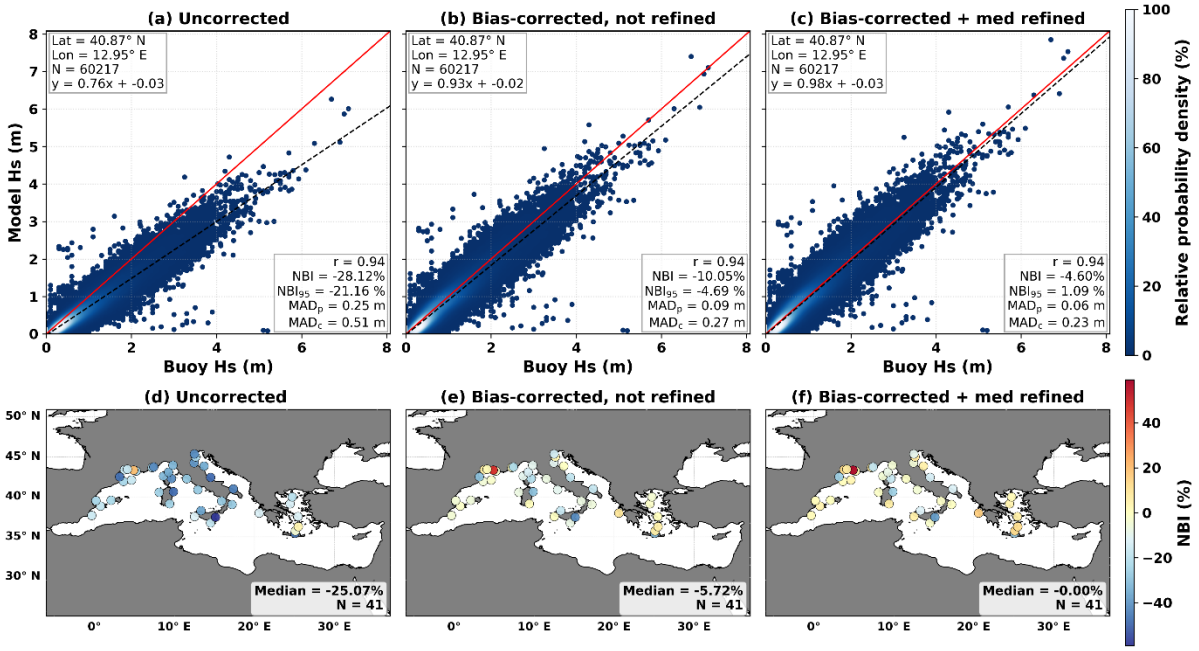

Figure S1: Effect of wave-height bias correction and Mediterranean regional refinement on model performance relative to buoy observations. Upper panels show scatter-density comparisons between modelled and observed significant wave heights (Hs) at the Ponza buoy location (Italy) for (a) the original hindcast, (b) the satellite bias-corrected dataset, and (c) the bias-corrected dataset after application of the Mediterranean refinement described in Sect. 2.2.3. The red line indicates the 1:1 relationship, and the dashed black line denotes the linear regression fit. Lower panels show the spatial distribution of normalized bias (NBI) across the 41 Mediterranean buoy locations for (d) the uncorrected hindcast, (e) the satellite bias-corrected dataset, and (f) the Mediterranean-refined dataset. The uncorrected comparison demonstrates that model biases are relatively small for low and moderate wave heights but increase progressively toward the upper tail of the distribution, providing the motivation for the percentile-based bias-correction framework. The satellite-based correction and subsequent Mediterranean refinement reduce the basin-wide median NBI from −25.1% to approximately zero, further improving agreement with buoy observations.

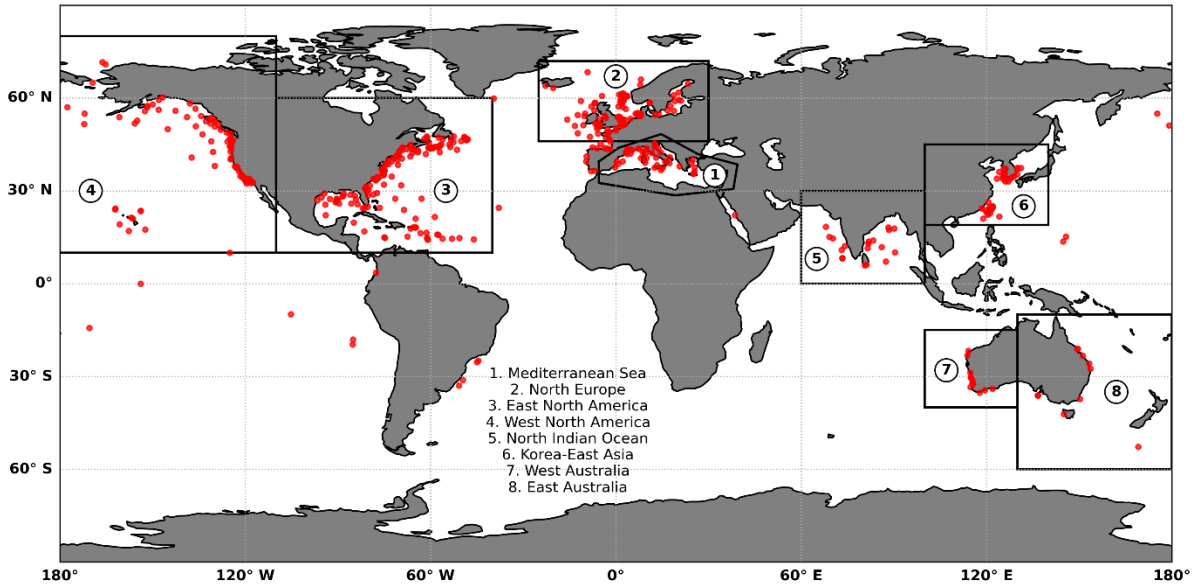

Figure S2: Regional grouping of buoy locations used for the computation of regional performance statistics presented in Table 2. Red circles indicate the 485 buoy locations used for validation of significant wave heights.

Boxes delineate the eight regions considered for the regional analysis: (1) Mediterranean Sea, (2) North Europe, (3) East North America, (4) West North America, (5) North Indian Ocean, (6) Korea–East Asia, (7) West Australia, and (8) East Australia. Regional groupings were defined a priori based on geographic proximity, basin characteristics, and buoy availability. Only regions containing sufficient observations to support robust regional statistics were considered.

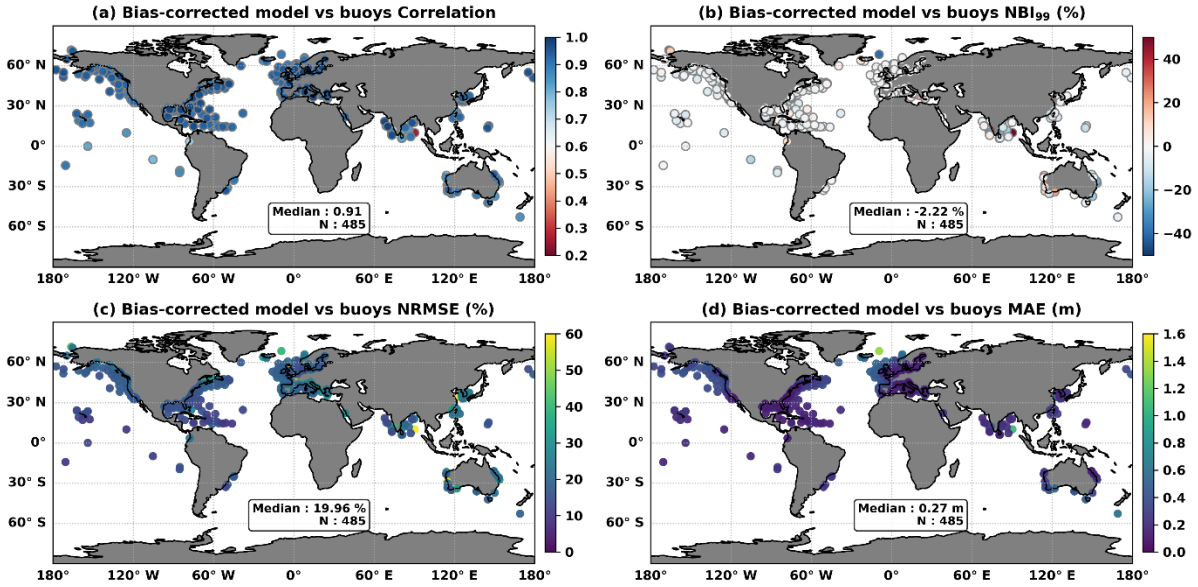

Figure S3: Additional performance metrics for the validation of bias-corrected significant wave heights against 485 buoy locations. Panels show (a) Pearson correlation coefficient, (b) normalized bias beyond the 99th percentile (NBI_{99}), (c) normalized root mean squared error (NRMSE), and (d) mean absolute error (MAE). NBI_{99} is computed following the same formulation as NBI_{95} described in Sect. 2.3.1, but using exceedances above the 99th percentile. The spatial patterns are broadly consistent with those reported in Figure 3, with highest skill observed in the Mediterranean Sea, North European region, and along the eastern coast of North America.

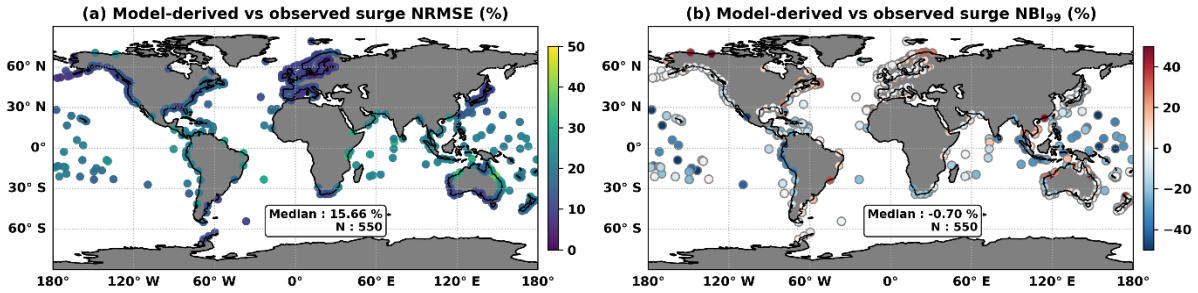

Figure S4: Additional performance metrics for the validation of model-derived storm surges against 550 GESLA3 tide gauges. Panels show (a) normalized root mean squared error (NRMSE) and (b) normalized bias beyond the 99th percentile (NBI_{99}). NBI_{99} is computed analogously to NBI_{95} described in Sect. 2.3.2, but using exceedances above the 99th percentile. The results confirm the spatial patterns identified in Figure 5, with higher skill generally observed in mid- and high-latitude regions and larger uncertainties in tropical environments.