Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P091-cd

A copula-based Bayesian Network to model wave climate multivariate uncertainty in the Alboran sea

Patricia Mares-Nasarre

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

Julio García-Maribona

DHI A/S, Copenhagen, Denmark. E-mail: jugm@dhigroup.com

Miguel A. Mendoza-Lugo

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

Oswaldo Morales-Nápoles

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

An accurate estimation of wind and wave variables is key for coastal and offshore applications. Recently, copulas have gained popularity for modelling wind and waves multivariate dependence, since accounting for the hydrodynamic relationships between them is needed to ensure reliable estimations of the required design values. In this study, copula-based Bayesian networks (BNs) are explored as a tool to model extreme values of significant wave height (H_s), wave period, wave direction, wind speed and wind direction. The model is applied to a case study located in the Alboran sea, close to the Spanish coast, using ERA5 database. Extreme values of H_s are sampled using Yearly Maxima and concomitant values of the missing variables are used. K-means clustering algorithm is applied to separate the different wave components and a BN is built for each of them. The assumption of modelling the dependence between the variables using Gaussian copulas and the structure of BNs are supported with the d-calibratioson score. Fitted marginal distributions are introduced in the nodes of the BNs and their performance is assessed using in-sample data and the coefficient of determination. The BN models proposed present high performance with a low computational cost proving to be powerful tools for modelling the variables under investigation. Future research will include different locations and databases.

Keywords: waves, wind, stochastic process, copulas, k-means, Bayesian networks

1. Introduction

Wind and waves are the main loads to consider when assessing offshore energy production, planning coastal and offshore operations or designing coastal and offshore structures, among others. Wind speed directly affects the bollard pull on mooring lines and the energy production of offshore turbines (Kim et al., 2022). Wave height is linked to wave energy, making higher waves more harmful for structures (Mares-Nasarre et al., 2022). Additionally, longer wave periods lead to higher run-up and overtopping, and, thus, more intense erosion rates on the beaches and dunes which protect the shoreline (Di Luccio et al., 2018). Moreover, a physical limitation exists, called maximum steepness condition, which states that too high waves break for given wave periods. Wind and waves do not approach structures or shorelines perpendicularly, but with an angle. This obliquity can significantly reduce wave loads, wave overtopping and wave run-up when compared to perpendicular waves (Mares-Nasarre and van Gent, 2020).

Waves are generated by wind friction on the sea surface. Thus, waves and wind arise from the same meteorological system. The characteristics of the resulting waves (height, period and direction) are related to the wind speed and direction in the generation area (sea waves). In addition, waves may travel along large distances after being generated, and their characteristics become less correlated to the local wind conditions (swell waves). Consequently, neglecting the probabilistic dependence between these variables may lead to the underestimation of design loads. Therefore, the multivariate uncertainty (probabilistic dependence) between them needs to be accounted to ensure reliable estimations.

In the last years, climate change has increased the focus on the uncertainty and probabilistic dependence of extreme climatic events (e.g. Esteves et al., 2011; Pham et al., 2021). A popular approach for modelling the dependency of climatic variables is based on copulas, which isolate the one-dimensional marginals from the dependence structure of random variables. Camus et al. (2019) and Lucio et al. (2020) proposed the use of clustering algorithms to increase the homogeneity in each sample (cluster) and describe the dependence structure within each cluster using Gaussian copulas. Other examples of the use of copula-based models to describe wind and wave climate can be found in Lin-Ye et al. (2016) or Jaeger and Morales-Nápoles (2017).

Copula-based Bayesian Networks (BNs) are probabilistic graphical models which represent high-dimensional probability distribution functions. BNs have been successfully used in different fields of civil engineering, such as to predict extreme river discharges (e.g. Paprotny and Morales-Nápoles, 2017) or to model weight-in-motion data (e.g. Mendoza-Lugo et al., 2022). However, to the authors' knowledge, BNs have not been applied to describe wind and wave climatic variables. Previous works have proven the adequacy of copula-based models to characterize the dependence between some wave or wind variables. Nevertheless, since all these variables are relevant for design purposes, a model to describe the dependence of all these variables simultaneously is needed. Here, the potential of copula-based Bayesian Network models to characterize the multivariate uncertainty of the extreme observations of wave and wind variables is explored. To do so, a case study in the Alboran Sea is used, as presented in Section 2. In Section 3, the modelling approach is described. In Section 4, the model is built and assessed, and its results are presented. Finally, in Section 5, conclusions are drawn.

2. Case study

The present study proposes a copula-based BN to characterize the dependence between extreme observations of wave and wind variables. A location in the Alboran Sea (Spain) is used as a case study. The Alboran sea is located in the South of Spain, at the East of the Strait of Gibraltar. The location of the case study has coordinates 36.4°N, -3.5°E.

Hourly measurements of aggregate wave and wind variables from 1959 to 2021 (63 years) are available in ERA5 database (Hersbach et al., 2018). The following variables are selected:

- Significant wave height, $H_s = H_{1/3}$ (m), the average height of the highest third of sea and swell waves.
- Mean wave period, T_m (s), the average time for two consecutive wave crests to pass through a fixed point.
- Mean wave direction, θ_H (°). North and east correspond to θ_H=0°and θ_H=90°, respectively.
- Mean wind speed, W_s (m/s), the magnitude of northward and eastward components of neutral wind at 10 metres above the sea surface.
- Mean wind direction, θ_w (°). North and east correspond to θ_w=0°and θ_w=90°, respectively.

For further reference on the definition of the above variables see Hersbach et al. (2018).

3. Modelling approach

In this section, the methodology to model the extreme observations of wave and wind storms using clustering and BNs is presented. H_s is selected as the dominant variable since it governs most of the coastal and offshore processes relevant within the civil engineering field. The methodology outline is as follows.

(i) Extreme observations of H_s are sampled using Yearly Maxima. Concomitant values (values observed at the same time as the main observation) of the other four variables are also sampled.

(ii) The main wave direction components are identified using clustering. Independent analysis is performed for each component.

(iii) The structure of the BN is defined based on the underlying physics and the correlation between the studied random variables. Correlation is assessed using the Spearman's correlation coefficient, r (Spearman, 1904; Glasser and Winter, 1961).

(iv) The performance of the bivariate Gaussian copulas and the quantification of the BNs are assessed using the d-calibration score (Morales-Nápoles et al., 2013) and in-sample simulations.

3.1. *Clustering*

The wave regime at a given location is usually a combination of different wave components caused by different drivers. Thus, each component needs to be identified and analyzed independently. Here, the k-means algorithm is applied along the values of θ_H. K-means is a simple unsupervised classification algorithm from signal processing science widely used in climate modelling (e.g. Camus et al., 2019; Lucio et al., 2020). The algorithm minimizes the intra-cluster variation defined as the sum of squared Euclidean distances between items and the centroid of the cluster. The total within-cluster variation, TV, is then the sum of the intra-cluster variation of each cluster.

$$TV = \sum_{k=1}^{K} W(C_k) = \sum_{k=1}^{K} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (1)$$

where $W(C_k)$ is the intra-cluster variation of cluster k, x_i is a data point in C_k and μ_k is the mean value of the points assigned to the cluster C_k.

Therefore, a suitable value for the hyper-parameter k, representing the number of clusters, must be selected for the k-means algorithm. Here, $k = 1, 2, ..., 5$ are considered, and the Gap statistic is used to select the optimal number of clusters (Tibshirani et al., 2001). The Gap statistic assesses the goodness of fit of the clustering measure comparing (log) TV with its expected value for a reference distribution without an obvious clustering. This reference distribution is built by generating each reference feature uniformly over the range of the observed values for that feature. The Gap statistic is defined as

$$Gap_n(k) = E_n^*[log(TV)^*] - log(TV) \quad (2)$$

where $E_n^*[log(TV)^*]$ is the expected value of $log(TV)$ for a reference distribution without a clear clustering with sample size n ($log(TV)^*$). $E_n^*[log(TV)^*]$ is then calculated as the average $log(W_k)^*$ from B resamples generated by bootstrapping from the reference distribution. Thus, the higher the Gap statistic, the better. The optimal k is defined as $Gap(k) \geq Gap(k+1) - s_{k+1}$ being s_{k+1} the standard error of the output, calculated as

$$s_k = sd(k)(1 + 1/B)^{0.5} \quad (3)$$

where $sd(k)$ is the standard deviation of $log(TV)^*$ in the B bootstrap resamples.

This analysis is performed using the R-library *Clusters* (Maechler et al., 2022).

3.2. *Copula model*

In this study, bivariate copulas are used to model the probabilistic dependence between variables. Bivariate copulas, or just copulas, are joint distributions with uniform marginal distributions in [0, 1]. According to Sklar (1959), any multivariate joint distribution can be described in terms of a set of univariate marginal distributions and a copula that models the dependence between the variables. For the bivariate case, a copula is defined as

$$H_{XY}(x, y) = C\{F_X(x), G_Y(y)\} \quad (4)$$

where $H_{XY}(x, y)$ for $(x, y) \in \mathbb{R}^2$ is a joint distribution with marginals $F_X(x)$ and $G_Y(y)$ in [0, 1] and a copula in the unit square $I^2 = ([0, 1] \times [0, 1])$, being Eq. (4) satisfied for all $(x, y) \in \mathbb{R}^2$.

In this study, Gaussian copulas, suitable to model variables without strong tail dependence, are considered. The hypothesis of no significant tail dependence between the studied variables is investigated by means of the d-calibration score in Section 4.4.

3.3. *Copula-based Bayesian Networks*

Bayesian Networks (Pearl, 2013) are high-dimensional probability distribution functions composed by a Directed Acyclic Graph (DAG) consisting of a set of nodes and a set of arcs. Each node of a Bayesian Network represents a random variable, while each arc connecting two nodes indicates probabilistic dependence. Thus, a Bayesian Network encodes the joint probability density on a set of variables by specifying conditional probability functions of each variable (child) given its direct preceding variables (parents). In this study, copula-based Bayesian Networks (BNs) are used, where parametric univariate distributions are assigned to the nodes and bivariate copulas are applied to describe the dependence between each pair of random variables. In this study, Gaussian copulas are used since there is previous evidence of their feasibility to model metocean variables (Camus et al., 2019; Lucio et al., 2020) and present computational advantages (Mendoza-Lugo et al., 2022). Further discussion on this can be found in Section 4.4.

BNs are implemented using the Python library BANSHEE (Paprotny et al., 2020; Koot et al., 2023).

4. Building the model

4.1. *Identified extreme wave storms*

Following the procedure indicated in Section 3, 63 extreme values of H_s are identified in the database described in Section 2, and concomitant values of the other 4 random variables are selected. The overview of the sampled extreme events is presented in Fig. 1.

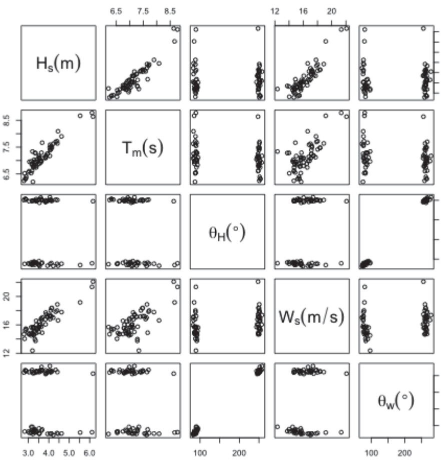

Fig. 1. Overview of sampled extreme wind and wave storms

In Fig. 1, two clear wave and wind components are identified: (1) west component ($\mu = 250.1°$), and (2) east component ($\mu = 88.3°$). This indicates that there are wave storms caused by different drivers and fetches. Applying the k-means algorithm described in Section 3.1, the wave systems are systematically separated. 33 observations correspond to the west component and 30 observations correspond to the east component. Fig. 2 presents the data for the west component.

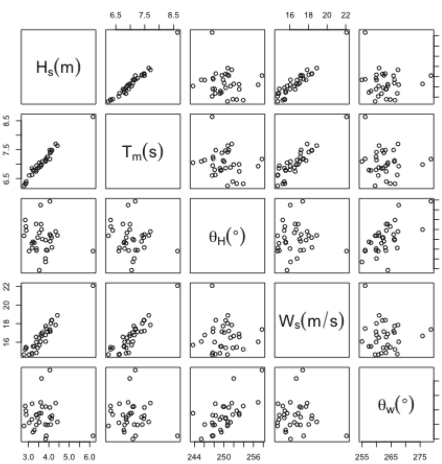

Fig. 2. Overview of the west component (cluster $\mu = 250.1$ °)

4.2. *Empirical correlation matrix*

The Spearman's correlation coefficient, r, is used to assess the dependence between the random variables for both wave components. $r \in [-1, 1]$, where r =1 and -1 represent perfect positive and negative monotonic dependence, respectively, is defined as

$$r = \frac{Cov[R(X), R(Y)]}{\sigma_{R(X)}\sigma_{R(Y)}} \quad (5)$$

where $Cov[R(X), R(Y)]$ is the covariance of the ranked variables, and $\sigma_{R(X)}$ and $\sigma_{R(Y)}$ are their standard deviations of the ranked variables.

Results for west and east components are shown in Tables 1 and 2, respectively.

Table 1. Empirical rank correlation matrix for west component.

r	H_s	T_m	θ_H	W_s	θ_w
H_s	1.00	0.981	-0.098	0.938	-0.027
T_m		1.00	-0.175	0.901	-0.050
θ_H			1.00	0.065	0.599
W_s				1.00	-0.012
θ_w					1.00

Table 2. Empirical rank correlation matrix for east component.

r	H_s	T_m	θ_H	W_s	θ_w
H_s	1.00	0.885	-0.220	0.811	-0.424
T_m		1.00	-0.067	0.552	-0.159
θ_H			1.00	-0.446	0.768
W_s				1.00	-0.654
θ_w					1.00

As shown in Table 1, strong correlations are observed between the pairs $H_s - T_m$, $H_s - W_s$ and $W_s - T_m$ for the west component. Similarly, the east component (Table 2) presents strong correlations between the pairs $H_s - T_m$ and $H_s - W_s$. However, a significantly lower correlation in the pair $W_s - T_m$ is obtained (0.901 vs. 0.551). This may indicate that the west component presents a more relevant contribution of local wind generation (sea waves), although it is still dominated by swell waves.

4.3. *Bayesian Network structure*

In this section, the DAG of the BN is defined attending to two criteria: (1) it must represent the underlying physics of wind and wave generation, and (2) the rank correlation matrix of the BN should approximate the empirical rank correlation matrix.

Wind is produced due to differences in temperature and air pressure in the atmosphere and is the main driver of wave generation. Wind blows along the sea surface, curling it and creating waves. These locally generated waves are called sea waves. Sea waves can propagate out of the generation zone, not being sustained by wind anymore, reaching distant locations. These are called swell waves. Therefore, the wave climate at a given location is the result of a mixture of sea and swell waves, and a relationship between the observations of wind and wave variables exists. For instance, the longer the distance along which the wind blows and the greater the wind speed, the higher the generated waves. This causal relationship is translated into the BN structure defining the wind variables (W_s and θ_w) as parents of wave variables. Figure 3 presents the BN structure defined in this study.

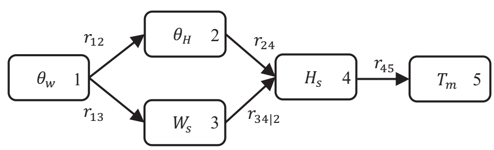

Fig. 3. Structure of the copula-based Bayesian Network defined in this study

4.4. *Model validation*

Two hypotheses must be validated when building a BN: (1) the dependence between the studied variables is well described by Gaussian copulas, and (2) the BN structure sufficiently represents the dependence structure. D-calibration score is used to this end, as mentioned in Section 3.

The d-calibration score (Morales-Nápoles et al., 2013; Morales-Nápoles et al., 2014) provides a measure of "how distant" the elements of two correlation matrices are. This score is 1 if matrices are equal and becomes closer to 0 as matrices differ from each other element-wise. The d-calibration score is given by

$$d(\Sigma_1, \Sigma_2) = 1 - \sqrt{1 - \eta(\Sigma_1, \Sigma_2)} \quad (6)$$

where Σ_1 and Σ_2 are correlation matrices and $\eta(\Sigma_1, \Sigma_2)$ is the Hellinger distance calculated (under the Gaussian copula assumption the term containing the mean vectors is neglected) as

$$\eta(\Sigma_1, \Sigma_2) = \frac{det(\Sigma_1)^{1/4} det(\Sigma_2)^{1/4}}{det\left(\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2\right)^{1/2}} \quad (7)$$

In order to validate the use of Gaussian bivariate copulas, the Empirical Rank Correlation matrices (ERC) are compared to the correlation matrices calculated from a saturated BN (NRC). Saturated BNs (SBNs) are those where all random variables are connected to each other. The validation of the BN structure in Figure 3 is done using a non-parametric Bayesian Network (NPBN) for each wave component. In this work, NPBNs refer to those where the random variables in the nodes are described using the empirical distribution functions. Following studies in the literature such as Mendoza-Lugo et al. (2022, 2019), the NRCs are compared with the rank correlation matrices of the NPBN (BNRC). Additionally, here, BNRCs are compared with ERCs. The calculated d-calibration scores range between 0.68 and 0.89, showing that both Gaussian copulas and the proposed BN structure are suitable to describe the dependence structure of the empirical data.

4.5. *Marginal distributions*

In this study, one-dimensional marginal distributions are fitted to each studied random variable. When dealing with extremes, extrapolation is needed to calculate return levels for very low exceedance probabilities, such as 10^{-6}. This extrapolation requires using parametric marginals. Thus, H_s is fitted using a Generalized Extreme Value distribution (GEV), since it was sampled using Block Maxima. To fit the distribution of the concomitant observations, Gamma, Normal (Norm.) and Lognormal (Logn.) distributions are considered, and the distribution providing the best fit in terms of Loglikelihood is selected. Fig. 4 presents an example of H_s fitting for the west component, while Table 3 shows a summary of the selected marginal distributions.

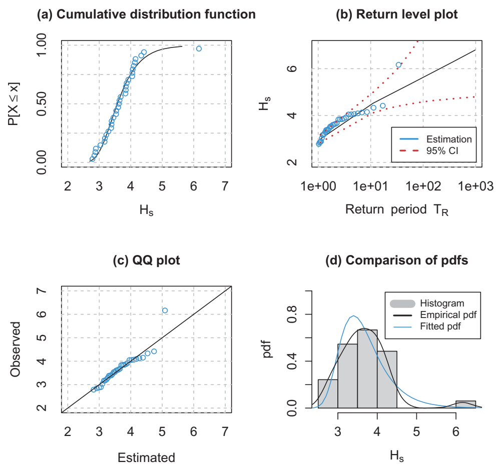

Fig. 4. Generalized Extreme Value marginal distribution, GEV, for H_s of west component.

Table 3. Summary of fitted marginal distributions.

Variable	Direct.	Distr.	μ	σ	ξ
H_s	West	GEV	3.40	0.47	0.02
	East	GEV	3.33	0.39	0.27
T_m	West	Logn.	1.95	0.06	
	East	Logn.	1.98	0.08	
θ_H	West	Logn.	5.52	0.01	
	East	Norm.	88.26	3.27	
W_s	West	Logn.	2.81	0.09	
	East	Logn.	2.75	0.11	
θ_w	West	Logn.	5.57	0.02	
	East	Logn.	4.40	0.07	

As shown in Table 3, the marginal distributions for T_m and W_s are similar for both wave components. However, smaller predictions for T_m and higher predictions of W_s are given for the west component. These results reinforce the observation that the east component is more dominated by swell waves (longer periods and less local wind), while the west component is more driven by local wind generation (sea waves). The fitted univariate distributions are applied to the nodes of the BN.

4.6. *Performance assessment*

The inferences that may be expected from the two developed copula-based BNs are assessed here. To do so, an in-sample procedure is applied. That is, BNs are conditionalized on the measured values of T_m, θ_H, W_s and θ_w, and 1,000 samples of $H_s|T_m, \theta_H, W_s, \theta_w$ are obtained for each case of conditional values. Based on those 1,000 samples of $H_s|T_m, \theta_H, W_s, \theta_w$, the mean and the 5% and 95% percentiles of the conditional distribution are obtained. Figure 5 compares the measured H_s with the mean and the 5% and 95% percentiles of the predictions given by the BN models.

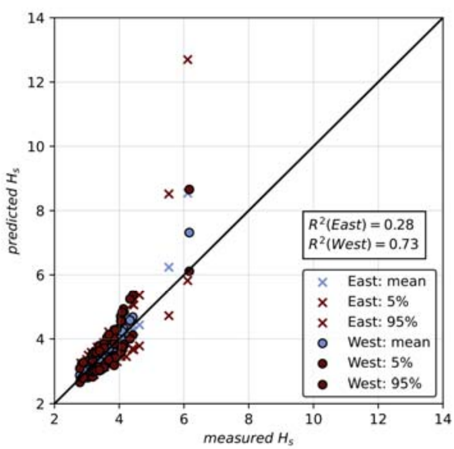

Fig. 5. Comparison between measured and estimated H_s using the BN models.

In-sample estimation of $H_s < 5m$ presents a very good agreement with the measured data, as well as a low variability. However, as H_s grows the BNs start to overestimate H_s, and the variability of the estimation increases. The coefficient of determination, R^2, is used to evaluate the performance of the mean predicted value for each wave component. $0 \leq R^2 \leq 1$ estimates the percentage of the variance explained by the model and is defined in Eq. (8).

$$R^2 = 1 - \frac{\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} (o_i - e_i)^2}{\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} (o_i - \overline{o})^2} \qquad (8)$$

where N_{obs} is the number of observations, o_i and e_i are the observations and estimations, respectively, and $\overline{o}$ is the mean of the observations.

Overall, the developed BNs are a satisfactory model for the purpose under investigation, especially for the west wave component.

5. Conclusions

In this study, the suitability of copula-based Bayesian Networks (BNs) to model extreme wind and wave climatic variables is studied. A location in the Alboran sea (south Spain) from ERA5 database was used as a case study. Yearly maxima is used to sample the extreme observations of wave height, and concomitant values of the remaining variables are taken. The k-means clustering algorithm is applied to identify the different wave components, and a BN is built for each of them. The use of the Gaussian copulas and the structure of the BN are validated using the d-calibration score, obtaining a good data agreement in statistical terms. Finally, fitted marginal distributions are introduced in the nodes of the BNs, and their performance is assessed using in-sample data and R^2. The developed BNs may be regarded as a good model with $R^2 = 0.28$ and $R^2 = 0.73$ for the east and west components, respectively. The results show that BNs are suitable for estimating extreme wind and wave climatic variables and, thus, further investigation with different locations and databases is recommended. Future research lines will also focus on introducing techniques to maximize the sampled maxima and explore the use of vine copulas to better account for possible asymmetries in the joint distributions.

Acknowledgement

This research was partially supported by the Mexican National Council for Science and Technology (CONACYT) under project number 2019-000021-01EXTF-00564 CVU 784544.

References

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.

Di Luccio, D., G. Benassai, G. Budillon, L. Mucerino, R. Montella, and E. Pugliese Carratelli (2018). Wave run-up prediction and observation in a micro-tidal beach. *Natural Hazards and Earth System Sciences 18*(11), 2841 – 2857.

Esteves, L., J. Williams, and J. Brown (2011). Looking for evidence of climate change impacts in the eastern irish sea. *Natural Hazards and Earth System Science 11*(6), 1641 – 1656.

Glasser, G. J. and R. F. Winter (1961). Critical values of the coefficient of rank correlation for testing the hypothesis of independence. *Biometrika 48*(3/4), 444–448.

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.

Kim, B., K. Lee, K. Ko, and J. Choi (2022). Offshore wind resource assessment off the coast of daejeong, jeju island using 30-year wind estimates. *Scientific Reports 12*(1).

Koot, P., M. A. Mendoza-Lugo, D. Paprotny, O. Morales-Nápoles, E. Ragno, and D. T. Worm (2023). Pybanshee version (1.0): A python implementation of the matlab toolbox banshee for non-parametric bayesian networks with updated features. *SoftwareX 21*, 101279.

Lin-Ye, J., M. Garcia-Leon, V. Gracia, and A. Sanchez-Arcilla (2016). A multivariate statistical model of extreme events: An application to the catalan coast. *Coastal Engineering 117*, 138–156.

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.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2022). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.4.

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).

Mendoza-Lugo, M. A., D. J. Delgado-Hernández, and O. Morales-Nápoles (2019). Reliability analysis of reinforced concrete vehicle bridges columns using non-parametric bayesian networks. *Engineering Structures 188*, 178–187.

Mendoza-Lugo, M. A., O. Morales-Nápoles, and D. J. Delgado-Hernández (2022). A non-parametric bayesian network for multivariate probabilistic modelling of weigh-in-motion system data. *Transportation Research Interdisciplinary Perspectives 13*, 100552.

Morales-Nápoles, O., A. M. Hanea, and D. Worm (2013). Experimental results about the assessments of conditional rank correlations by experts: Example with air pollution estimates. In *Proceedings 22nd European Safety and Reliability Conference "Safety, Reliability and Risk Analysis: Beyond the Horizon", ESREL 2013, Amsterdam, the Netherlands, 29–9 to 2–10*. Taylor & Francis Group, London.

Morales-Nápoles, O., D. J. Delgado-Hernández, D. De-León-Escobedo, and J. C. Arteaga-Arcos (2014). A continuous bayesian network for earth dams' risk assessment: methodology and quantification. *Structure and Infrastructure Engineering 10*(5), 589–603.

Paprotny, D. and O. Morales-Nápoles (2017). Estimating extreme river discharges in europe through a bayesian network. *Hydrology and Earth System Sciences 21*(6), 2615–2636.

Paprotny, D., O. Morales-Nápoles, D. T. Worm, and E. Ragno (2020). Banshee–a matlab toolbox for non-parametric bayesian networks. *SoftwareX 12*, 100588.

Pearl, J. (2013). A constraint propagation approach to probabilistic reasoning.

Pham, H. X., A. Y. Shamseldin, and B. W. Melville (2021). Projection of future extreme precipitation: a robust assessment of downscaled daily precipitation. *Natural Hazards 107*(1), 311 – 329.

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

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology 15*(1), 72–101.

Tibshirani, R., G. Walther, and T. Hastie (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(2), 411–423.