MS No.: egusphere-2026-1243

Response to Reviewers' Comments

Dear Reviewer,

We appreciate your time and effort in providing constructive feedback on our manuscript, "*Improved Estimation of Extreme Sea Levels via Non-asymptotic Statistical Methods for Coastal Hazard Assessment*." We address your comments here. The corresponding revisions have been implemented in the manuscript, and the revised manuscript will be submitted later as per instructions by the editorial office.

Note: Reviewer's comments are shown in black, while our responses, discussions, and proposed modifications are presented in blue. In certain responses, we have included context in the form of sentences from the original manuscript, for reference. These unchanged sentences are shown in red. The line numbers refer to the original manuscript.

The paper "Improved Estimation of Extreme Sea Levels via Non-asymptotic Statistical Methods for Coastal Hazard Assessment" by Sithara et al. presents a comparative analysis of four different approaches aimed at extrapolating extreme sea levels from gauges located in the Mediterranean Sea and the Atlantic Ocean. The paper is clear and well written, and the figures support the narrative effectively. The topic is well suited to NHESS; therefore, I am positive about its publication. There are, however, some aspects that I recommend addressing before the paper is published.

Comment 1: First, and most importantly, the authors refer to "new" declustering approaches (line 248) and "novel independent event selection methodologies" (line 434). However, the use of windows centred on the peak of an event has been adopted in a number of previous publications. Regarding the sorting method, I recall, for example, https://doi.org/10.1016/j.wace.2024.100701 and https://doi.org/10.1002/2016WR019426, among many others. It would perhaps be more accurate to state that such a selection procedure is applied for the first time in conjunction with the MEVD (if this is indeed the case), since the data extraction method is not new per se.

Reply 1: Thank you for pointing this out. We agree that our method presents similarities with the literature, in particular, the references pointed out by this reviewer. However, differences remain in how independent events are identified, and in how they are used in conjunction with MEVD. We have now modified the text in the Methodology section (section 3.4, after line 247) to highlight similarities and departures:

"The present study explored two declustering approaches for selecting IEs based on appropriate SLT and TWW values. *While conceptually similar approaches have been used in POT frameworks (Solari et al., 2017; Martín et al., 2024), the use of threshold optimization methodologies to identify independent events in non-asymptotic extreme-value modeling methods, such as the MEVD, is here explored for the first time."*

Comment 2: Second, a linear trend is used to detrend the time series (line 140). Is this assumption reliable? At some locations, acceleration effects may lead to exponential growth. Please comment on this aspect.

Reply 2: We agree that sea level rise may exhibit nonlinear behavior at global scales and over several decades. In fact, accelerations in sea level rise rates occur over long periods of time and require long records to be detected. In practice, observations that are longer than the lunar nodal cycle of about 19 years are needed to identify nonlinear changes in sea-level rise rates, to eliminate the effects of incomplete tidal periodicities. Many of the available observations cover just about 15 years: in such short records, low-frequency variability associated with tidal and coupled atmospheric-oceanic processes may dominate the signal, leading to unreliable estimates of higher-order trend components. Hence, we think that a linear detrending approach, widely used in extreme sea-level studies, is more reliable in removing the mean sea-level rise signal, while avoiding the introduction of a spurious signal. The short record length and the dominance of large, high-frequency, storm-driven variability also ensure that possible inaccuracies in the removal of mean sea-level changes, likely of the order of centimeters (out of a total sea level change of the order of 45 mm in 15 years), are small with respect to extreme sea-level events. Hence, the specific long-term trend removal method adopted does not significantly impact the results of extreme value analyses.

We modified the text in section 3.1 (after line 138) to improve clarity.

"Because sea level records at numerous tidal gauge stations were short, the detrended sea level here is obtained by simply removing a linear trend computed from the entire observed time series, as higher-order nonlinear trends cannot be reliably identified from such short records."

Comment 3: Finally, the core of the research is the identification of the best TSL–TWW combination based on the alignment between observed and predicted quantiles. However, Figure 3 suggests that different combinations yield very similar errors. Since the subsequent comparisons focus on the best-performing setting, this raises the question of whether a meaningful improvement is actually achieved. Perhaps colour bars with fewer classes would improve clarity, and this aspect should be discussed in greater detail. In addition, Figure 4 appears to show that the data are characterised by multiple ties. Does this introduce any bias in parameter estimation?

Reply 3: We agree that several SLT–TWW combinations produce RMSE* values that are close to the minimum. In this study, the optimal combination was defined as the one yielding the lowest RMSE*, as a standard and objective selection criterion.

Prompted by this and the other Reviewer's comment we have verified that the proposed methodology, using the minimum RMSE* values to select the optimal SLT–TWW combination, produces robust results and does not significantly influence the estimated return periods.

The following text has been added to the revised manuscript (after line 345) to illustrate the sensitivity analysis:

"These findings indicate that the selected SLT and TWW values exhibit variability across different stations and methodologies, underscoring their dependence on local conditions. The longer TWW values identified at some stations may be associated with the persistence of storm surges over several days and the occurrence of multiple exceedances associated with the same storm event. The selected TWW values may be analyzed in relation to the duration and clustering of meteorological and tidal forcings, as a lens to better understand the spatially varying drivers of storm surge extremes. An analysis that is left for subsequent developments and is not pursued further here. A sensitivity analysis was carried out by repeating the 1000 cross-validation analyses to obtain 20 independent realizations of extreme storm surge estimates for the four representative stations, to assess the sensitivity of the results to different optimal SLT–TWW combinations selected in different realizations. The results showed that the optimal SLT remained nearly unchanged across all methods except for the MEVD-sorting approach, where the largest difference

(approximately 35 cm) was observed at Dunkerque, representing about 6% of the local spring tidal range (~5.5 m). The optimal TWW exhibited variability across the 20 independent realizations for the four representative stations, with coefficients of variation (CVs) calculated from the selected TWW values ranging from 0 to approximately 60%, depending on the station and approach. Furthermore, the variability in estimated extreme sea-level quantiles for a 100-year return period was assessed by computing the CV from the 20 quantile estimates obtained from individual realizations. The results revealed low variability (CV ≤ 5.47%) for all four estimation approaches, indicating robustness with respect to the optimal SLT-TWW selection."

The apparent multiple ties in Figure 4 result from plotting the quantiles obtained from the 1000 cross-validation iterations on a single Q–Q plot. These repeated values reflect the resampling-based cross-validation procedure and empirical quantile estimation rather than ties in the original sea-level observations. Parameter estimation is performed independently within each cross-validation iteration before the Q–Q plots are constructed and is therefore not affected by these repeated values.

Also based on the above discussion, we decided to retain a common color scale for all four plots based on the overall minimum and maximum RMSE* values across the four methods to enable consistent visual comparison.

Minor comments:

In the abstract, line 8 ("This study" to "coastline") would be better placed later in the text. In the second line, it appears somewhat abruptly.

Reply: We moved this sentence to the end of the Introduction (after line 75), where additional context for the present study was added in response to Minor Comment 4.

I am not a native speaker, but I believe that "Threshold Sea Level" should be changed to "Sea Level Threshold". Consider making this change throughout the manuscript.

Reply: We agree. Thank you for catching this. This change has been made throughout the manuscript.

Line 53. Up to this point, the manuscript has referred to POT-GEV (see line 41); therefore, the GPD should be introduced and contextualised more clearly.

Reply: We added the following sentence in the revised manuscript (after line 49) for clarity.

"A significant challenge when using the traditional EVT-GEV methodology in investigating extreme sea levels is the scarcity of observations, both in terms of quantity and quality (Bernier et al., 2024; Fouad et al., 2025). Under these circumstances, the BM-GEV approach is critically inappropriate because it discards most of the already scarce available information and only makes use of block (typically yearly) maxima, resulting in an inefficient use of the data (Coles and Tawn, 2005; Haigh et al., 2010). The POT-GEV framework exploits information somewhat more effectively than the BM-GEV approach (Coles, 2001). Conveniently, when more than just a single yearly event is selected using a sufficiently large threshold, the distribution of the exceedances above the threshold converges to a Generalized Pareto Distribution (GPD) and is hence termed POT-GPD (Davison and Smith, 1990)."

More context about the present study should be provided in the Introduction, particularly from line 72 ("Here we study"). At a minimum, please introduce the dataset, even if it is described in detail in Section 2.

Reply: The following description has been added in the revised manuscript (after line 75)

"Here we study the performance of non-asymptotic (MEVD) approaches, relative to asymptotic GEV results, while expanding the MEVD application by integrating two novel independent event (IE) selection methodologies. The analysis uses a cross-validation method to quantify the predictive estimation uncertainty associated with different non-asymptotic/asymptotic methods and different independent event separation methodologies. To achieve these objectives, we analyzed hourly sea level observations from 39 tidal gauge stations distributed along the Mediterranean and Southern European Atlantic coastlines. These stations encompass diverse tidal and meteorological regimes, providing a suitable dataset for evaluating extreme sea-level models. The datasets analyzed were either from various national agencies or the Global Extreme Sea Level Analysis (GESLA) dataset. The shortest time series in the dataset analyzed here is 15 years long (2010-2024, Alcudia), while the longest record is 118 years long (1905-2022, Trieste). Using these observations, we identified the optimal SLT and TWW values for each station and assessed the predictive uncertainty of the statistical models via a cross-validation procedure involving independent calibration and validation samples."

Line 81. Add a reference to Figure 1 alongside Table 1.

Reply: We added the reference, as suggested. Thank you.

In Section 2, thirteen lines are devoted to introducing the Mediterranean basin, whereas only two lines describe the Atlantic coast. The discussion of the two study areas should be more balanced.

Reply: We agree and have modified the description in Section 2 as follows (after line 94):

"Hundreds of thousands of people live along the Atlantic coastline in Southern Europe, which hosts several high-value economic activities and extensive infrastructures that must be adapted to intensifying natural coastal hazards, including marine flooding (Breilh et al., 2014; Antunes et al., 2019). "The southern European Atlantic coast is highly exposed to storm surges and large tidal ranges. Previous studies have shown that this coastline experiences some of the most extreme sea levels in southern Europe, driven by the combined influence of astronomical tides and intense meteorological forcing associated with Atlantic storm systems (e.g., Marcos et al., 2009). These factors lead to substantial spatial variability in extreme sea levels and frequent occurrences of coastal flooding events."

"There is considerable variability in observed extreme sea levels, with values along the Atlantic coastline exceeding 250 cm above datum, while those in the Mediterranean are on average around 50–60 cm above the local datum, with high values of the order of 150-160 cm at numerous locations (e.g., Venice). Moving from the Bay of Biscay to Portugal and the southern coasts of Spain, the maximum observed values decrease by almost a meter. This spatial distribution is primarily driven by different tidal regimes (Tsimplis and Shaw, 2010). "Higher water levels for a given return period (200–300 cm for the 50-year return level) are observed along the Atlantic coasts due to the larger tides (Marcos et al., 2009)."

Line 250. Complete the sentence (the verb is missing).

Reply: Thank you, the sentence was updated as follows:

"First, all the down-crossings, i.e., intersections between the SLT and the detrended sea level line when the latter is characterized by a negative slope, are identified."

Line 324. Please explain the NSE index.

Reply: The following explanation has been added in the revised manuscript (after line 324).

"NSE measures the predictive performance of a model by comparing the variance of prediction errors with the variability of the process as expressed by the variance of the observations. NSE values range from -∞ to 1, with 1 indicating perfect agreement, 0 indicating a performance that is equivalent to just using the observed mean as a prediction, and negative values indicating an inferior performance with respect to this reference."

References

Marcos, M., Tsimplis, M. N., and Shaw, A. G. P.: Sea level extremes in southern Europe, J. Geophys. Res. Oceans, 114, https://doi.org/10.1029/2008JC004912, 2009.

Martín, A., Wahl, T., Enriquez, A. R., and Jane, R.: Storm surge time series de-clustering using correlation analysis, Weather Clim. Extrem., 45, https://doi.org/10.1016/j.wace.2024.100701, 2024.

Solari, S., Egüen, M., Polo, M. J., and Losada, M. A.: Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value, Water Resour. Res., 53, 2833–2849, https://doi.org/10.1002/2016WR019426, 2017.