

Article

Use of Machine-Learning Techniques to Estimate Long-Term Wave Power at a Target Site Where Short-Term Data Are Available

María José Pérez-Molina ^{1,2}  and José A. Carta ^{2,*} 

¹ Oceanic Platform of the Canary Islands (PLOCAN), 35214 Telde, Spain; mariaj.perez@plocan.eu

² Group for the Research on Renewable Energy Systems (GRRES), Department of Mechanical Engineering, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

* Correspondence: jose.cart@ulpgc.es

Abstract: Wave energy is a promising renewable resource supporting the decarbonization of energy systems. However, its significant temporal variability necessitates long-term datasets for accurate resource assessment. A common approach to obtaining such data is through climate reanalysis datasets. Nevertheless, reanalysis data may not accurately capture the local characteristics of wave energy at specific sites. This study proposes a supervised machine-learning (ML) approach to estimate long-term wave energy at locations with only short-term in situ measurements. The method involves training ML models using concurrent short-term buoy data and ERA5 reanalysis data, enabling the extension of wave energy estimates over longer periods using only reanalysis inputs. As a case study, hourly mean significant wave height and energy period data from 2000 to 2023 were analyzed, collected by a deep-water buoy off the coast of Gran Canaria (Canary Islands, Spain). Among the ML techniques evaluated, Multiple Linear Regression (MLR) and Support Vector Regression yielded the most favorable error metrics. MLR was selected due to its lower computational complexity, greater interpretability, and ease of implementation, aligning with the principle of parsimony, particularly in contexts where model transparency is essential. The MLR model achieved a mean absolute error (MAE) of 2.56 kW/m and a root mean square error (RMSE) of 4.49 kW/m, significantly outperforming the direct use of ERA5 data, which resulted in an MAE of 4.38 kW/m and an RMSE of 7.1 kW/m. These findings underscore the effectiveness of the proposed approach in enhancing long-term wave energy estimations using limited in situ data.

Keywords: Measure–Correlate–Predict; wave energy; machine learning; reanalysis data; wave period; significant wave height



Academic Editors: Rodrigo Carballo Sánchez, Iván López Moreira, Mario López Gallego, Carlos Pérez-Collazo and Victor Ramos

Received: 22 May 2025

Revised: 16 June 2025

Accepted: 17 June 2025

Published: 19 June 2025

Citation: Pérez-Molina, M.J.; Carta, J.A. Use of Machine-Learning Techniques to Estimate Long-Term Wave Power at a Target Site Where Short-Term Data Are Available. *J. Mar. Sci. Eng.* **2025**, *13*, 1194. <https://doi.org/10.3390/jmse13061194>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the International Renewable Energy Agency (IRENA), the need to use renewable energy in any transition toward a climate-friendly future is indisputable [1]. As IRENA reports, solar and wind energy continue to dominate the expansion of renewable capacity, together accounting for 96.6% of all net renewable additions in 2024. In addition, 2024 represents the year with the highest annual increase in renewable power generation capacity, as well as the greatest growth recorded—mainly solar energy—in percentage terms. Within the spectrum of renewable energy sources (solar, wind energy, hydropower, bioenergy, geothermal, and marine), IRENA identifies marine energy as the one that experienced the smallest increase in capacity in 2024 [2].

In this context, according to IRENA [3], “ocean energy is among the technologies that must be scaled up to achieve full decarbonization of the energy system”. This statement refers to the set of technologies designed to harness ocean energy as a resource.

With a projected global market potential of 350 gigawatts (GW) by 2050 [4], ocean energy can deliver clean, local, and predictable electricity to coastal countries and island communities worldwide. Within the broader category of ocean energy, wave energy—or the kinetic and potential energy of the ocean surface, captured and harnessed primarily by wave energy converters [5]—is abundant and offers several advantages: strong predictability, low intermittency, high energy density, and wide availability [6]. According to Astariz and Iglesias [7], wave energy is undoubtedly one of the most promising renewable sources. Moreover, there are islands and regions where the potential for land-based renewables is limited [8]. In such areas, wave energy may prove especially valuable, as reflected in a growing body of literature on the subject [8].

The feasibility of implementing a wave energy conversion system at a target site depends on multiple technical, socioeconomic, and environmental factors [9]. Among these, the available energy at the site stands out, as it is a key parameter in determining the levelized cost of electricity [7,8].

Given the inter- and intra-annual variability of wave energy highlighted by various authors [10–14], long-term wave data series are required to estimate the average characteristics of the resource over the lifetime of a wave energy project. As noted by Sun and Wang [5], the long-term variability of wave energy resources will affect the practical decadal deployment of wave energy converters. Access to extended wave datasets can help to reduce the degree of uncertainty in estimating the levelized cost of electricity during the planning phase of a wave energy project.

Given the difficulty of obtaining long-term records of wave data at target sites, reanalysis data are commonly used as an alternative. One of the most widely used reanalysis sources, as noted by Bessonova et al. [15], is the European Centre for Medium-Range Weather Forecasts Reanalysis (ERA5) [16]. In this context, Tong et al. [11] used ERA5 reanalysis to analyze wave energy resources in the South China Sea from 1 January 1979 to 31 December 2024. Similarly, Silva et al. [10] used ERA5 wind–wave data (1950–2020) to estimate the inter- and intra-annual wave energy variability along the northern coast of mainland Portugal. Ulazia et al. [17] employed hourly ERA5 data from 1981 to 2020 (a 40-year period) to estimate wave energy potential in the Canary Islands. Liu et al. [14] evaluated the long-term variability of wave power using ERA5 data spanning from 1940 to 2022. Mahmoodi et al. [18] analyzed the spatial and temporal characteristics of wave energy in the Persian Gulf based on ERA5 reanalysis data over an 18-year period (2000–2017), reporting clear evidence of seasonal variability in wave energy.

In order to assess how well reanalysis data represent measurements taken at target sites, several studies have been conducted. Silva et al. [10] compared ERA5 reanalysis data (10 January 2012–28 December 2019) with Leixões buoy data (tri-hourly records). The Leixões buoy is located off the northwest coast of mainland Portugal. According to the authors, ERA5 reanalysis data slightly underestimate the higher values of wave power observed by the buoy. Tong et al. [11] evaluated the accuracy of ERA5 reanalysis data for estimating wave power in the South China Sea, using in situ observations from a buoy positioned in the central part of the southern South China Sea. The buoy data cover the period from 22 February to 2 October 2021, amounting to just over seven consecutive months. According to the authors, the results suggest that ERA5 reanalysis data are reliable for calculating wave power in the South China Sea. Bessonova et al. [15] performed a global evaluation of the ERA5 significant wave height (H_s) against measurements from 444 buoys

worldwide. Their results indicate that ERA5 tends to underestimate H_s in the upper range. This conclusion has also been reported by other authors [19–21].

Li et al. [22] compared wave energy estimated using reanalysis data and observational data from a buoy located in the central area of the southern South China Sea, covering a continuous 16-month period (14 December 2018–12 March 2020). According to the authors, the H_s and energy period (T_e) from ERA5 generally match the buoy observations in terms of variation trends. However, during the summer monsoon season (May to September 2019), the ERA5 H_s values were higher than the observations. The ERA5 T_e values were consistently higher than those from the buoy throughout the entire period. Since both H_s and T_e are essential for calculating wave energy density, the authors conclude that further calibration is needed to ensure accurate wave energy assessment. They use a feed-forward neural network with a single hidden layer of eight neurons to calibrate these two parameters. Ayuso-Virgili et al. [23], similarly to Li et al. [22], also propose calibrating two parameters to improve wave energy assessment. Specifically, they focus on the calibration of H_s and the peak period (T_p), applying so-called Measure–Correlate–Predict (MCP) methods, which have been extensively used in the renewable energy literature to estimate long-term wind conditions at target sites based on short-term wind measurement campaigns [24]. The authors employ the most used MCP method, which relies on algorithms based on linear functions [24].

The scientific literature reviewed shows that the wave energy resource exhibits both seasonal and inter-annual variability at the various locations studied. Furthermore, it has been demonstrated that discrepancies may arise between reanalysis data and in situ buoy measurements.

In this context, it is justified that when planning the installation of wave energy technologies at a target site, a long-term wave dataset should be analyzed. If such a dataset is not available for the site, reanalysis data may be used instead, provided their accuracy is improved. In this regard, several authors have proposed calibrating the parameters H_s and T_e [22], or H_s and T_p [23], to enhance the accuracy of wave energy assessment.

Other approaches, such as benchmarking against similar regional studies or using regionally refined coupled numerical models, have also been proposed to address this resolution gap (e.g., Zhu et al., 2023 [25]), although they fall outside the scope of this study.

1.1. Aims and Originality of This Paper

The aim of this paper is to propose a strategy for estimating long-term wave power at a target site where a short-term measurement campaign of significant wave height (H_s) and energy period (T_e) has been conducted—specifically, at least one year of hourly mean data—allowing for the characterization of the seasonal behavior of the resource during that period [24].

The originality and scientific contribution of this work are reflected in the following points, which summarize its contributions to the body of knowledge:

Twenty-four years (2000–2023) of hourly mean H_s and T_e data from ERA5 reanalysis are used to represent long-term wave conditions. However, these data are transformed to improve the accuracy of wave power estimation by applying MCP methods based on supervised machine-learning (ML) techniques.

A range of ML techniques are evaluated to determine which produces the best performance metrics— $RMSE$, MAE , and R^2 —on test datasets that are not used during model training or validation. The target variable in all models is wave power.

The ML techniques considered in this study include Random Forest (RF), k-Nearest Neighbor (KNN), Multiple Linear Regression (MLR), Support Vector Regression (SVR),

Artificial Neural Networks (ANNs) [26,27], and extreme gradient boost (XGBoost) [28]. For each technique, the optimal hyperparameters are selected.

To analyze the influence of the year used for training and validation, each of the twenty-four years of data is individually used as the short-term dataset, while the remaining years serve as the long-term test dataset. This analysis is conducted using observations from a buoy moored far from the coast, in deep open waters. As a result, the wave measurements from the buoy sensors are not affected by local coastal effects and can be considered representative of large coastal areas.

1.2. Structure of the Paper

Section 2 presents the wave data used in this study, namely the ERA5 reanalysis data and the measurements from a buoy moored off the northern coast of Gran Canaria (Canary Islands, Spain). The proposed method is described in detail in Section 3. Section 4 presents and analyzes the results obtained from applying the method to the target site represented by the buoy. Finally, Section 5 outlines the main conclusions drawn from this work.

2. Materials

This section briefly describes the rationale behind the selection of the study area and the data used in the analysis.

2.1. Background

The Canary Archipelago, an outermost region of the European Union, is composed of seven main small islands—Lanzarote, Fuerteventura, Gran Canaria, Tenerife, La Gomera, and El Hierro. It is located off the northwest coast of the African continent, between the latitudes 27°37' and 29°25' (subtropical zone) and the longitudes 13°20' and 18°10' west of Greenwich (Figure 1).

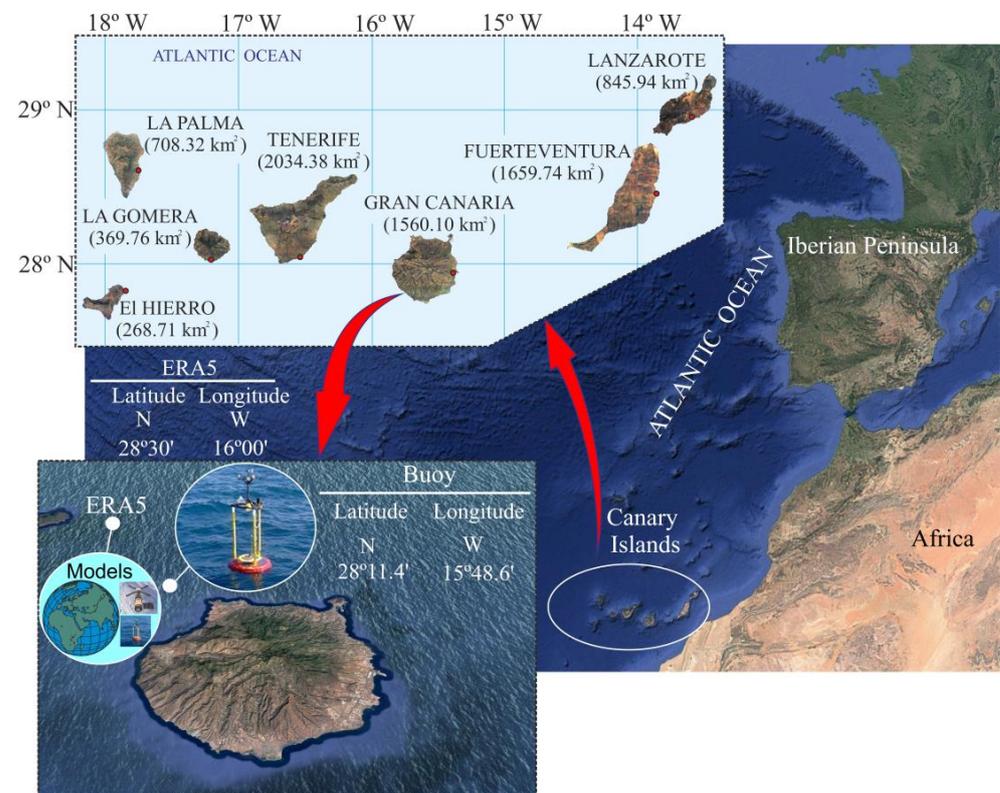


Figure 1. Location of the Canary Islands and position of the buoy and ERA5 point used in this study.

The Canary Archipelago presents certain peculiarities from an energy perspective [29]. The most notable characteristics in this regard are

- (a) Its geographical remoteness, which greatly hinders interconnection with the large energy supply networks of continental territories.
- (b) A lack of conventional energy sources, resulting in an almost total dependence on external supplies, primarily petroleum-based.
- (c) Its high wind and solar energy potential.

This set of circumstances has led the Canary Islands to develop their own energy strategy, which emphasizes the need to promote indigenous energy resources—namely, renewable energies. Given the technological maturity achieved by wind and solar power, these are currently the only sources being exploited on a large scale in the archipelago. However, other renewable sources could also be harnessed in the Canary Islands. Among them are energy sources potentially extractable from the Atlantic Ocean, in which the archipelago is located.

Historically, barriers to the exploitation of this type of energy in the Canary Islands have included the high installation and maintenance costs of these systems, as well as the limited reliability of early prototypes deployed globally—mainly due to damage caused by storms. However, at present, the technological improvements achieved by wave energy conversion devices [29–32], the reduction in associated costs, and the limited availability of land for wind energy development may make the installation of wave energy devices along the Canary coast feasible. To this end, prior studies on the potential of this energy source will be required.

In this regard, the Canary Islands' Energy Transition Plan (Plan de Transición Energética de CANarias—PTECAN¹) represents a major strategic commitment to achieving a sustainable, decarbonized, and self-sufficient energy system. One of its strongest pillars is the promotion of renewable energy sources—especially wind, solar, and marine energy—leveraging the archipelago's abundant indigenous resources. Given the islands' isolation from continental grids, the plan also prioritizes energy storage, smart grid systems, and hydrogen technologies to ensure supply reliability and environmental sustainability.

An essential actor in this transition is the Oceanic Platform of the Canary Islands (Plataforma Oceánica de CANarias—PLOCAN²), a public company dedicated to research, technological development, and innovation in the marine and maritime sectors. PLOCAN operates a multi-purpose offshore test site and platform (Figure 2), located off the east coast of Gran Canaria, that plays a key role in validating marine energy technologies under real-sea conditions.

The PLOCAN test site spans an area of 23 km² in waters ranging from 30 to 600 m deep. It is equipped with a dedicated submarine power and communications cable connected to the onshore electrical grid, allowing prototypes to be deployed, monitored, and even integrated into the grid during testing phases. This infrastructure enables the testing of wave energy converters, floating wind platforms, ocean current turbines, and hybrid systems combining multiple sources. PLOCAN, in line with PTECAN, exemplifies the region's commitment to becoming a global benchmark in island energy transition and marine renewable innovation.

2.2. Data Used

The buoy used in this study (code 2442), a SeaWatch type, belongs to the offshore network of Puertos del Estado [30] and is moored at a depth of 780 m off the northwest coast of Gran Canaria. The buoy's location (latitude: 28°11.4' N, longitude: 15°48.6' W; see Figure 1) qualifies it as being in deep waters, with its measurements largely unaffected by

shadowing effects—except for the northwest component, which may be distorted by the shadowing effect of Tenerife Island.



Figure 2. PLOCAN’s offshore test site and platform.

The ERA5 data correspond to the coordinates 28°30′ N, 16°00′ W (Figure 1), as this is the ERA5 grid point closest to the location of the Puertos del Estado buoy. The dataset consists of hourly mean values of H_s and T_e , covering the period from 1 January 2000 to 31 December 2023.

3. Method

A block diagram illustrating the proposed method, covering the process from data collection to results analysis, is shown in Figure 3.

3.1. Task 1 of the Method

The first step in the process consists of collecting data from the selected sources—namely, the H_s and T_e reanalysis data from ERA5 and the observed H_s and T_e data from the buoy considered in this study.

3.2. Task 2 of the Method

In this step, a comparison is performed between the reanalysis data and the buoy data, with the aim of identifying potential differences between them. The comparisons focus on the parameters H_s and T_e , as well as on the wave power (P_{wave}) which is estimated using Equation (1), commonly applied in several studies [11,13,23,31].

$$P_{wave} = \frac{\rho_{seawater} \cdot g^2}{64\pi} H_s^2 T_e \tag{1}$$

In Equation (1), ρ is the density of seawater (1025 kg/m³), and g is the acceleration due to gravity (9.81 m/s²). H_s is expressed in meters, T_e in seconds, and P_{wave} in kilowatts per meter (kW/m).

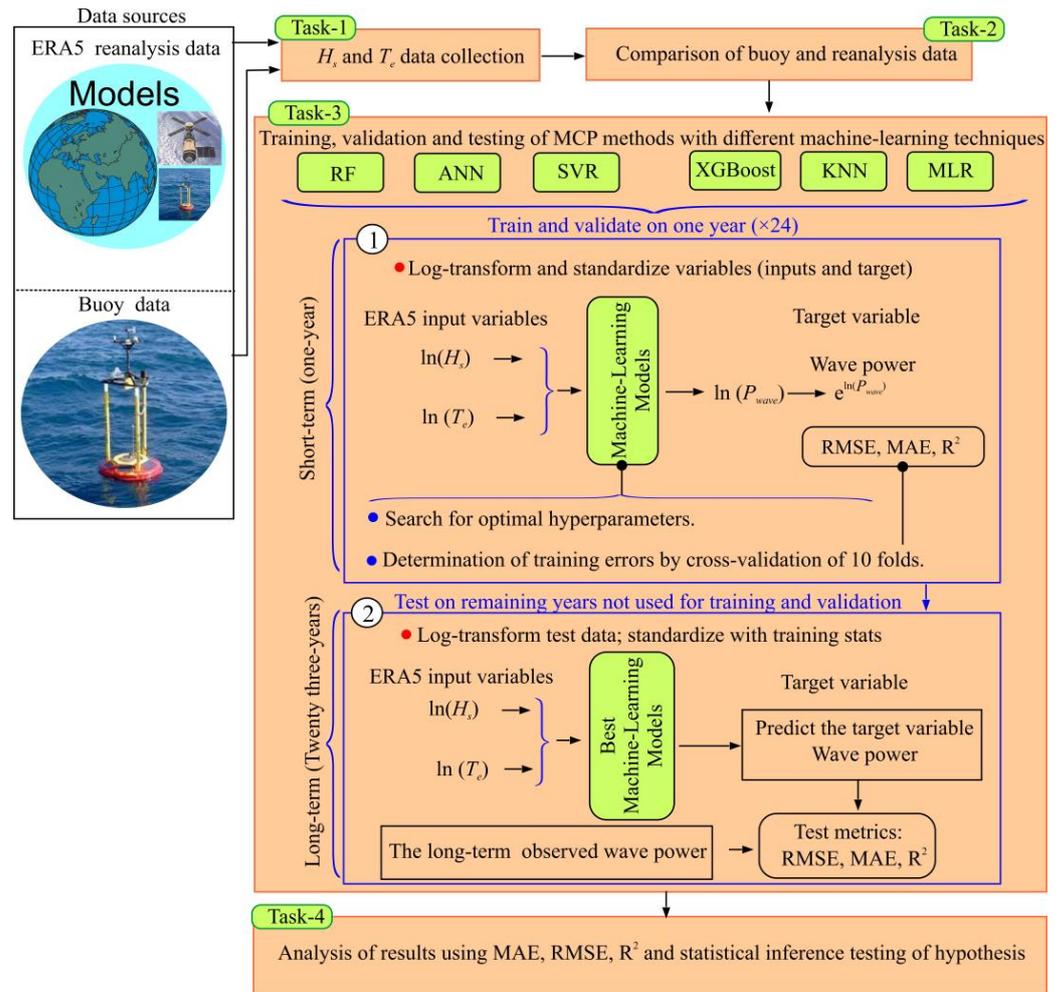


Figure 3. Schematic representation of the process used to estimate long-term wave power at a target site using MCP methods, with ERA5 reanalysis data as input.

3.3. Task 3 of the Method

In this step, six ML techniques are considered to construct the MCP models. The ML techniques employed are appropriately described in the literature: MLR, SVR, KNN, RF, and ANN in [26,27] and XGBoost in [28].

- MLR: A baseline linear model that assumes additive relationships between the inputs and the output.
- SVR: A non-linear model that identifies optimal hyperplanes in a transformed feature space, suitable for capturing complex patterns.
- KNN: A distance-based algorithm that predicts output values by averaging the nearest training points in the feature space.
- RF: An ensemble of decision trees trained on bootstrapped subsets of the data to enhance robustness and reduce overfitting.
- XGBoost: A powerful boosting technique that builds trees sequentially to correct the errors of previous trees.
- ANN: Layered computational models capable of learning non-linear relationships through training across multiple hidden units.

These methods were selected to cover a diverse set of modeling paradigms—linear (MLR), non-linear (SVR, KNN), ensemble-based (RF, XGBoost), and deep learning (ANN)—allowing for a comparison of their predictive performance and generalization ability. Each

technique offers distinct advantages in terms of flexibility, interpretability, and capacity to model complex relationships in oceanographic time series.

Each model is fed with H_s and T_e reanalysis variables to directly estimate the wave power at the target site—specifically, at the location where the buoy is moored.

The proposed models for estimating the target variable were developed using multiple regression, as shown in Equation (2).

$$Y_t = f(X_t) = f\left(\overbrace{\ln(H_{s,t}), \ln(T_{e,t})}^{ERA5}\right) \tag{2}$$

The datasets are preprocessed by calculating the logarithm of H_s and T_e , based on a logarithmic transformation of Equation (1) leading to a linear form, as expressed in Equation (3):

$$\ln(P_{wave}) = \ln\left(\frac{\rho_{seawater}g}{64\pi}\right) + 2 \cdot \ln(H_s) + \ln(T_e) \tag{3}$$

In the functional forms of the models, $X = (X_1, X_2)^T$ represents the input variables, the subscript t indicates the evaluated time step, and Y_t is the predicted response variable, corresponding to $\ln(P_{wave})$ at the target site. Once this variable has been estimated, a post-processing step is performed to obtain P_{wave} , as shown in Equation (4).

$$P_{wave} = \exp(Y_t) \tag{4}$$

The process is summarized in two steps, each represented by a number enclosed in a circle, as shown in Figure 3.

In the first step, all variables were standardized using z-score normalization—subtracting the mean and dividing by the standard deviation. The scaling parameters were computed from the training set and subsequently applied to the training, validation, and test data to avoid data leakage.

Next, the optimal hyperparameters for the selected model are determined (Figure 2). The data are divided into 10 folds to train and evaluate the model using cross-validation, ensuring robustness and minimizing the risk of overfitting. In each iteration, one fold is used as the validation set, while the remaining folds are used for training. The model is then defined using the selected ML technique, and the error metric to be used during the training and validation of the MCP model is specified.

For each ML technique used, a hyperparameter search space is defined, meaning that key values are explored to optimize model performance (Table 1).

Table 1. Regression models for wave power estimation.

Model	Tuned Hyperparameters	Preprocessing	Main R Packages
RF	'mtry' ∈ [1, p]; 'trees' ∈ [500, 3000]; 'max_depth' ∈ {3, 5, 7, 30}	Log transformation; centering and scaling	'randomForest' [32–34], 'ranger' [35,36], 'parsnip' [37–39], 'tune' [40,41], 'yardstick' [42,43]
ANN	Hidden layers ∈ [1, 3]; neurons per layer ∈ [2, 100]; epochs ∈ [100, 1500]	Log transformation; centering and scaling	'h2o' [44], 'tidymodels' [45], 'recipes' [46,47], 'yardstick'

Table 1. *Cont.*

Model	Tuned Hyperparameters	Preprocessing	Main R Packages
XGBoost	'min_n' ∈ [2, 40]; 'tree_depth' ∈ [1, 15]; 'learn_rate' ∈ [0.001, 0.3]; 'loss_reduction' ∈ [0, 10]	Log transformation; centering and scaling	'xgboost' [48–50], 'tidymodels' [45], 'tune', 'yardstick'
SVR	'C' ∈ {0.1, 1}; 'σ' ∈ {0.001, 0.01}; 'ε' = 0.001	Log transformation; centering and scaling	'caret' [51,52], 'kernlab' [53,54], 'yardstick', 'doParallel' [55,56]
KNN	'neighbors' ∈ [3, 50]; 'weight_func' ∈ {10 types}; 'deg_free' ∈ [2, 18]	Log transformation; splines; centering and scaling	'kknnc' [57,58], 'tidymodels', 'tune', 'yardstick'
MLR	None (standard linear regression: intercept, 'T _s ', 'H _s ' coefficients)	Log transformation; centering and scaling	'stats' [37], 'caret', 'rsample' [59,60], 'yardstick'

For hyperparameter optimization, we applied a grid search combined with 10-fold cross-validation. This procedure systematically explored the hyperparameter space defined for each machine-learning model (as detailed in Table 1) and selected the configuration that minimized the average root mean squared error (RMSE) on the validation folds. This strategy ensured that the selected hyperparameters were optimal within the explored space, guaranteeing model robustness and generalization capability.

In the case of RF, the hyperparameters explored include the number of trees (*trees*), the number of variables randomly selected at each split (*mtry*), where *p* is the total number of predictor variables, and the maximum tree depth (*max_depth*) [61].

For ANN, combinations of hyperparameters such as the number of hidden layers, the number of neurons per layer, and the number of training epochs are explored.

In the case of XGBoost, the defined hyperparameters include the minimum number of observations per node (*min_n*), the learning rate (*learn_rate*), the maximum tree depth (*tree_depth*), and the minimum loss reduction required to make a split (*loss_reduction*).

For SVR, different values of the key parameters—*C* (regularization parameter) and *σ* (kernel width or bandwidth parameter)—are tested using a predefined grid [34].

In KNN, the hyperparameters include the number of neighbors used for prediction (*neighbors*), the neighbor weighting function (*weight_func*: rectangular, inv, gaussian, triangular, Epanechnikov, biweight, triweight, cos, rank, optimal), and the degrees of freedom of the splines applied to the predictor variables (*deg_free*).

In MLR, parameters are adjusted based on the errors obtained on the validation data, without any hyperparameter tuning beyond the inherent structure of the linear model.

Subsequently, the model is trained with the best parameters and evaluated using all available short-term data. For the evaluation, we use the metrics *MAE*, (Equation (5)), *RMSE* (Equation (6)), and *R*² (Equation (7)).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n \left(y_i - \frac{\sum_{i=1}^n y_i}{n} \right)^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{7}$$

In Equations (5)–(7), y_i is the observed value, \hat{y}_i is the predicted value, and n is the total number of observations.

In Step 2, the best-performing model (with the optimal parameters) is used to estimate the values of the long-term target variable.

3.4. Task 4 of the Method

In the fourth task, a Wilcoxon test [62] is performed to evaluate the statistical significance of the metric values obtained by the different MCP models analyzed. The comparison focuses on the ML technique that generated the smallest errors, with a preset significance level of 0.05. The p_values are adjusted using the Benjamini–Hochberg method [62].

4. Results and Discussion

This section presents the results of the analyses conducted based on the tasks described in the methodology, as outlined in Figure 3.

4.1. Comparison of the Data Recorded in the Two Data Sources

Figure 4 shows the Pearson correlation coefficients between the different variables used in the analysis.

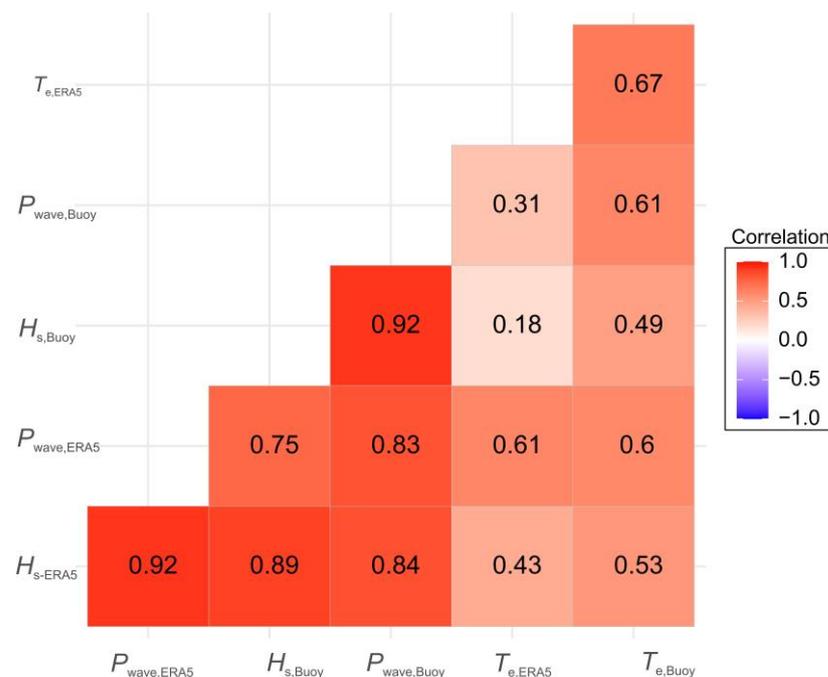


Figure 4. Pearson correlation coefficients between the variables recorded by the buoy and those from ERA5.

As shown in Figure 4, the Pearson correlation coefficient obtained from the hourly mean significant wave height data recorded by ERA5 and the buoy during the period 2000–2023 was $R = 0.89$. This value is similar to the one reported by Li et al. [22] ($R = 0.90$). However, the Pearson correlation coefficient obtained for the energy periods from both sources was $R = 0.67$, which is lower than the value reported by Li et al. [22] ($R = 0.9$).

The correlation coefficient between $P_{wave,Buoy}$ and $H_{s,ERA5}$ is 0.84, whereas the correlation between $P_{wave,Buoy}$ and $T_{e,ERA5}$ is only 0.31. Due to the greater influence of significant wave height compared to the energy period in wave power estimation, the correlation coefficient between $P_{wave,Buoy}$ and $P_{wave,ERA5}$ reaches a value of $R = 0.83$.

As shown in Figure 5, the mean significant wave height is higher at the target site (1.67 m) compared to the corresponding ERA5 mean (1.64 m). However, the mean energy period at the target site is lower (5.49 s) than the corresponding ERA5 value (8.33 s).

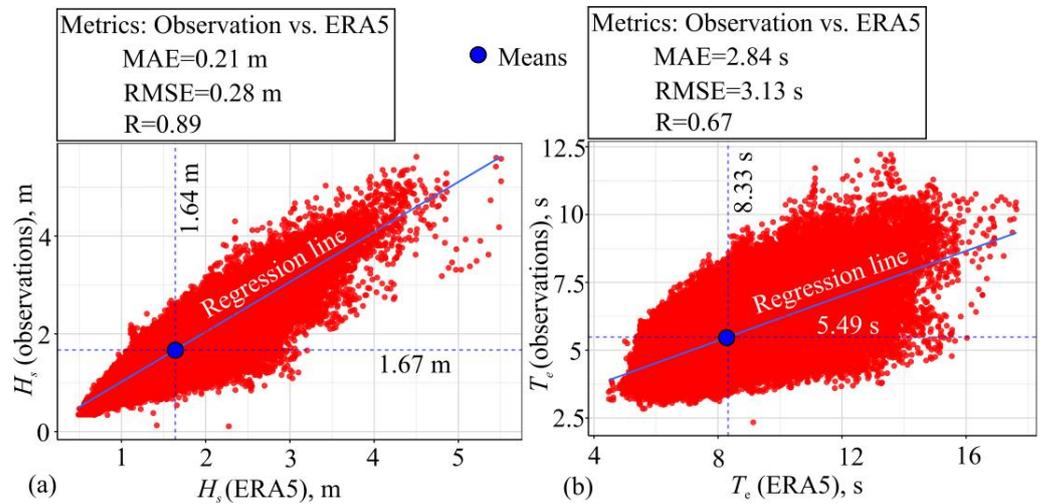


Figure 5. Scatter plots comparing wave parameters from buoy observations and ERA5 reanalysis data: (a) significant wave height (H_s); (b) energy period (T_e).

Although the coefficient of determination for wave power is moderate ($R^2 = 0.69$), the MAE and RMSE values were found to be 4.38 kW/m and 7.10 kW/m, respectively (Figure 6). The mean wave power estimated using ERA5 data was 12.5 kW/m, whereas the mean estimated from buoy data was 8.9 kW/m (Figure 6). In this context, reducing these errors is considered important in order to improve the accuracy of the results.

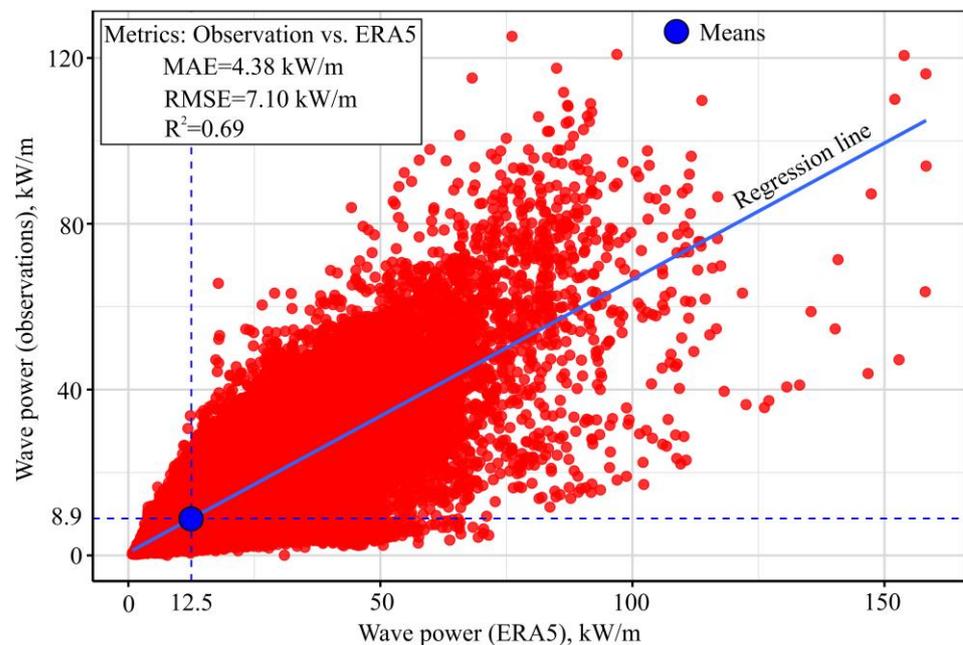


Figure 6. Comparison between observed and ERA5-estimated wave power values.

4.2. Results Obtained from the MCP Models

Figure 7 presents the values of the performance metrics (MAE, RMSE, and R^2) obtained during the testing phase (long-term) for the MCP models, using each of the ML techniques considered (RF, XGBoost, SVR, ANN, KNN, and MLR). Each boxplot contains 24 values,

corresponding to the test metric results obtained when each MCP model was trained and validated using 1 of the 24 individual years of data (2000–2023).

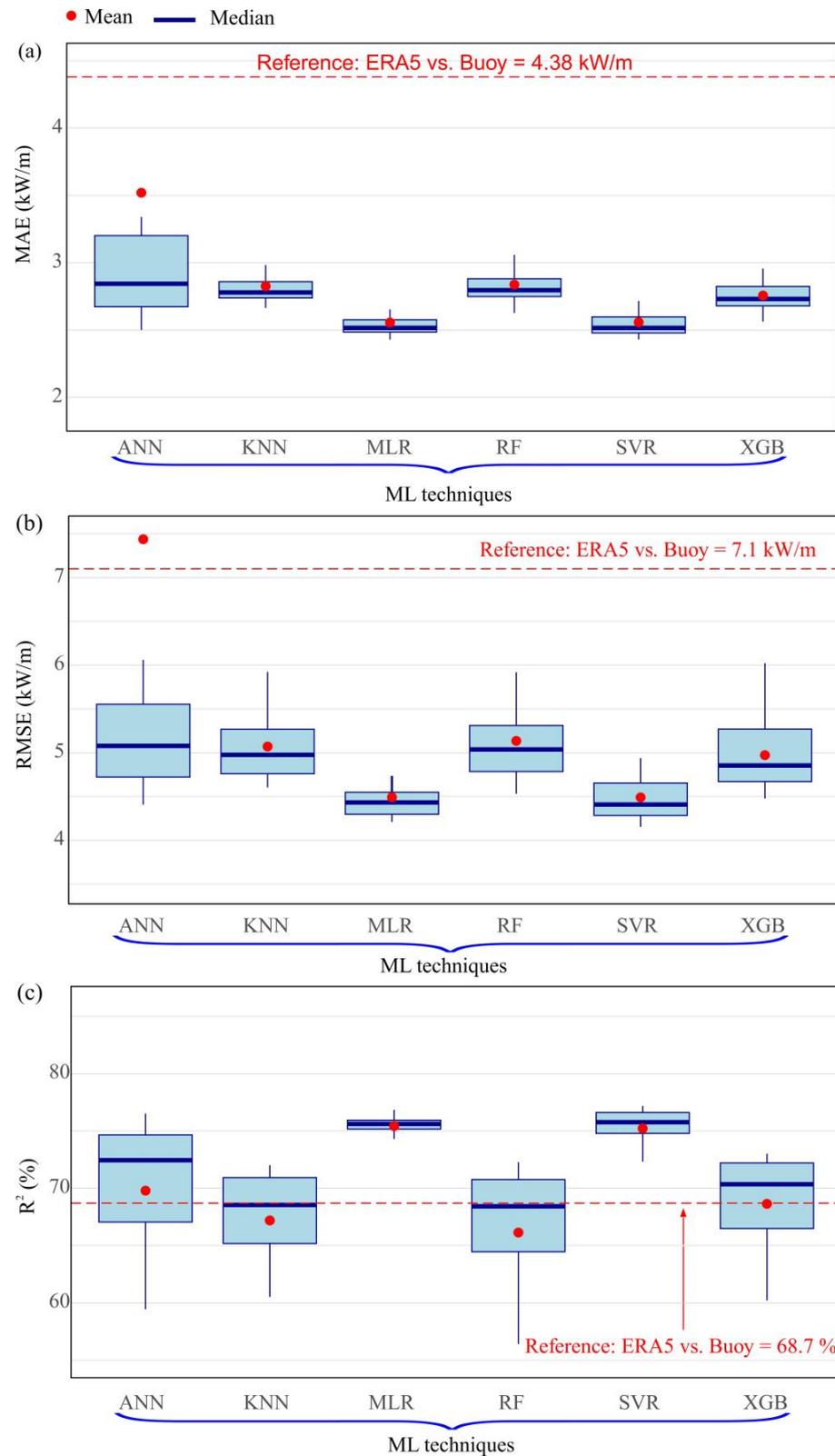


Figure 7. Performance metrics, (a) MAE, (b) RMSE, and (c) R², obtained during the testing phase for the MCP models trained with different ML techniques. Each boxplot represents the distribution of metric values obtained from 24 different training/validation years (2000–2023). Red dashed lines indicate the reference metric values obtained from the direct comparison between ERA5 and buoy data.

As shown in Figure 7, the ML techniques that yielded the best average metric values were MLR and SVR. Moreover, Figure 7 shows that, regardless of the year used for training and validation, these techniques consistently outperformed the reference metrics computed directly between ERA5 values and buoy measurements.

To assess whether the differences among the metric values obtained by the various ML techniques are statistically significant, the Wilcoxon test [62] was applied with a significance level of 0.05. The *p*-values were adjusted using the Benjamini–Hochberg method [62]. No statistically significant differences were observed between MLR and SVR for the MAE (*p*-value: 0.724), RMSE (*p*-value: 1.000), and R^2 (*p*-value: 0.407) metrics (Figure 8).

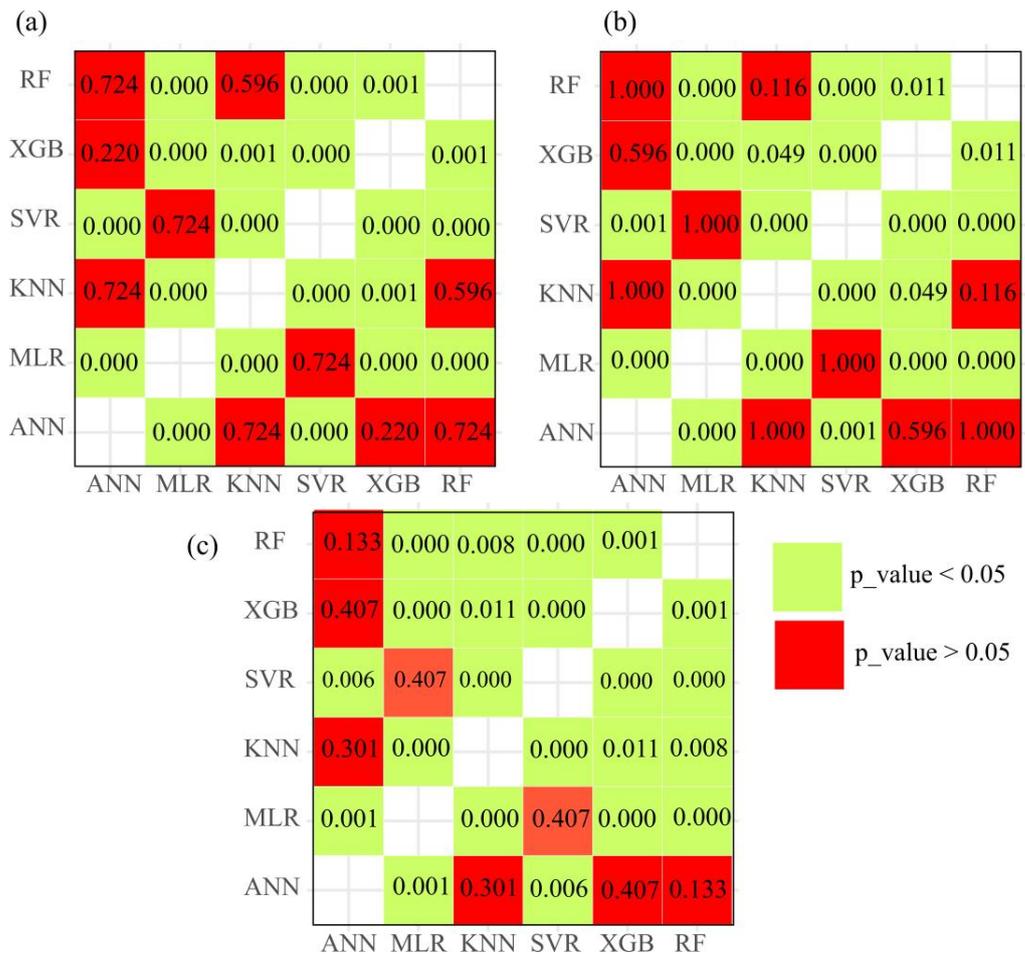


Figure 8. Pairwise Wilcoxon test results (*p*-values) for comparing ML techniques based on (a) MAE, (b) RMSE, and (c) R^2 . Statistically significant differences (*p*-value < 0.05) are highlighted in green while non-significant differences (*p*-value > 0.05) are highlighted in red.

Nevertheless, MLR was selected as the preferred model due to its lower computational complexity, higher interpretability, and ease of implementation. This choice aligns with the principle of parsimony, especially in contexts where model transparency is an important consideration.

Table 2 presents the parameter values of the MLR models obtained for each training/validation year. The parameters obtained for the remaining ML techniques are provided in Appendix A.

Table 2. Standardized and de-standardized parameters of the MLR models for each training year. The table shows the standardized coefficients (a' , b' , c') and the corresponding de-standardized parameters (A , b , and c) of the MLR models fitted using data from each year (2000–2023). The values of A are expressed in $J/(m^3 \cdot s^2)$, and the exponents b and c correspond to the power-law formulation of wave power estimation. The standardized intercept a' is theoretically zero due to the standardization of variables; small numerical deviations are within machine precision.

Year	Standardized Coefficients			De-Standardized Parameters		
	a'	b'	c'	A $J/(m^3s^2)$	b	c
2000	1.48×10^{-16}	0.9798	-0.1515	4.113	2.688	-0.415
2001	3.73×10^{-16}	0.9098	-0.1123	3.647	2.448	-0.302
2002	6.05×10^{-16}	0.9661	-0.1452	3.933	2.569	-0.386
2003	-1.20×10^{-15}	0.9673	-0.1129	3.572	2.524	-0.295
2004	1.76×10^{-16}	0.9007	-0.0157	2.285	2.749	-0.048
2005	6.93×10^{-17}	0.9362	-0.1914	5.719	2.527	-0.516
2006	-7.58×10^{-17}	0.9418	-0.1246	4.346	2.492	-0.330
2007	-3.55×10^{-16}	0.919	-0.1166	4.677	2.443	-0.310
2008	2.16×10^{-16}	0.9246	-0.0859	3.781	2.477	-0.230
2009	-1.25×10^{-16}	0.9506	-0.1008	3.610	2.566	-0.272
2010	-2.02×10^{-17}	0.9048	-0.0892	3.264	2.49	-0.245
2011	-2.51×10^{-16}	0.9017	-0.1020	3.738	2.419	-0.274
2012	1.81×10^{-17}	0.9238	-0.0911	3.337	2.616	-0.258
2013	-4.62×10^{-17}	0.9079	-0.0360	2.777	2.264	-0.090
2014	2.65×10^{-16}	0.9154	0.0368	1.834	2.315	0.093
2015	1.28×10^{-16}	0.9221	-0.0627	2.867	2.500	-0.170
2016	1.47×10^{-16}	0.9148	-0.0947	3.600	2.589	-0.268
2017	1.40×10^{-16}	0.9407	-0.1211	4.004	2.683	-0.345
2018	-7.57×10^{-16}	0.9715	-0.0840	3.320	2.475	-0.214
2019	-2.28×10^{-17}	0.9615	-0.2132	6.719	2.711	-0.601
2020	4.85×10^{-16}	0.9554	-0.1062	4.417	2.621	-0.291
2021	4.64×10^{-16}	0.9519	-0.1778	5.688	2.562	-0.479
2022	-5.64×10^{-16}	0.8899	-0.0635	3.172	2.443	-0.174
2023	-2.65×10^{-16}	0.9281	-0.1520	5.081	2.626	-0.430

The MLR model used with standardized data is shown in Equation (8), where the asterisks indicate that the variables are standardized (mean = 0, standard deviation = 1).

$$[\ln(P_{wave})]^* = a' + b' \cdot [\ln(H_s)]^* + c' \cdot [\ln(T_e)]^* \tag{8}$$

The coefficients a' , b' , and c' in Equation (8) cannot be directly interpreted in terms of the original physical equation, Equation (1), since the variables are standardized. Therefore, it is necessary to reverse the standardization process, as shown in Equation (9)

$$\ln(P_{wave}) = a + b \cdot \ln(H_s) + c \cdot \ln(T_e) \tag{9}$$

The values of a , b , and c are obtained from Equation (10), provided in Appendix B.

$$b = b' \cdot \left(\frac{\sigma_{\ln(P_{wave})}}{\sigma_{\ln(H_s)}} \right); \quad c = c' \cdot \left(\frac{\sigma_{\ln(P_{wave})}}{\sigma_{\ln(T_e)}} \right); \quad a = \mu_{\ln(P_{wave})} - b \cdot \mu_{\ln(H_s)} - c \cdot \mu_{\ln(T_e)} \tag{10}$$

In Equation (10), μ and σ represent the mean and standard deviation, respectively, of the unstandardized variables.

Equation (9) can be rewritten as Equation (11), which allows for comparison with the physical model in Equation (1).

$$P_{wave} = A \cdot H_s^b \cdot T_e^c \tag{11}$$

In Equation (11), the parameter A is derived from Equation (12).

$$A = \exp(a) \tag{12}$$

Equations (12) and (1) are equivalent when the following condition is met (Equation (13)):

$$A = \frac{\rho_{seawater} \cdot g^2}{64\pi} ; b = 2 ; c = 1 \tag{13}$$

The de-standardized parameters in Table 2 reveal the relationship between the wave power at the target site and the ERA5-based H_s and T_e parameters, through Equation (11).

Although the functional form of the regression model (Equation (11)) is based on the physical wave power equation (Equation (1)), the empirical fitting using ERA5 input data and buoy observations as targets yields different coefficients. In particular, $b > 2$, $c < 1$, and the values of A are significantly higher than $(\rho_{seawater} \cdot g^2)/64\pi$. These discrepancies can be explained by the systematic biases observed in ERA5 variables: underestimation of H_s and overestimation of T_e (Figure 6).

In this context, the model adjusts the exponents and the constant to compensate for these mismatches and to produce a more accurate estimate of actual wave power. The resulting model does not contradict physical theory; rather, it empirically adapts it to the specific characteristics of the available data.

Figure 9 shows the boxplots of MAE and RMSE errors obtained during training (short-term) and testing (long-term) of the MLR models, using each of the 24 available years as the training set in a temporal cross-validation process. In all cases, the test errors are higher than those observed during training, and Wilcoxon paired-sample tests confirm that these differences are statistically significant ($p_value < 0.05$). However, despite their statistical significance, the absolute differences between training and test errors are moderate. The medians, means, and interquartile ranges are similar, suggesting that the models exhibit good temporal generalization capability, with consistent performance when applied to years not included in the training process. This indicates that no relevant overfitting has occurred and supports the robustness of the proposed approach.

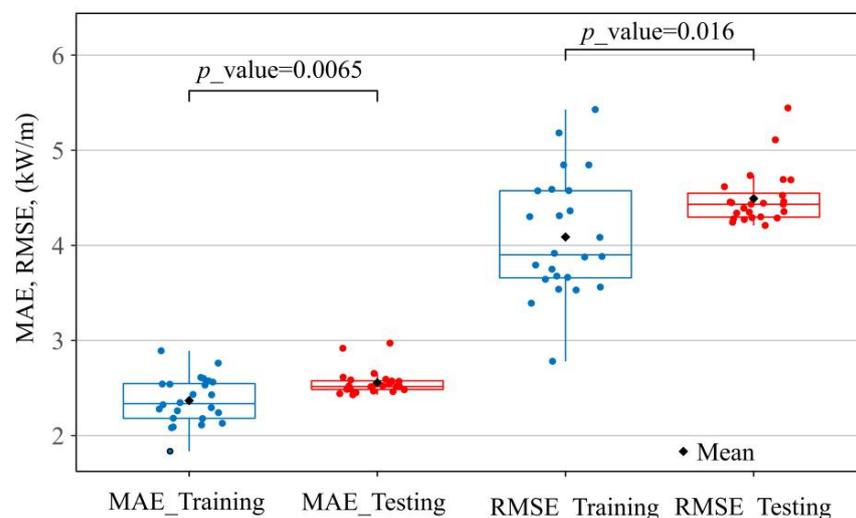


Figure 9. Comparison of training (blue) and testing (red) errors (MAE and RMSE) for MLR models using temporal cross-validation. Wilcoxon test p -values indicate statistically significant differences.

Figure 10 shows the evolution of long-term monthly mean wave power obtained from ERA5 and buoy data, as well as the values estimated using the MLR models trained on the year with the lowest *RMSE* (2022) and the year with the highest *RMSE* (2004).

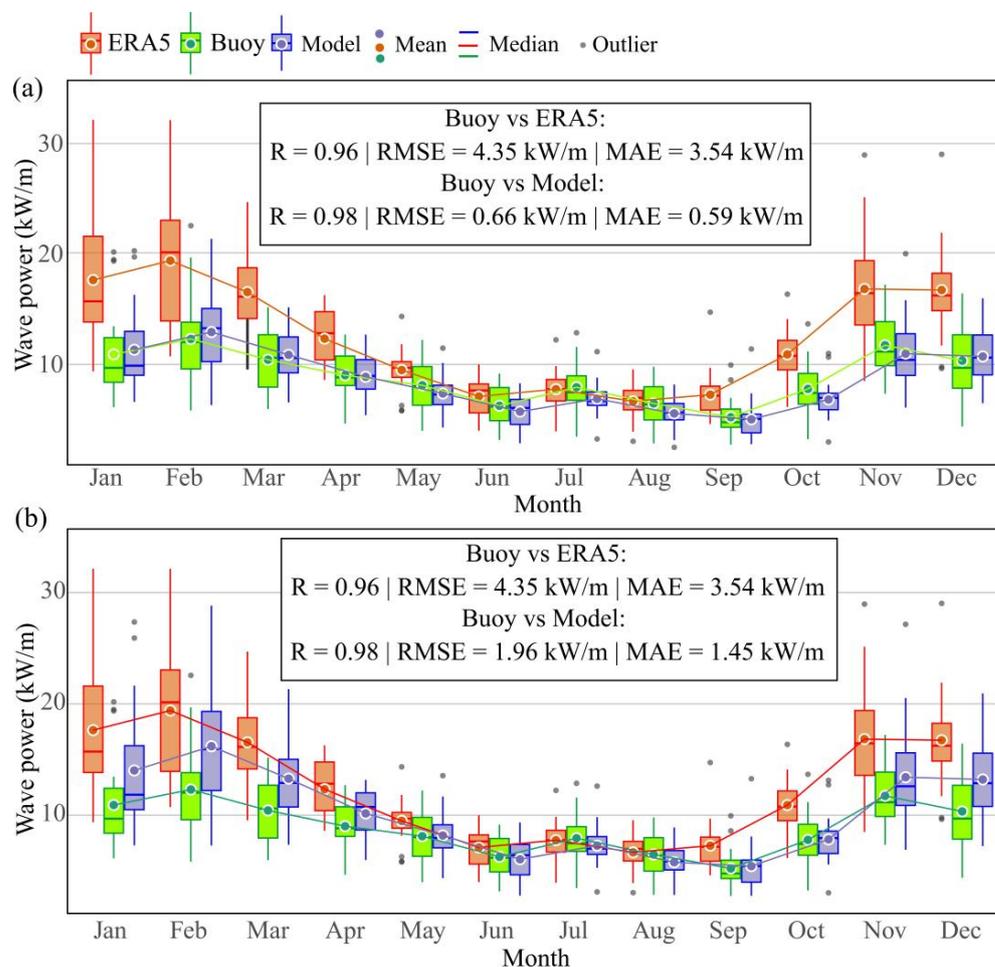


Figure 10. Monthly mean wave power from buoy data, ERA5 reanalysis, and MLR models trained on the years with the lowest (a) and highest (b) *RMSE*. Statistical metrics show the improved performance of MLR models compared to ERA5.

The highest observed monthly mean wave power values occur during the periods from January to March and from November to December. During these months, the differences between the reanalysis-based wave power and that recorded by the buoy are more pronounced. The ERA5 data systematically overestimate wave power during these peak months. Despite these discrepancies, the Pearson correlation coefficients between the buoy and ERA5 monthly means and those from the MLR model were 0.96 and 0.98, respectively, indicating a strong relationship between the seasonal variations in ERA5/MLR data and the buoy measurements.

As shown in Figure 10, the error metrics associated with the MLR models are lower than those of ERA5, even when using the model trained on the year that yielded the highest testing *RMSE*. This demonstrates that the MLR-based methodology not only captures the seasonal dynamics of wave power with high accuracy but also outperforms ERA5 in terms of predictive precision. Moreover, the ability of the model to accurately track peak wave power months reinforces its robustness and suggests its potential usefulness for seasonal wave energy resource planning in coastal areas. The analysis showed that the specific year chosen for model training (2000–2023) had only a minor impact on model performance. All models trained with one full year of data achieved consistent and robust results that

outperformed direct ERA5-based estimates. This finding confirms the suitability of using any single full year for training in similar contexts.

As shown in Figure 11, no clear diurnal pattern is observed in the hourly mean wave power. However, the interquartile range of ERA5 estimates is consistently wider than that of both the buoy data and the MLR model estimates, indicating higher uncertainty. ERA5 also exhibits a systematic overestimation of the wave power throughout the day. Although this bias is less pronounced in the worst-case MLR model (Figure 11b), the model still demonstrates better alignment with buoy observations in terms of central tendency and dispersion, further supporting its robustness for hourly-scale applications.

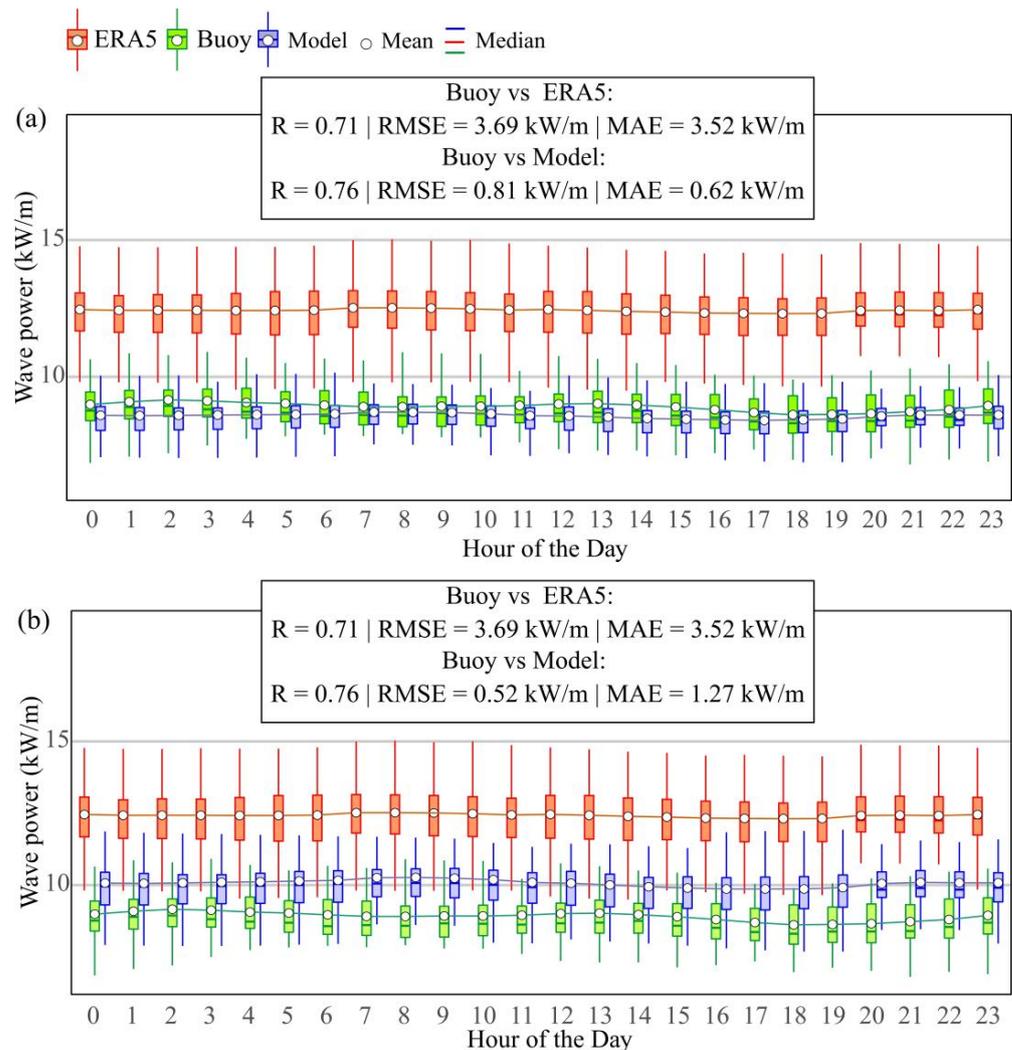


Figure 11. Hourly mean wave power from buoy, ERA5, and MLR models trained on the best (a) and worst (b) performing years. ERA5 shows greater interquartile variability and a systematic overestimation of wave power.

MCP methods are based on a set of assumptions [24], including knowledge of the seasonal variation pattern during the concurrent data period and climate stability. The short-term period used for training the models must be long enough to allow the extraction of seasonal wave behavior. The general recommendation is that this concurrent dataset should span at least one year. These methods ignore the effects of climate change and assume that the wave behavior during the energy project’s lifetime will be similar to that of the past—i.e., they assume the statistical stationarity of the datasets.

In this context, Figure 12 shows the evolution of hourly wave power throughout 2023, calculated from buoy data and estimated by the MLR model trained using data from the year 2000. This scenario represents a test of model performance trained at the beginning of the available dataset (2000–2023) and evaluated at its end. In this case study, the MLR model continued to outperform ERA5, providing better error and correlation metrics, thus supporting the temporal robustness of the proposed approach.

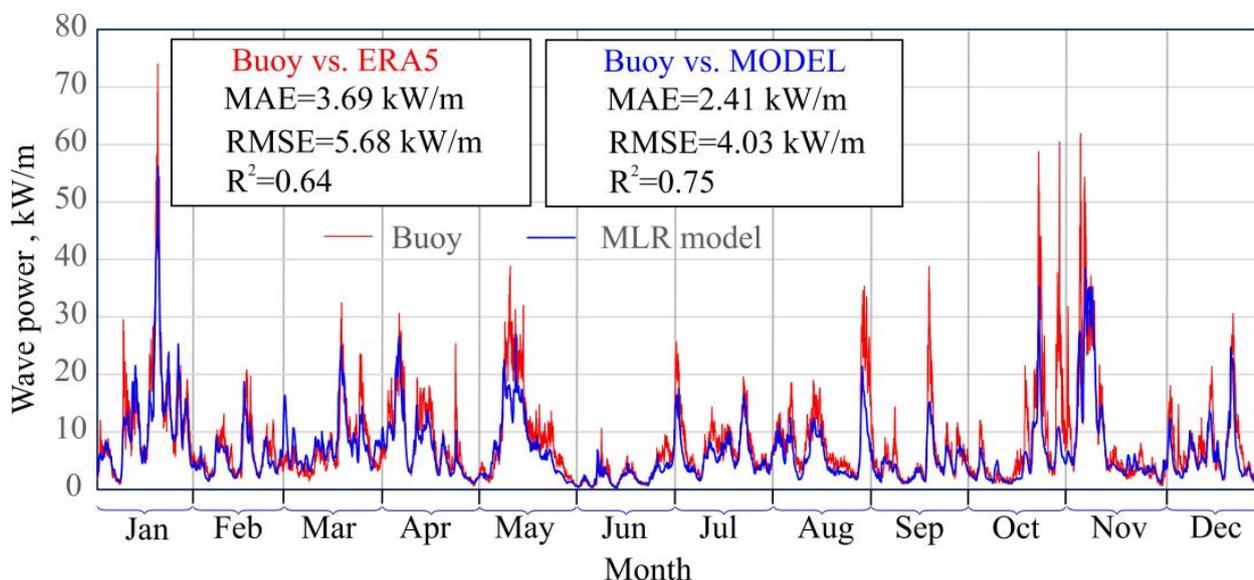


Figure 12. Hourly wave power in 2023 from buoy observations and estimates from the MLR model trained with data from 2000.

When MCP models are used to estimate long-term renewable energy resources, it is generally recommended to train the models for at least one full year [24]. Increasing the duration of the training period allows the models to better learn the relationship between the observed data and the reference data, potentially reducing uncertainty in long-term resource estimation. However, extending the measurement campaign also increases the associated costs. Therefore, determining the optimal duration of a measurement campaign should be based on a cost–benefit analysis tailored to the specific characteristics and constraints of each project [63]. It is also worth noting that in some regions, the minimum duration of resource assessment campaigns is regulated by official guidelines [64].

In the context of the present case study, Figure 13 illustrates the trends observed in long-term error and correlation metrics as a function of the number of years used to train the MLR models. The results show that increasing the training period from 1 to 4 years leads to a slight decrease in the average error metrics (*MAE* and *RMSE*) and a slight increase in the coefficient of determination (R^2). These findings confirm that longer training periods can marginally improve model performance, although with diminishing returns. The improvement observed by extending the training period from three to four years was below 1%, and is considered marginal from a practical standpoint. No statistical significance test was performed, as the magnitude of the difference does not suggest a meaningful impact.

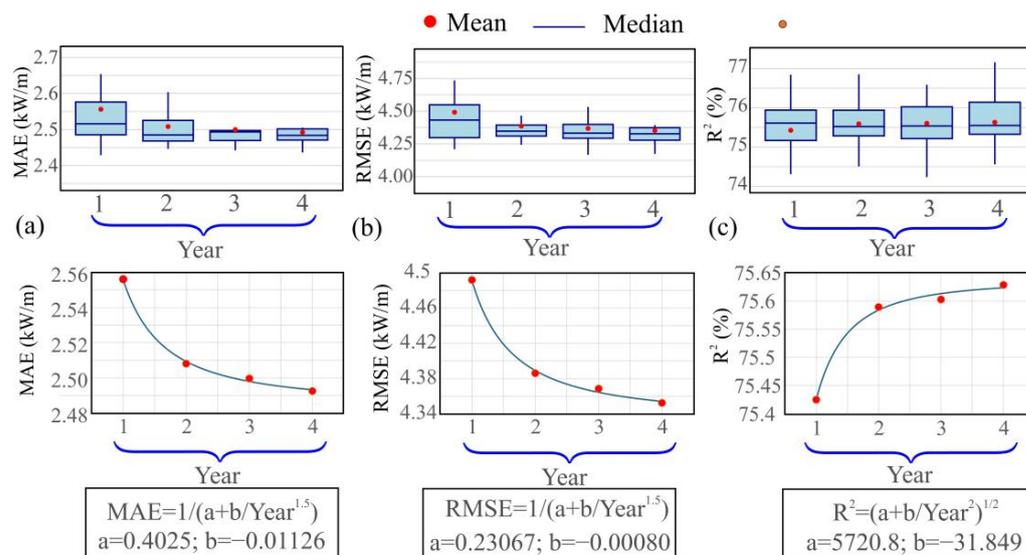


Figure 13. Influence of training period length (1–4 years) on the long-term performance metrics of MLR models: (a) MAE, (b) RMSE, and (c) R². Trend curves illustrate the marginal improvement in model performance with longer training durations.

5. Conclusions

This work presents a methodology based on supervised ML techniques to estimate long-term wave power at coastal locations where only short-term in situ wave data are available. The approach combines a limited measurement campaign with long-term ERA5 reanalysis data, applying MCP methods to develop predictive models for wave power at the target site.

The main findings can be summarized as follows:

- ERA5 reanalysis data systematically overestimate wave power compared to buoy observations, especially during high-energy months, and fail to accurately capture the daily and seasonal variability at the target site.
- Machine-learning techniques—particularly MLR and SVR—substantially improve wave power estimation, reducing error metrics (MAE, RMSE) and increasing the coefficient of determination (R²) relative to ERA5-based predictions.
- The MLR model offers a key advantage through its interpretable power-law form, enabling direct analysis of how ERA5 variables influence wave power estimates. The parameters A, b, and c provide insight into calibration bias and variable sensitivity, bridging data-driven modeling with physical understanding.
- The approach demonstrates strong temporal robustness, with MLR models trained on early-period data (e.g., 2000) still outperforming ERA5 when applied to recent years (e.g., 2023), even under evolving wave conditions.
- The methodology is fully reproducible and transferable to other coastal regions, relying on openly available reanalysis data and implemented entirely with transparent R-based tools, which supports its adoption in data-limited marine energy assessments.
- Additionally, in this case study, extending the training period from one to four years resulted in slight improvements in model performance. While this suggests that longer measurement campaigns may offer marginal benefits, such conclusions should be validated on a site-specific basis and balanced against the increased cost of data acquisition.

Author Contributions: Conceptualization, M.J.P.-M. and J.A.C.; methodology, M.J.P.-M. and J.A.C.; software, J.A.C.; validation, J.A.C.; formal analysis, M.J.P.-M. and J.A.C.; investigation, M.J.P.-M. and J.A.C.; resources, M.J.P.-M. and J.A.C.; data curation, M.J.P.-M.; writing—original draft preparation,

M.J.P.-M.; writing—review and editing, J.A.C.; visualization, M.J.P.-M. and J.A.C.; supervision, J.A.C.; project administration, M.J.P.-M. and J.A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-funded with ERDF funds through the INTERREG MAC 2021–2027 program in the RESMAC project (1/MAC/2/2.2/0011). No funding sources had any influence on study design, collection, analysis, or interpretation of data, manuscript preparation, or the decision to submit for publication.

Data Availability Statement: These data were derived from the following resources available in the public domain: Puertos del Estado data [<https://portus.puertos.es/#/>], accessed on 7 May 2025] and ERA5 [<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview>], accessed on 7 May 2025].

Acknowledgments: The authors acknowledge Puertos del Estado (Spain) for providing wave data from their buoy network, which were used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
ERA5	European Centre for Medium-Range Weather Forecasts Reanalysis
GW	GigaWatts
H _s	Significant wave height
IRENA	International Renewable Energy Agency
KNN	K-Nearest-Neighbor
KW/m	KiloWatts per meter
MAE	Mean absolute error
MCP	Measure–Correlate–Predict
ML	Machine learning
MLR	Multiple Linear Regression
PLOCAN	PLataforma Oceánica de CANarias—Oceanic Platform of the Canary Islands
PTECAN	Plan de Transición Energética de CANarias—Canary Islands’ Energy Transition Plan
P _{wave}	Wave power
RF	Random Forest
R ²	Coefficient of determination
RMSE	Root mean squared error
SVR	Support Vector Regression
T _e	Energy period
T _p	Peak period
XGBoost	eXtreme Gradient Boosting

Appendix A

Table A1. Parameters of the RF and ANN models.

Year	RF Parameters			ANN Parameters		
	Mtry	Trees	Max_Depth	Hidden Layers	Epochs	Neurons per Layer
2000	1	3000	30	1	604	2
2001	1	3000	30	1	1004	5
2002	1	3000	30	2	904	9-9
2003	1	3000	30	1	604	2
2004	1	3000	7	1	1004	9

Table A1. Cont.

Year	RF Parameters			ANN Parameters		
	Mtry	Trees	Max_Depth	Hidden Layers	Epochs	Neurons per Layer
2005	1	3000	30	1	904	8
2006	1	3000	30	2	804	6-6
2007	1	3000	30	2	504	8-8
2008	1	3000	7	2	903	3-2
2009	1	3000	30	1	704	2
2010	1	3000	30	1	904	3
2011	1	3000	30	2	905	5-5
2012	1	3000	30	1	604	3
2013	1	3000	30	2	1003	4-2
2014	1	3000	30	1	804	10
2015	1	3000	30	2	904	6-6
2016	1	3000	30	2	804	4-4
2017	1	3000	30	2	1004	3-2
2018	1	3000	30	1	803	2
2019	1	3000	30	2	504	10-10
2020	1	3000	30	2	1005	3-3
2021	1	3000	30	2	904	4-3
2022	1	3000	30	1	603	12
2023	1	3000	30	2	504	8-8

Table A2. Parameters of the SVR and XGBoost models.

Year	SVR Parameters			XGBoost Parameters			
	σ	C	ϵ	Min_n	Tree.Depth	Learn_Rate	Loss.Reduction
2000	0.01	1	0.001	8	8	0.0052511	5.85×10^{-5}
2001	0.01	1	0.001	30	13	0.00989746	0.34308064
2002	0.01	1	0.001	11	10	0.00602343	7.83×10^{-8}
2003	0.01	1	0.001	20	9	0.00746615	1.06×10^{-10}
2004	0.01	1	0.001	15	5	0.09011734	0.04092907
2005	0.01	1	0.001	8	8	0.00515634	1.74×10^{-10}
2006	0.01	1	0.001	21	11	0.01865139	0.8246732
2007	0.01	1	0.001	4	6	0.00828647	6.40×10^{-10}
2008	0.01	1	0.001	28	11	0.03561075	0.07489249
2009	0.01	1	0.001	36	7	0.05246322	0.10571014
2010	0.01	1	0.001	18	10	0.0075247	2.95×10^{-6}
2011	0.01	1	0.001	33	11	0.0079609	4.80×10^{-7}
2012	0.01	1	0.001	19	12	0.00557426	3.80×10^{-9}
2013	0.01	1	0.001	35	15	0.00771763	0.26072018
2014	0.01	1	0.001	3	9	0.06530932	0.11592727
2015	0.01	1	0.001	23	13	0.00589396	5.01×10^{-9}
2016	0.01	1	0.001	36	7	0.05246322	0.10571014
2017	0.01	1	0.001	21	11	0.01865139	0.8246732
2018	0.01	1	0.001	15	8	0.00967407	2.57×10^{-7}
2019	0.01	1	0.001	27	11	0.04011186	0.15475962
2020	0.01	1	0.001	16	12	0.00437732	0.00017414
2021	0.001	1	0.001	15	14	0.0055972	1.29×10^{-7}
2022	0.01	1	0.001	21	12	0.03830943	0.07938781
2023	0.01	1	0.001	32	11	0.00494699	0.00513382

Table A3. Parameters of the KNN models.

Year	KNN Parameters				
	Neighbors	weight_func	deg_free_H	deg_free_T	Neighbors
2000	26	triangular	18	18	26
2001	37	triangular	2	4	37
2002	39	triweight	5	6	39
2003	26	biweight	16	17	26
2004	36	triweight	15	18	36
2005	14	biweight	2	2	14
2006	34	triweight	2	3	34
2007	30	triweight	4	4	30
2008	50	biweight	17	14	50
2009	20	optimal	3	2	20
2010	40	triweight	2	2	40
2011	31	triweight	10	12	31
2012	41	biweight	10	11	41
2013	41	triweight	7	8	41
2014	50	triangular	10	10	50
2015	23	biweight	3	2	23
2016	36	biweight	2	3	36
2017	50	triangular	10	10	50
2018	33	biweight	18	18	33
2019	22	triangular	4	3	22
2020	14	inv	18	18	14
2021	14	inv	18	18	14
2022	50	triweight	13	17	50
2023	47	triweight	14	11	47

Appendix B

Standardization and Recovery of the Original Scale in the MLR Model.

The fitted model is expressed in Equation (A1), where the variables have been standardized according to Equations (A2)–(A4).

$$[\ln(P_{wave})]^* = a' + b' \cdot [\ln(H_s)]^* + c' \cdot [\ln(T_e)]^* \tag{A1}$$

$$[\ln(P_{wave})]^* = \frac{\ln(P_{wave}) - \mu_{\ln(P_{wave})}}{\sigma_{\ln(P_{wave})}} \tag{A2}$$

$$[\ln(H_s)]^* = \frac{\ln(H_s) - \mu_{\ln(H_s)}}{\sigma_{\ln(H_s)}} \tag{A3}$$

$$[\ln(T_e)]^* = \frac{\ln(T_e) - \mu_{\ln(T_e)}}{\sigma_{\ln(T_e)}} \tag{A4}$$

Here, μ and σ represent the mean and standard deviation, respectively, of the unstandardized variables.

The goal is to derive Equation (A5), where the model parameters are expressed at the original scale.

$$\ln(P_{wave}) = a + b \cdot \ln(H_s) + c \cdot \ln(T_e) \tag{A5}$$

From Equation (B2), we isolate $\ln(P_{wave})$, as shown in Equation (A6).

$$\ln(P_{wave}) = [\ln(P_{wave})]^* \cdot \sigma_{\ln(P_{wave})} + \mu_{\ln(P_{wave})} \tag{A6}$$

Substituting Equation (A1) into Equation (A6) gives (A7).

$$\ln(P_{wave}) = \{a' + b' \cdot [\ln(H_s)]^* + c' \cdot [\ln(T_e)]^*\} \cdot \sigma_{\ln(P_{wave})} + \mu_{\ln(P_{wave})} \quad (A7)$$

Substituting Equations (A3) and (A4) into Equation (A7), and rearranging, we obtain Equation (A8).

$$\ln(P_{wave}) = \underbrace{\mu_{\ln(P_{wave})} + a' \cdot \sigma_{\ln(P_{wave})} - b' \cdot \sigma_{\ln(P_{wave})} \cdot \frac{\mu_{\ln(H_s)}}{\sigma_{\ln(H_s)}} - c' \cdot \frac{\mu_{\ln(T_e)}}{\sigma_{\ln(T_e)}}}_{\substack{b \\ c}} + \underbrace{b' \cdot \frac{\sigma_{\ln(P_{wave})}}{\sigma_{\ln(H_s)}} \cdot \ln(H_s)}_b + \underbrace{c' \cdot \frac{\sigma_{\ln(P_{wave})}}{\sigma_{\ln(T_e)}} \cdot \ln(T_e)}_c \quad (A8)$$

Notes

¹ www.gobiernodecanarias.org/energia/info-publica/PTECan_VersionInicial/, accessed on 7 May 2025

² <https://plocan.eu/>, accessed on 7 May 2025

References

- Renewable Energy in Climate Change Adaptation: Metrics and Risk Assessment Framework. Available online: <https://www.irena.org/Publications/2025/Apr/Renewable-energy-in-climate-change-adaptation> (accessed on 24 April 2025).
- International Renewable Energy Agency. Renewable Capacity Highlights 2025. 2025. Available online: <https://unstats.un.org/unsd/methodology/m49/> (accessed on 24 April 2025).
- International Renewable Energy Agency. OCEAN ENERGY TECHNOLOGIES A Brief from the IRENA Collaborative Framework on Ocean Energy and Offshore Renewables. 2023. Available online: www.irena.org (accessed on 24 April 2025).
- International Renewable Energy Agency. World Energy Transitions Outlook 2022: 1.5 °C Pathway. Available online: www.irena.org/publications/2022/Mar/World-Energy-Transitions-Outlook-2022 (accessed on 24 April 2025).
- Sun, P.; Wang, J. Long-term variability analysis of wave energy resources and its impact on wave energy converters along the Chinese coastline. *Energy* **2024**, *288*, 129644. [[CrossRef](#)]
- Khojasteh, D.; Mousavi, S.M.; Glamore, W.; Iglesias, G. Wave energy status in Asia. *Ocean. Eng.* **2018**, *169*, 344–358. [[CrossRef](#)]
- Astariz, S.; Iglesias, G. The economics of wave energy: A review. *Renew. Sustain. Energy Rev.* **2015**, *45*, 397–408. [[CrossRef](#)]
- Satymov, R.; Bogdanov, D.; Dadashi, M.; Lavidas, G.; Breyer, C. Techno-economic assessment of global and regional wave energy resource potentials and profiles in hourly resolution. *Appl. Energy* **2024**, *364*, 123119. [[CrossRef](#)]
- Majidi, A.G.; Ramos, V.; Rosa-Santos, P.; Akpınar, A.; Neves, L.D.; Taveira-Pinto, F. Development of a multi-criteria decision-making tool for combined offshore wind and wave energy site selection. *Appl. Energy* **2025**, *384*, 125422. [[CrossRef](#)]
- Silva, K.; Abreu, T.; Oliveira, T.C.A. Inter- and intra-annual variability of wave energy in Northern mainland Portugal: Application to the HiWave-5 project. *Energy Rep.* **2022**, *8*, 6411–6422. [[CrossRef](#)]
- Tong, Y.; Li, J.; Chen, W.; Li, B. Long-Term (1979–2024) Variation Trend in Wave Power in the South China Sea. *J. Mar. Sci. Eng.* **2025**, *13*, 524. [[CrossRef](#)]
- Reguero, B.G.; Losada, I.J.; Méndez, F.J. A global wave power resource and its seasonal, interannual and long-term variability. *Appl. Energy* **2015**, *148*, 366–380. [[CrossRef](#)]
- Sun, Z.; Zhang, H.; Xu, D.; Liu, X.; Ding, J. Assessment of wave power in the South China Sea based on 26-year high-resolution hindcast data. *Energy* **2020**, *197*, 117218. [[CrossRef](#)]
- Liu, J.; Li, R.; Li, S.; Meucci, A.; Young, I.R. Increasing wave power due to global climate change and intensification of Antarctic Oscillation. *Appl. Energy* **2024**, *358*, 122572. [[CrossRef](#)]
- Bessonova, V.; Tapoglou, E.; Dorrell, R.; Dethlefs, N.; York, K. Global evaluation of wave data reanalysis: Comparison of the ERA5 dataset to buoy observations. *Appl. Ocean Res.* **2025**, *157*, 104490. [[CrossRef](#)]
- Hersbach, H.; Peubey, C.; Simmons, A.; Berrisford, P.; Poli, P.; Dee, D. ERA-20CM: A twentieth-century atmospheric model ensemble. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 2350–2375. [[CrossRef](#)]
- Ulazia, A.; Sáenz, J.; Saenz-Aguirre, A.; Ibarra-Berastegui, G.; Carreno-Madinabeitia, S. Paradigmatic case of long-term collocated wind–wave energy index trend in Canary Islands. *Energy Convers. Manag.* **2023**, *283*, 116890. [[CrossRef](#)]
- Mahmoodi, K.; Ghassemi, H.; Razminia, A. Temporal and spatial characteristics of wave energy in the Persian Gulf based on the ERA5 reanalysis dataset. *Energy* **2019**, *187*, 115991. [[CrossRef](#)]

19. Wang, J.; Wang, Y. Evaluation of the ERA5 Significant Wave Height against NDBC Buoy Data from 1979 to 2019. *Mar. Geod.* **2022**, *45*, 151–165. [[CrossRef](#)]
20. Hisaki, Y. Intercomparison of Assimilated Coastal Wave Data in the Northwestern Pacific Area. *J. Mar. Sci. Eng.* **2020**, *8*, 579. [[CrossRef](#)]
21. Shi, H.; Cao, X.; Li, Q.; Li, D.; Sun, J.; You, Z.; Sun, Q. Evaluating the Accuracy of ERA5 Wave Reanalysis in the Water Around China. *J. Ocean Univ. China* **2021**, *20*, 1–9. [[CrossRef](#)]
22. Li, B.; Chen, W.; Li, J.; Liu, J.; Shi, P.; Xing, H. Wave energy assessment based on reanalysis data calibrated by buoy observations in the southern South China Sea. *Energy Rep.* **2022**, *8*, 5067–5079. [[CrossRef](#)]
23. Ayuso-Virgili, G.; Christakos, K.; Lande-Sudall, D.; Lümmen, N. Measure-correlate-predict methods to improve the assessment of wind and wave energy availability at a semi-exposed coastal area. *Energy* **2024**, *309*, 132904. [[CrossRef](#)]
24. Carta, J.A.; Velázquez, S.; Cabrera, P. A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. *Renew. Sustain. Energy Rev.* **2013**, *27*, 362–400. [[CrossRef](#)]
25. Zhu, P.; Li, T.; Mirocha, J.D.; Arthur, R.S.; Wu, Z.; Fringer, O.B. A Moving-Wave Implementation in WRF to Study the Impact of Surface Water Waves on the Atmospheric Boundary Layer. *Mon. Weather Rev.* **2023**, *151*, 2883–2903. [[CrossRef](#)]
26. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2021. [[CrossRef](#)]
27. Hastie, T.; Friedman, J.; Tibshirani, R. *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2001. [[CrossRef](#)]
28. Kuhn, M.; Johnson, K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019; pp. 1–297. [[CrossRef](#)]
29. Calero, R.; Carta, J.A. Action plan for wind energy development in the Canary Islands. *Energy Policy* **2004**, *32*, 1185–1197. [[CrossRef](#)]
30. PORTUS (Puertos del Estado). Available online: <https://portus.puertos.es/#/> (accessed on 29 April 2025).
31. Bouhrim, H.; El Marjani, A.; Nechad, R.; Hajjout, I. Marine Science and Engineering Ocean Wave Energy Conversion: A Review. *J. Mar. Sci. Eng.* **2024**, *12*, 1922. [[CrossRef](#)]
32. CRAN: Package randomForest. Available online: <https://cran.r-project.org/web/packages/randomForest/index.html> (accessed on 7 May 2025).
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
34. Breiman, L.; Cutler, A.; Liaw, A.; Wiener, M. randomForest: Breiman and Cutlers Random Forests for Classification and Regression. CRAN: Contributed Packages. 2002. Available online: <https://rdrr.io/cran/randomForest/> (accessed on 7 May 2025).
35. CRAN: Package Ranger. Available online: <https://cran.r-project.org/web/packages/ranger/index.html> (accessed on 7 May 2025).
36. Wright, M.N. A Fast Implementation of Random Forests [R Package Ranger Version 0.17.0], CRAN: Contributed Packages. 2024. Available online: <https://rdrr.io/cran/ranger/> (accessed on 7 May 2025).
37. R: The R Stats Package. Available online: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html> (accessed on 7 May 2025).
38. CRAN: Package Parsnip. Available online: <https://cran.r-project.org/web/packages/parsonip/index.html> (accessed on 7 May 2025).
39. Kuhn, M.; Vaughan, D. A Common API to Modeling and Analysis Functions [R Package Parsnip Version 1.3.1], CRAN: Contributed Packages. 2025. Available online: <https://rdrr.io/cran/parsonip/> (accessed on 7 May 2025).
40. CRAN: Package Tune. Available online: <https://cran.r-project.org/web/packages/tune/index.html> (accessed on 7 May 2025).
41. Kuhn, M. Tidy Tuning Tools [R Package Tune Version 1.3.0], CRAN: Contributed Packages. 2025. Available online: <https://cran.r-project.org/web/packages/tune/> (accessed on 7 May 2025).
42. CRAN: Package Yardstick. Available online: <https://cran.r-project.org/web/packages/yardstick/index.html> (accessed on 7 May 2025).
43. Kuhn, M.; Vaughan, D.; Hvitfeldt, E. Tidy Characterizations of Model Performance [R Package Yardstick Version 1.3.2], CRAN: Contributed Packages. 2025. Available online: <https://rdrr.io/cran/yardstick/> (accessed on 7 May 2025).
44. Downloading and installing H2O-3. Available online: <https://github.com//h2oai/h2o-3/blob/master/h2o-docs/src/product/downloading.rst> (accessed on 7 May 2025).
45. tidymodels. Available online: <https://www.tidymodels.org/> (accessed on 7 May 2025).
46. CRAN: Package Recipes. Available online: <https://cran.r-project.org/web/packages/recipes/index.html> (accessed on 7 May 2025).
47. Kuhn, M.; Wickham, H.; Hvitfeldt, E. Preprocessing and Feature Engineering Steps for Modeling [R Package Recipes Version 1.3.1], CRAN: Contributed Packages. 2025. Available online: <https://cran.r-project.org/web/packages/recipes/> (accessed on 7 May 2025).

48. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. xgboost: Extreme Gradient Boosting, CRAN: Contributed Packages. 2014. Available online: <https://dmlc.r-universe.dev/xgboost> (accessed on 7 May 2025).
49. CRAN: Package Xgboost. Available online: <https://cran.r-project.org/web/packages/xgboost/index.html> (accessed on 7 May 2025).
50. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Toronto, ON, Canada, 13–17 August 2016; pp. 785–794. [CrossRef]
51. Kuhn, M. Classification and Regression Training [R Package Caret Version 7.0-1], CRAN: Contributed Packages. 2024. Available online: <https://rdrr.io/cran/caret/> (accessed on 7 May 2025).
52. CRAN: Package Caret. Available online: <https://cran.r-project.org/web/packages/caret/index.html> (accessed on 7 May 2025).
53. CRAN: Package Kernlab. Available online: <https://cran.r-project.org/web/packages/kernlab/index.html> (accessed on 7 May 2025).
54. Karatzoglou, A.; Smola, A.; Hornik, K. Kernel-Based Machine Learning Lab [R Package Kernlab Version 0.9-33], CRAN: Contributed Packages. 2024. Available online: <https://rdrr.io/cran/kernlab/> (accessed on 7 May 2025).
55. CRAN: Package doParallel. Available online: <https://cran.r-project.org/web/packages/doParallel/index.html> (accessed on 7 May 2025).
56. Corporation, M.; Weston, S. Foreach Parallel Adaptor for the “Parallel” Package [R Package doParallel Version 1.0.17], CRAN: Contributed Packages. 2022. Available online: <https://rdrr.io/cran/doParallel/> (accessed on 7 May 2025).
57. CRAN: Package kkn. Available online: <https://cran.r-project.org/web/packages/kkn/index.html> (accessed on 7 May 2025).
58. Schliep, K.; Hechenbichler, K. Weighted k-Nearest Neighbors [R Package kkn Version 1.4.1], CRAN: Contributed Packages. 2025. Available online: <https://rdrr.io/cran/kkn/> (accessed on 7 May 2025).
59. CRAN: Package rsample. Available online: <https://cran.r-project.org/web/packages/rsample/index.html> (accessed on 7 May 2025).
60. Frick, H.; Chow, F.; Kuhn, M.; Mahoney, M.; Silge, J.; Wickham, H. General Resampling Infrastructure [R Package Rsample Version 1.3.0], CRAN: Contributed Packages. 2025. Available online: <https://rdrr.io/cran/rsample/> (accessed on 7 May 2025).
61. Boehmke, B.; Greenwell, B. *Hands-On Machine Learning with R*; CRC Press: New York, NY, USA, 2019. [CrossRef]
62. Neuhauser, M. *Nonparametric Statistical Tests: A Computational Approach*; CRC Press: New York, NY, USA, 2011. [CrossRef]
63. Methods, C.; Triton, T. A Cost-Benefit Analysis of Additional Measurement Campaign Methods Reducing Uncertainty in Wind Project Energy Estimates Wind Project Financing. 2014. Available online: <https://www.vaisala.com/sites/default/files/documents/Triton-DNV-White-Paper.pdf> (accessed on 7 May 2025).
64. Miguel, J.V.P.; Fadigas, E.A.; Sauer, I.L. The Influence of the Wind Measurement Campaign Duration on a Measure-Correlate-Predict (MCP)-Based Wind Resource Assessment. *Energies* **2019**, *12*, 3606. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.