tidal residual).

Lines 49-51: It's not that the "difference" is uncorrelated; it's that skew surge is uncorrelated with measured high water (see Williams, 2016).

Lines 51-53: The SSJPM fitting a GPD to skew surges is not a reflection of there being fewer skew surge values compared to non-tidal residual values for a time series of the same length. Fitting the GPD (or any extreme value distribution), as opposed to an empirical distribution, has the advantage of providing probabilities for values that exceed the maximum observed value. In fact, the Revised Joint Probability Method (Tawn & Vassie 1989; Tawn, 1992) made this improvement by fitting a GEV to the non-tidal residuals (rather than an empirical distribution).

Lines 69-71: I would recommend defining these terms earlier in the introduction and using one consistent term for measured minus predicted water level. I recommend "residual," rather than storm surge (because the residual is often not storm surge) or random noise (because there are deterministic components of the residual).

Lines 54-57: See general comments above about the Batstone et al. (2013) manual intervention

Lines 128-134: This paragraph is confusing in a couple of places:

- Lines 129-130: Do you mean the reverse? i.e. that the extreme values are more difficult to determine?
- Lines 81-83 (and Equation 1) show Gumbel ranking in ascending order, but you say "substituting for Gumbel's descending rank with an ascending rank"

Lines 209-214: Why not fit a spline to the hourly data? Or you could show that it's not important to do this by comparing high waters over time periods with 15-minute data to time periods with hourly data.

Table 3:

- Is the "mean difference" the mean of the difference between AMAX or TMAX and Batstone et al. (2013) across all 35 stations? Which is subtracted from which? And is the standard deviation the standard deviation of the difference?
- I would suggest showing these results on a map for each individual gauge – especially because Batstone et al. (2013) discusses individual sites where the GPD fit to the skew surge distribution was not physically plausible. Essentially, show the information in Table 5 on a map.
- Why don't the means and standard deviations in Table 5 match the mean differences in the "TMAX to 2009" row of Table 3?

Figure 4: This is difficult to interpret the way the bins are labelled and without geographic information. It should be shown on a map with exact values reported. The same should be done for AMAX (and AMAX should somehow be compared to TMAX) to support the conclusion that one method provides more stable estimates than the other (see also general comments above).

Results presented in Table 5: There are 9 sites that have differences greater than or equal to 10 cm at the 20-year return period level, compared to the SSJPM (Avonmouth, Dover, Hinkley Point, Immingham, Newlyn, North Shields, Port Ellen, Tobermory, Workington). This seems like a relatively large difference, but it is difficult to interpret without information on geography, record length, or tidal range

Conclusions: I interpret the primary conclusions as 1) the TMAX method giving "a significantly better internal fit and reduced variance" compared to the AMAX method, and 2) the TMAX method being "at least as accurate as the MAX and SSJPM methods." These conclusions should be described in an expanded and quantitative discussion section that points to the specific results and/or analyses that support the conclusions. Topics such as how geography, tidal range, and record length impact the results should be discussed.

Comment: The mean difference is the mean of the difference between the TMAX or AMAX output and EA2011. i.e. $D[i] = (TMAX2009.Port[i] - EA2011.Port[i])$
Mean $= xm = (1/n) \Sigma_i D[i]$
Stdev $= \sqrt{(1/n-1)\Sigma(xi-xm)^2}$

Comment: ESL 100 year results were previously and are shown on a map. Note the map shows the results using data to 2018, because it seems sensible to show the results from using the whole tidal data set.

Comment: Actually they did match in the paper. Bottom two rows of Table 5 did match the bottom row of Table3: TMAX to 2018. In the revised paper, the tables have a different layout which I hope you will find an improvement.

Comment: Agreed. They are now split into 4 histograms.

Comment: That number of outliers is about what you would expect for the level of scatter described by the standard deviation given.

Comment: Agreed. The claims was somewhat over hyped and are removed. In the revised version I claim that the 1) confidence limits are narrower for TMAX than AMAX and 2) The TMAX shows broad agreement with EA2011 (SSJPM).

Technical corrections

Lines 48-49: Perhaps revise to "… difference between the maxima of measured and predicted water level for each tidal cycle…"

Batstone et al. (2013) is sometimes referred to as "Batstone" and sometimes referred to as "Batstone 2013."

The equations used in the TMAX method should be more clearly and concisely stated.

Figure 5: I recommend not using red and green for colour-blindness

Comment: Agreed.. This wording has been revised accordingly in lines 60 to 65.

Comment: This reference has been rationalised, using EA2011 as far as possible.

Comment: More equations describing TMAX have been provided.

Comment: Agreed. Since the section on sea-level-rise has been removed this no longer applies as the red and green coloured dots are now removed.

Reviewer #2

This paper is another contribution based on the application of extreme value theory (EVT) to sea level extremes. EVT is a wide field and in recent times since the early 21st century has been widely used in environmental studies , especially for extreme events and any modifications have been suggested thereof. Similarly this paper touches on a specific aspect of the EVT and suggests some new developments to the Gumbel type I AMAX approach and looks at the highest tide each tidal day and verifies the approach with comparison against SSJPM.

The methodology and the results are sound. However a consideration of the following points may benefit the manuscript:

- The Most important point would be regarding a more detailed treatment of the comparison of the TMAX method with the r-largest method , which arises as a limiting distribution to the GEV distribution , see Coles (2021). The r-largest method is well established, theoretically sound, and implemented in R packages with maximum likelihood estimation. Many studies implementing the r - largest method usually stick to top 3/top 5 annual maxima sea level. Within the TMAX method as well the author infers that n=5 leads to minimum variance. So More detail is required on how the TMAX is an added value over the r-largest method. Following this the results derived from comparison with the SSJPM is definitely meaningful.
- Following the above point , the introduction could be improved to include more about the r-largest method and how the proposed TMAX method is an added value.

Comment: Many thanks to the reviewer for taking the time and effort to prepare their comments.

Comment: Actually the selection of largest peaks is applied to the entire dataset, so it resembles more the Peaks Over Threshold POT rather than r-largest. 5 per year is used to create a total number required by multiplying by the duration in years. Then this selections is easily applied using the rank. This is simpler than applying a threshold, which will be specific to tidal pattern and location.. It is also more sensible. Consider this example. Say there were 8 large significant peaks in one year and say only 2 in another year. Then in r-largest one would be selecting 5 from each year which is surely less efficient than TMAX which would select the largest 10 altogether, irrespective of year.

Comment: I have provided a more detailed discussion of r-largest, Peaks over Threshold, AMAX and TMAX. See Sect.3.3.

Reviewer #3

Referee Report

Evaluation of Extreme Sea-Levels and Flood Return Period using Tidal Day Maxima at Coastal Locations in the United Kingdom

General comments

The preprint proposes "TMAX," a workflow based on tidal-day maxima and linear regression on Gumbel reduced variates, positioned as an automated alternative to SSJPM and other established approaches. The topic sits squarely within the scope of *Ocean Science*, and the paper is motivated by an operational need for transparent and automatable estimation of return levels. In its current form, however, the scientific framing and validation are not yet sufficient for design-grade use: key assumptions (tail type, independence/declustering, stationarity) are not stress-tested; uncertainty is not fully quantified at the site level; and reproducibility is limited by brief methodological descriptions (peak selection, QC, and parameter tuning).

Alignment with OS evaluation aspects (brief assessment).

ˆ **Scientific significance.** Addresses a relevant question and aims to deliver operational simplification; the novelty is primarily in workflow/automation rather than in new statistical theory. Substantial impact would require stronger validation and generalization beyond the UK testbed.

ˆ **Scientific quality.** Core assumptions (Gumbel tail, independence of selected peaks, uniform de-trending) need diagnostics and comparisons (e.g., GEV/POT baselines). Results are compared against SSJPM but without per-site uncertainty bands or sensitivity analyses.

ˆ **Presentation quality.** Generally clear narrative, but several essential methodological details are under-specified (peak selection, independence checks, QC). Figures would benefit from confidence bands and clearer indication of the fitted range.

ˆ **Reproducibility/traceability.** The text should provide enough algorithmic detail (and QC rules) to enable replication; parameters selected by heuristics (e.g., number of peaks per year) should be justified or selected by diagnostics.

Comment: Many thanks to the reviewer for taking the time and effort to prepare their comments.

Comment: The paper was not intended to provide certification for "design-grade use" to a certifcation authority. It is a description of a modification of Gumbel's Method, to enable the reader to assess its suitability for conversion of tide data into ESLs. Improvements have been made in this revision to address the issues raised.

Comment: Agreed.

Comment: Per site confidence bounds have been added for TMAX2009 as requested. Further statistical analysis is provided in Tables 1-3 and Figure 4a,b,c,d. It makes the table somewhat squashed and figures have been reduced to 1 decimal place so it fits on A4 sheet. If the editor prefers it can be split into two tables each to 2 decimal places in metres.

Comment: An improved description of peak selection and confidence in mean of differences is now provided.

Comment: An improved description of the algorithm which is used for peak detection and declustering is provided in this revision.

Specific comments

S1. Automation as primary justification vs. reliability for design. The manuscript motivates TMAX mainly by its potential for full automation as an alternative to SSJPM, which is presented as difficult to operationalize. If design decisions based on return levels may have severe consequences, methodological reliability should come first and automation second. Automation is valuable for broad regional screening, but for site-specific design one would typically prefer the best-supported method (SSJPM or POT/GEV with explicit diagnostics), even if harder to automate. Please either (i) provide stronger evidence that automated TMAX delivers design-grade performance (diagnostics, uncertainty, sensitivity, external validation) or (ii) clearly position TMAX as a rapid, approximate tool for comparative assessments rather than as a standalone basis for design.

S2. Geographical scope and transferability beyond the UK. The paper frames TMAX as broadly automatable, but all evidence and tuning are limited to United Kingdom gauges. Please clarify whether the method is intended as UK-specific or globally transferable. If transferable, add an external validation or stress test at a few non-UK sites spanning different tidal regimes (micro/meso/macro; mixed/diurnal/semidiurnal), sampling cadences (15-min vs. hourly), and surge climatologies. Otherwise, state clearly that conclusions are restricted to the UK, and discuss what would be required to extend elsewhere (re-tuning of n_A, declustering windows, trend handling, tail-shape checks).

S3. Assumption of Gumbel vs. GEV tail form (and relation to prior practice). By fitting a straight line on probability paper using the reduced variate, the analysis *assumes* a Gumbel (Type I) tail. The manuscript should test whether a GEV model with freely estimated shape ξ provides a better description (Gumbel, Fréchet, or Weibull). A POT (Pareto–Poisson/GPD) baseline is asymptotically consistent with GEV and would allow estimation of ξ. From a physical perspective, after de-trending, extreme still-water levels are plausibly bounded, which argues for exploring Weibull-type behavior ($\xi < 0$) and could explain curvature in low-rank portions of the plots. The paper contrasts with work where heavy tails were mitigated via spatial averaging; here the tail difficulty is sidestepped by *assuming $\xi = 0$*. Please justify this choice with tail diagnostics (GEV and POT/GPD fits, QQ/PP/return-level plots) or relax the assumption.

S4. Peak identification and independence of events. The peak-picking algorithm ("scan above MSL, store the largest value before returning below MSL") and the selection of the top $n = n_A n_y$ peaks are described briefly, but independence (declustering) is not demonstrated. Please report extremal index estimates and/or runs-declustering sensitivity, and show ACF/dispersion checks to confirm that the retained peaks are approximately independent across sites.

S5. Choice of sample vs. threshold modeling (link to *r*-largest/POT and threshold selection). Selecting the "top n_A per year" resembles *r*-largest approaches, and the choice of n_A plays a role analogous to *threshold selection* in POT. A single global variance minimum is unlikely to be optimal everywhere. Please provide sitewise sensitivity of return levels (with uncertainty) to $n_A \in \{3, \dots, 10\}$, and outline diagnostics for automatic selection grounded in stability/predictive criteria (in line with established threshold-selection practice).

Comment: The journal is not an application to a certification authority. It is a means for presenting work which may be of interest to the wider scientific community.

Comment: The author has not come across any suitable data sources and ESLs to extend the study to the whole world. Remember that we would need both the tide gauge data and the ESLs from respected organizations in each country for a valid comparison. Therefore the study is limited to a comparison study at UK tidal locations.

Comment: The paper is a description of a modification of Gumbel's Type 1 method which is the assumed distribution applying to higher tidal levels. Reasonable sd / confidence limits are derived which relax it - the results are provided given the assumption.

Comment: The revised paper now describes further detail of why the algorithm acts as a peak selector and also how it provides a degree of declustering.

Comment: Agreed - there are similarities with r-largest and POT. The revised paper now discusses these similarities in more detail.

S6. Stationarity and de-trending. All series are de-trended using a fixed 2 mm yr-1 with a specified origin date. This uniform correction may not match local relative sea-level change, and no diagnostic is provided for residual nonstationarity in extremes. Please add simple diagnostics (moving-window location/scale of daily maxima, change-point tests) or adopt a basic nonstationary model if indicated.

S7. Uncertainty quantification for return levels. Formulas for σ_m and σ_p are given, but site-level return curves lack confidence or prediction bands. Please add per-site intervals for key return periods (e.g., 20, 100, 200 years) and discuss extrapolation uncertainty for longer periods.

S8. Data curation and sampling resolution. Combining 15-min and hourly records is pragmatic, but potential peak under-sampling is asserted rather than quantified, and the QC procedure is described as manual. Please specify reproducible QC criteria and include a down-sampling experiment (e.g., degrade the 15-min periods to hourly and re-estimate return levels) to demonstrate that conclusions based on hourly data are consistent with those from 15-min data.

S9. Relationship to JPM/SSJPM and tide–surge dependence. TMAX sidesteps explicit modeling of conditional dependence between surge and tide. Please show whether selected peaks cluster near springs and whether dependence structures vary by site. If dependence persists after conditioning, consider incorporating it (e.g., via covariates in exceedance rate/parameters or via an extremal index correction).

Document End SET 22/12/2025

Comment: A detrend of 2mm per year was carried out for comparison with EA2011. Discussion of the use of the TMAX or AMAX for sea-level-rise has been removed from the revised paper as it is a distraction from this comparison. Hence SLR is not discussed except in so far as it affects comparison with Batsone's results.

Comment: Per site confidence intervals as SD are now provided. For 95% bound are approximately double these values.

Comment: A re-sampling experiment showed a change of ~1cm in 200 year ESL and 95% confidence bounds by ~2mm.

Comment: The reader can judge the effectiveness of TMAX from its results. The reviewer is suggesting that the assumption of a Gumbel Type 1 statistic is inappropriate. However most of the "plots" indicate a good fit to a straight line, as is evidenced by the low values of variance of mean and variance of new point and the reasonable values of SD of mean difference between TMAX results and those of EA2011/Batstone. Critiques of the TMAX method should therefore make their own argument that Gumbel Type 1 statistic is inappropriate.