Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p26065-cd

Bivariate copula modelling of significant wave height temporal dependence: a comparison study between buoy and reanalysis data

Daniel Guerra-Medina

Department of Physics, University of Las Palmas de Gran Canaria/i-UNAT FIMATA, Spain.
E-mail: daniel.guerra@ulpgc.es

Germán Rodríguez

Department of Physics, University of Las Palmas de Gran Canaria/i-UNAT FIMATA, Spain.
E-mail: german.rodriguez@ulpgc.es

Patricia Mares-Nasarre

Hydraulic Structures and Flood Risk, Delft University of Technology, Delft, the Netherlands.
E-mail: p.maresnasarre@tudelft.nl

A comprehensive understanding of the wave climate at a site of interest is essential for offshore and coastal engineering applications such as flood risk analysis, structural reliability assessment, or the design of coastal interventions. Specifically, the significant wave height, H_s, plays a central role in these designs and assessments, as it characterizes the severity of each sea state and, therefore, governs the loads on coastal and offshore structures, controls key coastal morphodynamic phenomena, among other processes.
To conduct these designs and assessments, practitioners need reliable wave data for statistical characterization. Several sources are available, such as buoys and reanalysis data. Buoy records are the ideal option but they are often spatially and temporally scarce. Therefore, reanalysis data has become more popular in recent decades, providing more continuous and longer time series. However, it is necessary to understand to what extent is reanalysis data representing reality.
This study, conducted in the Canary Islands, Spain, focuses on comparing the temporal evolution of wave conditions derived from reanalysis data and from a spatially overlapping buoy in probabilistic terms. Each database is clustered through the K-means++ algorithm to identify wave conditions by using the main metocean variables. In order to estimate the optimal number of clusters, the Elbow method is used. It yields a value of $k = 5$ for both datasets. Afterwards, an autocorrelation analysis is performed for each database and clusters; in general, higher correlations are observed for the reanalysis time series.
Bivariate copulas are used to model the temporal dependence of the significant wave height in each cluster, i.e., the joint distribution of $H_s(t)$ and $H_s(t+1)$. Both, the fitted copulas and empirical marginals are compared for each cluster and database observing significant differences. Overall, relevant differences are observed between reanalysis and buoy data from a probabilistic perspective that might have consequences in design.

Keywords: Bivariate copulas, temporal dependence, significant wave height, dataset validation, stochastic modelling, wave climate, probabilistic models.

1. Introduction

Offshore and coastal areas have high strategic value from both a social and an economic perspective (e.g., Guerra-Medina and Rodríguez, 2021). In order to plan, design, and assess the effectiveness and resilience of offshore and coastal infrastructures, it is fundamental to understand the wave conditions in the area of interest. The significant wave height, H_s is commonly employed by practitioners and researchers to describe the long-term wave conditions in a given location. H_s characterizes the severity of individual sea states and governs loads on coastal and offshore structures, and controls main coastal morphodynamic processes, among others (e.g., Mares-Nasarre et al., 2022; Branch et al., 2026; Mares-Nasarre and van Gent, 2020). Therefore, reliable oceanographic data is required to properly characterize the loads to plan and make well-informed decisions concerning coastal and offshore interventions.

In-situ buoy records are generally regarded as the most accurate source of wave information; however, they are spatially and temporally scarce, and commonly, also intermittent due to maintenance or operational problems. To overcome these limitations, reanalysis datasets derived from numerical wave models have become increasingly used over recent decades, offering multi-decadal temporal coverage and extensive spatial domains that support regional to global assessments (e.g, Gómez Lahoz and Carretero Albiach, 2005; C3S, 2018). Nevertheless, it is essential to evaluate the degree to which reanalysis data is representing real observations.

Local wave climate is often the result of various meteorological systems. Depending of the location, a site may experience local wind-waves or swell waves, both separately in time, as well as both simultaneously. Consequently, most methods in the scientific literature rely on the use of clustering to group similar wave conditions (e.g. Camus et al., 2019; Mares-Nasarre et al., 2023). This is, sea states generated by the same meteorological system and, thus, belonging to the same statistical population are identified to simplify the study of the local wave climate.

Copulas are a powerful statistical model that allow to model the statistical dependence between variables separately from their marginal distributions (Sklar, 1959). Their use has become increasingly popular to characterize the probabilistic relationship between oceanographic variables (e.g., Jaeger and Morales-Nápoles, 2017; Lucio et al., 2020). However, most of the existing studies assume that the dependence between the variables of interest is properly described using Gaussian copulas (e.g., Camus et al., 2019; Lucio et al., 2020). Nevertheless, other authors such as (Jaeger and Morales-Nápoles, 2017) or (Antão and Guedes Soares, 2014) have shown how the joint behavior of oceanographic variables may not be properly captured using Gaussian copulas, especially when dealing with extremes.

In this study, reanalysis data from the Spanish Ports Authority –called SIMAR– is compared to a spatially overlapping buoy record in probabilistic terms focusing on the temporal evolution of H_s. To do so, a methodology similar to the state of the art is applied to wind and wave data (see Camus et al., 2019; Lucio et al., 2020). First, the clustering algorithm K-means++ (Arthur and Vassilvitskii, 2006) is applied to each dataset to identify the different wave conditions (see Section 2). Once defined, clusters from each database are compared to each other through the empirical marginal distribution of H_s and the dependence of successive H_s using bivariate copulas. In summary, this manuscript presents a statistical comparison between reanalysis and buoy datasets with a special focus on the temporal behavior of H_s.

2. Datasets

This study employs overlapping datasets of buoy and reanalysis (SIMAR), at coordinates $28.20°$ N, $15.80°$ E. It is located in the archipelago of Canary Islands (Spain), specifically, to the northwest of Gran Canaria island. These datasets belong to the Spanish Port Authority.

Both datasets consist of hourly data. The SIMAR record begins in 1958, while the buoy was installed in 2003. To ensure a consistent comparison between the two, only the overlapping period (2003–2024) is considered. The selected metocean variables are:

- Significant wave height, H_s (m), both datasets provide the significant wave height estimated from the wave spectra.
- Peak wave period, T_p (s), the wave period associated with the most energetic component in the wave spectra.
- Mean wave direction, θ (°). Angles are ascendant in a clockwise direction with $0°$ and $90°$ being the N and E respectively.
- Wind speed, ws (m/s), defined as the magnitude of northward and eastward components of neutral wind at 3 and 10 meters above the sea surface for buoy and reanalysis data respectively.
- Mean wind direction, θ_{ws} (°). Angles are ascendant in a clockwise direction with $0°$ and $90°$ being the N and E respectively.

More detailed definitions and datasets information can be found in *Puertos del Estado*.

3. Methodology

The methodology is divided in the following subsections. First, the data is preprocessed to ensure the logic of the wave and wind angles, standardized and clustered. Afterwards, each dataset and cluster are compared in terms of the empirical marginals of H_s and dependence between consecutive H_s, as well as the autocorrelation of the time series of H_s. The dependence between H_s at time t and $t+1$ is modeled here using bivariate copulas. The best-fitting copula models for each cluster are finally validated using in-sample validation.

3.1. *Preprocessing and clustering*

3.1.1. *Data preprocessing*

The variables used here are H_s, T_p, ws, θ, and θ_{ws}, as explained in Section 2. These metocean variables are used only in the clustering step and are preprocessed and standardized for it. First, the directional variables, θ and θ_{ws}, are transformed. The circular means are used as reference and the values of the variables are wrapped around them to express the resulting angles within $[-180, 180]$. Mathematically,

$$\theta(i) = \text{mod}(\theta(i) - \overline{\theta} + 180°, 360°) - 180° \quad (1)$$

where $\overline{\theta}$ is the circular mean and mod stands for modulus, basic modular arithmetic notation which specifies a value over which angles must wrapped around (Proakis and Manolakis, 1988).

3.1.2. *Clustering*

Each database is clustered based on the metocean variables defined in Section 2, so that sea states generated by the same meteorological conditions are grouped together. For clustering, the K-means++ algorithm, an improved version of the K-means, is applied to each database.

The K-means is an unsupervised classification algorithm that minimizes the intra-cluster variation. This is, the algorithms aims to minimize the sum of squared Euclidean distances between the centroid of each cluster and the observations belonging to it. K-means algorithm requires the initialization of the centroids, typically done randomly. Differently to K-means, the K-means++ selects the initial centroids more effectively by randomly choosing the centroids well spread throughout the data, which translates into a faster convergence (Arthur and Vassilvitskii, 2006).

This algorithm requires the user to define the number of clusters, k. Then, an appropriate value of the hyperparameter k needs to be carefully selected. Here, $k = 2, 3, ..., 15$ are considered, and the Elbow method is used to estimate the optimal value. This method is based on the inertia computed for each clustering attempt, that is, for each k. Inertia is defined as the sum of squared distances of each observation to its closest centroid. Therefore, a lower value is desired. The goal of the Elbow method is to identify the inflection point in the curve of decreasing inertia values as k increases. This is, the point beyond which increasing k only yields a slight improvement of the inertia (Thorndike, 1953).

3.2. *Datasets comparison*

Once the clusters are identified, the study focuses on comparing the temporal dependence of H_s between datasets. Different comparisons are conducted based on autocorrelation, empirical marginal distributions, and bivariate copulas.

3.2.1. *Autocorrelation*

The temporal dependence of H_s is of special importance in many practical applications such as those related to sea state persistence (Goda, 2010). Therefore, reanalysis data, ideally, should reproduce the memory of the process given by the buoy.

In order to study the autocorrelation of H_s, the Spearman's rank correlation coefficient (Spearman, 2010) is estimated for lags between 1 and 48 hours. Similarly, once clustering is performed, the autocorrelation is estimated for each cluster and dataset and compared among them.

The Spearman's rank correlation coefficient, r, is defined as

$$r = \frac{Cov[R(X), R(Y)]}{\sigma_{R(X)}\sigma_{R(Y)}} \quad (2)$$

where $Cov[R(X), R(Y)]$, $\sigma_{R(X)}$ and $\sigma_{R(Y)}$ are the covariance and standard deviations of the ranked variables respectively. $r \in [-1, 1]$ where $r = -1$ and $r = 1$ mean a perfect negative and positive monotonic relationship between variables, respectively. $r = 0$ means no correlation.

3.2.2. *Empirical Marginals*

The empirical margins of the different clusters from each dataset are defined and compared visually through exceedance probability plots in logarithmic scale, as well as statistically using Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1948). Here, the two-sample Kolmogorov-Smirnov test is performed.

3.2.3. *Bivariate Copulas*

Bivariate copulas, or just copulas, are joint distributions whose marginal distributions are uniform in $[0, 1]$. (Sklar, 1959) states that any multivariate joint distribution can be quantified using two components: a set of univariate marginal distributions and a copula that captures the dependence between them.

In the bivariate case, the joint distribution function $H_{XY}(x, y)$ can be expressed as:

$$H_{XY}(x, y) = C\big(F_X(x), G_Y(y)\big), \qquad (3)$$

where $F_X(x)$ and $G_Y(y)$ are the marginal distribution functions of X and Y, respectively, and C is a copula defined in the unit square $I^2 = [0, 1] \times [0, 1]$. Eq. (3) holds for all $(x, y) \in \mathbb{R}^2$.

Bivariate copulas are used in this study to model the probabilistic dependence between H_s at time t, $H_s(t)$, and at time $t + 1$, $H_s(t + 1)$. First, for each cluster in each dataset, the empirical copula is visually compared to those of the other dataset. The empirical copula, C_n is defined as

$$C_n(u, v) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1} \left(\frac{R_k}{n+1} \leq u, \frac{S_k}{n+1} \leq v \right) \qquad (4)$$

where $\{(R_k, S_k)\}$ are the ranks associated to a random sample $\{(X_k, Y_k)\}$, $k = 1, ..., n$, $u, v \in \mathbb{I}$ and $\mathbf{1}$ is an indicator function (Salvadori et al., 2007).

Cramer-von-Misses statistic, CvM and its corresponding p-value (Genest and Favre, 2007) are used to compare the copulas for a given cluster and dataset to the copulas of all the clusters of the other dataset. In order to perform such comparison using CvM, the copulas need to be evaluated in the same joint probabilities. Therefore, here, parametric copulas are fitted to each cluster.

The best-fitting copula for each cluster in each dataset is first selected using the Python library Pyvinecopulib by Nagler and Vatter (2025), which fits 12 parametric copulas to the observations and chooses the best-fitting one based on the Akaike Information Criterion (AIC) (Akaike, 1998). The goodness of fit of this selected parametric copula is compared to the one provided by Gaussian, Gumbel, Frank, and Clayton copulas using CvM and its associated p-value, as well as the semi-correlation values (Joe, 2014).

Semi-correlations quantify the asymmetries in the dependence structure and assess whether those are captured by the fitted models. To this end, semi-correlations compute Pearson's correlation coefficient (Pearson, 1895) of the data transformed into standard normal space and divided into four quadrants. They are computed for both the data and simulations of each adjusted copula and compared.

3.3. *Validation*

In addition to the previously goodness of fit analysis, the performance of the best fitting copulas are assessed through in-sample validation. To do so, time series of consecutive 51 hours are extracted from each cluster. For each time series, $H_s(t)$, with $t = 0, 1, ..., 49$ $[h]$, is used as input to conditionalize the fitted copula and estimate the conditional distribution of $H_s(t + 1)$ using 500 samples. From the resulting simulated distributions, the 5^{th}, 50^{th}, and 95^{th} percentiles are computed. Then, the original time series (from $t = 1, 2, ..., 50$) are compared against the simulated percentiles, which are presented as error bars to visually assess the predictive performance of the copula models.

4. Results and discussion

In this section, the results derived from the methodology in Section 3 are described and discussed.

4.1. *Clustering results*

According to the Elbow method, an optimal number of clusters k=5 is obtained for each dataset. The clusters are named Bi for the buoy and Si for SIMAR with $i = 0, 1, ..., 4$.

Figures 1 and 2 show an overview of how the different observations are clustered for the studied variables. The diagonal panels show the distribution of each variable per cluster, while the off-diagonal panels are scatter plots of the different pairs of variables.

Some similarities and differences are observed in the clustering results of the two datasets. For example, $B1$ and $S2$ present similar conditions: $H_s < 2\,m$, small values of T_p, moderate WS, and θ and θ_{ws} mainly from N to NE. This corresponds to wind-waves generated by the wind trades of this region. However, other clusters present notable differences. For instance, in Figure 2, cluster $S4$ corresponds to wave conditions comprising moderate H_s and WS, mid-values of T_p and with θ and θ_{ws} ranging from around 180° to 359° and from 0° to 20°. This cluster is the least populated and seems to represent any wave conditions during which the wind flows from the south west, channeled by the orography of Gran Canaria and Tenerife. However, in the buoy dataset, the wave conditions in $S4$ seem to be splitted into $B0$ and $B4$. In this case, H_s, T_p and θ_{ws} are similar for both clusters with differences in θ and WS. $B0$ comprises wave conditions coming mainly from NNW to NNE ($\theta = 320°$ to $30°$) and mostly $WS > 6\ m/s$. In contrast, for $B4$, wave conditions come mainly from WSW to WNW ($\theta = 230°$ to $310°$) and mostly $WS < 6\ m/s$. Despite these differences, clusters in both datasets can be interpreted as representative for different wave conditions given in the area.

4.2. *Autocorrelation*

Figure 3 shows the autocorrelation of the significant wave height for lags between 0 and 48 hours for both datasets before clustering. Both datasets exhibit a similar pattern, indicating a similarity in the temporal dependence structures.

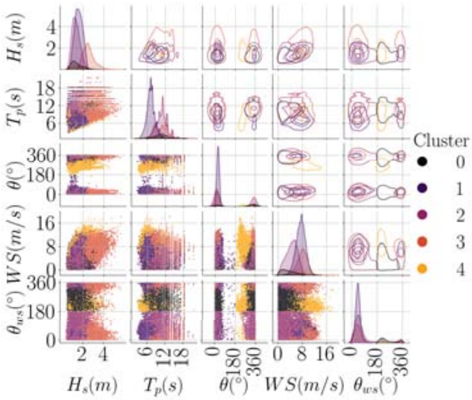

Fig. 1. Pairplot of the buoy metocean variables used in this study –H_s, T_p, θ, ws, and θ_{ws}– per cluster.

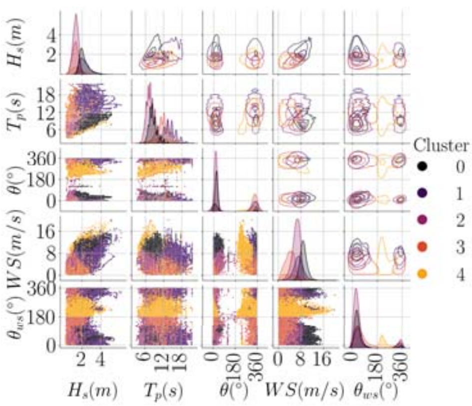

Fig. 2. Pairplot of the SIMAR metocean variables used in this study –H_s, T_p, θ, ws, and θ_{ws}– per cluster.

Autocorrelation decay is nearly identical, which might imply that the characteristic timescales over which the process losses memory are consistent between datasets. However, SIMAR dataset retains a slightly higher correlation ($\approx 4\%$) than the buoy dataset.

Although not presented in this document, the autocorrelations are also computed for the different clusters and compared between datasets. This is, each cluster from one dataset is compared against all the clusters from the other dataset. Results show that in 85% of the comparisons SIMAR retains more correlation than the buoy.

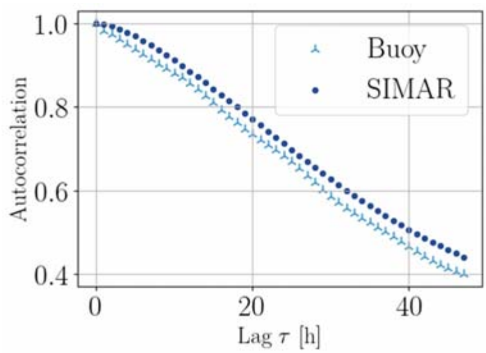

Fig. 3. Autocorrelation function of the significant wave height, H_s, computed for lags 0 to 48 hours.

4.3. *Empirical marginals*

Empirical margins are used to compare H_s of each cluster. Since the upper tail is of special interest, the survival function is presented in logarithmic scale, as displayed in Figure 4.

In Figure 4, the comparisons for the buoy cluster $B3$ are presented as an example. $B3$ is compared against all the SIMAR clusters, $S0$ to $S4$. As expected, for most cases, the exceedance probability curves are far from each other. For instance, while $B3$ consists of values of $H_s > 2\,m$, $S2$ and $S3$ are representative of milder wave conditions. $S0$ and $S1$ show the closest visual agreement, and $S4$ seems to have a similar behavior but shifted towards smaller values. However, regardless of the degree of agreement, the tails behave differently. Kolomogorov-Smirnov test was computed finding significant differences for all comparisons.

4.4. *Copula models validation*

Once the best fitting copulas for each cluster are selected, these are validated using an in-sample validation procedure as explained in Section 3.

Figure 5 shows an example for clusters $B4$ and $S4$. Note that each cluster is modeled separately and they do not correspond to the same time frame. They are plotted together for comparison of the predicted uncertainty between datasets.

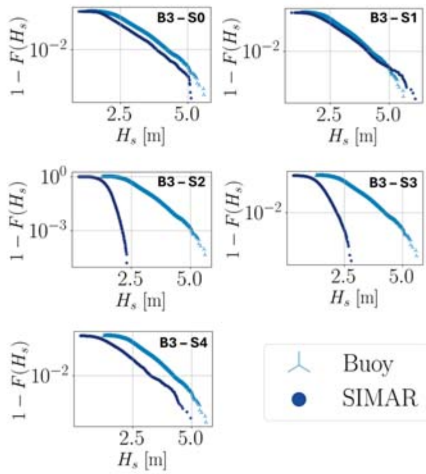

Fig. 4. Exceedance probabilities of the empirical marginals of H_s in logarithmic scale for buoy cluster 3, $B3$, and the different clusters from the SIMAR dataset.

Table 1. Copula distributions that best modeled the relationship between $H_s(t)$ and $H_s(t+1)$ for each cluster.

Cluster B/S	BUOY	SIMAR
0	Frank	Gumbel
1	Gumbel (rot=180)	Gumbel
2	Gaussian	Gaussian
3	Gaussian	Gaussian
4	Frank	Gaussian

As shown in Figure 5, SIMAR presents a smoother line with few oscillations, as well as narrower 90% intervals. The 5th and 95th are close to the median, with 90% interquantile ranges reaching a maximum of $30cm$. However, the time series from the buoy shows more oscillations. Consequently, the simulated buoy values show larger 90% interquantile ranges up to $96cm$. Despite the wider distribution simulated for values in $B4$, they tend to fall within the predicted 90% range. Similar patterns are observed for other clusters. Overall, SIMAR presents smoother time series due to a higher correlation which is translated to the bivariate copulas and the variability of the modeled values.

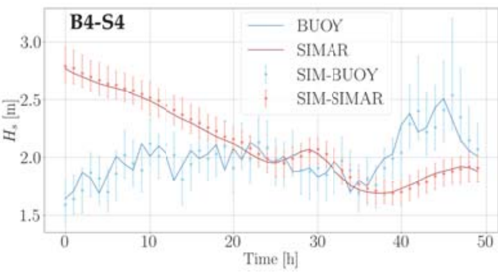

Fig. 5. Validation example of copula models for clusters $B4$, and $S4$. 5%, 50% and 95% percentiles of the inferred distribution of $H_s(t+1)$ are depicted.

5. Conclusions

In this study, a set of methods have been used for analysing and comparing the statistical structure of $H_s(t)$ and $H_s(t+1)$ for two different datasets, buoy and reanalysis data (SIMAR). Clustering and bivariate copulas were employed to identify climatic wave patterns and compare the dependence structure with views to further research towards the development of a wave generator. This study has been developed in Canary Islands (Spain), but the proposed methods can be applied to different locations and datasets.

It is concluded that there are significant statistical differences between datasets, with SIMAR presenting a higher correlation due to smoother time series with few oscillations. Therefore, SIMAR is not able to fully capture the natural oscillations of H_s over time which might lead to underestimations of extremes similar to other reanalysis datasets (e.g. Hersbach et al., 2018). Should extremes or those oscillations become relevant for the analyst, SIMAR data would not be the optimal dataset to use.

Regarding the dependence structure, clusters in the buoy dataset are mainly symmetric, with only one cluster presenting a notable lower tail dependence. However, two clusters in the SIMAR dataset present a notable upper tail dependence.

Finally, in-sample validation of the bivariate copulas fitted to the clusters is performed by estimating the distribution of $H_s(t+1)$ given $H_s(t)$; the methodology shows promising results as simulations is able to capture the oscillations observed in the original time series.

Acknowledgement

The authors acknowledge Puertos del Estado for providing the datasets employed. This research was partially supported by two ERASMUS scholarships: ERASMUS + Prácticas and ERASMUS + Fines Formativos. The first author acknowledges the financial support of the FPU programme (*Formación Profesorado Universitario*) funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (grant FPU2023-03776).

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pp. 199–213. Springer.

Antão, E. and C. Guedes Soares (2014). Approximation of bivariate probability density of individual wave steepness and height with copulas. *Coastal Engineering 89*, 45–52.

Arthur, D. and S. Vassilvitskii (2006). k-means++: The advantages of careful seeding. Technical report, Stanford.

Branch, R., J. R. McVey, F. Ticona Rollano, R. J. Cavagnaro, A. Turpin, E. Geon, B. Roberts, M. Steele, A. Nawaz, M. Wickett, and M. Wickett (2026). Wave energy converter buoy for arctic observations. *Renewable Energy 256*.

C3S (2018). ERA5 hourly data on single levels from 1940 to present.

Camus, P., A. Tomás, G. Díaz-Hernández, B. Rodríguez, C. Izaguirre, and I. Losada (2019). Probabilistic assessment of port operation downtimes under climate change. *Coastal Engineering 147*, 12–24.

Genest, C. and A.-C. Favre (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering 12*(4), 347–368.

Goda, Y. (2010). *Random seas and design of maritime structures*, Volume 33. World Scientific Publishing Company.

Guerra-Medina, D. and G. Rodríguez (2021). Spatiotemporal Variability of Extreme Wave Storms in a Beach Tourism Destination Area. *Geosciences 11*(6), 237.

Gómez Lahoz, M. and J. C. Carretero Albiach (2005). Wave forecasting at the Spanish coasts. *Journal of Atmospheric & Ocean Science 10*(4), 389–405.

Hersbach, H., B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. M. Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thépaut (2018). Era5 hourly data on single levels from 1959 to present. Technical report, Copernicus Climate Change Service (C3S), Climate Data Store (CDS). Accessed 1st Sept. 2022.

Jaeger, W. S. and O. Morales-Nápoles (2017). A vine-copula model for time series of significant wave heights and mean zero-crossing periods in the north sea. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3*, 04017014.

Joe, H. (2014). *Dependence Modeling with Copulas* (0 ed.). Chapman and Hall/CRC.

Kolmogorov, A. (1933). 19++. sulla determinazione empirica di una legge di distributione. *Giornale dell'Istituto Ntaliano degli Attuari 4*(8), 91.

Lucio, D., A. Tomás, J. Lara, P. Camus, and I. Losada (2020). Stochastic modeling of long-term wave climate based on weather patterns for coastal structures applications. *Coastal Engineering 161*, 103771.

Mares-Nasarre, P., J. Garcìa-Maribona, M. A. Mendoza-Lugo, and O. Morales-Nápoles (2023). A Copula-Based Bayesian Network to Model Wave Climate Multivariate Uncertainty in the Alboran Sea. In *Proceeding of the 33rd European Safety and Reliability Conference*, pp. 1053–1060. Research Publishing Services.

Mares-Nasarre, P., J. Molines, M. E. Gómez-Martín, and J. R. Medina (2022). Hydraulic stability of cube-armored mound breakwaters in depth-limited breaking wave conditions. *Ocean Engineering 259*, 111845.

Mares-Nasarre, P. and M. R. A. van Gent (2020). Oblique wave attack on rubble mound breakwater crest walls of finite length. *Water 12*(2).

Nagler, T. and T. Vatter (2025). pyvinecopulib.

Pearson, K. (1895). Vii. note on regression and inheritance in the case of two parents. *proceedings of the royal society of London 58*(347-352), 240–242.

Proakis, J. G. and D. G. Manolakis (1988). *Introduction to digital signal processing*. Prentice Hall Professional Technical Reference.

Salvadori, G., C. de michele, N. Kottegoda, and R. Rosso (2007). Extremes in nature: An approach using copulas. *Extremes in Nature: An Approach Using Copulas, By G. Salvadori, C. De Michele, N.T. Kottegoda, and R. Rosso. Berlin: Springer, 2007. 49*.

Sklar, M. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics 19*(2), 279–281.

Spearman, C. (2010). The proof and measurement of association between two things. *International Journal of Epidemiology 39*(5), 1137–1150.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika 18*(4), 267–276.