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 D2.4

Report and algorithms to infer wave period from altimeters

DOCUMENT INFORMATION

Deliverable number	D2.4
Deliverable title	Report and algorithms to infer wave period from altimeters
Work Package	WP2
Deliverable type[1]	Report
Dissemination level[2]	Public
Due date	30.01.2025 (Month 14)
Pages	
Document version[3]	2.0
Lead author(s)	Claudia Cecioni, ROMA3
Contributors	Edoardo Pasta, POLITO 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	03/12/2024	Claudia Cecioni, ROMA3	Creation
0.2	30/12/2024	Edoardo Pasta, POLITO	Writing organization
FIRST PEER REVIEW			
1.1	15/01/2025	Claudia Cecioni, ROMA3	Review
FINAL REVIEW			
2.0	30/01/2025	Giuseppe Giorgi, POLITO	Final Review

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

The Report and algorithms to infer wave period from altimeters serve as a foundational document for the AIMS (Artificial Intelligence for Marine Surveillance) project. This report describes in details the method and the algorithm to infer the wave period from the altimeters data, i.e. wave height and wind speed. The wave period inference is an important tool to get the full sea state characterization.

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.

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.

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	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). 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.

ABBREVIATIONS

Acronym	Description
AI	Artificial Intelligence
SWAN	Simulating WAves Nearshore
WW III	Wave Watch III
H_{m0}	Significant wave height
T_p	Peak wave period

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

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.

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. INTRODUCTION

1.1 Deliverable 2.4 inside AIMS project

The Deliverable D2.4 Report and algorithms to infer wave period from altimeters aims at describing and providing the methodology, i.e. the algorithm, to infer wave period from altimeters data. Altimeters installed on satellites are not able to directly measure the wave period, which is a crucial information for the design of offshore systems and understanding climate variations in the long term. AIMS develops a semi-empirical technique to indirectly estimate the wave period, based on direct measurements of wave height and wind speed. With the assistance of AI approach, an iterative learning method will be used to progressively refine the wave period estimation.

This deliverable concerns the research activities carried out within the WP2 that has the milestone (M2) of developing AI algorithms based on available satellite and numerical data. More in particular deliverable D2.4 is related to the task 2.2 “AI for inference of the wave period” of the WP2.

To collocated the present deliverable (D2.4) in the AIMS project framework a short resume of the main objectives is presented in this section.

The major objective of AIMS is to develop and validate a new framework for metocean surveying, based on novel Artificial Intelligence (AI) algorithms for gap-filling of remote monitoring via satellites. The main objective is shown at the bottom-right corner of the methodology Figure 1. Such a framework, achieved for metocean variables by the end of the project, has the potential to be applied to other fields beyond the project: flexibility and interoperability considerations are included to facilitate extrapolation to other domains.

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.

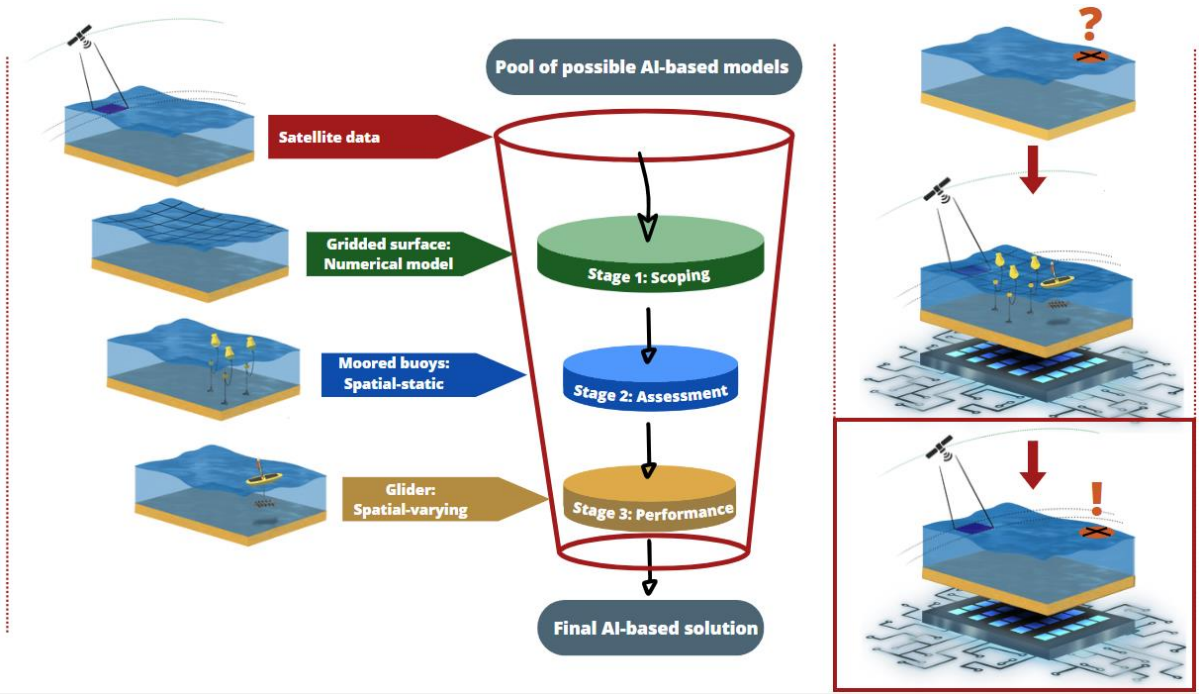

Figure 1. Schematic representation of AIMS methodology

To ensure a smooth and successful progress, and to provide transparent monitoring tools, the main objective is broken into the following SMART results, i.e. Specific, Measurable, Achievable, Realistic, and Time-bound:

- Building the underlying dataset for AI training and validation; it will be a unique composition of heterogeneous sources:
    - 2 refined numerical downscaling datasets of wave data: SMART, because already consolidated with ROMA3 and CNR, respectively in SWAN (Simulating WAves Nearshore) and WW3 (WAVEWATCH III), and the latter can be re-run to meet AI requirements.
    - Satellite data: SMART because data is available on Copernicus Open Access Hub, selecting altimeters from Sentinel-3A and -3B; other constellations will be considered to further populate this dataset (e.g., SARAL, Jason, Sentinel 6).
    - In-situ measurement, using multiple moored wave buoys installed in proximity in a gridded layout, to provide distributed spatial-static data. This is a unique type of dataset, since buoys are usually deployed singularly or in pairs. The objective is to obtain months of data.
    - In-situ measurement, using a glider, to provide spatial-varying data.

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.

- Performance of the AI algorithms, with a stage-gate approach. The final SMART objectives are:
    - Speedup surveying: **>** +40%. Estimated by comparing conventional full offshore surveys with the reduced time required for AI training.
    - Cost reduction: **>** -60%. Based on the reduction of surveying time and on common cost metrics for offshore operations.
    - Gap-filling accuracy: **>** 90% within a 10km distance from the satellite's direct measurements. Measured by comparing AI outputs with experimental validation subset. Note that AI algorithms provide information in any arbitrary point in space, while the accuracy is expected to decay with the distance from the direct measuring point (satellite orbit swath).
    - Time resolution: < 3h. Related to the typical period of stationarity of sea states, and represents at least a 4-fold increase in resolution with respect to typical revisit times from satellite orbits.
    - Wave period inferring: **>** 85% accuracy, by comparison with experimental datasets.

On the one hand, AIMS objectives are ambitious, since they lead to a substantial progress with respect to the state-of-the-art, impactful for science and technology; on the other hand, they are also achievable, since they are based on previous experience of the partners in offshore experimental campaigns, numerical models and AI solutions.

The originality of AIMS objectives is to inherently combine different sources of data to get the best value out of satellites, both in time and space: while usually gaps are thought only in time, AIMS rethinks gaps in both space and time, adding a new dimension and a realm of new possibilities, both within the ocean monitoring field and beyond. Moreover, AIMS leads to a more informative spatially-distributed information, rather than point-wise. AIMS new perspective opens a new way to cut costs down and accelerate monitoring campaigns, since it reduces the number of required expensive in-situ instruments and the execution time of the survey.

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.

1.2 Sea-state parameter

In general, sea states are identified by wave parameters, as wave height, wave period and wave direction. Irregular sea state needs synthetical and statistical values of these three parameters. The most common statistical parameters used are the significant wave height H_{m0}, the peak wave period T_p and the mean direction D_m. These statistical wave parameters can be derived from the water surface elevation time series, which usually are few tens of minutes long recorded with a sampling frequency of around 1 or 2 Hz. Both a time-domain or a frequency domain can be carried out in order to find the statistical parameters. From a frequency domain analysis, the significant wave height is defined as

$$H_{m0} = 4\sqrt{m_0}$$

Where m_0 is the zero-order integral of the energy density spectrum. The peak wave period T_p instead is the inverse of the frequency where the wave energy density is maximum. From a time domain analysis, the significant wave height is defined as the mean of the highest one-third apparent waves, and it is named as H_s, which is in general equivalent to H_{m0}.

Wind generated waves show higher wave height and longer wave period as faster is the wind velocity. However, wave height and period are also proportional to the fetch, the length of the sea area, where the wind blows parallel to the direction of the wind.

Waves generated locally, i.e. inside the fetch's area, by wind are known as *sea waves*. These waves propagate more or less in the wind direction and are formed by a crossing pattern of few wave trains propagating at a small angle around from the mean wind direction. Sea waves consist of many different wave heights and periods, they are therefore irregular wave pattern. These waves will travel also beyond the area of generation, for large bodies of water. While the waves travel such long distances, the energy is transferred from higher frequencies to lower ones, resulting in longer wave periods, smaller wave height and more pronounced wave grouping. These wave pattern, far from the generation area are called *swell waves*.

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.

The period of wind-generated waves, is shorter than 30 s. When waves are being generated by the local wind, i.e. *sea waves*, they are irregular and short-crested, with a wave period from few seconds to 10-12 seconds. When they leave the generation area, they take on a regular and long-crested appearance and are called *swell*. While travelling long distances away from the generation area, the sea waves energy is transferred from higher frequencies to lower ones, resulting in longer wave periods of the order of 15-20 seconds, smaller wave height and more pronounced wave grouping.

A mixed sea state of wind sea (short, irregular, locally generated waves) and swell (long, smooth waves, generated in a distant storm) may have the same significant wave height and period as a slightly higher wind sea without swell. To distinguish such conditions, more parameters are needed, for instance, a significant wave height and period for wind sea and swell separately. This is sometimes done and it may be adequate in some cases, but any small number of parameters would not, in general, completely characterise the wave conditions. For a complete description (in a statistical sense), the spectral technique is required. It is based on the notion that the random motion of the sea surface can be treated as the summation of a large number of harmonic wave components. The wave frequency spectrum describes the wave energy density distribution over the frequency, as shown in Figure 2.

Swell is generally of a much lower frequency than a young wind sea, so in this case the two wave systems are well separated, both in frequency and in direction. Moreover, swell is rather regular and long-crested, so its spectrum is narrow.

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.

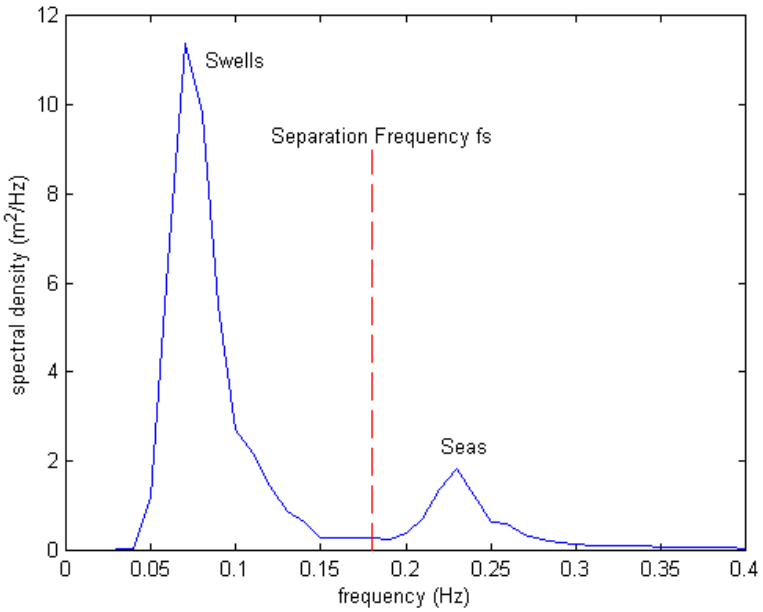

Figure 2: A typical wave spectrum with dominant swell waves, showing the separation frequency and the distribution of swell and wind-seas energy with frequency.

1.3 Wave period inference

Since altimeters wave data from satellite observation are limited to the significant wave height, a methodology is described in the next sections, to infer the wave period, as useful information to characterize the sea state.

From section 1.2, it results clear that, if free surface time series are available, (from in situ measurements) a complete characterization of the sea state can be obtained performing both a time or frequency domain analysis. Moreover, using wind data and defining the fetches distribution over all the possible wind directions, wave data can be hindcasted. Hindcasting wave data (H_{m0} and T_p) from wind data (wind velocity and direction) can be done using different methodologies: i) spectral hindcasting models, such as WAM, SWAN and WW III; ii) parametric models, which assume that the wave spectrum behaves as a known function (JONSWAP or Pierson-Moskowitz spectrum); iii) statistical models or iv) empirical models.

The methodology here proposed aims at inferring the wave period information based on the significant wave height and wind velocity information measured 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.

2. CORRELATION AMONG WAVE PARAMETERS

2.1 Introduction

Satellites or altimeters wave data, providing wave data continuously in time, are a useful source that is considered by the AIMS project.

However, altimeters wave data are limited to the wave height, in terms of significant wave height, and no further information are given for the wave period and the wave direction. Therefore, one aim of the AIMS project is that of inferring the wave period information, based on a detailed analysis of the available wave databases. These wave datasets consist on two refined numerical downscaling models, SWAN and WVAEWATCH III, and on localized measurements from in-situ measurements. As reported in Deliverable 1.1, the wave parameters computed by the numerical models are relative to an area of interest in the North Tyrrhenian sea. Numerical models allow to perform a large wave database distributed in space and discretised in time, however in order to apply these refined downscale models a big computational effort is needed. The accuracy of the numerical wave database can be checked and eventually calibrated with the more accurate in situ direct measurements. Therefore, machine learning techniques are trained with numerical wave dataset available to infer wave period from altimeters data.

2.2 Wave parameters correlations

The proposed methodology (described later in Section 3) is based on wave dataset analysis here described. The here reported analysis refers to wave data from ERA5 point with coordinates 43.5 °N and 9.5 °E. In particular wave data from 01/01/2018 to 31/12/2022 are analysed.

In Figure 5 a scatter diagram highlights the correlation between the peak wave period, T_p, and the significant wave height, H_{m0}. It is known, and it can be noted from the data in Figure 5, that peak wave periods increase with increasing significant wave heights. Among all these data, some sea states are relative to swell sea states, which have small wave heights, (smaller than 2.5 m for this case), and long wave periods, that reach the maximum peak

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.

period values. Other sea states are relative to sea states with higher wave heights and wave periods, which increase with a power correlation with H_{m0}.

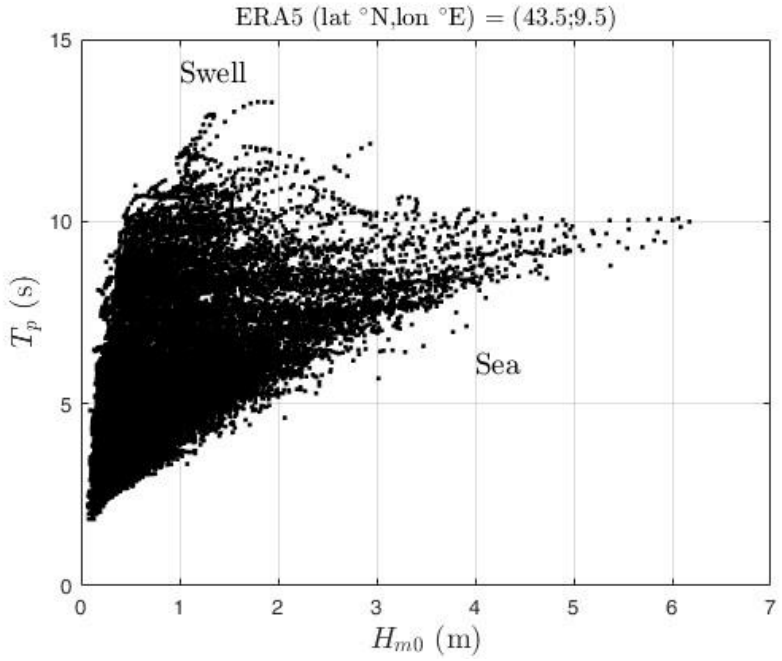

Figure 3. Scatter diagram of peak wave period vs significant wave height. Total of 43824 wave data from ERA5 database at node 43.5°N, 9.5°E.

In order to distinguish the swell from the sea-states, a spectrum analysis has to be performed, allowing to individuate the separation frequency, as reported in Figure 2. Moreover, wave steepness analysis allows to separate swell from sea waves, since swell waves present steepness lower than that of sea waves. Indeed, the wave steepness, ε, is defined as the ratio between wave height and wave length, the latter being proportional to the squared wave period. For the wave data of Figure 3, the steepness value that separate the two class of waves has been identified as $\varepsilon = 0.015$. Figure 4 highlights in black the swell waves ($\varepsilon < 0.015$) and in red the sea waves ($\varepsilon > 0.015$).

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.

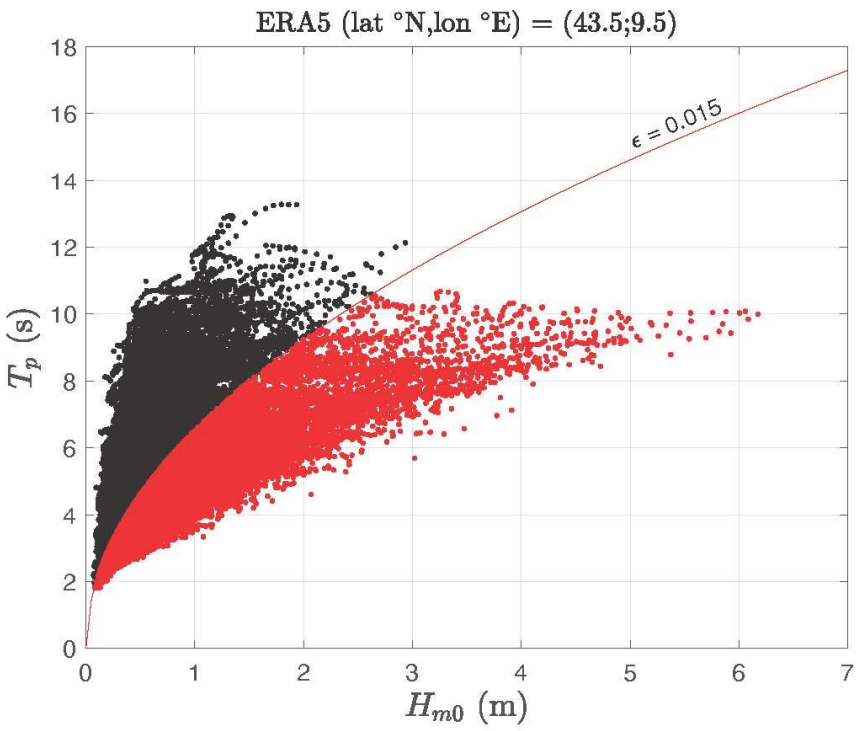

Figure 4 Scatter diagram of peak wave period vs significant wave height. Red dots are relative to sea state with a wave steepness higher than 0.015, while black dots refer to wave steepness lower than 0.015. Total of 43824 wave data from ERA5 database at node 43.5°N, 9.5°E.

The Italian wave Atlas [1], reports the results of activities based on the analysis of wave parameters measured by the National Wave Recording Network. The analysis concerns more specifically to the extreme sea states. Among other statistical analysis of recorded data, the Italian Wave Atlas reports also the wave parameters correlations for *sea waves*, stating that:

$$T_p = b\,(H_{m0})^c,$$

with the coefficient b and c fitted by wave data. In particular for La Spezia wave buoy recording the values of these coefficients are:

b_m = 6.604; $b_{84\%}$ = 9.559; $b_{16\%}$ = 3.641

c_m = 0.255; $c_{84\%}$ = 0.090; $c_{16\%}$ = 0.551

The curve obtained with b_m and c_m, is reported in Figure 5, with the continuous red line, while the dashed red lines indicates the curves obtained with $b_{84\%}$ and $c_{84\%}$ and $b_{16\%}$ and $c_{16\%}$.

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.

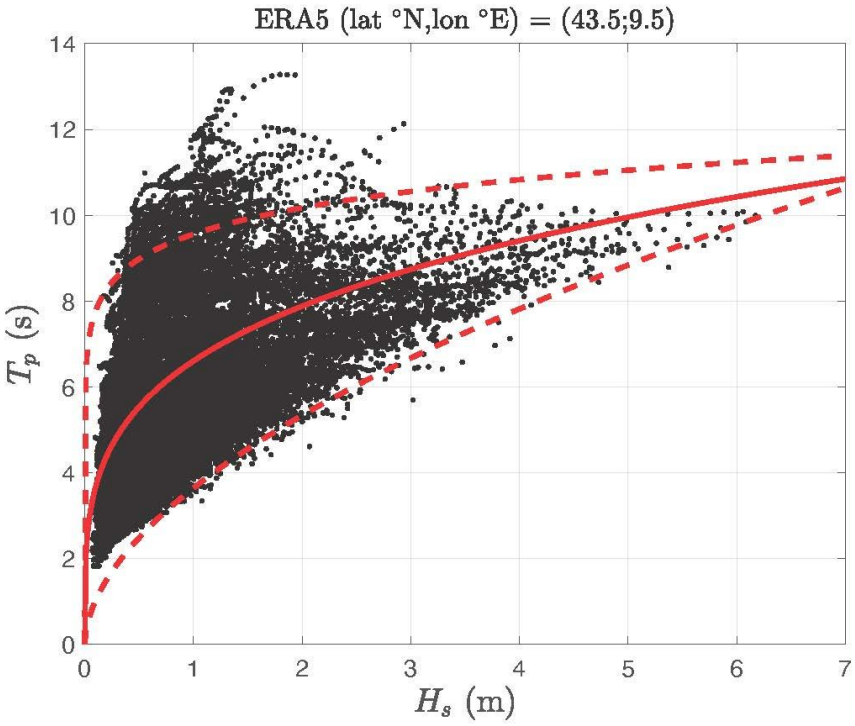

Figure 5. Scatter diagram of peak wave period vs significant wave height. Total of 43824 wave data from ERA5 database at node 43.5°N, 9.5°E.

Correlations between H_{m0} and T_p are known for sea states, different correlations can be found for swell sea state (even if data dispersion is higher). Therefore, it is not possible to infer the wave period exclusively from the wave height, unless it would be possible to know whether swell or sea states are expected.

Since from altimeters data, also the wind velocity is known, similar scatter plot has been obtained in Figure 6, considering the peak wave periods and the wind velocities. At higher wind velocities correspond higher peak periods; however, as for the correlation with wave heights, the longer wave periods are induced by smaller values of the wind velocities. Indeed, it is expected that faster wind velocities generate sea waves, on the contrary, swell waves propagate when wind stopped to blow or is blowing far away.

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.

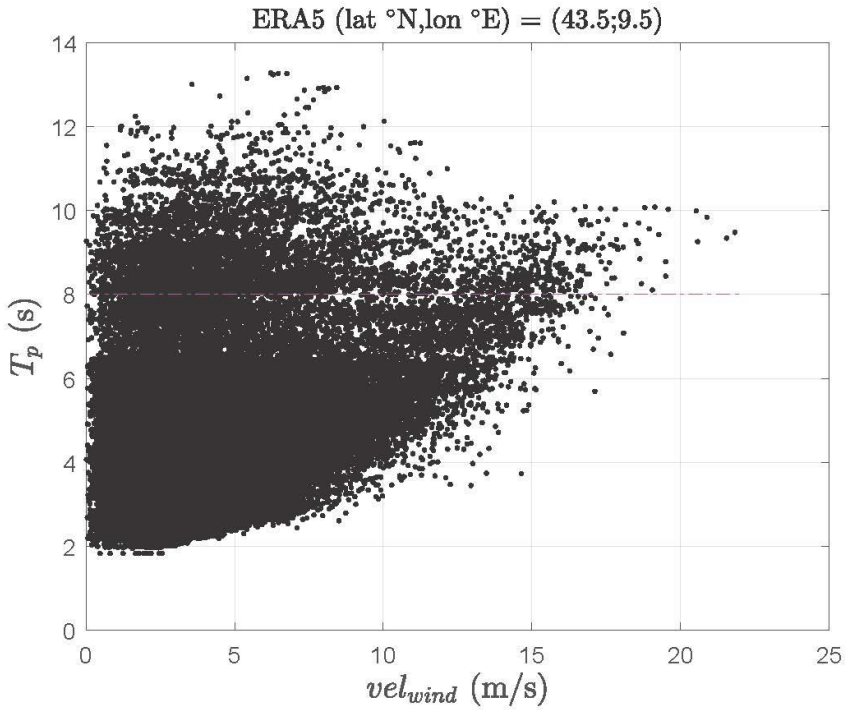

Figure 6. Scatter diagram of peak wave periods versus wind velocities. Total of 43824 wave data from ERA5 database at node 43.5°N, 9.5°E.

Three-dimensional scatter plots, as that in Figure 7, allow to identify the wave periods for a given couple of wave height and wind velocity values.

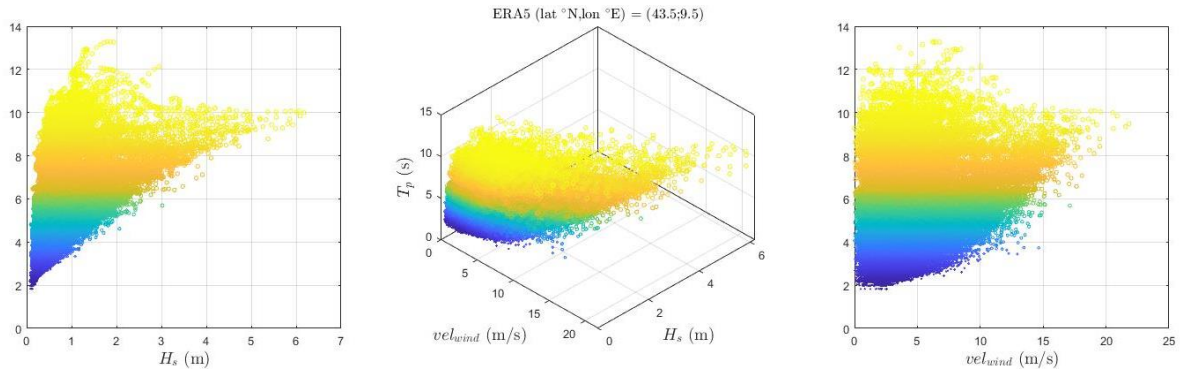

Figure 7. Scatter 3D diagram of peak wave period versus the significant wave height and wind velocity. The coloured markers indicate different scale of peak wave period values.

However, data are quite scattered and do not allow to infer almost univocally the wave period for given wind velocity and wave height values, i.e. it exists large variability for low values of H_s and vel_{wind}, where both swell and sea states can occur.

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.

A further scatter diagram of H_s and wind velocity is presented in Figure 8, highlighting a better correlation among these two parameters, as expected. Significant wave height and wind velocity are the two altimeters data available. Therefore, in order to include the wave period values, a table such the one proposed in Figure 9 can be performed. The significant wave height data have been divided into classes of 0.5 m, while the wind velocity data have been divided into classes of 1 m/s. Among all the wave data inside each class, the mean peak period has been calculated and it is reported in the table. A colored scale distinguishes the longer mean peak periods (green coloured) from the shorter ones (red coloured).

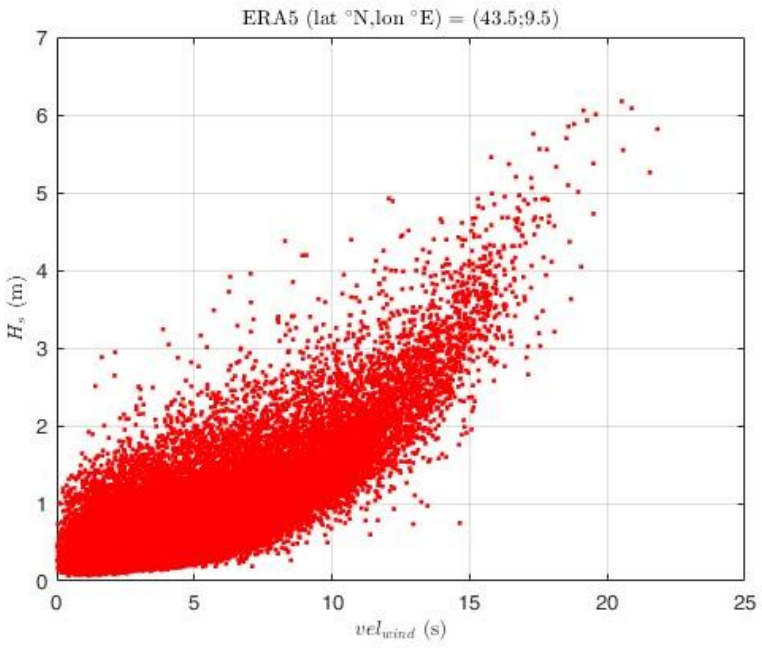

Figure 8. Scatter diagram of significant wave height and wind velocity. Total of 43824 wave data from ERA5 database at node 43.5°N, 9.5°E.

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.

$\overline{T_p}$ (s)	vel_{wind} (m/s)																				
H_s (m)	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	9 -10	10 - 11	11 - 12	12 - 13	13 - 14	14 - 15	15 - 16	16 - 17	17 - 18	18 - 19	19 - 20	20 - 21	21 - 22
6,5 - 7,0																						
6,0 - 6,5																				10,059	9,920	
5,5 - 6,0																		9,958	9,690	9,431	9,261	9,482
5,0 - 5,5																9,437	9,852	9,537	9,483	8,787		9,346
4,5 - 5,0													9,928	9,540	8,934	9,643	9,283	9,069	9,129	8,441		
4,0 - 4,5									9,971	9,137	10,023	9,879	9,571	9,431	9,283	9,035	8,754	8,683	8,851	8,110		
3,5 - 4,0							9,530	9,508	10,202	9,682	9,551	9,844	9,380	8,908	8,645	8,428	8,028	7,677	8,260			
3,0 - 3,5				9,324	9,750	9,392	9,366	9,875	9,819	9,244	9,233	8,986	8,752	8,066	7,812	7,756	7,826	7,143	7,080			
2,5 - 3,0		9,571	9,442		9,877	8,952	9,166	8,888	9,228	9,014	8,776	8,177	7,638	7,496	7,314	7,389	6,181	7,302				
2,0 - 2,5		9,336	9,324	8,729	8,845	8,642	8,788	8,810	8,272	8,045	7,588	6,838	6,766	6,572	7,035	6,794						
1,5 - 2,0	8,636	8,759	8,190	8,191	8,477	8,216	8,042	7,727	7,246	6,610	6,131	5,947	5,916	5,965	5,390	6,128						
1,0 - 1,5	8,029	8,218	8,198	7,909	7,541	7,280	6,830	6,246	5,620	5,281	5,163	5,228	5,245	4,080								
0,5 - 1,0	7,066	6,902	6,668	6,239	5,741	5,282	4,784	4,555	4,403	4,230	4,792	4,214	3,459	3,739	3,73027							
0,0 - 0,5	4,303	4,325	4,195	4,075	3,825	3,584	3,267	3,425	3,522	3,064												

Figure 9. Mean peak periods distribution along classes of significant wave height and wind velocity.

The Table in Figure 9, can also be represented in a H_{m0}-vel_{wind} map, as in Figure 10, with mean peak period values interpolated.

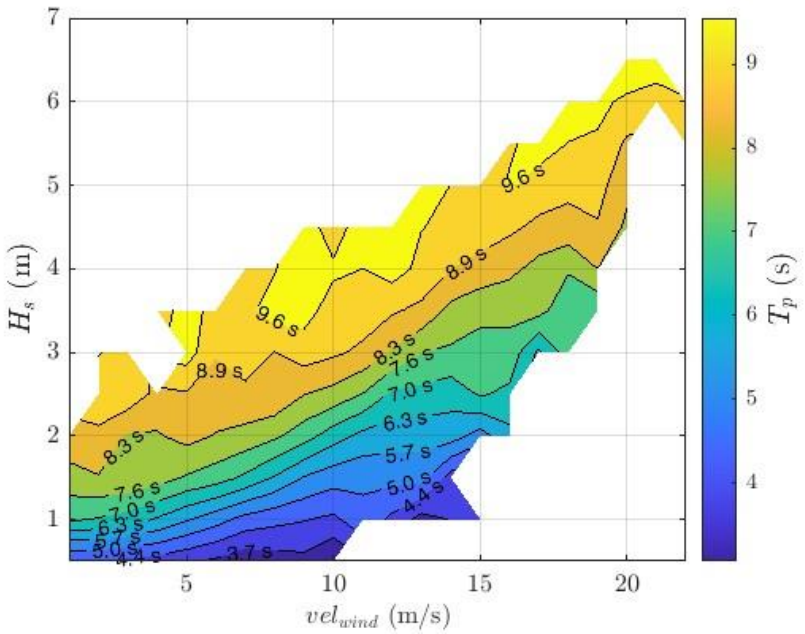

Figure 10. Wind and wave climate condition. Mean values of peak wave period among data classified as for Figure 9.

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.

Tables and Figures as those of Figure 9 and Figure 10, allow to get a mean peak period value, from the two altimeters data (H_{m0} and vel_{wind}). In order to provide an idea of the dispersion of the data, Figure 11 and Figure 12 report the boxplots of the wave periods as a function of H_{m0} and vel_{wind}.

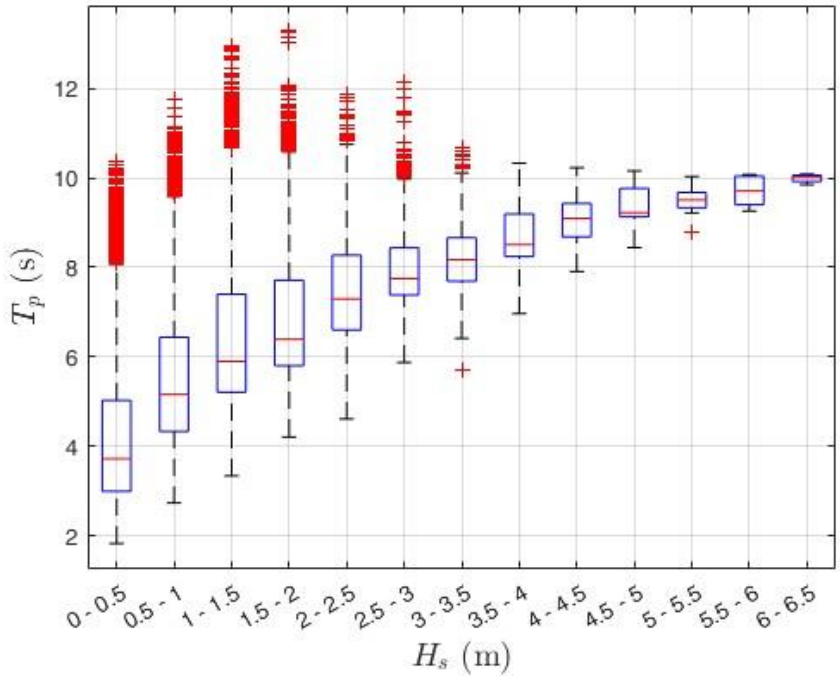

Figure 11. Boxplot of the peak period as a function of the significant wave heights

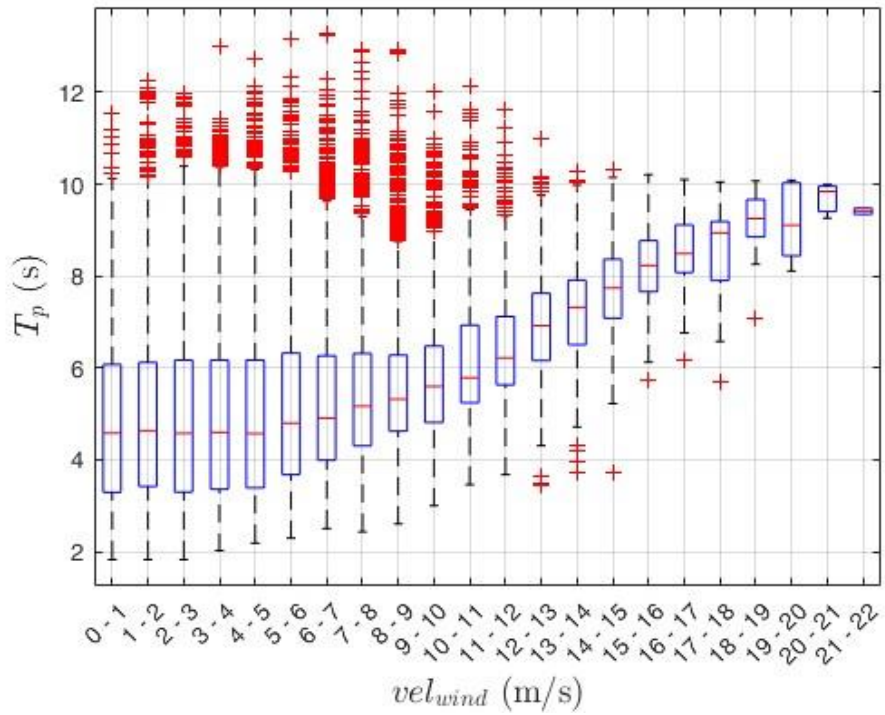

Figure 12. Boxplot of the peak period as a function of the wind velocity

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.

For both Figures, on each box, the central mark (red line) indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using red cross markers. A larger number of outliers occur for lower values of both H_{m0} and vel_{wind}. These outliners are all relative to peak wave period higher than the mean values, indeed as expected are relative to the swell sea states, i.e. large wave periods for smaller wave height and wind velocity.

Therefore, once H_s and vel_{wind} are known, an estimation of the mean peak wave period can be obtained. This estimation is more accurate as more extreme is the sea state; while, for lower values of both wind velocity and wave height, the accuracy of the estimation is lower. This is because both swell and sea state can be characterized by low H_{m0} and low vel_{wind}.

Indeed, recalling the swell and sea states separation used in Figure 4, a scatter plot between H_{m0} and vel_{wind} is now reported in Figure 13, using red dots for sea waves and black dots for swell waves.

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.

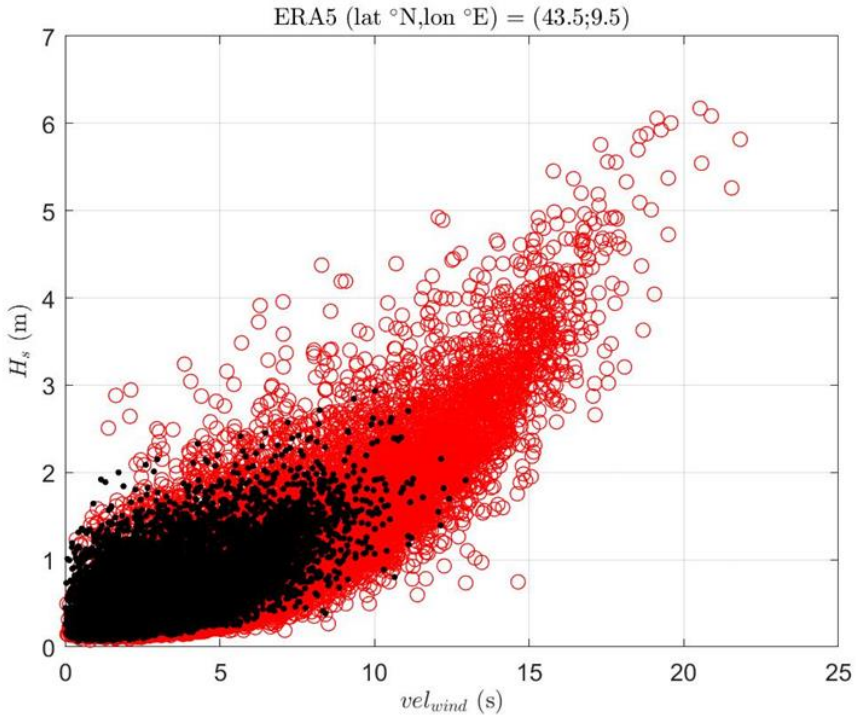

Figure 13 Scatter diagram of significant wave height and wind velocity. Red dots are relative to sea state with a wave steepness higher than 0.015, while black dots refer to wave steepness lower than 0.015. Total of 43824 wave data from ERA5 database at node 43.5°N, 9.5°E.

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. METHODOLOGY FOR WAVE PERIOD INFERENCE FROM SATELLITE DATA

The methodology described in this section, makes use of the physical considerations reported in section 2. To infer the wave period, T_p, using only significant wave height H_{m0} and wind speed vel_{wind} - parameters directly observable by satellite altimeters - the following two-step methodology is proposed. In the first step a classification among sea and swell wave is identified, based on a complete dataset of wave data. In the second step two different wave period inferences are obtained, one trained on swell waves and the other on sea waves. The final output would be identifying the wave type probability to be swell or sea wave, and further provide the wave period prediction.

This method capitalizes on numerical datasets and machine learning techniques to achieve accurate wave period predictions.

3.1 Step 1: Sea and Swell Classification

3.1.1 Data Preparation and Wave Classification

Dataset: The ERA5 dataset is analyzed for all nodes within the project's area of interest.

Wave Classification: For each data point in the dataset, the relationship between significant wave height (H_{m0}) and peak period (T_p) is evaluated.

A polynomial curve is defined to act as a boundary, dividing data into two wave classes:

- Swell waves: Generated far from the area of observation and characterized by long wave periods.
- Wind-sea waves: Locally generated, characterized by shorter wave periods.

Data points are labeled as either swell or wind-sea based on their position relative to this boundary.

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.

3.1.2 Training a Classification Tree for Wave Type Prediction

Inputs: Each labeled data point is associated with its spatial position, significant wave height (H_{m0}), and wind speed (vel_{wind}).

Classification Model:

- A classification tree (or probabilistic regression tree) is trained to predict whether a wave is swell or wind-sea.
- The model outputs:
  - A binary classification (swell or wind-sea).
  - A probabilistic score indicating the likelihood of the wave belonging to the swell class.

3.2 Step 2: Wave Period Estimation

3.2.1 Dataset Subdivision

The dataset is divided into two subsets based on the output of the classification model:

- Swell Waves Subset: Includes all data points classified as swell.
- Wind-Sea Waves Subset: Includes all data points classified as wind-sea.

3.2.2 Training Regression Models

For each subset, a regression tree is trained to predict the wave period (T_p).

Inputs:

- Spatial position.
- Significant wave height (H_{m0}).
- Wind speed (vel_{wind}).

Output:

- The regression tree provides an estimated wave period specific to the corresponding wave type (swell or wind-sea).

3.3 Final Output

Once the classification and regression models are trained, the methodology can process satellite observations (H_{m0} and vel_{wind}) to provide the following:

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. Wave Type Probability: The likelihood that the observed wave is a swell wave.
2. Wave Period Prediction:
    - A specific wave period (T_p) value for swell waves.
    - A specific wave period (T_p) value for wind-sea waves.

This structured approach ensures accurate wave period estimation while leveraging the distinct characteristics of swell and wind-sea waves. By combining numerical datasets and machine learning, the method maximizes the potential of available satellite data to improve offshore and metocean monitoring systems.

4. IMPLEMENTATION OF THE METHODOLOGY FOR WAVE PERIOD INFERENCE FROM SATELLITE DATA

In this section, the previously described methodology for inferring the peak wave period (T_p) using the satellite measurements (H_{m0} and vel_{wind}) has been implemented.

All the algorithms have been developed using Python 3.12.7 and the scikit-learn 1.5.1 library.

4.1: Random Forest for classifying sea and swell states

The first step for inferencing the wave period is that of discerning whether the consider state is a sea state or a swell state. Even if this is actually a classification problem, from an algorithm point of view, it was treated as a regression problem, where the probability of belonging to a swell state ($Prob_{swell}$) was considered as the target variable. Analogously, it would have been possible to consider as the target variable the probability of belonging to a sea state ($Prob_{sea}$), since these two probabilities are complementary:

$Prob_{sea} = 1 - Prob_{swell}$

Indeed, the targets of the random forest were either 1 (if it was a swell state) or 0 (if it was a sea state). The random forest had as input data for predicting the sea state the following variables:

- Significant Wave Height H_{m0};

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.

- Wind Speed vel_{wind};
- Spatial coordinates latitude and longitude;
- Time indication;

In order to capture more easily information about the seasonality and the daily cycles, the time indication was given in multiple variables:

- Days since the start of the training period (2018-01-01 00:00);
- The specific year;
- The specific month;
- The specific day;
- The specific hour.

The Random Forest was implemented with the 'RandomForestRegressor' command from scikit-learn. All the available data was partitioned a priori randomly into a training dataset, a validation one and a testing one, using an 80%-10%-10% partitioning. The output of the Random Forest was a number between 0 and 1. If the predicted number was higher than 0.5, it was considered as predicting a swell state, while if it was lower than 0.5, it was considered as predicting a sea state.

4.1.1 Classification Tree results

Below are reported the confusion matrices of the classification tree for both the validation and testing sets.

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.

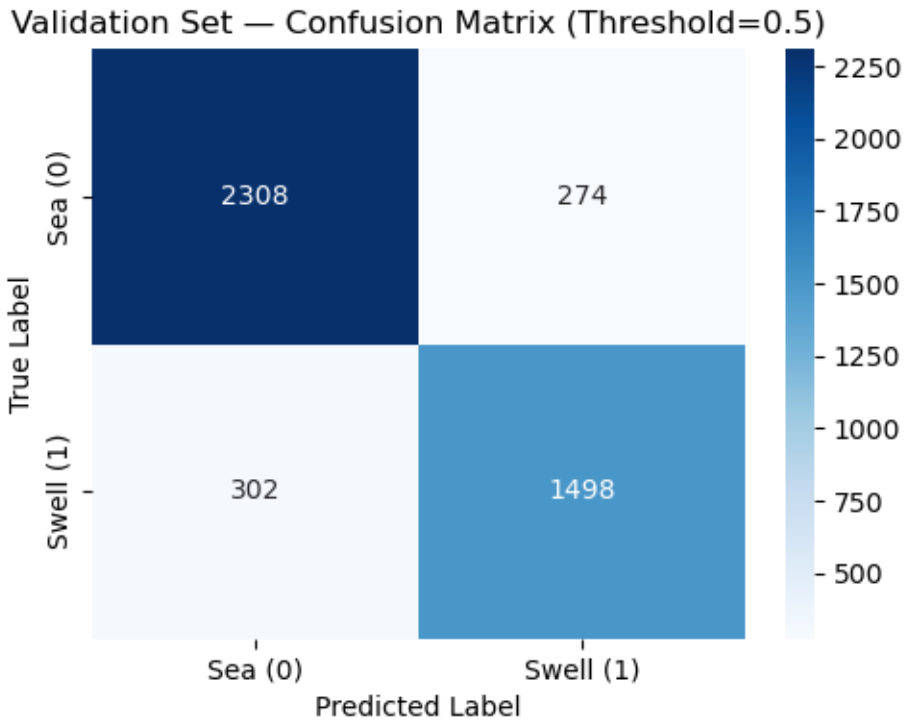

Figure 14 Confusion matrix of the classification tree on the validation set

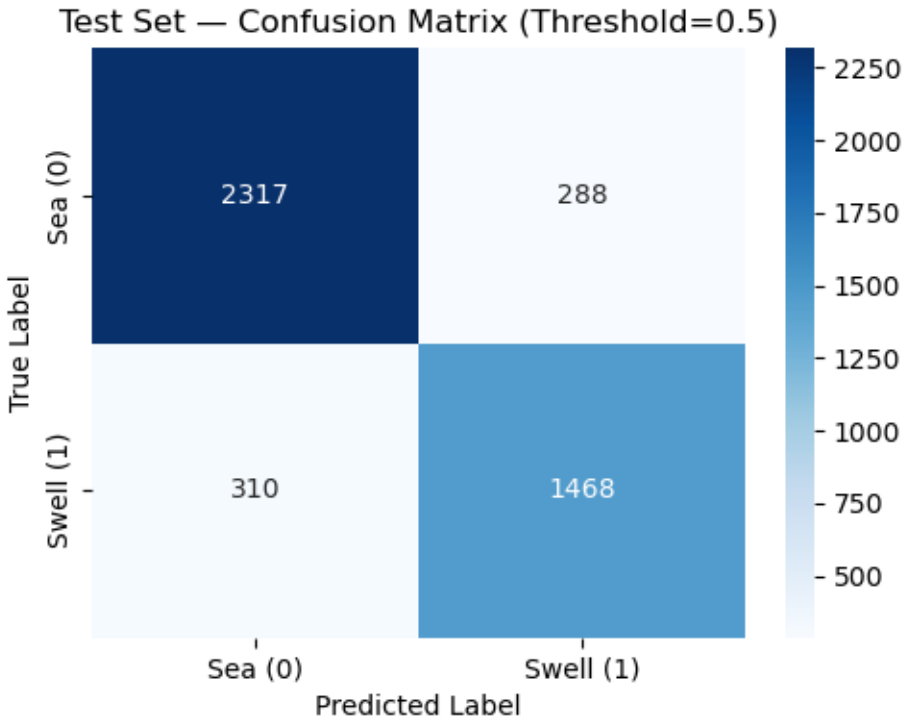

Figure 15 Confusion matrix of the classification tree on the test set

As can be seen from Figure 14 and Figure 15, the developed classification tree is able to correctly classify the majority of the unseen states, misclassifying

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.

only around 13% of the states in both validation and testing. The algorithm is also avoiding overfitting, considering the absence of a difference between the performances in validation and testing.

4.2 Random Forest for predicting wave period in sea states and in swell states

After having developed a classification tree for discerning whether the considered state is a swell state or a sea state, two different regression trees have been trained separately, one using only swell state data and the other using only sea state data. Both these trees aimed at predicting the peak wave period (T_p), again using as covariables the spatial coordinates, the parameters measured by the altimeter (H_{m0} and vel_{wind}) and the temporal information, as in the classification tree.

4.2.1 Regression Tree results

The obtained predictions and the errors of the previously described regression trees, along with the achieved metric performances, are shown in the figures and tables below.

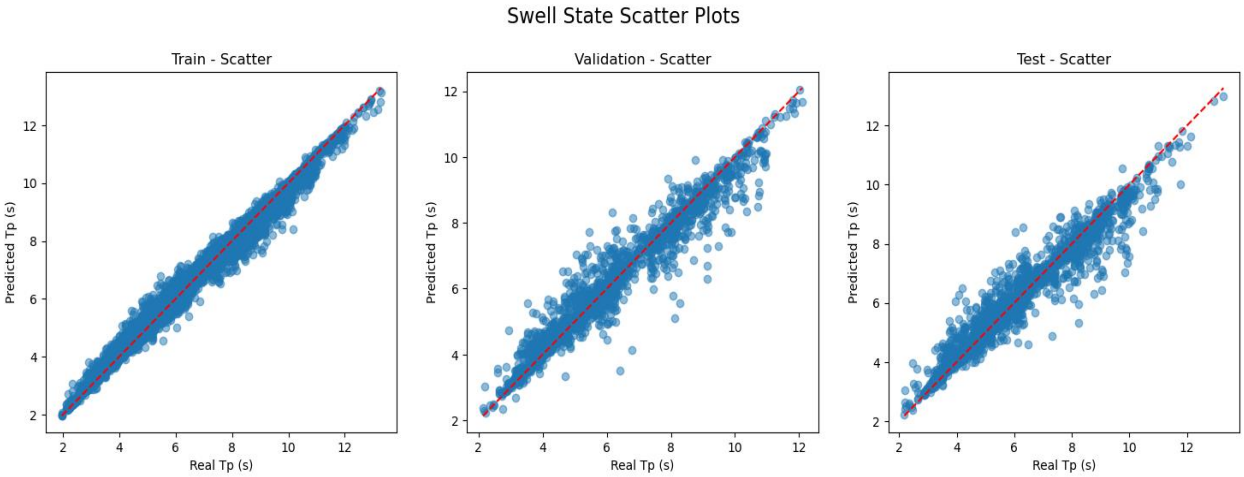

Figure 16 Scatter plots of the regression tree for the swell states

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.

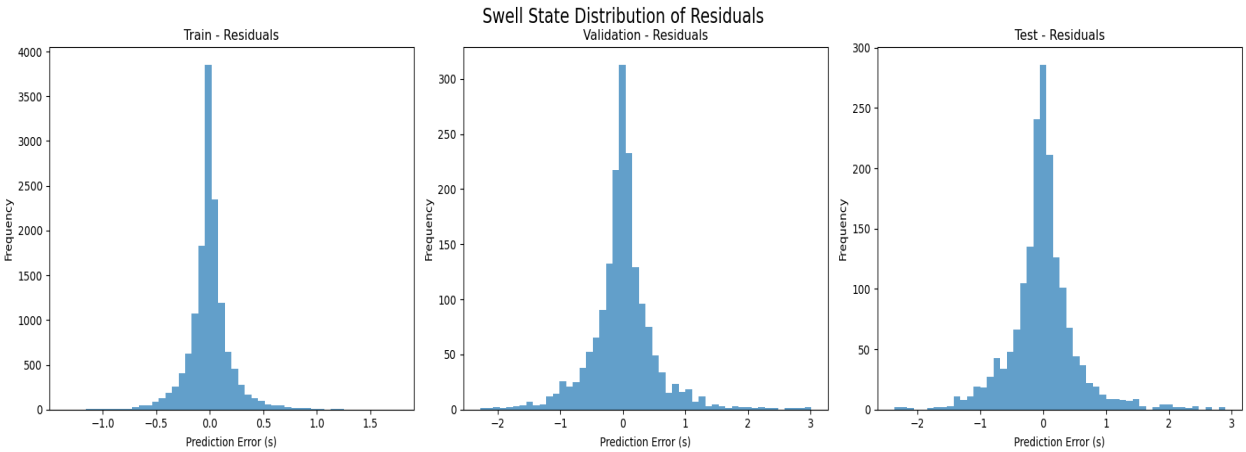

Figure 17 Residuals distributions of the regression tree for the swell states

Table 1. Obtained metric performances of the regression tree for the swell states

Performance metric	Training	Validation	Test
MAE (s)	0.1272	0.3425	0.3521
RMSE (s)	0.1987	0.5236	0.5349
R^2	0.9905	0.9355	0.9285

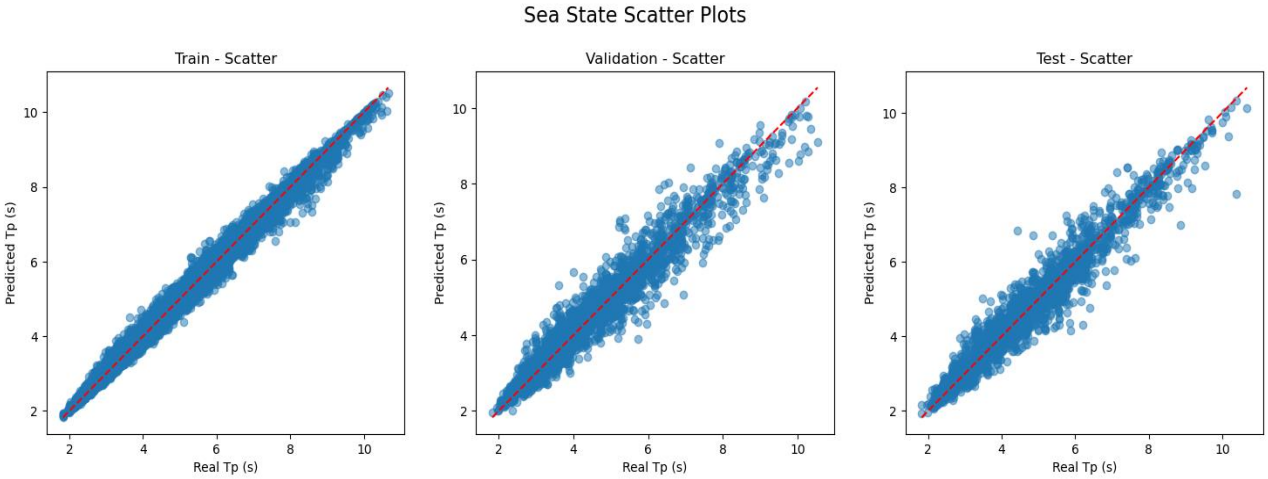

Figure 18 Scatter plots of the regression tree for the sea states

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.

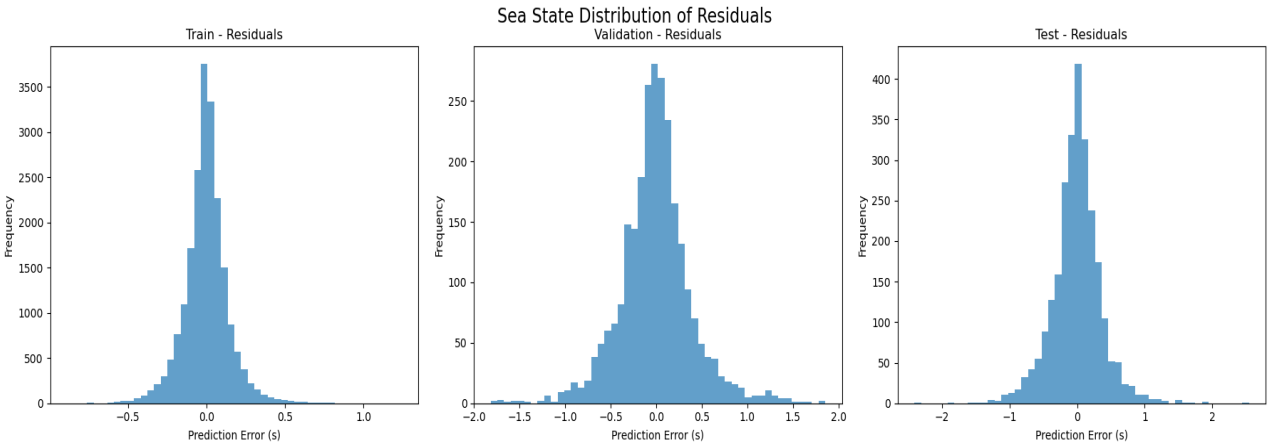

Figure 19 Residuals distributions of the regression tree for the sea states

Table 2. Obtained metric performances of the regression tree for the sea states

Performance metric	Training	Validation	Test
MAE (s)	0.1003	0.2828	0.2635
RMSE (s)	0.1430	0.3964	0.3689
R^2	0.9925	0.9435	0.9479

As shown by the reported figures and tables, the regression trees are able to correctly infer the peak period using spatial and temporal data along with the parameters measured by the altimeter. In doing so, almost all the variance of the predicted parameter is explained, also in the validation and test sets, and the predictions have a relatively low error of around 0.3s in the validation and test sets. Even if the results are satisfactory for both the swell states and the sea states, it is possible to note that the metric performances for the swell states are a bit lower, indicating how in these states it is more difficult to correctly infer the wave period.

4.3Random Forest for predicting wave period in sea state and in swell state with no temporal information

A final study was conducted on how the previously described regression trees performed in the same setup but removing the temporal information from the

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.

available covariables. The algorithms were run with the same data, only removing all the time variables.

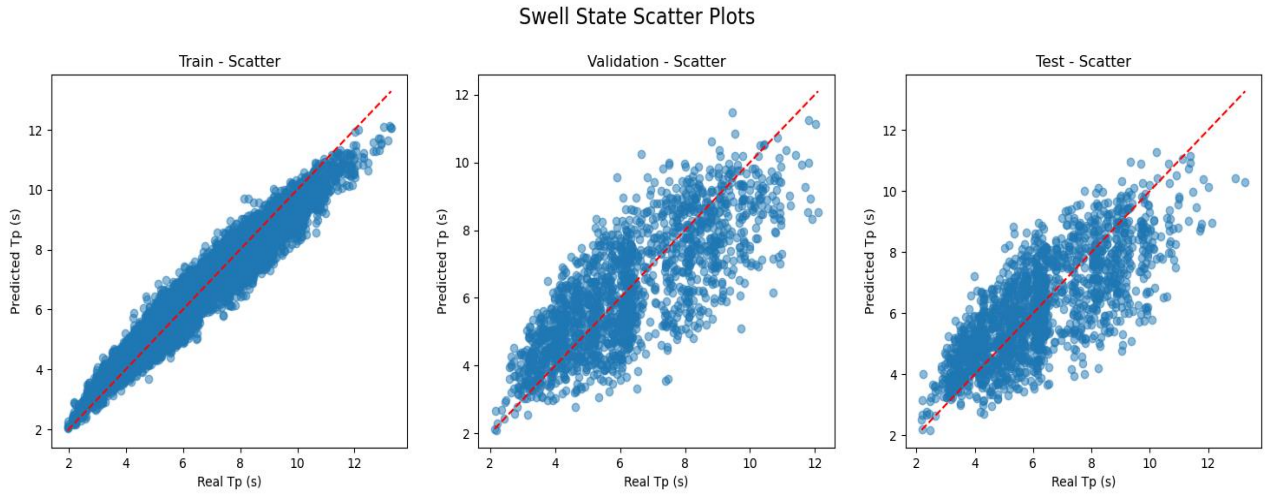

Figure 20 Scatter plots of the regression tree without time info for the swell states

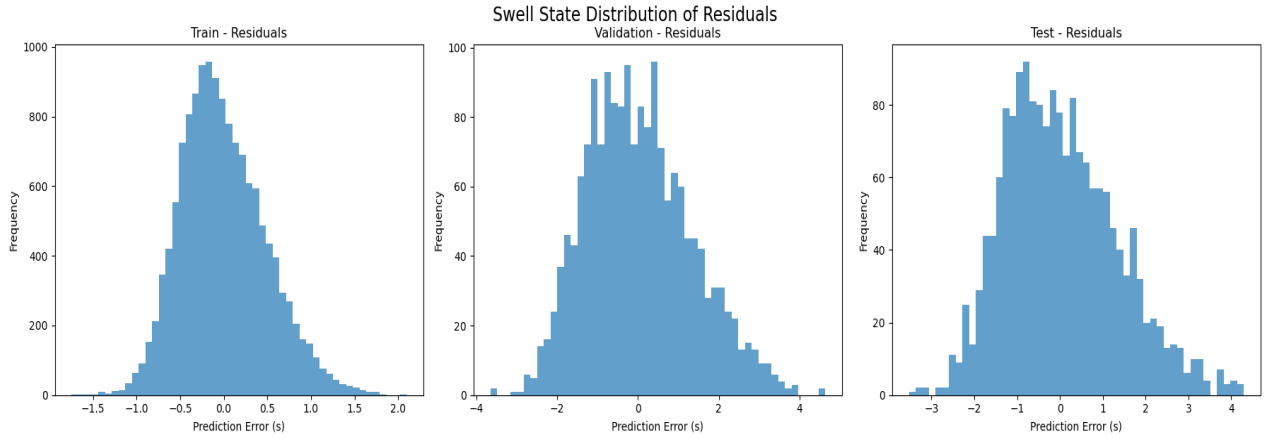

Figure 21 Residuals distributions of the regression tree without time info for the swell states

Table 3 Obtained metric performances of the regression tree without time info for the swell states

Performance metric	Training	Validation	Test
MAE (s)	0.3905	1.0565	1.0603
RMSE (s)	0.4880	1.3078	1.3161
R^2	0.9429	0.5977	0.5672

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.

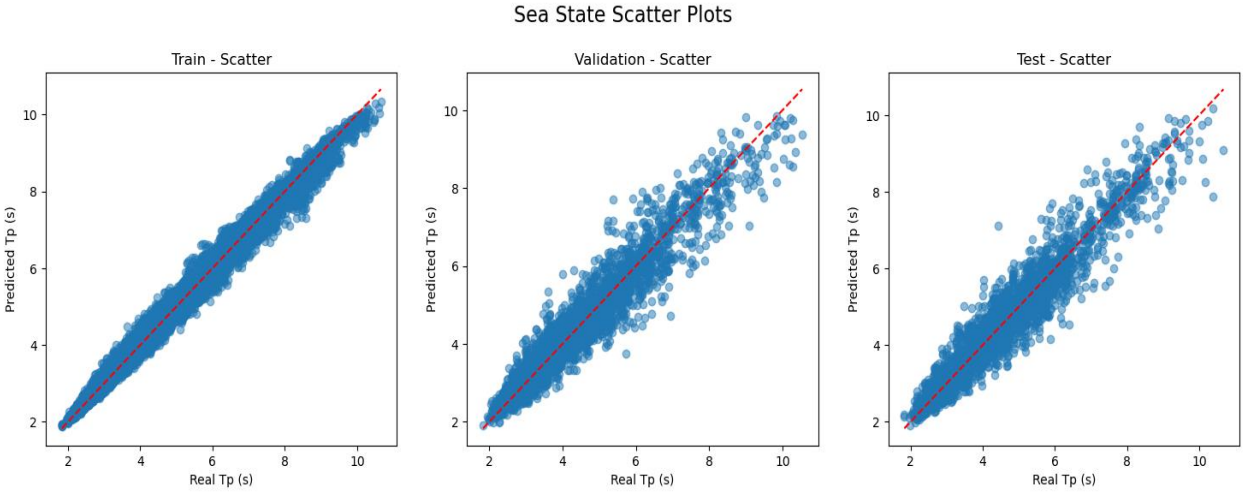

Figure 22 Scatter plots of the regression tree without time info for the sea states

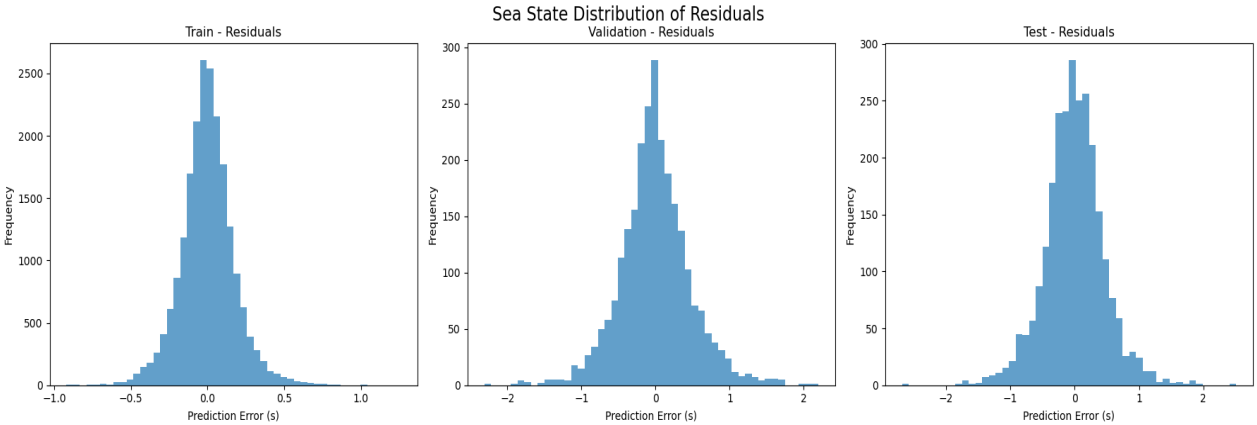

Figure 23 Residuals distributions of the regression tree without time info for the sea states

Table 4 Obtained metric performances of the regression tree without time info for the sea states

Performance metric	Training	Validation	Test
MAE (s)	0.1307	0.3555	0.3484
RMSE (s)	0.1770	0.4814	0.4651
R^2	0.9865	0.9167	0.9171

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 it could be expected, the results are generally worse than before, having the random forests less information available for inferring the wave period. However, this worsening is relatively small for the random forest of the swell states, which maintains accurate predictions, but much higher for the random forest of the sea state, where the mean error of the predictions almost triplicates without time information, indicating the higher complexity in inferring the wave period for those states.

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.

5. CONCLUSIONS

A methodology to infer wave period data, when only wave heights and wind velocity data are available, has been outlined. The methodology is applied to hindcast the wave period from altimeters data (i.e. wave height and wind velocity). The method is based on numerical datasets and machine learning techniques to achieve accurate wave period predictions.

The method is successful, since it achieves R^2 values of 94.79% and 92.85% for sea and swell states, respectively, which are higher than the 85% minimum threshold target initially set.

REFERENCES

[1] Atlante Delle Onde Nei Mari Italiani. Italian Wave Atlas. 2004. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici

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.