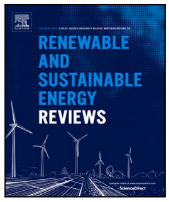
Renewable and Sustainable Energy Reviews 167 (2022) 112751

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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



A data-driven long-term metocean data forecasting approach for the design of marine renewable energy systems

Markel Penalba [a,c,*], Jose Ignacio Aizpurua [b,c], Ander Martinez-Perurena [a,b], Gregorio Iglesias [d,e]

[a] Fluid Mechanics Department, Mondragon University, Loramendi 4, 20500 Arrasate, Spain
[b] Signal Theory & Communications Department, Mondragon University, Loramendi 4, 20500 Arrasate, Spain
[c] Ikerbasque, Basque Foundation for Science, Euskadi Plaza 5, Bilbao, Spain
[d] School of Engineering and Architecture & MaREI, Environmental Research Institute, University College Cork, College Road, Cork, Ireland
[e] University of Plymouth, School of Engineering, Marine Building, Drake Circus, Plymouth, PL4 8AA, United Kingdom

ARTICLE INFO

Keywords:
Metocean data
Re-analysis data
Long-term trend
Wave forecasting
Machine learning
Regression algorithms
Classification algorithms

ABSTRACT

The potential of Marine Renewable Energy (MRE) systems is usually evaluated based on recent metocean data and assuming the stationarity of the MRE resource. Yet, different studies in the literature have shown long-term resource variations and even the connection between ocean warming and wave power variations. Therefore, it is crucial to accurately characterise the future resource, including these long-term variations. To that end, this paper presents a novel data-driven forecasting approach through the combination of machine-learning (ML) and oceanic engineering concepts. First, the historical resource is characterised in the Bay of Biscay, including the different long-term trends identified based upon the dataset obtained via the SIMAR model ensemble. Secondly, the most relevant features of the metocean dataset are extracted and selected via advanced statistical techniques. Finally, three different ML algorithms are designed, validated and tested. All three ML models demonstrate to adequately represent the overall pattern of the dataset, although showing difficulties with reproducing particular peak values. Accordingly, an alternative interval prediction approach is presented for three different wave height discretisation levels, showing a greater potential for long-term metocean data forecasting.

1. Introduction

In the last decade, climate change caused by anthropological activity has become a global priority. In fact, society's energy demand is in progressive growth, averaging around 1% to 2% per year [1]. Hence, despite the increase in renewable energy production, fossil fuel sources remain dominant of the energy sector, with carbon emissions increasing by 0.5% on an annual basis. In this sense, the energy transition needs a strong acceleration to reach a carbon-neutral energy system by 2050, where renewable energies become the main player. To reach carbon neutrality and contain global warming below the safety threshold of 1.5 °C determined by [2,3] estimates that the installed capacity of renewable energy systems would need to be increased up to 18.000 GW (with additional 14.000 GW of solar and wind capacity). In this context, marine renewable energies (MRE) can provide a significant boost to the sector, *e.g.* worldwide offshore wind capacity is expected to increase by 30 in the next 30 years, while [4] estimates a potential of tidal and wave energy devices of about 10% of Europe's power demand. Similarly, [5] suggests that harvesting 20% of the technical marine resource in the US coast would cover 10% of electricity demand in the USA.

Most of these studies are based on the mean power of the resource. However, intra- and inter-annual variations are crucial when designing MRE systems [6]. Different alternatives have included other relevant aspects to the characterisation of the wave energy resource. In a recent study, [7] proposes an alternative *wave exploitability index* defined as the ratio between the root-mean-square and maximum wave height to assess the suitability of different locations worldwide, moving beyond the classical mean wave power metric. Similarly, [8] suggests a resource classification system for US coastal waters that includes key attributes for energy planning and project development. Another alternative is presented in [9], where authors define a decision-making process that considers the energy generation capacity of wave energy converters via the combination of annual energy production and capacity factor, installation aspects with the bathymetry, and grid connection accessibility. In any case, all the studies in the literature only use past

* Corresponding author at: Fluid Mechanics Department, Mondragon University, Loramendi 4, 20500 Arrasate, Spain.
E-mail addresses: mpenalba@mondragon.edu (M. Penalba), jiaizpurua@mondragon.edu (J.I. Aizpurua), ander.martinezrd@alumni.mondragon.edu (A. Martinez-Perurena), gregorio.iglesias@ucc.ie (G. Iglesias).

https://doi.org/10.1016/j.rser.2022.112751
Received 29 December 2021; Received in revised form 9 May 2022; Accepted 29 June 2022
Available online 11 July 2022

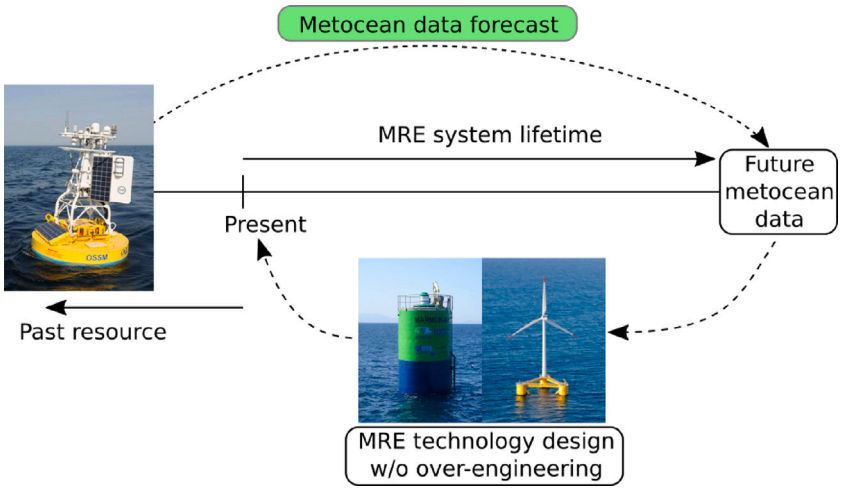


Fig. 1. The graphical framework of the present study.

data to assess the potential of different locations. Yet, [10] demonstrates global warming is affecting marine energy sources, increasing the power and frequency of extreme metocean conditions, as shown by [11]. In addition, a recent study has concluded that, despite the relatively high uncertainties, sea level rise caused by climate change will very likely affect the estuarine tidal energy due to variations in tidal and sediment dynamics [12]. This means that current potential assessment of MRE resources and MRE technology expansion roadmaps may become unreliable in a relatively short period of time. As a consequence, it may be the case that new MRE system designs may also become outdated, because these designs largely depend on accurate resource assessment, especially accurate extreme event estimation. Therefore, it is essential to accurately forecast the metocean conditions that MRE devices will face in the long-term future. Fig. 1 graphically illustrates the framework. Based on these predictions, MRE technology developers will be able to guarantee structural integrity of their devices avoiding over-engineering, classification societies will be able to update their certification and standards in order to better suit the requirements of the MRE industry, and decision-makers will plan sensible roadmaps for the worldwide expansion of MRE.

However, long-term forecasting of metocean conditions is a complex and disregarded task. The forecasting of renewable energy sources has been mostly focused on the forecasting of the resource variability within a relatively limited time horizon, as reviewed by [13]. Before going into more details, the forecasting horizons should be defined. Accordingly, Fig. 2 shows the different horizons, where short-term represents a few seconds, medium-term focuses on the prediction of metocean conditions between hours, days and weeks ahead, and long-term represents a temporal horizon of years and decades. Each forecasting horizon is necessary for different MRE-related applications:

- Short-term forecasting is particularly interesting for energy maximising control purposes and ship motion prediction in offshore operations;
- Medium-term prediction is relevant for maintenance operation planning, MRE device operation mode selection, *e.g.* power production mode or survivability mode, and bid placing on wholesale energy markets; and
- Long-term forecasting is important for MRE farm deployment site selection, feasibility studies and system design.

The literature shows several examples of short-term wave forecasting. Fusco and Ringwood [14] suggests different forecasting models, such as a cyclical model with time-invariant frequencies that becomes linear in the parameters and an extended version with time-variant frequencies based on the extended Kalman filter; autoregressive (AR) models; Artificial Neural Networks (ANN); and a Gaussian Process (GP) model. Recent publications show improvements of the same models for short-term forecasting, such as the revisited AR model suggested in [15] and the wave-spectrum-informed GP models proposed in [16]. The latter has been demonstrated to outperform AR and ANN methods, and includes the advantage of providing the uncertainty of the prediction [17]. Medium-term prediction approaches suggested in the literature include data-driven and physics-based methods, which tend to forecast wave statistical parameters, *e.g.* significant wave height (H_s) and peak period (T_p). The latter have traditionally been used via the well-known Simulating WAves Nearshore (SWAN) model [18]. James et al. [19] compares Machine Learning (ML) techniques to SWAM, concluding that similar results can be obtained for a fraction of the computational time. The authors suggest a convolutional neural network (CNN) and a Support Vector Regression (SVR) model, and conclude that CNN models are more appropriate. The use of ANNs in medium-term wave height forecasting is also analysed in several studies. Shamshirband et al. [20] concludes that the efficacy of the model especially relies on the training data size and the forecasting horizon. In contrast, [21] finds the SVR method to enable removing the need for model calibration. Other alternatives include Extreme Learning Machines, Genetic Algorithms, and Random Forest (RF), suggested by [22–24], respectively. Long-term trends of wave conditions have long been ignored in the literature due to the lack of interest for the community. In fact, the belief that past resource data is representative of the future resource is still present in the community. As a consequence, long-term forecasting approaches suggested in the literature are scarce. Therefore, the main goal of the present study, as highlighted in Fig. 1, is the development of a ML-based forecasting model to predict the statistical parameters of metocean conditions for the following

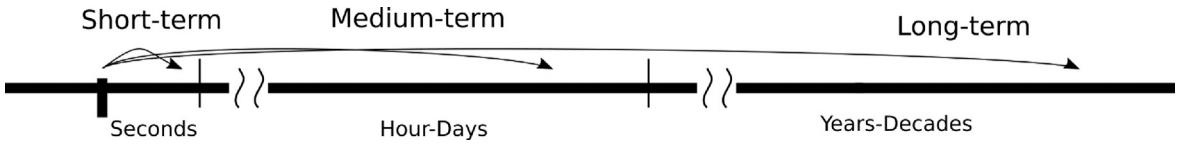


Fig. 2. Short, medium and long term forecasting horizons.

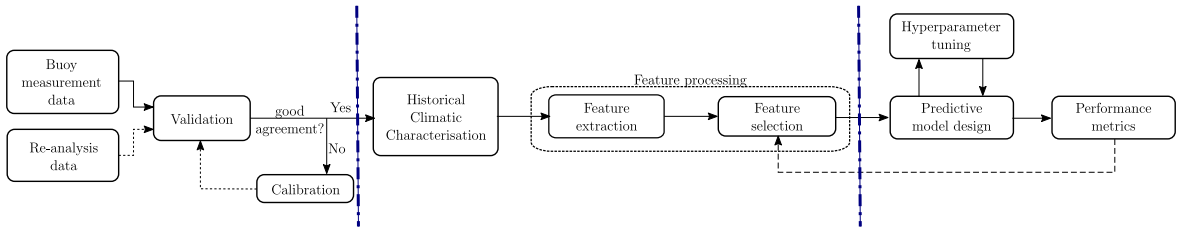


Fig. 3. Summarised flowchart of the methodology.

years and decades in a specific location. On the one hand, this approach involves the identification of meaningful predictive trends through an statistical analysis of local historical metocean data. On the other hand, an appropriate combination of the different statistical features should be found in order to synthesise the most valuable insights into the resource characteristics for the ML-based models. However, the authors do not aim to precisely predict future sea-states at a specific time, which would result unfeasible, but the determination of the future probability density functions and evolution of extreme events.

The present paper is organised as follows. Section 2 describes the methodology and the ML-based models suggested in this study, Section 3 defines the location selected for the analysis including the metocean data for that location, Section 4 presents the metocean data characterisation of the past data, Section 5 extracts and selects the most relevant parameters via statistical feature extraction techniques, Section 6 presents the results of the forecasting models, Section 7 discusses the potential improvements and Section 8 draws the main conclusions of the study.

2. Methodology

Fig. 3 illustrates the methodology adopted in the present study, which is comprised of three main stages: (i) metocean data collection and validation, (ii) dataset processing for the identification of statistical features, and (iii) predictive model design and evaluation. These three stages are further described in the following subsections.

2.1. Dataset validation

The first stage of the methodology is the analysis of the simulation/re-analysis dataset in order to verify that it is representative of the area of study. During the validation, the most relevant parameters obtained from *in-situ* buoy measurements and re-analysis datasets are compared by means of statistical metrics. The most common metrics employed in climatic studies are the Root Mean Square Deviation (RMSD), the standard deviation (σ_y) and the correlation coefficient. The RMSD is defined as follows:

$$RMSD = \sqrt{\frac{\sum_{n=1}^{N}\left(\hat{y}_n - y_n\right)^2}{N}}, \tag{1}$$

where $\hat{y}_n$ is the variable obtained from the *SIMAR* model (see Section 3), y_n is the measured variable and N the number of samples considered within the validation period.

The standard deviation is given as,

$$\sigma_y = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_n - \mu_y\right)^2}, \tag{2}$$

where μ_y represents the mean of all samples,

$$\mu_y = \frac{1}{N}\sum_{n=1}^{N} y_n. \tag{3}$$

The Pearson Correlation (PC) coefficient is specially relevant for the analysis of similarity and is given as follows,

$$PC = \frac{cov(\hat{y}, y)}{\sigma_{\hat{y}}\sigma_y}, \tag{4}$$

where $cov(\hat{y}, y)$ is the covariance. However, the PC coefficient shows the linear correlation. Thus, an alternative non-linear rank class correlation coefficient is also considered in this study. One of the most common nonlinear metrics is the non-parametric Spearman's Rank Correlation (SRC) coefficient, which defines the relationship between the rankings of two variables as follows,

$$SRC = \frac{cov_{rg}(\hat{y}, y)}{\sigma_{\hat{y}}^{rg}\sigma_y^{rg}}, \tag{5}$$

where $cov_{rg}(\hat{y}, y)$ is the covariance of the rank variables, and $\sigma_{\hat{y}}^{rg}$ and σ_y^{rg} are the standard deviations of the two variables' rankings. Based on these four metrics, the suitability of the re-analysis dataset and the need for calibration can be evaluated, as shown in Fig. 3.

2.2. Data-driven metocean data characterisation

Once the re-analysis dataset is validated, the second stage consists on the extraction of the statistical features that are relevant for the design of the forecasting model. In order to extract these statistical features, a preliminary characterisation of the historical metocean data is crucial. In fact, input data for prediction models must include all the possible information about the characteristics of the metocean data in a concise manner. However, traditional resource assessment methods such as (i) bi-variate $T_P - H_s$ distributions through *e.g.* copulas, (ii) predominant wind and wave directions via wind and wave roses, (iii) inter- and intra-annual variations, and (iv) rate of extreme events, are not sufficient for the construction of long-term forecasting models. Therefore, other less common statistical techniques are employed in the present paper for the characterisation of the metocean data over the last decades. Three different statistical techniques are used to identify the three particular characteristics of the ocean resource: stationarity, seasonality and randomness. These statistical techniques are common is statistics, but less so in ocean resource characterisation. In addition to these characteristics, forecasting models are also fed with other physical and non-physical features, which are identified by feature extraction techniques. Therefore, the statistical characteristics and techniques used in this paper are described in Section 2.2.1, while the information about feature extraction is provided in Section 2.2.2.

2.2.1. Resource characterisation and trend identification

Due to the particularities of ocean waves, the analysis of the data should consider the following statistical characteristics (C):

(C.i) *Stationarity* involves that statistical properties, such as the mean and variance, remain constant over time. Therefore, statistical properties of non-stationary signals vary with time. The identification of this parameter is important for pattern analysis, which will be employed in the subsequent forecasting model development. The traditional belief is that metocean data is rather stationary, despite the inter- and intra-annual seasonal variations.

(C.ii) *Seasonality* is a particular characteristic of time-series in which the fluctuations of the data occur at regular intervals and are recurrent within a determined period of time (day, seasons, year). These fluctuations generally form easily predictable patterns. In

the case of metocean data, according to the meteorological standards, there exists a strong annual seasonality that can be divided into four seasons: December–February, March–May, June–August and September–November.

(C.iii) *Randomness* means the lack of a pattern or logic and, thus, implies the lack of predictability, which in the case of metocean data may arise from different sources. On the sea-surface there exist spatial and temporal randomness. Short-term variability can be studied through GP models and the medium-term random variability of the sea-states occurs as a consequence of weather conditions.

Opposed to the general belief of stationary metocean data, different studies in the literature present significant long-term variations of the wave energy resource. In this direction, [10] presents the most recent and relevant study, where the increase in global wave energy is connected to oceanic warming. In fact, the authors show that these trends vary significantly geographically, demonstrating the relevance of the local studies. Thus, apart from characterising the seasonality and the randomness of the signal, identifying the overall long-term trend can be important.

Consequently, the characterisation of the resource is carried out using the following three techniques (T):

(T.i) *Signal decomposition* assumes that a signal consists of different intrinsic modes of oscillation, and allows for the extraction and isolation of these modes. Signal decomposition can be carried out, for example, by means of the Empirical Mode Decomposition (EMD), which is a data-driven method and does not make any assumption on the periodicity of the signal or the use of specific base function. In addition, the EMD method provides results in the time-domain, dividing the signal into three modes, each of which is related to the aforementioned statistical characteristics: the trend, the seasonality and the residual. The isolation of these modes is particularly interesting for the construction of the predictive models because the cyclical components (seasonality) and the trend can significantly hamper their performance.

(T.ii) *Envelopes* study the smooth curves that outline the extremes of an oscillating signal. In the study of historical metocean data, envelopes are useful for the analysis of the evolution of maximum and minimum sea-state values along relatively long periods of time.

(T.iii) *Conditional probability* evaluates the occurrence probability of an event conditioned on another event that has already occurred. In ocean engineering, conditioned probabilities enable the characterisation of the resource by dividing the resource in different regions, *e.g.* low-energetic, medium-energetic and high-energetic sea-states.

While signal decomposition is useful to extract the seasonality content and identifying the trend, envelopes and conditioned probabilities are useful for the analysis of extreme events. In fact, qualitative and quantitative characterisation of extreme events can be effectively carried out via envelopes and condition probabilities, respectively.

2.2.2. Statistical feature extraction and selection

In addition to the three characteristics defined above, other features are also necessary for the prediction of future metocean conditions. These features can be either (i) physical variables available in the different metocean datasets, *e.g.* wave height, period, sea-surface temperature (SST), wind speed and current velocity; or (ii) features inferred from the statistical post-precessing of the previous physical parameters, *e.g.* mean, variance, standard deviation, skewness, momentum and derivatives, which can provide inherent predictive information of the process. The post-processing of the statistical features requires dividing the datasets into smaller sets. In the present study, due to the multi-decadal analysis carried out, statistical features are inferred every month, as illustrated in Fig. 4.

Once different statistical features are extracted, the final number of features used in the forecasting model must be reduced significantly so as to reduce the computational cost and improve the performance of the model. This is achieved via a feature selection process that ranks these features with respect to their relevance for the prediction. One common measure to quantify the relevance of the different features is the PC coefficient, defined in Eq. (4). For the PC-based comparison of different features, the *heatmap* graphical representation is often used, providing a clear view of the different correlations. Alternatively, there exist feature selection algorithms that automatically eliminate the features that have low relevance according to a loss function. The *Recursive Feature Elimination* algorithm is one of the common ML tools that eliminates a feature in every iteration of the regression, quantifying the accuracy by means of different metrics (Normalised Mean Absolute Error (NMAE) and Normalised Mean Squared Error (NMSE), for example). Finally, the RF algorithm includes an *importance score* that quantifies the relevance of the different features for the prediction model.

2.3. Predictive models

Once the historical data is adequately characterised, and the most relevant information is extracted and selected, the critical point is the development of the forecasting model. The design of the predictive model consists on the selection, development and optimisation of the model. The optimisation of the model involves tuning the different model hyperparameters to improve the performance of the predictive model and generalise results. Similarly, the selection of the features included in the model is crucial for improving the performance of the predictive model. Therefore, the second and third stages of the methodology are inter-connected, as depicted in Fig. 3, where the feature selection and hyperparameter tuning form the design loop of the predictive model. In this study, three popular ML models are evaluated.

2.3.1. Random forests

Random Forest is an ensemble of recursive trees (see [25]). Each tree is generated from a bootstrapped sample and a random subset of descriptors is used at the branching of each node in the tree. Random Forests create a large number of trees by repeatedly resampling training data and averaging differences through voting. The RF model has been implemented through the `scikit-learn` package in Python. The hyperparameters include the number of regression trees (n_estimators) and the size of each tree (max_depth), minimum samples required to split a node (min_samples_split) and minimum samples required at each leaf (min_samples_leaf). These parameters have been optimised searching the best parameters from a predefined grid of parameters.

2.3.2. Support vector regression

The SVR maps input data into an m-dimensional feature space using a kernel function. The kernel translates a non-linearly separable problem into a feature space linearly separable by a hyperplane. A loss function (ϵ) that ignores the errors situated within a certain distance of the true value is defined and the model is parameterised through the choice of the kernel function. For a nonlinear problem, the Radial Basis Function (RBF) kernel is recommended:

$$k(x, x') = e^{(\gamma \|x - x'\|^2)}, \tag{6}$$

where γ is the RBF width, x and x' are training and testing data samples, and $\|d\|$ is the Euclidean norm. The SVR solves an optimisation problem maximising the distance from the hyperplane to the nearest training point and penalises the loss function with a cost variable c. The training of the model consists on calculating the hyperparameters c and γ. Model training is performed using the Python `sklearn` package and grid search was used to optimise c and γ.

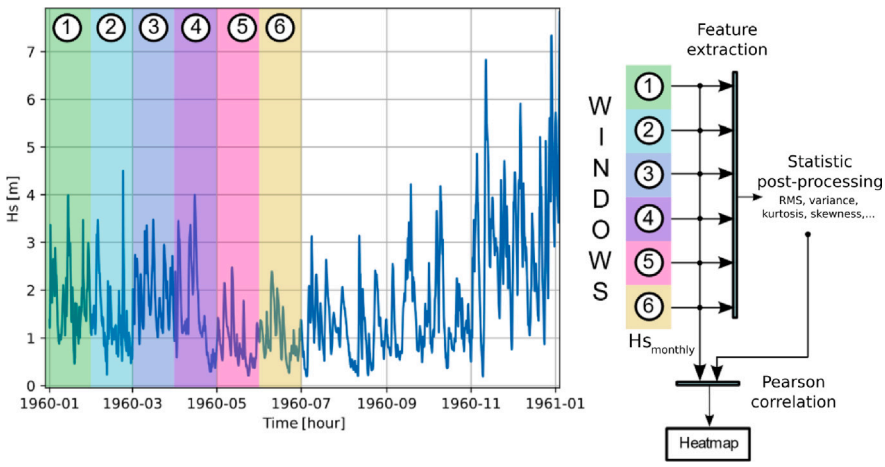


Fig. 4. Feature extraction and correlation analysis framework.

2.3.3. Artificial neural networks

ANNs are widely used for classification and regression tasks (see [26]). The multilayer perceptron (MLP) feedforward model was used in this work. The MLP is a multi-layer network (input, hidden, output) comprised of fully connected neurons. Each neuron performs a weighted sum of its inputs and passes the results through an activation function. All ANN models use a rectifier linear unit (ReLU) activation function for the hidden layers, linear activation function for output nodes and the Adam optimisation algorithm. Model training is performed using a back-propagation algorithm. The goal is to learn the neuron weights to generate the network output from the sample input, which minimises the error with respect to the target output. To select an optimal number of hidden nodes and hidden layers, a number of networks are trained varying the number of hidden nodes from 1 to 20, and hidden layers from 1 to 4. Hence, the network with the highest accuracy was selected. The ANN is implemented using the `TensorFlow` Python package.

The methodology described in this section is executed consecutively in the following four sections, where the dataset validation is carried out in Section 3, resource characterisation and feature extraction are described in Sections 4 and 5, respectively, and the predictive model is designed in Section 6.

3. Geographical location & metocean data validation

The geographical location selected for the present study is the Gulf of Biscay in the North Atlantic Ocean, where the potential for MRE farm deployment, particularly for wave energy, is shown to be significant [27].

Historical metocean data for specific locations is usually provided by national or international oceanographic agencies, such as the *National Oceanic and Atmospheric Agency* in the United States [28] or Puertos del Estado in Spain [29], which own data collection equipment in the areas of interest. In addition, these oceanographic agencies own *in-house* numerical models calibrated against these measurements.

Hence, historical metocean data from different sources is typically employed, collected via either *in-situ* measurements as in [30], satellite altimeter measurements, see *e.g.* [31], or atmospheric re-analyses of the *European Centre for Medium-Range Weather Forecasts* (ECMWF) as suggested in [32–34], among others. In fact, the combination of measurements and re-analysis methods is also a typical procedure. For example, [35,11] use *in-situ* measurements, which serve as validation/calibration datasets for re-analysis datasets.

Usually, buoy measurements are considered to be the ground truth. However, the spatial coverage of measurement buoys is limited and the maintenance operation requirements are expensive [36]. In that sense, and due to the significant development of the computational power, re-analysis datasets have gained greater relevance for long-term historical metocean data analysis, particularly because of the capacity to cover very wide spatial areas assimilating a large number of historical observations.

In the present paper, two different data sources are used:

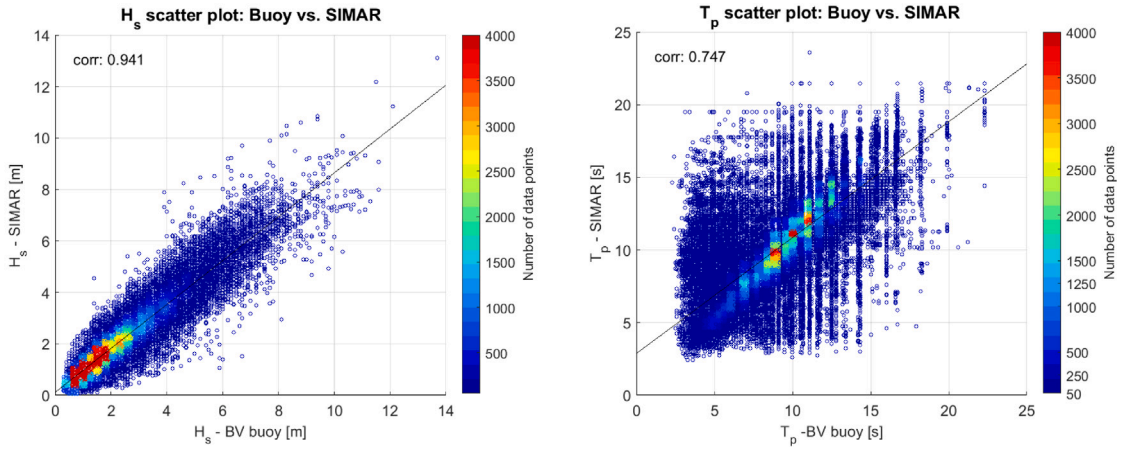
- *SIMAR* is an ensemble of modelling metocean data created upon a high-resolution numerical model by the Spanish Oceanographic Agency *Puertos del Estado*, which covers the coast along the Iberian Peninsula between 1958–2020 with a temporal resolution of 1 h.
- Bilbao–Vizcaya (BV) buoy measurements also provided by the Spanish Oceanographic Agency *Puertos del Estado*, which also provides metocean data of the wave and wind resource with a temporal resolution of 1 h. However, buoy measurements are available from 1990, meaning that only 30 years of *in-situ* measurements are available.

By means of the metrics described in Section 2.1, the validation is carried out for the two main statistical parameters that characterise sea-states: H_s and T_p. Fig. 5(a) and (b) represent the relationship between the hourly BV buoy measurements and the *SIMAR* model by means of a scatter plot, showing a good agreement for H_s and a significantly higher dispersion for T_p.

Although a good overall representation of the real measurements is provided by the *SIMAR* model, the discrepancy increases substantially at some peaks, which produces the dispersion shown in Fig. 5(b). Table 1 shows the statistical parameters for H_s and T_p, including the mean values over the period between 1990–2020. As in Fig. 5, dispersion is found to be larger for T_p. Similar results are found in the literature when comparing re-analysis datasets and buoy measurements, with higher discrepancies for T_p values, see [35,11]. Therefore, data from *SIMAR* is considered to be adequate for the analysis and is selected over the *BV-buoy* measurements due to its larger temporal coverage, reaching back in time until 1960.

4. Historical analysis of past metocean conditions

In order to accurately predict future metocean conditions, historical trends of the resource must be well-understood. The main objective of the historical analysis is the extraction of the statistical characteristics of the dataset focusing on the stationarity, seasonality and randomness characteristics (C.i–C.iii in Section 2.2). To that end, wave resource data for the last six decades is studied through signal decomposition, envelopes and conditional probability methods described in Section 2.2 (T.i–T.iii).



a) H_s scatter plot. b) T_p scatter plot.

Fig. 5. Metocean re-analysis data validation against the *BV-buoy*.

Table 1
Statistical parameters for the validation of the *SIMAR model with respect to BV-buoy*.

Dataset	T_p					H_s				
	Mean [s]	$RMSD$ [s]	PC [−]	SRC [−]	σ_y [s]	Mean [m]	$RMSD$ [m]	PC [−]	SRC [−]	σ_y [m]
BV-buoy	9.65	–	–	–	2.71	1.93	–	–	–	1.12
SIMAR-3155039	10.23	1.98	0.75	0.785	2.82	1.73	0.4	0.94	0.92	1.11

For a better understanding of the wave trends, the Wave Energy Flux (WEF) enables the study of the combined impact of T_p and H_s. Under the deep water approximation, the WEF is built up upon the combination between H_s and T_p as follows:

$$WEF = 0.49 T_e H_s^2 = 0.49 \alpha T_p H_s^2. \tag{7}$$

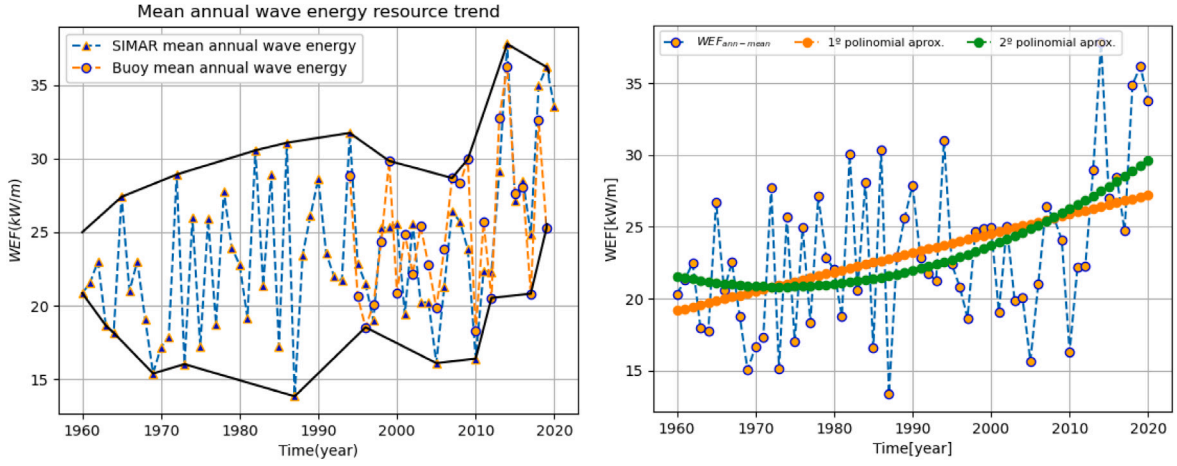
Since the original WEF equation includes the energy period (T_e), a correction has to be applied following the relationship defined in frequency spectra,

$$T_e = \alpha T_p, \tag{8}$$

where α depends on the frequency of the spectrum, which, considering a JONSWAP spectrum, has been considered to be $\alpha = 0.9$, as suggested in [37]. For the sake of simplicity, the directionality of the wave resource has been neglected, which is expected to be incorporated in future versions of the study.

Fig. 6(a) illustrates the evolution of the WEF along the last six decades based on the *SIMAR* dataset and the last three decades using the *BV-buoy* measurements. Although differences between the two datasets are noticeable, both show a similar increasing trend. These trends are detected despite the considerable fluctuations of the annual mean WEF values, which illustrate the inter-annual variations mentioned previously in Section 2.2. Fig. 6(b) shows the increasing trend represented by a linear and a second-order regression. The linear regression shows a decadal growth of about 1.3 kW/m (7%), while the second-order regression assumes an expanding growth rate with a rate of 2.5 kW/m in the last decade.

However, the regression models may be biased by the particular characteristics of the resource, such as the impact of seasonal variations



a) b)

Fig. 6. Wave energy flux variations along the decades (a) for the SIMAR and *BV-buoy* datasets and (b) the representation of trends by means of a linear and second-order regression.

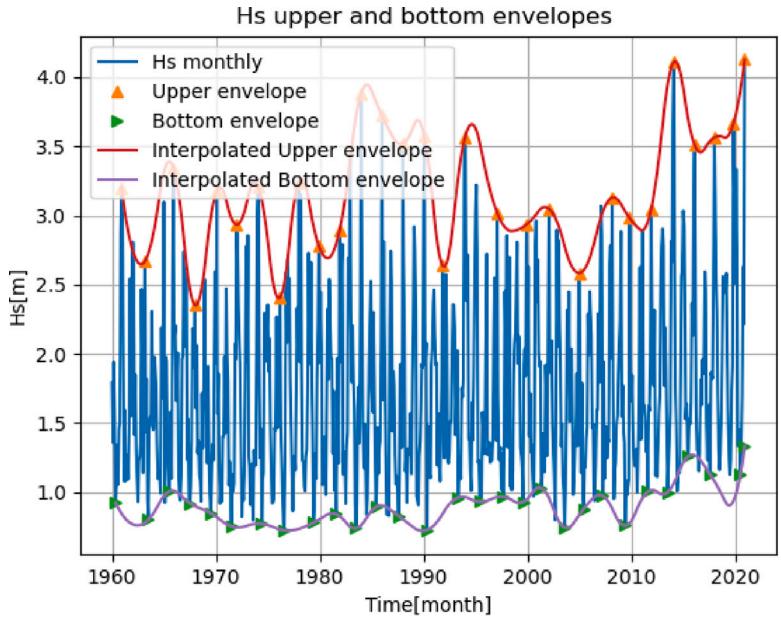


Fig. 7. Envelop analysis: wave height variations with upper and lower envelops.

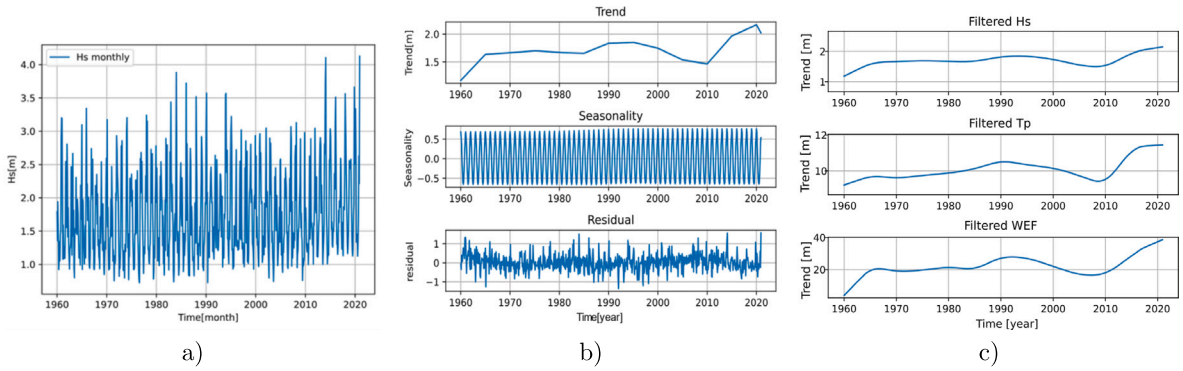


Fig. 8. Decomposition analysis of the H_s signal: (a) raw H_s, (b) three modes of the decomposition for wave heights, and (b) decomposition-based trends for wave height, period and energy flux.

and exceptional random events, as described in Section 2.2. Fig. 7 illustrates monthly-averaged H_s values computed based on the data extracted from the *SIMAR* model for the period 1960–2020, showing a highly fluctuating signal. In fact, due to these fluctuations, the trend identified in Fig. 6(b) for the WEF is hardly detectable for H_s.

Therefore, an adequate identification of the trend should isolate the seasonality and randomness of the historical dataset. To that end, the EMD method is applied, extracting the main modes as illustrated in Fig. 8(b) for H_s: trend (top), seasonality (middle) and residual (bottom). In this case, the trend shows minor oscillations which can be smoothed by filtering the trend itself via a second decomposition. Fig. 8(c) illustrates the smooth trends for H_s, T_p and WEF with a second filtering. On the other hand, the seasonality shows clearly the annual pattern of the wave resource, where the wave energy is mildest in summer and highest in winter. Finally, the residual shows the remaining part of the signal including the wave resource randomness.

As a consequence, as shown in previous studies, an increasing trend of the wave resource at Bay of Biscay is confirmed by an alternative statistical approach. In addition, this alternative approach particularly fits the requirements of the predictive models, *i.e.* feature extraction. Accordingly, the evolution of extreme events is analysed, which is a crucial aspect of the wave resource for the design of MRE systems. To that end, the lower and upper envelops are computed using minimum and maximum values, as presented in Fig. 7 for H_s. Although the lower envelop considerably increases in the final decade between 2010–2020, it is hard to identify a clear trend. Conversely, the upper envelop shows a clearer increasing trend. Fig. 9 top graph shows the annual maximum H_s values, and first- and second-order regression functions where that increasing trend is easily identified.

Yet, the frequency of the extreme events, defined as the occurrence of events exceeding a threshold value, is as important as the absolute value of maximum values of the extreme events. Accordingly, the sea-states are divided into four modes, with mild ($0 < H_s < 3$), intermediate ($3 < H_s < 5$), harsh ($5 < H_s < 8$) and extreme ($H_s > 8$) sea-states, and the conditioned probability of different states is studied during the last six decades, as presented in the middle graph in Fig. 9. The mild sea-state is clearly predominant, which is consistent with Fig. 5(a), and no clear trend can be identified. The lower graph in Fig. 9 zooms in the final mode, which is imperceptible in the middle graph. However, a clear trend cannot be identified either for this final mode. Lastly, based on the correlation between SST and WEF suggested by [10], a brief analysis is included in order to evaluate whether SST can be used as an input for the predictive models. The top graph in Fig. 10(a) compares the SST anomaly in the North Atlantic (based on the data provided by the European Environment Agency) and the Bay of Biscay, where a clear warming can be observed in the recent decades. On the

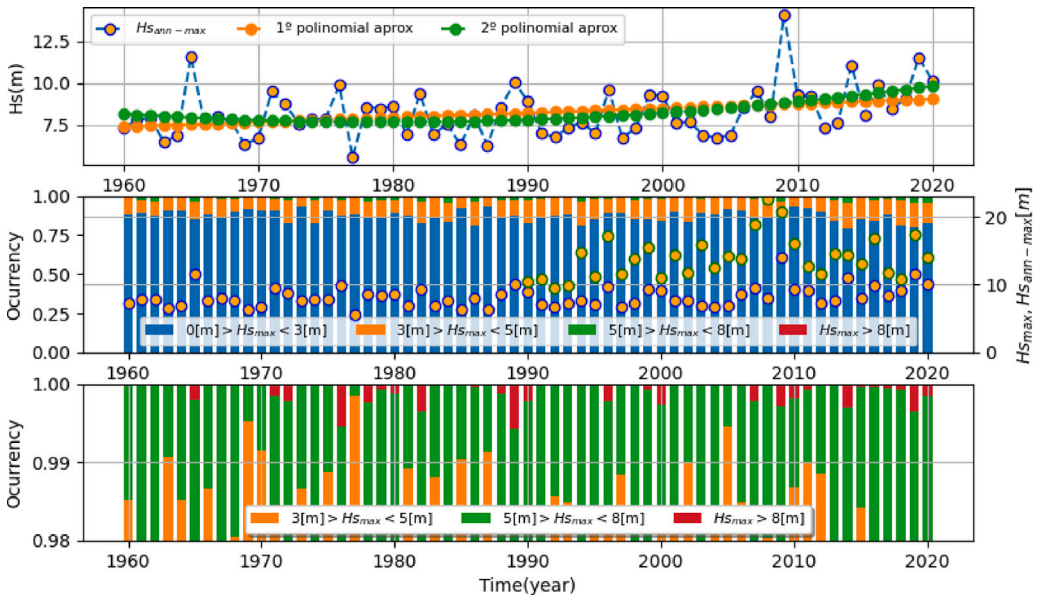
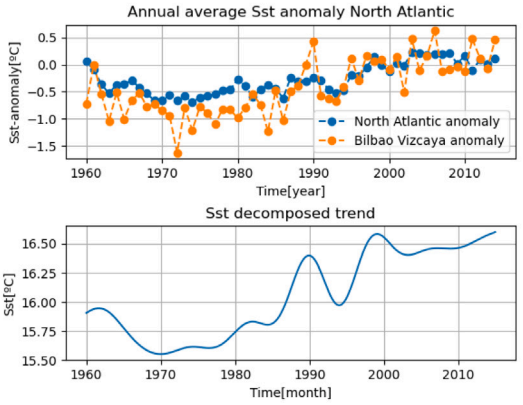
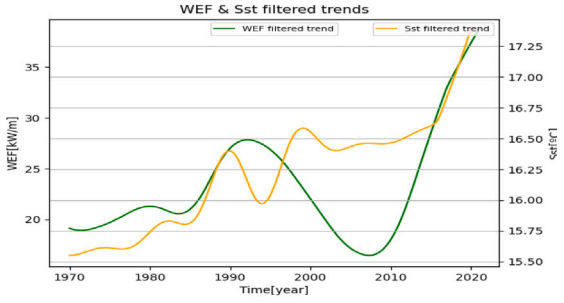


Fig. 9. Wave height analysis along the decades: (top) maximum wave heights, (middle) conditioned probability distribution of different modes with (bottom) a zoom at the most extreme events.



a) SST variations in the North Atlantic and the Bay of Biscay.



b) SST and WEF along the decades.

Fig. 10. Study of the correlation between SST and WEF along the decades.

other hand, the bottom graph shows the SST trend computed via the EMD, removing the influence of seasonality and residuals. As in [10], Fig. 10(b) shows the correlation between SST and WEF in the Bay of Biscay, suggesting that the evolution of the SST can be an appropriate input parameter for the predictive models.

5. Feature processing

The characterisation of the resource aims to understand the different trends along the previous decades, and identify the factors and features that can potentially aid on the prediction of future values. The correlation heatmap in Fig. 11 synthesises the linear relation between features. Some of the features are shown to have a very low correlation with H_s and, accordingly, they are discarded as inputs for the predictive models. For example, kurtosis and skewness values, the mode, the mean-to-peak parameter and all the derivatives show correlations of below 70%, which is considered to be in low relationship with H_s. In contrast, features like root-mean-square, standard deviation and percentiles, show a high correlation, being potentially valid parameters for the predictive modelling stage.

For the selection of the most relevant features, the Recursive Feature Elimination (RFE) technique is employed, which requires the pre-selection of a potentially relevant set of features. This pre-selection of features includes variables extracted from the metocean dataset, such as minimum, maximum, seasonality and long-term trend; the SST and its trend; and the different statistical features extracted from H_s, excluding those that have a very low correlation.

For the feature selection process via the RFE technique, the 100% of the dataset is employed, including the whole period between 1960–2020. That way, the idea is reconstructing the dataset with the minimum possible information, identifying the variables and features that provide that information. This analysis is carried out with a linear regression (LR), a RF and a SVR model. Unfortunately, the ANN models also described in Section 2.3 do not allow for a similar analysis. Fig. 12(a) shows the cross-validation of the dataset reconstruction as a function of the number of features considered in the reconstruction and using the NMAE as the scoring metric. The SVR algorithm seems to be the most efficient one, accurately reconstructing the signal with only 3 variables and features. The RF algorithm shows a similar performance, providing good results with 4 features. Finally, the LR model needs, at

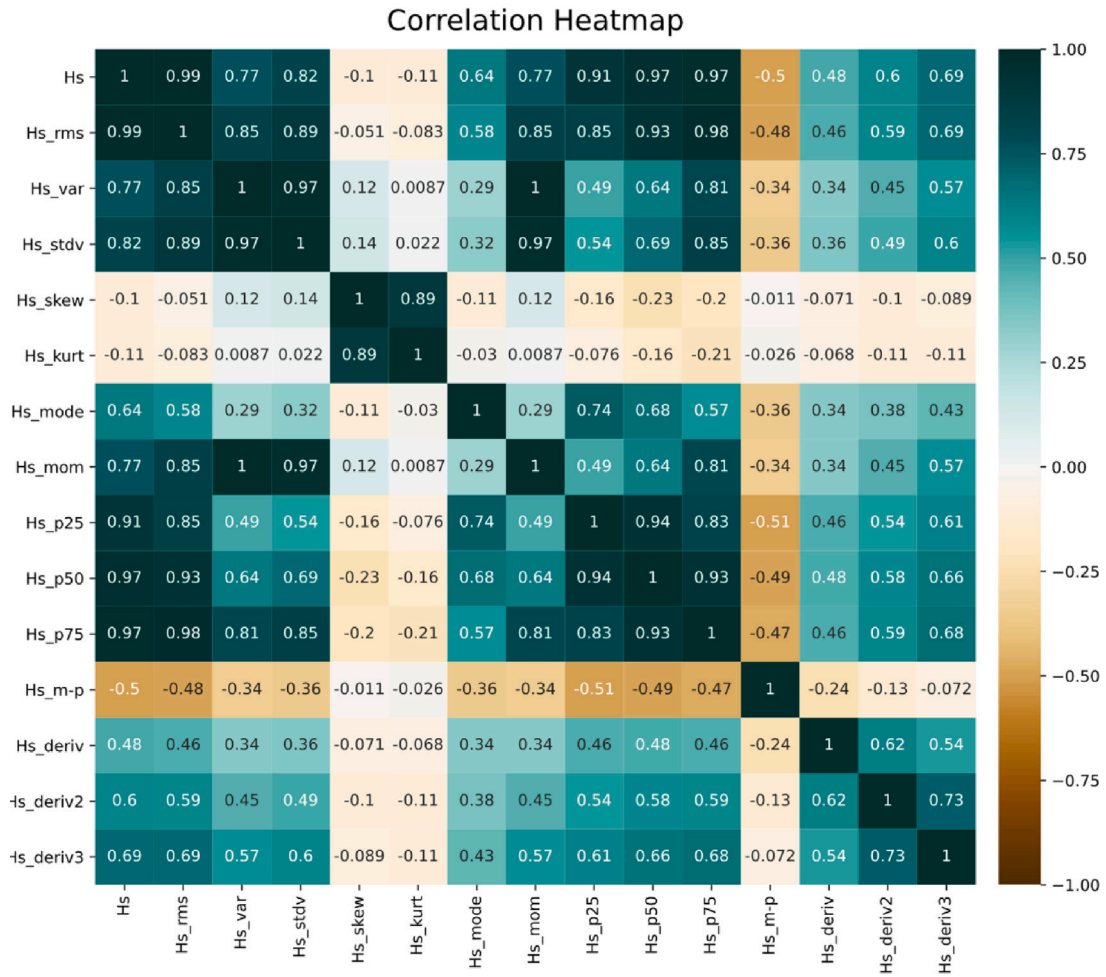
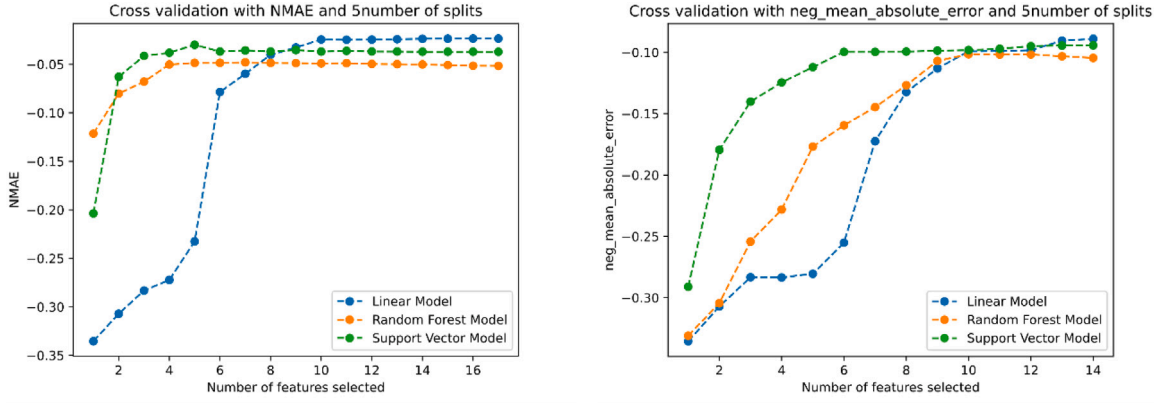


Fig. 11. Correlation heatmap of the different features.



(a) Using percentiles. (b) Neglecting percentiles.

Fig. 12. Evaluation of the NMAE for LR, RF and SVR models as a function of the number of considered features.

least, 8 features to provide an accuracy similar to that of the SVR and RF models.

To achieve the maximum accuracy, the LR model requires all the features, while 7 and 5 are needed for the RF and SVR models, respectively. The three percentiles are needed in all cases, since the combination of the 25%, 50% and 75% percentiles provides a very complete set of information. However, the use of percentiles in the prediction can be problematic due to the difficulty to estimate future percentiles. Therefore, the same exercise is carried out discarding the percentiles, which significantly increases the minimum number of features required to achieve a decent reproduction of the dataset: 9 in the case of LR and RF models, and 5 for the SVR. Once again, the SVR model seems to be the most efficient.

Therefore, the importance scores of the SVR model are computed, showing the most relevant features. Fig. 13 shows the importance of each variable, highlighting the 5 most important features: STD, minimum, mode, maximum and seasonality. The LR and RF models also show these same features among the most important ones. In addition,

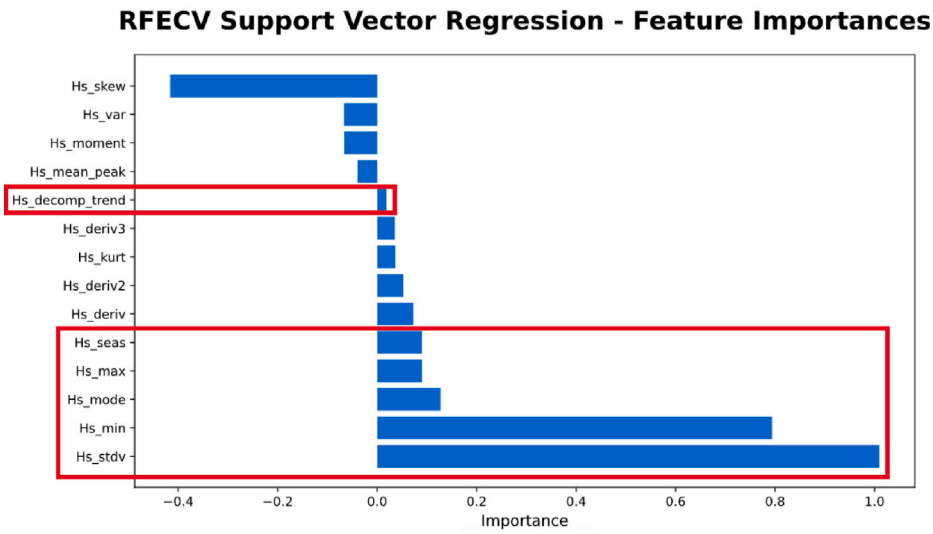


Fig. 13. Importance score of the SVR model.

although the importance of the trend seems to be considerably low for the re-construction of the dataset, the long-term variations identified for H_s and SST suggest that the projection of the trends into the future may be key in the correct prediction of future H_s values. Therefore, as highlighted in Fig. 13, different projections of these trends are also considered together with the most relevant features.

6. Predictive modelling results

After inferring the most relevant features from the Bay of Biscay dataset, the dataset is divided into training and testing sets in order to design ML models through a parameter-tuning process (cf. Section 2.3), and evaluate their predictive power. Two different prediction strategies are suggested. On the one hand, a traditional prediction method that attempts to forecast exact metocean data based upon the raw data, described in Section 6.1. On the other hand, an alternative discretised approach that intends to forecast the category of the metocean data previously classifying raw data into different ranks, as presented in Section 6.2.

In both cases, the training set consists of the first 80% of the dataset, which covers the time period between 1960 and 2008. The training set is used to fine-tune the algorithms identifying the hyperparameters for each algorithm that best represent the training data. The testing set consists of the rest 20% of the dataset, covering the most recent time period from 2008 to 2020. The testing set evaluates the predictive power of each algorithm using unseen data, which can be considered as a validation of the model. In this case, the accuracy of the predictive models is evaluated by means of the mean average error (MAE). More solid cross-validation strategies, such as k-fold cross validation [38], are left for future work.

The forecasting activity in the present paper focuses on wave heights, H_s, with the following input features: H_s seasonality, minimum H_s, maximum H_s, and H_s trend. The seasonality is expected to provide information about the annual seasonal pattern. Similarly, the minimum and maximum values are considered critical to inform the predictive model about peaks and valleys of the wave height signal. When evaluating the capacity of the predictive models to predict the H_s, the following assumptions are adopted:

1. The standard deviation will be repeated in the testing set.
2. Minimum H_s is considered that remains constant, *i.e.* the same as in the past.
3. The mode will be repeated in the testing set.
4. Maximum H_s is considered that increases as shown in Fig. 9(a)
5. The observed seasonality will be repeated in the testing set.

Finally, the trend is assumed to provide relevant predictive information so that the ML model can project the resource to the future informed by the long-term trends of the past. The following different H_s trend configurations are tested:

- `Mean`: H_s remained constant over the following decades.
- `Lin`: H_s increased following the linear regression.
- `Trend`: H_s increased following the decomposition-based trend.
- `SST`: H_s increased related to the increase in SST.

6.1. Predictive regression model

Raw data from the *SIMAR* model is used directly in the case of the traditional forecasting model, where the three methods presented in Section 2.3 are compared. An hyperparameter tuning exercise is carried out in each model in order to define the optimal number of layers and neurons. In the case of the ANN, the optimal configurations are obtained with 13 neurons (1 intermediate layer), 18 neurons (2 intermediate layers), 19 neurons (3 intermediate layers) and 19 neurons (4 intermediate layers). Overall, the lowest MAE is obtained for the model with a single intermediate layer which is selected for subsequent testing purposes.

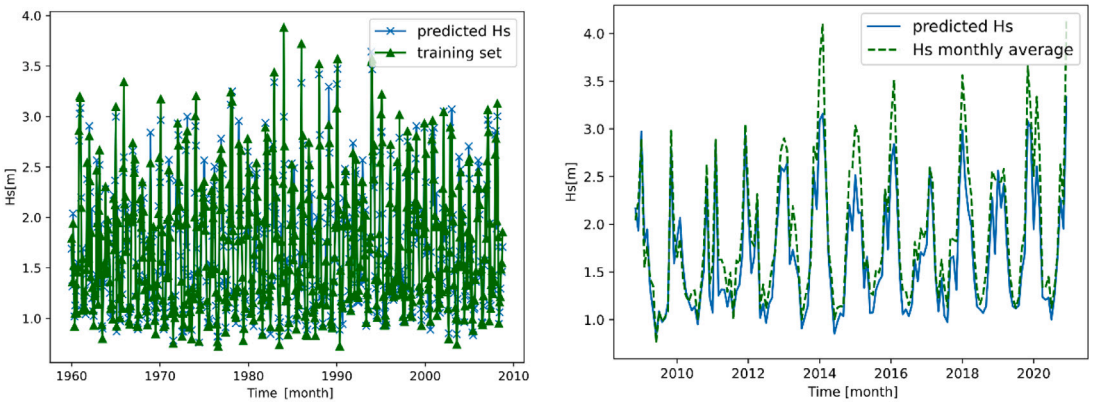
The hyperparameter tuning process for RF models focuses on searching the best parameters from a predefined grid (see also Section 2.3):

- n_estimators = [200:10:2000],
- max_depth = [10:10:110],
- min_samples_split = [1, 5, 10],
- min_samples_leaf = [1, 2, 4]

Similarly, the hyperparameters for the RBF kernel of the SVR method are fitted through grid search so as to optimise c and γ within c = [1e−1:10:1e3] and γ = [2e−5:2:2e2]. Table 2 summarises the MAE results for the different ML models, highlighting the most accurate ones for the training and testing periods.

Among all the testing results, the lowest MAE is obtained with the SVR model with the `mean` H_s configuration. As for the training set, the best configuration is obtained with the RF model with the `SST` configuration. Accordingly, Fig. 14(a) and (b) show the prediction results on the training set for the RF model and the testing set for the SVR model, respectively.

All the analysed predictive models provide very similar predictive capacity for the training and testing sets. In fact, all the models seem



(a) Training dataset (b) Validation dataset

Fig. 14. Prediction results for different input data configurations for the (a) training set via RF; (b) testing set via SVR.

Table 2
MAE for different models and configurations.

Config	*RF - train*	*RF - test*	*SVR - train*	*SVR - test*	*ANN - train*	*ANN - test*
Trend	0.137	0.225	0.174	0.225	0.235	0.239
Mean	0.144	0.229	0.175	**0.216**	0.19	0.235
Lin	0.139	0.233	0.173	0.219	0.186	0.254
SST	**0.135**	0.234	0.169	0.237	0.185	0.231

Table 3
Definition of the different discrete H_s ranks.

H_s	0.5 m	1 m	1.5 m	2 m	2.5 m	3 m	3.5 m	4 m	4.5 m
Rank # ($\Delta = 0.5$ m)	1	2	3	4	5	6	7	8	9
Rank # ($\Delta = 1$ m)	1		2		3		4		5
Rank # ($\Delta = 2$ m)	1				2				3

to have the same strengths and weak points. Although the three predictive models seem to accurately reproduce the overall behaviour of the dataset, all the models have difficulties in reproducing the peaks of the signal. Note that these peaks do not correspond to single extreme events or a single sea-state, but the mean of all sea-states every month. The peaks appear due to the seasonality of the dataset, showing lower energetic sea-states in Summer and higher energetic sea-states in Winter. Therefore, the capability to reach these peaks is crucial to accurately forecast future metocean conditions. In fact, the models seem to have problems to reach these peaks even in the training period, as illustrated in Fig. 14(a), showing a potential limitation of the mathematical structure of the selected predictive models. However, the reason why these predictive models fail to accurately reproduce the peaks is not determined yet and will need further investigation. However, an alternative method to overcome this difficulty is also presented in the following section.

6.2. Predictive classification model

Due to the difficulties found with the traditional exact wave height prediction approach, an alternative wave height interval prediction problem is suggested here. It is expected that turning an exact value prediction problem into a classification problem will address the difficulties in the prediction of exact peak values. Changing the regression problem into a classification problem requires the modification of the training procedure and the error quantification metrics. The most widely used metrics for classification are accuracy ($\mathcal{A}$) and precision. The latter provides the capacity of the classification model to estimate the correct rank, while the accuracy measures how close or far off the estimation is to the true value, which is defined as follows:

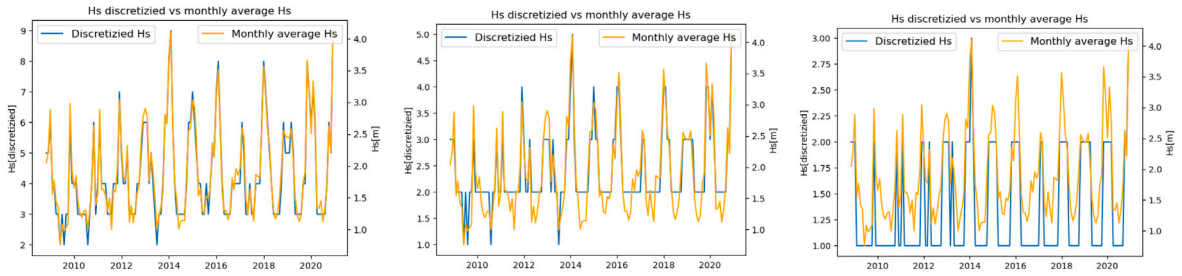
$$\mathcal{A}(y,\hat{y}) = \frac{\sum_{i=0}^{nb_{samples}-1} 1(\hat{y}_i = y_i)}{n_{samples}}, \tag{9}$$

where $\hat{y}_i$ is the estimates value and y_i the ground truth. The accuracy can be computed for different error ranges, and accordingly, the present study redefines the accuracy metrics for different scenarios: $\mathcal{A}cc1$ allows an error of ± 1 rank and $\mathcal{A}cc2$ allows an error of ± 2 ranks. In addition to the metrics, the monthly averaged H_s values must be discretised. In this study three different discretisation intervals are suggested (0.5 m, 1 m and 2 m), as described in Table 3 and depicted in Fig. 15.

Hence, this alternative predictive model only needs to determine the rank of future wave heights, which is in line with the objective of determining future occurrence matrices (also organised in ranks). For the preliminary classification analysis of this alternative approach, the predictive method that showed the best performance in Section 6.1 is selected and adapted: the Support Vector Classifier (SVC), equivalent to the SVR for the classification problem, based on a RBF kernel architecture. Similarly to the regression model, the hyperparameter optimisation is carried out via an extensive grid search, preparing the model to provide discretised predictions from a continuous dataset.

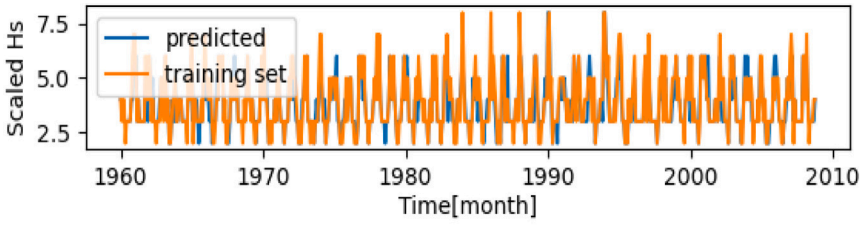
Based on the three discretisation ranges defined in Table 3, Fig. 16 shows the performance of the SVC model in its different stages and configurations. Fig. 16(a) shows the performance of the predictive model with the training set, where the capacity of the SCV model to reach the peaks of the discrete signal is demonstrated. In this sense, turning the prediction problem into a classification problem seems to partially solve the problem encountered in Section 6.1.

However, the prediction performance in the testing set is similar to the regression models, where the problem to reach the peaks still remains. Fig. 16(b) illustrates the comparison between the predicted and true values for $\Delta = 0.5$ m in the testing set. Despite the adequate representation of the overall pattern, the discretised predictive model fails to accurately predict the rank of future wave height. Yet, the confusion matrix presented in Fig. 16(c) for the $\Delta = 0.5$ m illustrates that when the predictive model fails to predict the correct rank, only fails by one rank in most of the cases. In fact, although the accuracy for the prediction of the exact rank ($\mathcal{A}cc0$) is below 0.6, the results for $\mathcal{A}cc1$ and $\mathcal{A}cc2$ improve significantly, reaching accuracy values of 0.92 and 0.99, respectively. Fig. 16(b) illustrates the samples for which the SVC prediction fails, where all mispredictions arise at maximum peaks, both for $\mathcal{A}cc1$ and $\mathcal{A}cc2$. Finally, although a wider discretisation can improve the accuracy of the predictor, it should be noted that the uncertainty within the predicted range also increases. Therefore, wider discretisation ranges have not been ignored in the present study.

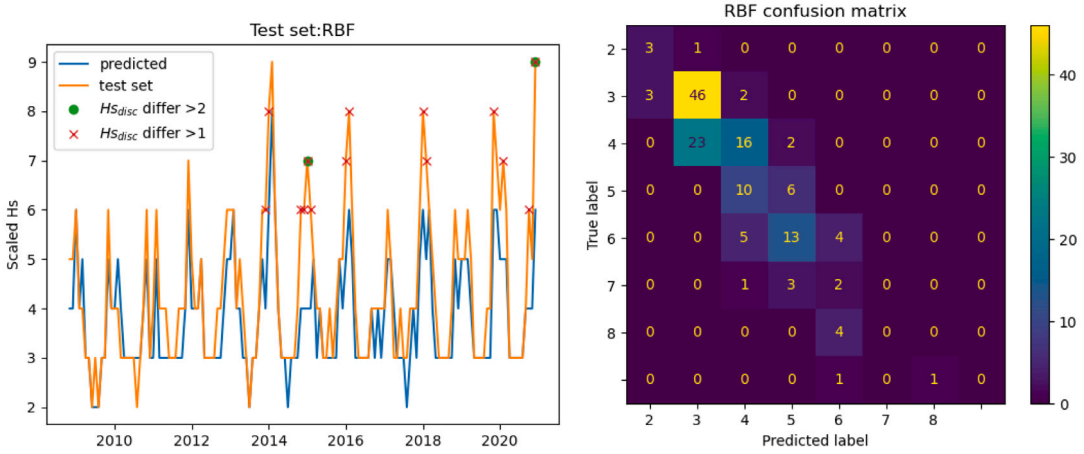


(a) Discretised H_s: Δ=0.5 m. (b) Discretised H_s: Δ=1 m. (c) Discretised H_s: Δ=2 m.

Fig. 15. Exact vs. discretised H_s dataset for the testing part (2008–2020).



(a) Prediction for $\Delta = 0.5$ m in the training set.



(b) Prediction for $\Delta = 0.5$ m in the testing set. (c) Confusion matrix.

Fig. 16. Performance of the SVC discrete forecasting model.

7. Discussion

The final objective of the proposed data-driven forecasting approach is to estimate probabilistically the long-term future metocean conditions of a local area. In this manner, developers will be able to accurately design the different MRE systems, reducing the current uncertainties in the load estimations and the excessive conservatism of the design process [39]. However, these future predictions should include the impact of climate change, particularly on the increase of the number and frequency of storms and extreme events. Fig. 17 depicts the strategy to design a ML-based forecasting model that provides information about the metocean conditions for the next 20–30 years (lifetime of a MRE system).

Yet, the results of the present study show that the objective of developing accurate site-specific wave height prediction models based on past metocean data is remarkably challenging. Apart from a traditional exact wave height prediction approach, an alternative wave height interval prediction method has also been suggested in the present paper. Particularly, the exact wave height prediction approaches showed difficulties to reproduce peak values of the signals. These problems appear both in the training and testing sets of the dataset, illustrating a significant limitation of the models analysed in this study. Changing the problem into a wave height classification problem, it was expected that prediction accuracy results would improve. The results in the training set showed that the SVC model is able to reproduce the peaks, indicating that the structural limitation of the exact value prediction models disappeared. However, the prediction results of the SVC model in the testing set do not show significant improvements compared to the predictive models based on the exact value prediction. It should be noted, though, that depending on the wave height groups, *i.e.* discretisation range, the prediction error is shown to be different. A more comprehensive sensitivity analysis in this direction is recommendable, in order to show the strengths and weaknesses of different models under different wave height prediction problems. In any case, a more thorough study of the alternative classification approach is expected to improve significantly the results shown in this paper and is the first future line the authors will follow.

Two main additional areas for improvement are also identified: (i) discretisation of the input variables used for the classification approach (H_s seasonality, minimum H_s, maximum H_s and H_s trend); and (ii)

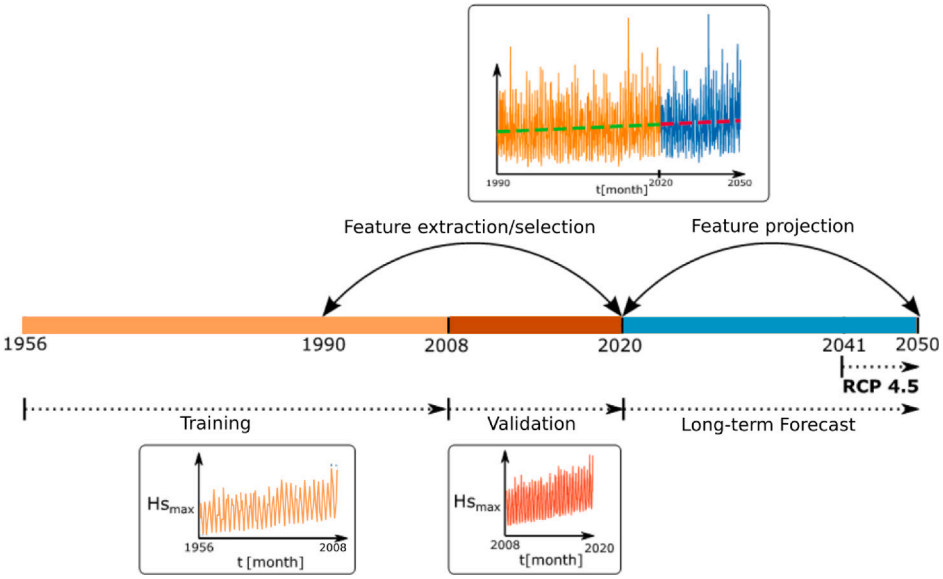


Fig. 17. Long-term forecasting approach.

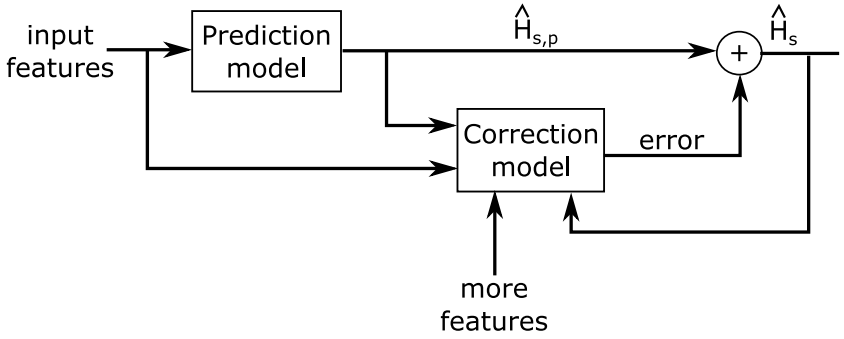


Fig. 18. Suggestion of a hybrid prediction approach.

integration of different classifiers through ensemble modelling strategies that allow handling different groups of H_s by means of different models [40]. In fact, the final objective of the long-term metocean data prediction is not the exact prediction of the metocean data at specific points in time, but the determination of the future probability density functions, *i.e.* $H_s - T_p$ joint occurrence matrix for the wave resource and the evolution of extreme events. Therefore, this alternative approach based on discrete intervals seems to be more appropriate.

Another alternative for future studies could be a hybrid prediction model that integrates a statistical prediction model with a long-term data-driven correction model [41]. Fig. 18 shows the potential hybrid configuration approach. The proposed hybrid prediction approach should focus on the integration of white-box statistical prediction models with machine-learning models so that prediction errors are corrected through an outer feedback loop. The feasibility of the approach should be tested for different prediction horizons.

8. Conclusions

The present paper presents a preliminary analysis of long-term metocean data forecasting at a local scale for marine renewable energy applications. To that end, historical metocean characteristics are identified, extracting all the relevant information of the last decades since 1960 based on atmospheric re-analysis datasets. Once these historical re-analysis datasets are validated against *in-situ* buoy measurements, the long-term variations are studied, where a clear annual seasonal pattern and an increasing trend have been identified by means of the empirical decomposition method. Although both seasonality and trends have been identified for wave heights, periods and energy flux, the analysis is focused on the long-term wave height characterisation, for the sake of simplicity. Hence, a clear long-term trend has been identified for the average wave heights and maximum wave heights. The latter is believed to denote an increase in extreme events. However, the frequency of this extreme events, characterised by means of a conditioned probability analysis, seems to have no clear trend. Finally, an increasing long-term trend of the sea surface temperature, which has been related to the variations of wave energy in the literature, has been identified.

This characterisation of the historical metocean conditions is used for the development of machine-learning-based predictive models. Random Forest, Support Vector Regression and Artificial Neural Network models have been tested using different combinations of the input data. All models and input combinations have provided similar results, highlighting the special importance of considering the information about seasonality, maximum and minimum wave heights. In contrast, the long-term trend, represented in different manners, seems to be irrelevant for the forecasting of long-term metocean data. Besides the traditional predictive model, an alternative classification problem is suggested.

Based on the results of the present study, it is believed that the traditional approach of predicting precise wave height values within a considerably long horizon is not adequate. The main reason is the incapacity to reach the maximum peaks, at least, with the methods and input features evaluated in this study. In contrast, the alternative

classification approach is expected to have the potential to significantly improve the prediction of long-term future metocean data, particularly when $H_s - T_p$ joint predictions come into play.

List of abbreviations

MRE	Marine Renewable Energy
AR	Autoregressive
ANN	Artificial Neural Networks
GP	Gaussian Process
SWAN	Simulating WAves Nearshore
ML	Machine Learning
CNN	Convolutional Neural Network
SVR	Support Vector Regression
RF	Random Forest
RMSD	Root Mean Square Deviation
PC	Pearson Correlation
SRC	Spearman's Rank Correlation
EMD	Empirical Mode Decomposition
SST	Sea-Surface Temperature
NMAE	Normalised Mean Absolute Error
NMSE	Normalised Mean Square Error
RBF	Radial Basis Function
MLP	Multilayer Perceptron
ReLU	Rectifier Linear Unit
ECMWF	European Centre for Medium-Range Weather Forecasts
BV	Bilbao–Viscaya
WEF	Wave Energy Flux
RFE	Recursive Feature Elimination
LR	Linear Regression
SVC	Support Vector Classifier
C	Statistical characteristics
T	Statistical techniques

List of symbols

H_s	Significant wave height
T_p	Peak period
$\hat{y}_n$	A variable of the SIMAR model
y_n	A variable measured by a wave measuring buoy
N	Number of samples in the signal
cov	Covariance
$\sigma_{\hat{y}}, \sigma_y$	Standard deviation
μ_y	Mean of all samples
$_{rg}$	Rank variables
k	Kernel
x, x'	Training and testing data samples
γ	RBF width
$\|\|d\|\|$	Euclidean norm
T_e	Energy period
α	Wave period correction coefficient
$\mathcal{A}$	Accuracy metric

CRediT authorship contribution statement

Markel Penalba: Conceptualization, Methodology, Data curation, Writing – original draft. **Jose Ignacio Aizpurua:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Ander Martinez-Perurena:** Software, Visualization. **Gregorio Iglesias:** Conceptualization, Writing – review & editing.

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.

Acknowledgements

The authors gratefully acknowledge the Spanish agency Puertos del Estado for supplying metocean data from the SIMAR model and the Bilbao–Vizcaya measuring buoy. In addition, this material is based upon works supported by the Basque Government under the Grant No. KK-2022/00090 (ELKARTEK) and J. I. Aizpurua is funded by Juan de la Cierva Incorporacion Fellowship (Spanish State Research Agency - grant number IJC2019-039183-I).

References

[1] Ritchie H. Energy production and consumption. 2021, Available in https://ourworldindata.org/energy-production-consumption. (Last Accessed 29 April 2022).
[2] IPCC. Global warming of 1.5°C. Tech. Rep., Intergovernmental Panel on Climate Change (IPCC); 2018.
[3] IRENA. Future of wind: Deployment, investment, grid integration and socio-economic aspects (A Global Energy Transformation paper). Tech. Rep., Abu Dhabi: International Renewable Energy Agency; 2019, Available in https://www.irena.org/publications/2019/Apr/Global-energy-transformation-A-roadmap-to-2050-2019Edition.
[4] Ocean Energy Europe. 2030 ocean energy vision. Tech. Rep., 2020, URL https://www.oceanenergy-europe.eu/wp-content/uploads/2020/10/OEE_2030_Ocean_Energy_Vision.pdf.
[5] NREL. Marine energy in the United States: An overview of opportunities. Tech. Rep. February, 2021, URL https://www.nrel.gov/docs/fy21osti/78773.pdf.
[6] Ulazia A, Penalba M, Ibarra-Berastegui G, Ringwood J, Sáenz J. Reduction of the capture width of wave energy converters due to long-term seasonal wave energy trends. Renew Sustain Energy Rev 2019;113:109267. http://dx.doi.org/10.1016/j.rser.2019.109267, URL https://www.sciencedirect.com/science/article/pii/S1364032119304757.
[7] Martinez A, Iglesias G. Wave exploitability index and wave resource classification. Renew Sustain Energy Rev 2020;134:110393. http://dx.doi.org/10.1016/j.rser.2020.110393, URL https://www.sciencedirect.com/science/article/pii/S136403212030681X.
[8] Ahn S, Haas KA, Neary VS. Wave energy resource classification system for US coastal waters. Renew Sustain Energy Rev 2019;104:54–68. http://dx.doi.org/10.1016/j.rser.2019.01.017, URL https://www.sciencedirect.com/science/article/pii/S1364032119300279.
[9] Choupin O, Pinheiro Andutta F, Etemad-Shahidi A, Tomlinson R. A decision-making process for wave energy converter and location pairing. Renew Sustain Energy Rev 2021;147:111225. http://dx.doi.org/10.1016/j.rser.2021.111225, URL https://www.sciencedirect.com/science/article/pii/S1364032121005128.
[10] Reguero BG, Losada IJ, Méndez FJ. A recent increase in global wave power as a consequence of oceanic warming. Nature Commun 2019;10(205). http://dx.doi.org/10.1038/s41467-018-08066-0.
[11] Penalba M, Ulazia A, Ibarra-Berastegui G, Ringwood J, Sáenz J. Wave energy resource variation off the west coast of Ireland and its impact on realistic wave energy converters' power absorption. Appl Energy 2018;224:205–19. http://dx.doi.org/10.1016/j.apenergy.2018.04.121, URL https://www.sciencedirect.com/science/article/pii/S0306261918306895.
[12] Khojasteh D, Lewis M, Tavakoli S, Farzadkhoo M, Felder S, Iglesias G, et al. Sea level rise will change estuarine tidal energy: A review. Renew Sustain Energy Rev 2022;156:111855. http://dx.doi.org/10.1016/j.rser.2021.111855, URL https://www.sciencedirect.com/science/article/pii/S1364032121011229.
[13] Widén J, Carpman N, Castellucci V, Lingfors D, Olauson J, Remouit F, et al. Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. Renew Sustain Energy Rev 2015;44:356–75. http://dx.doi.org/10.1016/j.rser.2014.12.019, URL https://www.sciencedirect.com/science/article/pii/S1364032114010715.
[14] Fusco F, Ringwood JV. Short-term wave forecasting for real-time control of wave energy converters. IEEE Trans Sustain Energy 2010;1(2):99–106. http://dx.doi.org/10.1109/TSTE.2010.2047414.
[15] Peña Sanchez Y, Mérigaud A, Ringwood JV. Short-term forecasting of sea surface elevation for wave energy applications: The autoregressive model revisited. IEEE J Ocean Eng 2020;45(2):462–71. http://dx.doi.org/10.1109/JOE.2018.2875575.
[16] Mérigaud A, Herterich J, Flanagan J, Ringwood J, Dias F. Incorporating wave spectrum information in real-time free-surface elevation forecasting: Real-sea experiments. IFAC-PapersOnLine 2018;51(29):232–7. http://dx.doi.org/10.1016/j.ifacol.2018.09.508, 11th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2018. URL https://www.sciencedirect.com/science/article/pii/S2405896318321979.
[17] Shi S, Patton RJ, Liu Y. Short-term wave forecasting using Gaussian process for optimal control of wave energy converters. IFAC-PapersOnLine 2018;51(29):44–9. http://dx.doi.org/10.1016/j.ifacol.2018.09.467, 11th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2018. URL https://www.sciencedirect.com/science/article/pii/S2405896318321566.

[18] SWAN. SWAN scientific and technical documentation. Technical report swan cycle iii version 40.51, Delft University of Technology; 2006.
[19] James SC, Zhang Y, O'Donncha F. A machine learning framework to forecast wave conditions. Coast Eng 2018;137:1–10. http://dx.doi.org/10.1016/j.coastaleng.2018.03.004, URL https://www.sciencedirect.com/science/article/pii/S0378383917304969.
[20] Shamshirband S, Mosavi A, Rabczuk T, Nabipour N, wing Chau K. Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. Eng Appl Comput Fluid Mech 2020;14(1):805–17. http://dx.doi.org/10.1080/19942060.2020.1773932.
[21] Salcedo-Sanz S, Nieto Borge JC, Carro-Calvo L, Cuadra L, Hessner K, Alexandre E. Significant wave height estimation using SVR algorithms and shadowing information from simulated and real measured X-band radar images of the sea surface. Ocean Eng 2015;101:244–53. http://dx.doi.org/10.1016/j.oceaneng.2015.04.041, URL https://www.sciencedirect.com/science/article/pii/S0029801815001225.
[22] Cornejo-Bueno L, Nieto-Borge JC, García-Díaz P, Rodríguez G, Salcedo-Sanz S. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm – Extreme learning machine approach. Renew Energy 2016;97:380–9. http://dx.doi.org/10.1016/j.renene.2016.05.094, URL https://www.sciencedirect.com/science/article/pii/S0960148116305110.
[23] Ali M, Prasad R. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. Renew Sustain Energy Rev 2019;104:281–95. http://dx.doi.org/10.1016/j.rser.2019.01.014, URL https://www.sciencedirect.com/science/article/pii/S1364032119300243.
[24] Altunkaynak A. Prediction of significant wave height using geno-multilayer perceptron. Ocean Eng 2013;58:144–53. http://dx.doi.org/10.1016/j.oceaneng.2012.08.005, URL https://www.sciencedirect.com/science/article/pii/S0029801812003216.
[25] Breiman L. Random forests. Mach Learn 2001;45(1):5–32. http://dx.doi.org/10.1023/A:1010933404324.
[26] Meireles MRG, Almeida PEM, Simoes MG. A comprehensive review for industrial applicability of artificial neural networks. IEEE Trans Ind Electron 2003;50(3):585–601.
[27] Weiss CVC, Guanche R, Ondiviela B, Castellanos OF, Juanes J. Marine renewable energy potential: A global perspective for offshore wind and wave exploitation. Energy Convers Manage 2018;177:43–54. http://dx.doi.org/10.1016/j.enconman.2018.09.059, URL https://www.sciencedirect.com/science/article/pii/S0196890418310628.
[28] NOAA. National Oceanographic and Atmospheric Agency, https://www.noaa.gov/.
[29] Puertos del Estado. Prediccion de oleaje, nivel del mar; Boyas y mareografos, URL http://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx.
[30] Ruggiero P, Komar PD, Allan JC. Increasing wave heights and extreme value projections: The wave climate of the US Pacific Northwest. Coast Eng 2010;57(5):539–52.
[31] Young IR, Zieger S, Babanin AV. Global trends in wind speed and wave height. Science 2011;332(6028):451–5.
[32] Bertin X, Prouteau E, Letetrel C. A significant increase in wave height in the North Atlantic Ocean over the 20th century. Glob Planet Change 2013;106:77–83. http://dx.doi.org/10.1016/j.gloplacha.2013.03.009.
[33] Zheng C, Shao L, Shi W, Su Q, Lin G, Li X, et al. An assessment of global ocean wave energy resources over the last 45 a. Acta Oceanol Sin 2014;33(1):92–101. http://dx.doi.org/10.1007/s13131-014-0418-5.
[34] Reguero BG, Losada IJ, Méndez FJ. A global wave power resource and its seasonal, interannual and long-term variability. Appl Energy 2015;148:366–80. http://dx.doi.org/10.1016/j.apenergy.2015.03.114.
[35] Ulazia A, Penalba M, Ibarra-Berastegui G, Ringwood J, Saénz J. Wave energy trends over the Bay of Biscay and the consequences for wave energy converters. Energy 2017;141. http://dx.doi.org/10.1016/j.energy.2017.09.099.
[36] Agarwal A, Venugopal V, Harrison GP. The assessment of extreme wave analysis methods applied to potential marine energy sites using numerical model data. Renew Sustain Energy Rev 2013;27:244–57. http://dx.doi.org/10.1016/j.rser.2013.06.049, URL https://www.sciencedirect.com/science/article/pii/S1364032113004401.
[37] Tucker MJ, Pitt EG. Waves in ocean engineering. Elsevier ocean engineering series, Elsevier Science; 2001, URL https://books.google.es/books?id=0v5RAAAAMAAJ.
[38] Aizpurua JI, McArthur SDJ, Stewart BG, Lambert B, Cross JG, Catterson VM. Adaptive power transformer lifetime predictions through machine learning and uncertainty modeling in nuclear power plants. IEEE Trans Ind Electron 2019;66(6):4726–37. http://dx.doi.org/10.1109/TIE.2018.2860532.
[39] Penalba M, Aizpurua JI, Martinez-Perurena A. On the definition of a risk index based on long-term metocean data to assist in the design of marine renewable energy systems. Ocean Eng 2021;242:110080. http://dx.doi.org/10.1016/j.oceaneng.2021.110080, URL https://www.sciencedirect.com/science/article/pii/S0029801821014086.
[40] Aizpurua JI, Catterson VM, Stewart BG, McArthur SDJ, Lambert B, Cross JG. Uncertainty-aware fusion of probabilistic classifiers for improved transformer diagnostics. IEEE Trans Syst Man Cybern A 2021;51(1):621–33. http://dx.doi.org/10.1109/TSMC.2018.2880930.
[41] Guo J, Li Z, Li M. A review on prognostics methods for engineering systems. IEEE Trans Reliab 2020;69(3):1110–29. http://dx.doi.org/10.1109/TR.2019.2957965.