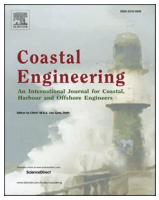
Coastal Engineering 192 (2024) 104574



Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng





A data model to forecast the morphological evolution of multiple beach profiles

Willian Weber de Melo [a,b,*], José Pinho [a], Isabel Iglesias [b]

[a] Centre of Territory, Environment and Construction (CTAC), School of Engineering, University of Minho, Campus de Azurém, Guimarães, 4800-058, Portugal
[b] Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Terminal de Cruzeiros Do Porto de Leixões, Av. General Norton de Matos, Matosinhos, 4450-208, Portugal

ARTICLE INFO

Keywords:
Beach profile evolution
Data modelling
Machine learning
Morphodynamics
Random forest

ABSTRACT

Beaches are a natural defense against extreme events, such as storms and hurricanes, whose intensity and frequency are expected to increase in the future due to climate change. In this context, models that forecast the morphological evolution of coastal areas can be used to anticipate the effects of future scenarios, allowing early action to mitigate the damage caused by extreme events. Hence, this study included data from three different monitoring programs in data models to simulate the seasonal morphological evolution of several Portuguese beaches. Two different data models were implemented using the Random Forest algorithm. One was fed with profile data and wave conditions while the other considered also sediment size data. Both models achieved suitable performances, but the inclusion of sediment data reduced the model errors and variance, and thus improved model performance. It was demonstrated that combining data from multidisciplinary campaigns can be a solution to generate reliable and robust morphological forecasting models.

1. Introduction

Coastal zones shelter about 10% of the world's population. These regions are among the most threatened areas exposed to the effects of storms and fluvial and coastal floods. Additionally, in the actual context of climate change, it is expected a sea level rise and an increase in the frequency and intensity of extreme events due to the augment of the global mean temperature (McGranahan et al., 2007; Pörtner et al., 2022). In the last decade, most European coastal cities experienced a sea level rise, which increased the flood risk exposure of these populations and their vulnerability (UN-Habitat UNHSP, 2022).

In coastal zones, beaches play a crucial role as natural barriers in protecting upland structures and infrastructure against the impact of extreme events, namely storms and wave action (Mil-Homens et al., 2022). The wave height and frequency, jointly with the coastal currents, the sediment size, and the beach type (dissipative, intermediate, or reflective), are the main morphodynamic drivers at coastal zones (Aagaard et al., 2013; Prodger et al., 2017). The front dunes at the shoreface dissipate part of the wave energy, reducing the wave sediment transport capacity and erosion on the backshore. However, their stability depends on a delicate sediment balance, considering all sediment particles entering or leaving the system. If this sum is positive, the area is being accreted; otherwise, it is being eroded.

Beach morphological evolution has been the focus of much research in the last decade, including studies that represent current and projected erosive/accretional patterns at several locations using different tools. Vousdoukas et al. (2020) investigated the trends of shoreline dynamics considering the effects of different climate change scenarios resulting from the analysis of synthetic data series generated by the Monte Carlo method, Karunarathna et al. (2012) implemented a cross-shore beach profile evolution predictor by merging a data-driven approach with a diffusion-type model, Gunawan et al. (2023) implemented a data model to forecast shoreline changes using satellite images, and Kalligeris et al. (2020) implemented two process-based numerical models to forecast the beach profile evolution in California. All the mentioned works identified that numerical models had low skill when predicting multiple beach profiles with a single set of parameters.

In this context, promoting monitoring projects in coastal areas is paramount to tracking coastal morphological evolution. While long-term observations can increase the knowledge about the causes of the morphological change and be a reliable source of information to decision-makers (Benveniste et al., 2019), short-term observations are

* Corresponding author. Centre of Territory, Environment and Construction (CTAC), School of Engineering, University of Minho, Campus de Azurém, Guimarães, 4800-058, Portugal.
E-mail address: id9257@alunos.uminho.pt (W. Weber de Melo).

https://doi.org/10.1016/j.coastaleng.2024.104574
Received 15 July 2023; Received in revised form 11 June 2024; Accepted 5 July 2024
Available online 6 July 2024
0378-3839/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

indispensable to monitoring the response of highly dynamic beaches and the effects of extreme events (Jayson-Quashigah et al., 2019). The data collected by monitoring projects is fundamental for the application of modelling methodologies, such as numerical and data modelling, which are highly dependent on the quality and quantity of field data.

Numerical modelling methodologies rely on monitoring data to calibrate the models and ensure their results are valid (Biolchi et al., 2022; Matheen et al., 2021; Suzuki and Cox, 2021). However, these tools need considerable computational resources, which will be even more extensive if different sediment classes are considered along the computational domain (Prodger et al., 2017; Weber de Melo et al., 2022). Morphodynamic numerical models, such as Delft3D and XBeach (Deltares, 2018; Roelvink et al., 2009), solve the shallow water equations coupled with sediment transport equations, updating the bathymetry after each hydrodynamic timestep. Although these models can achieve excellent results, their computational costs tend to be high, particularly for high-resolution models, which can be a critical restriction for adopting this solution as the main forecasting tool for operational systems.

Data modelling, which is based on artificial intelligence (AI) methods such as machine learning (ML) and deep learning (DL) techniques, depends not only on the data quality but also on its quantity to achieve adequate performance metrics (Razavi et al., 2012; Weber de Melo et al., 2023; Ye et al., 2021). These models have lower computational costs than numerical models to produce accurate solutions (Antolínez et al., 2016; Gharagozlou et al., 2022; Weber de Melo et al., 2023), even if different sediment classes are considered in the computation. AI-based models compute simpler equations resulting from training the model's parameters using input and output data. Therefore, data models should be considered an alternative and valuable tool for implementing methodologies capable of forecasting the morphological evolution of beaches.

AI algorithms have already been applied to assess shoreline evolution (McAllister et al., 2022), but most applications use satellite imagery as the main data source. Luijendijk et al. (2018) implemented a classification and regression tree model to determine the water-land transition using images from the Google Earth Engine, and Chen et al. (2023) implemented a physical-informed neural network to map the nearshore bathymetry also using satellite remote sensing data. These studies demonstrate the potential of AI to study the evolution of coastal morphology. However, few studies have combined these methods with data collected *in-situ*. While remote-sensing data is an easy and affordable method to monitor large areas, the resolution of publicly available images is restricted to 10 m (McAllister et al., 2022), which is inadequate to monitor the evolution of beach profiles with sufficient accuracy. This strongly limits the options for those who cannot afford to buy or measure beach profiles with higher-resolution approaches.

Hence, this study contributes to filling the gap in the applications of AI algorithms using *in-situ* measured beach profiles. The outcomes of multiple coastal monitoring programs were merged to implement two data models for forecasting the morphological evolution of multiple seasonal beach profiles along the northern Portuguese coast. The data models were implemented using the random forest (RF) algorithm. The first model (M1) used beach profile historical data, wave variables, and sand size data as inputs, and the second model (M2) used the same data as M1 but omitted sediment size data to understand the importance of this variable in the forecast. The outputs of both models were twelve beach profiles from eight different beaches in northern Portugal. Lastly, the forecast of beach slope and beach retreat/advance was computed to assess the reliability of the data model results. In addition, the models were tested under storm conditions to evaluate their performance outside the training data range.

2. Methodology

This study applied the RF algorithm to develop two data models that are able to predict the seasonal beach profile evolution due to wave action. This section describes the study area, the sources and characteristics of the data, and the main steps to implement the RF models.

The methodology framework (Fig. 1) summarizes the main tasks required to implement an AI model, as well as the specifications adopted in this study. The first task included the identification of data sources, the data collection, and the characterization of the study area. The second task addresses data processing, ensuring that the data type, values, and format are adequate for the selected algorithm. The data is usually normalized, outliers are removed, and features are combined to reduce data that may not be relevant to the model's objective. The necessity of normalization depends on the ML algorithm used. In the model implementation task, an algorithm and a training technique are defined according to the characteristics of the data and the algorithm capacity. The last task is focused on validating the obtained results using different metrics. Details about how each task was executed in this study are described in the following sections.

2.1. Study area

Northern Portugal has a Mediterranean climate dominated by the Azores' high-pressure system, with wet winters and dry summers (Sanderson et al., 2023). The region is characterized by high mesotidal conditions and wave-dominated dynamics, with waves exhibiting a predominantly northwesterly orientation. The significant wave heights in the region typically range between 2 m and 3 m, while the wave periods generally range between 8 s and 12 s (Alveirinho Dias and Nittrouer, 1984; Bio et al., 2020a). During storm events, low-pressure systems generated over the Atlantic Ocean can cross the region with associated *S*-SW winds. Extremely energetic conditions are produced during these events, with wave heights that can reach between 9 m and 12 m, although waves between 3 m and 6 m are most common during those events (Viitak et al., 2021). During these highly energetic conditions, which normally take place during winter, it was demonstrated that the Portuguese coastline is highly susceptible to severe erosion (Bio et al., 2022).

The Portuguese coast includes a wide range of morphological and sedimentary environments, such as beaches, cliffs, estuaries, lagoons, and barrier islands. Since the 19th century, an overall regressive trend of the coastline has been documented as being caused by human intervention (dams and coastal engineering structures construction, sand extraction, and agricultural practices, Santos et al., 2021). The NW Portuguese coast has been historically defended against the sea since the early 20th century. However, natural and artificial barriers can affect wave patterns, causing localized reversals in the prevalent longitudinal drift, mainly directed to the south.

This erosion has led to dune loss, the transformation of some sandy beaches into gravel beaches, and inefficient transverse structures (Granja and Pinho, 2012). Inside the Portuguese littoral, the NW coast is characterized by cliffs, low rocky coasts, sandy and gravel beaches, and extensive linear beaches backed by coastal dune systems. It was demonstrated that this region was prone to severe erosion rates that reached 8 m/yr (Ponte Lira et al., 2016). However, in the medium-term (last decade), the NW Portuguese shoreline presented a relatively stable behaviour with a trend towards erosion in recent years. This trend, combined with the projected sea level rise, changes in wave climate, and increased frequency and strength of extreme events, suggests that erosion and shoreline retreat could be a severe problem (Bio et al., 2022).

Northern Portugal's coastline sediments originated from igneous and metamorphic rocks in the Precambrian and Palaeozoic eras (Alveirinho Dias and Nittrouer, 1984). The sediment deposits of the region have a predominant marine origin below 50 m height and a fluvial origin between 50 m and 130 m (Araújo, 2000), with a transition character between 37 m and 60 m (Pedrosa, 2013).

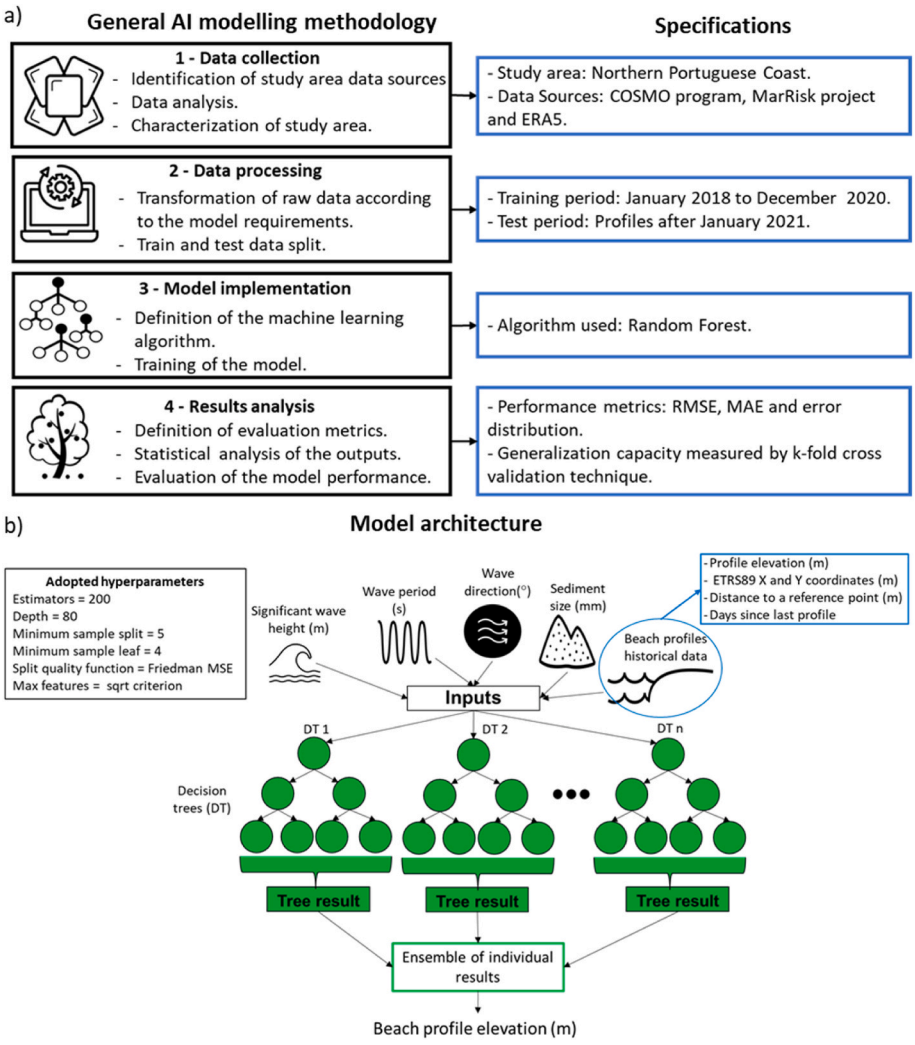


Fig. 1. Methodological approach: a) Methodology framework with the steps required to implement a data model. b) Architecture of the model and details on the input data and adopted hyperparameters.

2.2. Beach profile data

Two different datasets were selected to define the beach profiles that will feed the data models. The COSMO program monitored several beach profiles along the Portuguese coast. Meanwhile, the MarRisk project defined the coast type and the beach sediment sizes at several coastal locations. The combination of both datasets resulted in beach profile data from 12 sites on the NW coast of Portugal. These profiles are distributed along a coastal stretch of approximately 90 km in length between Moledo Beach, at Caminha municipality, and Paramos Beach, at Espinho municipality (Fig. 2). The beach profiles were selected by considering available data from two monitoring programs, the COSMO program (APA, Agência Portuguesa do Ambiente, 2023) and the MarRisk project (Bio et al., 2020b).

The COSMO project provided beach profiles measured between 2018 and 2021, with observations approximately every three months, resulting in 11–12 profiles for each location. Five parameters were selected from these surveys: the elevation referred to the mean sea level (Z_MSL, m), the x-coordinate (m), the y-coordinate (m), the distance to a reference point in the profile (m) and the time since the first observation (days). The coordinates were referenced using the ETRS89 reference system. The reference point was defined considering the profile below 6 m (MSL), corresponding to the profile extension exposed to wave action under average conditions. The horizontal errors in the profile survey varied from 1.1 cm to 2.5 cm, and the vertical errors ranged from 1.8 cm to 4.9 cm. The coordinates provide information about the geographical location of the profiles, allowing us to predict the elevation of points that were not previously measured as long as the coordinates are in the same range as the training coordinates.

Beach profile data is presented in Fig. 3, highlighting the different morphodynamic patterns of the beaches. Profiles such as the ones measured in Moledo, Âncora, and Árvore P2 revealed stronger morphodynamics than the ones observed in Neiva P2 and Paramos, whose profiles were relatively stable along the monitored period. Also, some profiles are wider than others. For instance, Neiva P2 extends to approximately 70 m, whereas Paramos extends to almost 200 m.

The active profile on some beaches begins precisely at the first point, at 6 m elevation (MSL). In contrast, others reveal an active profile below lower elevations, such as Neiva P2 and Ofir P3. Note that the COSMO program measured two profiles in Neiva and Árvore beaches and three profiles in Ofir beach. In the beaches that presented more than one profile at different locations, these profiles were identified as P1, P2, and P3. Most of the profiles are in an advance erosive state when the complete profile is observed (not shown), presenting erosion cliffs in the frontal dune, except at Moledo, Âncora, and Árvore P1, which presented a beach profile with a frontal dune, berms of variable extension, and a pronounced face.

Data from the MarRisk project were selected to define the coast type,

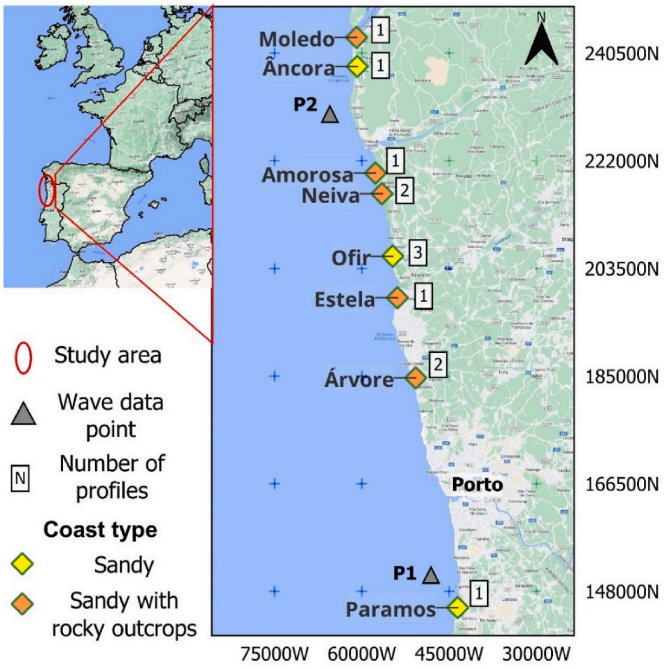


Fig. 2. Study area and location of the beaches where the profiles were measured. White rectangles indicate the number of profiles measured seasonally at each beach, and the rhombuses indicate the coast type of each beach. Coordinates are referred to the ETRS89 global reference system.

classifying it as sandy or sandy with rocky outcrops, and the beach sediment sizes for each location were obtained from sand samples collected at three different points for each profile (Table 1). These locations are identified in Table 1 by the number that follows the sediment size, in which "1″ refers to a berm intermediate position, "2″ refers to the berm edge, and "3″ refers to the beach face. Beach sediment sizes were analysed either by sieving and digital images (Pinho et al., 2024, in preparation).

It must be stressed that the COSMO program data had two profiles in Neiva and Árvore beaches and three profiles in Ofir, as mentioned before, but the MarRisk program only analysed the sediment size in a single profile of each one of these two beaches. Hence, the same sediment characteristics were assumed for the two profiles at Neiva beach and the three profiles at Ofir beach. Árvore beach is in the transition between the coast types, resulting in different classifications for each profile on that beach. It was assumed that the sand in these locations had the same D50 and D90, which correspond to the median diameter and the size in which 90% of the sediment mass is smaller, respectively. It was assumed that the sediment data remained constant over time due to the lack of sand samples collected in different seasons. The use of sediment data is relevant to include in the model the influence of the sediments on the beach morphology under a similar wave action.

The summary of the sediment size data and the profile locations are presented in Fig. 4. The sandy beaches had the lowest D50 and D90 with a more uniform granulometric distribution. Meanwhile, the sandy beaches with rocky outcrops had higher sediment sizes.

Results of both COSMO program and MarRisk project allowed the identification of four cliff profiles (VP Âncora, Ofir P1, Ofir P2, and Paramos) and eight Dean profiles, according to usual beach profiles classification (Jackson and Short, 2020). These classifications were based on the topographic survey conducted by the COSMO program and by using photogrammetric elevation data, aerial photography, and on-site observations collected in the MarRisk project.

The erosion/accretion trends of the selected beaches were presented using the profile data from the COSMO program (Fig. 5). The rates were computed by considering sequentially measured profiles, the first performed in one season and the last performed in the following season. Some profiles presented erosional and other accretional rates, but the sum of the rates for each seasonal pair is presented in Fig. 5. From this data, it can be inferred that accretional rates prevail between winter and spring and from spring to summer; meanwhile, erosion rates increase from autumn to winter and between summer and autumn. The beach accretion on the northern Portuguese coast is more intense for seasons with lower wave energy (winter to spring and spring to summer), reaching values of more than 60% of the occurrences along the monitored period. In contrast, erosion can occur during seasons with more intense wave energies (summer to autumn and autumn to winter), reaching frequencies between 58% and 69% along the monitored period.

2.3. Beach profile data processing

Beach profile data was divided into two sets. The first set, which covers the period between January 2018 and February 2021, was used to train the data model. The remaining data, from February 2021 to August 2021, was used to evaluate the model's performance. This partition resulted in ten to eleven profiles for each location for the model training and one for the model validation.

The profile variables used in the RF models were the Z_MSL (m), x-coordinate (m), y-coordinate (m), distance to a reference point in the profile (m), and number of days since a reference date (January 01, 2018) computed using the field campaign dates. The reference point in each profile was located 100 m before the first point in the profile, which presents a 6 m elevation (MSL). When the profiles have several points with 6 m elevation, only the points below the furthest 6 m elevation point were selected. This elevation value was adopted as the upper limit of the tide and wave action under average conditions (Fontán-Bouzas et al., 2022). This assumption restrained the available profile data to the area affected by tides and waves, reducing the amount of data to be passed to the RF model and filtering it to the profile extension under the influence of the coastal hydrodynamics. Thus, aeolian processes affecting the berm and frontal dunes dynamics will be left out of the analysis, as they are not the focus of this work.

Considering that the model was trained using the data of all the profiles simultaneously, the x and y coordinates are key parameters to provide the location information of the profiles along the study area. They allow the model to simulate the profiles at any given coordinate as long as the values are in the range of the profile coordinates. The distance to a reference point in the profile, which can also be referred to as the profile length, provided information on the potential magnitude of the wave effect at each profile point.

The mean and standard deviation (std) of the beach profile elevation and length are presented in Table 2. It can be observed that the elevation oscillated between 1.47 m and 3.70 m, and the profile's mean length varied between 142 m and 335 m. These metrics indicate that the profiles did not have any similar pattern because they were chosen according to the availability of data.

2.4. Wave data

Waves are the most relevant driver of morphological changes for the chosen profile extension in the study area. The wave data was included in the data models to provide them with information related to the hydrodynamics of the study area when trying to forecast the profile evolution. Wave data for this study was extracted from the Atlantic -Iberian Biscay Irish (IBI) Ocean Wave Reanalysis product managed by the E.U. Copernicus Marine Service Information (Copernicus Marine Service, 2022). Reanalysis datasets perform hindcast simulations that combine historical observations with climate models to generate consistent time series data. The temporal resolution of this database is hourly, containing records between January 1993 and December 2021 with a spatial resolution of 0.027° × 0.027°. The significant wave height

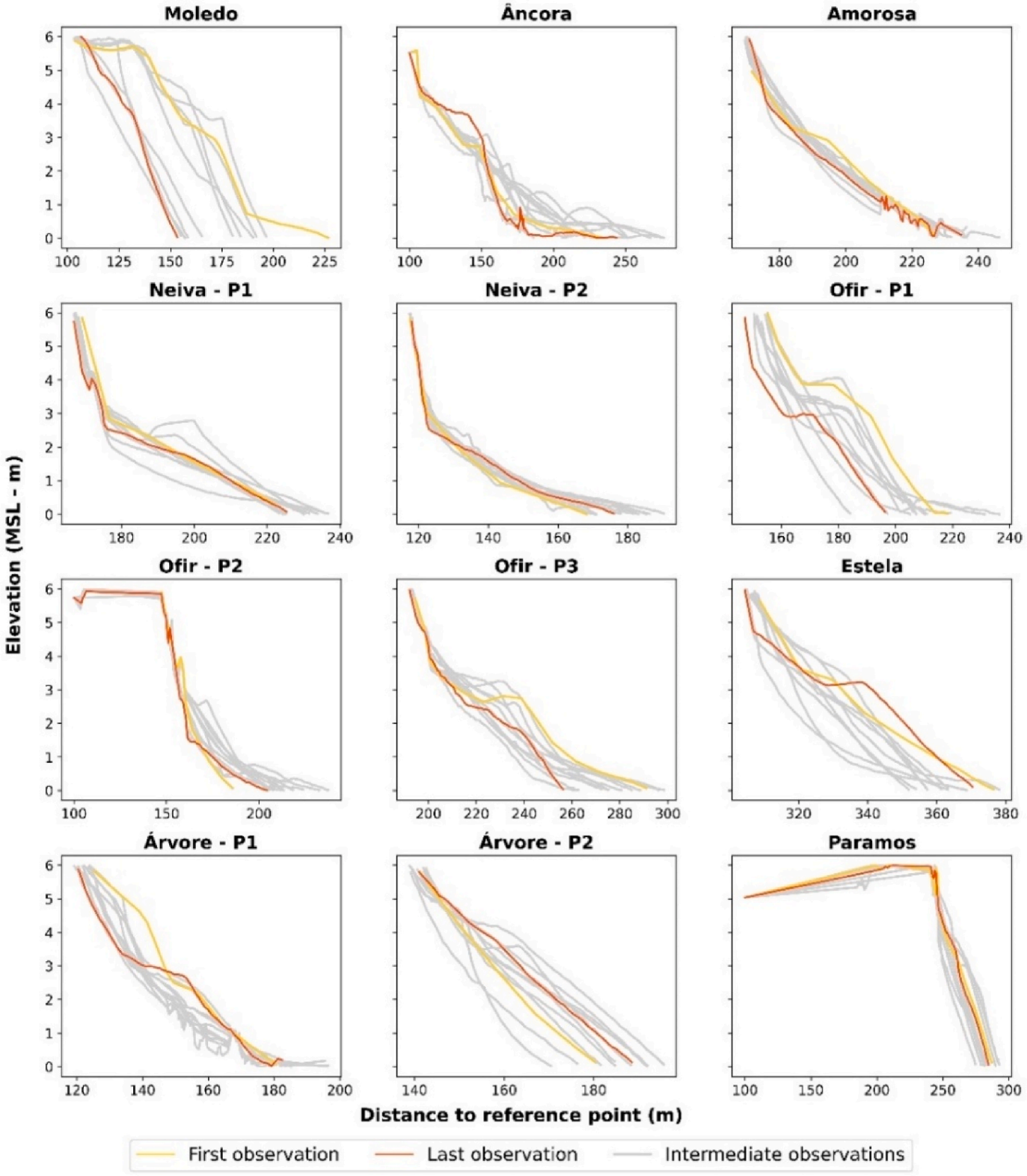


Fig. 3. COSMO beach profiles. The yellow line indicates the first profile surveyed, and the orange line is the last profile, which was used to assess the model's performance.

(HM0), wave period (WP), and wave direction (WD) variables were obtained for the period between January 01, 2018 and December 31, 2021, in agreement with the beach profile data. The wave data from the IBI product was extracted at two points, the first located near Paramos Beach (P1) and the other located between Âncora and Amorosa beaches (P2, Fig. 2). Fig. 6 represents the evolution of wave variables for the selected study period.

The wave data at both points present slight differences for HM0 but are identical for WP and WD for the analysed period. For WP and WD parameters, P1 and P2 lines overlap, and the mean and 90th percentile

Table 1
Beach sediments characteristic diameters (D50 and D90). The number following the diameter indicates where the sample was collected (1 – berm intermediate position; 2 – berm edge; 3 – beach face).

Beach	D50_1 (mm)	D50_2 (mm)	D50_3 (mm)	D90_1 (mm)	D90_2 (mm)	D90_3 (mm)	Coast type
Moledo	0.64	0.69	0.77	0.95	0.99	1.68	Sandy with rocky outcrops
Áncora	0.58	0.50	0.50	0.92	0.91	0.92	Sandy
Amorosa	0.37	1.48	0.42	0.61	4.67	2.58	Sandy with rocky outcrops
Neiva P1	0.36	0.29	0.22	0.76	0.49	0.44	Sandy with rocky outcrops
Neiva P2							
Ofir P1	0.35	0.45	0.27	0.49	1.38	0.75	Sandy
Ofir P2							
Ofir P3							
Estela	0.42	0.56	0.79	0.82	1.76	2.76	Sandy with rocky outcrops
Árvore P1	0.64	0.72	0.34	0.98	1.92	0.58	Sandy with rocky outcrops
Árvore P2							Sandy
Paramos	0.54	0.66	0.70	0.94	0.97	1.29	Sandy

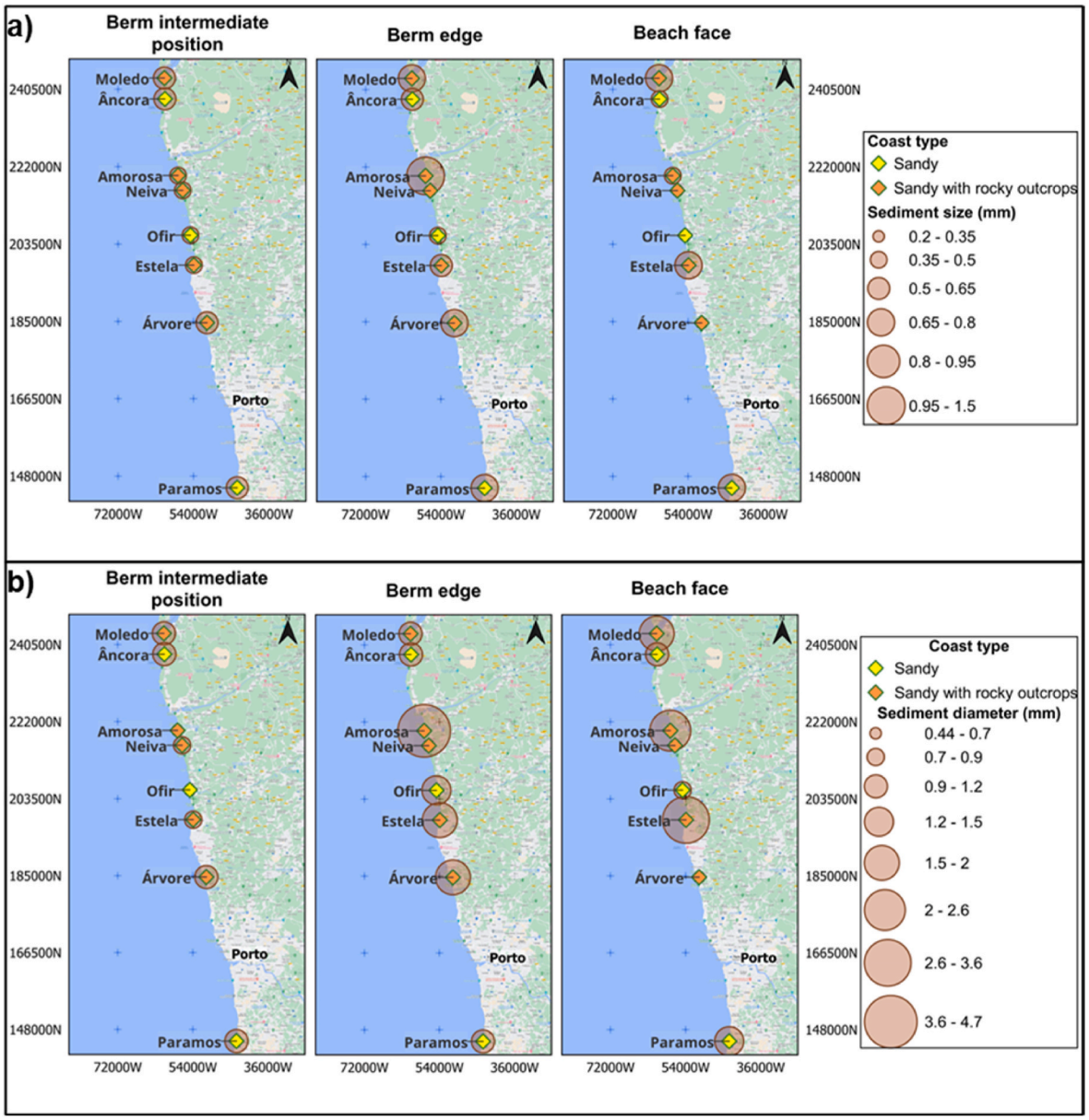


Fig. 4. Beaches sediment size. The sediment samples were collected at the berm intermediate position (first column), berm edge (second column), and at the beach face (third column). a) D50 results; b) D90 results.

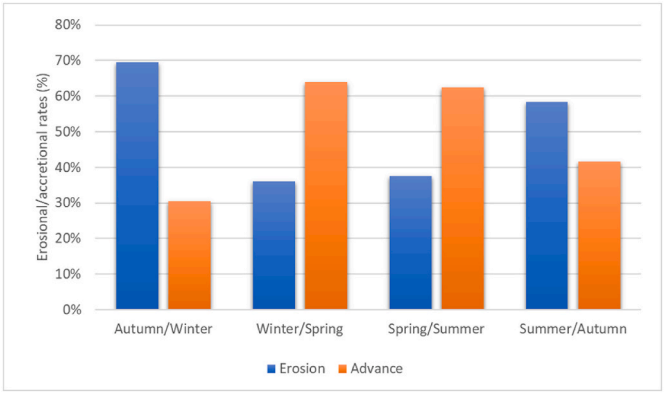


Fig. 5. Occurrence of erosion and accretion rates in northern Portugal beaches along the seasons, computed from the COSMO beach profiles data.

Table 2
Mean and std of beach elevation and profile length.

Profile	Mean elevation (m)	Elevation std (m)	Mean profile length (m)	Profile lenght std (m)
Moledo	3.7	1.9	142.2	24.5
Âncora	1.7	1.5	173.1	42.0
Amorosa	2.1	1.6	202.5	18.9
Neiva P1	1.9	1.4	197.0	18.6
Neiva P2	1.5	1.3	146.6	18.4
Ofir P1	2.3	1.6	182.9	19.9
Ofir P2	2.0	1.9	175.9	27.2
Ofir P3	2.1	1.6	235.5	26.6
Estela	2.4	1.7	334.6	17.7
Árvore P1	2.2	1.7	152.4	17.8
Árvore P2	2.8	1.6	163.5	14.2
Paramos	3.3	1.9	252.7	30.2

values are almost the same. The mean and the std values of the time series are presented in Table 3. The mean HM0 is 27 cm higher at P2, but the mean WP and WD have similar values. During the observation period, more energetic events during winter periods can be identified. The HM0 reached almost 8 m during the winters of 2019 and 2020, and the WP exceeded 20 s. Despite these values, it should be noted that there were no extreme storms during this period compared to historical data (Vieira et al., 2020).

It must be highlighted that the grid resolution of the IBI product is not fine enough to get accurate details for each one of the considered beach profiles. Therefore, P1 and P2 points were defined to provide wave data near the northern and southern limits of the study area (Fig. 2). The model learns how to process these inputs during the training, attributing the relevance of each considered input.

The distance between the beach profiles to P1 and P2 may indicate that some locations are more exposed to the influence of one of these points. The relevance of wave characteristics at P1 and P2 for each beach profile is determined during the training of the models. The parameters that associate the wave conditions at P1 with the output will present higher values when predicting a profile closer to P1, while the wave conditions at P2 will have lower weights when simulating a profile closer to P1. Hence, the associations that the model creates between the inputs and outputs are calibrated during the training phase.

Although the choice of only two data points may increase the uncertainties of the model, adding more wave data points would not significantly provide more information to the model because, as it was demonstrated in Fig. 6 and Table 3, the selected wave parameters present similar values and behaviour for both locations. Only two points were chosen to simplify the model and decrease the number of input variables and the computational cost. However, areas with higher wave data variability would need more points, and the adequate number of

Table 3
Mean value and std of the wave variables provided by the IBI reanalysis product at points P1 and P2.

Variable	P1		P2	
	Mean	Std	Mean	Std
HM0 (m)	1.81	0.97	2.08	1.06
WP (s)	10.9	2.7	10.7	2.7
WD (degree)	298.4	15.7	302.2	19.3

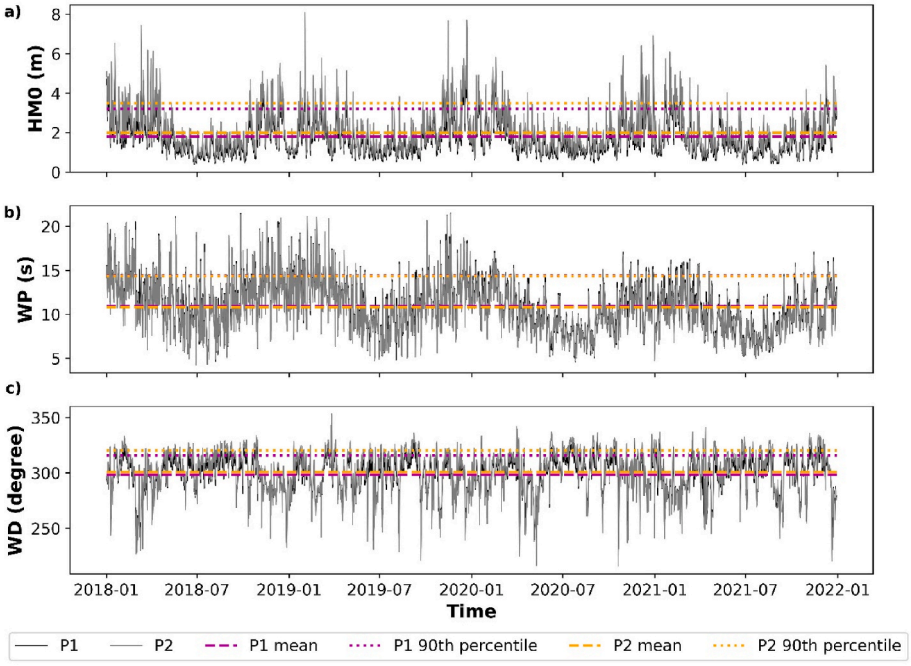


Fig. 6. IBI wave reanalysis data for the study period. a) Significant wave height; b) Wave period; c) Wave direction.

points for each study should be carefully selected to provide reliable data to the model and thus forecast the morphodynamic variables accurately.

The resulting wave data time series were then combined with the profile and the sediment size data. Particularly, the data model was fed with the 90th percentile of HM0 and WP and the mean WD, which were computed considering the time intervals between each survey. Hence, the wave variables summarized the hydrodynamic conditions during the interval between observations.

The 90th percentile was adopted because wave energy is one of the most important parameters for sediment transport and beach morphodynamics, which is directly associated with wave height and period (Jackson and Short, 2020). Therefore, if the mean HM0 value was considered instead of the 90th percentile, the effect of higher waves could be underestimated.

The input variables should have an adequate frequency according to the morphological output variable timescale to be predicted. For daily morphological changes, the water level variation is important for beach profile dynamics, especially for analysis of dynamics of morphological patterns that are present on the beach face and that are linked to cycles of spring and neap tides (Gomes et al., 2018; Jackson and Short, 2020). This is also the case for extreme events that could have a significant impact depending on the storm surge water level. However, for the seasonal timescale considered in this study, the tide effects were assumed to be less important than the wave action variables that were considered, with average values for the forecast period. This way, the model's simplicity was privileged since considering water levels will increase the model complexity.

Besides, the ocean water level variation due to tide and storm surge was not considered because these variables should present similar values along the study area. The seasonal frequency of the profile measurements did not capture tidal influence on the beach profile dynamics. In addition, profiles were not performed just before and after a storm event where its associated storm surge can have a strong impact. Hence, only wave action data were considered as the main drivers for the seasonal morphodynamic processes under simulation in this study. It is important to notice that tide and storm surge variability are key variables when modelling short-term morphological changes, considering hourly to daily timescales.

2.5. Random forest algorithm

The RF algorithm is an ensemble learning method that combines the results of multiple base models to reduce the results error. RF has the advantage of delivering satisfactory results even with few and no normalized data. This is an advantage for the data used in this work, which come from multiple sources. Moreover, RF uses a Decision Tree (DT) as a base model, an algorithm suitable for classification and regression problems. DTs were vastly used for different classification problems, for instance, to identify the category of ships according to their radar profile and to recognise numerical digits (Breiman et al., 2017). For the water resources scientific area, DTs were applied to estimate the flooding probability (Chen et al., 2008) and to predict the wave runup on vertical piles (Kazeminezhad and Etemad-Shahidi, 2015).

However, DTs models are easily hyper-fitted, which means they can present an excellent performance in the training dataset but have inferior performance when using different data. RFs can overcome this problem by combining the results of many DTs, which reduces the individual bias of each model and, consequently, the overfitting probability. Despite that, using RFs models does not guarantee that the result of the ensemble will be better than the result of the best individual model. However, it was demonstrated that model ensembles present smaller bias and variance than the individual solutions, thus improving the final prediction's accuracy, reliability, and consistency (Iglesias et al., 2022). Thus, the use of RFs models will decrease the likelihood of choosing a model with deficient performance (Cutler et al., 2012).

The RF models were implemented in Python version 3.10.9 using the open-source ML library Scikit-learn version 1.2.1 (Pedregosa et al., 2011) and the *RandomForestRegressor* function. The hyperparameter values were defined in an iterative process, which consisted of training the model and evaluating its performance. The model training consists of using the *fit(.)* function, which automatically adjusts the weights of the model to minimize the error between observed data and model output. During this process, the relevance of each input is determined, as well as the relevance of each DT to the output.

The models were trained and validated using the k-fold cross-validation method, which divided the dataset into five consecutive groups. This method uses one fold to evaluate the model performance and k-1 folds to train it. The process involves repeatedly refining the hyperparameters of the model and evaluating the results until a suitable level of performance is achieved. The data before January 21st, 2021, were used to train the model, and the remaining data were used for testing, resulting in ten to eleven profiles in the training dataset and one in the test dataset. The number of profiles considered depended on the location. At Estela, Moledo, and Paramos beaches, eleven profiles were considered in total, whereas twelve were considered for the other locations. The performance of the model was measured by the root mean squared error (RMSE), the mean absolute error (MAE), the coefficient of determination (R2), and relative error (RE) indicated by equations (1)–(4) respectively:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right| \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y)^2} \tag{3}$$

$$\mathrm{RE} = \frac{|y_i - \widehat{y}_i|}{y_i} \tag{4}$$

where n is the number of observations, y_i is the observed value, $\widehat{y}_i$ is the predicted value and y is the mean of the observed values. The maximum error was also determined for each validation fold.

Additionally, two different models were implemented with the same specifications but different input variables. The M1 model was fed with the beach profiles, the wave conditions, and the sediment size data. The M2 model considered only the waves and the profile data to assess the impact of including sediment size data on the profile evolution forecast. Fig. 1 presents the architecture of the RF model and the chosen hyperparameters.

The data used for training and testing were organised to enable the models to predict the profile at $t + 1$ using the elevation of the current profile at t. The input variables included the current profile's x-y coordinates, elevation, number of days until the next campaign, HM0 and wave period 90th percentiles, mean wave direction (M2), and sediment characteristic diameters (D50 and D90 -M1). The statistics of the wave data were calculated between t and t+1.

The final values of the RF hyperparameters were 200 estimators, the maximum depth of the DTs was 80, the minimum sample split was 5, and the minimum sample leaf was 4. The function to measure the quality of the splits was the Friedman mean squared error, and the maximum number of features to consider during each split used the *sqrt* criteria. These values were determined iteratively by observing how each hyperparameter affects the model's error. The models were trained, and a performance metric baseline was determined. After that, the value of a

single hyperparameter was changed, maintaining the values of the others, and its impact on the error was determined. When the minimal error for that configuration was determined, a second hyperparameter was assessed, following the same steps. The other hyperparameters used the default value. After training, the models' performance was evaluated using previously unseen data relative to the period after February 2021.

Although the sediment values were assumed to be constant at each location, the sediment data helped the model to differentiate each profile, considering that it was trained using all beach profiles simultaneously. This way, during the training phase, the model can learn how to associate the sediment size with each location and how the size values affect the morphological changes locally. More details on how the decision trees consider each input can be found in the literature (Cutler et al., 2012; Irizarry, 2019; Kubat, 2017). Ideally, a time series of the evolution of the beach profile granulometry at each location could also be used to achieve higher accuracy in modelling the morphological evolution of the cross-shore sections. Still, a simultaneous collection of sediment samples with the beach profile measurements is required.

2.6. Computation of beach slope, width and retreat/advance

After predicting the beach profiles with the data models, the resulting beach slope was computed to assess the differences between the observed and simulated slopes. The slope was determined considering the distance between the points with elevations of 0 m and 3 m (MSL). Moreover, these elevations were used as a reference to determine the beach width in each profile. This interval was selected according to the wave height values in the region, focusing on the elevations where the wave action is most relevant to the morphological change. The choice of reference heights directly influences the slope error. Nevertheless, these values are adequate for the profiles studied.

The profile lengths were divided into 5 m segments. At each segment, the slope was computed considering the elevation of each stretch's initial and final points. The elevation values were interpolated using the *interp1d* function from the SciPy library (Virtanen et al., 2020). Then, the mean slope of each profile was computed considering the slope of each segment.

The retreat/advance of the beach profile during the last monitoring period was computed considering four elevation references (1 m, 2 m, 3 m, and 4 m, MSL). As the profile elevation results were not necessarily equal to the adopted reference values, these elevations were added to the results of the models, and the correspondent distance was linearly interpolated between the closest values around the reference elevation. With the interpolated distance values, the advance/retreat was determined by computing the difference between the distances of the most recent and the previous profile at each elevation. For instance, if the distance difference at 1 m height is positive, it indicates accretion, whereas negative values represent erosion.

2.7. Storm conditions simulation

The model's capacity to forecast the storm's effects was also qualitatively evaluated. Two different scenarios were considered. The first scenario adopted an HM0 of 4 m and a WP of 15 s. It is the same scenario that Vieira et al. (2020) defined to study the performance of coastal protection structures on the Portuguese coast. The second scenario increased the HM0 to 8 m, maintaining the same value for the WP. The WD was the same for both scenarios, ranging between 286 and 305, according to the profile location.

The main objective of the proposed simulations is to produce insights about the model response using HM0 values in the range of the training data (~4 m) and additionally demonstrate the capacity of the model in predicting the morphological evolution considering wave conditions outside the training data range (the maximum HM0 was slightly under 8 m). Considering that RF is a multi-dimensional space, it is necessary to use additional techniques to inspect if they are considering each input variable adequately. Through this simulation, it is possible to evaluate if the model behaves as expected during a simulation of a storm event that was not covered in the training data, even if the available data is not adequate for short-term extreme events simulations, as explained before in section 2.4.

3. Results and discussion

In this section, the training performance of the models is presented, followed by an assessment of their test results. After, the results of the data models' computation of the beach slope, their retreat/advance, and simulation of a storm event are discussed.

3.1. Training performance

The results of the model training using the k-fold cross-validation method are presented in Table 4. M1 performed satisfactorily across all metrics, with practically equal R^2 values in the five considered folds. The RMSE and MAE also indicate that the model results had low errors. However, the maximum error range was from 0.96 m to 1.65 m, with a 1.33 m mean, which reveals that the implemented model may not be suitable for all the considered locations used to train the model.

M2 followed the M1 trend in RMSE, MAE, and R^2. Nonetheless, M2 maximum error was higher than M1, ranging from 1.34 m to 1.79 m, with a mean of 1.52 m. Hence, M1 performed better due to the lower maximum error. However, the other metrics did not present significant differences, which allowed us to conclude that the maximum error was associated with a few points in the profile data, as will be further discussed in the next section.

Both models performed highly, with similar R^2, RMSE, and MAE in all folds. However, the M1 maximum error was lower than that obtained with M2, with a mean difference of 20 cm. This result indicates that the sediment size data was useful in reducing the maximum error of the model in the training dataset. However, it did not lead to significant improvements in other metrics. Additionally, the models' performance demonstrated their accuracy regardless of the fold used for evaluation. The models have good generalising capacity and are not biased; consequently, they are reliable for forecasting using previously unseen data.

3.2. Prediction performance

The model performance for forecasting the profiles of the last field campaign is presented in Fig. 7. The best performances occurred in Amorosa and Neiva P1 beaches, with RMSE varying from 0.18 m to 0.26 m for both considered models, M1 and M2. The worst performance occurred in Moledo, with a RMSE higher than 1.35 m. Additionally, M1 had a better performance than M2 in eight of the 12 locations considered, namely at Âncora, Amorosa, Neiva P1 and P2, Ofir P3, Estela, Árvore P2 and Paramos, but M1 underperformed M2 in Moledo, Ofir P1, and Árvore P1. Both models presented equivalent results for the Ofir P2 profile, and the most significant improvement occurred in Neiva P2, in which M1 RMSE was 0.17 m and M2 RMSE was 0.49 m. The average M1 RMSE for all the considered profiles was 0.51 m, and the average M2

Table 4
Training results of the M1 and M2 models.

Model		Fold					
	Metric	1	2	3	4	5	Mean
M1	**R^2**	0.985	0.987	0.986	0.986	0.985	0.986
	Max error (m)	1.19	0.96	1.52	1.65	1.33	1.33
	RMSE (m)	0.04	0.04	0.04	0.04	0.04	0.04
	MAE (m)	0.15	0.15	0.148	0.14	0.15	0.15
M2	**R^2**	0.986	0.987	0.987	0.986	0.986	0.986
	Max error (m)	1.42	1.34	1.79	1.67	1.39	1.52
	RMSE (m)	0.04	0.04	0.04	0.04	0.04	0.04
	MAE (m)	0.14	0.14	0.14	0.14	0.14	0.14

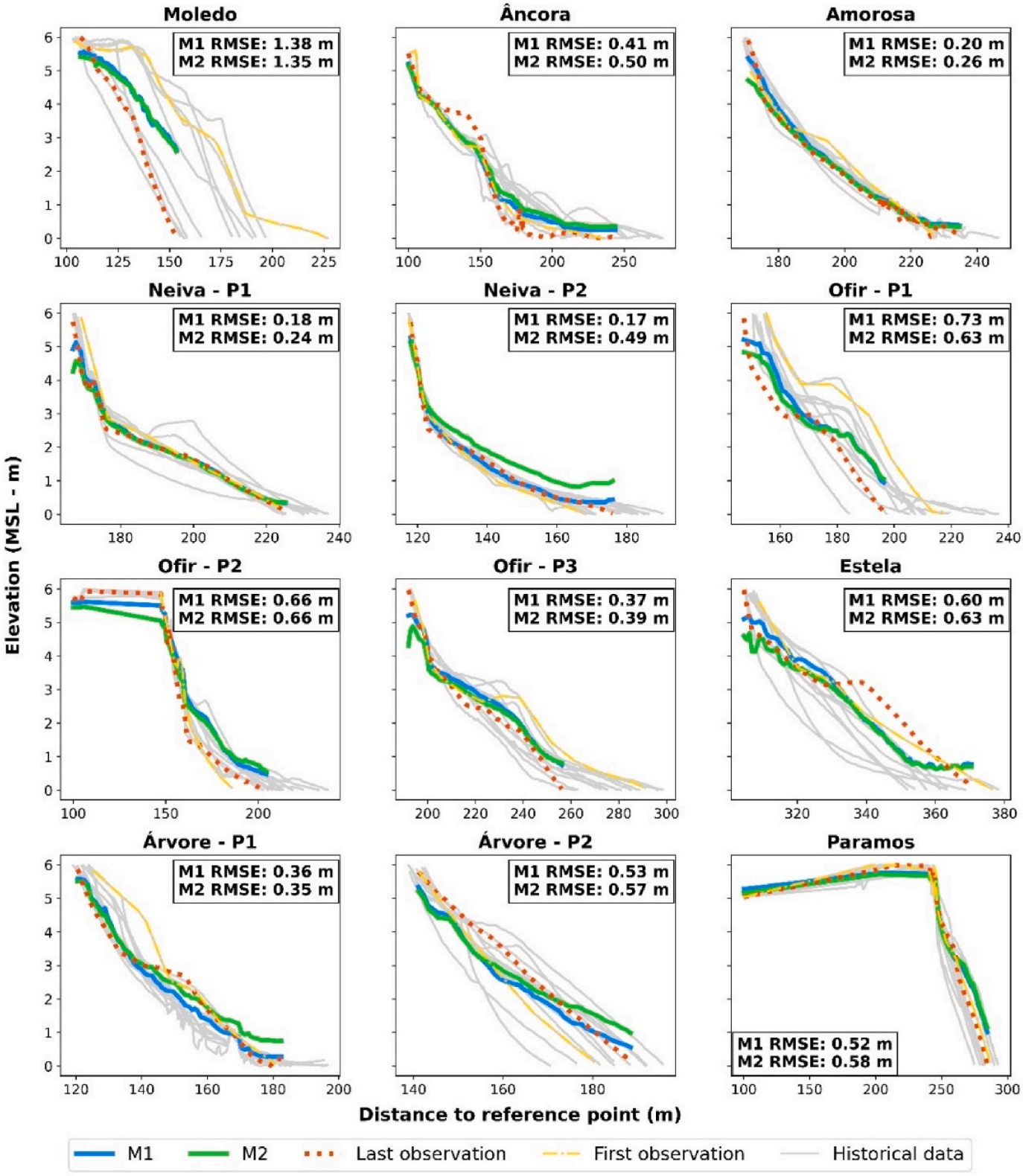


Fig. 7. Comparison between the results predicted by M1 (blue) and M2 (green) models and the observed values (red dots). The historical data used to train the models are represented in yellow (first observation) and grey.

RMSE was 0.55 m.

Additionally, both models achieved good results despite the temporal variability of the profiles. It highlights that the algorithm correctly reproduced the results in dynamic beaches, and not only in beaches with more stable trends. It can be observed that the M1 forecast at Amorosa, Neiva P1, and Neiva P2 overlapped with the observed data, having slight deviations. The results in Neiva P1 must also be emphasised due to higher variability in the historical data, demonstrating that the model prediction capacity is not limited to less dynamic areas.

In the case of the Moledo profile, it must be noted that it is located

very close to the Minho River estuary mouth, an important sediment source to the coastal platform that was not considered as input data. The narrow mouth of this estuary increases the current velocity, intensifying sediment transport from the estuary to the ocean (Melo et al., 2020). Therefore, the river flow may be a relevant variable that was not included in this study, and the AI models did not have enough information to forecast the profile evolution accurately in this location. Quantifying the contribution of fluvial inflows in the area may improve the performance in the locations close to estuary outlets.

Moreover, there is no clear pattern between the beaches in which the models performed well or not. The RMSE achieved in sandy beaches (Âncora, Ofir, and Paramos) ranged from 0.37 m to 0.73 m, whereas in sandy beaches with rocky outcrops, the RMSE varied from 0.17 m to 0.63 m (excluding Moledo). Hence, the lowest errors were obtained for sandy beaches, but the highest errors were not significantly different. In addition, proximity to the wave data points, geographical location, and sediment size did not affect the model's performance. These influences can be rejected when observing the performance of the models in simulating profiles at the same beaches (Neiva, Ofir, and Árvore), in which error had considerable local variability.

Despite that, it is possible to use explainable AI (XAI) frameworks that help to interpret the relationship between inputs and outputs. This concept aims to make the models more intelligible to humans (Gunning et al., 2019). Adopting such frameworks could enhance the reliability of AI models, allowing physical interpretations of the results and increasing their acceptance by domain experts. For instance, it would be possible to identify how the model associates each profile to each wave data point and which variables are the most relevant to the output. However, this study is not focused on that component of AI methodologies.

Furthermore, it was identified that the coast stretch type slightly influenced the M1 model error and had an effect on the averaged RMSE. The lowest errors were obtained in profiles measured at sandy beaches with rocky outcrops (Neiva P1, Neiva P2, and Amorosa). The average RMSE from the beach profiles located in sandy stretches with rocky outcrops was 0.48 m, whilst the average RMSE in the profiles in sandy stretches was 0.54 m. Regarding M2 results, an average RMSE of 0.55 m was obtained for sandy stretches with rocky outcrops and 0.56 m for sandy stretches. This result also highlights that the consideration of sediment size data on the models improved their performance. The accuracy of M1 increased mainly at the beginning of the profiles, as observed in the results of Amorosa, Ofir P1, Ofir P2, and Estela.

The capacity of the data models to forecast the evolution of beach profiles was compared with results obtained in previous studies. Using a diffusion model, Karunarathna et al. (2012) achieved an MAE range between 0.1 m and 0.5 m to predict profile changes, and Otto et al. (2021) used a spatiotemporal-functional approach to forecast the beach profile evolution, achieving RMSE ranging between 0.25 m and 0.9 m. The average MAE in the test dataset obtained in this study was 0.45 m for M1 and 0.49 m for M2 and, excluding Moledo, which is clearly an outlier, the MAE range is between 0.12 m and 0.62 m for M1 and between 0.14 m and 0.63 m for M2 (Table 5). Therefore, the values obtained are in the same range as those obtained by previous studies (Karunarathna et al., 2012; Otto et al., 2021).

The averaged RMSE for M1 was 0.54 m and for M2 0.58 m, and, excluding Moledo again, the RMSE range for M1 was from 0.17 m to 0.66 m and for M2 from 0.26 m to 0.66 m. Lee et al. (2022) implemented a numerical model to simulate long-term changes in beach profiles, achieving an RMSE of 3.7 m, a clearly higher value than the ones obtained in this study. It can be concluded that the errors reached in this study are similar or even lower to those obtained by other authors using different techniques, reinforcing the utility of the data models implemented here.

To further investigate the models' capacity to predict the observed data, the error distribution was assessed by computing the difference between the observed data and the models' results (Fig. 8). The error distribution plot is computed by determining the kernel density estimate plot to visualise the data distribution. This figure was done using the *kdeplot* function available in the Seaborn library (Waskom, 2021), which produces a continuous density estimation of the analysed variable using a Gaussian kernel. This methodology is suitable for comparing the models' errors. Density plots exhibit how frequent the values are in the analysed dataset.

Table 5
MAE and RMSE in the test dataset.

Profile	MAE (m)		RMSE (m)	
	M1	M2	M1	M2
Moledo	1.11	1.08	1.38	1.35
Âncora	0.35	0.45	0.41	0.50
Amorosa	0.16	0.14	0.20	0.26
Neiva - P1	0.11	0.13	0.18	0.24
Neiva - P2	0.12	0.54	0.17	0.49
Ofir - P1	0.62	0.55	0.73	0.63
Ofir - P2	0.58	0.63	0.66	0.66
Ofir - P3	0.31	0.32	0.37	0.39
Estela	0.50	0.49	0.60	0.63
Árvore - P1	0.30	0.29	0.36	0.35
Árvore - P2	0.46	0.51	0.53	0.57
Paramos	0.42	0.48	0.52	0.58
Average	0.46	0.50	0.55	0.59

These results were analysed by considering a global performance (all locations) and a local performance (assuming only the results at each location). Both models had a symmetric global error distribution centred near zero, indicating that most errors were near zero. However, the M1 density plot distribution indicates that M1 results had lower variance when compared to M2.

The values in the extremities of the plot indicate the highest and lowest error values obtained. Both models' maximum overestimations (negative values) were close to −2.7 m, and underestimations were close to 1 m for M1 and 2 m for M2. However, the low density of errors indicates that these values are uncommon. The global errors considering a 5% confidence interval were computed, removing the values in the extremities with frequency lower than 2.5%.

When analysing the local error distribution, a similar behaviour between both models can be identified, with a narrower tendency in the M1 results and more dispersion for M2. The highest error occurred in Ofir P2, which appeared in the profile's first 50 m (Fig. 7). Furthermore, both models had a normal error distribution, with the majority of values between −1 m and 1 m, and a mean of approximately 0 m. Both models best result was in Estela, in which the error distribution is a peaked curve centred near 0, indicating a good agreement between models and data.

Therefore, both models performed satisfactorily, achieving acceptable errors and low variance. The inclusion of sediment size data produced better results, indicating that it was an important variable for beach profile forecasting with data models. Moreover, the sediment size data helped to reduce the error in most beaches, regardless of the sand beach granulometric distribution.

The higher error observed in the forecast of sandy beach profiles can be explained due to their more dynamic behaviour. The rocky outcrops increase the stability of the coastal stretch, being more resilient to meteo-oceanographic conditions than the sandy stretches. The beach profile surveys were, perhaps, not long enough to capture significant modifications in these areas, whereas sandy beaches can experience significant changes in shorter periods. This could explain why M1 results in stretches with rocks were better. The model could learn its behaviour with the available data as the pattern changes slower. At the same time, a longer time series may be necessary to improve the model performance in more dynamic areas. However, the model learned the profile evolution of less stable beaches, as can be observed for Neiva P1 results.

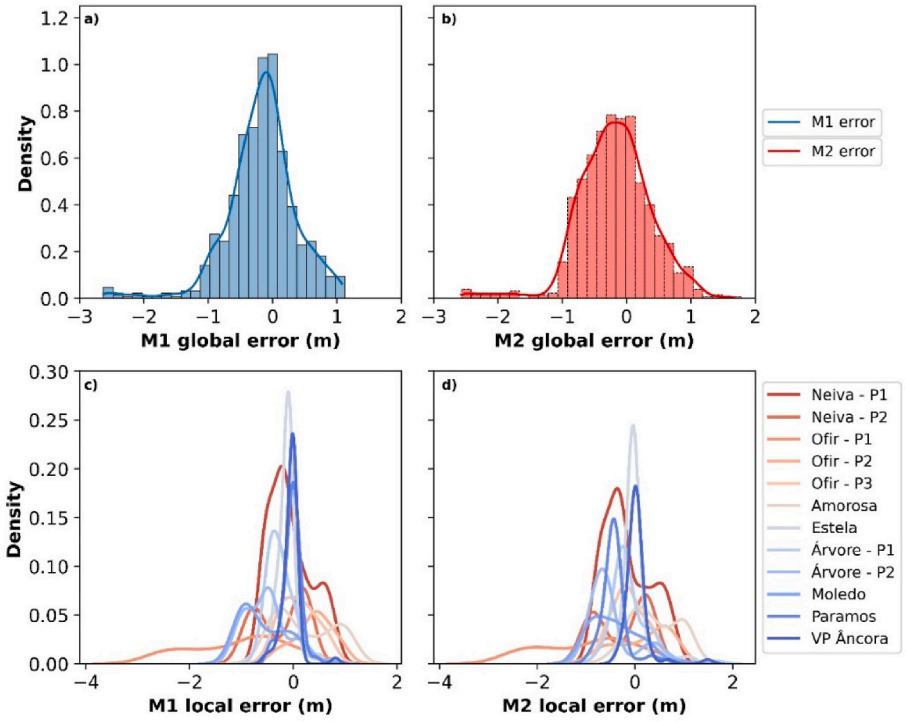


Fig. 8. Error distribution of the obtained results. a) M1 global error distribution; b) M2 global error distribution; c) M1 local error distribution; d) M2 local error distribution.

3.3. Beach slope, width and retreat/advance computation

Beach slopes were computed to assess whether the data models could reproduce this profile feature. The slope and width results are depicted in Fig. 9. As expected, the models' performance aligns with the beach profile's forecast performance. The best results were obtained in Amorosa, Neiva P1, and Neiva P2, while the worst was in Moledo. It must be noted that the stretches shown in this figure correspond to the beach faces of the most recent profiles presented in Fig. 7. For instance, in the Estela profile, the beach face corresponds to the stretch in which the models had lower performance when compared to the beach berm. These results clearly influence the slope error at this location. Despite that, the width error of both models was only 10 m.

The absolute error was computed to evaluate the slope error, while the relative error was utilized to assess the width error when comparing the models' performance in relation to beach slope and width (Table 6). Both models results were acceptable in reproducing the beach slope, with an absolute error lower than 2.4%. These results also showed that M1 overcame M2. Seven beaches had a slope error lower than 1% for M1 results, while M2 had five beaches within this error range. Additionally, M2 outperformed M1 only at Amorosa.

The beach width errors exhibited similar behaviour. Both models achieved perfect results in Amorosa, Neiva P1, Neiva P2, and Árvore P1. The mean relative errors for M1 and M2 are 19% and 18%, respectively. The width results for both models were identical for all profiles, except at Ofir P3, where M2 outperformed M1.

The slope predictions with errors lower than 1% presented the same pattern as the local RMSE. The best performances considering both metrics were obtained in Amorosa, Neiva P1, and Neiva P2, indicating that the results of the data models can help in the computation of relevant beach variables, such as the slope. It is important to note that all beach slopes were estimated considering the slope between 0 m and 3 m elevations, which may not be adequate for all considered beaches. However, interpolating the elevation every 5 m along the profile length and calculating the mean slope smooths the influence of the chosen elevation interval. This simplifies the analysis of the profile slope and focuses on the general performance of the model. When applying this methodology to a specific location, the best elevations should be carefully chosen, ensuring that the slope between the highest and lowest points is linear.

Finally, the profiles retreat/advance considering different reference elevations also demonstrated that M1 performs better than M2 (Fig. 10). Positive values represent an advance toward the ocean, while negative values represent a retreat. The best results occurred in Neiva P1, although the beach advance magnitudes at most profiles differ significantly. From a qualitative point of view (advance or retreat), both models correctly forecasted the behaviour of nine profiles at 2 m elevation and seven profiles at 3 m and 4 m height. The missing bars at some profiles at 1 m elevation indicate that the models did not have values at that height.

The higher errors at 1 m elevation can be related to poor accuracy of the model for that elevation, even for profiles with low RMSE and MAE. The performance of the models increased at the other elevations, probably because the lower profile section was more dynamic than the others, requiring more data for better forecasting by the RF model. Therefore, the advance and retreat results were not accurate enough if it were necessary to know the exact profile dynamics. Still, both models can indicate the behaviour of the profile. In addition, the advance/retreat results followed the same accretion/erosion behaviour observed in Fig. 5. Considering the results in which both models forecasted the same pattern of the observed trend, there are nine eroded and sixteen accreted profiles, representing 36% and 64%, respectively. There is a predominance of accretion over erosion in both model results.

These findings suggest that both models were effective in capturing significant features from the simulated profiles, particularly between 1 and 4 m in height. Both beach slope, width, and profile advance/retreat were effectively replicated on the majority of the beaches under consideration. As anticipated, the highest errors were observed in areas where the profile simulation also yielded higher errors.

In order to compare the behaviour of the profiles, the seasonal advance/retreat rate of the beach profiles was determined using the data from COSMO program (Fig. 11). There is no identifiable

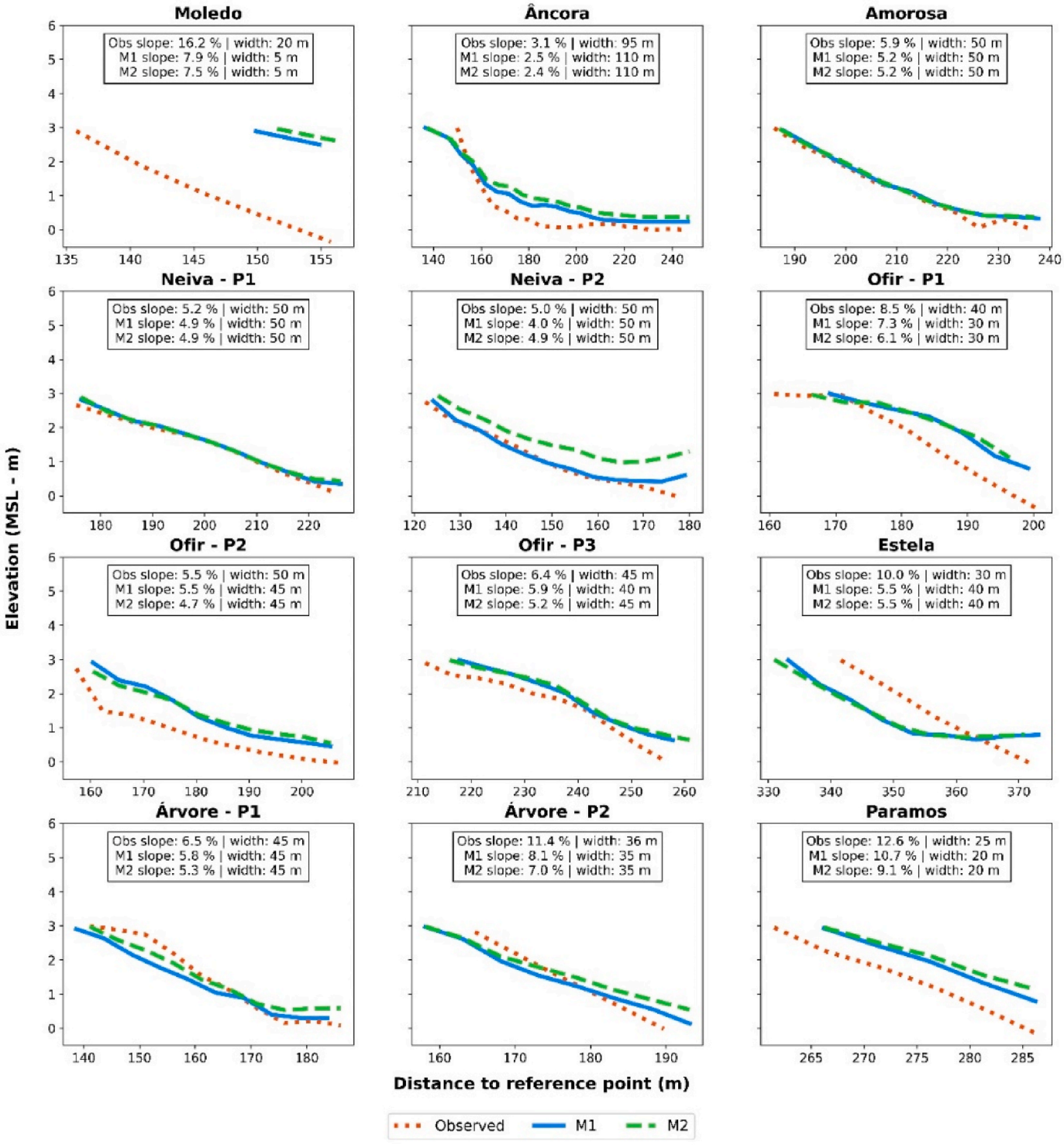


Fig. 9. Beach slope computation results. The red dotted line represents the observed data, whilst the blue and green lines represent the results of M1 and M2, respectively. The slope results are presented in the white boxes at each subplot.

morphodynamic behaviour considering all beaches. For instance, Amorosa presented a stable profile during all periods, whilst Moledo presented severe erosion after winters. At Árvore P2, the data showed erosion after the spring-summer surveys and accretion after winter. These results demonstrate that the RF model reproduced this variable with acceptable accuracy despite the absence of a clear pattern considering all beaches.

Although the quantitative error of the model is still considerable for some of the assessed beaches (e.g., Paramos and Estela), the qualitative prediction (advance/retreat) was accurate, indicating that the model can be reliable in indicating the trend. Besides, the presented results show a stable trend, except for Paramos. There is a balance between eroded and accreted profiles, with low profile changes. However, it must be stressed that the profile's advance/retreat was computed considering the time interval between the last two field surveys (approximately four months), which does not show the long-term trend of the profiles, only

Table 6
Mean beach slope and width results between elevations of 0 and 3 m.

Beach	Slope (%)					Width (m)				
	Obs	M1	M2	M1 AE	M2 AE	Obs	M1	M2	M1 RE (%)	M2 RE (%)
Moledo	16.2%	7.9%	7.5%	8.4%	8.7%	20	5	5	75%	75%
Âncora	3.1%	2.5%	2.4%	0.6%	0.8%	95	110	110	16%	16%
Amorosa	5.9%	5.2%	5.2%	0.8%	0.7%	50	50	50	0%	0%
Neiva - P1	5.2%	4.9%	4.9%	0.2%	0.3%	50	50	50	0%	0%
Neiva - P2	5.2%	4.9%	4.9%	0.2%	0.3%	50	50	50	0%	0%
Ofir - P1	8.5%	7.3%	6.1%	1.2%	2.3%	40	30	30	25%	25%
Ofir - P2	5.5%	5.5%	4.7%	0.0%	0.8%	50	45	45	10%	10%
Ofir - P3	6.4%	5.9%	5.2%	0.5%	1.2%	45	40	45	11%	0%
Estela	10.0%	5.5%	5.5%	4.5%	4.5%	30	40	40	33%	33%
Árvore - P1	6.5%	5.8%	5.3%	0.6%	1.1%	45	45	45	0%	0%
Árvore - P2	11.4%	8.1%	7.0%	3.3%	4.4%	25	35	35	40%	40%
Paramos	12.6%	10.7%	9.1%	1.9%	3.5%	25	20	20	20%	20%
Mean				1.9%	2.4%				19%	18%

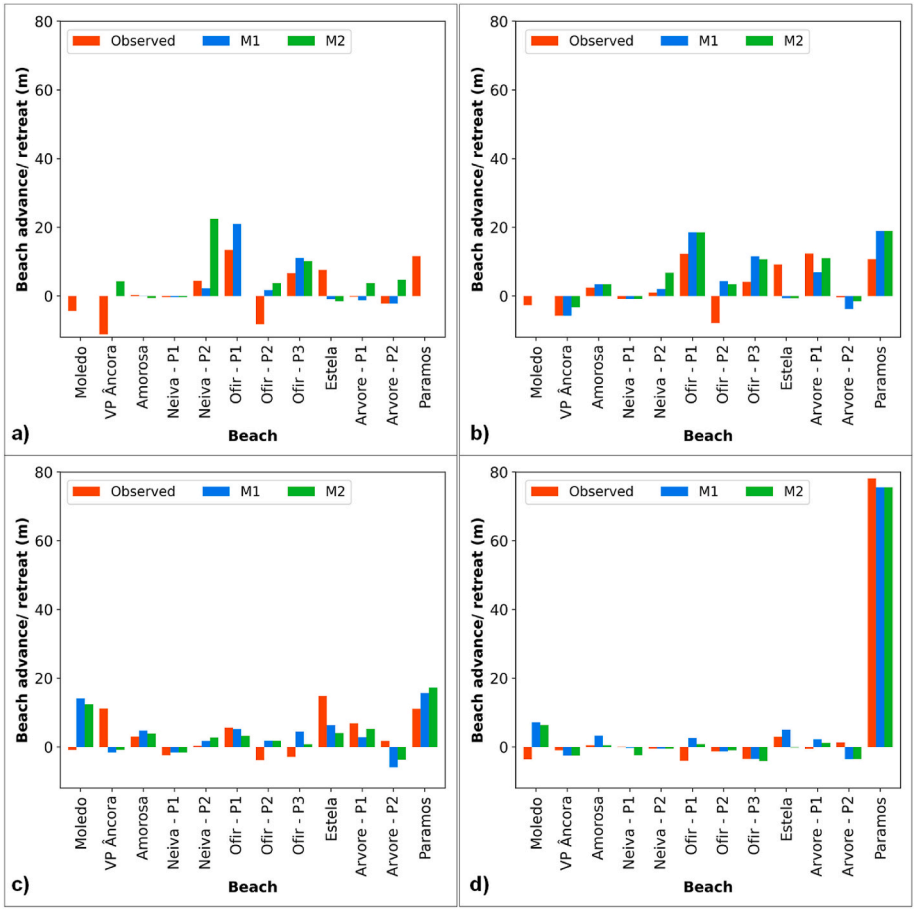


Fig. 10. Profile advance/retreat results. a) 1 m height; b) 2 m height; c) 3 m height; d) 4 m height.

seasonal trends. The same occurred when analysing the beach slope results, though the quantitative errors were lower.

Hence, the models results were useful to estimate the seasonal morphological evolution of profile features, achieving satisfactory error metrics. In the case of the profile advance/retreat, a classification model would have achieved excellent metrics, considering that the model correctly identified most of the accretion/erosion patterns.

3.4. Storm simulation scenario

The models' performance under storm conditions is presented in Fig. 12. The "_4″ and "_8″ refer to the simulation results with HM0 equal to 4 m and 8 m respectively. The previous profile line is the profile state before the storm simulation and serves as a baseline to assess the storm effect. As expected, increasing the HM0 value resulted in sharper profile changes when comparing the same model. It is also worth noting that M2 results simulated more intense changes compared to M1, as can be seen in Neiva P2 and Estela. Although the results were relatively close, the only scenario in which M1 forecasted more severe changes was in Ofir P1 and Árvore P2.

The model simulated the storm results differently at each location. At Moledo, Âncora, Neiva P2, Ofir P1, Ofir P3, and Árvore P1, the results showed an accretion predominance, while at the other profiles, erosion patterns prevailed. To evaluate these results, it would be necessary to

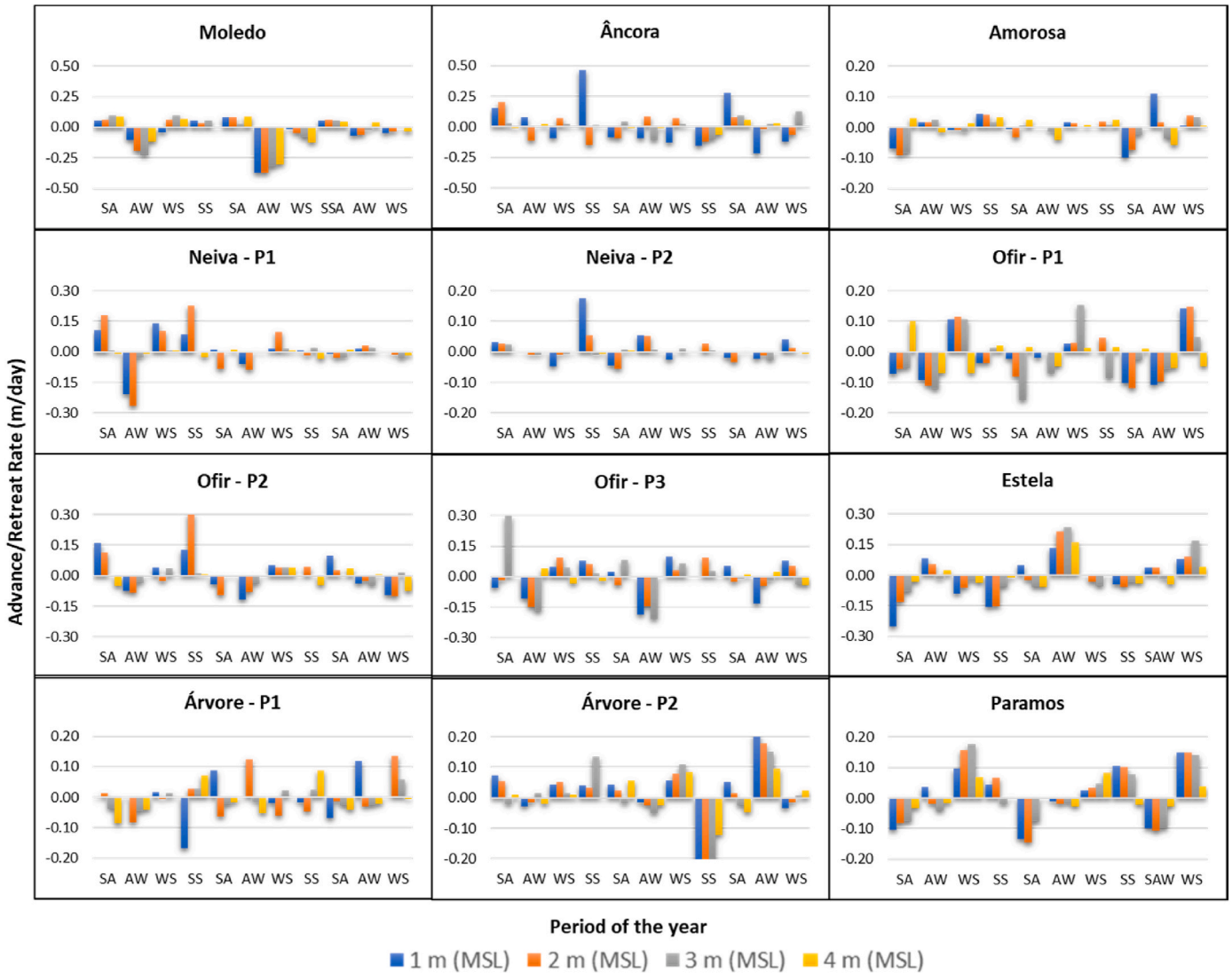


Fig. 11. Historical erosion/advance rates at the measured profiles. The capital letters represent the seasons in which the rates were computed, with A referring to autumn, W to winter, and S to summer and spring.

use process-based models to determine if the outputs are realistic.

In addition, the result of the 8 m HM0 simulation is proof of the robustness of data models. Although the HM0 value was never that high in the training dataset, the results are according to what was expected, showing sharper profile changes regardless of the model. However, it must be highlighted that when a model simulates conditions outside the training/calibration range, its uncertainties increase, and its performance diminishes. Hence, the results obtained in this section should be carefully considered. For instance, the model simulation resulted in more accreted profiles than eroded, which is not expected after storm events. However, considering the historical data, the simulation occurred between winter and spring, a period with accretion predominance (Fig. 5). Hence, the models reproduced the same morphological patterns for these seasons.

4. Discussion and conclusions

This section discusses the main challenges of this study, considering the use and processing of data from multiple sources and with different formats. Subsequentially, the limitations of this study's methodology are reviewed, and the main conclusions are presented.

4.1. Data models and the importance of open data

The latest advances in AI techniques have transformed data into one of the most valuable but still underused products. Relatively simple models can not only process different types of data but also discover patterns that may result in significant insights for the progress of beach morphodynamics understanding, and consequently, for the management of coastal areas. Hence, local monitoring programs can efficiently gather relevant information from any area of interest. However, there are still some barriers that must be overcome to implement this kind of solution.

Although global monitoring programs such as Copernicus provide essential data, they cannot accurately represent local features and dynamics due to the temporal and spatial resolution of the data. Local monitoring programs have the advantage of knowing the current situation of a specific study area, allowing the identification of critical points and key data to be observed/measured and also the needed timestep for those observations.

In this context, monitoring programs must provide open access to their data to allow the implementation of more robust data models that can merge relevant types of information. The models developed in this study combined numerical model reanalysis results from the Copernicus

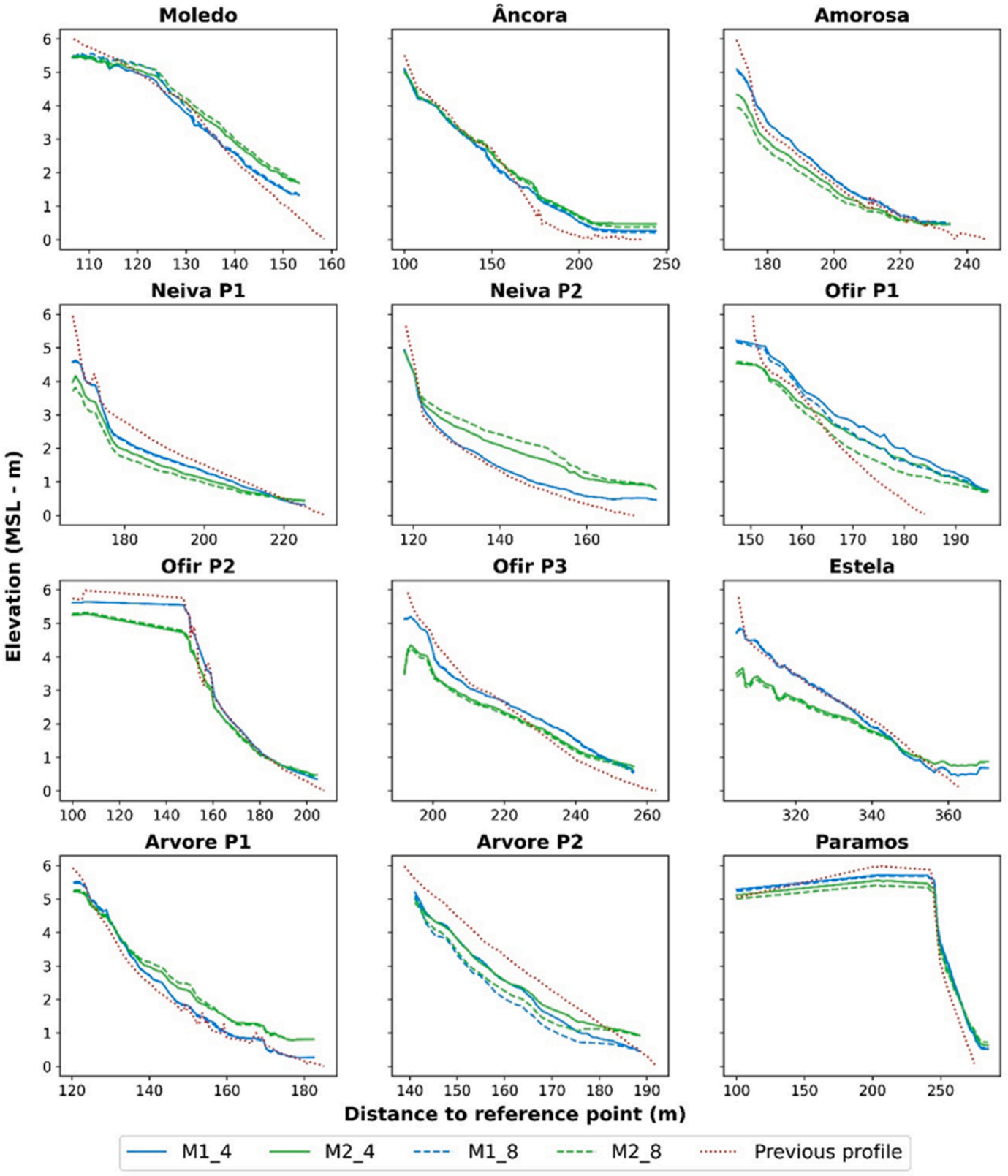


Fig. 12. Storm simulation results for HM0 are equal to 4 m (_4) and to 8 m (_8).

Marine Service, coastline profiles from the COSMO program, and sediment size data from the MarRisk project. However, one of the challenges of this study was to organise the data of the three sources according to date and location. Each one presents different time horizons and uses different data formats, which must be pre-processed to adequately implement a data model.

Nevertheless, it must be stressed that data models can integrate multidisciplinary data, making it possible to consider water quality, chemical, biological, urban, and anthropogenic data jointly with physical variables like the ones considered in this study. Interdisciplinary

approaches are essential to promote more integrated coastal management (de Schipper et al., 2020). It is just a question of determining which knowledge areas influence the phenomena under scrutiny and how each can collaborate to solve the problem from a more holistic point of view.

Therefore, creating open-access systems that concentrate all data types would be important in implementing transversal data models. This would allow any stakeholder to manipulate these data according to their necessities, goals, and expertise, which could be extremely relevant to those entities that cannot afford to invest in monitoring systems or field surveys. Moreover, part of the necessary work to implement a data model could be reduced by just adopting a pattern in the data types, and centralizing the data would facilitate its standardization, thereby reducing redundant costs. For example, under the COSMO and MarRisk projects, beach profile data were measured on the Northern Portugal coastline. Some beaches were surveyed only in one of the projects, others in both, and others in none.

Finally, it would be pertinent for national and international agencies to take the initiative to create a central multidisciplinary database. It would ensure that all the information could be accessed through a trustworthy agent, reducing the time required to find the desired data if it exists. The Copernicus Marine Service is an excellent example of this kind of platform, providing atmospheric, marine, land, and climate data. However, it is still not integrated into national databases. While it would be challenging to incorporate such efforts in an international platform, national agencies could start organizing their data to facilitate integrating national and international data sources for future projects, which would be extremely useful for implementing multidisciplinary data models.

4.2. Coastal processes and data models

This study combined beach profile survey data, sediment characteristics, and wave variables to implement an RF model capable of forecasting the morphological evolution of beach profiles. The achieved results are promising and demonstrate the capacity of data models to reproduce complex morphological patterns. A single model learned to differentiate the patterns at each location and forecasted consistent results using data from the profiles.

One of the main limitations of ML models is interpretability. Whilst drawing conclusions from numerical models is easy, they may underestimate some physical processes. For example, considering a beach with vegetation at the base of the dunes, the morphological evolution will be influenced by the main drivers, such as wave and wind actions, and the presence of vegetation. If the modeler ignores these features, uncertainties will be associated with the model. This problem can be solved by increasing the details and resolution of the models, but it also greatly increases the modelling efforts. For instance, the same can be considered for the granulometry of beach sediments. It varies for each profile, and sometimes, even within the same profile, simplifications are adopted to make the models feasible.

Data models have the advantage of intrinsically considering these features, as the influence of vegetation and sediment size would be reflected in the field data. This contribution does not need to be explicit, as is the case with numerical modelling methodologies. Whereas numerical models can easily extrapolate physical conditions, data models are limited to the training variables. If no data are acquired during an extreme event, the data model will be inappropriate for simulating such conditions.

Furthermore, XAI can solve the interpretability barrier of data models. It consists of a series of methods and tests that determine the relevance of each input to the output, thereby helping to understand the limitations of such methodologies. Therefore, by adopting these concepts, it will be possible to compare process-based models with data models, increase knowledge about the limitations of each model, and allow the creation of adequate modelling solutions for each case. The use of XAI frameworks is strongly encouraged in future research.

4.3. Methodology limitations

The data models implemented in this study considered wave characteristics, sediment size, and beach profile variables. However, the morphological evolution of beach profiles is also influenced by other variables that were not considered. For instance, at higher elevations, the wind can be an important factor to be considered in sediment transport, mainly for finner sediment beaches.

Additionally, as previously explained, one of the sources of the errors obtained in Moledo could be the lack of consideration of the Minho River discharge. Conversely, fluvial sediment transport's importance could also explain some of the errors in other beach profiles. One of the best results was in Neiva, which is only affected by the discharges of a small river. In northern Portugal, five main rivers can contribute to the fluvial sediment transport towards the ocean: the Minho River (influence Moledo and Âncora beaches), the Lima River (influence Amorosa beach), the Cávado River (influence Ofir beach), the Ave River and the Douro River (both affecting Árvore beach). Considering the river plumes' influence on the data models could reduce the errors mainly at the lower profile elevations.

Unfortunately, most river flow dischargers were unavailable for the period selected in this study and could not be incorporated into the implemented models. Depending on the data availability, two possible solutions can be considered to overcome this constraint. The first is to access data from environmental agencies that monitor the discharges of rivers, which directly feed this data to the data models. The second is the use of hydrological models to simulate these data based on historical precipitation records.

Furthermore, the COSMO data covered the morphological evolution of the profiles for approximately five years, which is a short period for long-term morphological process characterization. Hence, the models implemented in this study may not be adequate for forecasting long-term profile changes, only seasonal profiles. The data captured some trends in the dynamics of the profiles, such as in Âncora, Estela, Árvore P2, and Ofir P1. However, both models achieved acceptable metrics at these locations. It is necessary to apply the models to a longer-measured profile time series to assess the capacity of the model to reproduce long-term morphodynamic patterns, but unfortunately, no data are available to accomplish this task.

Additionally, the methodology could be further developed to simulate short-term morphological processes, such as the morphological changes induced by extreme events. The role of water level variations during daily intervals are more evident than during longer periods. For example, the effects of a storm surge will be highlighted if the data time interval is adequate for capturing this event. However, if the time window is too long, storm-induced water level fluctuations effects on the beaches morphology may be obscured. To capture short-term beach morphodynamics, it is necessary to have either a longer data series of beach profile measurements that include extreme events or a selective data series that precisely registers the morphologies before and after extreme events.

Another limitation of this model is the lack of data about beach nourishment during the study period. Human interventions significantly modify the beach profiles, changing their natural evolution dynamic. Beach fill operations have been increasing in Portugal in the last decades. According to Pinto et al. (2020), the northern Portuguese coast had more than 30 interventions between 1957 and 2017. Therefore, data about structural interventions on beaches could also increase the performance of data models, but this data was not available to the authors. However, quantitative data regarding this type of intervention is usually restricted to those who execute and supervise it. The few available open data are mostly qualitative, for instance, the date of the intervention and the place, but not its exact location and the nourishment volumes, which would be the information needed to be included in the data models.

4.4. Conclusions

This study combined data from different sources to create a model capable of forecasting the evolution of multiple beach profiles considering the influence of wave and sand size. Two different models were implemented using the RF algorithm, in which M1 used wave, beach profiles, and sediment size variables as inputs, whereas M2 considered the same variables except for the sediment size. The output variable was the elevation relative to the mean sea level, which was compared with observational datasets to depict the accuracy of the data models.

The training results showed that both models had low error metrics, achieved similar values, and demonstrated the suitability of the data models for representing beach morphology. The average RMSE was 0.04 m, and the MAE was lower than 0.15 m for both models. These metrics indicated that the difference in the model input data was negligible. The type of profile did not influence the performance of the models.

However, when the forecasting capacity of the models was tested, M1 outperformed M2. Considering the models' global performance, the average RMSE and MAE were respectively 0.54 m and 0.45 m for M1, and 0.58 m and 0.49 m for M2. When assessing the local errors, the best result occurred in Neiva P1, where M1 had an RMSE of 0.18 m while M2 had 0.24 m, values similar to those in Amorosa (0.20 m and 0.26 m, respectively). Hence, it can be concluded that the sediment size data improved the results, reducing the average maximum error by 19 cm in the training dataset and the RMSE and the variance of M1 when compared to M2.

Additionally, the profile slope and retreat/advance were computed to assess the difference between the observed and simulated results. Both models accurately predicted the slope of the profiles, but M1 outperformed M2, with an MAE of 2.1%. M2 had an MAE of 2.4%. The profile retreat/advance results were inaccurate, but both models had a good qualitative agreement. The best retreat/advance prediction results were obtained at an elevation of 2 m, in which the models correctly forecasted the behaviour of nine profiles. Additionally, the predominance of accretion in the profiles during this simulation followed the same pattern observed during the monitoring period. Hence, simulation results achieved satisfactory metrics in predicting the morphological features of the profiles.

The simulation under storm conditions showed that the model is reliable beyond the training data range, but their results must be carefully considered. Data obtained from physics-based numerical models can be included to enhance the robustness of data models. This approach will allow the complementation of the observed field data required to properly feed data models with consistent time series, such as for extreme events, thus increasing its applicability.

In summary, this study reached the objective of combining different data sources to implement a RF model to forecast the evolution of multiple beach profiles. The error metrics achieved by both models were within the error range obtained by previous studies. Therefore, data models can be a valuable tool for monitoring and forecasting the evolution of beach profiles as long as there is enough quality data to implement them. Future studies could assess the interpretability of ML models applied to coastal datasets and test other variables besides those used in this study, contributing to the knowledge advance about complex coastal morphodynamics processes.

CRediT authorship contribution statement

Willian Weber de Melo: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **José Pinho:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Isabel Iglesias:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the Doctoral Grant SFRH/BD/151383/2021 financed by the Portuguese Foundation for Science and Technology (FCT), and with funds from the Ministry of Science, Technology and Higher Education, under the MIT Portugal Program. I. Iglesias also acknowledge the FCT financing through the CIIMAR Strategic Funding UIDB/04423/2020 and UIDP/04423/2020, and the CEEC program (2022.07420 CEECIND).

The authors also want to acknowledge the data supplied by the Portuguese Environmental Agency (APA) obtained in the *Programa de Monitorização da Faixa Costeira de Portugal Continental* (COSMO), co-founded by the Operational Program Sustainability and Efficiency in the Use of Resources (POSEUR) and available at: https://cosmo.apambiente.pt, and also the data supplied by the European Union MarRISK project: *Adaptación costera ante el Cambio Climático: conocer los riesgos y aumentar la resiliencia* (0262_MarRISK_1_E).

This study has been conducted using E.U. Copernicus Marine Service Information; https://doi.org/10.48670/moi-00030.

References

Aagaard, T., Greenwood, B., Hughes, M., 2013. Sediment transport on dissipative, intermediate and reflective beaches. Earth Sci. Rev. 124, 32–50. https://doi.org/10.1016/j.earscirev.2013.05.002.

Alveirinho Dias, J.M., Nittrouer, C.A., 1984. Continental shelf sediments of northern Portugal. Continent. Shelf Res. 3, 147–165. https://doi.org/10.1016/0278-4343(84)90004-9.

Antolínez, J.A.A., Méndez, F.J., Camus, P., Vitousek, S., González, E.M., Ruggiero, P., et al., 2016. A multiscale climate emulator for long-term morphodynamics (MUSCLE-morpho). J. Geophys. Res. Oceans 121, 775–791. https://doi.org/10.1002/2015JC011107.

APA, Agência Portuguesa do Ambiente, 2023. COaStal MOnitoring Program (COSMO). https://cosmo.apambiente.pt/data. (Accessed 17 April 2023).

Araújo, M. da A., 2000. Depósitos continentais e marinhos na plataforma litoral da região do Porto Importância da tectónica na sua organização espacial. Cien. Terra 14, 111–122.

Benveniste, J., Cazenave, A., Vignudelli, S., Fenoglio-Marc, L., Shah, R., Almar, R., et al., 2019. Requirements for a coastal hazards observing system. Front. Mar. Sci. 6 https://doi.org/10.3389/fmars.2019.00348.

Bio, A., Gonçalves, J.A., Iglesias, I., Pinho, J.L., Vieira, L., Vieira, J.M., et al., 2022. Coastal erosion: a northern-Portuguese case study. In: Martí Ezpeleta, A., Lorenzo González, N., Royé, D., Díaz Poso, A. (Eds.), Madrid: Asociación Española de Climatología.

Bio, A., Gonçalves, J.A., Pinho, J., Vieira, L., Vieira, J., Smirnov, G., et al., 2020a. Indicadores de vulnerabilidade de erosão costeira: Um estudo de caso no Norte de Portugal. Revista de Gestão Costeira Integrada 20, 197–209. https://doi.org/10.5894/rgci-n337.

Bio, A., Gonçalves, J.A., Pinho, J., Vieira, L., Vieira, J., Smirnov, G., et al., 2020b. Indicadores de vulnerabilidade de erosão costeira: Um estudo de caso no Norte de Portugal. Revista de Gestão Costeira Integrada 20, 197–209. https://doi.org/10.5894/rgci-n337.

Biolchi, L.G., Unguendoli, S., Bressan, L., Giambastiani, B.M.S., Valentini, A., 2022. Ensemble technique application to an XBeach-based coastal early warning system for the northwest adriatic sea (Emilia-Romagna region, Italy). Coast. Eng. 173 https://doi.org/10.1016/j.coastaleng.2022.104081.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification And Regression Trees. Routledge. https://doi.org/10.1201/9781315139470.

Chen, J.C., Ning, S.K., Chen, H.W., Shu, C.S., 2008. Flooding probability of urban area estimated by decision tree and artificial neural networks. J. Hydroinform. 10, 57–67. IWA Publishing. https://doi.org/10.2166/hydro.2008.009.

Chen, Q., Wang, N., Chen, Z., 2023. Simultaneous mapping of nearshore bathymetry and waves based on physics-informed deep learning. Coast. Eng. 183, 104337 https://doi.org/10.1016/j.coastaleng.2023.104337.

Copernicus Marine Service, 2022. Atlantic -iberian Biscay Irish- Ocean Wave reanalysis. https://doi.org/10.48670/moi-00030.
Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. In: Zhang, C., Ma, Y. (Eds.), Ensemble Machine Learning. Springer New York, New York, NY, pp. 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5.
Deltares, 2018. Delft3D-Flow: simulation of multi-dimensional hydrodynamic flows and transport phenomena, including sediments - user Manual. Version 3.15.
Fontán-Bouzas, Á., Andriolo, U., Silva, P.A., Baptista, P., 2022. Wave impact analysis on a beach-dune system to support coastal management and nourishment works: the showcase of mira, Portugal. Front. Mar. Sci. 9 https://doi.org/10.3389/fmars.2022.861569.
Gharagozlou, A., Anderson, D.L., Gorski, J.F., Dietrich, J.C., 2022. Emulator for eroded beach and dune profiles due to storms. J. Geophys. Res. Earth Surf. 127 https://doi.org/10.1029/2022JF006620.
Gomes, A., Pinho, J.L.S., Granja, H.M., 2018. Morphodynamic modelling of a gravel beach at the NW Portuguese coast. IX Simpósio sobre a Margem Ibérica Atlântica, pp. 309–310. Coimbra.
Granja, H., Pinho, J.L., 2012. Coastal Defense in NW Portugal: the Improbable Victory, pp. 251–266. https://doi.org/10.1007/978-94-007-4123-2_15.
Gunawan, P.H., Iryanto, Ghozali AL., Ismantohadi, E., Baizal, Z.K.A., Satrio, A., 2023. Data-driven Shoreline Change Forecasting on Eretan Beach Using Random Forest. Institute of Electrical and Electronics Engineers (IEEE), pp. 1–4. https://doi.org/10.1109/icacnis57039.2022.10054972.
Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.-Z., 2019. XAI—explainable artificial intelligence. Sci. Robot. 4 https://doi.org/10.1126/scirobotics.aay7120.
Iglesias, I., Pinho, J.L., Avilez-Valente, P., Melo, W., Bio, A., Gomes, A., et al., 2022. Improving estuarine hydrodynamic forecasts through numerical model ensembles. Front. Mar. Sci. 9. https://doi.org/10.3389/fmars.2022.812255.
Irizarry, R.A., 2019. Introduction to data science. https://doi.org/10.1201/9780429341830.
Jackson, D., Short, A., 2020. Sandy Beach Morphodynamics. Elsevier.
Jayson-Quashigah, P.N., Appeaning Addo, K., Amisigo, B., Wiafe, G., 2019. Assessment of short-term beach sediment change in the Volta Delta coast in Ghana using data from Unmanned Aerial Vehicles (Drone). Ocean Coast Manag. 182 https://doi.org/10.1016/j.ocecoaman.2019.104952.
Kalligeris, N., Smit, P.B., Ludka, B.C., Guza, R.T., Gallien, T.W., 2020. Calibration and assessment of process-based numerical models for beach profile evolution in southern California. Coast. Eng. 158 https://doi.org/10.1016/j.coastaleng.2020.103650.
Karunarathna, H., Horrillo-Caraballo, J.M., Reeve, D.E., 2012. Prediction of cross-shore beach profile evolution using a diffusion type model. Continent. Shelf Res. 48, 157–166. https://doi.org/10.1016/j.csr.2012.08.004.
Kazeminezhad, M.H., Etemad-Shahidi, A., 2015. A new method for the prediction of wave runup on vertical piles. Coast. Eng. 98, 55–64. https://doi.org/10.1016/j.coastaleng.2015.01.004.
Kubat, M., 2017. An Introduction to Machine Learning. https://doi.org/10.1007/978-3-319-63913-0.
Lee, S., Lim, C., Lee, J.-L., 2022. Prediction of long-term beach profile evolution due to episodic wave incidence under tidal environment. Front. Mar. Sci. 9 https://doi.org/10.3389/fmars.2022.831262.
Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., Aarninkhof, S., 2018. The state of the world's beaches. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-018-24630-6.
Matheen, N., Harley, M.D., Turner, I.L., Splinter, K.D., Simmons, J.A., Thran, M.C., 2021. Bathymetric data requirements for operational coastal erosion forecasting using xbeach. J. Mar. Sci. Eng. 9 https://doi.org/10.3390/jmse9101053.
McAllister, E., Payo, A., Novellino, A., Dolphin, T., Medina-Lopez, E., 2022. Multispectral satellite imagery and machine learning for the extraction of shoreline indicators. Coast. Eng. 174, 104102 https://doi.org/10.1016/j.coastaleng.2022.104102.
McGranahan, G., Balk, D., Anderson, B., 2007. The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. Environ. Urbanization 19, 17–37. https://doi.org/10.1177/0956247807076960.
Melo, W., Pinho, J., Iglesias, I., Bio, A., Avilez-Valente, P., Vieira, J., et al., 2020. Hydro- and morphodynamic impacts of sea level rise: the Minho estuary case study. J. Mar. Sci. Eng. 8, 441. https://doi.org/10.3390/jmse8060441.
Mil-Homens, M., Brito, P., Magalhães, V., Rosa, M., Neres, M., Silva, Marta, et al., 2022. Integrated Geophysical and Sedimentological Datasets for Assessment of Offshore Borrow Areas: the CHIMERA Project (Western Portuguese Coast), 505. Geological Society, London, Special Publications, pp. 131–153. https://doi.org/10.1144/SP505-2019-100.
Otto, P., Piter, A., Gijsman, R., 2021. Statistical analysis of beach profiles – a spatiotemporal functional approach. Coast. Eng. 170 https://doi.org/10.1016/j.coastaleng.2021.103999.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pedrosa, A. de S., 2013. Risco de erosão no litoral norte de Portugal: uma questão de ordenamento do território. Sociedade & Natureza 25, 133–146. https://doi.org/10.1590/S1982-45132013000100011.
Pinho, J., Gomes, A., Weber de Melo, W., Iglesias, I., Bio, A., Avilez-Valente, P., et al., 2024. Comparison between traditional and digital image based approaches for beach sediment size characterization. A case study at the NW Portuguese coast. Preparation.
Pinto, C.A., Silveira, T.M., Teixeira, S.B., 2020. Beach nourishment practice in mainland Portugal (1950–2017): overview and retrospective. Ocean Coast Manag. 192, 105211 https://doi.org/10.1016/j.ocecoaman.2020.105211.
Ponte Lira, C., Nobre Silva, A., Taborda, R., Freire de Andrade, C., 2016. Coastline evolution of Portuguese low-lying sandy coast in the last 50 years: an integrated approach. Earth Syst. Sci. Data 8, 265–278. https://doi.org/10.5194/essd-8-265-2016.
Pörtner, H.-O., Roberts, D.C., Tignor, M., Poloczanska, E.S., Mintenbeck, K., Alegría, A., et al., 2022. IPCC, 2022: climate change 2022: impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. https://doi.org/10.1017/9781009325844. Cambridge, Uk and New York, NY, USA.
Prodger, S., Russell, P., Davidson, M., 2017. Grain-size distributions on high-energy sandy beaches and their relation to wave dissipation. Sedimentology 64, 1289–1302. https://doi.org/10.1111/sed.12353.
Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 https://doi.org/10.1029/2011WR011527.
Roelvink, D., Reniers, A., van Dongeren, A., van Thiel de Vries, J., McCall, R., Lescinski, J., 2009. Modelling storm impacts on beaches, dunes and barrier islands. Coast. Eng. 56, 1133–1152. https://doi.org/10.1016/j.coastaleng.2009.08.006.
Sanderson, M.G., Teixeira, M., Fontes, N., Silva, S., Graça, A., 2023. The probability of unprecedented high rainfall in wine regions of northern Portugal. Clim Serv 30, 100363. https://doi.org/10.1016/j.cliser.2023.100363.
Santos, A.I., Oliveira, A., Pinto, J.P., Freitas, M.C., 2021. Hydrodynamic and sediment transport patterns in the Minho and Douro estuaries (NW Portugal) based on ADCP monitoring data: Part 1-tidal sediment exchanges. Coasts 1, 31–55. https://doi.org/10.3390/coasts1010003.
de Schipper, M.A., Ludka, B.C., Raubenheimer, B., Luijendijk, A.P., ThomasA, Schlacher, 2020. Beach nourishment has complex implications for the future of sandy shores. Nat. Rev. Earth Environ. 2, 70–84. https://doi.org/10.1038/s43017-020-00109-9.
Suzuki, T., Cox, D.T., 2021. Evaluating XBeach performance for extreme offshore-directed sediment transport events on a dissipative beach. Coast Eng. J. 63, 517–531. https://doi.org/10.1080/21664250.2021.1976452.
UN-Habitat UNHSP, 2022. World Cities Report 2022. Nairobi, Kenya.
Vieira, B.F.V., Pinho, J.L.S., Barros, J.A.O., Antunes do Carmo, J.S., 2020. Hydrodynamics and morphodynamics performance assessment of three coastal protection structures. J. Mar. Sci. Eng. 8 https://doi.org/10.3390/jmse8030175.
Viitak, M., Avilez-Valente, P., Bio, A., Bastos, L., Iglesias, I., 2021. Evaluating wind datasets for wave hindcasting in the NW Iberian Peninsula coast. J. Oper. Oceanogr. 14, 152–165. https://doi.org/10.1080/1755876X.2020.1738121.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
Vousdoukas, M.I., Ranasinghe, R., Mentaschi, L., Plomaritis, T.A., Athanasiou, P., Luijendijk, A., et al., 2020. Sandy coastlines under threat of erosion. Nat. Clim. Change. https://doi.org/10.1038/s41558-020-0697-0 (in press).
Waskom, M., 2021. seaborn: statistical data visualization. J. Open Source Softw. 6, 3021. https://doi.org/10.21105/joss.03021.
Weber de Melo, W., Pinho, J., Iglesias, I., 2023. Coastal morphodynamic emulator for early warning short-term forecasts. Environ. Model. Software 165, 105729. https://doi.org/10.1016/j.envsoft.2023.105729.
Weber de Melo, W., Pinho, J., Iglesias, I., 2022. Emulating the estuarine morphology evolution using a deep convolutional neural network emulator based on hydrodynamic results of a numerical model. J. Hydroinf. https://doi.org/10.2166/hydro.2022.068.
Ye, Z., Shi, F., Zhao, X., Hu, Z., Malej, M., 2021. A data-driven approach to modeling subgrid-scale shallow marsh hydrodynamics. Coast. Eng. 166 https://doi.org/10.1016/j.coastaleng.2021.103856.