

Reply to reviewer 1

We would like to thank the reviewer for the constructive comments. In the following we recall the comments and we reply to each of the comments in turn (in green).

1- The study excludes wave overtopping (Andernos seafront) and river/coincidence events in the Leyre delta, focusing only on overflow. This is fine, but in several places the text could make it clearer that results shouldn't be generalized beyond that scope. Please highlight the likely direction of bias caused by these omissions (e.g., underestimation near overtopping zones, possible CAT misranking in areas with fluvial–marine interaction). Mention these limitations again in the Discussion or Conclusion.

⇒ Yes we agree that the discussion on the replicability of the method should mention this point. We propose to add a sentence to underline it in the last paragraph of the discussion as follows: “However, the methodological approaches and the associated conclusions regarding solution accuracy may vary depending on the environmental context, notably in wave overtopping prone areas and estuarine environments that are not considered in this study.”

2- The study assumes stationary metocean conditions through the tidal cycle, supported by tests showing less than 10 cm difference in SSHmax and similar flood extents for eight events. This is useful, but the sample is small and specific to this site.

- Provide uncertainty estimates for the “<10 cm” result (e.g., confidence intervals) and, if any, show one case where dynamic effects matter, or clearly state that none were found.
- Discuss when this assumption could fail, for example, under rapidly changing winds or asymmetric surge conditions.

⇒ To detail this point we propose to complete table 2 with the min and max differences between dynamic and stationary simulations for the 8 reference points in the basin. This will highlight that the maximum differences occur in the northeast of the basin (where the effect of strong winds on the amplification of storm surge is the most important). Domingos, being one of the most intense storms recorded, is representative (as Klaus and Fabien) of the worst situations.

3- Page 14, Line 26: The validation uses only a few historical and pseudo-historical events. Please give the exact number of cases per CAT level and per municipality so readers can see how balanced the data are. Currently, only totals are given.

⇒ The distribution of the pseudo-historical events into the CAT is illustrated in Fig. 6 (orange points). However, the caption does not mention it and the figure 6 is not referenced at this stage of the text. So, we propose to better reference the figure and complete it with a table indicating the number of events considered per CAT and municipality.

4- You mainly use per-node MAE and an Hmax-class confusion matrix. Consider adding flood extent metrics such as IoU/Jaccard or F1 (wet/dry) to complement these results.

⇒ Yes we agree that the analysis should be completed with F1. This will be done in the next version.

5- The GP models provide predictive variances, yet the results are shown as single-value predictions. Since this is meant for early warning, please show how uncertainty affects (i) CAT probabilities and (ii) exceedance probability maps for Hmax (e.g., $P[H_{\max} > 0.1 \text{ m}]$). Even one illustrative example for M2/M3 would help.

⇒ Regarding the water levels, we will discuss the results with respect to the uncertainty provided by the Gp models using the kriging variance as suggested.

⇒ Regarding Hmax, the uncertainty quantification requires accounting for multiple sources of uncertainty: the Gp uncertainty for the water levels, for uncertainty of the multioutput Gp models for Hmax, the modelling choices in the deep learning used for the map reconstruction, etc. This work goes beyond the scope of this study because it requires specific developments. These were analyzed in a recent paper (<https://www.researchsquare.com/article/rs-8758359/v1>). So we propose to reference to this paper in the text.

6- The tri-variate extreme analysis and clustering (50 metocean types \times 10 tide levels) are well explained. A figure showing the coverage of (U, Hs, SPM) space at each tide level would reassure readers that the training data cover the full range of conditions.

⇒ We are not sure to understand the question since the same 50 metocean events are applied to each 10 tide levels. So the the joint distribution (U, Hs, SPM) is the same for each tide level (but for low tide coefficients only strong storm events generate marine flooding).

⇒ Do you suggest to complete Fig 2 (A) with other patterns representing the 50 events with other parameters (like Hs/SPM or U/SPM)?

7- The Klaus 2.0 and 2.6 examples are strong. If possible, include scatter plots of simulated vs. predicted Hmax (with density shading) to better illustrate spatial performance.

⇒ Yes this is a good idea; scatter plots will be added.

8- The text says M2/M3 “minimise false positives but at the expense of false alarms.” Since a false alarm is a false positive, please clarify whether the models tend to over-predict (more false positives) or under-predict (more false negatives).

⇒ We mean that the models tend to over-predict with more false positives. This will be clarified in the text.

Minor Comments:

1- I see both “metamodelling” and “metamodeling” in the text. Pick one and use it throughout

=> OK

2- I recommend using “covariates” instead of “co-variables” where appropriate.

=> OK

3- Page 8, Line 6: Choose either “reanalyses” (UK) or “reanalyzes” (US) and use it consistently.

=> OK

4- Page 23, Table 3: Replace “Recommandation” with “Recommendation”, and “prepardness” with “preparedness.”

=> OK

5- Page 26, Line 18: Replace “bootleneck” with “bottleneck.”.

=> OK

6- Define NGF early and keep units consistent in text and figures.

=> OK

7- In Fig. 13 (confusion matrices), include overall accuracy and per-class F1; note which class accounts for most errors if possible.

=> OK

8- When describing CAT thresholds (10 points), include the actual SSHmax ranges in a small table for quick reference. This will help reproducibil

=> OK