



Article

Surrogate Modelling Approach for Estimating Extreme Significant Wave Heights Using Meteorological Records

Belkin Pereira-Olmos ¹, Oscar E. Coronado-Hernández ^{2,*} , Manuel Saba ¹ , Modesto Pérez-Sánchez ³ and Helena M. Ramos ⁴

¹ Faculty of Engineering, Universidad de Cartagena, Cartagena 130001, Colombia; bpereirao@unicartagena.edu.co (B.P.-O.); msaba@unicartagena.edu.co (M.S.)

² Instituto de Hidráulica y Saneamiento Ambiental, Universidad de Cartagena, Cartagena 130001, Colombia

³ Departamento de Ingeniería Hidráulica y Medio Ambiente, Universitat Politècnica de València, 46022 Valencia, Spain; mopesan1@upv.es

⁴ Civil Engineering, Architecture and Environment Department, CERIS, Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal; hramos.ist@gmail.com

* Correspondence: ocoronadoh@unicartagena.edu.co

Abstract

This study presents a comprehensive, evidence-based methodological framework for estimating extreme significant wave heights associated with several return periods using meteorological records (wind speed and air temperature) in combination with machine learning algorithms. The proposed framework integrates a sequence of preliminary diagnostic analyses prior to extreme value modelling, including homogeneity tests to identify potential change points and trend analyses to assess the presence of stationary or non-stationary behaviour, thereby enabling the selection of an appropriate frequency analysis approach for computing annual maximum significant wave heights across several return periods. The methodology is applied to annual maximum significant wave height records obtained from offshore buoy stations in the Antilles region of the Caribbean Sea. Results from the machine learning presets demonstrate a strong relationship between extreme significant wave heights and the corresponding wind speed and temperature records, yielding coefficients of determination (R^2) of 0.75 and 0.88 for the validation and testing stages, respectively. When frequency analysis is conducted using both the traditional approach and the proposed machine learning-based methodology, high agreement is observed for return periods between 2 and 20 years, with R^2 values ranging from 0.98 to 0.89. For longer return periods, the agreement decreases, which is consistent with the limited length of the available wave height records and the supervised nature of the machine learning algorithms. In this sense, the proposed framework provides a robust alternative for estimating extreme significant wave heights in regions where direct wave observations are unavailable but meteorological records are accessible.



Academic Editor: Lorenzo Carnevale

Received: 16 May 2026

Revised: 22 June 2026

Accepted: 6 July 2026

Published: 8 July 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Caribbean Antilles; frequency analysis; machine learning; nonstationary distribution; return period; significant wave height; stationary distribution

1. Introduction

Significant wave height is a key parameter in coastal and offshore engineering, as it governs wave loading, overtopping processes, and the performance of both conventional and ecosystem-based coastal defence systems. Coastal cities and their beaches constitute environments where the effects of meteorological phenomena are most evident, as well as

the impacts and influence that climate change may exert on the frequency and intensity of such events [1]. In most coastal regions, these events lead to flooding, erosion, alterations in coastal morphology, and degradation of water quality [2]. As a result, infrastructure, local populations, and ecosystems near coastal zones become increasingly vulnerable, reducing their capacity to manage associated damage and their resilience and recovery potential [3].

The climate of the Caribbean Sea basin is governed by the Intertropical Convergence Zone (ITCZ) [4], which is characterised by a low-pressure belt and two primary seasons: a dry season and a windy season. Non-Stationary Extreme Value Analysis (NEVA) has demonstrated notable effectiveness in hydraulic engineering by explicitly accounting for meteorological variability within traditional Extreme Value Analysis (EVA) frameworks [5]. Design parameters, such as the significant design wave height associated with a prescribed return period, are closely linked to climatic trends [6].

A major challenge in frequency analysis is selecting a probability distribution that adequately represents the observed data trends [7]. Although numerous probability distribution functions are available for frequency analysis, only a limited number provide satisfactory fits [8]. Among the most widely applied distributions for extreme event analysis are the Generalized Extreme Value (GEV), Gumbel, and Weibull distributions [9].

Several studies have applied non-stationary modelling approaches to the analysis of significant wave height, if hydrological processes consist of a time-dependent deterministic component and a stationary stochastic component [10]. Techniques such as Singular Spectrum Analysis (SSA) and Recurrence Quantification Analysis have been employed to decompose time series into deterministic and stochastic components [11].

Barlow [12] applied a non-stationary piecewise linear extreme value model to NORA10 hindcast data, which provides a time series of significant wave height, wave direction, and seasonality. Their study demonstrated the effectiveness of piecewise linear models for non-stationary extreme-value analysis of storm-peak significant wave heights, incorporating directional and seasonal covariates in a region of the North Sea.

Extreme significant wave height (H_s) is a primary parameter in coastal and ocean engineering because it governs wave loading, overtopping, and the performance of coastal and offshore infrastructure, including nature-based and hybrid defence systems. Accurate estimation of extreme wave conditions is critical for the design and resilience assessment of hydraulic structures, coastal protection measures, and ecosystem-based solutions. Extreme value theory provides the statistical framework for estimating rare wave events beyond observed records, allowing reliable extrapolation to design return periods. Robust characterisation of extremes is increasingly necessary as climatic and environmental conditions evolve, potentially influencing wave climate patterns and the reliability of stationary assumptions commonly employed in engineering design.

Methodological advances in extreme wave analysis have highlighted the importance of accounting for non-stationarity, multi-site variability, and environmental dependencies in coastal design and hazard assessment [13]. Non-stationary extreme value models, such as time-dependent Generalized Extreme Value (GEV) and Peaks-Over-Threshold (POT) approaches, have been increasingly applied to regional wave climates to enhance the estimation of return values under changing conditions [14], with significant impacts on design wave heights and infrastructure risk evaluations [15,16]. Likewise, regional frequency analysis methods allow the incorporation of spatial heterogeneity in wave climate characteristics by pooling information from homogeneous regions [17], thereby reducing uncertainty and improving the robustness of extreme wave height quantile estimates across coastal domains [13]. In addition, recent research has explored machine-learning and data-driven techniques to validate or complement classical extreme value models, linking wave extremes with environmental drivers such as wind velocity, sea surface temperature,

and storm dynamics to improve predictive performance and ensure physical consistency with observed metocean processes [18]. Artificial neural networks have been employed for the forecasting of coastal tsunami events [19].

These methodological developments are directly relevant to ecosystem-based and hybrid coastal defence strategies, which require reliable characterisation of extreme wave conditions to optimise design, resilience, and multifunctionality of nature-based solutions alongside engineered structures [20]. By integrating homogeneity and trend analyses with both statistical and machine-learning techniques, researchers and practitioners can derive design wave estimates that are statistically robust and physically meaningful for engineering applications across diverse coastal environments [21]. Such integrated approaches provide critical foundations for evaluating the performance of coastal infrastructure and nature-based solutions under extreme events, supporting sustainable and resilient coastal management in the face of climate variability and change [22].

De Leo [23] proposed a methodology based on annual maxima of significant wave height and peak period series spanning from 1979 to 2018. Both stationary EVA and NEVA were conducted using a three-parameter stationary Generalized Extreme Value (GEV) distribution and a four-parameter non-stationary Generalized Extreme Value (N-GEV) distribution. The Akaike Information Criterion (AIC) was employed to compare the GEV and N-GEV models, selecting the distribution with the lowest AIC value as the best-performing model.

In the insular zone of Latin America [23], changes in significant wave height along the Cuban coastline between 1950 and 2022 are shown using ERA5 reanalysis data with a temporal resolution of 6 h and a spatial resolution of 0.5° latitude. Their results indicate that interannual variability in significant wave height predominantly occurs between November and March, when wave heights exceed 1 m. This variability is attributed to the influence of extratropical lows and migratory anticyclones affecting the archipelago during this period. The highest values were observed in the first quarter of the year, particularly in February, reaching 1.6 m due to increased frequency and intensity of synoptic-scale meteorological systems. Conversely, the lowest values were recorded in August and September, with wave heights below 0.8 m, likely resulting from weak pressure gradients associated with the Azores–Bermuda subtropical anticyclone.

In this study, a methodological framework is proposed to jointly analyse stationary and non-stationary trends in annual maximum significant wave heights for the Caribbean Sea basin. Annual maxima were analysed using Gumbel, Generalized Extreme Value (GEV), and Pearson III, and Log-Pearson III probability distributions. Comparable approaches have been adopted in precipitation frequency studies [24], where annual maximum 24 h precipitation series from multiple meteorological stations were analysed using several probability distributions and evaluated using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [25].

Additionally, XLSTAT 2026 software was employed to assess the homogeneity of annual maximum significant wave height series using the Pettitt, Standard Normal Homogeneity Test (SNHT), Buishand, and Von Neumann tests [26]. These homogeneity tests have been widely applied in climatological studies, such as that conducted by [24], who analysed precipitation and temperature data from meteorological stations in Mexico to validate data consistency and identify years exhibiting significant variability through *p*-value analysis. Given the variety of approaches proposed by several researchers, their integration is required to enhance the robustness of analyses aimed at predicting extremely significant wave height events.

In this context, this study proposes a comprehensive methodology for predicting annual maximum significant wave heights based on machine learning presets and meteorological data.

logical records of wind speed and temperature. The novelty of this research is particularly valuable because missing values in significant wave height series can be reliably imputed using ML models trained on wind speed and temperature data. The proposed framework integrates homogeneity analysis and trend detection, and derives annual maximum significant wave heights from meteorological records modelled using 28 distinct machine learning (ML) configurations. Finally, the study compares traditional methods for computing extreme significant wave height values for several return periods with those obtained using the proposed methodology. This paper is organised as follows: Section 2 presents the generalisable methodology proposed in this study, which can be applied to different regions; Section 3 describes the application of the proposed methodology to the Antilles region of the Caribbean Sea; and Section 4 presents the conclusions of the study.

2. Materials and Methods

2.1. Methodology

The methodology adopted in this study is summarised in Figure 1 and is structured into three main stages. First, the study area is defined, and wave buoy records are selected. The variable of interest corresponds to the annual maximum values of significant wave height, H_s (Stage I). In addition, wind speed and temperature values need to be collected. In Stage II, a sequence of statistical analyses is conducted to assess data consistency and temporal behaviour. Initially, homogeneity tests are applied to the annual maximum H_s series to detect potential change points or inhomogeneities. The proposed methodology is generalisable and can be applied to different regions.

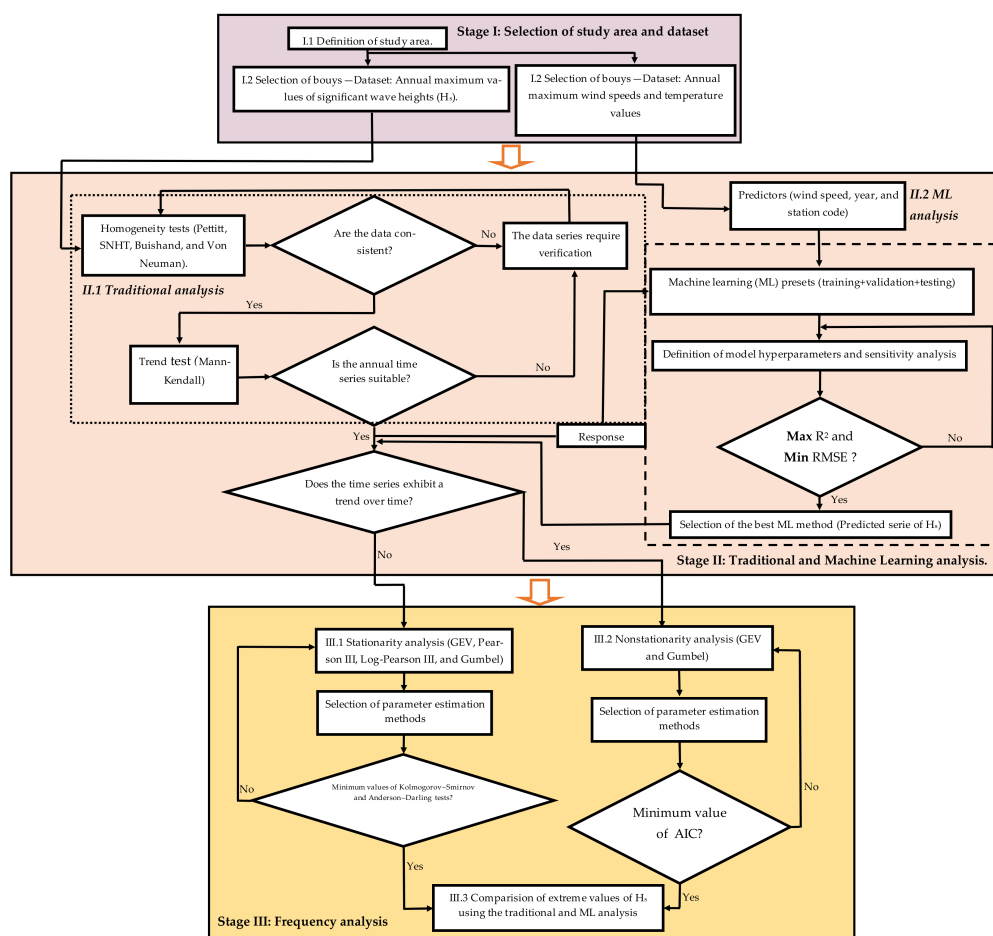


Figure 1. Flowchart of the methodology adopted in this study for the analysis of extreme significant wave heights.

These tests include the Pettitt test, the SNHT, the Buishand range test, and the Von Neumann ratio test. The Mann–Kendall test is employed to assess the presence and statistical significance of monotonic trends in the H_s time series. Subsequently, wind speed records at each station are analysed in relation to the corresponding significant wave heights to investigate their interdependence and the role of atmospheric forcing employing various machine learning algorithms. In Stage III, extreme value analysis is conducted to estimate significant wave heights for multiple return periods using both traditional approaches and the methodology proposed in this study, in which machine learning techniques are shown to effectively impute missing H_s values, thereby enabling more reliable estimation of extreme wave conditions across a range of return periods.

2.1.1. Homogeneity Test

Homogeneity tests were applied to the annual maximum significant wave height (H_s) series to assess their statistical consistency and to identify potential change points. The analyses were performed using the XLSTAT software (trial version). The Pettitt test, the SNHT proposed by Alexandersson [27], the Buishand range test, and the Von Neumann ratio test were employed to evaluate whether the time series exhibited homogeneous behaviour over time.

The hypothesis-testing framework considered a null hypothesis (H_0) assuming homogeneity of the series, and an alternative hypothesis (H_1) indicating the presence of inhomogeneities or structural breaks. Decisions regarding acceptance or rejection of the null hypothesis were based on a predefined significance level, representing the probability of incorrectly rejecting H_0 . These tests are particularly suitable for detecting abrupt changes or shifts in the statistical properties of hydrometeorological time series.

The Pettitt test is a non-parametric rank-based method derived from the Mann–Whitney test, designed to identify the most probable point in time at which a change occurs within a series [28]. By analysing the ranks of the observations, the test determines the location of a potential break without assuming any specific probability distribution.

The SNHT was applied to detect heterogeneities in the form of abrupt changes in the mean value of the series. An extended version of this test, proposed by Alexandersson and Moberg (1997) [27], allows for the identification of linear trends in the mean by comparing the average of the first k observations with that of the remaining $n - k$ observations.

The Buishand test, which has Bayesian foundations, was used to detect shifts in the mean of the data series by means of cumulative deviations from the average. This test provides critical values for assessing homogeneity and can be applied regardless of the underlying probability distribution of the variable.

Finally, the Von Neumann ratio test, a parametric approach, was employed to detect loss of randomness in the series caused by unspecified deterministic components [29]. This test is particularly useful for identifying serial dependence or structural irregularities in hydrometeorological records.

2.1.2. Machine Learning Methodology

A total of 28 machine learning algorithms were evaluated to assess their suitability for modelling extreme significant wave height (H_s) time series. For each wave buoy station, wind speed (V_w) was used as the predictor variable and related to the corresponding H_s values associated with extreme events. In addition, the year and station code were used as predictors for a general analysis. The selection of predictors was based on data availability at each station.

The dataset was split into training and test sets, with 90% allocated to training. A five-fold cross-validation strategy was implemented to ensure robust model performance

and minimise overfitting. The remaining 10% of the dataset was reserved for independent testing and model evaluation. The validation scheme adopted in this study is particularly suitable for relatively small datasets. Prior to partitioning, the dataset was randomly shuffled to ensure that different wave conditions were adequately represented in the training, validation, and testing subsets. Since extreme significant wave heights are treated as random variables rather than as a time series intended for forecasting, preserving the temporal sequence of observations was not necessary during the data-splitting process.

Table 1 summarises the machine learning models considered in this study, grouped by methodological category. These include linear regression models, decision tree models, ensemble tree methods, support vector machines with different kernel functions, Gaussian process regression models, artificial neural networks with varying architectures, and kernel-based regression approaches.

Table 1. Machine learning models per category.

Machine Learning Category	Preset/Model
Linear Regression Models	Linear Interactions Linear Robust Linear Stepwise Linear Efficient Linear Least Squares
Decision Tree Models	Fine Tree Medium Tree Coarse Tree
Ensemble Tree Methods	Boosted Trees Bagged Trees
Support Vector Machines (SVM)	Linear SVM Quadratic SVM Cubic SVM Efficient Linear SVM
Gaussian Kernel SVM	Fine Gaussian SVM Medium Gaussian SVM Coarse Gaussian SVM
Gaussian Process Regression (GPR)	Squared Exponential GPR Matern 5/2 GPR Exponential GPR Rational Quadratic GPR
Artificial Neural Networks (ANN)	Narrow Neural Network Medium Neural Network Wide Neural Network Bilayered Neural Network Trilayered Neural Network
Kernel-based Regression Methods	SVM Kernel Least Squares Regression Kernel

All computations were performed using MATLAB R2024b, specifically through the Regression Learner App. Model performance was assessed using two widely adopted statistical indicators: the Root Mean Square Error (RMSE) and the coefficient of determination (R^2).

The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (H_{s,T} - H_{s,P})^2} \quad (1)$$

where $H_{s,T}$ represents the observed (true) values of significant wave height, $H_{s,P}$ denotes the predicted values, and N is the total number of observations.

The coefficient of determination (R^2) is given by [30]:

$$R^2 = 1 - \frac{\sum_{i=1}^N (H_{s,T} - H_{s,P})^2}{\sum_{i=1}^N (H_{s,T} - \overline{H_{s,T}})^2} \quad (2)$$

where $\overline{H_{s,T}}$ is the mean of the observed H_s values.

These metrics were used to identify the most suitable machine learning models for reproducing extreme wave heights at each station.

In this research, ensemble tree-based methods achieved the most accurate performance for the analysed problem [30]. This preset was configured with a minimum leaf size of 8, a total of 30 learners, and a learning rate of 0.1. Accordingly, the underlying structure of the algorithms is outlined below. Four adaptive boosting schemes are considered.

- Adaptive boosting for binary classification

This method is a binary algorithm commonly applied in regression-based classification problems. It computes the weighted classification error as [30]:

$$\varepsilon_t = \sum_{n=1}^N d_n^{(t)} \mathbb{I}(y_n \neq h_t(x_n)) \quad (3)$$

where x_n denotes the predictor vector, y_n is the true class label, h_t represents the prediction of the t -th weak learner, $\mathbb{I}(\cdot)$ is the indicator function, and $d_n^{(t)}$ corresponds to the weight assigned to the n -th observation at iteration t .

- Adaptive boosting for multiclass classification

For problems involving K classes, this algorithm minimises a weighted pseudo-loss, defined as [30]:

$$\varepsilon_t = \frac{1}{2} \sum_{n=1}^N \sum_{k \neq y_n} d_n^{(t)} (1 - h_t(x_n, y_n) + h_t(x_n, k)) \quad (4)$$

where $h_t(x_n, k)$ denotes the confidence associated with assigning observation x_n to class k .

- Gentle adaptive boosting

This approach integrates characteristics of the two previous algorithms and evaluates performance using the mean squared error, expressed as [30]:

$$\varepsilon_t = \sum_{n=1}^N d_n^{(t)} (\tilde{y}_n - h_t(x_n))^2 \quad (5)$$

where $\tilde{y}_n$ represents the transformed response variable.

- Adaptive logistic regression

This binary classification method focuses on minimising the binomial deviance, which is given by [30]:

$$\varepsilon_t = \sum_{n=1}^N w_n \log(1 + \exp(-2y_n f(x_n))) \quad (6)$$

where w_n denotes the normalised observation weights and $f(x_n)$ corresponds to the predicted classification score.

2.1.3. Trend Test

The Mann–Kendall test was employed to assess whether the significant wave height (H_s) time series exhibits a monotonic increasing or decreasing trend. This nonparametric test is widely used in hydrology and coastal engineering to detect systematic changes in environmental time series without requiring assumptions about the underlying probability distribution [31].

For a time series $x_1, x_2, \dots, x_n$, the Mann–Kendall test statistical S is defined as:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_i) \quad (7)$$

where x_j represents the observed value in year j and x_i corresponds to the value observed in year i .

The sign function evaluates whether the difference between paired observations is positive, negative, or zero. The null hypothesis of the Mann–Kendall test assumes that the data are independent and randomly ordered over time. However, the presence of positive autocorrelation within the series can increase the probability of detecting a statistically significant trend when none exists, whereas negative autocorrelation may mask existing trends. Although this limitation is well recognised, it is often neglected in practical applications, and only a limited number of studies explicitly account for autocorrelation effects in trend analysis [32].

2.1.4. Frequency Functions and Statistical Measures

For the stationary extreme-value analysis of the annual-maximum significant wave height series, three commonly used probability distributions were adopted: the Gumbel, GEV, Pearson III, and Log-Pearson III distributions. These frequency functions are widely applied to characterise the probabilistic behaviour of extreme events and to estimate design wave heights associated with prescribed return periods.

The selected distributions allow the quantification of the probability of extreme H_s values exceeding those observed in the historical record. Estimation of distribution parameters and computation of exceedance probabilities were performed using the HEC-SSP 2.3.1 software package.

- Gumbel Distribution

The Gumbel distribution is one of the most widely used probability models in extreme value theory for representing maxima. It has proven particularly effective when the underlying data are approximately normally or exponentially distributed. The Gumbel distribution has been extensively applied to the analysis of extreme environmental variables, including significant wave height, wind speed, rainfall, offshore conditions, and climate-related extremes [32].

The probability density function (PDF) of the Gumbel distribution is given by:

$$f(x) = \frac{1}{a} \exp \left[-\frac{x-u}{a} - \exp \left(-\frac{x-u}{a} \right) \right] \quad (8)$$

where a is the scale parameter and u is the location (mode) parameter, which represents the central tendency of the distribution.

- GEV Distribution

The GEV distribution constitutes a family of continuous probability distributions within extreme value theory. It unifies the Gumbel, Fréchet, and Weibull distributions into a single framework, commonly referred to as Type I, Type II, and Type III extreme value distributions, respectively. This unification enables greater flexibility in modelling extreme events by allowing the shape of the distribution tail to vary. The GEV distribution is defined by three parameters: location (u), scale (a), and shape (k). When the shape parameter $k = 0$, the GEV reduces to the Gumbel distribution; when $k > 0$, it corresponds to the Fréchet distribution; and when $k < 0$, it corresponds to the Weibull distribution.

According to the extreme value theorem, the GEV distribution represents the limiting distribution of properly normalised maxima derived from sequences of independent and identically distributed random variables. Consequently, it is widely used as an approximation for modelling the maxima of long, finite time series of environmental data, including extreme wave heights [33].

The probability density function adopted in this study is expressed as:

$$f(x) = \frac{1}{a} \left[1 - \frac{k}{a}(x - u) \right]^{\frac{1}{k}-1} \exp \left\{ - \left[1 - \frac{k}{a}(x - u) \right]^{1/k} \right\} \quad (9)$$

where k is the shape parameter.

- Pearson III and Log-Pearson III distributions

These distributions have been used to characterise significant wave heights under extreme conditions. The probability density function of the Pearson III distribution is given by:

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} (x - m)^{\lambda-1} \exp[-\alpha(x - m)] \quad (10)$$

where α , λ , and m are the parameters of the distribution. Γ corresponds to the gamma function.

Similarly, the Log-Pearson III distribution is yielded by:

$$f(x) = \frac{\alpha^\lambda}{x\Gamma(\lambda)} (\ln x - m)^{\lambda-1} \exp[-\alpha(\ln x - m)] \quad (11)$$

- Nonstationary distributions

To account for potential temporal variability and long-term climatic trends in extreme significant wave heights, non-stationary versions of the aforementioned distributions were considered. In this framework, selected parameters of the stationary distributions were allowed to vary as explicit functions of time.

The location parameter was modelled as a linear function of time:

$$u(t) = \gamma_0 + \gamma_1 t \quad (12)$$

while the scale parameter was expressed as an exponential function of time to ensure positivity:

$$a(t) = \exp(\gamma_0 + \gamma_1 t) \quad (13)$$

where t denotes time, and γ_0 and γ_1 are constant coefficients estimated from the data. In this study, the shape parameter was assumed to remain constant throughout the analysis, a common assumption in non-stationary extreme value modelling when data availability is limited, and vice versa for the location parameter.

The non-stationary extreme value analysis of the annual maximum significant wave height series was carried out using the R programming environment, which offers com-

prehensive tools for statistical modelling and extreme value analysis. Specifically, the *nsextremes* package was employed to estimate time-varying parameters and to evaluate non-stationary probability distributions [34].

2.1.5. Goodness-of-Fit Tests and Parameter Estimation Methods

Model selection was conducted using the Akaike Information Criterion (AIC), implemented within the R environment for comparing stationary and non-stationary functions. The AIC facilitates an objective comparison among competing frequency distributions by balancing goodness of fit against model complexity through penalisation of the number of estimated parameters. Consequently, distributions with lower AIC values are considered to provide a better representation of the observed data. This test is modelled as follows:

$$AIC = 2k - 2\ln(L) \quad (14)$$

where k = number of estimated parameters in the model and L = maximum value of the likelihood function of the model.

For stationary distributions, the parameters were estimated using the L-moments method, as it has been shown to perform reliably in the analysis of rare extreme events. The method is defined by the following expressions:

$$\hat{\lambda}_1 = \hat{\beta}_0 \quad (15)$$

$$\hat{\lambda}_2 = 2\hat{\beta}_1 - \hat{\beta}_0 \quad (16)$$

$$\hat{\lambda}_3 = 6\hat{\beta}_2 - 6\hat{\beta}_1 - \hat{\beta}_0 \quad (17)$$

$$\hat{\lambda}_4 = 20\hat{\beta}_3 - 30\hat{\beta}_2 - 12\hat{\beta}_1 - \hat{\beta}_0 \quad (18)$$

where $\hat{\lambda}_L$ correspond to L-moments and $\hat{\beta}_0$ is the observed H_s .

After that, the Kolmogorov–Smirnov and the Anderson–Darling tests were performed to select the best stationary distribution. The analysis was conducted using the software HEC-SPP version 2.3.1 developed by the U.S. Army Corps of Engineers.

2.2. Study Area and Data

The Antilles constitute an extensive archipelago in the Caribbean region, forming a natural boundary between the Caribbean Sea and the Atlantic Ocean. The archipelago extends from areas near the southern tip of the Florida Peninsula (USA) and the Yucatán Peninsula (Mexico) to the northern coast of South America, particularly Venezuela. Owing to their location, the Antilles are exposed to a wide range of metoceanographic conditions influenced by trade winds, tropical storms, and large-scale atmospheric circulation.

For this study, eight wave buoys located in the Caribbean Sea and the adjacent Atlantic Ocean around the Antilles were selected (Figure 2). These buoys provide long-term records suitable for analysing extreme wave conditions. The analysis focuses on the annual maximum significant wave height (H_s) series, which were examined using both stationary and non-stationary frequency distribution functions. This approach enables the assessment of wave climate variability and the estimation of design wave heights for different return periods.

Data for this study were obtained from the National Oceanic and Atmospheric Administration (NOAA). NOAA is the United States federal agency responsible for meteorological and oceanographic observations, forecasting, and long-term environmental records, and it plays a key role in supporting international initiatives related to oceanography, fisheries, climate, space, and atmospheric sciences. Table 2 lists the wave buoy stations selected for this research. The available records for these stations range from 2005 to 2009 and extend to

the present. Accordingly, a common study period spanning from 2005 to 2023 was adopted for the analysis of annual maximum significant wave heights.

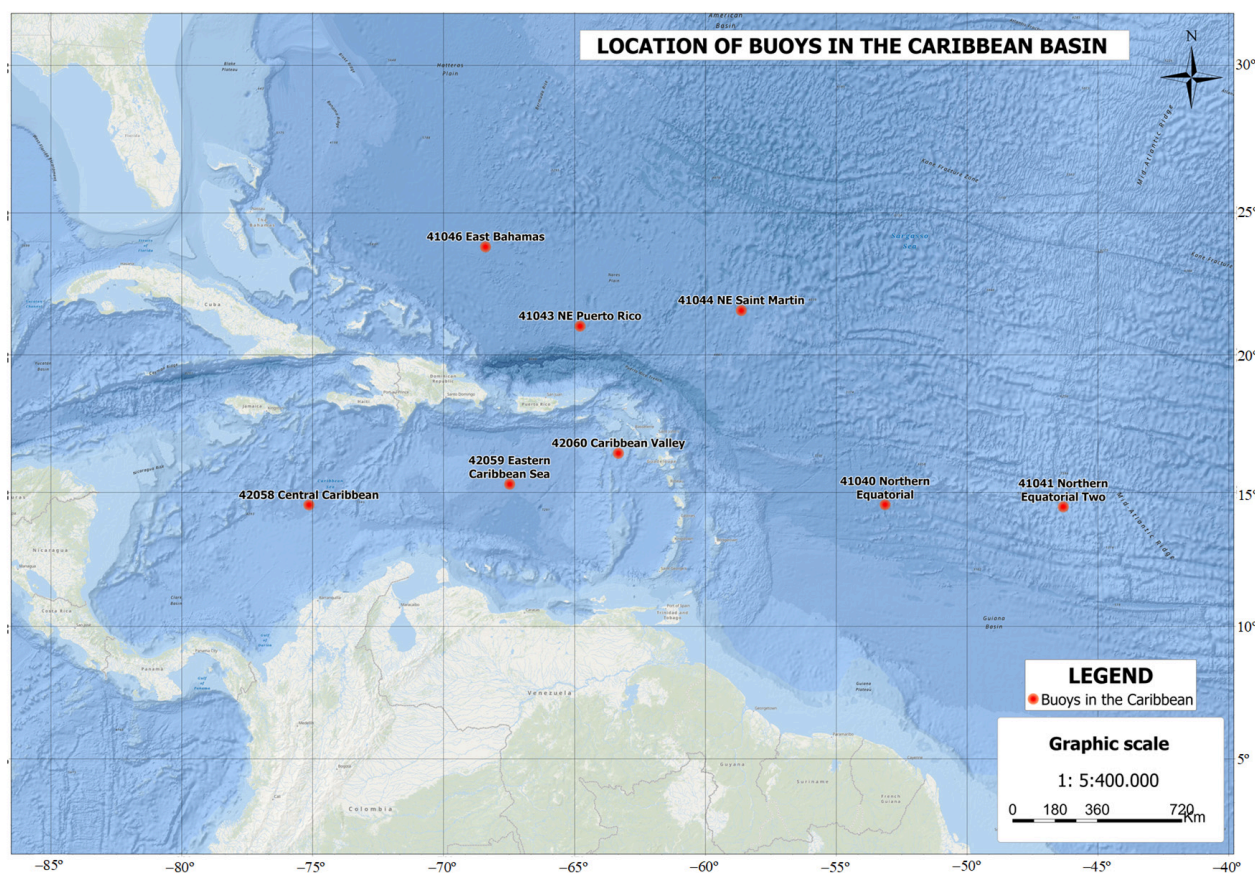


Figure 2. Location of buoys.

Table 2. Geographic location of buoys.

No	Station Name	Code	Location Coordinates		Year of Buoy Installation
			Latitude	Longitude	
1	Northern Equatorial—470 NM east of Martinique	41040	14°32'9" N	53°8'11" W	2005
2	Northern Equatorial Two—890 NM east of Martinique	41041	14°27'12" N	46°19'36" W	2005
3	NE Puerto Rico—170 NM NNE of San Juan, PR	41043	21°1'33" N	64°47'36" W	2007
4	NE Saint Martin—330 NM NE of Saint Martin Island	41044	21°34'57" N	58°37'48" W	2009
5	East Bahamas—335 NM east of San Salvador Island, Bahamas	41046	23°49'18" N	68°23'36" W	2007
6	Central Caribbean—210 NM SSE of Kingston, Jamaica	42058	14°31'31" N	75°8'25" W	2005
7	Eastern Caribbean Sea—180 NM SSW of Ponce	42059	15°18'0" N	67°28'58" W	2007
8	Caribbean Valley—63 NM WSW of Montserrat	42060	16°26'2" N	63°19'43" W	2009

3. Results and Discussion

3.1. Statistical Analysis and Homogeneity Test

This section presents the statistical characterisation and homogeneity assessment of the significant wave height (H_s) series in the Antilles region of the Caribbean Sea. Table 3 summarises the descriptive statistical parameters derived from the original annual maximum H_s series provided by NOAA.

Table 3. Statistical parameters of the significant wave height annual maximum series.

Station Name	Code	Mean (cm)	Standard Deviation (cm)	Coefficient of Variation (cm)
Northern Equatorial	41040	479.42	111.93	4.28
Northern Equatorial Two	41041	524.94	165.18	3.18
NE Puerto Rico	41043	592.41	233.94	2.53
NE Saint Martin	41044	725.67	258.22	2.81
East Bahamas	41046	672.00	262.85	2.56
Central Caribbean	42058	484.71	165.42	2.93
Eastern Caribbean Sea	42059	478.24	211.54	2.26
Caribbean Valley	42060	366.64	172.02	2.13

The results indicate a clear spatial variability in wave conditions across the study area. Stations in the northern sector of the region generally record higher mean significant wave heights, which can be attributed to greater exposure to Atlantic swells and stronger wind forcing. In particular, the NE Saint Martin (41044), East Bahamas (41046), and NE Puerto Rico (41043) stations exhibit the highest mean values. This spatial pattern suggests an increase in extreme wave heights towards the northern part of the Caribbean–Atlantic transition zone. In contrast, the lowest mean significant wave height was observed at the Caribbean Valley station (42060), located near the Lesser Antilles west of Montserrat and Guadeloupe, where wave conditions are comparatively more sheltered.

To assess the consistency and reliability of the H_s time series, homogeneity tests were applied using the XLSTAT software. These tests included the Pettitt test, the Standard Normal Homogeneity Test (SNHT), the Buishand range test, and the Von Neumann ratio test, to identify potential change points or structural inhomogeneities in the data.

The results of the homogeneity tests are presented in Table 4. The Northern Equatorial Two (41041), Central Caribbean (42058), and Eastern Caribbean Sea (42059) stations exhibit comparable behaviour, with relatively high p -values in the Pettitt and Buishand tests (ranging approximately between 0.95 and 0.55), suggesting no strong evidence against the null hypothesis of homogeneity. In contrast, the SNHT and Von Neumann tests yielded lower p -values for these stations (between 0.62 and 0.11), indicating greater sensitivity to subtle changes in the series. Notably, the latter two stations are located in the central Caribbean basin, offshore Colombia and Venezuela, respectively, where regional atmospheric and oceanographic processes may influence wave variability.

Similarly, the NE Puerto Rico (41043), East Bahamas (41046), and Caribbean Valley (42060) stations show comparable trends across the applied tests. For the East Bahamas and Caribbean Valley stations, the Pettitt test yielded a relatively low p -value (approximately 0.18), whereas the SNHT, Buishand, and Von Neumann tests yielded more consistent, higher values. Among these stations, East Bahamas exhibited the highest p -values, ranging from 0.98 to 0.88, followed by NE Puerto Rico, with values ranging from 0.93 to 0.72. The Caribbean Valley station presented slightly lower but still comparable values, ranging from 0.63 to 0.57.

In this analysis, the significance level (β) represents the probability of rejecting the null hypothesis of homogeneity. A threshold value of 5% was adopted, corresponding to the commonly used confidence level in hydrometeorological studies.

Table 4. Results of homogeneity tests for all wave buoy stations.

Station Name	Test			
	Pettitt	SNHT	Buishand	Von Neumann
Northern Equatorial	0.739	0.448	0.472	0.645
Northern Equatorial Two	0.804	0.529	0.775	0.346
NE Puerto Rico	0.703	0.936	0.872	0.722
NE Saint Martin	0.182	0.043	0.059	0.020
East Bahamas	0.184	0.983	0.983	0.889
Central Caribbean	0.947	0.617	0.919	0.339
Eastern Caribbean Sea	0.611	0.110	0.547	0.291
Caribbean Valley	0.261	0.639	0.616	0.572

The NE Saint Martin station (41044) exhibited the most atypical behaviour among the analysed series. This station recorded the lowest p -values across all applied tests, ranging from 0.18 to 0.02, and showed clear indications of inhomogeneity according to the SNHT and Von Neumann criteria. These results suggest a higher probability of structural changes or regime shifts within the time series, which may be associated with localised meteorological influences, data inconsistencies, or changes in measurement conditions. Consequently, this station warrants particular attention in subsequent trend and extreme value analyses.

Figures 3 and 4 illustrate the behaviour of the significant wave height time series and the outcomes of the homogeneity tests for two contrasting stations in the Caribbean basin: Caribbean Valley (42060) and NE Saint Martin (41044).

Figure 3 shows the annual evolution of significant wave heights for the Caribbean Valley station (42060). The time series exhibits moderate interannual variability, with values generally oscillating around the mean and no persistent upward or downward trend. Although isolated peaks are observed, they do not appear to persist over time. This relatively stable behaviour is consistent with the homogeneity test results, which yield values ranging from 0.261 to 0.639. As these values exceed the 5% significance level, the null hypothesis of homogeneity cannot be rejected, indicating that the series does not exhibit statistically significant change points.

The corresponding homogeneity test bars in Figure 3 further support this interpretation. While the Pettitt test yields a lower value compared to the other tests, it remains above the critical threshold. The SNHT, Buishand, and Von Neumann tests yield moderate to high p -values, reinforcing the conclusion that the Caribbean Valley series is homogeneous over the analysed period.

In contrast, Figure 4 highlights markedly different behaviour at the NE Saint Martin station (41044). The time series displays pronounced variability, characterised by sharp fluctuations and abrupt changes from one year to the next. Such behaviour suggests the possible presence of structural changes in the data, which may be associated with extreme wave events, local exposure to Atlantic swell systems, or changes in measurement conditions.

This visual evidence is corroborated by the homogeneity test results presented in Figure 4. The SNHT and Von Neumann tests yield p -values below the 5% significance level, indicating non-homogeneity and suggesting a change point in the series. Furthermore, the Pettitt and Buishand tests return values of 0.182 and 0.059, respectively, which are the lowest among all the analysed stations. These results confirm that the NE Saint Martin series exhibits the strongest deviation from homogeneity.

The analysis of Figures 3 and 4 demonstrates a clear contrast between the central Caribbean station (42060), which exhibits stable, homogeneous behaviour, and the northern

Caribbean station (41044), which shows significant non-homogeneity. The red and green dashed lines indicate the upper and lower thresholds, respectively.

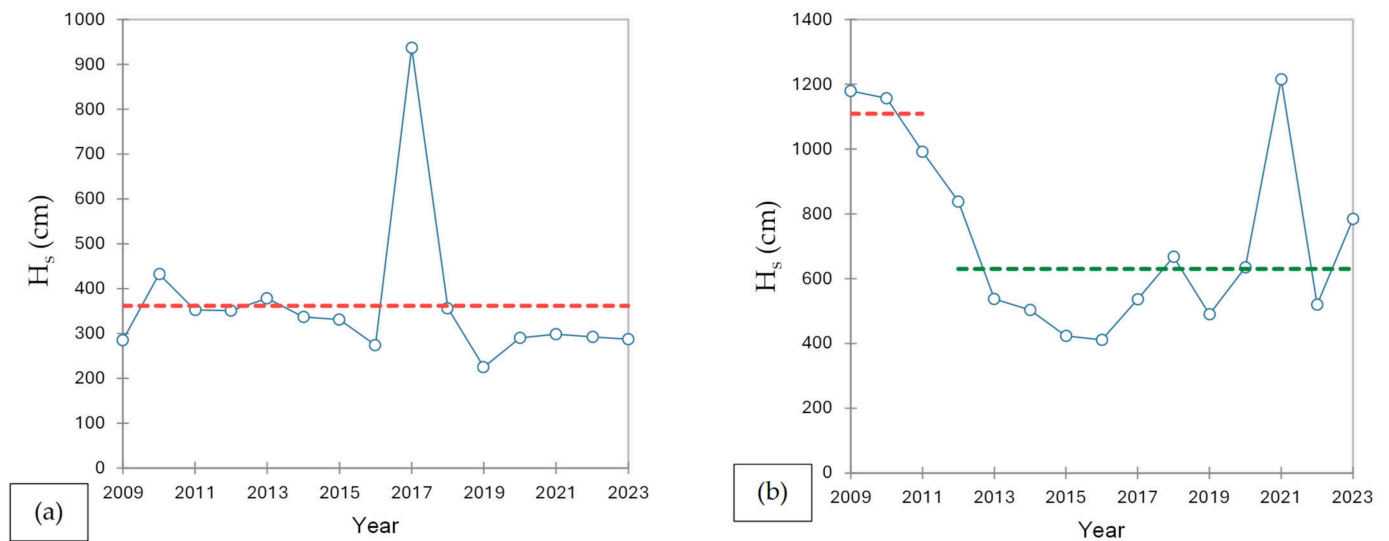


Figure 3. SNHT homogeneity test results over the study period for: (a) the Caribbean Valley station (42060); and (b) the NE Saint Martin station (41044).

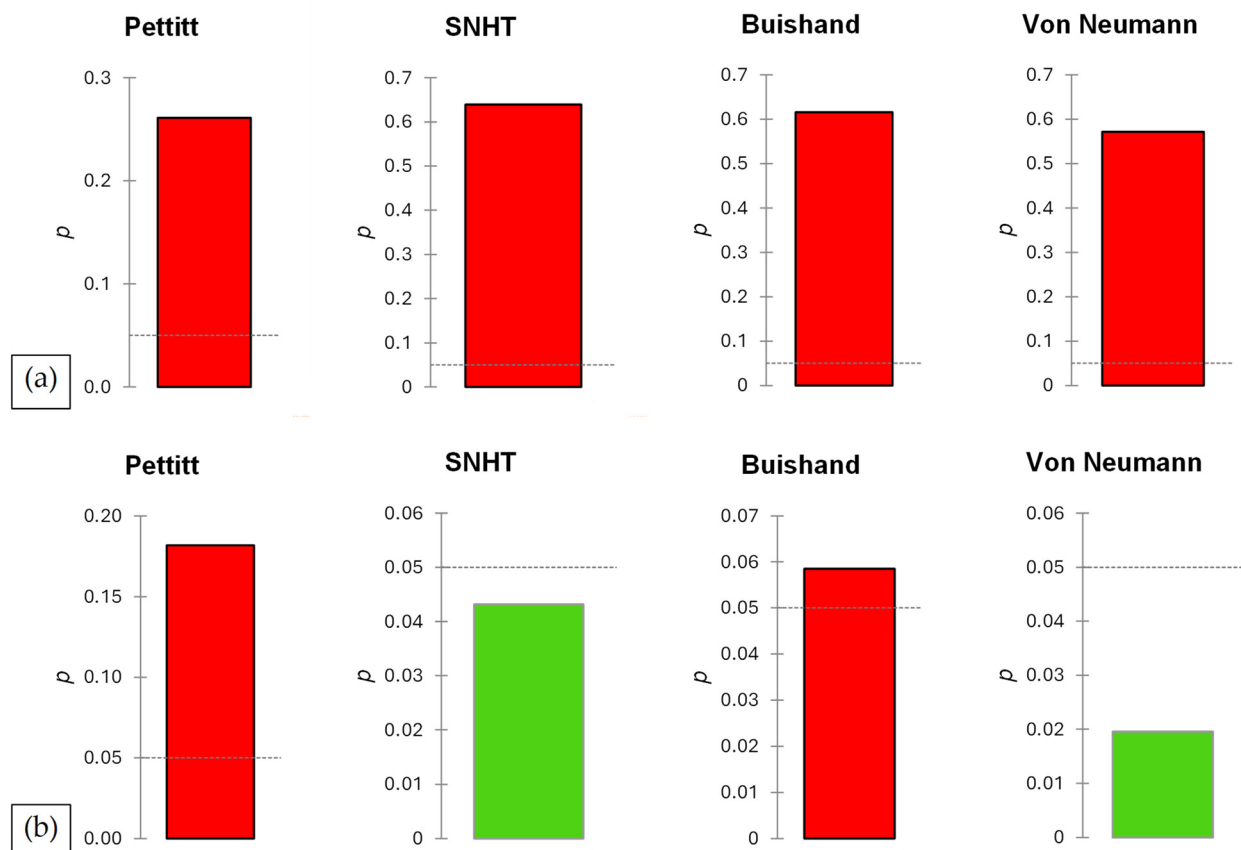


Figure 4. Probability (p -value) results of the Pettitt, SNHT, Buishand, and Von Neumann homogeneity tests for: (a) the Caribbean Valley station (42060); and (b) the NE Saint Martin station (41044).

3.2. Trend Test Results

The trend analysis was carried out using the XLSTAT 2026 software, applying the non-parametric Mann–Kendall test. Table 5 summarises the results obtained for all analysed

stations. According to the Mann–Kendall criterion, none of the time series exhibits a statistically significant trend. Consequently, all stations are classified as showing no trend (NT), indicating the absence of a consistent monotonic increase or decrease in the annual maximum significant wave height (H_s) over the study period. These results suggest that, to date, the analysed records have remained statistically stationary in their long-term temporal behaviour.

Table 5. Results of the Mann–Kendall trend test for annual maximum significant wave height (H_s).

Station Name	Trend Classification	<i>p</i> -Value
Northern Equatorial	NT	1.00
Northern Equatorial Two	NT	0.753
NE Puerto Rico	NT	0.41
NE Saint Martin	NT	0.276
East Bahamas	NT	0.773
Central Caribbean	NT	0.519
Eastern Caribbean Sea	NT	0.484
Caribbean Valley	NT	0.113

Note: NT = no statistically significant trend detected at the 5% significance level.

The application of the Mann–Kendall test therefore confirms that the annual maximum H_s series do not display significant temporal trends, supporting the assumption of stationarity adopted in the subsequent frequency analyses.

Although the non-stationary analysis was not found to be fully suitable for the stations analysed in the Antilles Sea region, the Akaike Information Criterion (AIC) values were nonetheless computed for both Gumbel and Generalised Extreme Value (GEV) distributions, allowing for time-varying parameters. Successful model fitting was achieved for only three time series—East Bahamas, NE Saint Martin, and Northern Equatorial Martinique—whereas numerical convergence issues or data processing errors were encountered for the remaining stations.

The results of the AIC-based model comparison are summarised in Table 6. For the East Bahamas station (41046), the non-stationary GEV model with a time-varying scale parameter provided the best fit, yielding the lowest AIC value (224.79). In contrast, for the NE Saint Martin station (41044), the stationary Gumbel distribution yielded the lowest AIC (212.11), indicating no improvement when incorporating non-stationarity. A similar behaviour was observed for the Northern Equatorial Martinique station, where stationary models outperformed their non-stationary counterparts.

Table 6. Akaike Information Criterion (AIC) values for stationary and non-stationary Gumbel and GEV distributions.

Probability Distribution	Function	AIC Test		
		East Bahamas	Saint Martin	Martinica
Gumbel	Stationary	230.8	212.1	209.5
Gumbel with $u(t)$	Nonstationary	232.8	213.2	211.5
Gumbel with $a(t)$	Nonstationary	232.5	212.5	211.4
GEV	Stationary	225.9	212.6	211.5
GEV with $u(t)$	Nonstationary	225.8	214.6	213.5
GEV with $a(t)$	Nonstationary	224.8	212.6	213.4

Note: The grey cells represent the best model.

These results further support the Mann–Kendall trend analysis, confirming that the annual maximum significant wave height series at the analysed stations predominantly exhibit stationary behaviour over the study period.

3.3. Machine Learning Results

This section presents the obtention of a ML model capable of simulating extreme annual events of significant wave height (H_s), based on the corresponding wind speed (V_w), temperature, station code, and the year of the occurrence of the value of H_s . Figure 5 presents the time series for the Puerto Rico station, where a clear correspondence between peaks and troughs in both variables is evident, indicating a strong physical linkage between extreme wave generation and wind forcing.

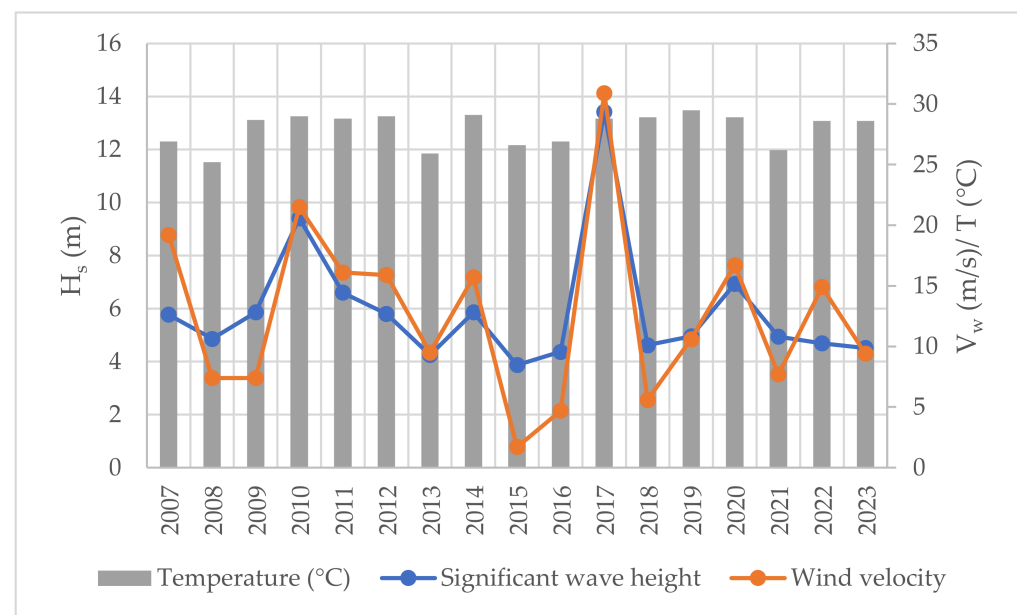


Figure 5. Temporal comparison of annual maximum significant wave height (H_s), wind speed (V_w), and temperature (T) at the Puerto Rico station, highlighting the correspondence between extreme wave and wind events.

The highest joint extreme event occurred in 2017, when H_s reached 13.42 m and V_w peaked at 30.9 m/s, reflecting the influence of intense atmospheric conditions on wave development. Conversely, the lowest values were recorded in 2015, with H_s and V_w decreasing to 3.87 m and 1.7 m/s, respectively. This coherent behaviour across the time series supports the suitability of wind velocity as a predictor variable for extreme wave heights in the study region. For the same record, the air temperature remains relatively stable over the analysed period, oscillating within a narrow range of approximately 11–14 °C, and shows no significant correlation with H_s .

To extend this assessment to all annual maximum significant wave height series, a suite of 28 machine-learning (ML) models was implemented using wind speed, temperature, station code and the year as the input variables. A dataset of 130 values was employed for the analysis. The performance metrics for the validation (V) and testing (T) stages are summarised in Table 7.

Among the evaluated models, the Bagged Trees preset yielded the most consistent and robust performance, achieving RMSE values of 1.15 m (validation) and 0.78 m (testing), and coefficients of determination (R^2) of 0.75 and 0.88 for validation and testing, respectively. These results indicate a strong predictive capability and limited overfitting, confirming the suitability of ensemble-based tree methods for capturing the non-linear relationship between the response H_s and the predictors (V_w , T , station code, and the year of occurrence).

Comparable performance was achieved with Boosted Trees, Fine Trees, and several linear models, all of which exhibited high R^2 values. The ML analysis confirms that several data-driven approaches can produce reliable estimates of annual maximum significant wave heights solely from the selected predictors.

Table 7. Performance metrics of the machine-learning models used to predict the annual maximum significant wave height.

Preset	RMSE (V)	R^2 (V)	RMSE (T)	R^2 (T)
Linear	1.22	0.72	0.85	0.86
Interactions Linear	1.24	0.71	0.93	0.83
Robust Linear	1.25	0.71	0.87	0.85
Stepwise Linear	1.20	0.73	0.96	0.82
Fine Tree	1.21	0.73	1.25	0.69
Medium Tree	1.31	0.68	1.06	0.78
Coarse Tree	1.99	0.25	1.35	0.64
Linear SVM	1.30	0.68	0.89	0.84
Quadratic SVM	1.33	0.66	0.94	0.83
Cubic SVM	1.61	0.51	0.94	0.83
Fine Gaussian SVM	2.13	0.14	2.00	0.21
Medium Gaussian SVM	1.32	0.67	1.08	0.77
Coarse Gaussian SVM	1.45	0.61	1.25	0.69
Efficient Linear Least Squares	2.30	0.00	2.25	0.00
Efficient Linear SVM	2.37	−0.06	2.33	−0.07
Boosted Trees	1.21	0.72	0.82	0.87
Bagged Trees	1.15	0.75	0.78	0.88
Squared Exponential GPR	1.27	0.69	0.89	0.84
Matern 5/2 GPR	1.28	0.69	0.89	0.84
Exponential GPR	1.27	0.70	0.88	0.85
Rational Quadratic GPR	1.27	0.69	0.89	0.84
Narrow NN	1.43	0.61	1.04	0.79
Medium NN	2.64	−0.31	1.71	0.42
Wide Neural Network	2.98	−0.67	2.93	−0.70
Bilayered NN	1.74	0.43	2.32	−0.06
Trilayered NN	3.23	−0.96	2.62	−0.35
SVM Kernel	1.47	0.59	1.17	0.73
Least Squares Regression Kernel	1.38	0.64	1.03	0.79

Note: V = validation stage, T = testing stage, and NN = Neural Network. The grey cells represent the best model.

Figure 6 illustrates the performance of the Bagged Trees model during the validation and testing stages by comparing predicted and observed significant wave height (H_s) values. Although some dispersion is observed for higher H_s values—particularly during the validation stage—the overall trend demonstrates that the Bagged Trees preset provides a suitable model for making predictions of extreme events of H_s .

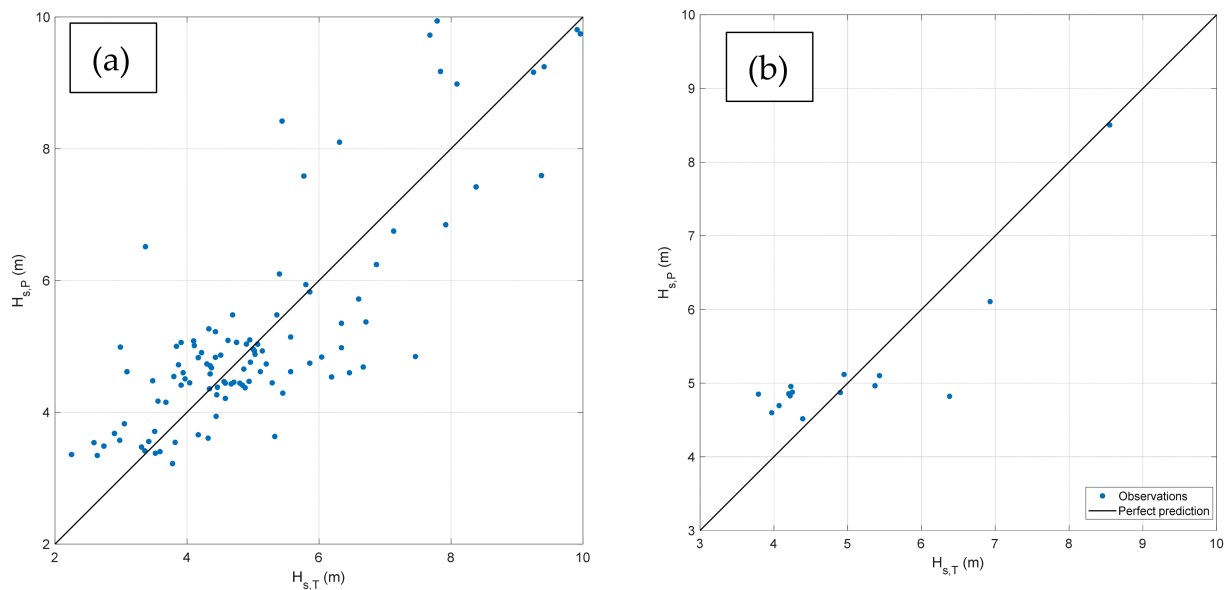


Figure 6. Comparison of H_s values between predicted and true values using the Bagged Trees algorithm: (a) validation stage; and (b) testing stage.

Figure 7 presents the SHAP values obtained with the selected ML preset, offering insight into the relative importance of the four predictors in predicting significant wave height. The SHAP analysis reveals a strong and consistent contribution of V_w to the model output, with higher wind velocities systematically associated with increased H_s values. This result not only confirms the physical consistency of the machine-learning model but also enhances its interpretability by explicitly quantifying the influence of wind forcing on wave extremes. Consequently, the SHAP results support the suitability of the Bagged Trees model. The remaining predictors (station code, T , and the year of occurrence) presented low SHAP values.

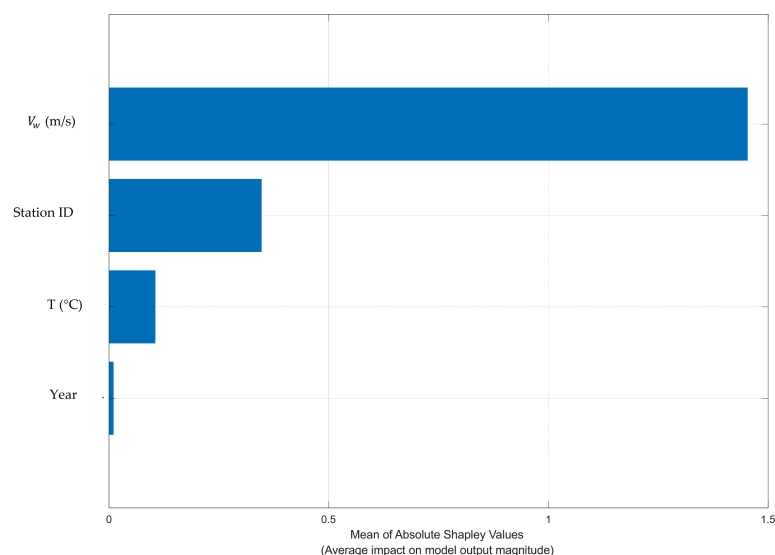


Figure 7. SHAP values for the adjustment of H_s .

To ensure that the generated significant wave height values preserve the statistical characteristics of the original extreme-value distribution, the absolute error was computed for the entire dataset at each station. The minimum and maximum values of the average absolute error and standard deviation are summarised in Table 8. The results confirm that

the errors are in an appropriate range, indicating that the generated values successfully reproduce the statistical properties of the original significant wave height series.

Table 8. Comparison of absolute error between ML model and recorded values.

Reference	Average Absolute Error (m)	Standard Deviation (m)
Minimum	0.01	0.01
Maximum	0.53	0.87

3.4. Frequency Analysis Using Traditional and ML Methods

This section presents the frequency analysis of significant wave height values performed using both the traditional approach and the methodology proposed in this study.

Once the optimal machine learning (ML) configuration was identified, namely the Bagged Trees model, annual maximum significant wave height (H_s) values were assessed using this approach. The trained model was exported to Simulink and employed to generate additional H_s data based on wind speed, air temperature, year, and station code. Figure 8 illustrates the Bagged Trees model implemented within the Simulink environment.

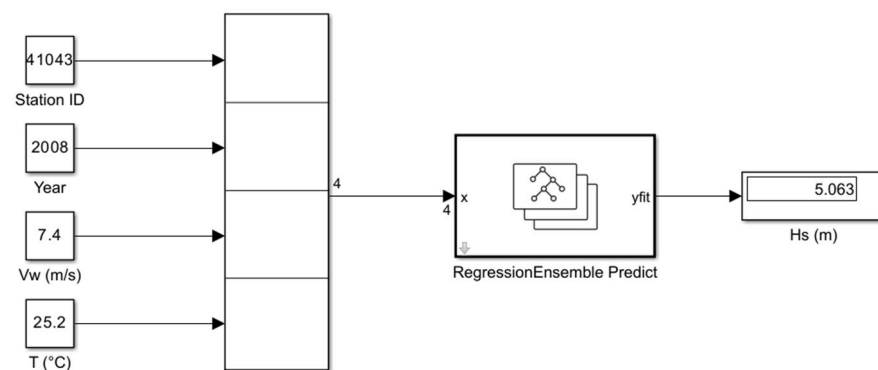


Figure 8. Bagged trees algorithm exported to Simulink.

Given the absence of statistically significant trends at all analysed stations, the extreme values of the H_s series were modelled under the assumption of stationarity. The Gumbel, Generalized Extreme Value (GEV), Pearson Type III, and Log-Pearson Type III distributions were considered, as they are commonly used to characterise the stochastic behaviour of extreme hydrometeorological variables and to estimate the probability of rare events exceeding the observed range.

To identify the probability distribution that best represents the H_s series recorded at the selected offshore buoys in the Caribbean Sea and adjacent regions, the Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) goodness-of-fit tests were applied. The results of these tests are summarised in Table 9. In grey, the best probability distributions are highlighted. The GEV distribution consistently exhibits the lowest Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) statistics when applied to observed significant wave height (H_s) records using the traditional methodology, indicating superior overall performance. Similarly, under the proposed machine learning (ML)-based methodology, the GEV distribution provides the best fit for stations 41040, 41043, 42058, and 42060. In contrast, the Pearson Type III distribution yields the best results for stations 41041 and 41044, while the Log-Pearson Type III distribution achieves the best performance for stations 41046 and 42059. In many cases, these distributions either failed to converge or produced non-applicable (NaN) results, particularly when fitted to the traditional datasets. A comparison between the traditional and ML-derived H_s series indicates that the goodness-of-fit statistics remain broadly consistent across both approaches.

Table 9. Goodness-of-fit results for stationary probability distributions.

Distribution	Test			
	Kolmogorov–Smirnov	Anderson–Darling	Kolmogorov–Smirnov	Anderson–Darling
	Traditional		ML	
Station code: 41040				
Generalized Extreme Value	0.095	0.221	0.115	0.17
Pearson III	NaN	NaN	0.213	
Log-Pearson III	NaN	NaN	NaN	NaN
Gumbel	0.178	0.798	0.264	1.998
Station code: 41041				
Generalized Extreme Value	0.112	0.261	0.294	0.982
Pearson III	NaN	NaN	0.222	
Log-Pearson III	0.116	0.255	NaN	NaN
Gumbel	0.125	0.28	0.373	1.661
Station code: 41043				
Generalized Extreme Value	0.144	0.217	0.129	0.301
Pearson III	NaN	NaN	NaN	NaN
Log-Pearson III	NaN	NaN	NaN	NaN
Gumbel	0.207	0.864	0.19	0.957
Station code: 41044				
Generalized Extreme Value	0.174	0.422	0.257	1.321
Pearson III	NaN	NaN	0.229	1.12
Log-Pearson III	NaN	NaN	NaN	NaN
Gumbel	0.182	0.481	0.265	1.369
Station code: 41046				
Generalized Extreme Value	0.098	0.244	0.200	0.739
Pearson III	NaN	NaN	NaN	NaN
Log-Pearson III	NaN	NaN	0.164	0.207
Gumbel	0.181	0.771	0.258	1.319
Station code: 42058				
Generalized Extreme Value	0.145	0.356	0.146	0.374
Pearson III	NaN	NaN	0.281	
Log-Pearson III	0.183	0.184	0.227	
Gumbel	0.225	1.164	0.319	2.097
Station code: 42059				
Generalized Extreme Value	0.089	0.174	0.138	0.455
Pearson III	0.134	0.069	0.213	1.228
Log-Pearson III	NaN	NaN	0.130	0.347
Gumbel	0.147	0.429	0.140	0.386
Station code: 42060				
Generalized Extreme Value	0.178	1.22	0.192	0.391
Pearson III	NaN	NaN	NaN	NaN
Log-Pearson III	0.193	0.296	0.240	0.470
Gumbel	0.252	1.174	0.289	1.591

Note: The grey cells represent the best model.

Considering the best-fitting probability distribution identified in Table 8, the frequency analysis was conducted using HEC-SPP version 2.3.1. Figure 9 compares the frequency analysis of annual maximum significant wave height (H_s) at station 41041 obtained using (a) the traditional methodology based on observed data (Figure 9a,b) and (b) the proposed machine learning (ML)-based approach (Figure 9b). For a 2-year return period, H_s values of 4.80 m and 4.87 m are obtained using the traditional and ML-based approaches, respectively. As the return period increases, the ML-based methodology tends to yield higher extreme H_s estimates at this station, particularly for longer return periods.

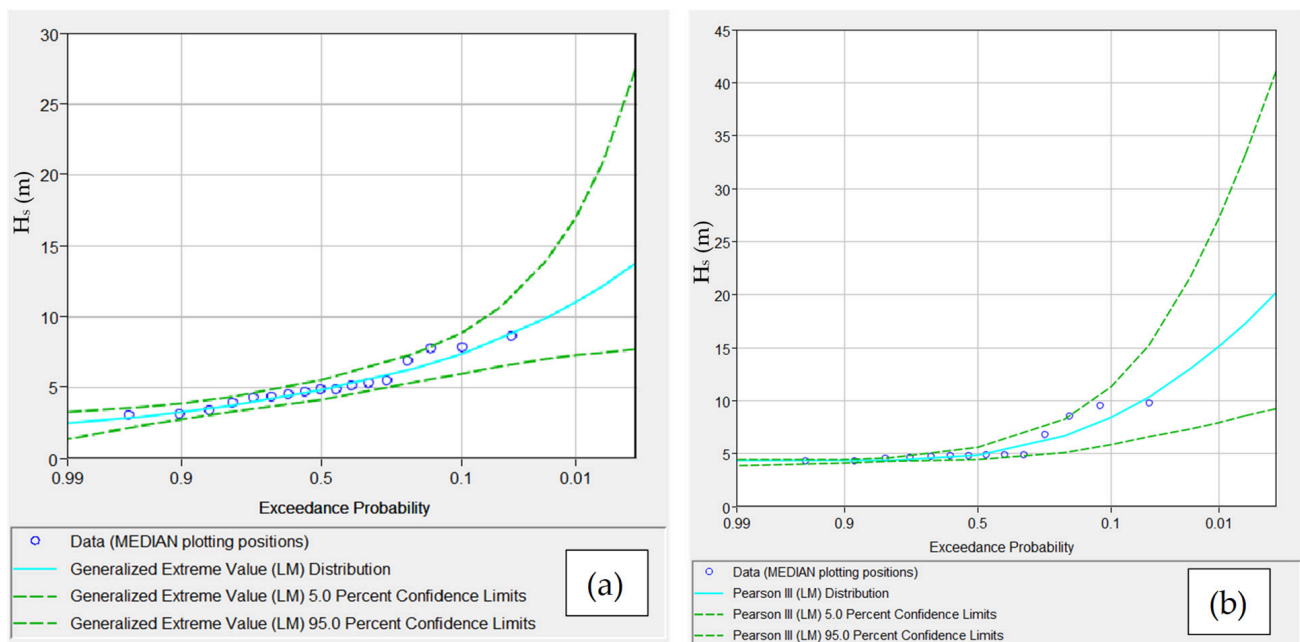


Figure 9. Frequency analysis for the station 41041 using: (a) the traditional methodology; and (b) the ML approach.

Figure 10 compares the significant wave height (H_s) values obtained using the traditional frequency analysis and the proposed machine learning (ML)-based approach for different return periods. The black line represents perfect agreement. Figure 10a,b, corresponding to return periods of 2 and 5 years, show excellent agreement between the two H_s estimates, indicating that the ML-based methodology accurately reproduces the traditional results for frequent events. For intermediate return periods of 10 and 20 years (Figure 10c,d), the agreement remains strong, although a moderate dispersion between the two approaches becomes apparent. For longer return periods of 50 and 100 years (Figure 10e,f), the differences between the methodologies are more pronounced.

These results are particularly relevant because the ML-based methodology can be reliably applied for low and intermediate return periods using only wind speed, air temperature, year, and station code as input variables. This capability is of paramount importance in situations where measured H_s data are unavailable or incomplete, as it enables the estimation of extreme wave conditions while preserving physical consistency with atmospheric forcing.

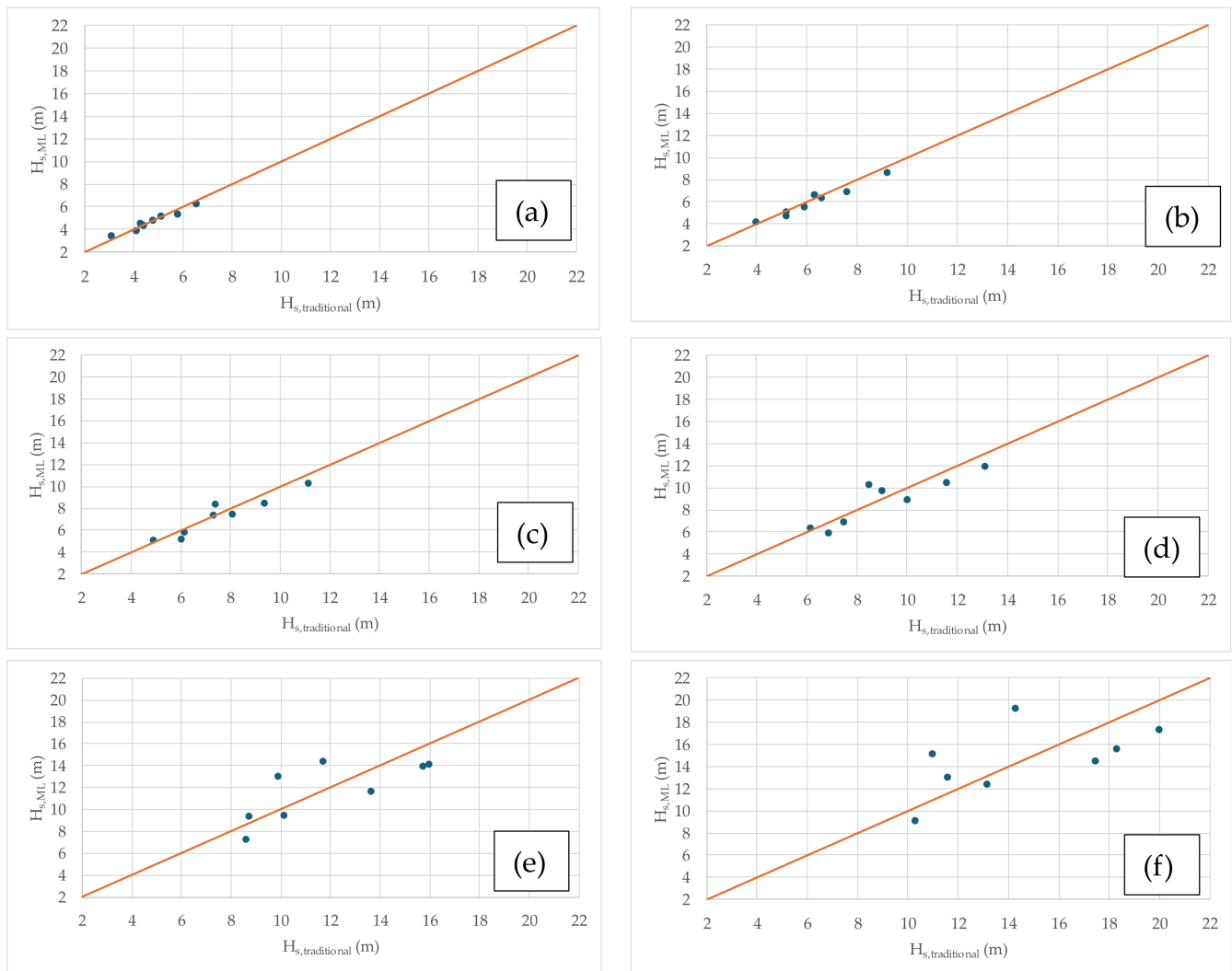


Figure 10. Comparison of H_s values between traditional and ML approaches for a return period of: (a) 2 years; (b) 5 years; (c) 10 years; (d) 20 years; (e) 50 years; and (f) 100 years.

Finally, to evaluate the level of agreement achieved using the machine learning (ML)-based approach, the coefficient of determination (R^2) was computed, as shown in Figure 11. Consistent with the results presented in Figure 10, very high R^2 values are obtained for low return periods (2–5 years), with $R^2 \approx 0.98$, indicating excellent agreement between significant wave heights (H_s) estimated using the traditional and ML-based methodologies. For intermediate return periods of 10 and 20 years, the R^2 values decrease to 0.95 and 0.89, respectively, but still reflect a good level of fit.

In contrast, for longer return periods of 50 and 100 years, the differences between the two approaches become more pronounced, with R^2 values of 0.74 and 0.58, respectively. The observed reduction in R^2 with increasing return period is expected, as ML models are primarily trained on observed data and become increasingly sensitive when extrapolated to rare extreme events.

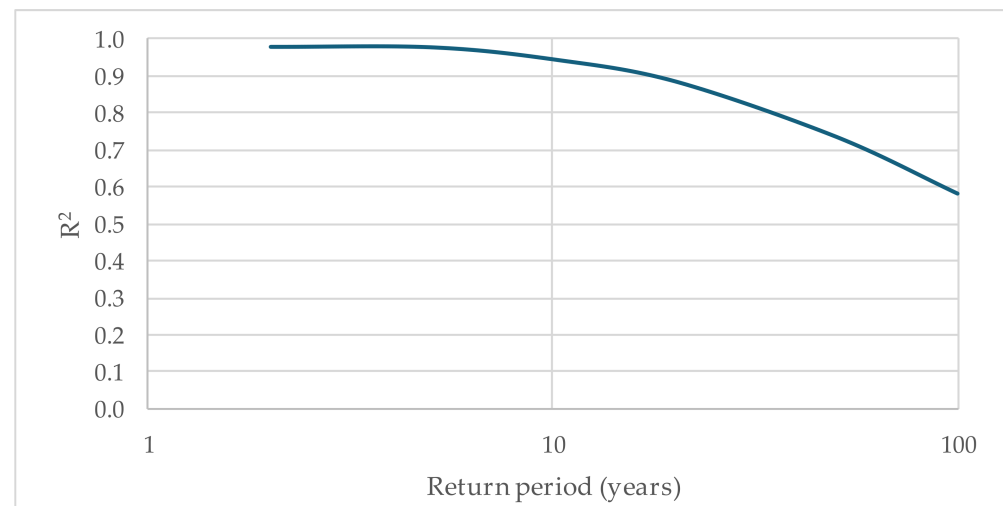


Figure 11. Computation of R^2 between the traditional ML results for several return periods.

3.5. Discussion

The methodology developed in this research is of significant importance, as it allows the estimation of maximum significant wave height for a given year based on a previously developed surrogate model. By considering the spatial location of a site and time-varying wind speed measurements, the proposed framework enables forecasting analyses relevant to coastal engineering applications.

The machine learning models were trained, validated, and tested across a representative range of observed conditions. In this context, the uncertainty associated with the relatively limited length of the dataset for each analysed station (approximately 18 years of observations) may be reduced in future applications by incorporating stations with longer and more complete records. Under these conditions, the estimation of extreme events for different return periods becomes comparable with traditional statistical approaches.

One of the main advantages of the proposed methodology is its capability to serve as a practical and rapid decision-support tool, reducing the computational burden associated with complex hydrodynamic models. This feature makes it particularly suitable for operational applications where fast assessments are required. The obtained information can be applied in various fields, such as navigation safety, coastal hazard assessment, and the design of coastal infrastructure.

Although the extreme significant wave height series analysed in this study exhibited stationary behavior, it remains essential for engineers and designers to also consider non-stationary trends in order to account for the potential impacts of climate change in other regions of the world. One of the limitations of the proposed methodology is that predictions associated with very rare events should be interpreted cautiously when used for the engineering evaluation and design of coastal engineering projects.

It is important to emphasise the computational efficiency of the selected machine learning model (Bagged Tree), which required only 7.6 s for prediction, thereby demonstrating its suitability as a rapid and practical tool for predictive applications. Figure 12 shows the prediction time for all ML models employed in this research.

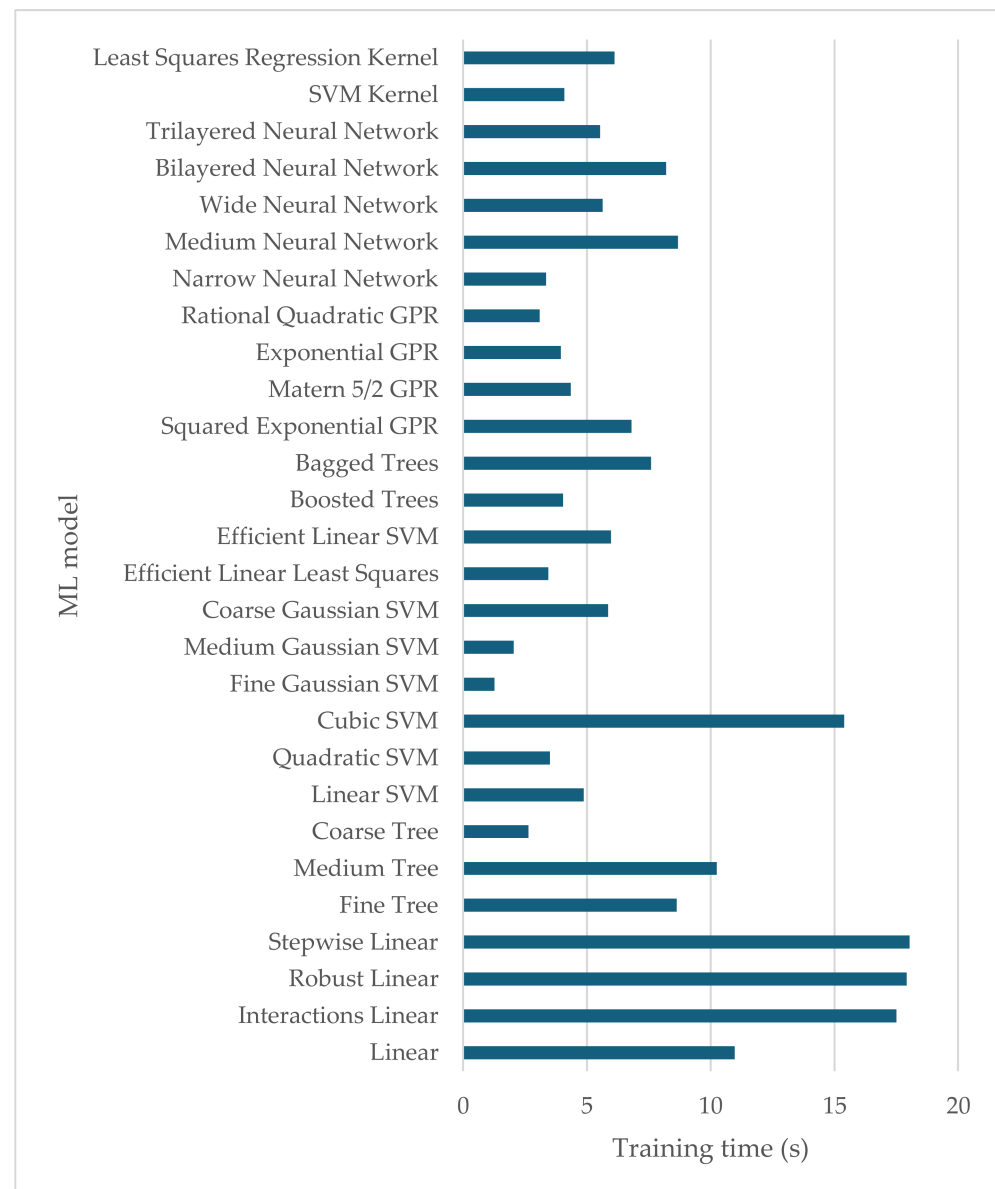


Figure 12. Training times of ML models.

4. Conclusions

This research proposes a comprehensive and evidence-based methodological framework for the estimation of extreme significant wave heights associated with different return periods using machine learning presets, with wind speed and air temperature records employed as predictor variables. The methodology incorporates a preliminary diagnostic analysis aimed at ensuring both statistical consistency and physical coherence of the input data.

The proposed framework can be applied in regions where significant wave height (H_s) records are incomplete or unavailable, provided that wind speed and air temperature data are available. The H_s values estimated for various return periods using the proposed approach were compared with those obtained through traditional extreme value analysis within a case study of the Antilles region in the Caribbean Sea.

The results indicate that the proposed methodology yields comparable estimates to the traditional approach for short return periods ranging from 2 to 5 years. For intermediate return periods between 10 and 20 years, the comparison also shows good agreement. Larger discrepancies were observed for longer return periods, particularly between 50

and 100 years. These results are consistent with the limited length of the available H_s records, which span approximately 18 years, and with the inherent characteristics of supervised machine learning methods, whose performance is constrained by the range and representativeness of the training data.

Author Contributions: Conceptualization, B.P.-O. and M.S.; methodology, B.P.-O., O.E.C.-H. and M.S.; validation, H.M.R.; formal analysis, B.P.-O., M.P.-S. and O.E.C.-H.; writing—original draft preparation, B.P.-O. and O.E.C.-H.; supervision, M.P.-S. and H.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by FCT, UID/6438/2025 CERIS, in the Hydraulic Laboratory, for experiments on pumped storage performance, and the project HY4RES (Hybrid Solutions for Renewable Energy Systems) EAPA_0001/2022 from the INTERREG ATLANTIC AREA PROGRAMME.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article.

Acknowledgments: The authors thank the Master's Program in Civil Engineering at the University of Cartagena.

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

Abbreviations

The following abbreviations are used in this manuscript:

a	Scale parameter
$f(x)$	Probability density function
H_s	Significant wave height
k	Corresponds to the shape parameter
N/n	Number of observations
P	Refers to a predicted value (-)
p	p -value
S	Mann–Kendall test statistic
T	Refers to a true value (-)
t	Time
u	Corresponds to the location or mode parameter
x	Observed value
β	Significance level
χ^2	Chi-Square test
γ	Coefficient for considering nonstationary values.
AIC	Akaike Information Criterion
EVA	Extreme Value Analysis
GEV	Generalized Extreme Value
GPR	Gaussian Process Regression
ML	Machine Learning
NEVA	Non-Stationary Extreme Value Analysis
NN	Neural Network
SNHT	Standard Normal Homogeneity Test
SSA	Singular Spectrum Analysis
SVM	Support Vector Machines
RMSE	Root Mean Squared Error
R^2	Coefficient of determination

References

1. Lizarbe-Palacios, M.; Aponte, H.; Botero, C.M. Multitemporal Scenic Evaluation of Urban Coastal Sites: A Peruvian Case Study. *Water* **2022**, *14*, 2336. [\[CrossRef\]](#)
2. Hassen Musa, M.K. The Effects of Climate Change on Geological Structures in coastline Regions. *Afr. J. Adv. Pure Appl. Sci.* **2025**, *4*, 123–129. [\[CrossRef\]](#)
3. Sealey, K.S. Assessing coastal vulnerability and climate resilience for Caribbean small island states using ecological principles. *Int. J. Disaster Risk Reduct.* **2024**, *105*, 104410. [\[CrossRef\]](#)
4. Schneider, T.; Bischoff, T.; Haug, G. Migrations and dynamics of the intertropical convergence zone. *Nature* **2014**, *513*, 45–53. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Corbella, S.; Stretch, D.D. Predicting coastal erosion trends using non-stationary statistics and process-based models. *Coast. Eng.* **2012**, *70*, 40–49. [\[CrossRef\]](#)
6. Lu, B.; Wang, W.; Wu, C.; Liu, Y. Wave climate on the southwestern coast of Lake Michigan: Perspectives from wave directionality. *Ocean Eng.* **2026**, *343*, 123306. [\[CrossRef\]](#)
7. Zhang, L.; Hua, L. Major Issues in High-Frequency Financial Data Analysis: A Survey of Solutions. *Mathematics* **2025**, *13*, 347. [\[CrossRef\]](#)
8. Strong, R.; Borgstroem, O.; Nathan, R.; Wasko, C.; O'Shea, D. Global Applicability of the Kappa Distribution for Rainfall Frequency Analysis. *Water Resour. Res.* **2025**, *61*, e2024WR039035. [\[CrossRef\]](#)
9. Domínguez, E.A.J. Identification and Separation of Environmental Events in Extreme Value Analysis for Return Period Estimation. Ph.D. Thesis, Escuela Politécnica Nacional, Quito, Ecuador, 2017.
10. Machado, S.; Zhu, X. Statistical and Machine Learning Approaches for Electrical Energy Forecasting. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2025**, *15*, e70033. [\[CrossRef\]](#)
11. Huang, W.; Zhu, X.; Jin, Y.; Shen, X. Nonstationary modelling of significant wave height using time series decomposition method. *Ocean Eng.* **2024**, *310*, 118731. [\[CrossRef\]](#)
12. Barlow, A.; Mackay, E.; Eastoe, E.; Jonathan, P. A penalised piecewise-linear model for non-stationary extreme value analysis of peaks over threshold. *Ocean Eng.* **2023**, *267*, 113265. [\[CrossRef\]](#)
13. Hiles, C.E.; Robertson, B.; Buckham, B.J. Extreme wave statistical methods and implications for coastal analyses. *Estuar. Coast. Shelf Sci.* **2019**, *223*, 50–60. [\[CrossRef\]](#)
14. Vanem, E. Uncertainties in extreme value modelling of wave data in a climate change perspective. *J. Ocean Eng. Mar. Energy* **2015**, *1*, 339–359. [\[CrossRef\]](#)
15. Galiatsatou, P.; Makris, C.; Krestenitis, Y.; Prinos, P. Nonstationary Extreme Value Analysis of Nearshore Sea-State Parameters under the Effects of Climate Change: Application to the Greek Coastal Zone and Port Structures. *J. Mar. Sci. Eng.* **2021**, *9*, 817. [\[CrossRef\]](#)
16. Vanem, E. A regional extreme value analysis of ocean waves in a changing climate. *Ocean Eng.* **2017**, *144*, 277–295. [\[CrossRef\]](#)
17. Pool, M.; Post, V.E.A.; Simmons, C.T. Effects of tidal fluctuations and spatial heterogeneity on mixing and spreading in spatially heterogeneous coastal aquifers. *Water Resour. Res.* **2015**, *51*, 1570–1585. [\[CrossRef\]](#)
18. McElman, S.; Verma, A.S.; Goupee, A. Quantifying tropical-cyclone-generated waves in extreme-value-derived design for offshore wind. *Wind Energy Sci.* **2025**, *10*, 1529–1550. [\[CrossRef\]](#)
19. Wang, Y.; Imai, K.; Miyashita, T.; Ariyoshi, K.; Takahashi, N.; Satake, K. Coastal tsunami prediction in Tohoku region, Japan, based on S-net observations using artificial neural network. *Earth Planets Space* **2023**, *75*, 154. [\[CrossRef\]](#)
20. Perricone, V.; Mutalipassi, M.; Mele, A.; Buono, M.; Vicinanza, D.; Contestabile, P. Nature-based and bioinspired solutions for coastal protection: An overview among key ecosystems and a promising pathway for new functional and sustainable designs. *ICES J. Mar. Sci.* **2023**, *80*, 1218–1239. [\[CrossRef\]](#)
21. Lorente, P.; Sotillo, M.G.; Aouf, L.; Amo-Baladrón, A.; Barrera, E.; Dalphinnet, A.; Toledano, C.; Rainaud, R.; De Alfonso, M.; Piedracoba, S.; et al. Extreme Wave Height Events in NW Spain: A Combined Multi-Sensor and Model Approach. *Remote Sens.* **2017**, *10*, 1. [\[CrossRef\]](#)
22. Caruso, M.F.; Marani, M. Extreme-coastal-water-level estimation and projection: A comparison of statistical methods. *Nat. Hazards Earth Syst. Sci.* **2022**, *22*, 1109–1128. [\[CrossRef\]](#)
23. De Leo, F.; Besio, G.; Briganti, R.; Vanem, E. Non-stationary extreme value analysis of sea states based on linear trends. Analysis of annual maxima series of significant wave height and peak period in the Mediterranean Sea. *Coast. Eng.* **2021**, *167*, 103896. [\[CrossRef\]](#)
24. Onen, F.; Bagatur, T. Prediction of Flood Frequency Factor for Gumbel Distribution Using Regression and GEP Model. *Arab J. Sci. Eng.* **2017**, *42*, 3895–3906. [\[CrossRef\]](#)
25. Kyojo, E.A.; Mirau, S.S.; Osima, S.E.; Masanja, V.G. Frequentist and Bayesian Approaches in Modeling and Prediction of Extreme Rainfall Series: A Case Study from Southern Highlands Region of Tanzania. *Adv. Meteorol.* **2024**, *2024*, 8533930. [\[CrossRef\]](#)

26. Zeinaldine, R.; Dahech, S. Climate Warming in the Eastern Mediterranean: A Comparative Analysis of Beirut and Zahlé (Lebanon, 1992–2024). *Urban Sci.* **2025**, *9*, 247. [[CrossRef](#)]
27. Toreti, A.; Kuglitsch, F.G.; Xoplaki, E.; Della-marta, P.M.; Aguilar, E.; Prohom, M. A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. *Int. J. Climatol.* **2010**, *31*, 630–632. [[CrossRef](#)]
28. Plant, H.; Conte, L.C.; Bayer, D.M.; Bayer, F.M. Bootstrap Pettitt test for detecting change points in hydroclimatological data: Case study of Itaipu. *Hydrol. Sci. J.* **2019**, *64*, 1312–1326. [[CrossRef](#)]
29. Bartels, R. The rank version of von Neumann's ratio test for Randomness. *J. Am. Stat. Assoc.* **1982**, *77*, 40–46.
30. Miškuf, M.; Michalik, P.; Zolotová, I. Data mining in cloud usage data with Matlab's statistics and machine learning toolbox. In Proceedings of the 15th International Symposium on Applied Machine Intelligence and Informatics (SAMII), Herl'any, Slovakia, 26–28 January 2017; pp. 377–382. [[CrossRef](#)]
31. Karuppusamy, B.; Kumar, D.; Lalmalsawma, P. Effect of climate change and deforestation on vector borne diseases in the North-Eastern Indian State of Mizoram bordering Myanmar. *J. Clim. Chang. Health* **2021**, *2*, 100015. [[CrossRef](#)]
32. Hamed, K.H.; Rao, A.R. A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol.* **1998**, *204*, 182–196. [[CrossRef](#)]
33. Chen, X.; Shao, Q.; Xu, C.-Y.; Zhang, J.; Zhang, L.; Ye, C. Comparative Study on the Selection Criteria for Fitting Flood Frequency Distribution Models with Emphasis on Upper-Tail Behavior. *Water* **2017**, *9*, 320. [[CrossRef](#)]
34. Obeysekera, J.; Asce, M.; Salas, J.D. Frequency of Recurrent Extremes under Nonstationarity. *J. Hydrol. Eng.* **2016**, *21*, 04016005. [[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.