EMI R&D PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2026
Project Title:	Enhancing efficiency and accuracy of coastal storm surge and wave forecasts by means of ensemble clustering and wind forcing correction.
Computer Project Account:	spitbarb
Principal Investigator(s):	Francesco Barbariol, Luigi Cavaleri, Luciana Bertotti, Alvise Benetazzo, Christian Ferrarin, Marco Bajo
Affiliation:	CNR-ISMAR (Venice, Italy)
Name of ECMWF scientist(s) collaborating to the project (if applicable)	………………………………………………………………… …………………………………………………………………
Start date of the project:	01/01/2026
Expected end date:	31/12/2028

Computer resources allocated/used for the current year and the previous one
(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	-	-	5,000,000	1,424,655
Data storage capacity	(Gbytes)	-	-	25,000	16.7

Summary of project objectives (10 lines max)

This project aims to improve coastal hazard forecasting in the Mediterranean and Adriatic Seas by enhancing the accuracy and efficiency of storm surge and wave predictions. It will develop machine-learning-based methods to correct biases in ECMWF IFS surface wind forecasts using satellite observations, with particular focus on coastal and semi-enclosed basins. The corrected winds will be used to assess improvements in surge and wave simulations. In parallel, the project will investigate ensemble reduction techniques based on clustering and machine learning to identify a small set of representative ensemble members. The goal is to preserve forecast uncertainty information while substantially reducing computational costs for high-resolution surge and wave forecasting. The methodology will be tested on different marine storm events using the SHYFEM-WW3 modelling system forced by ECMWF deterministic and ensemble forecasts.

Summary of problems encountered (10 lines max)

We have not encountered problems in using the available HPC resources for the project objectives.

Summary of plans for the continuation of the project (10 lines max)

During the rest of the current year we will continue with exploring ensemble reduction, using both random selection (or pseudo-random, e.g., considering the first members) and clustering algorithms. In the following years we will focus on the other objective of the project, which is the correction of global surface winds for storm surge and wave forecast in coastal and semi-enclosed areas, using statistical downscaling. We will pursue this by means of learning-based methods. A rationale for correction is to let the method learn how the model-minus-observation error in wind speed and direction depends upon the location (distance from the coast, basin area, etc) and wind regime.

List of publications/reports from the project with complete references

A manuscript to be submitted to the Natural Hazard and Earth System Sciences journal is in preparation. It will include the results of this project.

Summary of results

During the first year of the project, we have used the HPC to perform dynamical numerical simulations of waves in the Adriatic Sea. We have thus set-up a WAVEWATCH III model of the Adriatic Sea, and we have forced it using IFS-ECMWF EPS 10-m wind components. The 50-member and the control were run for 7 winter months between 2025 and 2026. Then, the wind ensemble was reduced using random selection and different clustering algorithms, to test the capability of reductions to represent the probabilistic features of the full ensemble. The simulated waves at a buoy station were extracted to build a full and different reduced wave ensembles, which were first intercompared and then compared against observations.

In the following, we show some figures illustrating the model domain (including the buoy position), the control validation against observations and the spread-skill relationship of the ensemble system.

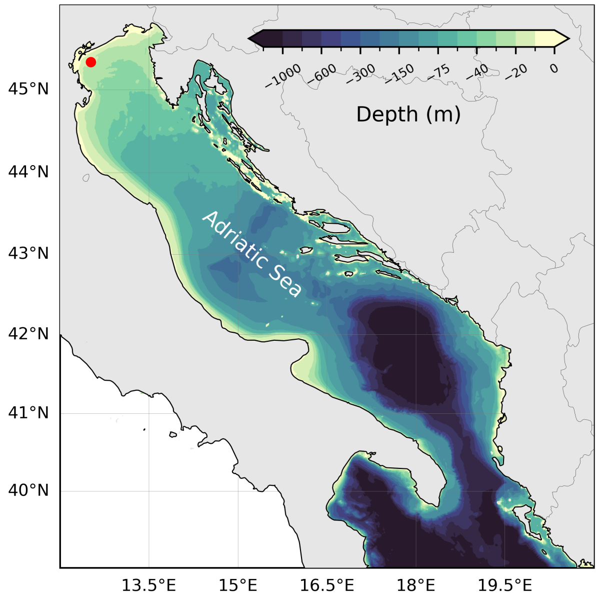

Figure 1 – Adriatic Sea computational domain, focus of the study. The red dot is in the position of the buoy (Italian National Buoy Network, ISPRA) where sea-truth data were collected.

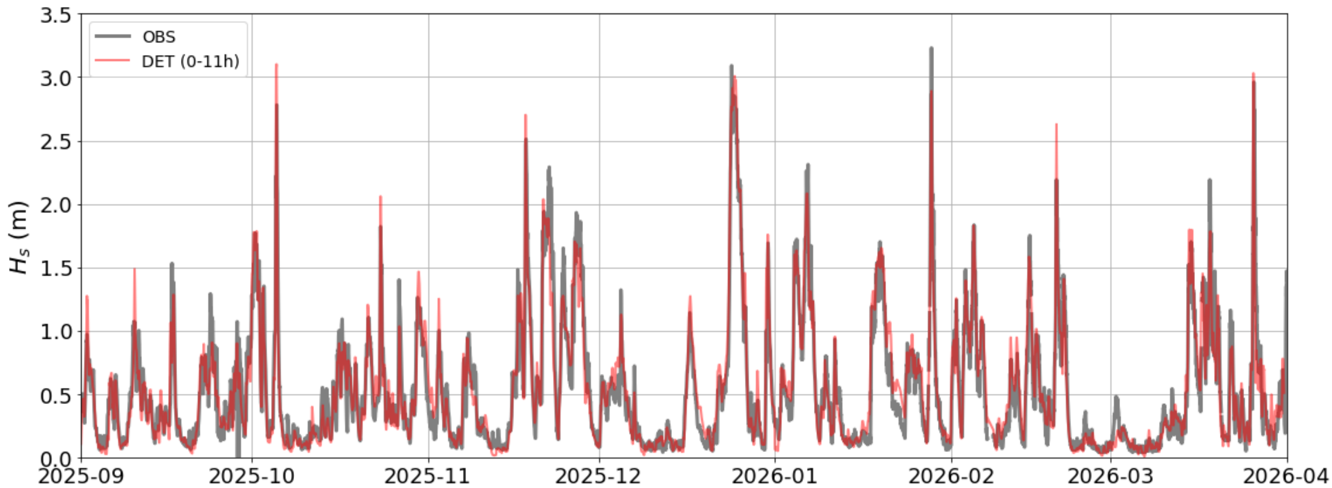

Figure 2 – Deterministic forecast assessment: significant wave height (H_s) observed (OBS; Italian National Buoy Network, ISPRA) and forecasted (DET) at the buoy position in Figure 1. Only the closest forecasts to observations are considered (0-11h leadtime, for each reference time).

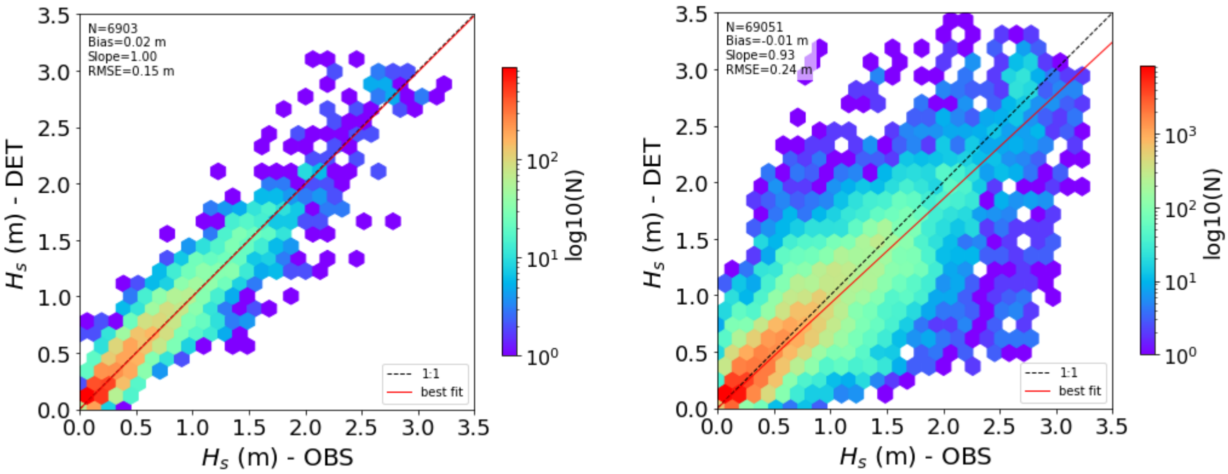

Figure 3 - Deterministic forecast assessment: scatter plots of forecasted (DET) versus observed (OBS) H_s, for the closest collocations (0-11h window, left panel) and for all the collocations (0-120h window, right panel). Assessment metrics are: Bias, slope of the linear fit line, root-mean-square error (RMSE) and number of forecast-observation collocations (N).

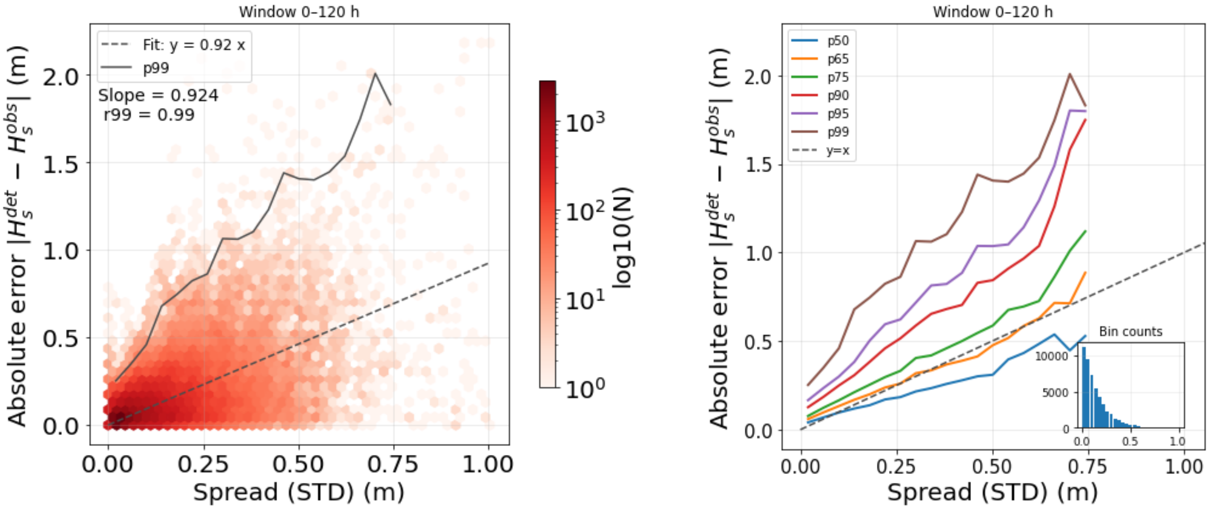

Figure 4 – Ensemble forecast assessment: spread-skill relationship. Scatter plot of the control run (det) absolute error on H_s, versus the ensemble spread (left panel), with reference linear fit (gray dashed line) and 99th percentile of the error for each spread bin (p99, gray solid curve). Slope of linear fit (Slope) and Pearson correlation coefficient of 99th percentile (r) are in the plot. Percentiles of the error versus spread (right panel) with reference 1:1 line (gray dashed line) and spread bin counts (inbox histogram). Spread-skill relationship is assessed over the whole 0-120h leadtime window.