Reply to the Reviewers

Re: Manuscript ID Preprint egusphere-2025-2998

"Assessing extreme total water levels across Europe for large-scale coastal flood analysis"

Camila Cotrim, Alexandra Toimil, Iñigo Losada, Melisa Menéndez, and Hector Lobeto

Natural Hazards and Earth System Sciences – NHESS

We would like to thank the Reviewers for their careful revision of our manuscript and for the thoughtful and constructive comments provided. We appreciate the time and effort dedicated to the review process, and we believe the suggestions have significantly improved the quality and clarity of the study. Below, we provide a list of main relevant changes made to the manuscript and detailed, point-by-point response to each comment, outlining the corresponding revisions made in the manuscript

Relevant changes made to the manuscript

Revised structure of the manuscript:

- Introduction has been reviewed, parts of the results have been moved to the Discussion, the main limitations of the study have been included in the Discussion, and a Conclusions section was included.

Focus on the validation of the databases used and the modeling chain:

- A section summarizing the validations performed has been included to the manuscript and more details have been included in the Supplementary Material.

Response to Reviewer 1

Review Report

The manuscript presents results from developing and analyzing a high resolution hindcast of total water levels for Europe. The authors address the important topic of including wave contributions into TWL calculations and subsequent flood impact/risk analysis. They use different approaches to derive wave setup estimates and quantify the effect of the methodological choice on the resulting flood extent. They also test the effect of using different thresholds in the extreme value analysis that is applied to derive the 100-year TWL levels, which are then used for the static flood mapping. The manuscript is very well written and I commend the authors for the comprehensive analysis. I have few general comments and some minor ones listed below which I believe should be addressed to further improve the quality of the paper.

General comments:

1. The authors emphasize that their new hindcast has an unprecedented spatial resolution of 1km. This is great! However, it is never shown how going from a coarser resolution (for which other hindcasts already exist) to a finer resolution affects the flood hazard assessment, does it make any difference whether I use 1km or 10km or 25km? Personally, I would find this a more interesting sensitivity analysis than, for example, the sampling in the POT approach (which has been done in previous studies for different locations).

A sensitivity analysis on the TWL resolution has been included in the Supplementary Material.

"Sensitivity analysis: TWL resolution

The adoption of large-scale (offshore) wave conditions has been discussed throughout the study, while the effects of the TWL spatial resolution are examined in this sensitivity analysis. To this end, two alternative resolutions have been tested at spatial resolutions of 10 km and 25 km. In contrast to the 51,010 CTPs resulting from the 1 km resolution, these led to 4,039 and 1,324 CTPs, respectively. To check whether this loss in TWL resolution and detail would affect the resulting 100-yr TWL and the corresponding flooded area (FA), a sensitivity analysis was performed. Results from all three cases were compared based on the methodology adopted in this study: nearshore wave conditions, Sunamura foreshore slope, Stockdon static wave setup, POT with lambda 2, 72h minimum interval, and an exponential fit.

Table S3 displays the 100-yr TWL variability per European basins and regions, based on the corresponding average and standard deviation values. For the entire study, the effect of the TWL resolution increases the more detail we add to the analysis. For example, for the entire study area, 100-yr TWL is 2.0 ± 1.3 m and 2.1 ± 1.4 m under 1 and 25 km spatial resolution, respectively. Meanwhile, in the Central Baltic Sea basin the 100-yr TWL changes from 1.4 ± 0.3 m with 1 km resolution to 1.7 ± 0.2 m with 25 km resolution. Overall, the higher the TWL resolution the lower are the return values, which indicate a possible overestimation of extreme TWL when lower resolution TWL is used instead. These results possibly represent a greater level of detail captured by the 1 km spatial resolution data as opposed to the one provided by 10 and 25 km resolutions.

For the European study area, the flooded areas vary by 0.5% and 1% when lowering the TWL resolution to 10 and 25 km, respectively. As with other analyses in this study, these values represent a balance of what occurs across different areas within the European floodplain. For example, in the Kattegat Bay the flooded area increases 0.5% with a lower TWL resolution, whereas in the NE Atlantic basin it decreases 14% with a 25 km resolution compared to the 1 km resolution adopted. In most basins, lower TWL resolution results in smaller flood extent. The most sensitive basins to the TWL resolution are located in the Atlantic region, which also showed to be the most sensitive region to the wave data selection in the sensitivity analysis of the wave setup components."

Table S3 Average ± standard deviation of 100-yr TWL (m) and the corresponding flooded area relative to the area of the corresponding floodplain (%) per European basin, region, and study area. Results are shown for three scenarios of TWL spatial resolution: 1 km, 10 km, and 25 km.

EU basin / region	100-yr TWL (m)			100-yr relative FA (%)		
	1 km	10 km	25 km	1 km	10 km	25 km
Norway Sea	2.4 ± 0.4	2.5 ± 0.4	2.5 ± 0.4	35.0	35.0	35.0
NE Atlantic	3.9 ± 1.2	3.9 ± 1.1	4.0 ± 1.2	28.8	26.9	24.8
North Sea	3.5 ± 1.1	3.7 ± 1.1	3.6 ± 1.1	57.0	56.8	56.9
English Channel	5.1 ± 1.6	5.0 ± 1.7	5.0 ± 1.7	47.6	47.5	47.6
Iberia and Biscay	3.0 ± 0.6	3.0 ± 0.6	3.0 ± 0.6	34.2	33.9	33.5
Macaronesia	1.8 ± 0.2	1.9 ± 0.2	2.0 ± 0.2	12.3	12.0	11.3
Atlantic coast	**3.2 ± 1.2**	**3.3 ± 1.2**	**3.4 ± 1.3**	**49.4**	**49.0**	**48.8**
Kattegat Bay	1.7 ± 0.2	1.8 ± 0.2	1.9 ± 0.2	19.8	19.7	19.9
Central Baltic	1.4 ± 0.3	1.7 ± 0.3	1.7 ± 0.3	22.8	22.7	22.7
Gulfs	1.7 ± 0.3	1.9 ± 0.3	1.9 ± 0.3	17.3	17.3	17.3
Baltic Sea	**1.6 ± 0.3**	**1.8 ± 0.3**	**1.8 ± 0.3**	**20.7**	**20.6**	**20.6**
Central Mediterranean	0.7 ± 0.1	0.8 ± 0.1	0.8 ± 0.1	16.2	16.3	16.2
Ionian Sea	1.1 ± 0.2	1.1 ± 0.2	1.1 ± 0.3	11.1	11.0	11.0
Adriatic Sea	0.7 ± 0.1	0.8 ± 0.1	0.8 ± 0.2	37.1	37.0	37.0

Mediterranean Sea	**0.8 ± 0.2**	**0.9 ± 0.2**	**0.9 ± 0.2**	**24.4**	**24.3**	**24.3**
EUROPE	**2.0 ± 1.3**	**2.1 ± 1.4**	**2.1 ± 1.4**	**36.8**	**36.6**	**36.5**

The TWL resolution was also addressed in the Discussion section. Lines 569 – 577: ”Besides the adoption of nearshore wave conditions in the estimation of the wave setup, the 1 km TWL hindcast has an unprecedented spatial resolution for this study area. When working with nearshore (downscaled) wave conditions, higher resolution is preferable to fully exploit the quality of the available information. Nearshore wave conditions capture local-scale variability, and the higher the resolution, the more faithfully the methodology represents these processes, making the best possible use of the data. A sensitivity analysis of the TWL resolution showed that the reduction in spatial resolution had minimal effect on the 100-yr TWL, although the flood extent varied up to 14% in some areas (Supplementary Table S3, sensitivity analysis on TWL resolution). The basins more sensitive to TWL resolution are located in the Atlantic region likely because of the increased need for accurate data when modeling a wider range of wave conditions given the highly energetic and variable wave climate in this region (Lobeto et al., 2024).”

2. The authors go into a lot of detail about different EVA methods, specifically the choice of distribution and threshold selection, but pass over other relevant aspects quickly such as the declustering window to derive independent events; they use a constant value following other studies (but no references are provided); that’s also an assumption (e.g., https://doi.org/10.1016/j.wace.2024.100701) with potential impact (as shown in the Arns et al., 2023 study).

A sensitivity analysis on the declustering time (minimum interval) adopted in POT has been included in the Supplementary Material:

“Sensitivity analysis: POT interval

The interval used when declustering events to guarantee their independence can be spatially constant or variable, depending on local climatic characteristics. For example, for waves in the Atlantic Ocean, previous evidence suggests that a minimum interval of at least 3 days (or 72h) is required, even though the Poisson assumption is well satisfied with intervals up to 6 days (Méndez et al., 2006). Meanwhile, a storm surge study along the German Bight showed that the influence of the declustering time on the return level outcomes is minimal when adopting POT, compared to AM (Arns et al., 2013). Ultimately, in a POT analysis, the independence criterion should reflect the geophysical origin and duration of the extreme events under study. On the one hand, when defining a spatially constant value, the interval is well-established with 72h in storm surge studies (Caspers et al., 2025; Dullaart et al., 2023; MacPherson et al., 2019; Pupić Vurilj et al., 2025; Vousdoukas, Voukouvalas, Annunziato, et al., 2016) although there are applications ranging from 1.5 days (Arns et al., 2013) to 6 days (Martín et al., 2024). In wave storm studies, a 48h interval has been commonly adopted (Lobeto et al., 2024) as well as a 12 – 24h calm period (Martzikos et al., 2023; Martzikos et al., 2021). On the other hand, spatially variable intervals have not been as explored, especially in large-scale studies. Similar to the adoption of a spatially variable POT threshold, the minimum interval would likely affect the estimation of return values as well as coastal flooding projections. However, the development of a methodology to define such heterogeneous interval values is out of the scope of the present study.

Following the decision of a spatially constant interval, a sensitivity analysis was performed to attest for the robustness of the adoption of the 72h instead of 48h and 96h. Figure S8 presents the EVA results obtained considering the different possibilities of declustering time. The 100-yr TWL variability resulting from the different POT minimum intervals tested is low compared to the other sensitivity analyses performed in this study (Figure S8a). The 95th confidence intervals of the 100-yr TWL indicate that 96h is the most uncertain option (Figure S8c), while 48h presents the lowest uncertainty (Figure S8d). However, the Anderson-Darling test analysis, indicates that the 72h interval leads to the highest proportion of CTPs adequately adjusting to the exponential fit (Figure S8b). Additionally, out of the 3,298 CTPs in which the 72h interval does not result in robust distributions, 2,098 CTPs rejected the null-hypothesis with all three POT minimum intervals tested.

The results indicate that, among the different steps of the methodology tested, the independence time between events has the least influence on the estimated return periods. Yet, this does not imply that the POT interval is irrelevant. Our analysis shows that it is still relevant, though less so than other methodological choices."

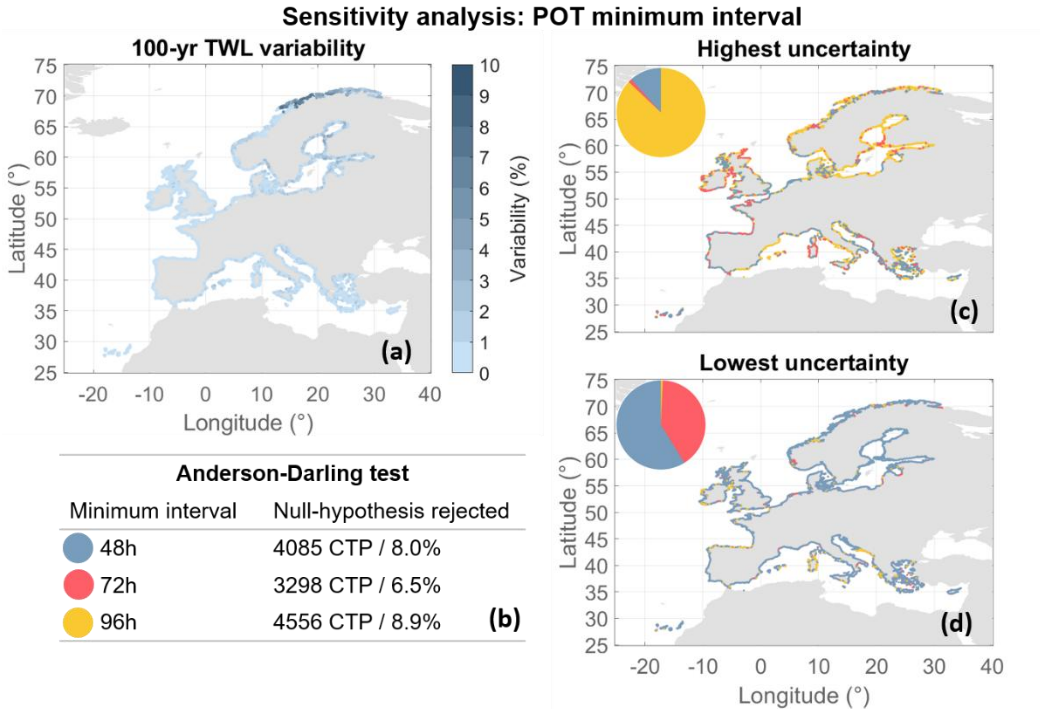

Figure S8 Sensitivity analysis of the POT interval selected (48h vs. 72h vs. 96h). Spatial variability 100-yr TWL variability (a). Anderson-Darling test results when adopting the different intervals (b). A rejected null hypothesis indicates that the sample does not fit the exponential fit. Uncertainty of results indicating which interval leads to the highest uncertainties (c) and the lowest uncertainties (d) of return levels, based on the 95th confidence level.

3. Overall, the paper is very long, especially introduction and discussion (but then there are no conclusions). The intro, for example, includes a lot of textbook style information about extreme value analysis, while the discussion repeats some info that is already touched on in the results section. I don't mind reading a longer paper, but the large amount of (sometimes tangential) information distracts a bit from what I think is the biggest novelty, namely the effect of using different approaches to account for wave contributions. I think with some additional effort it can become more concise while highlighting the key novelties and related conclusions.

A detailed revision of the text has been performed.

Introduction section. The introduction was revised to replace what the reviewer referred to as "textbook style information" with a discussion of recent advances in extreme value modeling. Lines 71 – 99: "Regarding the second key challenge, extreme value modeling can be approached through a variety of statistical frameworks, ranging from the well-established extreme value analysis (EVA), which provides the theoretical foundation for estimating return levels (Coles, 2001), to more refined methods designed to address different challenges. For example, dependence between variables can be treated with copulas (Bevacqua et al., 2020; MacPherson et al., 2019), spatial dependence can be included using Bayesian hierarchical models (Calafat & Marcos, 2020), and limited number of observations can be supported with regional frequency analysis (Collings et al., 2024). Yet, EVA remains the benchmark due to its low computational demands, relative simplicity, and long-standing use making it the baseline for comparison, especially in large-scale studies where reproducibility is essential.

EVA comprises two main steps: the identification of exceedances in a time series and a distribution family to which the identified sample is fitted to estimate return levels. The most commonly used approaches for identifying extreme events include the annual maxima (AM) and the peak-over-threshold (POT) methods, which differ primarily in sample size (Bezak et al., 2014). For an appropriately chosen threshold, POT generally yields larger samples of extremes than AM (Kirezci et al., 2020; Le Gal et al., 2023). However, threshold selection in POT is not straightforward. It must be high enough to exclude non-extreme events but low enough to ensure a sufficiently large sample for statistical analysis (Harley, 2017). In addition, the selected maxima peaks must represent independent extreme events, so a minimum time interval must be considered between two consecutive peaks. A stable threshold based on objective criteria, rather than arbitrary decisions, ensures consistency and more reliable results (Arns et al., 2013). Previous studies have tested different percentiles as the threshold and the 98th percentile (P98) of the TWL time series was used globally (Kirezci et al., 2020) while the 97th percentile (P97) was adopted for the European coast (Le Gal et al., 2023). Alternatively, a fixed TWL value can be used, although not suitable for large-scale studies where extreme magnitudes vary widely. To address spatial variability, Vousdoukas et al. (2016) adopted a threshold corresponding to an average number of independent events per year across the European coastline, while Vitousek et al. (2017) adjusted the selection of AM to the 3-largest events per year in a global study.

The choice of distribution families for estimating return levels also carries important implications that need to be carefully evaluated depending on regional and data characteristics. For example, Paprotny et al. (2016) applied a Gumbel distribution to the coast of Europe while Vitousek et al. (2017) applied a GEV, both based on a sample of extremes extracted with AM. Meanwhile, Kirezci et al. (2020) applied a Generalized Pareto Distribution (GPD) to a POT selected sample at the global scale. Nevertheless, the key features of an EVA application go beyond the method adopted to sample extreme events or the statistical model used to fit the data. Depending on regional climatic characteristics, data available, and variables considered (e.g., individual sources, combined drivers, and outputs of coastal hazards), the most suitable approach may vary (Coles, 2001)."

Results section. Some parts from sections 3.4 and 3.6 were transferred and properly adjusted into the Discussion.

Discussion section. Parts of the Discussion have been rewritten to incorporate the text removed from the Results sections mentioned above as well as to make the manuscript more concise, to highlight the improvement in TWL resolution, and to discuss the main limitations of the study. Lines 498 – 615: "The large-scale TWL calculation is complex and presents several challenges. We have presented a methodology for estimating large-scale TWL extremes and applied it to the European coast at the highest resolution to date. For the first time, this was done using nearshore wave conditions, particularly important for semi-enclosed European seas, where offshore data can lead to TWL overestimation. The methodology developed addresses the key challenges introduced regarding the inclusion of the wave contribution in the TWL, the TWL reconstruction, and the EVA method selected to characterize extreme events. As a consequence, it also introduces limitations inherited from some of the approaches adopted.

Concerning the wave contribution, even though the use of static wave setup as a component of TWL was validated, we acknowledge that this component may be slightly overestimated in areas such as the Baltic and Mediterranean Seas. Notably, this may also influence the relative contributions of the TWL components as well. Several factors influence the accuracy of wave contribution estimates. First, we use a semi-empirical setup formulation developed for open-coasts and beaches. It should be noted that although both our results and the literature indicate semi-empirical formulations as more appropriate and representative of current best practice, they also present limitations (Dodet et al., 2019). For example, the formulation adopted here was not designed for use in all types of wave regimes or beach profiles (Plant & Stockdon, 2015). Second, we rely on modeled foreshore slopes with capping and normalization, which may not fully capture the range and distributions of observed beach slopes in different coastal environments. This simplification can lead to the underestimation of the wave contribution in steep slopes or overestimation in gentler ones (Melet et al., 2020). Third, the adoption of a constant sediment grain size in the application of the Sunamura (1984) foreshore slope approximation is a necessary generalization at this scale and cannot fully represent the entire study area. This decision might overestimate the resulting foreshore slopes in areas such as wetlands and mudflats which have finer sediment grain sizes. However, although beach types and sediment materials are diverse across the European continent, the value adopted here follows the global study by Rueda et al. (2017) and agrees with values observed in smaller scale studies across the study area (Anthony & Héquette, 2007; Duo et al., 2020; Egon et al., 2025; Horn & Walton, 2007).

With respect to the TWL reconstruction, the linear approach adopted inevitably simplifies some coastal processes, particularly in areas such as bays and estuaries, for example. Although hydrodynamic modeling would be the ideal method to solve local processes, it is not yet feasible at this scale due to data and computational requirements. Nevertheless, although our TWL reconstruction does not explicitly resolve non-linear interactions, these are partly accounted for in the underlying databases. For instance, the storm surge database includes non-linear effects with the astronomical tide and the nearshore wave dataset considers sea level changes due to tide. Moreover, the TWL validation as a composite variable is inherently challenging, particularly at the large scale. While SWL can be validated with tide gauges, wave conditions can be validated with buoys. However, the increase in coastal water level as a result of the wave contribution is not captured by wave buoys. An alternative is to use instruments such as pressure sensors or ADCPs that capture local observations. Even these, however, are limited by the need for corrections, regular maintenance, and their localized nature. The lack of observational data and instruments allowing for the validation of the TWL has led previous studies to validate TWL indirectly, by validating each component individually. This strategy was adopted in China (Liu et al., 2025), New Zealand (Dalinghaus et al., 2025), and globally (Vousdoukas, Mentaschi, et al., 2018), for example. Therefore, given the successful validation presented for the TPXO, ROMS, offshore wave, and nearshore wave databases as well as the foreshore slope, we consider the TWL hindcast to be robustly validated for the purposes of this study. Nevertheless, a TWL validation was presented by comparing historical storms identified by previous studies. Even though most of the observed storms were referenced to tide gauges, which do not capture wave contribution, our analysis showed that the extreme events analyzed were well represented by the TWL hindcast. Given the lack of data to validate TWL, the use of tide gauge information to validate TWL is common practice in the literature (Kirezci et al., 2020; Le Gal et al., 2023; Treu et al., 2024; Wing et al., 2024). Similar to the results found here, when comparing validation results with and without wave contribution, Kirezci et al. (2020) detected more accurate estimates when including wave setup. The authors debate whether this is an underestimation of the SWL or the influence of waves given that these results are even more pronounced under extreme than average conditions. Even though the reasons behind these results remain inconclusive, it is clear that when considering the wave contribution, extreme events are better captured by TWL reconstructions.

Regarding the EVA methodology, its adoption in tide-dominated regions require careful evaluation. Tides are deterministic, periodic, and autocorrelated, which contradicts the EVA assumption of a sample of independent, abnormally extreme events. However, by applying EVA to the TWL composed of both deterministic and stochastic variables, random exceedances emerge due to the interaction of residuals with tides, introducing variability in the extremes. Therefore, fitting an extreme value distribution to the aggregated TWL, as performed in this study, remains a valid and informative approach and allows the tail behavior to be characterized even in regions dominated by astronomical tide.

One key outcome is the strong influence that methodological choices concerning the wave contribution to TWL reconstruction have on the results. The proportion of the European floodplain which is inundated under various TWL approaches varies by up to 10%, depending on wave dataset selection, foreshore slope approach, and wave setup formulation, with the largest differences arising from the choice of wave setup formulation. Using the Guza & Thornton (1981) method yields 4 – 7% more flooded area than the Stockdon et al. (2006) formulation, regardless of the foreshore slope assumption, whereas using a constant foreshore slope yields 1% more flooded area than incorporating variable slopes. Similarly, nearshore TWL produces 3% more flooded area than offshore TWL. The regional differences in the importance of the three components of wave setup can be explained by the geometry and exposure of the different regions. On the one hand, the steeper bathymetry of the Atlantic coast enhances wave shoaling processes, leading to higher waves as they approach the shore. In contrast, the shallower waters and gentler slopes in the Baltic and Mediterranean Seas tend to result in greater wave energy dissipation, as waves encounter the seabed earlier and lose energy more quickly. Although not accounted for in this study, non-linear processes involved in wave transformation, such as refraction and wave-breaking mechanisms, could also help explain such differences as wave transformation processes affect TWL mainly in areas with steep beach slopes and complex offshore bathymetry (Serafin et al., 2019). On the other hand, the Atlantic coast has a larger fetch, allowing for the development of higher waves, and is more exposed to coastal winds, which amplify wave energy. Conversely, the Mediterranean Sea has a smaller fetch, producing more localized and less energetic waves. These factors also result in greater discrepancies in wave conditions within the semi-enclosed seas compared to the Atlantic coast, where offshore and nearshore wave conditions are more similar. While these effects merit further analysis at a higher resolution and on a smaller scale, the results suggest that relying on offshore wave conditions, common in the literature, may lead to an underestimation of the actual flood extent.

Besides the adoption of nearshore wave conditions in the estimation of the wave setup, the 1 km TWL hindcast has an unprecedented spatial resolution for this study area. When working with nearshore (downscaled) wave conditions, higher resolution is preferable to fully exploit the quality of the available information. Nearshore wave conditions capture local-scale variability, and the higher the resolution, the more faithfully the methodology represents these processes, making the best possible use of the data. A sensitivity analysis of the TWL resolution showed that the reduction in spatial resolution had minimal effect on the 100-yr TWL, although the flood extent varied up to 14% in some areas (Supplementary Table S3, sensitivity analysis on TWL resolution). The basins most sensitive to TWL resolution are located in the Atlantic region likely because of the increased need for accurate data when modeling a wider range of wave conditions given the highly energetic and variable wave climate in this region (Lobeto et al., 2024).

A second outcome is the application of an EVA methodology that is appropriate for large-scale studies. The POT approach focuses on high-magnitude events, unlike annual maxima (AM) methods, which assume that a single extreme event occurs per year. Our findings show that threshold selection greatly influences sample size, particularly when applying a constant percentile threshold. Comparing our thresholds to previous studies, we found that 77.1% of the study area exhibits more than 12 events per year when the threshold of P97 used by Le Gal et al. (2023) is adopted. Meanwhile, if the thresholds of P98 and P98.5 used by Kirezci et al. (2020) and Paprotny et al. (2016) are adopted instead, this value decreases to 61.9% and 49.8%, respectively. These percentages indicate that large portions of the coast would not be adequately represented using these thresholds, leading to inconsistent analysis.

A third key insight is the importance of understanding the sources of extreme TWL events. Dominant TWL components provide insight into storm behaviors and potential impacts. The spatial patterns we observed have also been reported previously. The more exposed, tide-dominated Atlantic coast and the more sheltered and storm surge-dominated Baltic and Mediterranean Seas were identified by previous studies, even when neglecting waves (Merrifield et al., 2013). However, our results differ from others in the literature primarily because of our wave setup characterization. For example, compared to Vitousek et al. (2017) our storm surge contributions were higher for the European coast. This is perhaps due to the wave setup formulation adopted by the authors, which is specific to dissipative beaches and tends to overestimate wave setup. Additionally, their 111 km spatial resolution excluded marginal seas, where we observed the highest storm surge contributions. Meanwhile, we identified similar patterns of storm surge contributions in the Baltic Sea and of wave setup dominance in the Mediterranean Sea, when compared with Melet et al. (2018). Yet, the authors found lower tidal contributions than what we encountered for the Atlantic coast, likely due to the inclusion of the swash component, which gives more importance to the wave contributions and decreases the tidal influence. In our study, we excluded swash by using static wave setup, as swash operates on a scale of seconds to minutes, whereas flood events typically last hours to days (Hinkel et al., 2021; Parker et al., 2023).

Finally, we highlight the importance of considering TWL as a combination of its three components. For example, as we move towards more extreme return levels, the relative contribution of storm surge increases, exposing the coast to prolonged high TWL, which can also heighten wave setup processes (Su et al., 2024). However, we acknowledge that wave setup alone cannot drive coastal flooding. On the one hand, wave setup represents an increase in mean sea level of only a few centimeters to a couple of meters (Idier et al., 2019). On the other hand, the width of the coast affected by this increase in mean sea level is only tens to a couple hundred meters wide (Dodet et al., 2019). The volume of water being propagated towards the coast potentially leading to coastal flooding is not large when compared to tides and storm surges, which increase mean sea level over several kilometers of coastal extent (Woodworth et al., 2019). Additionally, although our study shows that astronomical tide modulates extreme TWL in many regions, it should be pointed out that the main drivers of TWL associated with coastal flooding are the unexpected extreme sea levels due to waves or storm surges as a result of storm conditions. This is because in physical terms, the tide is an expected oscillation to which coastal communities are well adjusted to. However, our analysis shows that without the inclusion of astronomical tide, coastal flooding would probably not occur in many regions of the study area. The results show that even in parts of the Baltic and Mediterranean Seas the tide reaches a contribution of more than 20%, being particularly relevant under average conditions (see Supplementary Fig. S16). Therefore, astronomical tide becomes crucial during the most extreme conditions, even in microtidal areas. Lastly, storm surges, often the primary source of extreme TWL, tend to sustain elevated water levels for extended periods. Therefore, neglecting any one of these components may lead to an underestimation of TWL, particularly if a storm coincides with a spring high tide, thereby increasing coastal risk."

A new Conclusions section has been added to the manuscript to place more emphasis on the key novelties and limitations of the study. Lines 616 – 642: "Currently, there remains a strong need to provide coastal flooding maps at large geographic scales. TWL is the primary input for coastal flood modeling and must be characterized appropriately for the scale of analysis. Without a proper assessment of TWL, estimating extreme conditions becomes more uncertain (Rohmer et al., 2021; Toimil et al., 2020). To address this need, the methodology presented incorporates spatial variability typical of large-scale coastal studies. This was achieved by using spatially and temporally variable foreshore slopes during TWL reconstruction and by applying spatially variable thresholds in the POT method. An analysis of relative contributions helped to interpret extreme TWL behavior along the European coastline. The characterization of three macro-regions (Atlantic coast, Baltic Sea, and Mediterranean Sea) supported the understanding of the degrees of uncertainty observed across different regions at distinct steps of the methodology.

Extreme TWL events can cause severe coastal impacts as water overflows inland, reaching communities, infrastructure, assets, and buildings (Vousdoukas, Bouziotas, et al., 2018). Proper reconstruction of TWL and its extreme events is the first step for accurate flood hazard assessment. Large-scale evaluations help us identify hotspots for more detailed risk assessments. Two key limitations of large-scale studies remain, which also point to priorities for future work. First, the computational demands of working with large datasets are high. However, improvements in data availability and computational efficiency have enabled this study to deliver high-resolution, high-quality TWL extreme estimates across Europe. Second, simplifications and assumptions are often required to handle diverse coastal environments. Here, representing TWL as a linear sum of its components overlooks possible additional inputs such as river discharges and non-linear interactions between TWL components, which are important in regions with wide continental shelves and enclosed lagoons (Bertin et al., 2012; Lorenz et al., 2023), such as the Baltic and Mediterranean Seas. According to Arns et al. (2020), not considering non-linear interactions between tide and storm surge, for example, can lead to a 30% increase in estimated extreme water levels, a 16% increase in coastal flooding costs, and an 8% increase in exposed people globally. A strategy to address this issue at the large-scale is to run a hydrodynamic model with both tidal and meteorological forcings combined (Haigh et al., 2014b). This approach, however, does not consider the contribution of waves.

Finally, we point out that although this study presents the highest-resolution estimates of extreme TWL for the entire European coastline to date, one could find a considerable variety of beach profiles and types, from sandy shores to rocky formations, within the 1 km distance adopted. Thus, while the TWL resolution adopted here is unprecedented at this scale, it remains insufficient for local-scale applications, where higher-resolution data are needed to support detailed planning."

Specific comments:

31 river discharge effects are not included, something worth mentioning somewhere, maybe when discussion limitations. Additional effects have been included. Lines 631 – 634: "Here, representing TWL as a linear sum of its components overlooks possible additional inputs such as

river discharges and non-linear interactions between TWL components, which are important in regions with wide continental shelves and enclosed lagoons (Bertin et al., 2012; Lorenz et al., 2023), such as the Baltic and Mediterranean Seas."

75-105 This part includes a lot of basic information that can be found in text books. At the same time it doesn't touch on more recent developments in extreme value modelling (for example, the approach proposed by Calafat and Marcos: https://www.pnas.org/doi/10.1073/pnas.1913049117, or approaches where distributions are fitted to the stochastic components of TWL after removing deterministic tides). The introduction was revised to replace what the reviewer referred to as "textbook style information" with a discussion of recent advances in extreme value modeling. Lines 71 – 78: "Regarding the second key challenge, extreme value modeling can be approached through a variety of statistical frameworks, ranging from the well-established extreme value analysis (EVA), which provides the theoretical foundation for estimating return levels (Coles, 2001), to more refined methods designed to address different challenges. For example, dependence between variables can be treated with copulas (Bevacqua et al., 2020; MacPherson et al., 2019), spatial dependence can be included using Bayesian hierarchical models (Calafat & Marcos, 2020), and limited number of observations can be supported with regional frequency analysis (Collings et al., 2024). Yet, EVA remains the benchmark due to its low computational demands, relative simplicity, and long-standing use making it the baseline for comparison, especially in large-scale studies where reproducibility is essential."

157-159 something is wrong with this sentence. Sentence modified.

166 "global median beach slopes reported locally" sounds very strange and I am not sure what it is supposed to mean. Clarified.

205-210 I got very confused here because the same numbering style is used for the list and the equations; I suggest switching to i, ii, iii or a, b, c for the list and keep (1) (2) style for the equations. Numbering style of the list has been modified to i, ii, and iii.

215 Refer to table 1 when mentioning approach A. I was wondering what it is and only realized on the next page. Table 2 referred to.

235 "two extreme events per year". Text modified.

242 "reconstructed…hindcast"; I would drop one. Text modified.

277 "Elbe Estuary". Text modified.

310-315 (and other places) When the tides are the main driver, does it make sense to fit an extreme value distribution because its assumptions are not met? I know it's a commonly used approach, but that doesn't make it necessarily right.

The reviewer raises a valid concern and it has been included as a limitation of the study. Lines 544 – 549: "Regarding the EVA methodology, its adoption in tide-dominated regions require careful evaluation. Tides are deterministic, periodic, and autocorrelated, which contradicts the EVA assumption of a sample of independent, abnormally extreme events. However, by applying EVA to the TWL composed of both deterministic and stochastic variables, random exceedances emerge due to the interaction of residuals with tides, introducing variability in the extremes. Therefore, fitting an extreme value distribution to the aggregated TWL, as performed in this study, remains a valid and informative approach and allows the tail behavior to be characterized even in regions dominated by astronomical tide."

348 delete "the" Deleted.

Figure 7 (and related text) How does it look when no wave contributions are included (which is still often the approach used for large-scale assessments)

We have included SWL to the sensitivity analysis on the wave contribution component. Lines 380 – 398: "Figure 7a presents the variability of the 100-yr TWL for different options of the wave contribution component: still water level (SWL), SWL combined with static wave setup, SWL with dynamic wave setup, and SWL with wave runup. The highest levels of variability are observed in the Mediterranean Sea, more specifically in Central Mediterranean and Ionian Sea, which could be an indication of relatively larger contributions of waves to TWL compared to the remaining two components, or a response to low ranges of TWL, indicating that even a slight change in the wave contribution is reflected in the estimation of TWL extremes. Meanwhile, the lowest variabilities are found in the Baltic Sea, probably a response to its limited exposure to incoming waves.

Figure 7b presents the FA per European basin, relative to their respective floodplain area, under the 100-yr TWL when adopting different wave contribution components. Overall, the highest increase in the FA occurs when changing from a static wave setup based-TWL to a dynamic wave setup based-approach. The basins most affected include Central Mediterranean and Iberia and Biscay. The former is located in a region typically known for coastal wave storms (Lobeto et al., 2024), while the latter appears to be a region sensitive to infragravity waves, which had been observed in the TWL reconstruction validation as well. Meanwhile, the least affected basins are Kattegat Bay and the Gulfs, both located in the Baltic Sea. Besides presenting low values of TWL, these basins are also amongst the steepest floodplains. However, when looking at the macro EU regions, the most affected one is the Mediterranean Sea, followed by the Atlantic coast and the Baltic Sea (Figure 7c). These results highlight a clear spatial variability of the wave contribution to TWL. Finally, the European FA with SWL is 36.09%, when including static wave setup it increases to 36.82%, dynamic wave setup to 40.40%, and wave runup to 41.23%. These results show that while the large-scale results do not change dramatically, it is important to zoom in to smaller regions and basins to identify areas in which such decisions might have the greatest impact."

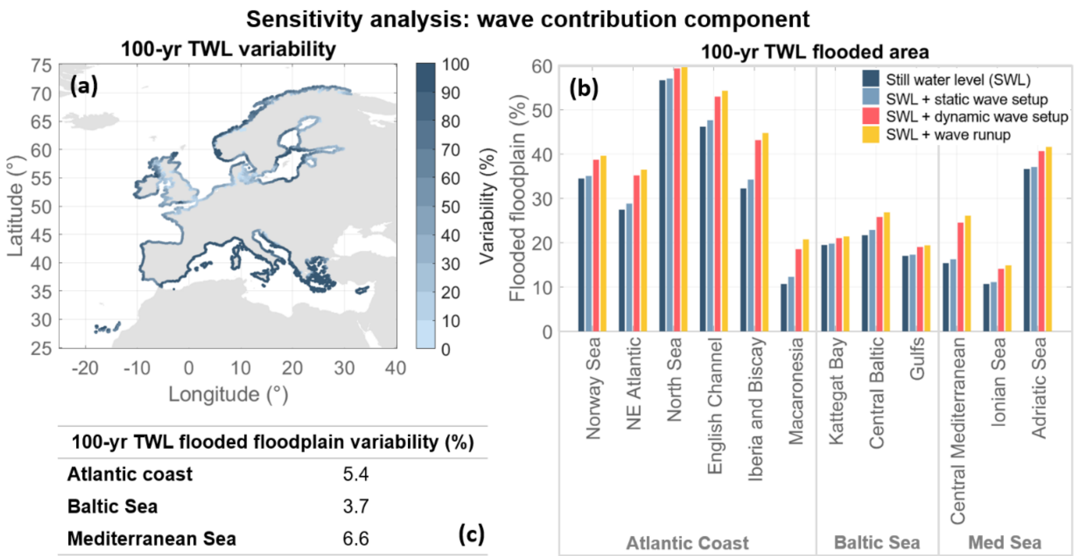

Figure 7 Sensitivity analysis of the wave contribution considered in the TWL reconstruction: still water level (SWL) vs SWL combined with static wave setup vs SWL with dynamic wave setup vs SWL with wave runup. (a) 100-yr TWL variability found in the results. (b) Results of maximum flooded area proportional to the floodplain in each European basin considering the 100-yr TWL. (c) Variability of 100-yr TWL flooded area relative to floodplain area per European region.

392-403 This and other parts of the results section read more like Discussion. Section moved to the discussion.

492 delete “an”. Deleted.

506 “applying a specially variable” (or “thresholds”). Text modified.

Supp. Fig. 8 “significancy” should be “significance” in the title. Figure title and caption modified.

Response to Reviewer 2

Review Report

In this paper, titled "Assessing extreme total water levels across Europe for large-scale coastal flood analysis", the authors present a modelling chain for computing high-resolution extreme sea levels along the European coast. I believe this manuscript may be of interest to NHESS readers; however, I do not recommend it for publication in its current form. Several aspects, mainly related to the manuscript structure, methodology, and model validation, need to be improved. I recommend resubmission, subject to the authors addressing the main specific comments outlined below:

1. Manuscript structure. This study is based on the results of a model chain. Therefore, in my opinion, a detailed description of the modelling system and its validation must be included in the main manuscript rather than relegated to the supplementary material. A Conclusions section is missing.

We understand the main concerns raised by the reviewer summarize to the large-scale aspect of the study which limits some of the methods adopted and possibilities for data validation given the lack of observational records at the scale. In the revised manuscript, we have expanded our discussion on the role and limitations of available observational records.

We would like to highlight that the focus of the study is to propose a methodology to estimate extreme TWL for the European coastline with the goal of providing the primary input of coastal flood studies, while taking considerable steps to improve the state-of-the-art characterization of large-scale TWL.

The scope of the study has been clarified in the manuscript. Line 104: "Focus is placed on the generation of the main input necessary for large-scale coastal flooding studies, the extreme TWL."

A new Conclusions section has been added to the manuscript to place more emphasis on the key novelties and limitations of the study. Lines 616 – 642: "Currently, there remains a strong need to provide coastal flooding maps at large geographic scales. TWL is the primary input for coastal flood modeling and must be characterized appropriately for the scale of analysis. Without a proper assessment of TWL, estimating extreme conditions becomes more uncertain (Rohmer et al., 2021; Toimil et al., 2020). To address this need, the methodology presented incorporates spatial variability typical of large-scale coastal studies. This was achieved by using spatially and temporally variable foreshore slopes during TWL reconstruction and by applying spatially variable thresholds in the POT method. An analysis of relative contributions helped to interpret extreme TWL behavior along the European coastline. The characterization of three macro-regions (Atlantic coast, Baltic Sea, and Mediterranean Sea) supported the understanding of the degrees of uncertainty observed across different regions at distinct steps of the methodology.

Extreme TWL events can cause severe coastal impacts as water overflows inland, reaching communities, infrastructure, assets, and buildings (Vousdoukas, Bouziotas, et al., 2018). Proper reconstruction of TWL and its extreme events is the first step for accurate flood hazard assessment. Large-scale evaluations help us identify hotspots for more detailed risk assessments. Two key limitations of large-scale studies remain, which also point to priorities for future work. First, the computational demands of working with large datasets are high. However, improvements in data availability and computational efficiency have enabled this study to deliver high-resolution, high-quality TWL extreme estimates across Europe. Second, simplifications and assumptions are often required to handle diverse coastal environments. Here, representing TWL as a linear sum of its components overlooks possible additional inputs such as river discharges and non-linear interactions between TWL components, which are important in regions with wide continental shelves and enclosed lagoons (Bertin et al., 2012; Lorenz et al., 2023), such as the Baltic and Mediterranean Seas. According to Arns et al. (2020), not considering non-linear interactions between tide and storm surge, for example, can lead to a 30% increase in estimated extreme water levels, a 16% increase in coastal flooding costs, and an 8% increase in exposed people globally. A strategy to address this issue at the large-scale is to run a hydrodynamic model with both tidal and meteorological forcings combined (Haigh et al., 2014b). This approach, however, does not consider the contribution of waves.

Finally, we point out that although this study presents the highest-resolution estimates of extreme TWL for the entire European coastline to date, one could find a considerable variety of beach profiles and types, from sandy shores to rocky formations, within the 1 km distance adopted. Thus, while the TWL resolution adopted here is unprecedented at this scale, it remains insufficient for local-scale applications, where higher-resolution data are needed to support detailed planning."

2. Modelling system. The total water level is computed by linearly summing tidal, storm surge, and wave contributions. These three components are simulated using different numerical

models, each with distinct domains, resolutions, and bathymetry. This could lead to inconsistencies at the selected coastal locations, and this issue must be discussed in the manuscript.

The inherent limitations associated to the regional scale of the TWL estimation have been included in the Discussion. Lines 520 – 522: "With respect to the TWL reconstruction, the linear approach adopted inevitably simplifies some coastal processes, particularly in areas such as bays and estuaries, for example. Although hydrodynamic modeling would be the ideal method to solve local processes, it is not yet feasible at this scale due to data and computational requirements."

Furthermore, the supplementary material indicates that the ROMS model is used to simulate sea levels induced by tidal and meteorological forcing. This approach captures the non-linear interactions between tides and storm surges (i.e., tidal fluctuations affecting storm surges and vice versa). It is therefore unclear why the authors de-tided the ROMS sea levels and used only the residual component to estimate the still water level. Is ROMS less accurate than TXPO9 in reproducing tidal dynamics? The authors must justify this choice and provide evidence that the selected approach yields the best results.

The developed approach to obtain the storm surge database includes two simulations with similar domain and setup of the ROMs. One simulation considering astronomical tide and meteorological forcings and other, a simpler tide-only simulation. Although the two model configurations were calibrated and the two outputs (tide and still water level) were validated previous to simulating the > 30 years (1985 – 2021) hindcast database, the tide-only outputs were not processed and saved in a database. They were used in the post-processing step to remove the tide component from the tide-atmospheric forced simulation outputs, thereby generating a 'storm surge' database considering non-linear interactions between these sea level dynamics. This approach has been applied in other studies to analyze the 'storm surge' data in the coastal area at regional and global scale, such as Fernández-Montblanc et al. (2019), Muis et al. (2020), Chaigneau et al. (2022), Irazoqui Apecechea et al. (2025), among others.

Since the time series of tide-only values from the ROMs are not available, we cannot compare these values against TPXO9 tide values. We chose the TPXO9 tide database because of its ability to reproduce the astronomical tide, which was verified by comparing against tide-gauge records in Europe. TPXO9 has a high coastal resolution (1/30º), assimilates satellite altimetry data during the simulation, and has been generated using a recognized global tide model by the scientific community.

Moreover, the coarse resolution of the ROMS model (5-11 km as stated in line 144) is a limiting factor in the reconstruction of the TWL at a 1 km resolution. Since the model results are extracted at the coast, the authors should specify the minimum bathymetric values used in the grids of the different models.

The minimum depth considered in the European ROMs numerical domain is 10 meters. We have included this information in the SP text describing the ROMs setup description.

We agree that the limitations of the applied regional pan-European approach must be indicated in the manuscript, and these have now been discussed for each main theme analyzed: wave contribution (Lines 504 – 519), TWL reconstruction method (Lines 520 – 543), and EVA (Lines 544 – 549).

However, we also note that validation of the astronomical tidal time series generated from TPXO9 (Figure S1 and Table S1) and of the ROMs-based storm surge database (Figures S1 and S2, Table S2) has been carried out using 48 tide-gauge records over Europe, demonstrating high skill in reproducing the tide magnitude and phase, as well as maxima storm surges.

3. Tide-surge-wave non-linear interactions. According to Arn et al. (2020; https://doi.org/10.1038/s41467-020-15752-5), tide-surge non-linear interactions are relevant in several European locations. However, in line 120, the authors state that they reconstruct TWL by "linear summing time series of wave setup, storm surges, and astronomical tides". This suggests that non-linear interactions among tides, surges, and waves are not explicitly accounted for in the TWL reconstruction. Nevertheless, the supplementary material indicates that the ROMS model is used to simulate sea levels induced by tidal and meteorological forcing, which is intended to capture non-linear interactions between tides and storm surges (see my comment 2). It is therefore unclear which specific non-linear processes are considered in this study and how they are incorporated into the overall analysis.

We reconstructed a TWL hindcast on an hourly basis by considering time series of wave setup, storm surges, and astronomical tides. However, this does not mean that the generation of the databases of these variables did not take into account non-linear interactions between them. For example, the storm surge database includes nonlinear effects with the tide, as it is explained in the SM (section 'Storm surge'), and the nearshore wave dataset generated using the DOW method (SM section 'Wind-waves') has considered sea level changes due to tide for the wave domains under meso- and macro-tidal ranges.

This information has been clarified in the manuscript. Lines 522 – 525: "Nevertheless, although our TWL reconstruction does not explicitly resolve non-linear interactions, these are partly accounted for in the underlying databases. For instance, the storm surge database includes non-linear effects with the astronomical tide and the nearshore wave dataset considers sea level changes due to tide."

4. Wave contribution to sea level. A critical aspect of computing total water levels at the coast is evaluating the wave contribution (wave setup and run-up). I commend the authors for their thorough sensitivity analysis of the wave contribution formulation. However, the approach adopted in this study uses a foreshore slope parameter derived from modelled wave properties rather than actual coastal characteristics. The authors should validate their slope estimation using literature data or values from topo-bathymetric datasets (e.g., EMODNET). Additionally, site-specific geomorphological features must be considered. For example, gravel beaches and rocky cliffs - common in many Mediterranean coastal locations - require different methodologies for estimating wave run-up. Ignoring coastal characteristics could lead to unrealistic wave setup estimations, resulting in erroneous evaluations of the relative wave contribution to total water levels, both locally and at regional or basin scales.

The reviewer raises a valid concern, and the validation of foreshore slopes has been included in the Supplementary Material:

"Validation: foreshore slope

The foreshore slope is one of the main uncertainty sources in coastal flooding studies that consider wave contributions. Therefore, we validated the adopted approach to ensure its suitability at larger scales and identified the regions that are most sensitive to such variable.

The estimated foreshore slope data was validated with local-scale topo-bathymetry data and locally observed values for a series of sites across the study area. Figure S6 compares the foreshore slopes estimated with the Sunamura (1984) formulation against those obtained with a traditional method based on high-resolution data or field observations. The Sunamura formulation estimates foreshore slope as a spatially and temporally variable function of wave conditions. By contrast, the traditional method derives slopes from high-resolution DEM profiles (perpendicular to the coastline) in Sites 1 – 8, calculated as the intertidal slope between the mean low and high tides at each site. DEM data were obtained at the highest resolution available for the different sites (DEFRA, 2023; PDOK, 2023; IGN, 2021; IGN, 2019; Geoportale Emilia-Romagna, 2014). Meanwhile, the validation of Sites 9 – 12 was performed with locally observed slopes reported by previous studies (Nuyts et al., 2024; Carneiro-Barros et al. 2025); Jarmalavičius et al., 2025; McGlashan et al., 2004; respectively).

Overall, the different methods show good agreement, with the best agreements observed in Site 5, where the high-resolution DEM is in fact a topo-bathymetry allowing for a more precise slope estimate, and Site 9 where the reference slope was based on field observations of intertidal slope. The Sunamura formulation showed to be conservative (i.e., smoother slopes) in half of the sites used for validation. However, it is considered that the differences are small enough that they would not be reflected in the resulting extreme TWL estimations. The findings show that the Sunamura (1984) formulation is a robust method that performs well across diverse coastal settings, making it suitable for application at continental scale. Given that the traditional method requires field work or high-resolution topo-bathymetric data, which is not consistently available across Europe, the Sunamura approach represents a practical and sufficiently accurate alternative for defining foreshore slopes.

An identification of rocky coastlines has been conducted to acknowledge regions in which the misrepresentation of the foreshore slope could lead to a possible unrealistic estimation of the wave setup. According to EUROSION (2004), 16.7% of the study area is comprised of rocky cliffs, most of which is located along the United Kingdom (UK), Sweden, and Finland. Given the low wave climate energy in the Baltic Sea (Björkqvist et al., 2024), it is assumed that the possible inaccuracy of the foreshore slope in Sweden and Finland would likely not influence the extreme TWL results or the relative contributions of TWL components. Meanwhile, UK was included in the foreshore slope validation confirming its adequate representation."

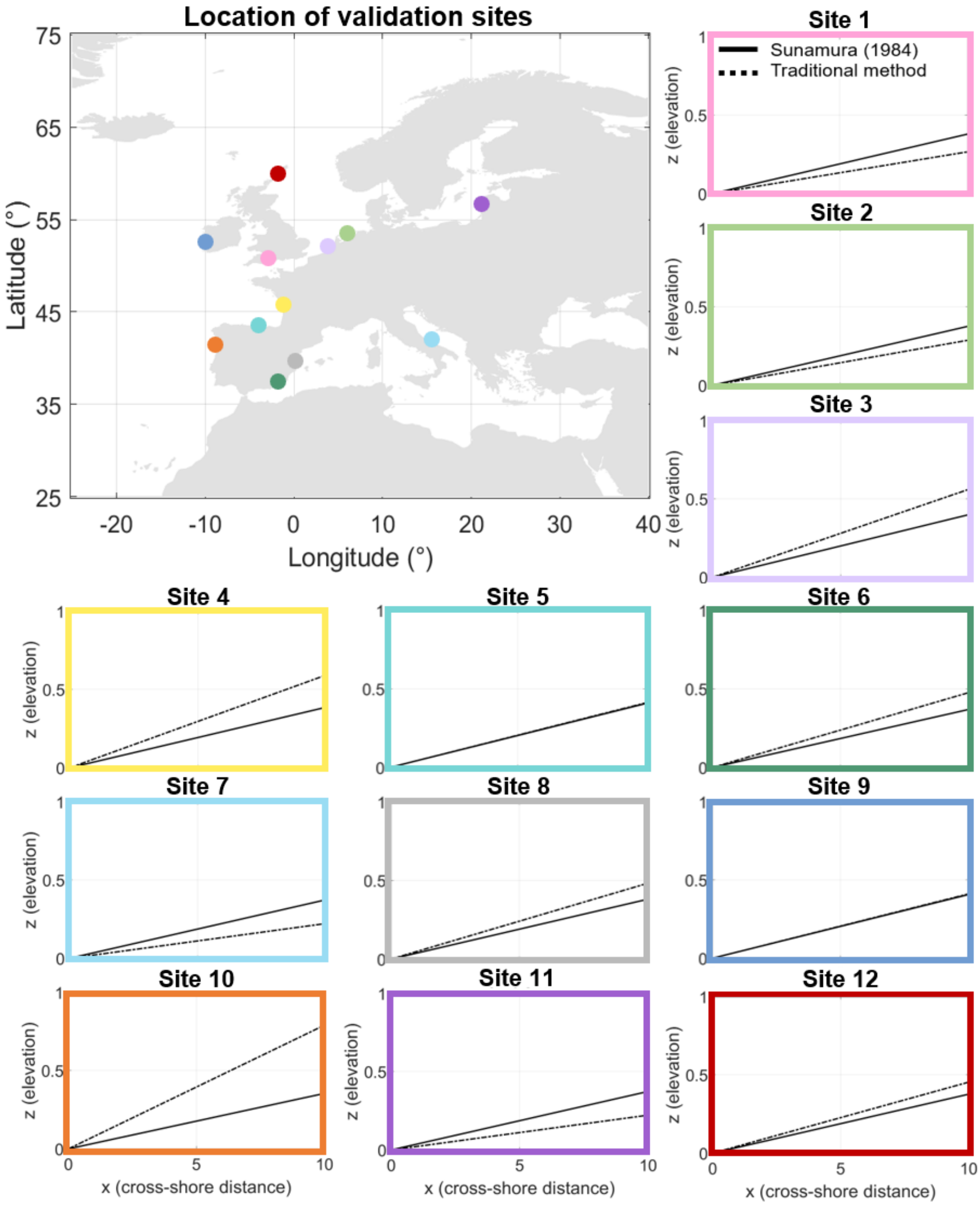

Figure S6 Comparison of mean foreshore slopes resulting from the Sunamura (1984) approach and a traditional method per site. Colored boxes refer to the location of sites shown in the map. The traditional slope values in Sites 1 – 8 were estimated with high-resolution DEM ranging from 1 m spatial resolution in Site 1 to 5 m in Site 2, for example. The traditional slope values in Sites 9 – 12 were obtained from local observations reported by previous studies.

5. Model validation. The supplementary material presents an incomplete validation of the numerical models used. I strongly suggest improving this section as outlined below and including the main validation results in the manuscript.

Thank you for the suggestion. The validation description of the databases used in the Supplementary Material has been extended with more detailed information (Figures S1, S4, and S5 as well as Tables S1 and S2).

As including detailed information on the validations performed in the manuscript would deviate the main goal and scope of the study, which is to provide extreme TWL input to large-scale coastal flooding studies, we chose to include a summary of the main validation results in the manuscript. Lines 177 – 199: “2.4 Model Validation Summary

The modeling chain used to reconstruct TWL combines independent, state-of-the-art datasets for tides, storm surges, and waves, each of which was validated against available observations prior to their integration. Table 1 presents a summary the validation results. Astronomical tides derived from TPXO9v5 were validated against 48 tide-gauge records distributed along the European coastline, showing high skill in reproducing tidal amplitudes and phases, with root mean square errors typically below 10 cm, biases close to zero, and Pearson correlation coefficients exceeding 0.9 at most locations (Table S1, Fig. S1). The storm surge hindcast generated with ROMS was validated against non-tidal residuals from the same tide-gauge network, yielding RMSE values generally between 6 and 15 cm, low bias, and correlation coefficients above 0.7 across most stations (Table S2, Fig. S2).

Offshore and nearshore wave hindcasts were validated against buoy observations, showing average biases below ±0.15 m, RMSE values typically below 0.35 m, and correlation coefficients exceeding 0.8 for the majority of stations (Figs. S4–S5). The foreshore slope estimation used to parameterize wave setup was independently validated against high-resolution topo-bathymetric data and field-based observations across representative European coastal settings, confirming its suitability for large-scale applications (Fig. S6).

Direct validation of TWL as a composite variable remains challenging at continental scale due to the scarcity of observations capturing wave-induced water-level contributions. Consistent with previous large-scale studies, TWL reliability is therefore assessed through the validation of its individual components and by comparison with documented historical extreme events. This approach provides confidence that the reconstructed TWL hindcast is robust and appropriate for the large-scale flood hazard analyses targeted in this study.”

Table 1 Model validation summary obtained from the different validation analyses performed for the different databases composing TWL, including the foreshore slope.

Component	Dataset / Model	Observations used	Key validation metrics (typical ranges)	Reference
Astronomical tide	TPXO9v5	48 tide gauges (Europe)	RMSE < 10 cm; Bias ≈ 0 cm; r > 0.9	Table S1, Fig. S1
Storm surge	ROMS (tide + meteo)	Non-tidal residuals from tide gauges	RMSE 6–15 cm; Bias < ±2 cm; r > 0.7	Table S2, Fig. S2

Component	Dataset / Model	Observations used	Key validation metrics (typical ranges)	Reference
Offshore waves	WW3-based hindcast	84 wave buoys	Bias < ±0.1 m; RMSE ≈ 0.3 m; r > 0.9	Fig. S4
Nearshore waves	Downscaled wave hindcast	Coastal wave buoys	Bias < ±0.15 m; RMSE 0.2–0.35 m; r > 0.8	Fig. S5
Foreshore slope	Sunamura-based estimate	DEMs & field studies	Good agreement across sites; conservative estimates	Fig. S6
Total Water Level	Composite (TWL)	Historical extreme events	Qualitative agreement with documented extremes	Table S4

- Tidal model. It is stated "Comparison of the tide simulated from the TPX09v5 outputs and measured by the tide gauges shows a near-perfect agreement, indicating a good non-stationary reconstruction of both tidal amplitudes and phases". However, no statistical metrics are provided. Are the tidal water levels or the tidal amplitudes and phases being evaluated? Why is the term 'non-stationary reconstruction' used? How is non-stationarity assessed? The authors must provide a comprehensive validation of the tidal model.

More details on the validation of the tidal model have been included in the Supplementary Material. Moreover, with the 'non-stationary reconstruction' wording, we wanted to indicate that the series accurately reproduce both the magnitude and phase of tidal cycles through time. We have re-written this sentence to avoid any misunderstanding.

"Astronomical tide from TPXO was validated against 48 tide-gauge records along the European coast. The validation was performed using comparisons of time series, quantile–quantile plots, and several statistical metrics (e.g., root mean square error – RMSE, bias, and Pearson correlation). Table S1 presents statistical metrics of the comparison of the tide series simulated from the TPXO9v5 outputs and measured by the tide gauges. Localized examples of quantile-quantile plots are presented in Figure S1 show the near-perfect agreement observed across the validation of the tidal model. With most of the cases presenting a RMSE below 10 cm, bias below 1 cm, and a Pearson correlation above 0.9, the results indicate the series accurately reproduce both the magnitude and phase of tidal cycles through time."

Table S1 Information on the tide-gauge stations used for validation of the tidal model. Tide-gauge station names, coordinates, RMSE, bias, and Pearson correlation values at each location.

	Tide gauge	Location		RMSE (cm)	Bias (cm)	Correlation
		Latitude	Longitude			
1	Vigo	42.243	-8.726	6.991	-0.563	1.00
2	A Coruña	43.364	-8.399	6.172	-0.610	1.00

	Tide gauge	Location		RMSE (cm)	Bias (cm)	Correlation
		Latitude	Longitude			
3	Ferrol	43.463	-8.326	5.476	-0.611	1.00
4	Gijón	43.558	-5.698	6.269	-0.637	1.00
5	Santander	43.461	-3.791	7.617	-0.644	1.00
6	Bilbao	43.357	-3.05	5.689	-0.644	1.00
7	Barcelona	41.342	2.163	3.851	-0.110	0.78
8	Catania	37.498	15.094	3.487	0.081	0.84
9	Livorno	43.546	10.299	3.817	-0.078	0.90
10	Ancona	43.625	13.506	4.713	-0.268	0.92
11	Imperia	43.878	8.019	3.511	-0.045	0.88
12	Genova	44.41	8.925	3.138	-0.100	0.92
13	Ravenna	44.492	12.283	7.328	-0.435	0.94
14	Venezia	45.418	12.426	7.699	-0.506	0.95
15	Trieste	45.649	13.759	9.318	-0.522	0.94
16	Nice	43.695	7.285	3.123	-0.104	0.91
17	Monaco	43.733	7.424	3.044	-0.104	0.91
18	Port Bloc	45.568	-1.062	17.923	-0.711	0.99
19	La Rochelle	46.148	-1.226	23.011	-1.349	0.99
20	Brest	48.383	-4.495	17.012	-0.905	1.00
21	Cherbourg	49.651	-1.635	11.205	0.181	1.00
22	Calais	50.969	1.868	19.824	1.374	0.99
23	Dunkerque	51.048	2.367	20.374	1.380	0.99
24	St Marys	49.918	-6.315	9.727	-0.281	1.00
25	Newlyn	50.103	-5.543	10.155	-0.133	1.00
26	Cromer	52.934	1.301	10.479	2.369	1.00
27	Whitby	54.483	-0.616	8.525	1.129	1.00
28	North Shields	55.007	-1.439	8.159	1.549	1.00
29	Aberdeen	57.15	-2.083	8.734	1.958	1.00
30	Stornoway	58.207	-6.389	9.017	-0.987	1.00
31	Wick	58.433	-3.083	8.214	0.206	0.99
32	Lerwick	60.154	-1.138	6.509	-0.424	0.99
33	Castletownbere	51.649	-9.903	7.094	-0.338	1.00
34	Malinhead	55.367	-7.333	9.274	-0.823	0.99
35	Cuxhaven	53.868	8.717	46.100	0.140	0.92
36	Helgoland	54.179	7.89	47.125	-0.001	0.85
37	Aarhus	56.15	10.217	8.160	-0.121	0.87
38	Goteborg	57.683	11.8	7.466	0.201	0.59
39	Stockholm	59.325	18.082	1.020	-0.088	0.65
40	Tregde	58.006	7.566	6.147	-0.069	0.72
41	Helgeroa	58.995	9.856	8.320	-0.062	0.69
42	Viker	59.036	10.949	9.111	-0.066	0.67
43	Rorvik	64.867	11.25	32.878	-0.003	0.86
44	Kabelvaag	68.212	14.482	15.806	-0.001	0.98
45	Andenes	69.326	16.135	20.360	-0.002	0.92
46	Vardo	70.333	31.1	58.073	0.004	0.71

Tide gauge		Location		RMSE (cm)	Bias (cm)	Correlation
		Latitude	Longitude			
47	Honningsvag	70.98	25.973	43.329	-0.002	0.79
48	Reykjavik	64.15	-21.933	7.282	-0.150	1.00

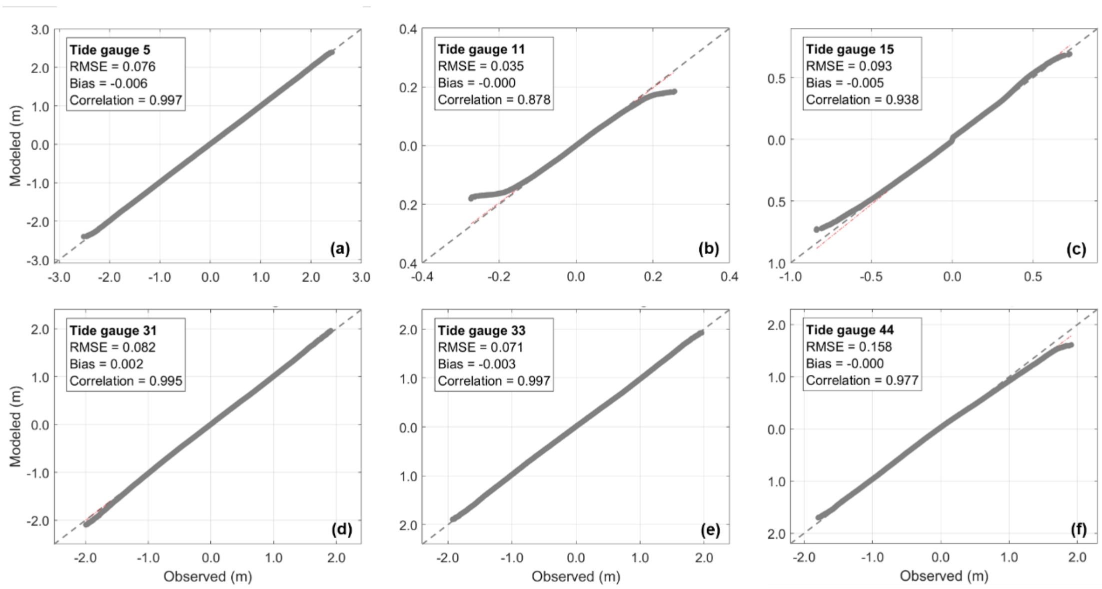

Figure S1 Quantile–quantile plots comparing tide-gauge observations (measured) with the tidal model (modeled) for six locations. Gray dashed lines indicate a 1:1 relationship and red-dashed lines correspond to a linear regression fitted to the paired quantiles. RMSE and bias values are shown in meters. (a) Santander (Spain), (b) Imperia (Italy), (c) Trieste (Italy), (d) Wick (Scotland), (e) Castletownbere (Ireland), and (f) Kabelvaag (Norway).

- Storm surge model. See my comment 2. A correlation coefficient between model results and observations should be provided, along with RMSE and BIAS. What are the reference datums of the tide gauges and the storm surge model used to compute BIAS?

The Pearson correlation has been added to the statistical validation metrics provided in Table S2. The reference levels used to compute the bias are: the local mean sea level of the ROMs model output for the hindcast (baroclinic sea level anomalies are not included) and the local zero obtained from the non-tidal-residual after subtracting the mean sea level as the mean value of each tide-gauge record.

Table S2 Information on the tide-gauge stations used for validation of the storm surge hindcast. Tide-gauge station names, coordinates, RMSE, bias, and Pearson correlation values at each location.

Tide gauge		Location		RMSE (cm)	Bias (cm)	Correlation
		Latitude	Longitude			
1	Vigo	42.243	-8.726	7.447	0.023	0.74
2	A Coruña	43.364	-8.399	6.840	-0.073	0.77
3	Ferrol	43.463	-8.326	7.434	0.000	0.78
4	Gijón	43.558	-5.698	7.130	-0.054	0.76

	Tide gauge	Location		RMSE (cm)	Bias (cm)	Correlation
		Latitude	Longitude			
5	Santander	43.461	-3.791	6.732	-0.053	0.72
6	Bilbao	43.357	-3.05	7.124	0.044	0.72
7	Barcelona	41.342	2.163	7.430	0.006	0.69
8	Catania	37.498	15.094	6.193	-0.149	0.63
9	Livorno	43.546	10.299	6.736	-0.351	0.70
10	Ancona	43.625	13.506	8.046	-0.639	0.74
11	Imperia	43.878	8.019	6.087	-0.241	0.71
12	Genova	44.41	8.925	6.105	-0.331	0.73
13	Ravenna	44.492	12.283	8.555	-0.652	0.75
14	Venezia	45.418	12.426	8.989	-0.691	0.74
15	Trieste	45.649	13.759	9.582	-0.523	0.72
16	Nice	43.695	7.285	6.024	-0.234	0.72
17	Monaco	43.733	7.424	5.939	-0.192	0.72
18	Port Bloc	45.568	-1.062	9.130	-0.630	0.72
19	La Rochelle	46.148	-1.226	9.707	-0.628	0.71
20	Brest	48.383	-4.495	8.209	-0.753	0.78
21	Cherbourg	49.651	-1.635	8.678	-0.703	0.78
22	Calais	50.969	1.868	13.121	-0.903	0.74
23	Dunkerque	51.048	2.367	13.277	-1.151	0.77
24	St Marys	49.918	-6.315	7.851	-0.594	0.77
25	Newlyn	50.103	-5.543	8.189	-0.613	0.77
26	Cromer	52.934	1.301	10.326	-1.292	0.87
27	Whitby	54.483	-0.616	8.982	-1.079	0.84
28	North Shields	55.007	-1.439	8.940	-1.074	0.82
29	Aberdeen	57.15	-2.083	8.657	-1.103	0.82
30	Stornoway	58.207	-6.389	9.307	-1.238	0.84
31	Wick	58.433	-3.083	9.002	-1.252	0.83
32	Lerwick	60.154	-1.138	8.383	-1.030	0.82
33	Castletownbere	51.649	-9.903	7.751	-0.564	0.81
34	Malinhead	55.367	-7.333	8.204	-1.219	0.87
35	Cuxhaven	53.868	8.717	16.696	-2.172	0.88
36	Helgoland	54.179	7.89	13.364	-2.319	0.90
37	Aarhus	56.15	10.217	9.878	-0.951	0.81
38	Goteborg	57.683	11.8	10.088	-1.190	0.82
39	Stockholm	59.325	18.082	10.385	-0.363	0.71
40	Tregde	58.006	7.566	8.812	-0.861	0.78
41	Helgeroa	58.995	9.856	9.464	-1.034	0.83
42	Viker	59.036	10.949	10.632	-0.988	0.82
43	Rorvik	64.867	11.25	9.463	-1.176	0.82
44	Kabelvaag	68.212	14.482	9.637	-1.261	0.83
45	Andenes	69.326	16.135	9.576	-0.969	0.78
46	Vardo	70.333	31.1	9.147	-0.820	0.78
47	Honningsvag	70.98	25.973	9.026	-0.940	0.79
48	Reykjavik	64.15	-21.933	7.730	-0.560	0.82

o Wave models. The statistical metrics for both the offshore and the nearshore datasets must be provided.

Validation figures comparing the offshore and nearshore hindcasts against buoy records have been incorporated into the Supplementary Material. For both datasets, the bias, Pearson correlation, and RMSE metrics are provided. The statistical metrics have been included under the section ‘Wind-waves’:

“For the offshore wave validation, comparison of significant wave height from the hindcast against 84 offshore buoys resulted in an average bias of –0.07 m, RMSE of 0.34 m, and Pearson correlation of 0.93 (Figure S4). Bias, RMSE, and Pearson correlation values resulting for the nearshore hindcast compared with coastal buoy observations are shown in Figure S5. The spatial distribution of the bias shows both positive and negative values, which in most locations do not exceed an absolute value of 0.15 m. Positive biases are predominant in the Baltic Sea, while negative biases prevail in the Mediterranean Sea and the Canary Islands. Other regions present a mix of positive and negative values, with several coastal areas showing biases close to zero. RMSE values between 0.2 and 0.3 m are observed in the North Sea, the Baltic Sea, and most locations in the Mediterranean Sea. Higher values are found in the Aegean Sea, Cantabrian Sea, English Channel, and around the British Isles. Pearson correlation coefficients remain above 0.6 for all buoys analyzed, with 85% of the buoys exhibiting values above 0.8.”

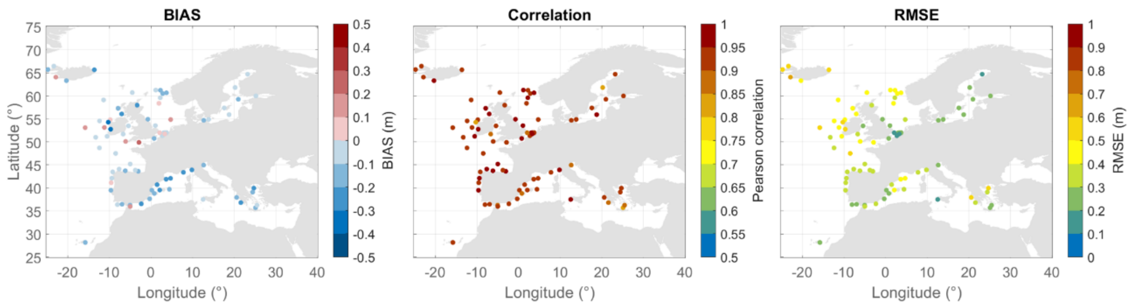

Figure S4 Spatial distribution of bias (left), Pearson correlation (middle) and RMSE (right) in significant wave height (Hs, m) from the offshore wave hindcast, evaluated at the locations of wave buoys.

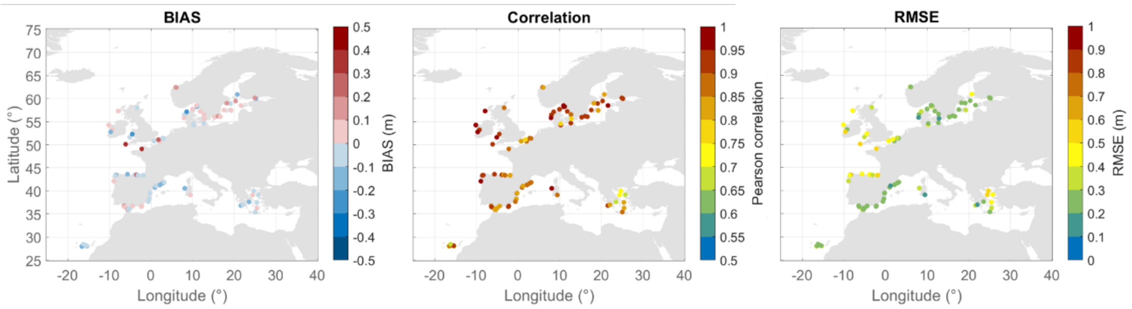

Figure S5 Spatial distribution of bias (left), Pearson correlation (middle) and RMSE (right) in significant wave height (Hs, m) from the nearshore wave hindcast, evaluated at the locations of wave buoys.

- TWL reconstruction. The authors validated the TWL reconstruction by comparing the estimated values against previously observed historical storms across the study area, based on previous studies (the list of which is reported in the supplementary material). It is not clear at all how this validation was performed. Did the authors consider only the maximum values during specific storm events? If so, the date of the events must be provided as well. Were the values taken directly from the publications or from tide gauge databases? My main concern regarding the TWL validation is the assumption that the observations account for the wave contribution to the sea level. However, several tide gauges (Catania, Venice, Marina di Campo, Santander and others) are located in protected harbors or behind jetties, and therefore are not exposed to wave action. It appears that the addition of the wave setup contribution merely compensates for an underestimation of the still water level. In conclusion, these observations cannot be used, and a proper TWL validation is missing.

Firstly, we confirm that we considered only the maximum value during specific storm events and that the values were taken directly from published studies. As suggested by the reviewer, the dates of the events have been provided in Table S4.

Table S4 References used in the TWL reconstruction validation indicating the locations analyzed and the corresponding dates of the extreme events considered.

References	Location validated	Date of extreme event
Breilh et al. (2013)	Brouage (FR)	28/02/2010
Cabrita et al. (2024)	Porto Garibaldi (IT)	06/02/2015
IHCantabria (2021a)	Santander (ES)	03/03/2014
Irazoqui Apecechea et al. (2023)	Hoek van Holland (NL)	02 – 04/01/2018
	Huelva (ES)	26/02 – 05/03/2018
	Kiel (DE)	01 – 03/01/2019
	Marina Di Campo (IT)	27 – 30/10/2018
	Venice (IT)	11 – 18/11/2019
	Valencia (ES)	20 – 23/01/2020
Kiesel et al. (2023)	Koserow (DE)	02/01/2019
	Sassnitz (DE)	02/01/2019
	Schlei Fjord (DE)	02/01/2019
	Wismar (DE)	02/01/2019
	Klagshamn (SE)	02/01/2019
Koks et al. (2023)	Ebel Estuary (DE)	05 – 06/12/2013
Lemos et al. (2025)	Faro (PT)	04/01/2014
Perini et al. (2015)	Porto Corsini (IT)	06/02/2015
Scicchitano et al. (2021)	Catania (IT)	24 – 28/12/2009
	Catania (IT)	05 – 09/11/2014
	Catania (IT)	10 – 12/02/2015

References	Location validated	Date of extreme event
	Portopalo di CapoPassero (IT)	27 – 28/09/2018
	Portopalo di CapoPassero (IT)	28 – 30/10/2016
	Portopalo di CapoPassero (IT)	11 – 12/11/2019
	Sicily (IT)	27 – 28/09/2018
	Malta (MT)	17 – 18/09/2020
Wadey et al. (2012)	Lymington (UK)	17/12/1989
	Solent (UK)	10/03/2008
Wolski & Wiśniewski (2021)	Wismar (DE)	20 – 23/11/2020
	Korsor (DK)	20 – 23/11/2020
	Parnu (EE)	20 – 23/11/2020
	Ristna (EE)	20 – 23/11/2020
	Hamina (FI)	20 – 23/11/2020
	Kemi (FI)	20 – 23/11/2020
	Swinoujscie (PL)	20 – 23/11/2020
	Skanor (SE)	20 – 23/11/2020
	Kungsholmsfort (SE)	20 – 23/11/2020

Secondly, we acknowledge that the TWL validation represents a key limitation in the study and this has been included in the discussion. Lines 525 – 544: "Moreover, the TWL validation as a composite variable is inherently challenging, particularly at the large scale. While SWL can be validated with tide gauges, wave conditions can be validated with buoys. However, the increase in coastal water level as a result of the wave contribution is not captured by wave buoys. An alternative is to use instruments such as pressure sensors or ADCPs that capture local observations. Even these, however, are limited by the need for corrections, regular maintenance, and their localized nature. The lack of observational data and instruments allowing for the validation of the TWL has led previous studies to validate TWL indirectly, by validating each component individually. This strategy was adopted in China (Liu et al., 2025), New Zealand (Dalinghaus et al., 2025), and globally (Vousdoukas, Mentaschi, et al., 2018), for example. Therefore, given the successful validation presented for the TPXO, ROMS, offshore wave, and nearshore wave databases as well as the foreshore slope, we consider the TWL hindcast to be robustly validated for the purposes of this study. Nevertheless, a TWL validation was presented by comparing historical storms identified by previous studies. Even though most of the observed storms were referenced to tide gauges, which do not capture wave contribution, our analysis showed that the extreme events analyzed were well represented by the TWL hindcast. Given the lack of data to validate TWL, the use of tide gauge information to validate TWL is common practice in the literature (Kirezci et al., 2020; Le Gal et al., 2023; Treu et al., 2024; Wing et al., 2024). Similar to the results found here, when comparing validation results with and without wave contribution, Kirezci et al. (2020) detected more accurate estimates when including wave setup. The authors debate whether this is an underestimation of the SWL or the influence of waves given that these results are even more pronounced under extreme than average conditions. Even though the reasons behind these results remain inconclusive, it is clear that when considering the wave contribution, extreme events are better captured by TWL reconstructions."

My additional minor suggestions for ameliorating the manuscript are listed here:

- Line 8: … the primary driver of coastal **flooding**.

Modified. Lines 7 – 8: "A key component of these assessments is the spatial characterization of total water level (TWL), the primary driver of coastal flooding."

- Lines 15-17: Please reformulate these sentences. It is not the Atlantic coast affected by the wave dataset, but the TWL estimation along the Atlantic coast that is affected by the wave dataset. And so on.

Reformulated. Lines 14 – 15: "The tide-dominated Atlantic coast is where the TWL is most affected by the wave dataset."

- Line 18: "no regions have extreme events dominated by wave setup". Maybe this is a problem of the adopted approach, which is not able to properly assess the wave setup contribution.

Given the difficulty of validating both the wave setup and the TWL at this scale, statements such as the one mentioned should be treated with caution. However, by validating the marine databases and the foreshore slope, we can safely assume that the wave setup and the TWL variables are reasonably well represented.

The sentence has been clarified. Lines 17 – 19: "A classification of TWL extremes revealed that no regions have extreme events dominated by wave setup when compared to the remaining components, while those dominated by tides show the highest return levels."

A mention to this issue has been included in the limitations of the study. Lines 505 – 506: "Notably, this may also influence the relative contributions of the TWL components as well."

- Line 154: Why do you use the nearest grid point instead of using an interpolation method?

The adoption of nearest grid point instead of interpolation was adopted to lower computational time required when selecting the astronomical tide and storm surge data. We highlight that the resolution of the different databases is higher than the resolution at which such variables usually vary. The astronomical tide TPXO spatial resolution is around 3.5 km and the storm surge spatial resolution is 10 – 15 km. Meanwhile, astronomical tide and storm surge processes present a spatial variability of approximately hundreds to thousands of kilometers (Woodworth et al., 2019).

Nevertheless, a sensitivity analysis was performed by assessing how the data selection with nearest grid point and interpolation influence the P95 results of astronomical tide, storm surge, and TWL in a selection of CTPs. Figure R1 presents the results for three locations, one in each of the three European regions identified in the study. The largest differences in astronomical tide and TWL are found in the Baltic Sea, while the largest differences in storm surge are found along the Atlantic coast. However, the statistical metrics analyzed indicate that such differences are too small to play a relevant role in the estimation of extreme values as R^2 range from 0.88 to 1, RMSE from 0 to 0.002 m, and bias from -0.001 to 0.001. Therefore, the different results obtained with both approaches do not justify the use of interpolation methods in this case, particularly at this scale when there are 51,010 points in which the method would have to be applied.

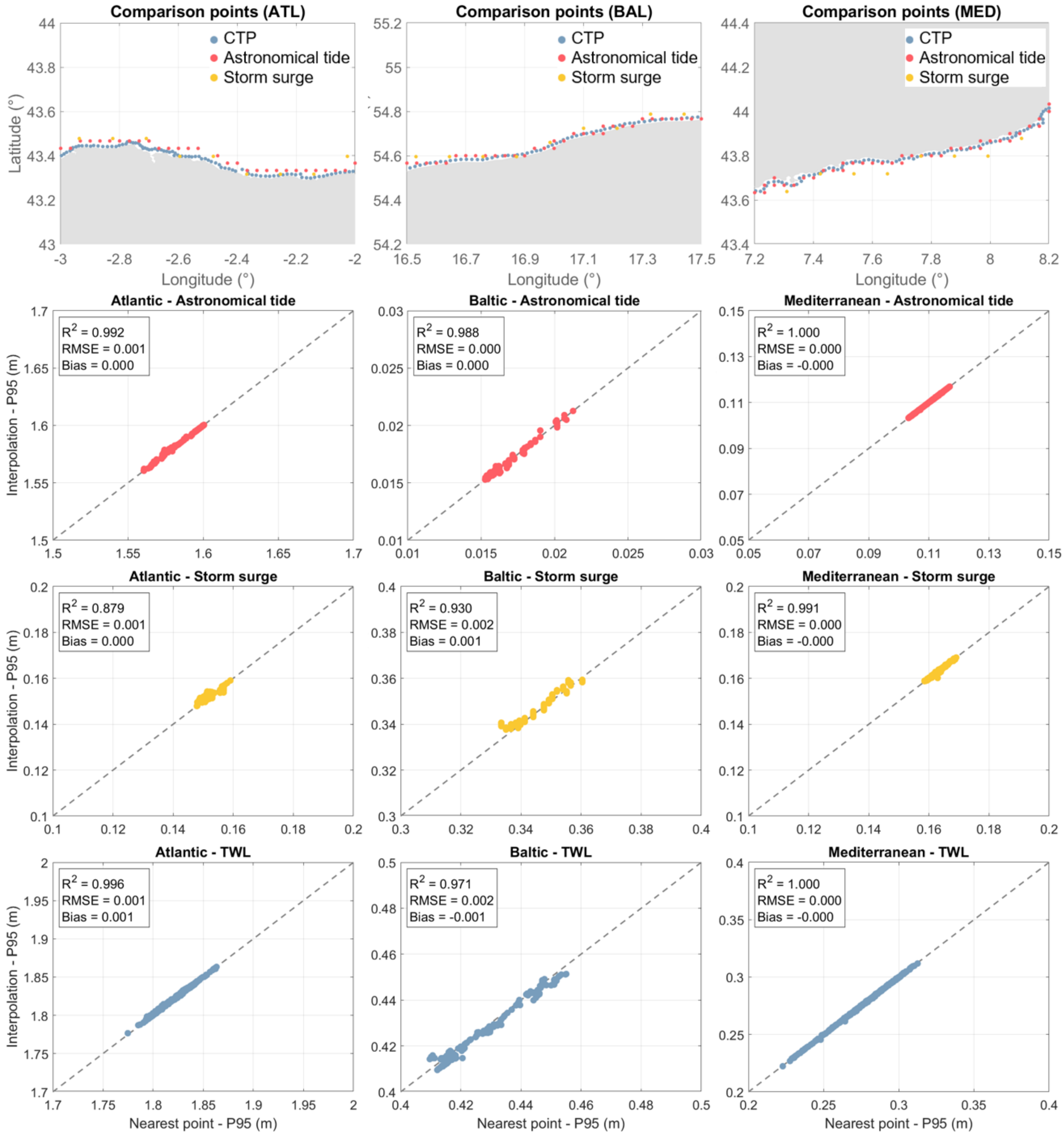

Figure R1 Sensitivity analysis of selection of corresponding marine database information based on P95. Examples are shown for smaller areas located in each of the three European regions identified in the study: Atlantic coast (left column), Baltic Sea (middle column), and Mediterranean Sea (right column).

- Line 176: You should justify the use of a 72h time interval for separating different extreme events.

A sensitivity analysis on the declustering time (minimum interval) adopted in POT has been included in the Supplementary Material:

"Sensitivity analysis: POT interval

The interval used when declustering events to guarantee their independence can be spatially constant or variable, depending on local climatic characteristics. For example, for waves in the

Atlantic Ocean, previous evidence suggests that a minimum interval of at least 3 days (or 72h) is required, even though the Poisson assumption is well satisfied with intervals up to 6 days (Méndez et al., 2006). Meanwhile, a storm surge study along the German Bight showed that the influence of the declustering time on the return level outcomes is minimal when adopting POT, compared to AM (Arns et al., 2013). Ultimately, in a POT analysis, the independence criterion should reflect the geophysical origin and duration of the extreme events under study. On the one hand, when defining a spatially constant value, the interval is well-established with 72h in storm surge studies (Caspers et al., 2025; Dullaart et al., 2023; MacPherson et al., 2019; Pupić Vurilj et al., 2025; Vousdoukas, Voukouvalas, Annunziato, et al., 2016) although there are applications ranging from 1.5 days (Arns et al., 2013) to 6 days (Martín et al., 2024). In wave storm studies, a 48h interval has been commonly adopted (Lobeto et al., 2024) as well as a 12 – 24h calm period (Martzikos et al., 2023; Martzikos et al., 2021). On the other hand, spatially variable intervals have not been as explored, especially in large-scale studies. Similar to the adoption of a spatially variable POT threshold, the minimum interval would likely affect the estimation of return values as well as coastal flooding projections. However, the development of a methodology to define such heterogeneous interval values is out of the scope of the present study.

Following the decision of a spatially constant interval, a sensitivity analysis was performed to attest for the robustness of the adoption of the 72h instead of 48h and 96h. Figure S8 presents the EVA results obtained considering the different possibilities of declustering time. The 100-yr TWL variability resulting from the different POT minimum intervals tested is low compared to the other sensitivity analyses performed in this study (Figure S8a). The 95th confidence intervals of the 100-yr TWL indicate that 96h is the most uncertain option (Figure S8c), while 48h presents the lowest uncertainty (Figure S8d). However, the Anderson-Darling test analysis, indicates that the 72h interval leads to the highest proportion of CTPs adequately adjusting to the exponential fit (Figure S8b). Additionally, out of the 3,298 CTPs in which the 72h interval does not result in robust distributions, 2,098 CTPs rejected the null-hypothesis with all three POT minimum intervals tested.

The results indicate that, among the different steps of the methodology tested, the independence time between events has the least influence on the estimated return periods. Yet, this does not imply that the POT interval is irrelevant. Our analysis shows that it is still relevant, though less so than other methodological choices."

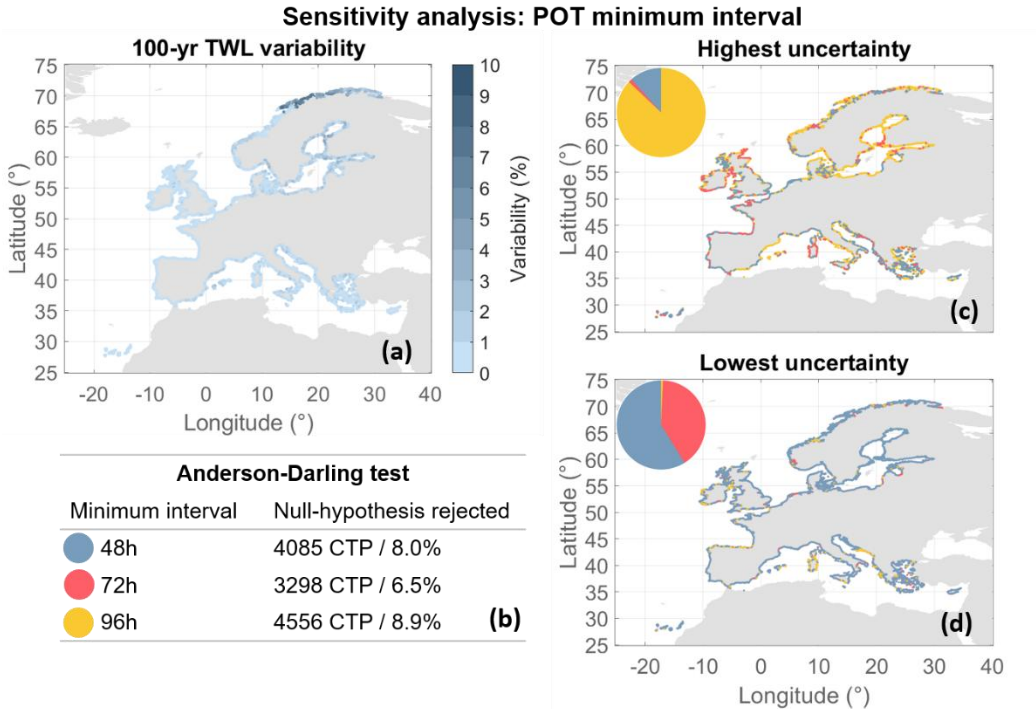

Figure S8 Sensitivity analysis of the POT interval selected (48h vs. 72h vs. 96h). Spatial variability 100-yr TWL variability (a). Anderson-Darling test results when adopting the different intervals (b). A rejected null hypothesis indicates that the sample does not fit the exponential fit. Uncertainty of results indicating which interval leads to the highest uncertainties (c) and the lowest uncertainties (d) of return levels, based on the 95th confidence level.

- Lines 201-203: the flood model must be described together with the modelling systems.

Clarified. Lines 234 – 236: "The flood model applied was a simple GIS-based bathtub approach, in which any area below a certain water level is considered flooded, provided it is hydraulically connected to the sea. This method was only used as a first-order approximation of the flooded area for comparative purposes in the sensitivity analyses."

- I suggest merging sections 2.6.1 and 2.6.2 and creating a table listing all TWL reconstruction approaches.

Indeed, both analyses focus on the wave contribution to the TWL. However, the reason for separating these two sections lies in their different objectives and in the type of analyses performed. Section 2.7.1 describes the sensitivity analysis regarding whether or not to include the wave contribution, as well as the way in which waves are represented (still water level vs. static wave setup vs. dynamic wave setup vs. wave runup). Section 2.7.2 goes a step further: once the static wave setup is selected as the preferred representation of the wave contribution, it describes the sensitivity analysis of different elements (wave dataset, foreshore slope approximation, static wave setup formulation). The influence of these elements on the alternative dynamic wave setup and wave runup was not considered. Therefore, by merging both analyses, we believe there could blur this distinction and lead to confusion in the interpretation of the results. For these reasons, we consider that keeping Sections 2.7.1 and 2.7.2 separate improves the clarity of the manuscript.

- Equations 5 to 7 should be moved to section 2.3.

The reason for presenting Equations 5 to 7 separately is that section 2.3 introduces the methodology proposed in this study to reconstruct the TWL, while the mentioned equations describe alternative approaches for representing the wave contribution. Therefore, including equations 5 to 7 alongside equations 2 to 4 could be misleading, as they are not part of the main methodology adopted in this work.

- Line 244: The location of the test points should be shown in the manuscript and not in the supplementary material.

The locations of the remaining test points were included in Fig 2g.

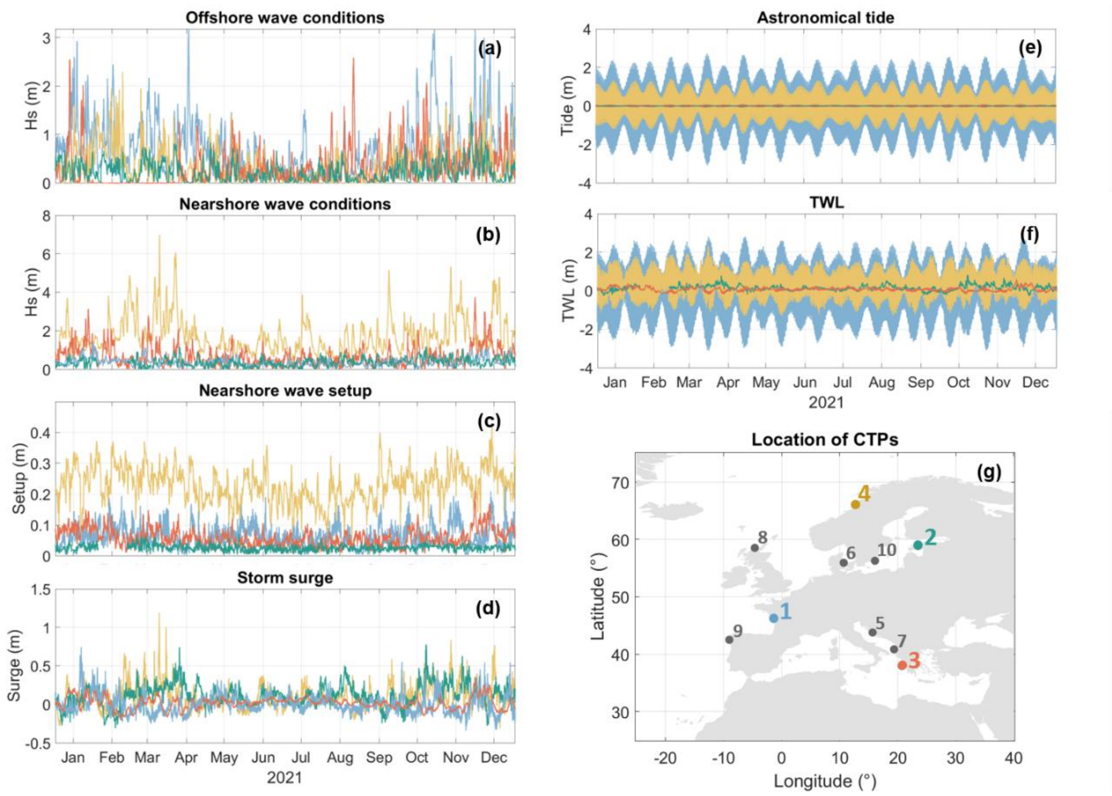

Figure 2 Hindcast time series referred to 2021 for the following variables: offshore significant wave height (a), nearshore significant wave height (b), nearshore wave setup (c), storm surge (d), astronomical tide (e), and TWL (f). Location of the CTPs used as examples (g). Time series are represented in the same colors as the corresponding CTP. Gray CTPs represent the remaining six test points selected with K-means.

- Line 246: More details on the clustering techniques must be provided.

Details have been included in the Supplementary Material:

"Selection of test points with K-means

To facilitate the development of the methodology, a set of representative (test) points were selected using a K-means clustering technique (Camus et al., 2011) applied to the mean relative contributions of the TWL components. These points were only used in preliminary tests in EVA and served as graphical examples throughout the study.

Figure S7 displays the results of the clustering analysis, which does not isolate extreme events nor does it consider geographical location during its application. The relative contributions of the TWL components reflect the spatial distribution of their individual patterns (see Section 3.2.2). When adopting a 10-cluster selection (Figure S7a), it is possible to see a grouping of CTPs per European regions along the Atlantic coast (clusters 1, 4, 8, and 9), the Baltic Sea (clusters 2, 6, and 10), and the Mediterranean Sea (clusters 3, 5, and 7). Even though similar grouping patterns are observed with 100 clusters (Figure S7b) it is noticeable the smaller number of clusters in the Baltic Sea, in which only 10 clusters appear. Oppositely, the Atlantic coast presents 43 dominating clusters and the Mediterranean Sea, 47. The amount of clusters observed in each region provides an estimate of dominance, although clusters that are dominant in one region may still appear elsewhere at smaller frequencies. These results indicate that the Baltic Sea might be more homogeneous regarding the spatial variability of the marine climate drivers composing TWL than the two remaining regions."

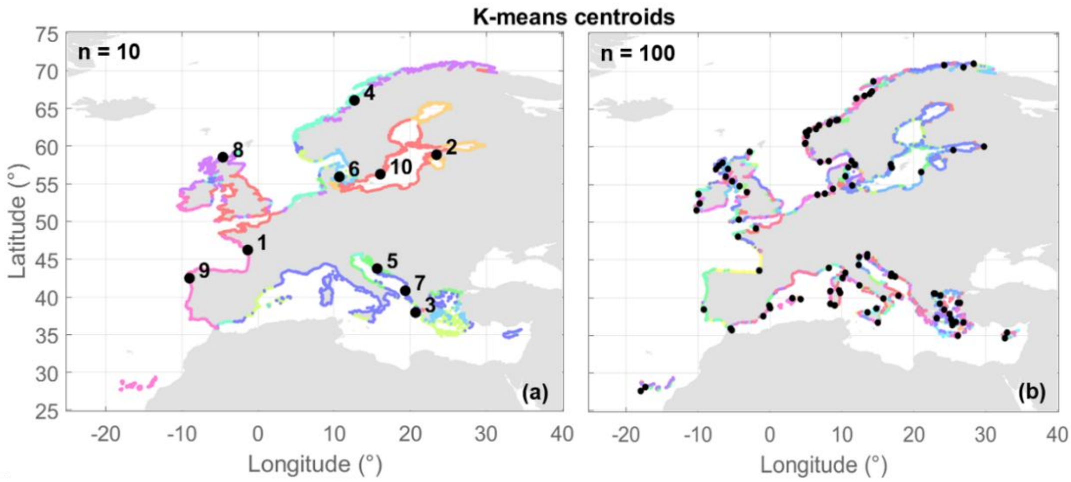

Figure S7 K-means clustering based on the average relative contributions of TWL components per CTP. (a) Selection of 10 clusters with their respective centroids highlighted in black dots. (b) Selection of 100 clusters with their respective centroids highlighted in black dots.

- Figure 3 and lines 267-272: I suggest removing panel (e) and the mentioned lines since there is no reason to mix the different approaches.

Panel e has been removed. Figure 3 is updated as follows:

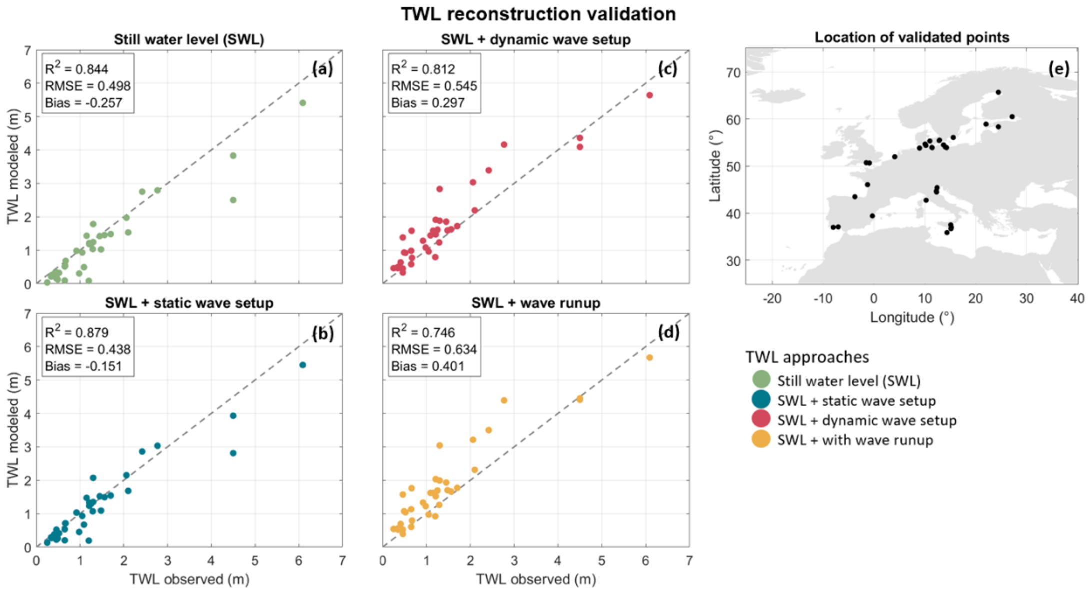

Figure 3 TWL reconstruction validation based on TWL with still water level (a), static wave setup (b), dynamic wave setup (c), and wave runup (d) when comparing historical coastal storms observed in different points across the study area (e), according to previous studies.

- Line 293: More details on ice modelling must be provided.

We have added more details in the Methods section regarding how the presence of sea ice was considered.

Details have been provided. Lines 169 – 176: "Since the European coastline reaches high latitudes, the presence of sea ice may affect the impact of the different coastal dynamics considered in this study (e.g., in the Baltic Sea). As all atmospheric-driven dynamics originate from a common forcing, ERA5, the sea ice cover provided by this reanalysis was used to account for the presence or absence of ice. This variable represents the fraction of each grid cell covered by ice. The temporal evolution of ice cover was extracted at the nodes closest to the coastal points. Ice presence was assumed when the coverage exceeded a fraction of 0.5, following the simple ice-blocking scheme used in wave propagation models such as WW3 (e.g., Kumar et al., 2025). Accordingly, when ice was present, the corresponding data at CTPs were excluded from the analysis, as neither waves nor tides effectively reach the coast under such conditions."

- Lines 352-353: It is not clear what the authors meant by "average conditions." Which supplementary material should provide insights into this?

Clarified. Lines 611 – 612: “The results show that even in parts of the Baltic and Mediterranean Seas the tide reaches a contribution of more than 20%, being particularly relevant under average conditions (see Supplementary Fig. S16).”

- Figure 6: Tide, Storm surge and Wave setup should also be reported in panels g, h and i.

Figure 6 updated.

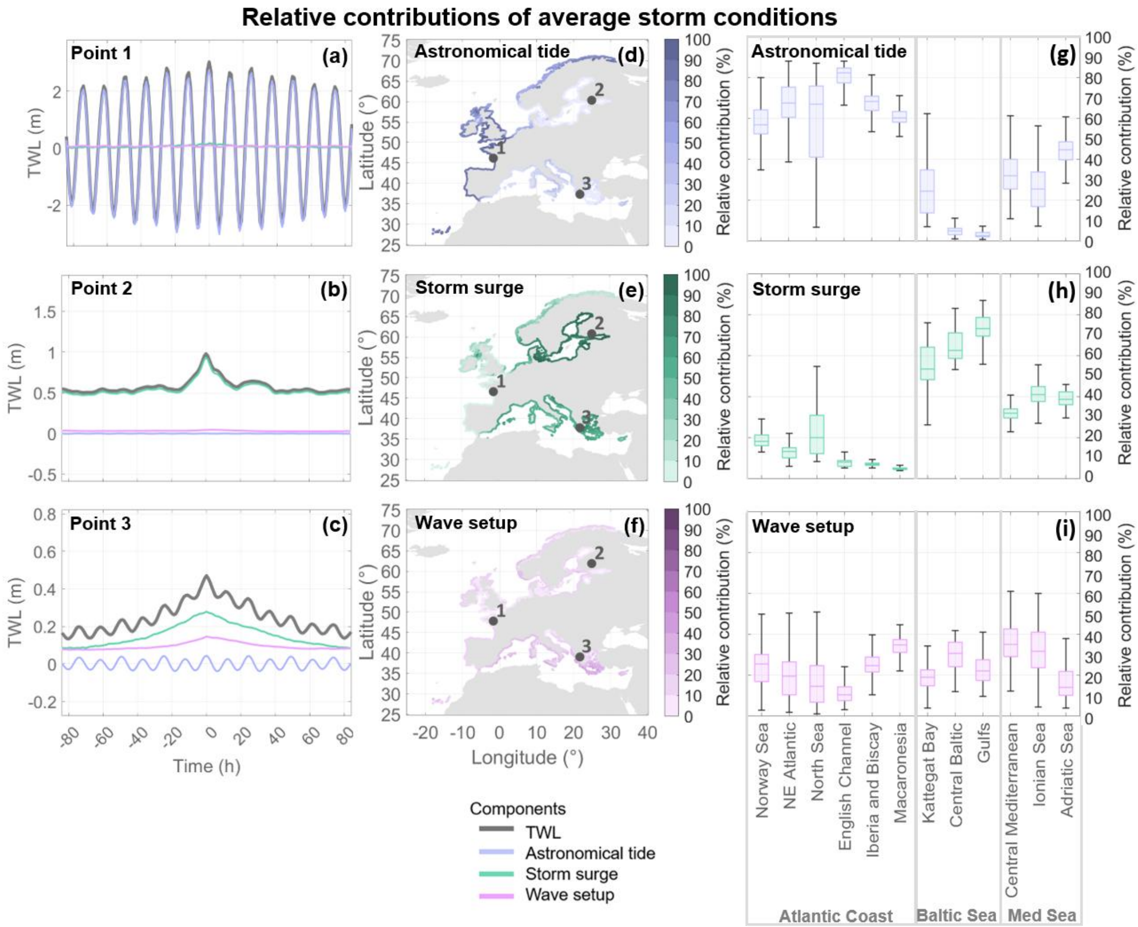

Figure 6 Examples of individual CTP with average storm conditions per TWL component (a, b, c). Spatial distribution of mean relative contributions of astronomical tide (d), storm surge (e), and wave setup (f) under extreme conditions. Dispersion of relative contributions of TWL components per European basin, grouped per European region (g, h, i).

- Figure 9: Doesn't the POT threshold with λ = 1 equal the Annual Maxima (AM) method?

Indeed, the selection of exceedances with POT threshold with λ = 1 and with Annual Maxima (AM) result in the same sample sizes. However, the actual events selected are not necessarily the same. AM selects events based on their temporal distribution, whereas POT selects events based on their magnitude. While AM assumes that every year presents an extreme event, POT is more flexible by considering that a specific year might have more than one extreme event and other might have none.

Nevertheless, the reviewer raises an interesting question for which a sensitivity analysis was performed. A comparison between AM and POT with $\lambda = 1$ was conducted, as well as with $\lambda = 2$, given that this is proposed methodology in the study. Figure R2 presents the comparison of the resulting exceedance threshold (i.e., lowest exceedance sampled), the percentage of extreme events sampled in both strategies being compared, and the confidence intervals of the corresponding 100-yr TWL. Since the proposed methodology adopts the 2-parameter exponential fit, comparison with AM were performed with the 2-parameter Gumbel distribution. Both comparisons indicate that AM does not select the most extreme events in a time series as it prioritizes a constant temporal distribution of exceedances, rather than their magnitudes.

When comparing POT $\lambda = 1$ with AM, it is confirmed that despite having the same sample size, not all AM exceedances are included as the most extreme ones according to POT. Thus confirming, that not every year presents an extreme event. On average, 62.3% of the AM exceedances are identified by POT $\lambda = 1$. Comparing the corresponding threshold, AM displays a slightly lower value, although the relative confidence intervals to a 100-yr TWL is smaller which would indicate lower uncertainties. When comparing the methodology proposed in the study (POT $\lambda = 2$) with AM, there is a higher percentage of overlapped exceedances sampled (83.1%, on average) and the corresponding threshold is also more similar. Meanwhile, the confidence interval for POT $\lambda = 2$ is generally smaller than AM, indicating lower uncertainty of return level estimates.

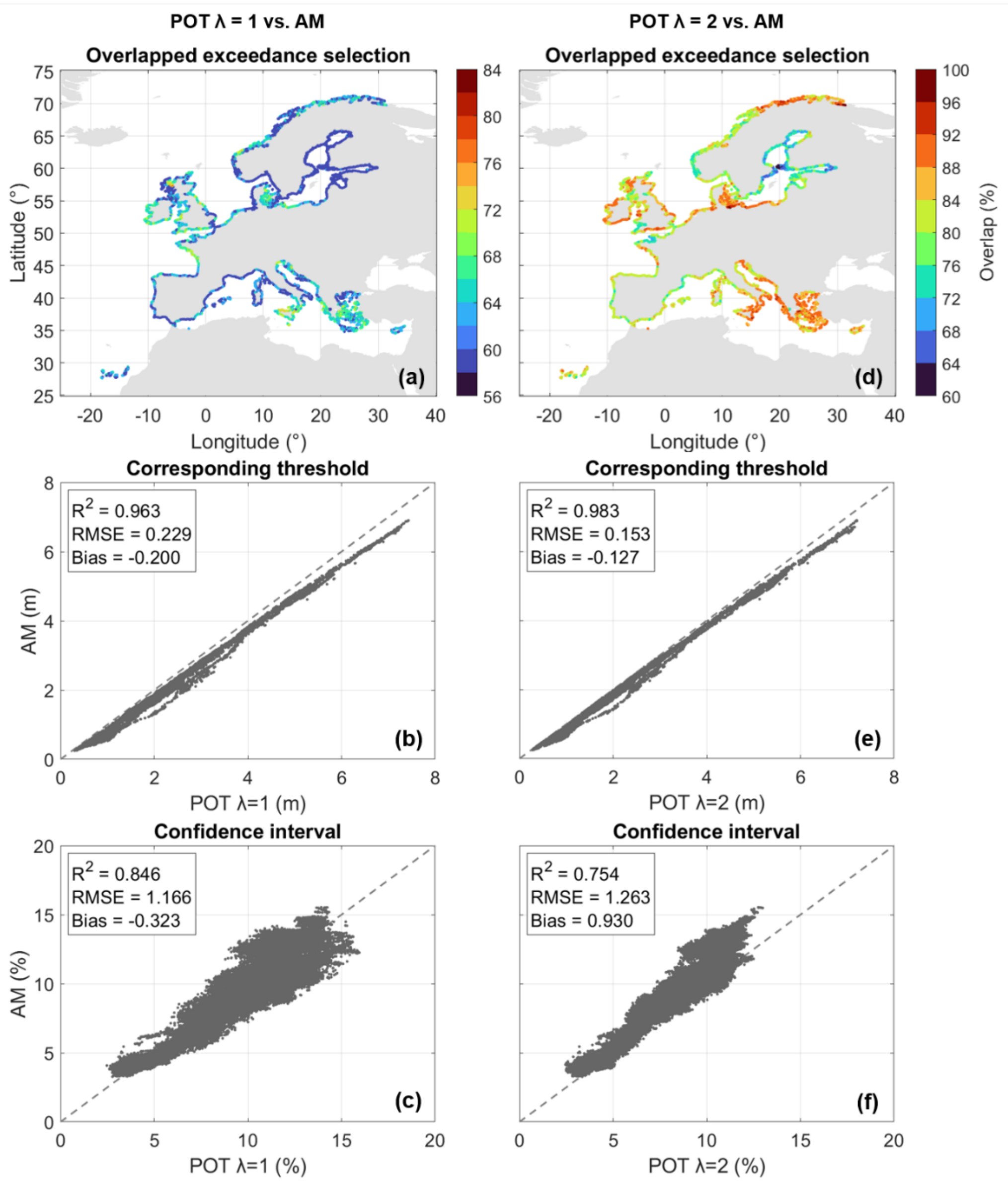

Figure R2 Comparison of POT and AM methods when selecting exceedances in EVA. Left column: POT λ = 1 vs. AM. Right column: POT λ = 2 vs. AM. (a, d) Overlapped exceedances sampled. (b, e) Corresponding threshold, or lowest exceedance sampled. (c, f) Relative confidence intervals for a 100-yr TWL event.

- Lines 499-509: These sentences summarise the results and would be more appropriately placed in the Conclusions section rather than in the Discussion.

Sentences moved to the newly added Conclusions section. Lines 617 – 624: "Currently, there remains a strong need to provide coastal flooding maps at large geographic scales. TWL is the primary input for coastal flood modeling and must be characterized appropriately for the scale of analysis. Without a proper assessment of TWL, estimating extreme conditions becomes more uncertain (Rohmer et al., 2021; Toimil et al., 2020). To address this need, the methodology presented incorporates spatial variability typical of large-scale coastal studies. This was achieved by using spatially and temporally variable foreshore slopes during TWL reconstruction and by applying spatially variable thresholds in the POT method. An analysis of relative contributions helped to interpret extreme TWL behavior along the European coastline. The characterization of three macro-regions (Atlantic coast, Baltic Sea, and Mediterranean Sea) supported the understanding of the degrees of uncertainty observed across different regions at distinct steps of the methodology."

- Lines 524-525: This sentence is not clear. Please reformulate it.

Sentence reformulated. Lines 581 – 584: "Comparing our thresholds to previous studies, we found that 77.1% of the study area exhibits more than 12 events per year when the threshold of P97 used by Le Gal et al. (2023) is adopted. Meanwhile, if the thresholds of P98 and P98.5 used by Kirezci et al. (2020) and Paprotny et al. (2016) are adopted instead, this value decreases to 61.9% and 49.8%, respectively."

- Line 540: 111 km resolution?

Clarified. Lines 592 – 593: "Additionally, their 111 km spatial resolution excluded marginal seas, where we observed the highest storm surge contributions."

References

Anthony, E. J., & Héquette, A. (2007). The grain-size characterisation of coastal sand from the Somme estuary to Belgium: Sediment sorting processes and mixing in a tide- and storm-dominated setting. *Sedimentary Geology*, *202*(3), 369–382. https://doi.org/10.1016/j.sedgeo.2007.03.022

Arns, A., Wahl, T., Haigh, I. D., Jensen, J., & Pattiaratchi, C. (2013). Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise. *Coastal Engineering*, *81*, 51–66. https://doi.org/10.1016/j.coastaleng.2013.07.003

Arns, A., Wahl, T., Wolff, C., Vafeidis, A. T., Haigh, I. D., Woodworth, P., Niehüser, S., & Jensen, J. (2020). Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-15752-5

Bertin, X., Bruneau, N., Breilh, J. F., Fortunato, A. B., & Karpytchev, M. (2012). Importance of wave age and resonance in storm surges: The case Xynthia, Bay of Biscay. *Ocean Modelling*, *42*, 16–30. https://doi.org/10.1016/j.ocemod.2011.11.001

Bevacqua, E., Vousdoukas, M. I., Zappa, G., Hodges, K., Shepherd, T. G., Maraun, D., Mentaschi, L., & Feyen, L. (2020). More meteorological events that drive compound coastal flooding are projected under climate change. *Communications Earth and Environment*, *1*(1). https://doi.org/10.1038/s43247-020-00044-z

Bezak, N., Brilly, M., & Šraj, M. (2014). Comparaison entre les méthodes de dépassement de seuil et du maximum annuel pour les analyses de fréquence des crues. *Hydrological Sciences Journal*, *59*(5), 959–977. https://doi.org/10.1080/02626667.2013.831174

Björkqvist, J.-V., Kanarik, H., Tuomi, L., Niskanen, L., & Kankainen, M. (2024). Event-based wave statistics for the Baltic Sea. *State of the Planet*, *4-osr8*, 1–8. https://doi.org/10.5194/sp-4-osr8-10-2024

Breilh, J. F., Chaumillon, E., Bertin, X., & Gravelle, M. (2013). Assessment of static flood modeling techniques: Application to contrasting marshes flooded during Xynthia (western France). *Natural Hazards and Earth System Sciences*, *13*(6), 1595–1612. https://doi.org/10.5194/nhess-13-1595-2013

Cabrita, P., Montes, J., Duo, E., Brunetta, R., & Ciavola, P. (2024). The Role of Different Total Water Level Definitions in Coastal Flood Modelling on a Low-Elevation Dune System. *Journal of Marine Science and Engineering*, *12*(6), 1003. https://doi.org/10.3390/jmse12061003

Calafat, F. M., & Marcos, M. (2020). Probabilistic reanalysis of storm surge extremes in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(4), 1877–1883. https://doi.org/10.1073/pnas.1913049117

Camus, P., Mendez, F. J., Medina, R., & Cofiño, A. S. (2011). Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, *58*(6), 453–462. https://doi.org/10.1016/j.coastaleng.2011.02.003

Carneiro-Barros, J. E., Majidi, A. G., Plomaritis, T., Fazeres-Ferradosa, T., Rosa-Santos, P., & Taveira-Pinto, F. (2025). Coastal Flooding Hazards in Northern Portugal: A Practical Large-Scale Evaluation of Total Water Levels and Swash Regimes. *Water (Switzerland)*, *17*(10). https://doi.org/10.3390/w17101478

Caspers, J. J., Kindermann, P. E., Rongen, G. W. F., & Geerse, C. P. M. (2025). Storm surge hydrographs and characteristics along the Dutch coast. An analysis from simulation data. *Coastal Engineering*, *201*(May). https://doi.org/10.1016/j.coastaleng.2025.104776

Chaigneau, A. A., Reffray, G., Voldoire, A., & Melet, A. (2022). IBI-CCS: A regional high-resolution model to simulate sea level in western Europe. *Geoscientific Model Development*, *15*(5), 2035–2062. https://doi.org/10.5194/gmd-15-2035-2022

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer series in statistics: London (UK).

Collings, T. P., Quinn, N. D., Haigh, I. D., Green, J., Probyn, I., Wilkinson, H., Muis, S., Sweet, W. V., & Bates, P. D. (2024). Global application of a regional frequency analysis to extreme sea levels. *Natural Hazards and Earth System Sciences*, *24*(7), 2403–2423. https://doi.org/10.5194/nhess-24-2403-2024

Dalinghaus, C., Coco, G., & Higuera, P. (2025). Assessing ToTal Water Level Uncertainties Using Global Sensitivity Analysis. *Journal of Geophysical Research: Oceans*, *130*, 460–466.

https://doi.org/10.1029/2025JC023011

DEFRA, D. for E. F. and R. A. (2023). *LIDAR Composite Digital Terrain Model (DTM)*. https://environment.data.gov.uk/dataset/13787b9a-26a4-4775-8523-806d13af58fc

Dodet, G., Melet, A., Ardhuin, F., Bertin, X., Idier, D., & Almar, R. (2019). The Contribution of Wind-Generated Waves to Coastal Sea-Level Changes. In *Surveys in Geophysics* (Vol. 40, Issue 6). Springer Netherlands. https://doi.org/10.1007/s10712-019-09557-5

Dullaart, J. C. M., Muis, S., De Moel, H., Ward, P. J., Eilander, D., & Aerts, J. C. J. H. (2023). Enabling dynamic modelling of coastal flooding by defining storm tide hydrographs. *Natural Hazards and Earth System Sciences*, *23*(5), 1847–1862. https://doi.org/10.5194/nhess-23-1847-2023

Duo, E., Sanuy, M., Jiménez, J. A., & Ciavola, P. (2020). How good are symmetric triangular synthetic storms to represent real events for coastal hazard modelling. *Coastal Engineering*, *159*(May). https://doi.org/10.1016/j.coastaleng.2020.103728

Egon, A., Farrell, E., Iglesias, G., & Nash, S. (2025). Measurement and Modelling of Beach Response to Storm Waves: A Case Study of Brandon Bay, Ireland. *Coasts*, *5*(3), 1–24. https://doi.org/10.3390/coasts5030032

EUROSION. (2004). *Living with Coastal Erosion in Europe: Sediment and Space for Sustainability. Part-1 Major Findings and Policy Recommendations of the EUROSION Project. Guidelines for implementing local information systems dedicated to coastal erosion management. Service* .

Fernández-Montblanc, T., Vousdoukas, M. I., Ciavola, P., Voukouvalas, E., Mentaschi, L., Breyiannis, G., Feyen, L., & Salamon, P. (2019). Towards robust pan-European storm surge forecasting. *Ocean Modelling*, *133*(May 2018), 129–144. https://doi.org/10.1016/j.ocemod.2018.12.001

Geoportale Emilia-Romagna, R.-R. (2014). *DTM 5x5*. https://geoportale.regione.emilia-romagna.it/catalogo/dati-cartografici/altimetria/layer-2

Guza, R. T., & Thornton, E. B. (1981). Wave set-up on a natural beach. *Journal of Geophysical Research*, *86*(C5), 4133–4137. https://doi.org/10.1029/JC086iC05p04133

Haigh, I. D., Wijeratne, E. M. S., MacPherson, L. R., Pattiaratchi, C. B., Mason, M. S., Crompton, R. P., & George, S. (2014). Estimating present day extreme water level exceedance probabilities around the coastline of Australia: Tides, extra-tropical storm surges and mean sea level. *Climate Dynamics*, *42*(1–2), 121–138. https://doi.org/10.1007/s00382-012-1652-1

Harley, M. (2017). Coastal Storm Definition. In *Coastal Storms: Processes and Impacts* (Issue April 2017). https://doi.org/10.1002/9781118937099.ch1

Hinkel, J., Feyen, L., Hemer, M., Le Cozannet, G., Lincke, D., Marcos, M., Mentaschi, L., Merkens, J. L., de Moel, H., Muis, S., Nicholls, R. J., Vafeidis, A. T., van de Wal, R. S. W., Vousdoukas, M. I., Wahl, T., Ward, P. J., & Wolff, C. (2021). Uncertainty and Bias in Global to Regional Scale Assessments of Current and Future Coastal Flood Risk. *Earth's Future*, *9*(7), 1–28. https://doi.org/10.1029/2020EF001882

Horn, D. P., & Walton, S. M. (2007). Spatial and temporal variations of sediment size on a mixed sand and gravel beach. *Sedimentary Geology*, *202*(3), 509–528. https://doi.org/10.1016/j.sedgeo.2007.03.023

Idier, D., Bertin, X., Thompson, P., & Pickering, M. D. (2019). Interactions Between Mean Sea Level, Tide, Surge, Waves and Flooding: Mechanisms and Contributions to Sea Level Variations at the Coast. *Surveys in Geophysics*, *40*(6), 1603–1630. https://doi.org/10.1007/s10712-019-09549-5

IGN, I. G. N. (2019). *Modelo Digitial del Terreno (5 metros) de España*.

IGN, I. N. de L. G. et F. (2021). *RGE Alti*. https://geoservices.ign.fr/rgealti

IHCantabria, I. de H. de C. (2021). *Elaboración de la evaluación de impactos y riesgos de inundación y erosión debidos al cambio climático en la costa murciana*.

Irazoqui Apecechea, M., Melet, A., & Armaroli, C. (2023). Towards a pan-European coastal flood awareness system: Skill of extreme sea-level forecasts from the Copernicus Marine Service . In *Frontiers in Marine Science* (Vol. 9).

Irazoqui Apecechea, M., Melet, A., Reffray, G., & Thieblemont, R. (2025). Impact of Dynamical Downscaling on Sea Level Projections in the North-East Atlantic Ocean and Regional Seas. *Journal of Geophysical Research: Oceans*, *130*(10). https://doi.org/10.1029/2025JC022737

Jarmalavičius, D., Žilinskas, G., Pupienis, D., & Janušaitė, R. (2025). Resilience of beach morphometric characteristics on decadal time scale: a case study from the Lithuanian Baltic Sea. *Oceanologia, Polish Academy of Sciences*, *67*(3). https://doi.org/https://doi.org/10.5697/KSMY5385

Kiesel, J., Lorenz, M., König, M., Gräwe, U., & Vafeidis, A. T. (2023). Regional assessment of extreme sea levels and associated coastal flooding along the German Baltic Sea coast. *Natural Hazards and Earth System Sciences*, *23*(9), 2961–2985. https://doi.org/10.5194/nhess-23-2961-2023

Kirezci, E., Young, I. R., Ranasinghe, R., Muis, S., Nicholls, R. J., Lincke, D., & Hinkel, J. (2020). Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st Century. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-67736-6

Koks, E. E., Le Bars, D., Essenfelder, A. H., Nirandjan, S., & Sayers, P. (2023). The impacts of coastal flooding and sea level rise on critical infrastructure: a novel storyline approach. *Sustainable and Resilient Infrastructure*, *8*(sup1), 237–261. https://doi.org/10.1080/23789689.2022.2142741

Kumar, R., Lemos, G., Semedo, A., & Li, J. G. (2025). A global high-resolution CMIP6 ensemble of wave climate simulations and projections using a coastal multigrid: Configuration and performance evaluation. *Ocean Modelling*, *197*(March), 102566. https://doi.org/10.1016/j.ocemod.2025.102566

Le Gal, M., Fernández-Montblanc, T., Duo, E., Montes Perez, J., Cabrita, P., Souto Ceccon, P., Gastal, V., Ciavola, P., & Armaroli, C. (2023). A new European coastal flood database for low-medium intensity events. *Natural Hazards and Earth System Sciences*, *23*(11), 3585–3602. https://doi.org/10.5194/nhess-23-3585-2023

Lemos, G., Soares, P. M. M., Simões, R., Antunes, C., Bosnic, I., & Pinto, C. (2025). A physical climate storyline for the Hercules storm in Portugal: Extreme coastal flooding in southwestern Europe under a changing climate. *Science of the Total Environment*, *971*(March). https://doi.org/10.1016/j.scitotenv.2025.179050

Liu, P., Din, A. H., & Hamden, M. H. (2025). Reconstructing hourly coastal total sea levels and assessing current and future extreme sea levels threats to the Coast of China. *Scientific Reports*, 1–16. https://doi.org/10.1038/s41598-025-22493-2

Lobeto, H., Semedo, A., Lemos, G., Dastgheib, A., Menendez, M., Ranasinghe, R., & Bidlot, J.-R. (2024). Global coastal wave storminess. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-51420-0

Lorenz, M., Arns, A., & Gräwe, U. (2023). How Sea Level Rise May Hit You Through the Backdoor: Changing Extreme Water Levels in Shallow Coastal Lagoons. *Geophysical Research Letters*, *50*(21), 1–11. https://doi.org/10.1029/2023GL105512

MacPherson, L. R., Arns, A., Dangendorf, S., Vafeidis, A. T., & Jensen, J. (2019). A Stochastic Extreme Sea Level Model for the German Baltic Sea Coast. *Journal of Geophysical Research: Oceans*, *124*(3), 2054–2071. https://doi.org/10.1029/2018JC014718

Martín, A., Wahl, T., Enriquez, A. R., & Jane, R. (2024). Storm surge time series de-clustering using correlation analysis. *Weather and Climate Extremes*, *45*(April). https://doi.org/10.1016/j.wace.2024.100701

Martzikos, N., Malliouri, D., & Tsoukala, V. (2023). Shape investigation and probabilistic representation of coastal storms. Applications to Mykonos and Barcelona. *Applied Ocean Research*, *135*(April). https://doi.org/10.1016/j.apor.2023.103563

Martzikos, N. T., Prinos, P. E., Memos, C. D., & Tsoukala, V. K. (2021). Statistical analysis of Mediterranean coastal storms. *Oceanologia*, *63*(1), 133–148. https://doi.org/10.1016/j.oceano.2020.11.001

McGlashan, D. J., Duck, R. W., & Reid, C. T. (2004). The foreshore: Geographical implications of the three legal systems in Great Britain. *Area*, *36*(4), 338–347. https://doi.org/10.1111/j.0004-0894.2004.00235.x

Melet, A., Almar, R., Hemer, M., Le Cozannet, G., Meyssignac, B., & Ruggiero, P. (2020). Contribution of Wave Setup to Projected Coastal Sea Level Changes. *Journal of Geophysical Research: Oceans*, *125*(8). https://doi.org/10.1029/2020JC016078

Melet, A., Meyssignac, B., Almar, R., & Le Cozannet, G. (2018). Under-estimated wave contribution to coastal sea-level rise. *Nature Climate Change*, *8*(3), 234–239. https://doi.org/10.1038/s41558-018-0088-y

Méndez, F. J., Menéndez, M., Luceño, A., & Losada, I. J. (2006). Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *Journal of Geophysical Research: Oceans*, *111*(7), 1–13. https://doi.org/10.1029/2005JC003344

Merrifield, M. A., Genz, A. S., Kontoes, C. P., & Marra, J. J. (2013). Annual maximum water levels from tide gauges: Contributing factors and geographic patterns. *Journal of Geophysical Research: Oceans*, *118*(5), 2535–2546. https://doi.org/10.1002/jgrc.20173

Muis, S., Apecechea, M. I., Dullaart, J., de Lima Rego, J., Madsen, K. S., Su, J., Yan, K., & Verlaan, M. (2020). A High-Resolution Global Dataset of Extreme Sea Levels, Tides, and Storm Surges, Including Future Projections. *Frontiers in Marine Science*, *7*(April), 1–15. https://doi.org/10.3389/fmars.2020.00263

Nuyts, S., Farrell, E. J., Fennell, S., & Nash, S. (2024). An Assessment of the Role of the Timex Sampling Strategy on the Precision of Shoreline Detection Analysis. *Coasts*, *4*(2), 347–365. https://doi.org/10.3390/coasts4020018

Paprotny, D., Morales-Nápoles, O., & Nikulin, G. (2016). Extreme sea levels under present and future climate: A pan-European database. *E3S Web of Conferences*, *7*. https://doi.org/10.1051/e3sconf/20160702001

Parker, K., Erikson, L., Thomas, J., Nederhoff, K., Barnard, P., & Muis, S. (2023). Relative contributions of water-level components to extreme water levels along the US Southeast Atlantic Coast from a regional-scale water-level hindcast. *Natural Hazards*, *117*(3), 2219–2248. https://doi.org/10.1007/s11069-023-05939-6

PDOK, P. D. O. de K. (2023). *Current Elevation Model Netherlands (AHN)*. https://www.pdok.nl/

Perini, L., Calabrese, L., Lorito, S., & Luciani, P. (2015). Il rischio da mareggiata in Emilia-Romagna: l'evento del 5-6 Febbraio 2015. *Il Geologo Volume*, *53*(January), 8–17.

Plant, N. G., & Stockdon, H. F. (2015). How well can wave runup be predicted? Comment on Laudier et al. (2011) and Stockdon et al. (2006). *Coastal Engineering*, *102*, 44–48. https://doi.org/10.1016/j.coastaleng.2015.05.001

Pupić Vurilj, M., Antolínez, J. A. Á., Muis, S., & Morales Napoles, O. (2025). Storm surge hydrographs from historical observations of sea level along the Dutch North Sea coast. *Natural Hazards*, *Ipcc 2021*. https://doi.org/10.1007/s11069-025-07351-8

Rohmer, J., Lincke, D., Hinkel, J., Le Cozannet, G., Lambert, E., & Vafeidis, A. T. (2021). Unravelling the importance of uncertainties in global-scale coastal flood risk assessments under sea level rise. *Water (Switzerland)*, *13*(6), 1–18. https://doi.org/10.3390/w13060774

Rueda, A., Vitousek, S., Camus, P., Tomas, A., Espejo, A., Losada, I. J., Barnard, P. L., Erikson, L. H., Ruggiero, P., Reguero, B. G., & Mendez, F. J. (2017). A global classification of coastal flood hazard climates associated with large-scale oceanographic forcing. *Scientific Reports*, *7*, 5038.

Scicchitano, G., Scardino, G., Monaco, C., Piscitelli, A., Milella, M., De Giosa, F., & Mastronuzzi, G. (2021). Comparing impact effects of common storms and Medicanes along the coast of south-eastern Sicily. *Marine Geology*, *439*(March), 106556. https://doi.org/10.1016/j.margeo.2021.106556

Serafin, K. A., Ruggiero, P., Barnard, P. L., & Stockdon, H. F. (2019). The influence of shelf bathymetry and beach topography on extreme total water levels: Linking large-scale changes of the wave climate to local coastal hazards. *Coastal Engineering*, *150*(January), 1–17. https://doi.org/10.1016/j.coastaleng.2019.03.012

Stockdon, H. F., Holman, R. A., Howd, P. A., & Sallenger, A. H. (2006). *Empirical parameterization of setup , swash , and runup*. *53*, 573–588. https://doi.org/10.1016/j.coastaleng.2005.12.005

Su, J., Murawski, J., Nielsen, J. W., & Madsen, K. S. (2024). Coinciding storm surge and wave setup: A regional assessment of sea level rise impact. *Ocean Engineering*, *305*(January). https://doi.org/10.1016/j.oceaneng.2024.117885

Sunamura, T. (1984). Quantitative Predictions of Beach-Face Slopes. *Bulletin of the Geological Society of America*, *95*(2), 242–245. https://doi.org/10.1130/0016-7606(1984)95<242:qpobs>2.0.co;2

Toimil, A., Losada, I. J., Nicholls, R. J., Dalrymple, R. A., & Stive, M. J. F. (2020). Addressing the challenges of climate change risks and adaptation in coastal areas: A review. *Coastal Engineering*, *156*(July 2019). https://doi.org/10.1016/j.coastaleng.2019.103611

Treu, S., Muis, S., Dangendorf, S., Wahl, T., Oelsmann, J., Heinicke, S., Frieler, K., & Mengel, M. (2024). Reconstruction of hourly coastal water levels and counterfactuals without sea level rise for impact attribution. *Earth System Science Data*, *16*(2), 1121–1136. https://doi.org/10.5194/essd-16-1121-2024

Vitousek, S., Barnard, P. L., Fletcher, C. H., Frazer, N., Erikson, L., & Storlazzi, C. D. (2017). Doubling of coastal flooding frequency within decades due to sea-level rise. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-01362-7

Vousdoukas, M. I., Bouziotas, D., Giardino, A., Bouwer, L. M., Mentaschi, L., Voukouvalas, E., & Feyen, L. (2018). Understanding epistemic uncertainty in large-scale coastal flood risk assessment for present and future climates. *Natural Hazards and Earth System Sciences*, *18*(8), 2127–2142. https://doi.org/10.5194/nhess-18-2127-2018

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., & Feyen, L. (2018). Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nature Communications*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-04692-w

Vousdoukas, M. I., Voukouvalas, E., Annunziato, A., Giardino, A., & Feyen, L. (2016). Projections of extreme storm surge levels along Europe. *Climate Dynamics*, *47*(9–10), 3171–3190. https://doi.org/10.1007/s00382-016-3019-5

Vousdoukas, M. I., Voukouvalas, E., Mentaschi, L., Dottori, F., Giardino, A., Bouziotas, D., Bianchi, A., Salamon, P., & Feyen, L. (2016). Developments in large-scale coastal flood hazard mapping. *Natural Hazards and Earth System Sciences*, *16*(8), 1841–1853. https://doi.org/10.5194/nhess-16-1841-2016

Wadey, M. P., Nicholls, R. J., & Hutton, C. (2012). Coastal flooding in the solent: An integrated analysis of defences and inundation. *Water (Switzerland)*, *4*(2), 430–459. https://doi.org/10.3390/w4020430

Wing, O. E. J., Bates, P. D., Quinn, N. D., Savage, J. T. S., Uhe, P. F., Cooper, A., Collings, T. P., Addor, N., Lord, N. S., Hatchard, S., Hoch, J. M., Bates, J., Probyn, I., Himsworth, S., Rodríguez González, J., Brine, M. P., Wilkinson, H., Sampson, C. C., Smith, A. M., … Haigh, I. D. (2024). A 30 m Global Flood Inundation Model for Any Climate Scenario. *Water Resources Research*, *60*(8). https://doi.org/10.1029/2023WR036460

Wolski, T., & Wiśniewski, B. (2021). Characteristics and long-term variability of occurrences of storm surges in the baltic sea. *Atmosphere*, *12*(12). https://doi.org/10.3390/atmos12121679

Woodworth, P. L., Melet, A., Marcos, M., Ray, R. D., Wöppelmann, G., Sasaki, Y. N., Cirano, M., Hibbert, A., Huthnance, J. M., Monserrat, S., & Merrifield, M. A. (2019). Forcing Factors Affecting Sea Level Changes at the Coast. In *Surveys in Geophysics* (Vol. 40, Issue 6). Springer Netherlands. https://doi.org/10.1007/s10712-019-09531-1