https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

An XBeach-informed machine-learning surrogate for rapid coastal flood prediction

Maria Emanuela Mihailov[1,2], Alina Spinu[1]*, Alexandru Cindescu[1], Lucian Dutu[2]

[1]National Institute for Marine Research and Development "Grigore Antipa" (NIMRD), Constanța, Romania
[2]Maritime Hydrographic Directorate (MHD), Constanța, Romania

Correspondence to: Maria Emanuela Mihailov (emihailov@alpha.rmri.ro, emanuela.mihailov@dhmfn.ro) and Alina Spinu (aspinu@alpha.rmri.ro)

Abstract. Rapid assessment of coastal inundation along the Western Black Sea coast requires computationally efficient methods that can screen multiple compound storm scenarios while preserving a clear link to the controlling physical drivers. This study develops and evaluates an XBeach-informed machine-learning surrogate for coastal inundation and threshold-based flood-extent screening. The framework combines high-frequency sea-level observations from the Maritime Hydrographic Directorate monitoring network, Copernicus Marine wave and Black Sea physical products, ERA5 atmospheric forcing, Global Runoff Data Centre Danube discharge, bathymetric and coastal-elevation descriptors, and an intermediate cross-shore physical-response layer. The modelling dataset consists of 5,000 Monte Carlo forcing scenarios, 22 forcing and derived predictors, and a 10 × 10 spatial target grids, corresponding to 100 inundation-depth nodes per scenario. Four tree-based residual models, namely Random Forest, Gradient Boosting, Extra Trees, and Histogram Gradient Boosting, are combined using validation-derived weights.

The conservative held-out node-level evaluation shows moderate quantitative skill for continuous inundation-depth prediction. The ensemble reaches $R^2 = 0.409$, RMSE = 1.181 m, MAE = 0.678 m, NSE = 0.409, KGE = 0.235, Pearson $r = 0.914$, and Spearman $\rho = 0.923$. In contrast, threshold-based flood-extent classification above the 0.30 m operational inundation threshold is substantially stronger, with F1 = 0.995, AUC-ROC = 0.9998, Matthews correlation coefficient = 0.989, and Cohen's $\kappa = 0.989$. Split conformal prediction intervals provide near-nominal 90 % marginal coverage, with empirical coverage of 0.911, but the mean 90 % interval width is large, 3.25 m, indicating limited sharpness for local quantitative depth estimation. Extreme-event diagnostics show systematic underprediction of upper-tail inundation depths, with a mean bias of approximately −2.59 m for events above the 90th percentile.

The present configuration is therefore most appropriate for rapid flood-extent screening, scenario ranking, and identification of cases where more detailed process-based simulations are required. It should not yet be interpreted as a stand-alone operational predictor of maximum inundation depth. The study demonstrates how in situ

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

observations, Copernicus Marine products, physical-response modelling, machine-learning emulation, and uncertainty diagnostics can be combined into a transparent coastal-hazard screening workflow for the Western Black Sea coast.

Keywords: coastal flooding; Black Sea; XBeach; machine-learning; storm surge; surrogate model, spatial flood prediction; Copernicus Marine Service.

1 Introduction

Coastal flood prediction has moved rapidly from deterministic event-based simulations towards probabilistic, multi-driver and data-informed frameworks. Full-physics hydrodynamic models remain the reference tools for resolving storm surge, wave setup, run-up, inundation pathways and local bathymetric control, but their computational cost limits their use for large ensembles, rapid scenario screening and real-time operational forecasting (Jafarzadegan et al., 2023; Yang et al., 2020). This limitation is particularly important in coastal early-warning systems, where decision makers require not only accurate water-level estimates, but also timely information on uncertainty, exceedance probability and spatially distributed flood impacts (Clare et al., 2022; Irazoqui Apecechea et al., 2023).

Recent literature also emphasizes that coastal flooding is rarely controlled by a single driver. Compound flood events may result from the interaction of storm surge, waves, river discharge, precipitation, local wind setup and antecedent water-level conditions, with the relative contribution of each driver varying strongly between coastal regions (Camus et al., 2021; Couasnon et al., 2020; Han and Tahvildari, 2024). Hybrid statistical–dynamical frameworks now combine stochastic forcing generators, hydrodynamic simulations and machine-learning surrogates to explore a much larger space of compound flood conditions than would be feasible with process-based models alone (Wang et al., 2025). This direction is especially relevant for deltaic and low-lying coastal systems, where marine and fluvial drivers may interact non-linearly and where flood depth, duration and spatial extent depend on local morphology and boundary-condition phasing (Seidenfaden et al., 2025).

For the Mediterranean and Black seas, the scientific context is different as both basins are microtidal and coastal extreme sea levels are primarily controlled by atmospherically driven sea-level variability, storm surge, seiches, wave setup and high-frequency sea-level oscillations rather than by astronomical tides (Pérez Gómez et al., 2022). Tide-gauge observations remain essential for detecting and validating these processes, particularly because high-frequency coastal sea-level variability is not always fully resolved by regional operational models (Irazoqui Apecechea et al., 2023; Pérez Gómez et al., 2022). The western Black Sea has been identified as one of the exposed sectors of the basin to storm-surge conditions, while the Western coast adds further complexity through the proximity of the Danube Delta, the contrast between low-lying deltaic sectors and cliff-backed coasts, and the

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

influence of freshwater discharge on shelf circulation and stratification (Ciliberti et al., 2021; Pérez Gómez et al., 2022).

The Copernicus Marine Service (CMEMS) has significantly improved the availability of operational forcing and validation data for coastal applications, including sea level, waves, currents, temperature and regional Black Sea analyses and forecasts (Ciliberti et al., 2021; Irazoqui Apecechea et al., 2023). These products provide an important baseline for downstream coastal-flood services, but recent assessments show that local coastal applications still require in situ calibration, high-frequency sea-level records and locally adapted modelling or downscaling strategies, particularly in areas where short-period oscillations, harbour-scale effects or unresolved nearshore morphology influence total water level (Irazoqui Apecechea et al., 2023; Pérez Gómez et al., 2022). Wave contribution remains another important component, as wave setup and wave-driven water-level anomalies can significantly increase coastal hazard and show strong sensitivity to shoreline geometry and nearshore bathymetry (Chaigneau et al., 2023; Soomere et al., 2020).

Machine-learning surrogate models, also referred to as emulators, provide one pathway for reducing the computational blockage. Surrogates are trained offline on a representative set of numerical or physics-informed simulations and then used to approximate the model response surface in near-real time and have become a solution for reducing the computational problem of high-resolution storm-surge and inundation modelling. Tree ensembles such as Random Forests, Gradient Boosting, Extra Trees and Histogram Gradient Boosting are suitable for tabular coastal-forcing datasets because they represent non-linear threshold behaviour, require limited preprocessing, and allow interpretable feature-importance diagnostics (Breiman, 2001; Friedman, 2001; (Breiman, 2001; Geurts et al., 2006).

Recent studies show that Random Forests (RF), Gradient Boosting (GB), Neural Networks (NN), Gaussian-process emulators and multifidelity emulators can reproduce hydrodynamic-model outputs or observed storm-surge responses with major gains in computational efficiency (Ayyad et al., 2022; Ma et al., 2022; Qin et al., 2023; Yang et al., 2020; Zahura et al., 2020). Current developments increasingly focus on emulating physically based models, preserving physical coherence, improving spatial generalisation and quantifying uncertainty in predictions (Longo et al., 2026; Wilkinson et al., 2026). However, most present-day applications stay site-specific, many are focused on tropical-cyclone surge environments, and relatively few studies integrate sea-level observations, wave forcing, atmospheric reanalysis and river discharge into a single operational surrogate framework for microtidal, freshwater-influenced basins such as the western Black Sea.

The Western Black Sea coast extends from the Danube Delta to the southern cliffed coast near Vama Veche and includes low-lying deltaic sectors, engineered harbour areas, tourist beaches, wetlands, and urban coastal assets. Although astronomical tides are small, coastal water levels can increase significantly during wind-driven storm-surge events, particularly under persistent north-easterly forcing. In the northern sector, positive sea-level

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

anomalies can affect low-gradient deltaic areas, while in the central and southern sectors the main operational concerns are beach erosion, harbour disruption, and combined wave-surge impacts.

Full-physics hydrodynamic and wave models are necessary for process understanding and detailed operational forecasting, but their computational cost limits rapid ensemble exploration and probabilistic scenario screening. Machine-learning surrogates can complement these models by learning the response surface of a validated numerical or physics-based simulation chain and producing near-instantaneous estimates once trained. Their value is highest when they retain the controlling physical predictors, including uncertainty information, and remain interpretable for operational users.

The Maritime Hydrographic Directorate (MHD) sea-level monitoring network provides high-frequency observations at Constanta, Gura Portitei, Mangalia, and Midia, supporting calibration of sea-level forcing and storm-surge exceedance statistics. In this paper, we use this observational basis together with Copernicus Marine wave and physical products (Buongiorno Nardelli et al., 2010, 2013; Jansen et al., 2024; Ricker et al., 2024; Staneva et al., 2022), ERA5 atmospheric forcing and Global Runoff Data Centre (GRDC) Danube discharge to construct a fast, physically interpretable machine-learning surrogate for coastal flood prediction for the Western Black Sea coast. By combining Monte Carlo forcing scenarios, process-informed feature engineering, multi-model regression, classification of severe flood conditions and prediction intervals, the framework contributes to the emerging generation of operational coastal-flood emulators while remaining tailored to the hydro-meteorological and geomorphological setting of the Western Black Sea coast.

The objective of this paper is to develop and evaluate a physically informed machine-learning surrogate for rapid coastal flood-depth prediction along the Western Black Sea coast. More specifically, the study aims to: (i) construct an XBeach-informed reference framework linking storm-surge, wave, atmospheric, riverine, and coastal-morphology controls to spatial flood-depth response; (ii) generate a physically plausible Monte Carlo scenario space constrained by in situ sea-level observations, Copernicus Marine products, ERA5 atmospheric forcing (Hersbach et al., 2020, 2023), and Danube discharge data (GRDC, 2026); (iii) train and compare tree-based surrogate models for rapid prediction of spatial flood-depth fields and threshold-based flood occurrence; (iv) assess model skill using regression, classification, residual, quantile-based, and hydrological performance metrics, including Root mean square error (RMSE), mean absolute error (MAE), bias, Nash–Sutcliffe efficiency (NSE), Kling–Gupta efficiency (KGE), quantile–quantile (Q–Q) diagnostics, and extreme-event errors; (v) identify the dominant physical predictors controlling the surrogate response, including the relative influence of sea-level anomaly, wave forcing, wind direction, sea-surface temperature, and the XBeach-informed baseline (Roelvink et al., 2009); and (vi) evaluate the potential and limitations of the resulting framework as a rapid screening tool for coastal-flood hazard assessment and Copernicus downstream decision-support applications.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

The evaluation is organised around three distinct scientific questions: first, whether the surrogate can reproduce spatial inundation depth; second, whether it can reliably identify the flood extent, expressed as grid nodes exceeding an operational inundation threshold; and third, whether its prediction intervals are sufficiently calibrated and sharp to support decision-relevant interpretation. This difference is required because estimating the spatial occurrence of flooding and predicting continuous inundation depth represent different levels of predictive difficulty. In the present configuration, the model shows robust ability in identifying flooded nodes and therefore in delineating threshold-based flood extent, whereas quantitative prediction of the deepest inundation values remains more uncertain.

2 Study area and data

2.1 Western Black Sea coastal setting

The study area covers the Western Black Sea coast from the Danube Delta to the southern Romanian coastal sector. The northern coast is influenced by deltaic and lagoonal morphology, shallow bathymetry and the freshwater input of the Danube. The central sector includes the Constanta-Midia harbour system and adjacent low-lying or engineered shorelines, while the southern sector includes Mangalia and Vama Veche. The coastal water-level regime is controlled mainly by meteorological forcing, wave setup and shelf-scale circulation. Danube discharge also influences nearshore density structure, plume extent and stratification, although its direct contribution to local flood-depth prediction is weaker than the sea-level and wave predictors in the residual-surrogate analysis (Figure 1).

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

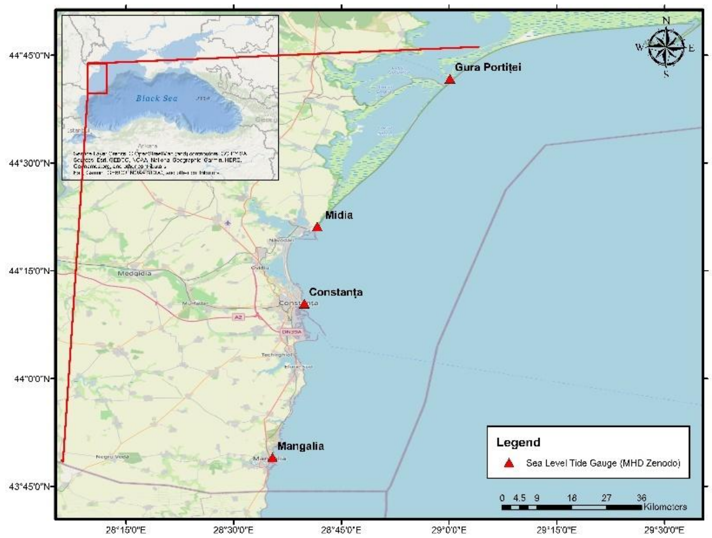

Figure 1. Study area and sea-level monitoring network along the Western Black Sea coast. Red triangles indicate the Maritime Hydrographic Directorate (MHD) sea-level stations used in this study: Gura Portiței, Midia, Constanța, and Mangalia. Map created by the authors using ArcGIS, using station data available from Zenodo (Dutu et al., 2026). Basemap sources: © OpenStreetMap contributors; Esri, GEBCO, NOAA, National Geographic, Garmin, HERE, Geonames.org, and other contributors | Powered by Esri.

2.2 Sea-level observations and station network

Sea-level elevation (SLEV) and atmospheric pressure observations were obtained from the Maritime Hydrographic Directorate (MHD) monitoring network at Constanta, Gura Portitei, Mangalia, and Midia (Figure 1). The four stations cover the main Western coastal sectors, from the Danube Delta-influenced northern coast to the central and southern harbour and beach areas. The records provide the observational basis for characterising the distribution of sea-level anomalies used in the Monte Carlo scenario generator (Table 1). Quality-controlled time series (Figure 2) were used to characterise the local distribution of sea-level anomalies and to support the evaluation of the surrogate-model forcing space (Dutu et al., 2026).

Table 1. Maritime Hydrographic Directorate (MHD) station metadata and summary sea-level statistics used to define the sea-level forcing envelope. SLEV = sea-level elevation relative to local datum (Dutu et al., 2026).

Station	Lat. (deg N)	Lon. (deg E)	mean SLEV (m)	std SLEV (m)	max SLEV (m)
Constanța	44.14	28.66	0.0055	0.0461	1.278
Gura Portiței	44.68	29.27	0.0053	0.0471	0.7240
Mangalia	43.80	28.59	0.0059	0.0461	0.4890
Midia	44.32	28.69	0.0062	0.0468	1.151

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

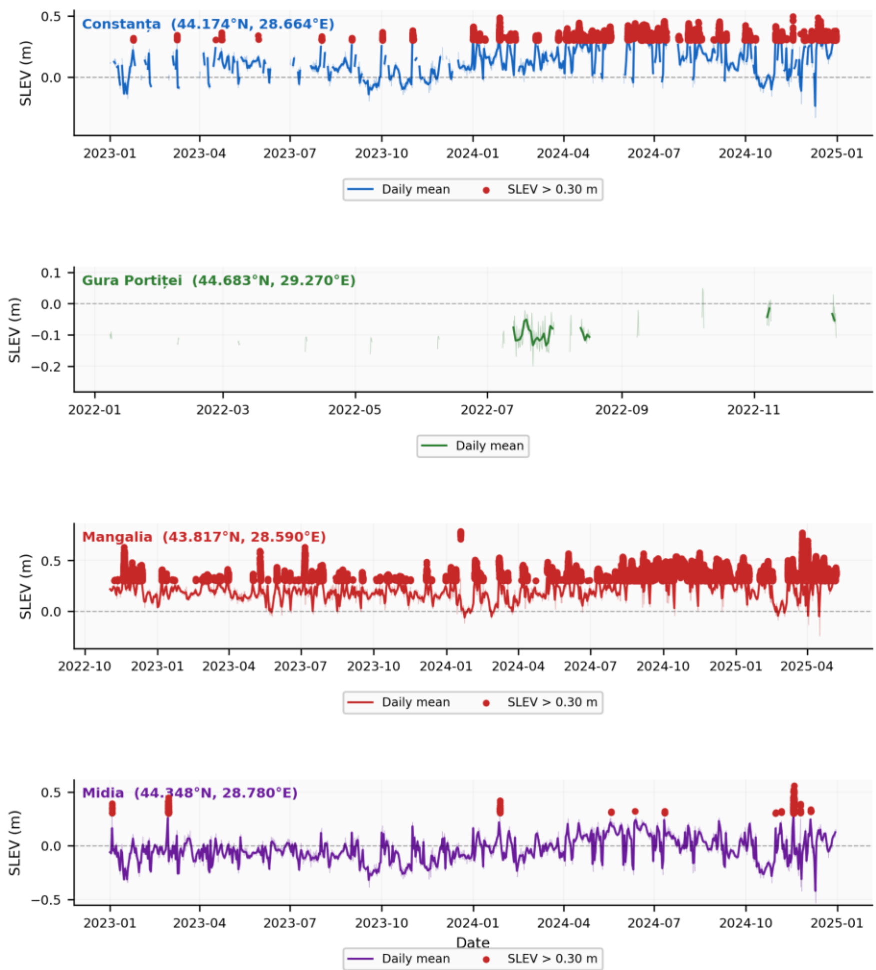

Figure 2. Daily mean sea-level elevation from the MHD stations at Constanța, Gura Portitei, Mangalia and Midia (Dutu et al., 2026). Red markers denote values exceeding 0.30 m. The records provide the empirical basis for storm-surge sampling and threshold classification.

2.3 Forcing and auxiliary datasets

The forcing dataset combines local sea-level observations with gridded marine, atmospheric, river-discharge, thermal, and simplified coastal-morphology information (Table 2). Station sea-level elevation and atmospheric pressure from the MHD network (Dutu et al., 2026)were used to constrain the local storm-surge and pressure-related variability. Offshore wave conditions were obtained from the Copernicus Marine Black Sea Waves Analysis and Forecast product, including significant wave height, peak wave period, and wave direction (Ricker et al., 2024; Staneva et al., 2022). The regional physical sea state was described using the Copernicus Marine Black

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

Sea Physics Analysis and Forecast product, which provides sea surface height, currents, temperature, salinity, and related hydrodynamic variables over the Black Sea basin (Jansen et al., 2024). Sea surface temperature was included as an additional thermal and seasonal-state predictor using the Copernicus Marine Black Sea high-resolution Level-4 SST product.

Atmospheric forcing was derived from ERA5 hourly single-level reanalysis fields, including 10 m wind components and mean sea-level pressure (Hersbach et al., 2020, 2023). Danube discharge was obtained from the Global Runoff Data Centre daily river-discharge archive and was included to represent the potential influence of freshwater input on nearshore stratification, plume-related conditions, and compound coastal response (GRDC, 2026). Bathymetric and coastal-elevation information was used as a static descriptor of the nearshore morphology. These data supported the definition of cross-shore coastal profiles, local slope, relative elevation, and station-dependent exposure used in the XBeach-informed response layer. In this framework, bathymetry is not treated as a time-varying forcing variable, but as a controlling boundary condition for wave transformation, water-level response, and the spatial distribution of flooding.

The final model input matrix contains 22 forcing and derived predictors describing sea-level anomaly, wave forcing, wind direction and intensity, atmospheric pressure effects, freshwater discharge, sea surface temperature, and simplified location- and morphology-related controls.

These variables are not intended to replace process-resolving hydrodynamic modelling. Instead, they define the dynamic forcing, environmental state, and static coastal-geometry descriptors, including bathymetry and coastal elevation, required to emulate the XBeach-informed coastal-response surface within a rapid surrogate-modelling framework.

Table 2. Data sources and their role in the surrogate modelling workflow.

Source	Variables	Role in workflow
FOCCUS/MHD sea level (Dutu et al., 2026)	SLEV, ATMS	storm-surge and sea-level forcing
Copernicus Marine waves (Ricker et al., 2024; Staneva et al., 2022)	Hs, Tp, wave direction	wave setup, wave power and erosion-related forcing
Copernicus Marine physics/SST (Jansen et al., 2024)	sea-surface state, SST	sea-level context and thermal forcing
ERA5 (Hersbach et al., 2020, 2023)	wind speed, wind direction, MSLP	wind setup, wind stress and inverse barometer
GRDC (GRDC, 2026)	Danube discharge	freshwater/plume-related forcing
Bathymetry/topography (EMODnet Bathymetry Consortium, 2026)	coastal elevation, slope, cross-shore geometry	spatial context and XBeach-informed baseline

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

3 Methods

3.1 Integrated model chain

The modelling system follows a chained physical-statistical structure (Figure 3). First, in situ observations and gridded marine, atmospheric, river-discharge, and coastal-geometry datasets define the forcing and environmental conditions. Second, the XBeach-informed physical baseline represents the nearshore response to storm-surge, wave, and morphology controls, providing a reference spatial inundation field. Third, Monte Carlo simulation is used to expand the range of physically plausible forcing combinations. Finally, a tree-based residual surrogate learns the correction between the physical baseline and the target inundation response, allowing rapid estimation of inundation depth and threshold-based flood extent. The downstream visualisation prototype is used only to display trained-model outputs and explore scenarios; the scientific assessment presented in this paper is based on the modelling chain, validation metrics, residual diagnostics, and uncertainty analysis.

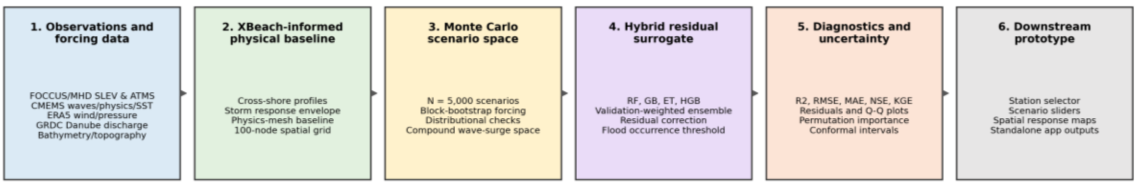

Figure 3. Chained physical–statistical workflow used to develop and evaluate the XBeach-informed machine-learning surrogate.

3.2 Monte Carlo scenario generation

A Monte Carlo simulation (MCS) framework was used to generate 5,000 forcing scenarios covering a broad but physically plausible range of coastal hazard conditions. The scenario matrix includes both sampled variables and process-derived predictors, including wind speed and direction, significant wave height, peak wave period, wave direction, sea-level elevation, mean sea-level pressure, Danube discharge, sea-surface temperature, storm duration, inverse-barometer contribution, wave-power-related terms, and interaction terms between sea-level anomaly and wave forcing. Where the dependence structure between variables is relevant, empirical or block-bootstrap sampling is used to preserve part of the observed co-variability in the forcing record. The consistency of the sampled variables with the observed or aligned forcing data is evaluated using distributional comparisons, including histogram and kernel-density estimates, together with Earth Mover's Distance (EMD) as a quantitative measure of marginal distributional agreement (Table 3, Figure 4).

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

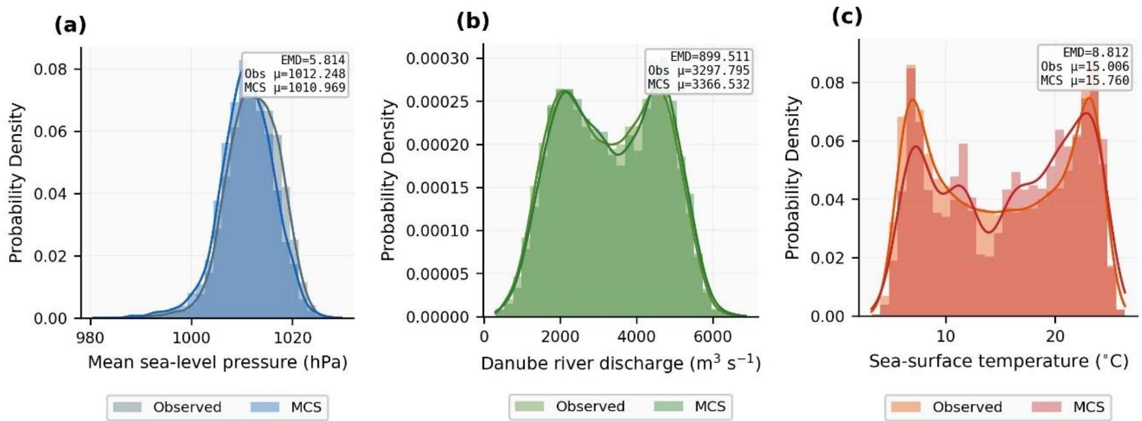

Figure 4. Monte Carlo distributional validation for selected forcing variables: (a) mean sea-level pressure, (b) Danube River discharge and (c) sea-surface temperature. Histograms and kernel-density estimates compare observed or aligned forcing values with Monte Carlo realisations. The metrics shown in each panel provide distributional quality-control information rather than model-skill statistics.

Table 3. Summary validation of Monte Carlo forcing distributions against the aligned forcing dataset.

Variable	Unit	Observed mean	MCS mean	Observed std	MCS std
wind_speed_ms	m s^{-1}	8.061	8.412	2.357	3.176
Hs_m	m	1.273	1.355	0.3809	0.5434
pressure_hPa	hPa	1012.2	1011.0	5.213	5.331
danube_Q_m3s	m^3 s^{-1}	3297.8	3366.5	1299.6	1301.8
SST_C	degC	15.01	15.76	6.415	6.208
sea_level_m	m	0.2076	0.2962	0.2828	0.3002

3.3 XBeach-informed physical baseline and spatial target

The XBeach-informed component provides an intermediate physical-response layer used before surrogate-model training. It combines simplified representative cross-shore profiles with design-of-experiments forcing combinations to describe how sea-level anomaly and wave forcing translate into nearshore inundation response (Figure 5). The cross-shore profiles represent the main differences in coastal exposure and nearshore morphology among the monitored sectors, while the forcing combinations define a physically interpretable response surface for water-level and wave-driven inundation.

This component is used as a physics-informed baseline rather than as a fully calibrated, site-specific XBeach implementation for every western Black Sea coastal segment. Its role is to provide a consistent reference response against which the machine-learning surrogate can learn residual corrections. The spatial target is represented on a 10 × 10 grid, corresponding to 100 output nodes for each Monte Carlo scenario. The surrogate therefore predicts spatial inundation depth at grid-node level and supports the identification of threshold-based flood extent.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

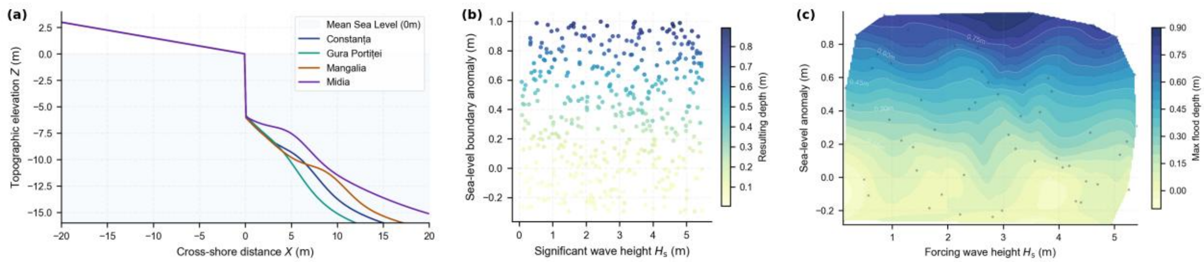

Figure 5. XBeach-informed intermediate response layer: (a) representative cross-shore transects, (b) design-of-experiments forcing matrix linking sea-level anomaly and significant wave height to flood-depth response, and (c) storm-sequence response used to compare the physics baseline with the surrogate target envelope.

3.4 Hybrid residual surrogate

The final surrogate was formulated as a residual correction to the XBeach-informed physical baseline. In this formulation, the physical layer first provides a baseline estimate of inundation depth, and the machine-learning component then predicts the residual term, defined as the difference between the target inundation response and the baseline estimate. This structure preserves the physical interpretability of the model because the baseline response remains explicitly visible, while the residual model accounts for the part of the inundation signal that is not captured by the simplified physical-response layer.

Four tree-based regression models were trained: Random Forest (RF), Gradient Boosting (GB), Extra Trees (ET), and Histogram Gradient Boosting (HGB). The individual model predictions were combined into a validation-weighted ensemble, using weights derived from their validation performance (Table 4). The resulting weights are nearly equal, indicating that the four models contributed similarly to the ensemble prediction. This ensemble formulation reduces dependence on a single algorithm and provides a more stable residual correction across the Monte Carlo scenario space.

The residual-surrogate design also supports model interpretation. When the predicted residual is small, the XBeach-informed baseline explains most of the simulated inundation response. When the residual correction is large, feature-importance diagnostics can be used to identify which forcing variables or morphology-related descriptors contribute most strongly to the correction.

Table 4. Validation-derived model weights used in the ensemble residual surrogate.

Model	Validation weight
Random Forest (RF)	0.2499
Gradient Boosting (GB)	0.2501
Extra Trees (ET)	0.2499
Histogram Gradient Boosting (HGB)	0.2501

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

3.5 Validation metrics and uncertainty

Model performance was evaluated using complementary regression, classification, residual, and uncertainty diagnostics. Continuous inundation-depth predictions were assessed using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean bias, Nash–Sutcliffe efficiency (NSE), Kling–Gupta efficiency (KGE), Pearson correlation, and Spearman rank correlation. These metrics jointly describe the magnitude of prediction errors, systematic overprediction or underprediction, linear and rank-based association, and model skill relative to simple reference predictions.

Threshold-based flood extent was evaluated by treating each grid node as flooded or non-flooded according to an operational inundation threshold of 0.30 m. Binary classification skill was quantified using accuracy, precision, recall, F1 score, Matthews correlation coefficient (MCC), Cohen's kappa, area under the receiver operating characteristic curve (AUC-ROC), and average precision. This separate classification assessment is necessary because identifying flooded nodes and reproducing continuous inundation depth represent different levels of predictive difficulty.

Residual behaviour was analysed using residual histograms, fitted normal-density curves, residual-versus-predicted plots, and quantile–quantile (Q–Q) diagnostics. These diagnostics were used to identify systematic bias, non-normal error structure, heteroscedasticity, and tail errors associated with the largest inundation values.

Predictive uncertainty was quantified using split conformal prediction intervals. Under the assumption of exchangeability between calibration and test data, conformal intervals provide distribution-free marginal coverage. However, their practical value for coastal-hazard interpretation also depends on interval sharpness, calibration across the inundation-depth range, and robustness under possible distribution shifts, particularly for rare or extreme flood scenarios.

4 Results

4.1 Dataset structure and spatial target

The final modelling dataset consists of 5,000 Monte Carlo scenarios, each described by 22 forcing and derived predictors. The predictors represent the main dynamic and static controls used in the surrogate framework, including sea-level anomaly, wave forcing, wind conditions, atmospheric pressure, Danube discharge, sea-surface temperature, and simplified coastal-geometry descriptors. For each scenario, the target response is represented on a 10×10 spatial grid, corresponding to 100 output nodes. The resulting target matrix therefore has a dimension of $5{,}000 \times 100$, with each row representing one forcing scenario and each column representing one spatial grid node.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

The target variable is inundation depth at grid-node level. This structure allows the surrogate to be evaluated both as a continuous predictor of spatial inundation depth and as a threshold-based classifier of flood extent.

4.2 Continuous inundation-depth skill

The conservative held-out node-level evaluation indicates moderate quantitative skill for continuous inundation-depth prediction. The ensemble improves substantially over the XBeach-informed physical baseline, reducing RMSE from 1.868 to 1.181 m and MAE from 1.319 to 0.679 m (Table 5). The ensemble reaches $R^2 = 0.409$, NSE = 0.409, KGE = 0.235, Pearson r = 0.914, and Spearman $\rho = 0.923$. These values indicate that the surrogate preserves much of the event ranking and spatial response structure, but does not fully reproduce the magnitude of the largest inundation depths.

The parity and residual diagnostics confirm this interpretation. The ensemble follows the dominant inundation-response pattern, while the residual distribution and Q–Q plot reveal non-Gaussian errors, systematic negative bias, and tail departures associated with extreme-depth nodes (Figure 6). The model should therefore be interpreted cautiously for quantitative estimation of maximum inundation depth under the most severe conditions.

Table 5. Conservative node-level test performance for continuous spatial inundation-depth prediction. RMSE and MAE are expressed in centimetres. NSE = Nash–Sutcliffe efficiency; KGE = Kling–Gupta efficiency.

Model	R2	RMSE (cm)	MAE (cm)	NSE	KGE	Pearson r	Spearman rho
RF	0.4082	118.1	68.18	0.4082	0.2339	0.9150	0.9258
GB	0.4090	118.1	67.67	0.4090	0.2351	0.9140	0.9271
ET	0.4080	118.2	68.20	0.4080	0.2349	0.9133	0.9264
HGB	0.4091	118.1	67.62	0.4091	0.2357	0.9131	0.9271
Ensemble	0.4087	118.1	67.85	0.4087	0.2348	0.9142	0.9230
Physics baseline	-0.4791	186.8	131.9	-0.4791	-0.5740	0.0260	0.0583

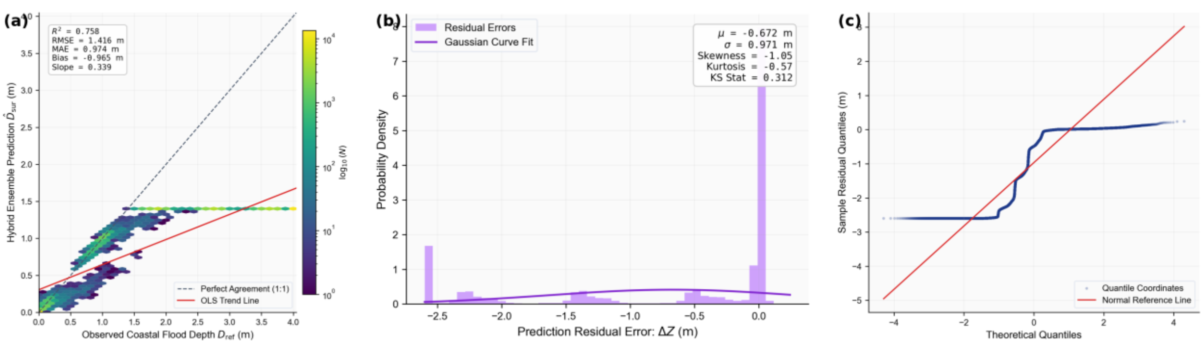

Figure 6. Core validation diagnostics for the spatial residual surrogate: (a) density-coloured predicted-versus-reference flood-depth scatter, (b) residual error distribution and (c) normal Q-Q plot. The diagnostics show that the

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

model captures the dominant event structure but that residuals remain non-Gaussian and biased for extreme-depth nodes

4.3 Threshold-based flood-extent classification

Flood extent was evaluated separately by classifying each grid node as flooded or non-flooded using the 0.30 m operational inundation threshold. This threshold-based evaluation shows substantially higher skill than the continuous inundation-depth regression. The ensemble reaches F1 = 0.995, MCC = 0.989, Cohen's κ = 0.989, and AUC-ROC = 0.9998 (Table 6), while the thresholded physics baseline remains close to no-skill behaviour.

These results indicate that the surrogate is highly effective in identifying flooded grid nodes and delineating threshold-based flood extent. In its present configuration, the model is therefore most robust for rapid flood-occurrence screening, scenario ranking, and identifying locations or forcing combinations where more detailed process-based simulations may be required. This classification ability should be interpreted separately from the continuous inundation-depth results because detecting threshold exceedance is less demanding than reproducing the exact magnitude of the maximum inundation depths.

Table 6. Threshold-based flood-extent classification skill at the 0.30 m operational inundation threshold. MCC = Matthews correlation coefficient; κ = Cohen's kappa; AUC-ROC = area under the receiver operating characteristic curve.

Model	F1	MCC	kappa	AUC-ROC
RF	0.9930	0.9859	0.9858	0.9995
GB	0.9963	0.9925	0.9925	0.9999
ET	0.9930	0.9859	0.9858	0.9991
HGB	0.9963	0.9925	0.9925	0.9998
Ensemble	0.9948	0.9895	0.9894	0.9998
Physics baseline	0.4856	-0.0028	-0.0028	0.4991

4.4 Residual structure and extreme events behaviour

Residual diagnostics show that the main limitation of the surrogate is concentrated in the upper tail of the inundation-depth distribution. Across the full node-level test dataset, the mean residual is approximately −0.672 m, indicating a systematic tendency towards underprediction. The residual distribution is negatively skewed, and the Q–Q diagnostics depart from the normal reference line, confirming that the errors are not normally distributed and that the largest deviations occur for the highest reference inundation depths.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

This behaviour is most evident for events above the 90th percentile of the reference inundation-depth distribution. For these upper-tail cases, the ensemble shows a mean bias of approximately −2.59 m and an RMSE of approximately 2.59 m (Table 7). The similar values of MAE, RMSE, and bias indicate a systematic rather than random underestimation of the most severe inundation responses. This upper-tail bias is therefore a key constraint on the present model configuration and limits its use for direct quantitative prediction of maximum inundation depth. The model remains more reliable for threshold-based flood-extent screening than for estimating the exact magnitude of extreme inundation depths.

Table 7. Extreme-event diagnostics for the ensemble model. Negative bias indicates underprediction of upper-tail flood-depth nodes.

Quantile	Threshold (m)	n events	MAE (cm)	RMSE (cm)	Bias (cm)
0.9000	3.966	11520.0	259.4	259.4	-259.4
0.9500	4.000	6034.0	260.0	260.0	-260.0
0.9900	4.000	6034.0	260.0	260.0	-260.0

4.5 Feature importance and physical controls

Permutation-importance diagnostics indicate that the XBeach-informed physical baseline is the dominant predictor in the residual surrogate (Table 8; Figure 7). This result confirms that the intermediate physical-response layer carries most of the first-order information on coastal geometry, water-level response, and spatial inundation structure. The residual machine-learning component therefore acts mainly as a correction to this baseline rather than as an entirely independent predictor of inundation.

After the physical baseline, the most influential predictors are wind direction, sea-surface temperature, and the interaction between sea-level anomaly and significant wave height. This ranking is physically consistent. Wind direction controls coastal exposure, onshore forcing, and storm-surge setup along different coastal sectors. Sea-surface temperature acts as a seasonal and hydrographic-state proxy, reflecting the broader environmental conditions under which storm events occur. The sea-level–wave-height interaction captures the compound effect of elevated water levels and wave forcing, which is central to nearshore inundation and threshold exceedance.

The remaining predictors, including wave direction, wind stress, sea-level anomaly, significant wave height, onshore wind speed, and wave-power-related terms, have smaller individual contributions in the permutation test. Their low individual importance does not imply that these processes are physically irrelevant; rather, part of their effect is already represented through the XBeach-informed baseline and through correlated compound predictors.

Overall, the feature-importance results show that the surrogate response remains physically interpretable, but they also confirm that the present model limitation is not solved by predictor ranking alone. The upper-tail

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

underprediction identified in Sect. 4.4 is more likely related to the simplified baseline representation, limited morphology descriptors, and insufficient sampling of the most severe inundation conditions..

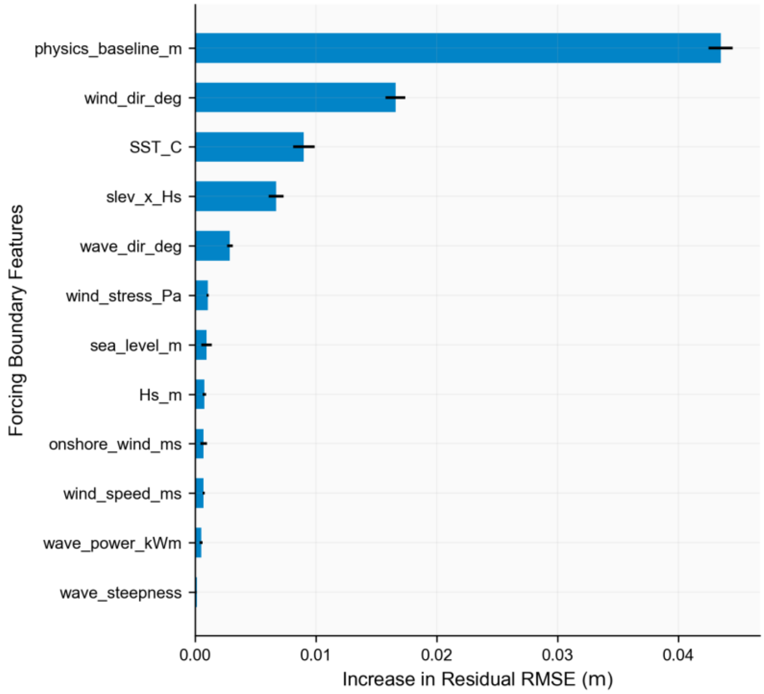

Figure 7. Permutation importance for the residual surrogate. Importance is expressed as the increase in residual RMSE when each predictor is permuted. The XBeach-informed physical baseline is the dominant predictor, followed by wind direction, sea-surface temperature, and the sea-level–wave-height interaction. Error bars indicate the variability of the permutation estimate across repeated permutations.

Table 8. Ten leading predictors from the permutation-importance analysis. Importance is reported as the increase in residual RMSE after predictor permutation. Larger values indicate stronger influence on the residual-surrogate prediction.

Feature	Increase in residual RMSE (m)	Std.
physics_baseline_m	0.0435	0.0010
wind_dir_deg	0.0166	0.0008
SST_C	0.0090	0.0009
slev_x_Hs	0.0067	0.0006
wave_dir_deg	0.0029	0.0002
wind_stress_Pa	0.0011	0.0001
sea_level_m	0.0009	0.0004
Hs_m	0.0008	0.0001
onshore_wind_ms	0.0007	0.0003
wind_speed_ms	0.0007	0.0001

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

4.6 Uncertainty calibration

Split conformal prediction intervals were used to evaluate whether the surrogate provides uncertainty estimates that are consistent with the observed test errors. The reliability diagram shows that the empirical coverage is close to the nominal level for the 90 % interval, with PI90 reaching an empirical coverage of 0.911 for a nominal coverage of 0.90 (Figure 8; Table 9). This indicates that the conformal intervals are approximately calibrated in a marginal sense.

However, interval sharpness remains limited for decision-relevant quantitative inundation-depth prediction. The PI90 interval has a mean width of 3.25 m, and the PI95 interval reaches full empirical coverage only with a similarly wide interval. In contrast, the PI50 interval is much narrower but slightly undercovers the test data, with empirical coverage of 0.458 for a nominal coverage of 0.50. These results indicate that the conformal intervals are useful for communicating uncertainty magnitude and for identifying predictions that should be interpreted cautiously, but they should not be treated as precise local confidence bands for inundation depth.

The large width of the higher-coverage intervals is consistent with the residual diagnostics and upper-tail underprediction described in Sect. 4.4. In the present model configuration, conformal uncertainty estimates therefore provide a conservative indication of predictive uncertainty, particularly for severe inundation scenarios, but further calibration by depth range, station, or morphology class would be needed before operational quantitative use.

Table 9. Split conformal prediction interval coverage and sharpness on the test set. PI50, PI90, and PI95 denote prediction intervals with nominal coverage of 50 %, 90 %, and 95 %, respectively.

Interval	Nominal coverage	Empirical coverage	Mean width (cm)	Absolute residual quantile (m)
PI50	0.5000	0.4578	2.806	0.0167
PI90	0.9000	0.9112	324.7	2.580
PI95	0.9500	1.000	326.7	2.600

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

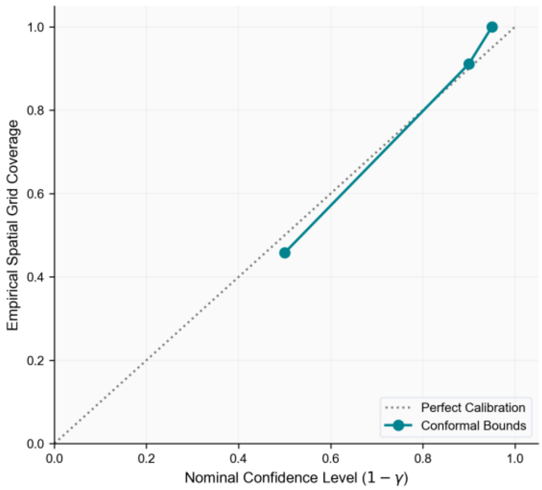

Figure 8. Conformal reliability diagram for the spatial surrogate. The dashed line indicates perfect calibration, while the markers show empirical spatial-grid coverage for the conformal prediction intervals. Coverage is close to nominal for the 90 % interval, but interval width remains large; therefore, the intervals are interpreted as uncertainty diagnostics rather than as precise local confidence bands for inundation depth.

4.7 Coastal-site and seasonal validation

Site-specific diagnostics indicate that the ensemble performance is relatively consistent across the evaluated coastal locations, but with a persistent negative bias at all sites (Table 10). RMSE values remain within a narrow range of approximately 1.15–1.20 m, while MAE values are generally between 0.65 and 0.70 m. The highest site-specific R^2 is obtained at Mamaia ($R^2 = 0.428$), whereas the lowest value occurs at Midia ($R^2 = 0.396$). These differences are modest relative to the total error magnitude, suggesting that the current limitation is not restricted to a single station or coastal sector.

The site-specific residual structure confirms the systematic underprediction already identified in the node-level diagnostics (Figure 9a). The larger spread of residuals at Gura Portitei and Midia indicates that the surrogate is less stable in sectors where local morphology, exposure, or boundary-condition variability may be less well represented by the simplified physical baseline and static descriptors. This reinforces the need for richer morphology information and event-based local calibration before quantitative operational use.

Seasonal validation also shows a relatively stable pattern across the year (Table 11). R^2 values remain close to 0.40 for all seasons, with slightly better performance in spring (MAM; $R^2 = 0.420$) and slightly weaker performance in winter (DJF; $R^2 = 0.400$). The absence of a strong seasonal contrast indicates that the main source of error is structural rather than confined to a particular season. In other words, the model limitation appears to arise primarily from the representation of the physical baseline, the spatial target, and the upper-tail inundation response, rather than from a seasonal failure of the forcing data.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

The storm-sequence diagnostic provides an additional process-based check (Figure 9b). The surrogate target follows the timing and shape of the reference XBeach-type response more closely than the simplified physics-mesh baseline, particularly during the rising limb and peak of the storm sequence. However, this diagnostic should be interpreted as a consistency check within the modelling workflow, not as independent observational validation. It supports the usefulness of the residual-surrogate approach while confirming that further event-based validation is required for operational quantitative inundation-depth prediction.

Table 10. Site-specific validation metrics for the ensemble spatial surrogate. Negative bias indicates systematic underprediction of inundation depth.

Station	n node samples	R2	RMSE (cm)	MAE (cm)	Bias (cm)
Constanța	15500	0.4057	118.6	68.35	-67.57
Eforie Nord	15800	0.4166	116.8	66.62	-66.12
Gura Portiței	15000	0.4081	118.2	67.91	-67.24
Mamaia	18000	0.4279	114.6	64.56	-64.07
Mangalia	16400	0.4025	119.2	68.85	-68.24
Midia	18600	0.3962	120.2	70.05	-68.93
Vama Veche	15900	0.4035	119.0	68.63	-67.99

Table 11. Blocked seasonal validation metrics for the ensemble spatial surrogate. Seasons are defined as DJF, MAM, JJA, and SON.

Season	n node samples	R2	RMSE (cm)	MAE (cm)
DJF	21600	0.3998	119.7	69.84
JJA	36000	0.4047	118.8	68.45
MAM	28800	0.4200	116.0	65.81
SON	28800	0.4091	118.0	67.65

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

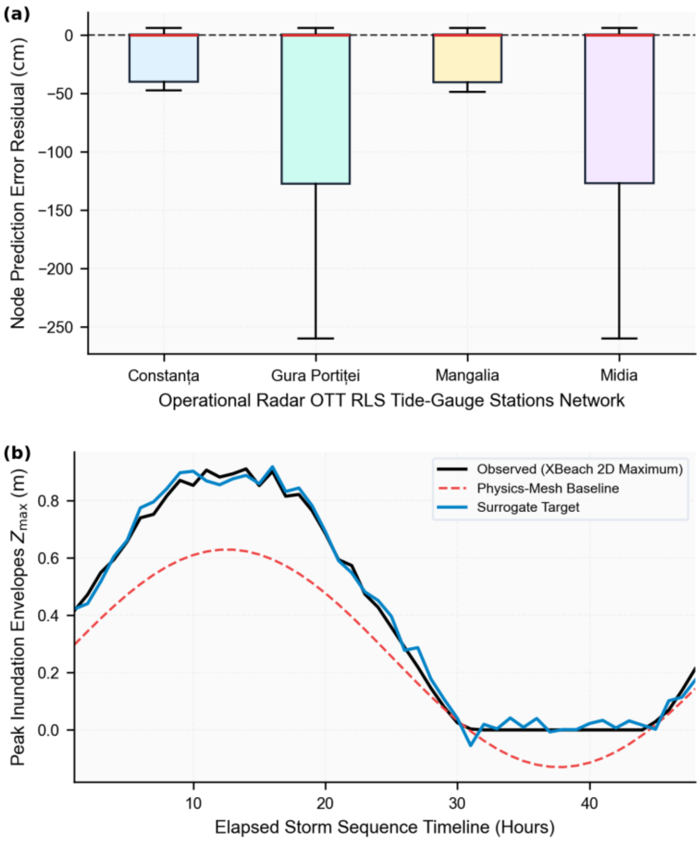

Figure 9. Site-specific and storm-sequence validation diagnostics for the ensemble surrogate. (a) Residual distribution at selected coastal locations, expressed as node-level prediction error. (b) Storm-sequence response comparing the reference XBeach-type maximum inundation response, the simplified physics-mesh baseline, and the surrogate target response. The diagnostics indicate that the surrogate improves the representation of storm-sequence behaviour relative to the baseline, while systematic negative residuals remain evident across the evaluated coastal sites.

5 Discussion

5.1 Scientific interpretation of the model skill pattern

The main scientific result is not represented by a single performance metric, but by the contrast between threshold-based flood-extent classification and continuous inundation-depth prediction. The surrogate identifies flooded and non-flooded grid nodes with high skill above the 0.30 m operational inundation threshold, but it remains less accurate in reproducing the magnitude of the largest inundation depths. This distinction is important because flood occurrence and inundation-depth magnitude represent different prediction problems. The first depends mainly on

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

whether the model captures threshold exceedance, whereas the second requires an accurate representation of the full spatial and upper-tail response.

The observed skill pattern indicates that the surrogate learns the dominant inundation-response structure and preserves much of the event ranking, but does not yet fully resolve the most severe inundation conditions. The systematic underprediction of upper-tail values likely reflects a combination of factors: the limited representation of rare severe events in the training space, the simplified nature of the XBeach-informed physical baseline, the idealised spatial grid, and the absence of richer local morphology descriptors such as dune-crest elevation, beach width, backshore slope, harbour exposure, lagoon connectivity, and surface roughness. As a result, the model is better suited in its present form to rapid flood-extent screening and scenario prioritisation than to stand-alone quantitative prediction of maximum inundation depth.

5.2 Implications for coastal early warning

The present surrogate is best interpreted as a rapid screening and prioritisation tool rather than as a deterministic predictor of absolute inundation depth. Its main operational value lies in the rapid identification of likely flooded grid nodes, the ranking of forcing scenarios according to relative hazard level, and the identification of coastal sectors where more detailed process-based simulations may be required. In this role, the surrogate can complement, rather than replace, full XBeach or hydrodynamic modelling by reducing the number of scenarios that require computationally expensive high-resolution simulations.

This type of framework is relevant for Copernicus downstream coastal services because it links in situ observations, gridded marine and atmospheric products, physical-response modelling, and uncertainty diagnostics within a computationally efficient workflow. The model can provide rapid, interpretable guidance for scenario exploration, sensitivity testing, and preliminary flood-extent assessment. However, operational use for quantitative inundation-depth forecasting would require further validation against independent flood, overtopping, or impact observations, together with locally calibrated XBeach simulations and improved representation of site-specific morphology.

5.3 Why the extreme tail remains difficult

The reduced skill for upper-tail inundation depths is likely caused by a combination of sampling, physical-representation, and spatial-description limitations. First, the most severe inundation cases represent only a small fraction of the training dataset. As a result, the surrogate has fewer examples from which to learn the response of rare but high-impact compound events. In addition, the concentration of target values near the imposed upper bound of approximately 4 m creates a plateau in the target distribution, which limits the ability of the model to distinguish between different levels of severe inundation.

Second, the XBeach-informed physical baseline is intentionally simplified and does not resolve all local processes that may control extreme inundation. These include wave overtopping pathways, backshore storage, small-scale topographic barriers, harbour geometry, lagoon connectivity, and local variations in surface roughness. Such processes can strongly affect maximum inundation depth, especially during severe wave–surge events, but they are only partly represented in the current baseline.

Third, the spatial grid and morphology descriptors remain idealised. They provide a consistent framework for rapid screening, but they do not yet fully distinguish between dune-backed beaches, engineered harbour areas, low-lying lagoonal sectors, cliff-backed coasts, and urbanised backshore environments. This limitation helps explain why the surrogate captures flood occurrence more reliably than the exact magnitude of upper-tail inundation depths.

Future development should therefore prioritise event-based XBeach calibration at representative coastal sites, targeted sampling of severe compound wave–surge scenarios, improved morphology descriptors, and independent validation against observed flood, overtopping, or impact data. These steps are necessary before the framework can be used for operational quantitative prediction of maximum inundation depth.

These limitations do not reduce the value of the present surrogate as a rapid screening tool, but they define the boundary between exploratory flood-extent assessment and operational quantitative inundation-depth forecasting. Further improvement would require stronger event-based calibration of the XBeach-informed baseline, better representation of local morphology, and independent validation against observed flood or overtopping impacts.

6 Conclusions

In this paper we developed and evaluated an XBeach-informed machine-learning surrogate for rapid coastal inundation and flood-extent screening along the Western Black Sea coast. The framework combines FOCCUS/MHD sea-level observations, Copernicus Marine wave and physical products, ERA5 atmospheric forcing, GRDC Danube discharge, bathymetric and coastal-elevation descriptors, an intermediate XBeach-informed physical-response layer, and tree-based residual emulation.

The results show a clear distinction between continuous inundation-depth prediction and threshold-based flood-extent classification. At node level, the ensemble provides moderate quantitative skill for inundation-depth prediction, but it substantially improves on the simplified physical baseline. Its strongest performance is obtained for identifying flooded grid nodes above the 0.30 m operational threshold, indicating that the framework is most reliable for rapid flood-occurrence screening, scenario ranking, and preliminary decision-support analysis.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

Residual diagnostics and extreme-event evaluation identify the main limitation of the current configuration. The model systematically underpredicts upper-tail inundation depths, with an average bias of approximately −2.6 m for the most severe cases. This limitation reflects the combined effects of rare-event sampling, the simplified XBeach-informed baseline, idealised spatial representation, and limited morphology descriptors. Consequently, the present surrogate should not be interpreted as a stand-alone operational tool for quantitative prediction of maximum inundation depth.

Despite these limitations, the model chain represents a useful and reproducible step towards rapid coastal-hazard screening for the Black Sea coast. It provides a reproducible, interpretable and computationally efficient framework that can be improved incrementally while preserving a clear link between observations, physical forcing, surrogate prediction and uncertainty diagnostics. In this role, the framework can support Copernicus downstream applications by rapidly identifying potentially affected coastal sectors and by guiding where more detailed event-based XBeach or hydrodynamic simulations are needed.

Appendix A. Auxiliary diagnostic figures

This appendix provides supplementary diagnostic figures supporting the modelling workflow described in the main text. The figures are included to document the construction of the Monte Carlo forcing space and the auxiliary checks used during the development of the XBeach-informed physical-response layer. They are not used as independent validation results, but they provide additional methodological transparency and help clarify how the forcing scenarios and simplified physical baseline were constructed.

Figure A1 shows the joint Monte Carlo scenario space for storm-surge sea level and significant wave height. Each point represents one Monte Carlo realisation, while the density shading indicates the concentration of sampled sea-level–wave combinations. The vertical and horizontal dashed lines mark the 95th-percentile thresholds for storm-surge sea level and significant wave height, respectively. The upper-right quadrant therefore identifies compound high-water and high-wave scenarios, which are relevant for severe coastal inundation screening. This diagnostic confirms that the scenario generator covers both the central forcing space and the more hazardous compound tail of the distribution.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

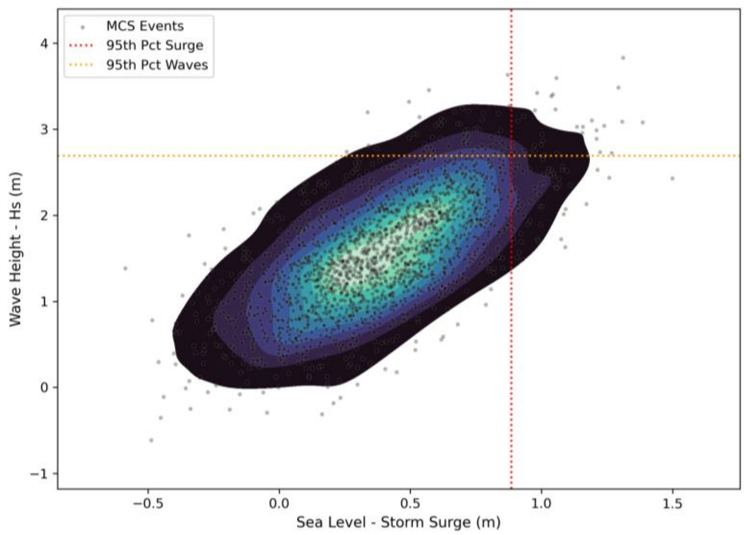

Figure A1. Compound sea-level and wave-height scenario space generated by the Monte Carlo simulation framework. Grey points represent individual Monte Carlo scenarios, while the density shading indicates the concentration of sampled forcing combinations. The vertical and horizontal dashed lines mark the 95th-percentile thresholds for storm-surge sea level and significant wave height, respectively. The upper-right sector identifies compound high-water and high-wave conditions relevant for severe coastal inundation screening.

Figure A2 summarises auxiliary lookup and design-of-experiments checks used during construction of the XBeach-informed physical-response layer. These diagnostics were used to examine how combinations of sea-level anomaly and wave forcing translate into the simplified inundation-response surface used as the physical baseline. The purpose of this figure is to document the intermediate response layer that links the forcing variables to the spatial inundation target before residual-surrogate training. It therefore supports the methodological traceability of the model chain, while the quantitative performance of the final surrogate is assessed in the main Results section.

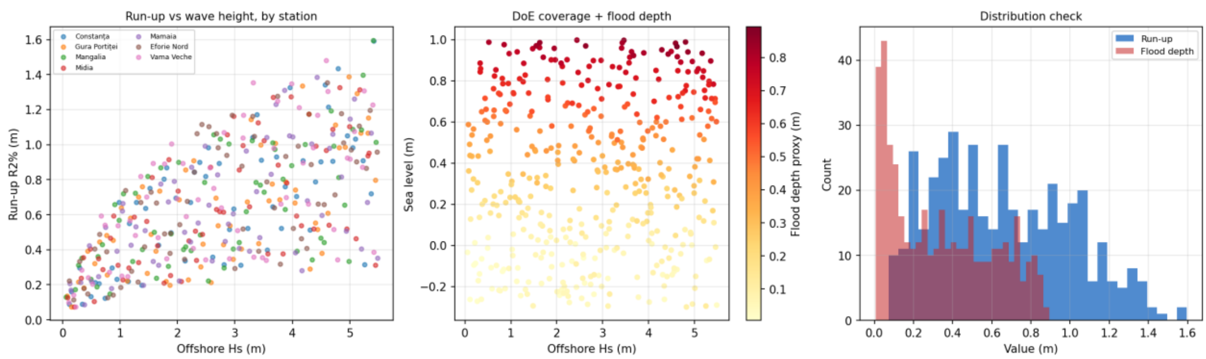

Figure A2. Auxiliary lookup and design-of-experiments diagnostics used during development of the XBeach-informed physical-response layer. These diagnostics support the construction of the simplified physical baseline and illustrate

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

how combinations of sea-level anomaly and wave forcing are translated into an inundation-response surface before residual-surrogate training.

Appendix B. Decision-support visualisation interface

A prototype decision-support visualisation interface was developed to demonstrate how the trained surrogate outputs can be explored in an interactive setting. The interface links user-defined forcing conditions with the trained model response and displays simplified diagnostic outputs relevant for coastal-flood screening. Users can select a coastal site, adjust forcing variables such as sea level, wave height, storm duration, atmospheric pressure, Danube discharge, and wind speed or direction, and inspect the resulting flood-occurrence class and associated response indicators.

The interface is designed as a communication and scenario-exploration layer rather than as an operational warning system. Its role is to illustrate how the surrogate model could be translated into a practical downstream tool for rapid screening, visual inspection, and comparison of plausible coastal-hazard scenarios. The scientific evaluation of the model remains based on the validation metrics, residual diagnostics, feature-importance analysis, and uncertainty assessment presented in the main text. Therefore, the interface outputs should be interpreted as demonstrative visualisations of the trained surrogate response, not as independent validation results or operational forecasts.

Figure B1 summarises the interface architecture. User controls define the scenario input variables, including station or site selection, sea level, wave height, storm duration, atmospheric pressure, Danube discharge, and wind forcing. These inputs are passed through the dashboard callback engine, which synchronises selected events, computes total water-level indicators, updates maps, profiles, and diagnostic cards, and exports standalone visual outputs. The decision-output layer displays cross-shore flood profiles, simplified spatial response layers, coastal flooding cards, erosion and port-operability indicators, and plume or sea-surface-temperature stress indicators.

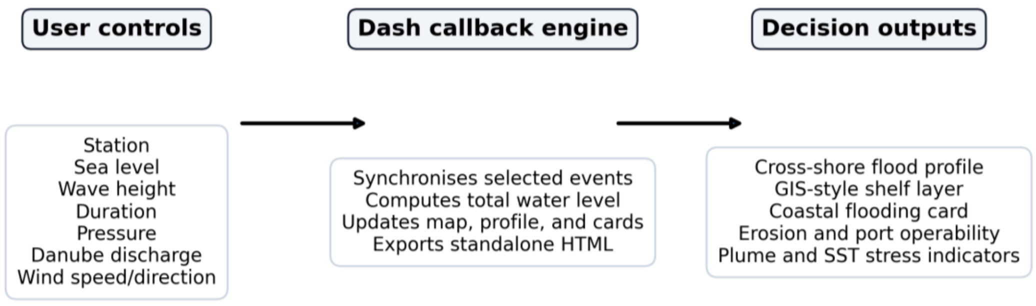

Figure B1. Decision-support interface architecture linking model inputs, surrogate prediction and scenario visualisation.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

Figure B2 illustrates an example view of the prototype coastal-hazard visualisation interface (.html format) developed as a downstream demonstrator of the surrogate workflow. The interface allows the user to select a coastal site or forcing scenario and to adjust the main environmental drivers, including sea level, storm-surge or wave-height forcing, storm duration, atmospheric pressure, Danube discharge, wind speed, and wind direction. The central panel displays the corresponding cross-shore response, while the lower diagnostic cards summarise simplified indicators for coastal flooding, beach erosion, port operability, compound hazard, Danube plume salinity, and thermal stress.

The interface is intended for scenario exploration, communication, and rapid screening of trained-surrogate outputs. It is not used as independent validation evidence and should not be interpreted as an operational warning system. The scientific assessment of the model remains based on the quantitative validation metrics, residual diagnostics, feature-importance analysis, and uncertainty evaluation presented in the main text.

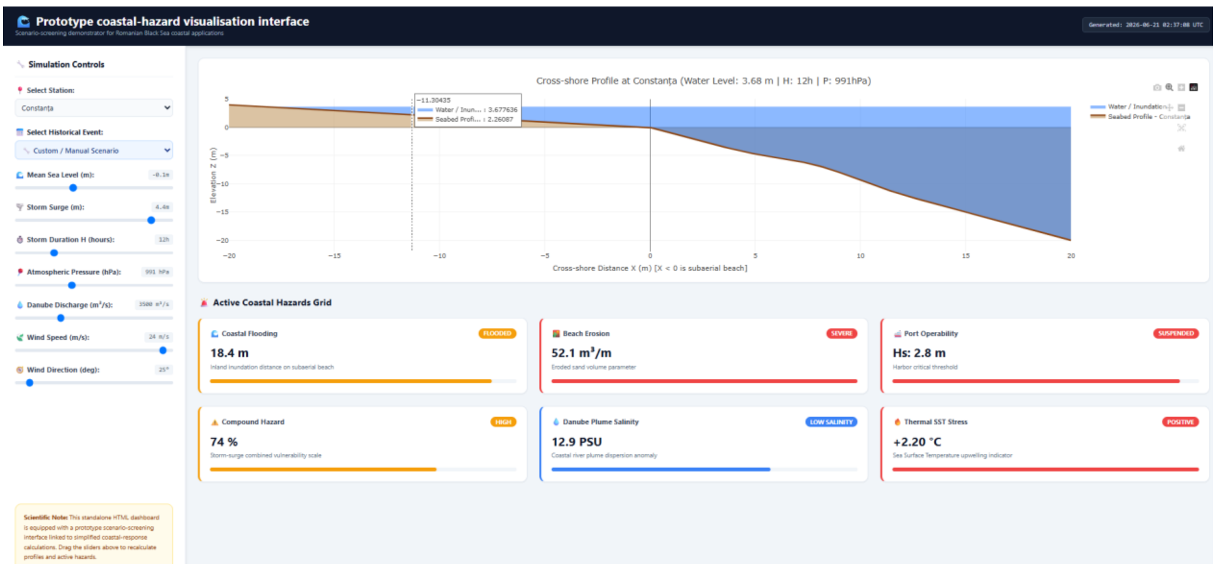

Figure B2. Example screenshot of the prototype coastal-hazard visualisation interface. The left panel contains user controls for site selection and forcing conditions, the central panel shows the cross-shore water-level and inundation response, and the lower cards summarise simplified coastal-hazard indicators. The interface is included as a downstream demonstrator for scenario screening and communication of trained-surrogate outputs; it is not used as independent model-validation evidence. Source and image credit: authors' own web-based prototype, developed in Python using Dash and Plotly; screenshot produced by the authors.

Appendix C. Supporting performance and input-diagnostic tables

This appendix provides additional diagnostic tables that support the interpretation of the surrogate-model results presented in the main text. These tables are included for transparency and reproducibility, but they are not used as

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

the primary basis for the scientific conclusions. The main interpretation of model skill, uncertainty, and limitations is presented in Sects. 4 and 5.

Table C1. Grouped predictor-sensitivity diagnostics for the residual surrogate. The table reports the change in RMSE after grouped predictor perturbation. Positive ΔRMSE values indicate that the predictor group contributes useful information to the model prediction, while values close to zero indicate limited additional influence in the present configuration.

Feature group	Base RMSE (cm)	Perturbed RMSE (cm)	ΔRMSE (cm)	Number of features (n)
XBeach-informed baseline	118.1	118.2	0.0964	1
Wind and pressure	118.1	118.2	0.0848	6
Sea-surface temperature (SST)	118.1	118.2	0.0494	1
Sea-level (SLEV)	118.1	118.2	0.0208	2
Waves	118.1	118.2	0.0185	7
Location and morphology	118.1	118.1	0.0001	3
Danube discharge	118.1	118.1	-0.0072	2

Table C2. Summary statistics and quality-control overview for the input forcing variables used in the analysis. The table reports the mean, standard deviation, minimum, maximum, and number of valid records after alignment of the forcing dataset.

Variable	Mean	Standard deviation (Std)	Minimum (Min)	Maximum (Max)	Number of valid records
Wind speed ($m\ s^{-1}$)	8.061	2.357	4.511	27.94	35064
Wind direction (°)	120.8	115.7	0.0048	360.0	35064
Mean sea-level pressure (hPa)	1012.2	5.213	980.6	1029.1	35064
Significant wave height, Hs (m)	1.273	0.3809	0.4844	4.846	35064
Peak wave period, Tp (s)	4.997	0.5544	2.983	8.100	35064
Wave direction (°)	124.9	116.0	0.0179	360.0	35064
Sea-surface temperature (°C)	15.01	6.415	3.229	26.38	35064
Danube discharge ($m^3\ s^{-1}$)	3297.8	1299.6	500.0	6858.5	35064
Sea-level elevation (m)	0.2076	0.2828	-0.2563	1.101	35064

Code and data availability

The computational workflow was implemented in Python using a notebook-based workflow developed in the Jupyter environment (Kluyver Thomas et al., 2016) and managed through the Anaconda Python distribution. The workflow includes data alignment, Monte Carlo scenario generation, XBeach-informed response processing, residual-surrogate training, validation diagnostics, uncertainty analysis, and figure production. The processed forcing matrix, spatial target tables, validation tables, trained model objects, and figure-generation scripts associated with this analysis will be made available in a DOI-bearing repository accompanying the final publication, subject to third-party data-licence constraints. The FOCCUS/MHD sea-level observations are cited

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

through the published dataset of (Dutu et al., 2026). All metrics, figures and tables reported in this manuscript are based on the aligned observational and gridded forcing dataset described in Sects. 2 and 3.

Author contributions

M.E.M. led the manuscript preparation, scientific framing, coastal-oceanographic interpretation, model development. A.S. and A.C. contributed to data preparation, and hydrographic interpretation. M.E.M. and L.D. contributed to validation design. M.E.M., A.S. and A.C. contributed to figure/table checking and interpretation of coastal-flood diagnostics. All authors reviewed the manuscript and approved the submitted version.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors acknowledge the FOCCUS project and the Maritime Hydrographic Directorate (MHD) sea-level monitoring network for the observational data used in this study, cited through the published Zenodo dataset (Dutu et al., 2026). The study was initiated by M.E.M. at MHD within the framework of the "Forecasting and Observing the Open-to-Coastal Ocean for Copernicus Users" — FOCCUS project (Forecasting and Observing the Open-to-Coastal Ocean for Copernicus, 2026). This work also made use of E.U. Copernicus Marine Service Information, ERA5 reanalysis data from the Copernicus Climate Change Service, Danube discharge records from the Global Runoff Data Centre (GRDC), and institutional support from the National Institute for Marine Research and Development "Grigore Antipa" (NIMRD) and MHD.

Financial support

The research of M.E.M. was carried out with financial support from the "Forecasting and Observing the Open-to-Coastal Ocean for Copernicus Users" — FOCCUS project, funded by the European Union (Grant Agreement No. 101133911). The research of M.E.M., A.S., and A.C. was also supported by the Nucleu Programme SMART-BLUE 2023–2026, funded by the Ministry of Research, Innovation and Digitalization, Romania, under contract no. 33N/2023 project PN23230101.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

References

Ayyad, M., Hajj, M. R., and Marsooli, R.: Machine learning-based assessment of storm surge in the New York metropolitan area, Sci. Rep., 12, 19215, https://doi.org/10.1038/s41598-022-23627-6, 2022.

Breiman, L.: Random Forests, 5–32 pp., https://doi.org/https://doi.org/10.1023/A:1010933404324, 2001.

Buongiorno Nardelli, B., Colella, S., Santoleri, R., Guarracino, M., and Kholod, A.: A re-analysis of Black Sea surface temperature, Journal of Marine Systems, 79, 50–64, https://doi.org/10.1016/j.jmarsys.2009.07.001, 2010.

Buongiorno Nardelli, B., Tronconi, C., Pisano, A., and Santoleri, R.: High and Ultra-High resolution processing of satellite Sea Surface Temperature data over Southern European Seas in the framework of MyOcean project, Remote Sens. Environ., 129, 1–16, https://doi.org/10.1016/j.rse.2012.10.012, 2013.

Camus, P., Haigh, I. D., Nasr, A. A., Wahl, T., Darby, S. E., and Nicholls, R. J.: Regional analysis of multivariate compound coastal flooding potential around Europe and environs: sensitivity analysis and spatial patterns, Natural Hazards and Earth System Sciences, 21, 2021–2040, https://doi.org/10.5194/nhess-21-2021-2021, 2021.

Chaigneau, A. A., Law-Chune, S., Melet, A., Voldoire, A., Reffray, G., and Aouf, L.: Impact of sea level changes on future wave conditions along the coasts of western Europe, Ocean Science, 19, 1123–1143, https://doi.org/10.5194/os-19-1123-2023, 2023.

Ciliberti, S. A., Grégoire, M., Staneva, J., Palazov, A., Coppini, G., Lecci, R., Peneva, E., Matreata, M., Marinova, V., Masina, S., Pinardi, N., Jansen, E., Lima, L., Aydoğdu, A., Creti', S., Stefanizzi, L., Azevedo, D., Causio, S., Vandenbulcke, L., Capet, A., Meulders, C., Ivanov, E., Behrens, A., Ricker, M., Gayer, G., Palermo, F., Ilicak, M., Gunduz, M., Valcheva, N., and Agostini, P.: Monitoring and Forecasting the Ocean State and Biogeochemical Processes in the Black Sea: Recent Developments in the Copernicus Marine Service, J. Mar. Sci. Eng., 9, 1146, https://doi.org/10.3390/jmse9101146, 2021.

Clare, M. C. A., Leijnse, T. W. B., McCall, R. T., Diermanse, F. L. M., Cotter, C. J., and Piggott, M. D.: Multilevel multifidelity Monte Carlo methods for assessing uncertainty in coastal flooding, Natural Hazards and Earth System Sciences, 22, 2491–2515, https://doi.org/10.5194/nhess-22-2491-2022, 2022.

Couasnon, A., Eilander, D., Muis, S., Veldkamp, T. I. E., Haigh, I. D., Wahl, T., Winsemius, H. C., and Ward, P. J.: Measuring compound flood potential from river discharge and storm surge extremes at the global scale, Natural Hazards and Earth System Sciences, 20, 489–504, https://doi.org/10.5194/nhess-20-489-2020, 2020.

Dutu, L., Mihailov, M. E., González Pérez, L., Criado Sudau, F. F. , S.-G. E. , R. R. E. , M. C. C., Soriano-González, J., Andrescu, D., Fernandez-Mora, A., Figueiredo, L., and Almeida, L. P.: Sea Level Romanian Black Sea, Zenodo, https://doi.org/https://doi.org/10.5281/zenodo.19133361, 2026.

EMODnet Bathymetry Consortium: EMODnet Bathymetry, European Marine Observation and Data Network, 2026.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

Forecasting and Observing the Open-to-Coastal Ocean for Copernicus: https://foccus-project.eu/, last access: 21 June 2026.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach. Learn., 63, 3–42, https://doi.org/10.1007/s10994-006-6226-1, 2006.

GRDC: https://portal.grdc.bafg.de/, last access: 20 June 2026.

Han, S. and Tahvildari, N.: Compound Flooding Hazards Due To Storm Surge and Pluvial Flow in a Low-Gradient Coastal Region, Water Resour. Res., 60, https://doi.org/10.1029/2023WR037014, 2024.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), Copernicus Climate Change Service, Climate Data Store, https://doi.org/https://doi.org/10.24381/cds.adbb2d47, 2023.

Irazoqui Apecechea, M., Melet, A., and Armaroli, C.: Towards a pan-European coastal flood awareness system: Skill of extreme sea-level forecasts from the Copernicus Marine Service, Front. Mar. Sci., 9, https://doi.org/10.3389/fmars.2022.1091844, 2023.

Jafarzadegan, K., Moradkhani, H., Pappenberger, F., Moftakhari, H., Bates, P., Abbaszadeh, P., Marsooli, R., Ferreira, C., Cloke, H. L., Ogden, F., and Duan, Q.: Recent Advances and New Frontiers in Riverine and Coastal Flood Modeling, Reviews of Geophysics, 61, https://doi.org/10.1029/2022RG000788, 2023.

Jansen, E., Martins, D., Causio, S., Maslo, A., Ilicak, M., Cretì, S., Lecci, R., Lima, L., Sözer, A., Trotta, F., and Ciliberti, S. A.: Black Sea Physical Analysis and Forecast (Copernicus Marine Service BLK-PHY, EAS7 system) (Version 1) , Copernicus Marine Service, https://doi.org/https://doi.org/10.48670/mds-00355, 2024.

Kluyver Thomas, Ragan-Kelley Benjamin, Pérez Fernando, Granger Brian, Bussonnier Matthias, Frederic Jonathan, Kelley Kyle, Hamrick Jessica, Grout Jason, Corlay Sylvain, Ivanov Paul, Avila Damián, Abdalla Safia, Willing Carol, and Jupyter Development Team: Jupyter Notebooks – a publishing format for reproducible computational workflows, in: Positioning and Power in Academic Publishing: Players, Agents and Agendas, IOS Press, https://doi.org/10.3233/978-1-61499-649-1-87, 2016.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

Longo, E., Ficchì, A., Verlaan, M., Muis, S., and Castelletti, A.: A Deep Learning Framework for Extreme Storm Surge Modeling Under Future Climate Scenarios, Earths Future, 14, https://doi.org/10.1029/2025EF007072, 2026.

Ma, P., Karagiannis, G., Konomi, B. A., Asher, T. G., Toro, G. R., and Cox, A. T.: Multifidelity Computer Model Emulation with High-Dimensional Output: An Application to Storm Surge, J. R. Stat. Soc. Ser. C Appl. Stat., 71, 861–883, https://doi.org/10.1111/rssc.12558, 2022.

Pérez Gómez, B., Vilibić, I., Šepić, J., Međugorac, I., Ličer, M., Testut, L., Fraboul, C., Marcos, M., Abdellaoui, H., Álvarez Fanjul, E., Barbalić, D., Casas, B., Castaño-Tierno, A., Čupić, S., Drago, A., Fraile, M. A., Galliano, D. A., Gauci, A., Gloginja, B., Martín Guijarro, V., Jeromel, M., Larrad Revuelto, M., Lazar, A., Keskin, I. H., Medvedev, I., Menassri, A., Meslem, M. A., Mihanović, H., Morucci, S., Niculescu, D., Quijano de Benito, J. M., Pascual, J., Palazov, A., Picone, M., Raicich, F., Said, M., Salat, J., Sezen, E., Simav, M., Sylaios, G., Tel, E., Tintoré, J., Zaimi, K., and Zodiatis, G.: Coastal sea level monitoring in the Mediterranean and Black seas, Ocean Science, 18, 997–1053, https://doi.org/10.5194/os-18-997-2022, 2022.

Qin, Y., Su, C., Chu, D., Zhang, J., and Song, J.: A Review of Application of Machine Learning in Storm Surge Problems, J. Mar. Sci. Eng., 11, 1729, https://doi.org/10.3390/jmse11091729, 2023.

Ricker, M., Behrens, A., and Staneva, J.: The operational CMEMS wind wave forecasting system of the Black Sea, Journal of Operational Oceanography, 17, 165–186, https://doi.org/10.1080/1755876X.2024.2364974, 2024.

Roelvink, D., Reniers, A., van Dongeren, A., van Thiel de Vries, J., McCall, R., and Lescinski, J.: Modelling storm impacts on beaches, dunes and barrier islands, Coastal Engineering, 56, 1133–1152, https://doi.org/10.1016/j.coastaleng.2009.08.006, 2009.

Seidenfaden, I. K., Skjerbæk, M. R., Henriksen, H. J., Kjeldsen, K. K., and Sonnenborg, T. O.: Compound Flooding From Storm Surges, Precipitation, Rivers, and Groundwater—Hydrodynamic Modeling in a Coastal Catchment, Water Resour. Res., 61, https://doi.org/10.1029/2024WR037563, 2025.

Soomere, T., Pindsoo, K., Kudryavtseva, N., and Eelsalu, M.: Variability of distributions of wave set-up heights along a shoreline with complicated geometry, Ocean Science, 16, 1047–1065, https://doi.org/10.5194/os-16-1047-2020, 2020.

Staneva, J., Ricker, M., and Behrens, A.: Black Sea Waves Analysis and Forecast (CMEMS BS-Waves, EAS5 system) (Version 1) [Data set]., Copernicus Monitoring Environment Marine Service (CMEMS), https://doi.org/https://doi.org/10.25423/CMCC/BLKSEA_ANALYSISFORECAST_WAV_007_003_EAS5, 2022.

Wang, Z., Leung, M., Mukhopadhyay, S., Sunkara, S. V., Steinschneider, S., Herman, J., Abellera, M., Kucharski, J., Nederhoff, K., and Ruggiero, P.: A hybrid statistical–dynamical framework for compound coastal flooding analysis, Environmental Research Letters, 20, 014005, https://doi.org/10.1088/1748-9326/ad96ce, 2025.

https://doi.org/10.5194/egusphere-2026-3628
Preprint. Discussion started: 15 July 2026

Wilkinson, H., Bates, P., Lucas, C., Quinn, N., Haigh, I. D., Collings, T., and Watson, P.: Physically Coherent Machine Learning for Tropical Cyclone Storm Surge Emulation, https://doi.org/10.5194/egusphere-2026-781, 23 February 2026.

Yang, K., Paramygin, V. A., and Sheng, Y. P.: A Rapid Forecasting and Mapping System of Storm Surge and Coastal Flooding, Weather Forecast., 35, 1663–1681, https://doi.org/10.1175/WAF-D-19-0150.1, 2020.

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., and Behl, M.: Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community, Water Resour. Res., 56, https://doi.org/10.1029/2019WR027038, 2020.