Response to Reviewer 2

We thank the Reviewer for the detailed and scientifically important critique. The manuscript has been substantially restructured. The central change is that the paper no longer treats tide-gauge sea level as direct evidence of land inundation. The revised work is outlined explicitly as a machine-learning emulator of fixed-bed XBeach simulations forced by two documented storms. The former synthetic 10 × 10 target, residual physics baseline, and pre-XBeach Monte Carlo construction have been removed. The main quantitative validation is now an independent temporal test: models are trained on November 2023 XBeach states and tested on the complete December 2024 event.

Major comment 1. *My major concern relates to the fact that the methodology for predicting coastal flooding has been developed and validated using tide-gauge observations of sea level rather than direct inundation data. Tide gauges measure sea level near the coast and, because they are typically installed within harbors or near coastal structures, they cannot adequately represent coastal flooding processes. In particular, they do not account for wave-induced effects such as wave setup and runup, and their records may be influenced by harbor resonance. This appears to be the case in the present study, where the tide gauges at Constanța, Mangalia, and Midia are located within well-protected harbors, while the coordinates reported for Gura Portiței indicate an offshore location. Indeed, Figures 5 and 7 suggest that waves have a negligible influence on the estimated inundation depth. Moreover, the selection of the flooding threshold should be related to the specific characteristics of each coastal segment. In the present work, however, a single threshold value has been adopted for all investigated coastal sectors. This also raises the question of how sensitive the flood predictions are to the choice of this threshold. For these reasons, although the proposed modelling framework may be conceptually sound, it cannot be properly applied to and validated for the western Black Sea coast using the available data. I therefore strongly encourage the authors to seek and incorporate direct inundation observations. In my opinion, without such data, the methodology cannot be adequately tested and the manuscript is not suitable for publication.*

Response: We agree with the important criticism. The previous formulation overstated what tide-gauge data could validate. We have removed the tide gauges from the land-inundation validation chain and removed all labels such as “observed coastal flood depth”. The revised target is actual land wetting calculated from XBeach zs and bed elevation along the cross-shore profiles. We have not obtained spatially georeferenced field maps of inundation for the two storms, so we do not claim end-to-end observational validation. Instead, the paper is explicitly scoped as an XBeach emulator. To strengthen the physical test, we use 14 documented-event XBeach runs, an entirely independent December 2024 temporal test, event-peak diagnostics, source-aware terrain QA, and wet/dry-threshold sensitivity. The title, Abstract, Methods, Discussion, and Conclusions have been revised accordingly.

> *Revised manuscript text: “The revised study is explicitly framed as an emulator of fixed-bed XBeach simulations rather than as an observationally validated inundation-forecast system. Tide-gauge records are therefore not used as a proxy for land flooding. Instead, the surrogate targets are extracted directly from shoreline-connected wetting simulated by XBeach for two documented storm events. Because spatially georeferenced field observations of inundation depth or flood extent are not available for these events, the XBeach simulations themselves are not claimed to be validated against direct inundation observations; this limitation defines the present scope of the study.”*

Response: The threshold issue is also addressed directly. The former universal 0.30 m classification threshold has been removed from the XBeach land-wetting diagnostic. A 0.05 m numerical wet/dry criterion is used for the surrogate target, and the resulting connected flood extent is recomputed at 0.01, 0.05, 0.10, and 0.20 m for both storms. Figure 6 now shows the sector-specific sensitivity.

Revised manuscript text: "Sensitivity to the wet/dry criterion was assessed by recomputing shoreline-connected flood extent at thresholds of 0.01, 0.05, 0.10, and 0.20 m. For sectors with connected flooding, increasing the threshold from 0.01 to 0.20 m reduces mean event-peak extent by about 12.8% for November 2023 and 17.6% for December 2024, with strong site-to-site variation. The 0.05 m value is retained as the operational wet/dry diagnostic used for surrogate training, while the sensitivity curves are reported to make the threshold dependence explicit."

Major comment 2. *Secondly, the manuscript lacks a clear and comprehensive description of the methodology. This aspect should be substantially improved to enable readers to fully understand the proposed approach. In particular: i) the Methods section should be reorganized to follow the workflow illustrated in Fig. 3 (observations, XBeach model, Monte Carlo scenarios, surrogate model, etc.); ii) it is not clear whether the Monte Carlo scenarios were generated using all sea-level observations from the different tide gauges collectively, or whether separate scenarios were developed for each coastal sector; iii) the description of the XBeach model should also provide details of the simulation setup, including the model domain, boundary and forcing conditions, spatial resolution, time step, and the representation of coastal and harbor structures; iv) the characteristics of the 10 × 10 spatial target grid used by the surrogate model should be reported, including its extent, resolution, and the bathymetric and topographic datasets employed; v) a clear description on how the dataset have been used for training and testing the surrogate models is missing; vi) it is unclear how the 0D tide-gauge observations, the 1D XBeach profiles, and the 2D surrogate-model grid are integrated.*

Response: The Methods section has been fully reorganised to follow the revised workflow (new Fig. 3): documented forcing → final terrain profiles → 14 fixed-bed XBeach simulations → time-resolved ML ensemble → event-conditioned Monte Carlo. The previous 10 × 10 grid has been removed, so there is no longer an unexplained 0D–1D–2D mapping. The surrogate predicts three diagnostics extracted directly from each 1-D XBeach state: maximum land water depth, shoreline-connected flood extent, and number of wet land nodes. We now report the XBeach boundary forcing, grid design, 73 h duration, 600 s output interval, fixed-bed configuration, lack of explicit harbour structures, 0.05 m wet/dry criterion, one-hour spin-up exclusion, predictor definitions, leave-one-sector-out model-development procedure, and fully independent December 2024 test.

Revised manuscript text: "In this study, inundation depth is defined as the positive XBeach-modelled water-column thickness above the local bed elevation at landward grid nodes, h = max(zs - zb, 0). A land node is considered wet when h is at least 0.05 m. The contiguous flood extent is the landward distance from the shoreline to the last consecutively wet node before the first dry interruption. These quantities are XBeach model diagnostics; they are not tide-gauge sea level and they are not direct field observations of inundation."

Major comment 3. *Thirdly, the authors should pay greater attention to the use of appropriate technical terminology. Below are some examples of terms that are either unclear or potentially misleading:*

3.1 *Coastal flood depth, inundation depth, inundation field, inundation response, spatial inundation. In my opinion, the term inundation is not appropriate in this context, since both the observations and the modelling framework appear to represent the (total) sea level rather than actual land flooding. A clear definition of inundation depth should be clearly stated at the beginning of the Methods section.*

3.2 *XBeach-informed physical baseline, physics baseline. These terms appear to refer simply to the outputs generated by the XBeach model. I therefore suggest replacing them with a more straightforward expression such as XBeach outputs or XBeach simulations. The terms informed, physical, and baseline add unnecessary complexity and may confuse the reader.*

3.3 *Storm surge, storm-surge setup, sea-level anomaly. Tide gauges record the total sea level and additional processing is required to separate and quantify the different components of sea-level variability (e.g., storm surge, sea-level anomaly, astronomical tide). The manuscript should clearly explain how these quantities were derived and used.*

3.4 *Storm sequence: It is unclear what is meant by this term. The authors may be referring to storm duration, a storm event, or a storm case. Please either use a more precise term or clearly define its meaning in the manuscript.*

Response: Agreed. Inundation is now defined only from positive XBeach water depth on land, not from tide-gauge sea level. The definition appears at the beginning of Methods.

> *Revised manuscript text: "In this study, inundation depth is defined as the positive XBeach-modelled water-column thickness above the local bed elevation at landward grid nodes, h = max(zs - zb, 0). A land node is considered wet when h is at least 0.05 m. The contiguous flood extent is the landward distance from the shoreline to the last consecutively wet node before the first dry interruption. These quantities are XBeach model diagnostics; they are not tide-gauge sea level and they are not direct field observations of inundation."*

Response: Agreed. The terms "XBeach-informed physical baseline" and "physics baseline" have been removed. The final surrogate is trained directly on XBeach outputs; the former residual-baseline formulation and its negative R^2 are no longer part of the paper.

Response: Agreed. We no longer call the CMEMS boundary series a tide-gauge-derived storm surge. The manuscript consistently uses "CMEMS sea-level forcing" or "CMEMS zos". No tide/surge/harbour-resonance decomposition is performed or claimed.

> *Revised manuscript text: "The revised manuscript uses the terms CMEMS sea-level forcing, XBeach simulation/output, documented storm event, maximum land water depth, contiguous flood extent, and wet land nodes. No decomposition of tide-gauge total water level into astronomical tide, storm surge, or harbour-resonance components is used in the surrogate workflow."*

Response: Agreed. "Storm sequence" has been replaced by "documented storm event", "event time", or "time-resolved event response".

Major comment 4. *Lastly, the results indicate that the different surrogate models exhibit very similar performance. I therefore suggest presenting only the ensemble mean results in the main manuscript and moving the complete set of performance statistics to the Supplementary Material. This would improve the readability of the paper without affecting the interpretation of the results. Furthermore, the authors should consider that these findings may be strongly influenced by the methodological limitations highlighted in the above concerns. Therefore, the reported model performance should be interpreted with caution until the proposed framework is validated against appropriate inundation observations.*

Response: Agreed. Only the validation-weighted ensemble is discussed quantitatively in the main Results. Individual RF, ET, GB, and HGB performance is relegated to Supplementary Fig. S3, while Table 4 in the main manuscript reports only the weights used to construct the ensemble. We also added explicit cautions that the reported skill is emulation fidelity to XBeach, not validation against field inundation observations.

> *Revised manuscript text: "Surrogate skill is evaluated against XBeach outputs, not against tide-gauge observations. The November 2023 event provides 3,031 post-spin-up states for model development, while the December 2024 event is held out in its entirety as an independent temporal test with another 3,031 states. On that independent event, the validation-weighted ensemble gives $R^2 = 0.976$ and RMSE = 0.033 m for maximum land water depth, $R^2 = 0.988$ and RMSE = 6.25 m for contiguous flood extent, and $R^2 = 0.987$ and RMSE = 1.10 nodes for the number of wet land nodes. These statistics quantify emulation fidelity to XBeach and must not be interpreted as observational flood-validation skill."*

Additional comments

1. *As it stands, the abstract reads more like that of a paper intended for a national or regional journal. Moreover, it contains unnecessary details on the datasets and statistical results, including unexplained acronyms. The abstract should be streamlined to emphasize the study's objectives, methodology, key findings, and significance.*

Response: The Abstract has been rewritten and shortened. It now states the objective, revised XBeach-to-ML workflow, independent temporal test, principal ensemble results, and the explicit limitation that direct inundation observations are unavailable. Unnecessary dataset acronyms and the former long list of metrics were removed.

> *Revised manuscript text: "A wet/dry-threshold sensitivity analysis and an event-conditioned Monte Carlo experiment are used to quantify diagnostic robustness and the surrogate applicability domain. The revised framework does not use tide-gauge sea level as a proxy for land inundation and does not claim validation against direct flood observations, which are unavailable for the analysed events. It is therefore intended as a rapid emulator and scenario-screening tool for XBeach coastal-flood diagnostics rather than as a stand-alone operational predictor of absolute inundation depth."*

2. *Lines 113-126: details about the datasets, model performance metrics and predictors should not be reported here.*

Response: The former introduction-level details on predictors and performance metrics have been removed. The retained Introduction now provides only the scientific motivation, revised scope, and objectives; implementation details are in Methods and quantitative results are in Sect. 4.

> *Revised manuscript text: "The objective is to test whether a machine-learning ensemble can reproduce time-resolved XBeach coastal-flood diagnostics at seven Western Black Sea sectors and then use the trained emulator for rapid, event-conditioned scenario screening. The revised workflow therefore separates (i) documented marine and atmospheric forcing, (ii) sector-specific cross-shore terrain and fixed-bed XBeach simulations, (iii) surrogate training and independent temporal testing, and (iv) event-conditioned Monte Carlo application."*

3. *Lines 128-129: the flood extent, expressed as grid nodes exceeding an operational inundation, does not implies continuity of the flooding area. Please discuss the limitation of this approach.*

Response: Agreed. The former count of threshold-exceeding 2-D grid nodes has been removed. Flood extent is now defined as a shoreline-connected 1-D cross-shore extent, which explicitly requires consecutive wet nodes from the shoreline. Its limitation as a transect metric rather than a 2-D polygon is stated.

> *Revised manuscript text: "The revised flood-extent metric explicitly enforces one-dimensional shoreline connectivity along each cross-shore profile: only consecutive wet land nodes beginning at the shoreline are counted. This removes the former ambiguity associated with counting isolated threshold exceedances. It should nevertheless be interpreted as a 1-D transect extent, not as a continuous two-dimensional inundation polygon or an alongshore-connected flooded area."*

4. *exceeding an operational inundation threshold*

Response: The wording has been corrected throughout. The revised manuscript uses "wet/dry criterion" and "shoreline-connected flood extent" rather than the grammatically incomplete phrase in the previous version.

5. *Lines 141-144: is this a result of the present study? If so, it should be moved to the Results section. Otherwise, an appropriate reference should be provided.*

Response: The sentence in question has been removed from the Introduction. Event-specific findings are now confined to the Results section.

6. *Line 154: main Romanian Western coastal sectors.*

Response: The geographical wording has been standardised to "Romanian Western Black Sea coast" or "Western Black Sea sectors", depending on context.

7. *Table 1: include the temporal extension and the number of valid data for each station.*

Response: The former tide-gauge statistics table has been removed because tide gauges are no longer used as the inundation-validation target. Revised Table 1 instead documents all seven XBeach sectors, and Methods reports the two event windows and the 433 post-spin-up states per sector per event (3,031 states per event).

8. *Figure 2: This figure suggests that the tide gauges use different vertical reference datums, while Table 1 reports very similar mean SLEV values. Please clarify whether the data were adjusted to a common datum before the analysis.*

Response: The former tide-gauge time-series Fig. 2 has been removed from the validation chain. The revised Fig. 2 shows CMEMS–ERA5 forcing for the two documented storms. We no longer compare local tide-gauge datums as if they were directly interchangeable. The remaining vertical-reference limitation is explicitly stated.

> *Revised manuscript text: "The former cross-station comparison of tide-gauge elevations has been removed from the validation chain. Terrain is expressed in a local shoreline-relative model datum with z = 0 at the shoreline, while CMEMS zos provides the time-varying offshore water-level boundary. A geodetic transformation that would place all land-survey elevations and CMEMS sea level in a single independently verified vertical reference is not available for every sector; this is therefore treated as a limitation of absolute depth interpretation rather than concealed through datum pooling."*

9. *Line 175: what do you mean with related hydrodynamic variables. You should only mention the variables that have been used in the present study.*

Response: The vague phrase "related hydrodynamic variables" has been removed. Revised Table 2 lists only variables actually used: Hs, Tp, wave direction, CMEMS zos, ERA5 wind speed/direction, and static terrain sources.

10. *Line 182: provide the reference for the topo-bathymetric dataset.*

Response: Topo-bathymetric sources are now specified explicitly: EMODnet DTM for marine bathymetry, Copernicus DEM GLO-30 for land elevation, shoreline/GNSS anchors for coastline registration, and the Zenodo RTK-GNSS/multibeam dataset for Vama Veche.

> *Revised manuscript text: "The profile-generation workflow was revised to remove the artificial offshore vertical-wall artefact identified by the reviewer. The shoreline is imposed at z = 0 using the available GNSS/shoreline anchor, bathymetry is taken from EMODnet (with the Vama Veche nearshore profile supplemented by the RTK-GNSS/multibeam Zenodo dataset), and land elevation is taken from Copernicus DEM GLO-30. Source-aware interpolation is used only across the unresolved nearshore transition. Quality control confirms no offshore re-emergent land nodes in the final source-aware profiles; steep nearshore sectors are retained with explicit visual-QA flags rather than being silently smoothed."*

11. *Table 2: many variables (SLEV, ATMS, Hs, Tp, sea-surface state, SST, MSLP) and roles (storm surge, erosion-related forcing, plume related forcing, …) need to better defined.*

Response: Revised Table 2 defines every variable and its modelling role. Variables from the old framework that are not used by the final surrogate (e.g. SST, Danube discharge, MSLP as a direct surrogate predictor) have been removed from the core Methods.

12. *Line 203: define "target inundation response".*

Response: The phrase "target inundation response" has been replaced by explicit XBeach-derived targets: maximum land water depth, contiguous shoreline-connected flood extent, and wet-land-node count.

> *Revised manuscript text: "In this study, inundation depth is defined as the positive XBeach-modelled water-column thickness above the local bed elevation at landward grid nodes, h = max(zs - zb, 0). A land node is considered wet when h is at least 0.05 m. The contiguous flood extent is the landward distance from the shoreline to the last consecutively wet node before the first dry interruption. These quantities are XBeach model diagnostics; they are not tide-gauge sea level and they are not direct field observations of inundation."*

13. *Lines 214-216: the variables storm duration, inverse-barometer contribution, wave-power-related terms, and interaction terms between sea-level anomaly and wave forcing must be clearly defined.*

Response: The final predictor definitions are now explicit. The previous vague terms were replaced by elapsed event time, Hs^2Tp as a wave-power proxy, zos×Hs as the sea-level–wave product, cos(θwind−θwave) as wind–wave alignment, and max[cos(θwave),0] as the onshore-wave factor. The old generic "storm duration" and inverse-barometer terms are not part of the final surrogate feature set.

14. *Table 3: The mean values obtained from the Monte Carlo simulations (MCS) are significantly higher than the observed means. This discrepancy is particularly evident for sea level, for which the simulated mean exceeds the observed mean by more than 40%. The authors should provide an explanation for this difference and discuss its potential implications for the modelling results.*

Response: The former 5,000-scenario pre-XBeach Monte Carlo table and the 40% mean sea-level mismatch have been removed. Monte Carlo is now an event-conditioned post-surrogate application based on the documented forcing states with small perturbations and an explicit applicability-domain filter.

> *Revised manuscript text: "The Monte Carlo stage was also reformulated. It no longer generates a synthetic forcing climatology before XBeach. Instead, 10,000 states per sector are sampled conditionally from the two documented-event forcing states with small perturbations, and only samples that remain inside the empirical surrogate-training envelope are used in the principal summaries. Depending on sector, 93.68–98.30% of generated states satisfy this applicability criterion. These results describe event-conditioned surrogate uncertainty and response ranges; they are neither a flood climatology nor an observational validation."*

15. *Lines 229-230: define "design-of-experiments forcing combinations".*

Response: The former "design-of-experiments forcing combinations" concept has been removed from the final workflow.

16. *Line 230: why sea-level anomaly?*

Response: The final workflow does not derive a tide-gauge sea-level anomaly. It uses CMEMS zos directly as the time-varying offshore sea-level boundary and labels it as sea-level forcing.

> *Revised manuscript text: "The revised manuscript uses the terms CMEMS sea-level forcing, XBeach simulation/output, documented storm event, maximum land water depth, contiguous flood extent, and wet land nodes. No decomposition of tide-gauge total water level into astronomical tide, storm surge, or harbour-resonance components is used in the surrogate workflow."*

17. *Line 231: specify the topo-bathymetric dataset used for implementing the XBeach cross-sections.*

Response: The terrain sources used for the XBeach cross-sections are now specified in Sect. 2.2/3.2, Table 2, and the revised profile figure.

> *Revised manuscript text: "The profile-generation workflow was revised to remove the artificial offshore vertical-wall artefact identified by the reviewer. The shoreline is imposed at z = 0 using the available GNSS/shoreline anchor, bathymetry is taken from EMODnet (with the Vama Veche nearshore profile supplemented by the RTK-GNSS/multibeam Zenodo dataset), and land elevation is taken from Copernicus DEM GLO-30. Source-aware interpolation is used only across the unresolved nearshore transition. Quality control confirms no offshore re-emergent land nodes in the final source-aware profiles; steep nearshore sectors are retained with explicit visual-QA flags rather than being silently smoothed."*

18. *Lines 234-236: I do not agree with this approach. Each component of the modelling system must be calibrated and verified to obtain a reliable prediction system.*

Response: We agree that a predictive system requires calibration and verification of each component. Because direct spatial inundation observations are unavailable for these two events, we have not represented the fixed-bed XBeach simulations as observationally calibrated. Instead, the paper is explicitly limited to XBeach emulation and screening; direct field calibration is identified as a prerequisite for operational forecasting.

Revised manuscript text: "The present framework remains intentionally limited. It uses one-dimensional fixed-bed XBeach transects, does not resolve harbour geometry, alongshore exchange, morphodynamic feedbacks, or two-dimensional overtopping pathways, and has not been calibrated against direct field maps of inundation depth or extent. Consequently, the surrogate is appropriate for rapid emulation, scenario ranking, and screening of the XBeach response space, but not yet for stand-alone operational prediction of absolute coastal flood depth."

19. *Lines 236-238: this part should be moved to section 3.4.*

Response: The old text has been removed during the complete Methods restructuring; the final surrogate formulation is now contained in Sect. 3.4.

20. *Figure 5: The left panel clearly shows that the XBeach profiles provide an unrealistic representation of the offshore/submerged part of the coastal profile, including a vertical wall exceeding 5 m in height. The authors should explain the origin of this feature and discuss its potential effects on the model results.*

Response: This issue prompted a full revision of the profile-construction workflow. The former figure containing the artificial offshore wall has been replaced by source-aware, shoreline-registered profiles. The new QA verifies the marine/land transition, removes offshore re-emergent nodes, and flags steep sectors for visual inspection.

Revised manuscript text: "The profile-generation workflow was revised to remove the artificial offshore vertical-wall artefact identified by the reviewer. The shoreline is imposed at z = 0 using the available GNSS/shoreline anchor, bathymetry is taken from EMODnet (with the Vama Veche nearshore profile supplemented by the RTK-GNSS/multibeam Zenodo dataset), and land elevation is taken from Copernicus DEM GLO-30. Source-aware interpolation is used only across the unresolved nearshore transition. Quality control confirms no offshore re-emergent land nodes in the final source-aware profiles; steep nearshore sectors are retained with explicit visual-QA flags rather than being silently smoothed."

21. *Line 291: from where do you get observations of the inundation depth at grid-node level?*

Response: There are no observations of inundation depth at grid-node level. The revised manuscript now states this explicitly. The former 2-D grid target is removed, and all ML targets are extracted directly from XBeach output.

Revised manuscript text: "In this study, inundation depth is defined as the positive XBeach-modelled water-column thickness above the local bed elevation at landward grid nodes, h = max(zs - zb, 0). A land node is considered wet when h is at least 0.05 m. The contiguous flood extent is the landward distance from the shoreline to the last consecutively wet node before the first dry interruption. These quantities are XBeach model diagnostics; they are not tide-gauge sea level and they are not direct field observations of inundation."

22. *Lines 296, tables 5 and 6, and somewhere else: always specify "mean" when referring to average value of the surrogate models.*

Response: Agreed. The main manuscript now refers to the "validation-weighted ensemble" rather than an ambiguous "average model". Individual-model statistics are placed in the Supplement.

23. *Table 5: why the physics baseline has negative R2?*

Response: The former physics baseline and its negative R^2 have been removed. The final model is a direct multi-output surrogate of XBeach simulations rather than a residual correction to a simplified baseline.

24. *Figure 6: What's the Observed Coastal Flood Depth? Tide gauge measure sea level not flood depth.*

Response: The label "Observed Coastal Flood Depth" has been removed. Revised Figs. 7 and 8 explicitly use "XBeach" on the reference axis/series and "ML ensemble" on the prediction axis/series.

25. *Captions of figures 6, 7, 8 and 9: do not report comments in the captions.*

Response: Figure captions have been rewritten as descriptive captions only; interpretive statements have been moved to the Results text.

26. *Section4.5: explain how the permutation-importance diagnostics have been computed.*

Response: The computation is now described explicitly: on the independent December 2024 event, each predefined predictor group is permuted while the trained ensemble is fixed; the operation is repeated five times; importance is the increase in RMSE relative to the unpermuted prediction.

> *Revised manuscript text: "Permutation importance is computed on the independent December 2024 event by randomly permuting each predefined predictor group while leaving the trained ensemble unchanged, repeating the perturbation five times, and reporting the increase in RMSE relative to the unpermuted prediction. The grouped analysis shows that profile geometry and site identity dominate all three outputs, with sea level providing the next consistent contribution. Wave forcing and the explicit sea-level–wave interaction have small additional permutation importance in this two-event experiment; no claim of a robust compound sea-level–wave interaction is therefore made."*

27. *Line 353: clarify how the interaction between sea-level anomaly and significant wave height have been defined.*

Response: The interaction feature is now defined explicitly as the product zos×Hs. Importantly, the revised grouped permutation analysis shows negligible additional importance of this interaction in the current two-event experiment.

> *Revised manuscript text: "Permutation importance is computed on the independent December 2024 event by randomly permuting each predefined predictor group while leaving the trained ensemble unchanged, repeating the perturbation five times, and reporting the increase in RMSE relative to the unpermuted prediction. The grouped analysis shows that profile geometry and site identity dominate all three outputs, with sea level providing the next consistent contribution. Wave forcing and the explicit sea-level–wave interaction have small additional permutation importance in this two-event experiment; no claim of a robust compound sea-level–wave interaction is therefore made."*

28. *Line 356: no compound sea-level–wave interaction effects can be reliably assessed in this study, as the observations were collected from tide gauges located within harbors, and the XBeach simulations include an unrealistic 5 m vertical wall structure.*

Response: Agreed. We no longer claim a reliably quantified compound sea-level–wave interaction. The revised profile artefact has been addressed, and the new importance analysis shows that the explicit interaction contributes little beyond geometry, site identity, and sea level. This is stated as a result rather than interpreted as a robust physical interaction.

> *Revised manuscript text: "Permutation importance is computed on the independent December 2024 event by randomly permuting each predefined predictor group while leaving the trained ensemble unchanged, repeating the perturbation five times, and reporting the increase in RMSE relative to the unpermuted prediction. The grouped analysis shows that profile geometry and site identity dominate all three outputs, with sea level providing the next consistent contribution. Wave forcing and the explicit sea-level–wave interaction have small additional permutation importance in this two-event experiment; no claim of a robust compound sea-level–wave interaction is therefore made."*

29. *It is not necessary to include both Figure 7 and Table 8 as they report the same information.*

Response: The redundant figure/table pair has been removed. The main manuscript retains the grouped permutation-importance figure; detailed individual feature values remain supporting material rather than duplicating the figure in the main text.

30. *Section 4.7: from where these other stations come from?*

Response: All seven sectors are now introduced at the start of Sect. 2 and listed in revised Table 1 and Fig. 1. They are Constanța, Mamaia, Midia, Eforie Nord, Mangalia, Vama Veche, and Gura Portiței.

31. *Figure 9b: do not mix observation with XBeach model output.*

Response: Agreed. The revised validation figures no longer mix “observation” with XBeach output. They compare XBeach reference outputs directly against the ML ensemble, and the text repeatedly states that this is simulator-emulation validation, not observational flood validation.

> *Revised manuscript text: “Surrogate skill is evaluated against XBeach outputs, not against tide-gauge observations. The November 2023 event provides 3,031 post-spin-up states for model development, while the December 2024 event is held out in its entirety as an independent temporal test with another 3,031 states. On that independent event, the validation-weighted ensemble gives $R^2 = 0.976$ and RMSE = 0.033 m for maximum land water depth, $R^2 = 0.988$ and RMSE = 6.25 m for contiguous flood extent, and $R^2 = 0.987$ and RMSE = 1.10 nodes for the number of wet land nodes. These statistics quantify emulation fidelity to XBeach and must not be interpreted as observational flood-validation skill.”*