Artificial Intelligence to Monitor our Seas

Funded by the European Union

**PRIN 2022 PNRR Call for Proposals (D.D.1409 of 14/09/2022)**

# AIMS

Artificial Intelligence to Monitor our Seas

*Project number: P2022587FM*

*Starting date: 30th November 2023 – Duration: 24 months*

# Deliverable D3.1
# Report on Training and Validation on Spatial-static data

## DOCUMENT INFORMATION

| | |
|---|---|
| **Deliverable number** | D3.1 |
| **Deliverable title** | Report on Training and Validation on Spatial-static data |
| **Work Package** | WP3 |
| **Deliverable type**[1] | R |
| **Dissemination level**[2] | P |
| **Due date** | 31.07.2025 (Month 20) |
| **Pages** | 18 |
| **Document version**[3] | 2.0 |
| **Lead author(s)** | Leonardo Gambarelli (POLITO), Edoardo Pasta (POLITO) |
| **Contributors** | Giuseppe Giorgi (POLITO) |

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

1 Type: R: Report; D: Dataset
2 Dissemination level: C: Confidential; P: Public
3 First digit: 0: draft; 1: peer review; 2: peer review 3: coordinator approval; 4: final version

# DOCUMENT CHANGE HISTORY

| Version | Date | Author | Description |
|---|---|---|---|
| **DRAFT** | | | |
| 0.1 | 03.07.2025 | Leonardo Gambarelli (POLITO) | Creation |
| **FIRST PEER REVIEW** | | | |
| 1.0 | 17.07.2025 | Edoardo Pasta (POLITO) | Consolidation of input from reviewers |
| **FINAL VERSION** | | | |
| 2.0 | 27.07.2025 | Giuseppe Giorgi (POLITO) | Final and formal review, version ready for submission |

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

## SHORT ABSTRACT FOR DISSEMINATION PURPOSES

**Abstract**

This deliverable extends the gate-clearance procedure initiated in D 2.2 and D2.5, by benchmarking the five spatial interpolation/regression algorithms that passed the first and second gates – TPS, RBF, GPR, RF and ANN – on both simulated and buoy-based observations for December 2022. Quantitative performance metrics (Mean Absolute Error) for Significant Wave Height and Peak Period are used to evaluate, compare and rank the algorithms.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca).

# TABLE OF CONTENTS



AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

# LIST OF PARTNERS

| Nº | Logo | Name | Short Name | City |
|---|---|---|---|---|
| 1 | Politecnico di Torino 1859 | Politecnico di Torino | POLITO | Torino |
| 2 | ROMA TRE UNIVERSITÀ DEGLI STUDI | Universitá degli studi di Roma Tre | ROMA3 | Roma |
| 3 | Italian National Research Council | Consiglio Nationale delle Ricerche | CNR | Firenze |

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca).

# ABBREVIATIONS

| Acronym | Description |
|---|---|
| **ANN** | Artificial Neural Network |
| **GPR** | Gaussian Process Regression |
| **MAE** | Mean Absolute Error |
| **RBF** | Radial Basis Function |
| **RF** | Random Forest |
| **TPS** | Thin Plate Spline |

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

## LIST OF FIGURES



## LIST OF TABLES

# EXECUTIVE SUMMARY

This deliverable deepens the "gate clearance" analysis begun in D 2.2 and D2.5 by applying a shared benchmark to the five algorithms that survived the sequential screenings.

At first the discrepancy between satellite and buoy data is assessed by comparing their respective collocated measurements.

Then, using the hindcast archive outlined in D 1.1 (2017-2022) together with the seven operational wave buoys deployed under D1.3, the study concentrates on the single month common to both sources, December 2022. Each of the 248 three-hourly snapshots were spatially subsampled, with 300 of 754 model grid points and five buoys forming the training set, while the remaining two buoys served exclusively for validation. Performance was assessed by the Mean Absolute Error at these two holdout buoys for Significant Wave Height (Hs) and Peak Period (Tp).

The results reveal a clear hierarchy. For Hs the TPS method records the lowest mean absolute error at 0.065 m, followed by RF, GPR, and RBF whereas ANN lags furthest behind at 0.156 m. For Tp, RF leads with a mean error of 0.790 s, TPS comes second, GPR again third, then ANN and RBF posts the poorest performance at 1.366 s.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca).

# 1. Introduction

This deliverable builds on the first-gate and second-gate clearance reports D2.2 [1] and D2.5 [2] by subjecting the algorithms that advanced past that gate to further testing and head-to-head comparison.

The candidate methods are:

- Thin-Plate Spline (TPS)
- Radial Basis Functions (RBF)
- Gaussian Process Regression (GPR)
- Random Forest (RF)
- Artificial Neural Networks (ANN)

## *1.1 Experiment description*

We employ the same numerical data set described in D1.1 [3]—already used in D2.2—supplemented by in-situ measurements from deliverable D1.3 [4]. Each algorithm is tasked with reconstructing Significant Wave Height (Hs) and Peak Period (Tp) from a limited, spatially sparse set of known points.

As in D2.2, reconstructions are performed purely in space; every snapshot is treated independently, with no temporal autocorrelation assumed.

Because the D1.1 numerical archive spans 2017–2022 whereas the nine buoys from D1.3 were deployed only in November 2022, we restrict both data sets to their common period—December 2022. Subsampling every three hours yields 248 snapshots for evaluation.

For each snapshot, 300 of the 754 offshore grid points in the numerical simulation are randomly assigned to the training set.

Of the nine deployed buoys, two failed before December 2022; thus only seven provided usable data. Five of these seven buoys supply additional training observations, while the remaining two serve exclusively for validation. Figure 1 depicts the buoy layout, highlighting the units allocated to training versus testing.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca).

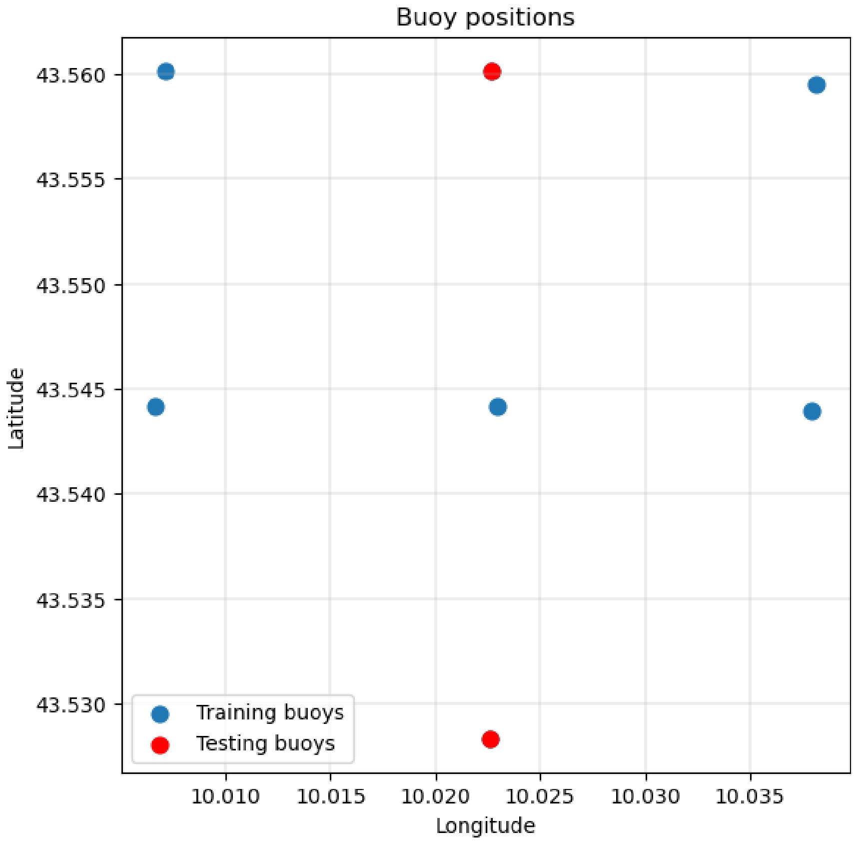


**Figure 1 - Disposition and partitioning of the measuring buoys**

Model performance was quantified, for every experimental setup, by the Mean Absolute Error (MAE) computed at the two hold-out buoys.

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|z_i - \hat{z}_i|$$

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca).

## 2. Comparison between satellites and buoys measurements

Before benchmarking reconstruction algorithms, we verified the trustworthiness of the in-situ buoy measurements used in this deliverable by cross-checking them against near-coincident satellite Significant Wave Height (SWH) observations. The aim is to ensure that buoy Hs is consistent with an independent observing system within reasonable co-location windows, thereby supporting its use as ground truth in the subsequent analyses. This validation was performed considering always December 2022. The satellite data used were the ones from the deliverable D1.2 [5].

A buoy measurement and a satellite measurement have been considered collocated if their latitude and longitude differences were less than 0.2° and if they time difference was less than 3 hours. Below are reported 2 scatterplots, the first one plotting for each satellite measurement all the valid collocated buoys measurement, using different colours for different buoys, and the second one plotting each satellite measurement against the average of the valid collocated buoys measurements.

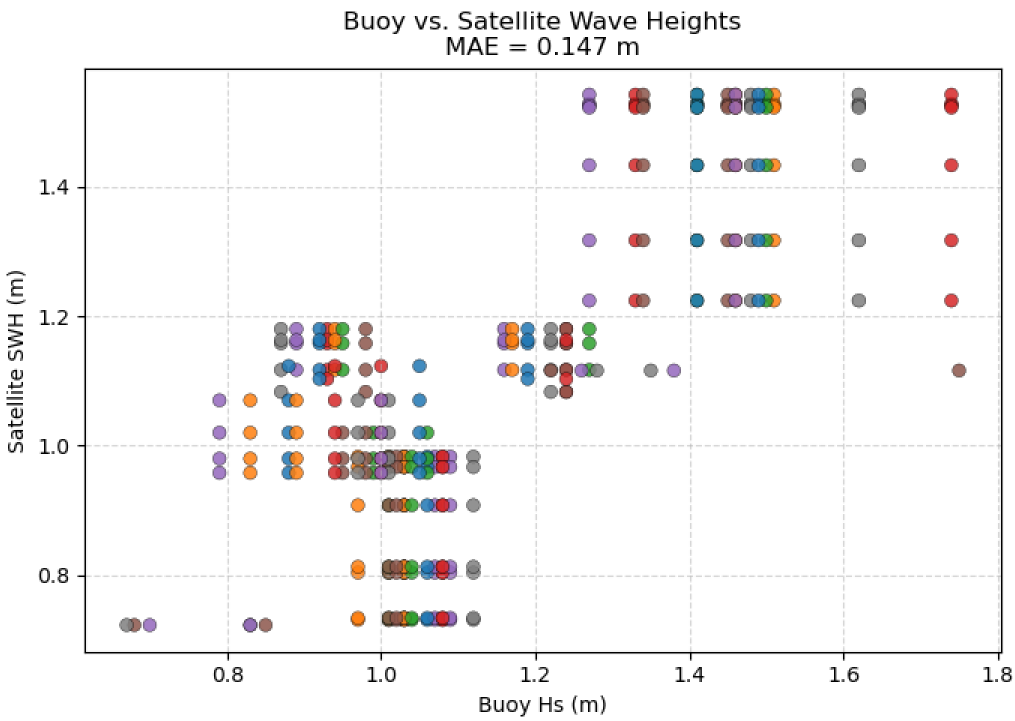


**Figure 2 - Satellite vs buoy measurements scatterplot.**

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

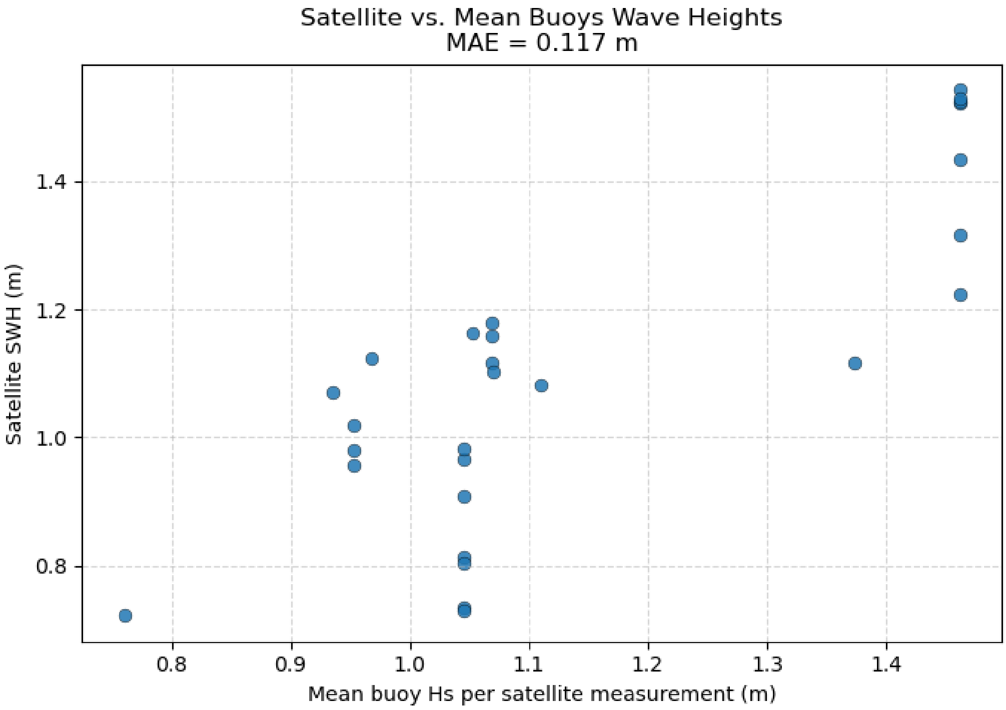


**Figure 3 - Satellite vs average buoys measurements scatterplot.**

The two scatter plots indicate a good agreement between satellite SWH and in-situ Hs. Using all buoy–satellite pairs, the mean absolute error (MAE) is 0.147 m. When we first average all valid co-located buoy measurements per satellite observation, the MAE drops to 0.117 m—an improvement of roughly 20%. This reduction is consistent with averaging mitigating small-scale spatial/temporal mismatch and instrument noise. No clear systematic bias is evident over the observed range, and the remaining spread is compatible with footprint differences and short-term sea-state variability within the ±3 h co-location window.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

## 3. Results

After having run the algorithms for all the available snapshots and for both parameters, it was recorded the error obtained on the test buoys and it was possible to assign a MAE for each of them over the whole month of December 2022.

The main results are shown in the following figures and summary table.

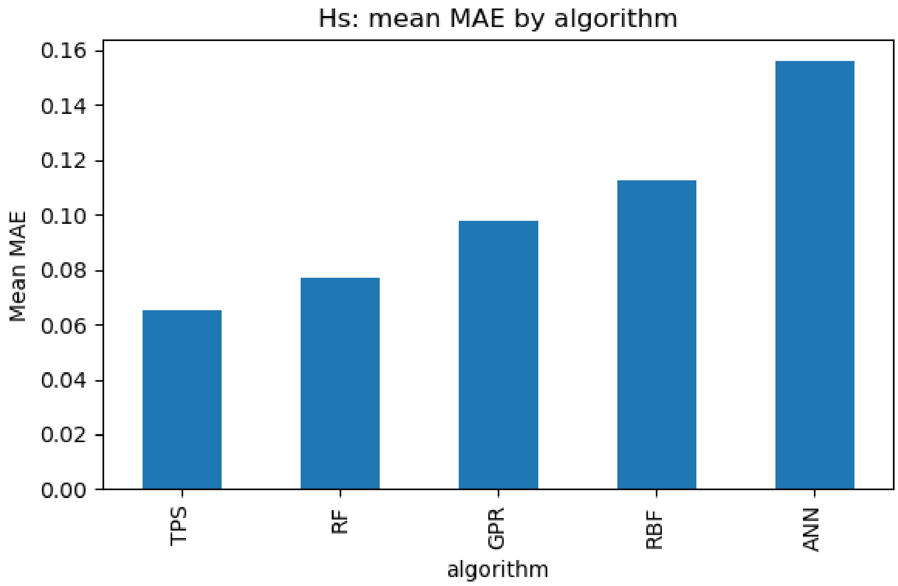


**Figure 4 - Mean MAE achieved by the algorithms when reconstructing Hs.**

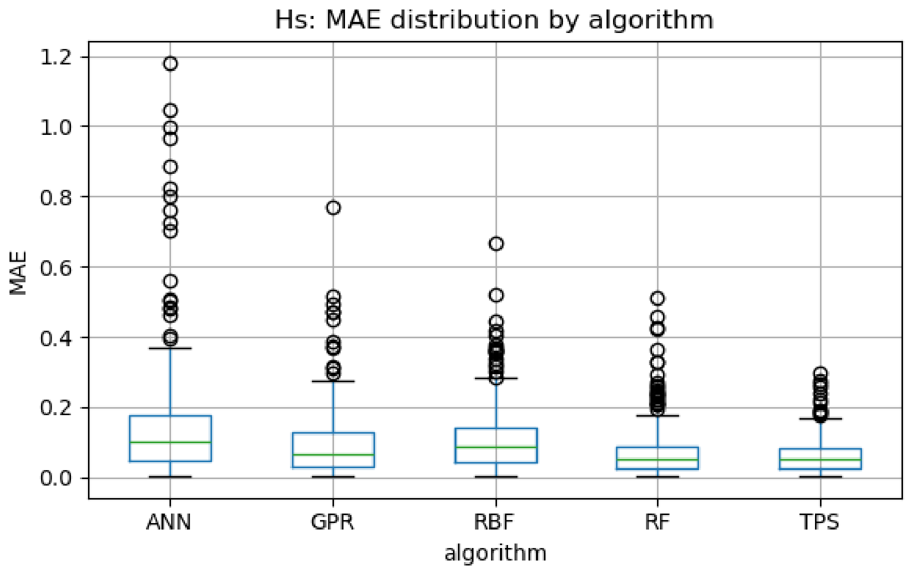


**Figure 5 - Boxplot of the MAE achieved by the algorithms when reconstructing Hs.**

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

As can be seen from Figure 2 and Figure 3, TPS and RF achieved the lowest reconstruction errors for Hs, with instead the RBF and particularly the ANN achieving the highest ones.

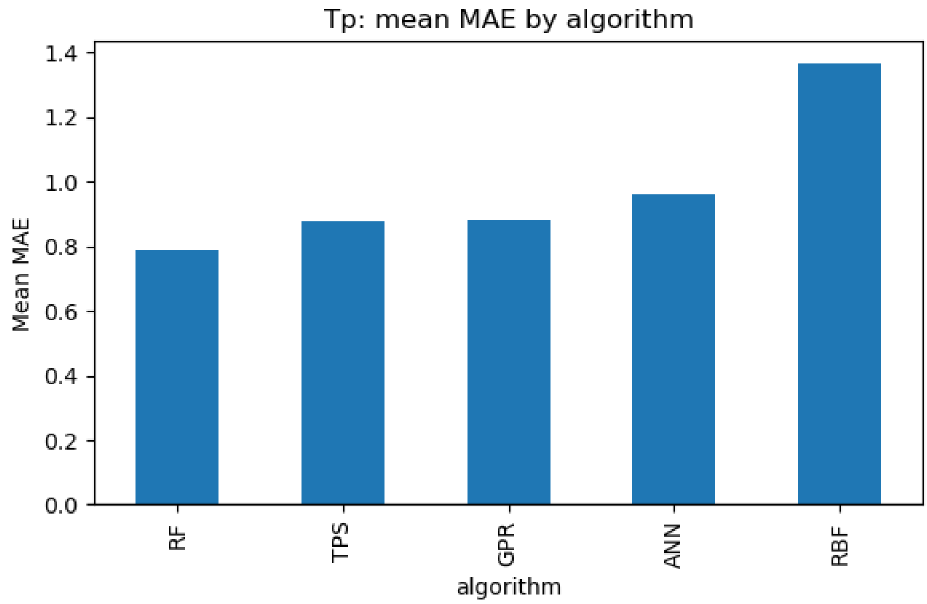


**Figure 6 - Mean MAE achieved by the algorithms when reconstructing Tp.**

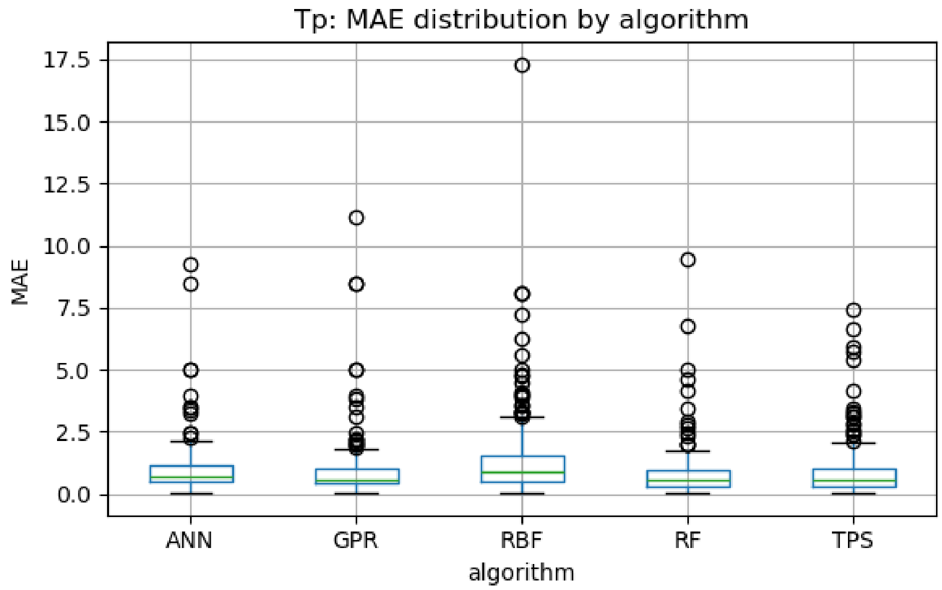


**Figure 7 - Boxplot of the MAE achieved by the algorithms when reconstructing Tp.**

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

Figure 4 and Figure 5 confirm the results obtained for Hs also for Tp, with again the RF and the TPS being the best performing algorithms while the ANN and the RBF result to be the worst performing ones. The GPR instead achieved middle performances for both parameters. Anyway it can be noted that both the best and the worst algorithm are different for the 2 parameters: the TPS achieves the best results for Hs, while the RF for Tp; instead the ANN obtains the worst performance for Hs, while the RBF for Tp.

**Table 1 - Average MAE achieved by the reconstruction algorithms.**

| Algorithm | Hs (m) | Tp(s) |
|---|---|---|
| TPS | 0.0649 m | 0.8788 s |
| RBF | 0.1125 m | 1.3664 s |
| GPR | 0.0977 m | 0.8821 s |
| RF | 0.0768 m | 0.7904 s |
| ANN | 0.1558 m | 0.9594 s |

Part of the parameter-dependent spread in model skill stems from the intrinsic character of the two fields. As already observed in D2.2, Tp in the numerical hindcast is almost quantised—jumping between a limited set of discrete values—so its spatial surface is peppered with sharp discontinuities. In contrast, Hs varies much more smoothly. Radial Basis Functions, which rely on smoothly blending local kernels, struggle to track these abrupt Tp jumps and therefore suffer a marked loss of accuracy. For Hs, where continuity prevails, RBF's underlying assumptions are far less violated, narrowing the performance gap.

In addition, certain snapshots—most noticeably in the Tp series—show error spikes well above the typical range. These surges point to brief intervals dominated by strong non-linear dynamics or extreme events (e.g., rapidly intensifying storm systems) that lie outside the algorithms' representational envelope.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca).

## 4. Conclusions

Benchmarking against December 2022 buoy observations consolidates the ranking tentatively established in D2.2 and D2.5.

Overall, the evidence justifies advancing TPS and RF to the final selection phase, keeping GPR as a pragmatic back-up option, and dropping RBF and ANN from further development.

Thin-Plate Spline looks like the most dependable choice for reconstructing Hs thanks to its capacity to capture smooth spatial variability with minimal error, whereas RF delivers the best Peak Period estimates by accommodating sharp, nonlinear changes across the domain. GPR holds a respectable middle ground and retains value where uncertainty quantification is essential. Artificial Neural Networks and Radial Basis Functions persistently perform worst, and the extra tuning effort they require is no longer defensible.

The project will therefore carry TPS and RF forward into the operational integration stage, with GPR maintained as a reserve.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.

## REFERENCES

[1] Deliverable D2.2 Report on the first gate clearance of AI algorithms.
[2] Deliverable D2.5 Report on the second gate clearance of AI algorithms.
[3] Deliverable D1.1 Report and dataset from wave models.
[4] Deliverable D1.3 Report and dataset from moored gridded wave buoys.
[5] Deliverable D1.2 Report and dataset from satellites.

AIMS: Artificial Intelligence to Monitor our Seas is funded by the European Union - NextGeneration EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This document reflects only the authors' view, and the Commission and Ministry cannot be considered responsible for any use that may be made of the information it contains.