MS No.: egusphere-2026-1243

Response to Reviewer's 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: Following Reviewer 2's suggestion, the term **Threshold Sea Level (TSL)** was changed throughout the manuscript to **Sea Level Threshold (SLT)**. For consistency, the revised terminology (SLT) is used in the present response to Reviewer 1. Reviewer's comments are shown in black, while our responses, discussions, and proposed modifications are presented in blue. In certain responses, we have included a few surrounding sentences from the original manuscript to provide context for the proposed revisions. These unchanged sentences are shown in red. The line numbers refer to the original manuscript.

In this paper, titled "Improved Estimation of Extreme Sea Levels via Non-asymptotic Statistical Methods for Coastal Hazard Assessment", the authors performed a comparative assessment of extreme value analysis approaches to evaluate the likelihood of coastal extreme sea levels. The paper is well written and well structured, and the results are clearly presented, with figures and tables being all relevant. Despite this, the manuscript presents some points that need to be clarified before being considered for publication. In what follows, some comments and suggestions are given.

Comment 1: My first concern relates to the lack of a clear description of the cross-validation procedure used to select the optimal TSL and TWW, as well as to evaluate the performance of the MEVD and POT-GPD approaches. I strongly recommend clarifying which variables are used to compute the performance indicators (e.g., those presented in Figures 3 and 4).

Reply 1: Agreed, the revised manuscript now more clearly describes the cross-validation procedure used to select optimal SLT and TWW, and the performance evaluation of the MEVD and POT-GPD approaches.

The description of the methodology, Section 3.5.1 (after line 313), was modified as follows:

"A value of the Root Mean Square Error (RMSE) was computed using the extreme sea level quantile estimates from the test sample for RP values less than or equal to the length of the calibration sample (here, 4 years), for all 1000 randomly generated cross validation-test pairings."*

We also modified Section 3.5.2 (after line 323) as follows:

"The performance of the POT-GPD and MEVD methods was assessed graphically and via statistical metrics. Quantile-Quantile (Q-Q) plots were used to visually assess the agreement between observed and model-predicted extreme sea-level estimates. The slope of the best-fit line through the origin between the observed and model-predicted quantiles in the test sample (here, 1000 test samples; all plotting together) was used as a measure of agreement. The RMSE and Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) were computed across all extreme sea level quantiles in the 1000 cross-validation test samples. 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. Additionally, a Quantile Ratio ($QR = \frac{q_{predicted}}{q_{observed}}$) was calculated on the highest quantiles in the test samples to quantify the uncertainty in estimating the magnitude of longer RP events."

The following revisions have been incorporated into the second (after line 337) and third (after line 356) paragraphs of the Results and Discussion section.

"Among the 39 stations, results for four stations are presented here for detailed illustration: Venice-AAOT, Sete, Dunkerque, and Ponta Delgada. The four stations were selected as representative examples of the diverse tidal and storm regimes within the study domain. Together, these stations encompass a broad range of tidal amplitudes, storm climatologies, and sea-level variability: (i) Venice-AAOT is located in a shallow Adriatic Sea area dominated by storm surges from sirocco winds, low tidal range, but compound tide-surge interactions, (ii) Sete is in the western Mediterranean sea, with a microtidal regime, affected by Mediterranean cyclones, (iii) Dunkerque is characterized by a large tidal range and a strong North Atlantic storm influence, (iv) Ponta Delgada is an open-ocean Atlantic island, exposed to North Atlantic swells and cyclones, with a relatively low tidal range. Figure 3 depicts the best SLT-TWW combinations selected for these stations, based on the cross-validation results. Extreme sea-level quantiles for RP less than or equal to the size of the calibration sample (here, 4 years) are estimated from calibrations on 1000 randomly selected training subsets. The deviations between the model-estimated quantiles and the corresponding observed quantiles of the test sample are then used to compute the RMSE values for each of the four methodological approaches being compared at each station."*

"Figure 4 shows the Q-Q plots of the extreme sea level events based on the optimal SLT-TWW selection for all methods for Venice-AAOT, Sete, Dunkerque, and Ponta Delgada stations, as illustrative and representative examples. Quantiles predicted by the MEVD approach match observed ones well, as evidenced by a good alignment along the 1:1 line in the Q-Q plots. This alignment appears tighter for MEVD predictions than for POT-GPD-predicted quantiles, both for high and low values of extreme surges (Fig. 4, panels C and D). Similar results were observed for other stations. Additionally, for 33 out of 39 stations, the slope of the least-squares regression line through the origin is closer to 1 for the MEVD methods than for the POT approaches."

Comment 2: Secondly, the authors should discuss in greater detail the variability and uncertainty associated with the selection of the optimal TSL-TWW combination. According to the results presented in Figure 3, several TSL-TWW combinations - even those with substantially different values - yield very similar RMSE*. This raises the question of how sensitive the return period estimation is to the choice of TSL-TWW combination.

Reply 2: 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. However, prompted by this comment we have verified that 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."

Comment 3: Thirdly, the authors should more clearly link the results regarding the selection of independent events to the underlying physical processes governing extreme sea levels at the coast, namely tides and storm surges. In particular, the choice of an appropriate TWW should be discussed in relation to the dynamics of the meteorological drivers (primarily extratropical cyclones) that control extreme sea levels in the Atlantic and the Mediterranean Sea. In this context, I suggest that the authors assess whether the use of relatively large TWW values (8-10 days) may inadvertently mask independent events.

Reply 3: As the Reviewer points out, the selected independent events correspond to extreme sea-level occurrences driven by the combined action of astronomical tides and storm surges. The lower and upper limits adopted for TWW selection (2–10 days) are indeed justified and discussed in relation to known physical tidal and meteorological time scales. As discussed in the manuscript, a minimum of 2 days ensures separation beyond tidal-cycle variability in both Atlantic and Mediterranean regimes, while larger values (8–10 days) reflect the longest storm surge events, specifically the time scales associated with extratropical cyclones and Mediterranean storm systems. Beyond this physical justification of the approach, however, a systematic analysis, at each tide gauge, of the specific physical drivers that may lead to the determination of optimal threshold values, would require an extensive exploration of dominant atmospheric conditions corresponding to each event, which would require a large amount of time. The results of such an analysis would take the manuscript in a significantly different direction and, we believe, would be beyond the objective of developing the novel statistical extreme storm-surge analysis presented. We thus think it best not to include this additional analysis, which may be pursued subsequently by leveraging the methods developed in the present paper.

Regarding the use of relatively large TWW, we agree it may result in masking out a few independent events. However, our analysis shows that model predictions are less sensitive to the value of TWW compared with

SLT. Therefore, even if a few independent events may be masked out by choosing a large time window of 8-10 days, this will not significantly alter the results. We agree that this aspect was not sufficiently clear in the original manuscript, and we now include the following text (after lines 345 and 365) to clarify the issue.

"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."

"Further analyses were conducted to better understand of the variability in SLTs and TWWs. Figure 5 illustrates the violin plots displaying the variation of the Quantile Ratio (QR) evaluated for the highest quantile in cross-validation experiments, across different methods and for various SLT values, while maintaining a fixed TWW value of 4, 9, 6, and 3 days for the representative stations, respectively. QR values vary significantly as a function of the thresholds, underscoring the role of threshold selection in extreme value analysis. QR values for the highest quantile show some variability as a function of TWW (Fig. 6). At some sites, e.g., Dunkerque, there is more noticeable variation in QR across the methods. However, the relatively consistent median QR values across different TWW values suggest that model predictions are less influenced by TWW compared to SLT (Fig. 6). Therefore, although longer TWW values may potentially combine a few closely spaced events, this does not significantly influence the estimated extreme sea level quantiles."

Minor comments:

Line 7-8: The first sentence of the abstract is out of topic since the effects of climate change on extreme sea levels are not assessed.

Reply: Thank you for catching this. The sentence has been removed from the revised abstract.

Line 29: The reference to Favaretto et al. (2025), which refers to a local study, is somewhat forced in this international context. I suggest removing it.

Reply: We agree, the reference has been removed as suggested.

Lines 87-88: This sentence is not clear and probably not correct. The rate of sea level rise is quite homogeneous over the Mediterranean Sea.

Reply: The reviewer is correct. We should have used the term "relative sea-level rise rate" for the sentence to be correct. The sentence was modified for better clarity as follows:

Moreover, relative sea-level rise, which includes both changes in sea level and vertical land motion (e.g., land subsidence or uplift), varies considerably across the Mediterranean Sea.

Line 91: Do you mean that coastal flooding is exacerbated by SLR or that a significant change in storminess has been detected?

Reply: The original text was not sufficiently clear. We do not imply that changes in storminess have been detected in general. This is typically site-dependent. We have now more clearly limited our considerations to sea-level change:

Hence, coasts are highly exposed to extreme sea level events, whose impacts are amplified by rising mean sea level (Toomey et al., 2022), exacerbating coastal hazards such as flooding, shoreline retreat, and saltwater intrusion, and posing a significant threat to heavily populated cities and important infrastructure.

Line 97: "There is considerable variability in observed extreme sea levels, …".

Reply: The sentence has been modified, as suggested.

Lines 141-142: This is not correct as several authors (e.g., Arns et al., 2020; Ferrarin et al., 2022) detected a significant negative correlation between tidal and non-tidal components of the sea levels.

Reply: The sentence has been modified in the revised manuscript as follows

After removing this long-term trend from sea-level observations, it is possible to study the detrended sea level as the sum of storm-surge and tidal contributions.

Lines 256-257: This is not correct as the Mediterranean Sea is mainly characterized by a mixed semi-diurnal tidal regime (Ferrarin et al., 2018).

Reply: Thank you for pointing this out. The sentence has been corrected as follows:

The tidal regimes in the Mediterranean Sea and Southern European Atlantic Ocean are predominantly mixed semi-diurnal and semidiurnal, respectively (Dias et al., 2021; Fanjul et al., 1998; Ferrarin et al., 2018).

Lines 338-339: Please justify the choice of these four stations.

Reply: We agree and have now added the following explanation (after line 339):

The four stations were selected as representative examples of the diverse tidal and storm regimes within the study domain. Together, these stations encompass a broad range of tidal amplitudes, storm climatologies, and sea-level variability: (i) Venice-AAOT is located in a shallow Adriatic Sea area dominated by storm surges from sirocco winds, low tidal range, but compound tide-surge interactions, (ii) Sete is in the western Mediterranean sea, with a microtidal regime, affected by Mediterranean cyclones, (iii) Dunkerque is characterized by a large tidal range and a strong North Atlantic storm influence, (iv) Ponta Delgada is an open-ocean Atlantic island, exposed to North Atlantic swells and cyclones, with a relatively low tidal range.

Line 339: The correct name of the Venice-CNR station is Acqua Alta Oceanographic Station (AAOT).

Reply: Thank you for pointing this out. The name has been corrected throughout the manuscript.

Lines 409-411: The authors should provide a physical explanation of the estimation error variability over the Atlantic-Mediterranean study site.

Reply: We thank the reviewer for this valuable suggestion. We agree that the variability in estimation error is influenced by the physical processes governing extreme sea levels. In the revised manuscript, we have added a discussion providing the general physical context of the Atlantic and Mediterranean study areas.

The observed estimation-error variability, however, does not exhibit a clear basin-scale pattern, as stations with comparatively larger uncertainty occur in both the Atlantic and Mediterranean regions. This suggests that the estimation error is likely influenced by a combination of local factors, including tidal regime, storm characteristics, coastal geometry, and bathymetry, rather than by basin-scale differences alone. A detailed station-by-station attribution of the observed estimation-error variability to specific physical processes would require a dedicated investigation of local hydrodynamic and meteorological conditions, which is beyond the scope of the present study.

The following text has been added to the revised manuscript (after line 413).

The differences in estimation-error variability across the Atlantic–Mediterranean study area likely reflect regional variations in the physical processes governing extreme sea levels. Along the Southern European Atlantic coast, extreme sea levels are influenced by the interaction of semidiurnal tides with energetic storm surges and waves associated with North Atlantic storms. In contrast, much of the Mediterranean Sea is microtidal and semi-enclosed, with extreme sea levels primarily driven by meteorological forcing, except in regions such as the Northern Adriatic, where tidal and storm-surge influences are more pronounced. The observed estimation-error variability, however, does not show a clear basin-scale pattern, indicating that it is likely controlled by a combination of local tidal, meteorological, and coastal characteristics rather than by basin-scale differences alone.

References

Dias, J. M., Pereira, F., Picado, A., Lopes, C. L., Pinheiro, J. P., Lopes, S. M., and Pinho, P. G.: A comprehensive estuarine hydrodynamics-salinity study: Impact of morphologic changes on Ria de Aveiro (atlantic coast of Portugal), https://doi.org/10.3390/jmse9020234, 1 February 2021.

Fanjul, A. E., Gomez, B. P., Carretero, J. C., Rodriguez, I., and Arevalo, S.: Tide and surge dynamics along the Iberian Atlantic coast, Oceanologica Acta, 21, 131–143, 1998.

Ferrarin, C., Bellafiore, D., Sannino, G., Bajo, M., and Umgiesser, G.: Tidal dynamics in the inter-connected Mediterranean, Marmara, Black and Azov seas, Prog. Oceanogr., 161, 102–115, https://doi.org/10.1016/j.pocean.2018.02.006, 2018.