Probabilistic estimation of the mean wave overtopping discharge on mound breakwaters

Mares-Nasarre, Patricia

DOI
10.1016/j.coastaleng.2025.104792

Publication date
2025

Document Version
Final published version

Published in
Coastal Engineering

Citation (APA)
Mares-Nasarre, P. (2025). Probabilistic estimation of the mean wave overtopping discharge on mound breakwaters. *Coastal Engineering*, *201*, Article 104792. https://doi.org/10.1016/j.coastaleng.2025.104792

Coastal Engineering 201 (2025) 104792

Contents lists available at ScienceDirect

Coastal Engineering

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

Probabilistic estimation of the mean wave overtopping discharge on mound breakwaters

Patricia Mares-Nasarre

Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.15790655

Keywords:
Overtopping
Wave overtopping
Mean wave overtopping discharge
Mound breakwater
Coastal structure
Copula
Bayesian network

ABSTRACT

This study develops a probabilistic model, a Gaussian copula-based Bayesian Network (BN), to explain the joint probability distribution of the dimensionless mean wave overtopping discharge ($Q = q/\sqrt{gH_{m0}^3}$, being q the mean wave overtopping discharge, g the gravity acceleration and H_{m0} the spectral significant wave height) and a set of explanatory variables on mound breakwaters. This model estimates the distribution of Q conditional to the values of (all or some of) the explanatory variables. The goal of this model is to allow the incorporation of the uncertainties of the structural response and the overtopping phenomenon to probabilistic frameworks. Given a tolerable Q value, a probability of failure can be directly computed from the distribution of Q estimated by the developed BN, differently to current methods in the literature which are deterministic. To develop the BN, a subset of CLASH database focused on mound breakwaters is used (3,179 tests), using 80% of those tests for training and 20% for statistical and performance testing. Ten dimensionless explanatory variables are selected with the following experimental ranges: bottom slope, $7.6 \leq m \leq 1000$; wave attack angle, $0 \leq \beta \leq 80°$; roughness factor, $0.38 \leq \gamma_f \leq 1.00$; dimensionless crest freeboard, $0 \leq R_c/H_{m0} \leq 4.37$; wave steepness, $1.31 \cdot 10^{-3} \leq s_{-1,0} \leq 0.069$; dimensionless width of the crest berm, $0 \leq G_c/H_{m0} \leq 6.67$; dimensionless height of the crest berm, $0 \leq A_c/H_{m0} \leq 4.2$; dimensionless width of the crest of the toe berm, $0 \leq B_t/H_{m0} \leq 15.9$; dimensionless water depth at the toe of the structure, $1.03 \leq h/H_{m0} \leq 17.6$; and armor slope, $1.19 \leq cot\alpha \leq 4$. Empirical cumulative distribution functions are used to quantify the nodes of the BN. The Gaussian copula assumption is successfully validated using the training subset. The proposed model is evaluated using the testing subset in both statistical and performance terms. In statistical terms, the proposed model seems to satisfactorily capture the dependence structure between the studied variables. In performance terms, the predicted mean of the distribution of Q is a reasonable estimator of Q ($R^2 = 0.78$) and the percentage of the observations that lay within the predicted 90% confidence intervals is close to the expected 90%. Finally, the use of the model for the probabilistic design of the crest elevation of mound breakwaters is also illustrated through one example. It should be noted that the less information provided to the model, the wider the estimated distribution of Q as the uncertainty is higher.

1. Introduction

Ports are essential components of the global supply chain, playing a vital role in economic activities such as international trade and logistics (Haralambides, 2017). Ports are typically located in low-lying coastal areas, where they are exposed to a range of natural hazards, including storm surges, wave storms, and sea level rise (Verschuur et al., 2023; Lucio et al., 2024). To guarantee safe and continuous operations within these environments, coastal protection structures, such as mound breakwaters, are fundamental. In response to increasing environmental uncertainties and the need for optimized resource allocation, the design of coastal structures is progressively shifting from traditional deterministic approaches to probabilistic design frameworks (e.g.: ROM:0.0-01, 2001; ROM:1.0-09, 2009), which better account for the aleatoric uncertainty in loading conditions. Although various frameworks in the literature focus on the probabilistic modeling of environmental loadings (e.g.: Lucio et al., 2020, 2024), relatively few studies provide approaches to account for the uncertainty in the structural response of coastal defenses (e.g.: Mares-Nasarre et al., 2024b).

Wave overtopping is a stochastic process that results of the interaction of individual waves with the mound breakwater which, after suffering a number of transformations, may overtop. Specifically for mound breakwaters, wave energy transforms when interacting with the

E-mail address: p.maresnasarre@tudelft.nl.

https://doi.org/10.1016/j.coastaleng.2025.104792
Received 5 February 2025; Received in revised form 23 May 2025; Accepted 25 May 2025
Available online 2 July 2025

structure by being transmitted through the breakwater via overtopping and flow through the porous medium, and dissipated through wave breaking on the seaward slope and crest of the structure, and turbulent flow and friction inside the porous medium (Clavero et al., 2020). These processes are influenced by a large number of uncertain variables in practice, such as the porosity and placement of the elements in the core, filters, and armor, or the geometry of the overall structure and the different layers that compose the mound breakwater, between others (Medina et al., 2014; Díaz-Carrasco, 2023). During the design phase of a mound breakwater, the tolerable mean wave overtopping discharge (q) is usually the applied criterion to determine the crest elevation of mound breakwaters. q is then used as a representative variable of overtopping that aims to summarize the result of a stochastic process (individual waves in a wave storm) interacting with the mound breakwater with a number of uncertainties.

Numerous methods to estimate q can be found in the literature, from empirical equations (e.g.: EurOtop, 2018; Molines and Medina, 2016) to numerical models (e.g.: Irías Mata and van Gent, 2023; Molines et al., 2019) and machine learning algorithms (e.g.: den Bieman et al., 2021; Carro et al., 2024). To the author's knowledge, all the reported methods are of a deterministic nature. However, Romano et al. (2015) already reported wave overtopping as a highly uncertain phenomenon as q can vary by up to an order of magnitude for different time series realizations from the same wave spectrum. In order to account for the natural variability of q, methods in the literature quantify confidence intervals based on observed errors in a limited set of observations (Mares-Nasarre et al., 2020; Molines and Medina, 2016), Gaussian-distributed parameters in empirical equations (EurOtop, 2007, 2018), or ensembles of models built using a bootstrap strategy (van Gent et al., 2022; den Bieman et al., 2021). Although these confidence intervals provide an estimate of the observed error, they have two main limitations. First, they do not quantify the distribution of q. Therefore, the provided estimations of q cannot be embedded into a probabilistic design framework which accounts for the uncertainty in the structural response and provides with a probability of failure as a result. Second, the provided percentiles are only based on the value of q, not accounting for the influence of the explanatory variables on the distribution of q. This is, two situations with a comparable predicted value of q by a deterministic method can be caused by different situations. For instance, a more extreme wave storm with oblique wave attack and a milder wave storm with perpendicular wave attack on the same structure can have a similar predicted q. However, the distribution of q can be significantly different; the situation with a more extreme wave storm can lead to a more asymmetrical distribution of q with a longer right tail due to the higher uncertainty in the more extreme waves in the wave storm and, thus, a different probability of failure. Therefore, tools capable of estimating the distribution of q as function of the explanatory variables are needed to capture the stochastic nature of overtopping and improve the design of mound breakwaters under wave attack.

Copula-based models are widely used to model uncertainty in the literature (e.g.: Leontaris et al., 2016; Torres-Alves and Morales-Nápoles, 2020; Mares-Nasarre et al., 2024a). Specifically in the Coastal Engineering field, they have been applied to model the joint probability distribution of wave height and period (Antão and Guedes Soares, 2014; Jaeger and Morales-Nápoles, 2017) or the dependence between the hydrodynamic variables of wave overtopping on mound breakwaters (Mares-Nasarre et al., 2024b), between other applications. Within the existing copula-based models, Gaussian copula-based Bayesian Networks (BNs) (Kurowicka and Cooke, 2004; Hanea et al., 2006; Hanea et al., 2010, 2015) are gaining popularity as they enable the definition of joint distributions in high-dimensional spaces, combining the flexibility of copulas with efficient computations made possible by the Gaussian copula assumption. Examples of application of BNs range from the joint distribution of wind and wave variables (Mares-Nasarre et al., 2023) to the characterization of the variability of the corrosion depth in the elements of a steel bridge (Barros et al., 2024).

This study proposes a probabilistic model, a BN, to describe the joint probability distribution of q and a number of explanatory variables on conventional mound breakwaters based on part of CLASH database (CLASH database, 2005). Thus, the main contribution of this paper is the development of a probabilistic model based on Gaussian copulas capable of estimating the distribution of q given a number of explanatory variables. Multivariate probabilistic models relate the probability distribution of different variables and, thus, allow to compute the distribution of one variable given the value of other variables in the model. Specifically in this study, the proposed model enables the computation of the probability distribution of q given the values of (all or some of) the explanatory variables. The paper is structured as follows. In Section 2, an overview of the existing methods in the literature to estimate q is presented. In Section 3, the selection of the observations from CLASH database and the explanatory variables for q are described. In Section 4, the theoretical concepts related to BNs are exposed. In Section 5, the definition and validation of the model are explained. In Section 6, an example of an application of the model is presented. Finally, in Section 7, conclusions and recommendations for future research are drawn.

2. Methods for the estimation of the mean wave overtopping discharge

Since Owen (1980) proposed an empirical exponential formulation to estimate q as a function of the non-dimensional crest freeboard R_c/H_{m0}, several authors have derived similar empirical expressions for the prediction of q. Based on the findings of studies such as van der Meer and Janssen (1994) and van Gent (2001), TAW (2002) manual recommended the following formulation

$$Q = \frac{q}{\sqrt{gH_{m0}^3}} = a_1 \xi_{-1,0} exp\left(-b_1 \frac{R_c}{H_{m0}} \frac{1}{\gamma_f \xi_{-1,0}}\right) \tag{1a}$$

with a maximum of

$$Q = a_2 \; exp\left(-b_2 \frac{R_c}{H_{m0}} \frac{1}{\gamma_f}\right) \tag{1b}$$

where Q is the dimensionless mean wave overtopping discharge, $a_1 = 0.067$ and $b_1 = 4.74$, and $a_2 = 0.2$ and $b_2 = 2.6$ for Eqs. (1a) and (1b), respectively, are empirical coefficients, $\xi_{-1,0} = \frac{tan\alpha}{2\pi H_{m0}/(gT_{-1,0}^2)}$ is the surf similarity parameter or Iribarren number computed using the spectral significant wave height $H_{m0} = 4m_0^{0.5}$ and the spectral wave period $T_{-1,0} = \frac{m_{-1}}{m_0}$, where m_i is the ith spectral moment, R_c is the structure crest freeboard, and γ_f is the roughness factor. The variability of the fitting to the observations of q is proposed to be modeled by assuming the coefficient b_1 as Gaussian-distributed with a standard deviation of 0.5.

EurOtop (2007) recommended using Eq. (1b) and correct the provided estimate using the reduction factor proposed by Besley (1999), Cr, if $G_c > 3D_{n50}$, where G_c is the width of the berm of the crest of the structure and D_{n50} = is the nominal diameter of the elements in the armor. The reduction factor Cr is computed as

$$Cr = min\left[1.0, 3.06 \; exp\left(-1.5 \frac{G_c}{H_{m0}}\right)\right] \tag{2}$$

The uncertainty of the fitting of Eq. (1b) recommended in EurOtop (2007) is proposed to be modeled by assuming the coefficient b_2 as Gaussian-distributed with a standard deviation of 0.35.

Between 2001 and 2003, the European project CLASH collected 10,532 overtopping tests of different coastal structures and laboratories (Steendam et al., 2005). Each of the tests were described using a number of variables containing information about the geometry of the structure and the tested hydraulic conditions. Fig. 1 illustrates the main explanatory variables of q considered in the present study. In addition, two factors were defined regarding the complexity of the tested structure and the reliability of the test: the Complexity Factor,

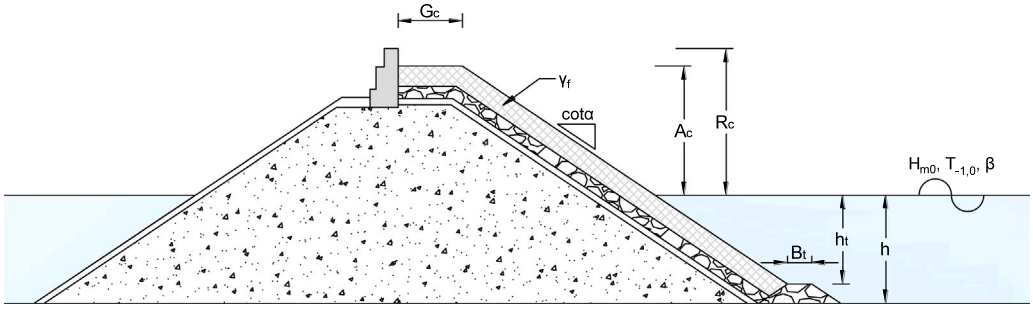

Fig. 1. Cross-section of a conventional mound breakwater and variables definition.

$1 \leq CF \leq 4$, being $CF = 1$ a 'very simple' structure where the variables in the database describe accurately the cross-section, and $CF = 4$ a 'very complex' structure, and the Reliability Factor, $1 \leq RF \leq 4$, being $RF = 1$ a 'very reliable' test where all the required information was directly available, and $RF = 4$ for a 'non-reliable' test.

Based on part of that database (filtering including $CF < 4$, $RF < 4$ and $q > 0$), van Gent et al. (2007) trained and validated Multi-layer Feed-forward Neural Networks (NNs) with three layers to estimate q. Making use of bootstrapping, 500 NNs were created and used to compute confidence intervals for the predictions. The developed model, named CLASH NN, is still between the most accurate predictors for q according to scientific literature (e.g.: Mares-Nasarre et al., 2020; Molines and Medina, 2016).

Zanuttigh et al. (2016) extended CLASH database with tests reported by the authors and Besley et al. (1993), Oumeraci et al. (2007), Andersen et al. (2009), Victor and Troch (2012), and Van Doorslaer et al. (2015). Based on the new database, Zanuttigh et al. (2016) developed a new predictive model based on a classifier and three NNs. Confidence intervals in the predictions of q were computed in a similar fashion as van Gent et al. (2007) based on bootstrapping.

Molines and Medina (2016) used CLASH NN and CLASH database to develop an explicit estimator for q on conventional mound breakwaters in non-breaking conditions using only 6 explanatory dimensionless variables: R_c/H_{m0}, G_c/H_{m0}, where G_c is the width of the crest berm, $\xi_{-1,0}$, R_c/h, where h is the water depth at the toe of the structure, A_c/R_c, where A_c is the freeboard of the crest berm, and a toe berm variable based on R_c/h. A factor was also proposed to account for oblique wave attack.

EurOtop (2018) added a power law to Eq. (1b) to account for freeboards close to 0 and modified γ_f as

$$Q = 0.09 exp\left(-\left(1.5\frac{R_c}{H_{m0}\gamma_{f,mod}}\right)^{1.3}\right) \tag{3a}$$

with

$$\gamma_{f,mod} = \gamma_f + \frac{(\xi_{-1,0} - 5)(1 - \gamma_f)}{5} \tag{3b}$$

van Gent (2022) noted that the power also affects the influence factors. The variability of the fitting of Eq. (3) is given by assuming both coefficients in Eq. (3a) as Gaussian-distributed with standard deviations $\sigma(0.09) = 0.0135$ and $\sigma(1.5) = 0.15$.

One of the last methods to be applied to the estimation of q is the XGBoost algorithm in den Bieman et al. (2021). This machine learning algorithm is based on ensembles of regression trees and makes use of the principle that a combination of weak predictions can form a strong predictor. Confidence intervals for q were computed based on 500 models built using 500 bootstrap resamples, following the methodology in van Gent et al. (2007).

In this section, models in the literature to estimate q on mound breakwaters have been briefly summarized; all the proposed methods are deterministic and cannot quantify the distribution of q given a number of explanatory variables. Therefore, further research is needed to develop probabilistic estimators of q that account for the stochastic nature of the phenomenon and can improve the design of mound breakwaters by incorporating structural uncertainty in probabilistic design frameworks.

3. Dataset and variable selection

This section first describes the data used in this research. Afterwards, the selection of the explanatory variables of q is presented based on the existing literature.

3.1. Dataset based on CLASH database

Here, the database collected in the European CLASH project (Steendam et al., 2005) is used. However, the CLASH database includes a wide variety of typologies of coastal structures, while this research is focused on conventional mound breakwaters. Therefore, a subset of CLASH database is used here. The following filtering criteria are applied: the slope needs to be continuous without changes in the slope angle (slopes in the upper and lower slope, $cot\alpha_{up} = cot\alpha_{down}$) or berms in the slope (width of the berm, $B = 0$, slope of the berm, $tan\alpha_b = 0$ and the depth of the berm, $h_b = 0$), the slope angles need to be in ranges reasonable for conventional mound breakwaters ($1.19 \leq cot\alpha \leq 4$, Molines and Medina, 2016) and the structure needs to be "simple" ($CF = 1$). Moreover, since erroneous data can degrade the performance of the developed models, the tests with low reliability were removed from the dataset as $RF \leq 2$ and only tests with significant overtopping were considered, $Q = q/\sqrt{gH_{m0}^3} > 10^{-6}$. After the filtering of the dataset, a total of 3179 tests were obtained.

The obtained dataset is splitted in training and testing subsets for further analysis. From the obtained dataset, 2582 tests were performed under perpendicular wave attack, while 597 were conducted under oblique wave attack. Therefore, in order to guarantee the presence of tests with oblique waves in both training and testing subsets, the dataset is first split by wave angle of incidence as $\beta = 0$ and $\beta > 0$. From those two subset, 80% of the observations are used for training, and 20% for testing. Thus, 2544 tests were used for training and 635 tests were used for testing, having 2066 tests and 516 tests under perpendicular wave attack, respectively.

3.2. Selection and definition of dimensionless variables

This study is focused on the mean wave overtopping discharge, q, which is made dimensionless as $Q = q/\sqrt{gH_{m0}^3}$ (e.g.: TAW, 2002; EurOtop, 2018). The following dimensionless explanatory variables are selected based on the literature:

1. $m = cot\alpha_{bottom}$, is the foreshore slope. It determines the type of wave breaking on the toe of the structure and has been included as an explanatory variable in previous studies to estimate Q (Zanuttigh et al., 2016; den Bieman et al., 2021) or the distribution of individual wave overtopping volumes (Mares-Nasarre et al., 2024b).

2. β, is the angle of wave attack. Oblique wave attack is known to reduce wave loading in general (e.g.: van Gent, 2014; Yu et al., 2002) and wave overtopping in particular (e.g. Lykke Andersen and Burcharth, 2009; Galland, 1994).
3. γ_f, is the roughness factor. This factor accounts for the armor unit, the number of layers of the armor, the porosity of the armor, between other characteristics of the structure (Molines and Medina, 2015; Pepi et al., 2022).
4. R_c/H_{m0}, is the dimensionless crest freeboard. It is the most widely accepted dimensionless variable to describe the mean wave overtopping discharge. It can be found as a primary variable in empirical formulas (e.g.: Owen, 1980; EurOtop, 2018) and machine learning algorithms (e.g.: van Gent et al., 2007; den Bieman et al., 2021) to estimate Q.
5. $s_{-1,0} = H_{m0}/L_{-1,0}$, is the wave steepness calculated with $L_{-1,0} = gT^2_{m-1,0}/2\pi$. Wave steepness is one of the key variables that determine the hydraulic behavior of mound breakwaters, namely the reflected, dissipated and transmitted energy through the structure (Díaz-Carrasco et al., 2020; Díaz-Carrasco, 2023). Studies such as Koosheh et al. (2022) and van Gent et al. (2022) highlighted the role of wave steepness and included it in the derived empirical expressions to compute Q. Regarding the choice of wave period to compute the wave steepness, van Gent (2001) showed that $T_{-1,0}$ can be applied to accurately estimate wave overtopping for a number of spectral shapes.
6. G_c/H_{m0} is the dimensionless width of the crest berm. Wider crest berms lead to a reduction of Q as pointed out in studies such as Besley (1999) or Molines and Medina (2016).
7. A_c/H_{m0} is the dimensionless height of the crest berm. Similar to R_c/H_{m0}, the higher the crest berm, the lower the Q (e.g.: Molines and Medina, 2016; van Gent et al., 2007).
8. B_t/H_{m0} is the dimensionless width of the toe berm. B_t has already been reported as significant by van Gent et al. (2007), Molines and Medina (2016) or den Bieman et al. (2021).
9. h/H_{m0}, is the dimensionless water depth at the toe of the structure. This variable is applied as an indicator of whether waves are depth-limited (e.g.: van Gent, 1999, defined it using the deep water H_{m0}). Thus, it is included in methods to estimate Q (e.g.: Koosheh et al., 2022; van Gent et al., 2007).
10. $cot\alpha$ is the slope of the structure on the sea side. The slope angle determines the length of the slope and, thus, the energy dissipated through the armor layer. Thus, it has a great impact on Q as highlighted in Irías Mata and van Gent (2023) or Altomare et al. (2016).

Most of these variables are also depicted in Fig. 1. In addition of the aforementioned explanatory variables, h_t/H_{m0}, where h_t is the water depth over the toe berm, was also analyzed. However, it was discarded as the observed (rank) correlation between h/H_{m0} and h_t/H_{m0} was almost 1, indicating that h_t/H_{m0} did not provide additional information when h/H_{m0} was already included.

The experimental ranges of the selected observations from CLASH database used in this study are $7.6 \leq m \leq 1000$, $0 \leq \beta \leq 80°$, $0.38 \leq \gamma_f \leq 1.00$, $0 \leq R_c/H_{m0} \leq 4.37$, $1.31 \cdot 10^{-3} \leq s_{-1,0} \leq 0.069$, $0 \leq G_c/H_{m0} \leq 6.67$, $0 \leq A_c/H_{m0} \leq 4.2$, $0 \leq B_t/H_{m0} \leq 15.9$, $1.03 \leq h/H_{m0} \leq 17.6$, and $1.19 \leq cot\alpha \leq 4$.

4. Methodology

In this research, a Gaussian copula-based Bayesian Networks (BN) is implemented as a model to describe the uncertainty of the mean wave overtopping discharge, q. In this section, the concept of bivariate copula is introduced. Afterwards, the definition and validation of BNs are explained.

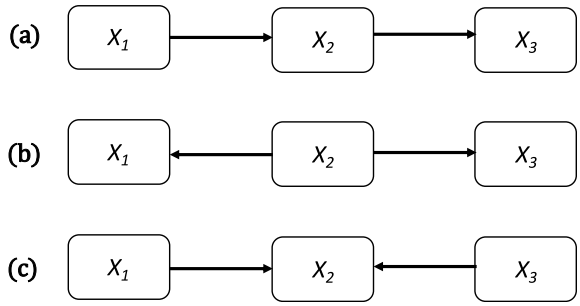

Fig. 2. Possible DAGs for 3-nodes Bayesian Networks.

4.1. Concept of bivariate copula

Bivariate copulas, or just copulas, are bivariate distributions with uniform margins in [0, 1]. Following Sklar (1959)'s theorem, any multivariate distribution of continuous variables can be described as a set of univariate margins and a copula that models the dependence. The definition of copula for the bivariate case is given by

$$H_{X,Y}(x, y) = C\{F_X(x), G_Y(y)\} \tag{4}$$

where $H_{X,Y}(x, y)$ for $(x, y) \in \mathbb{R}^2$ is a bivariate 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 $(x, y) \in \mathbb{R}^2$.

Different families of copulas exist in the literature (see Czado, 2019). One of the features that distinguishes between copula families is tail dependence which characterizes the correlations in the tails of the distributions of two random variables. The upper tail dependence coefficient is defined as $\lambda_{upper} = \lim_{t\to 1^-} P(X_2 > F_2^{-1}(t)|X_1 > F_1^{-1}(t))$ (Sibuya et al., 1960; Joe, 1997). In this research, the Gaussian copula is used, which is a symmetric model and, thus, does not present tail dependence. For further information about copulas the reader is referred to Nelsen (2006). The bivariate Gaussian copula is given by

$$H_{XY}(x, y) = \Phi_2\left(\Phi^{-1}(x), \Phi^{-1}(y)|\rho\right) \tag{5}$$

where $\Phi_2\{.,.|\rho\}$ is the cumulative distribution function of the bivariate normal distribution with 0 expectation, unit variance and ρ Pearson correlation coefficient (Pearson and Galton, 1895).

4.2. Gaussian copula-based Bayesian networks

Bayesian Networks (Pearl, 2013) are high-dimensional probability distribution functions composed by a directed acyclic graph (DAG) composed of a set of nodes and a set of arcs. Nodes represent random variables, while arcs connecting two nodes indicate the probabilistic dependence between the connected random variables. Therefore, a Bayesian Network encodes the joint probability density on a set of random variables by specifying conditional probability functions of each variable (child) given its direct preceding variables (parents). Fig. 2 shows the three possibilities of 3-nodes DAGs for Bayesian Networks. Conditional independence statements can be read as follows. In Fig. 2(a) and (b), X_1 and X_3 are not independent ($X_1 \not\perp X_3$). However, X_1 and X_3 are conditionally independent given X_2 ($X_1 \perp X_3|X_2$). In Fig. 2(c), X_1 and X_3 are independent ($X_1 \perp X_3$), as no arc is connecting them, but they become not independent given X_2 ($X_1 \not\perp X_3|X_2$).

In this study, copula-based Bayesian Networks (BNs) are used, as introduced in Kurowicka and Cooke (2004) and extended in Hanea et al. (2006) and Hanea et al. (2010), where parametric univariate distributions are assigned to the nodes and bivariate copulas are applied to describe the dependence between each pair of random variables. BNs build the joint distribution function of a set of random variables by coupling the marginal distributions with the dependence structure in bivariate pieces so that the conditional dependence statements given

by the DAG are preserved. Thus, quantifying a BN implies a number of marginal distributions equal to the number of nodes in the DAG and a number of (conditional) dependence parameters equal to the number of arcs. Here, empirical distribution functions are used as univariate margins, as fitting parametric distributions is only beneficial when inferring extremes (probabilities that have not been observed yet) is necessary. In principle, different bivariate copulas can be used to quantify the arcs. In this study, the bivariate Gaussian copula (see Eq. (5)) is adopted for its advantages in computations and inference of complex problems (Mendoza-Lugo et al., 2022; Hanea et al., 2015). Therefore, the protocol given in Hanea et al. (2006) with the Gaussian-copula assumption is applied in this research. The open-source toolbox BANSHEE (Paprotny et al., 2020; Koot et al., 2023) in Python is applied to implement the BNs.

4.2.1. Validation of Gaussian copula-based Bayesian networks

When working with BNs, two main validations need to be performed: (1) the assumption of the Gaussian copula to describe the dependence between the variable pairs is reasonable, and (2) the defined DAG succeeds in capturing the dependence structure between the random variables.

In order to validate the assumption of the use of the Gaussian copula, the empirical copula defined from the observations of a pair of random variables is compared with: (1) Gaussian, (2) Frank, (3) Clayton, and (4) Gumbel. The objective of this comparison is to detect whether copulas with tail-dependence (Clayton and Gumbel) arise as better models and, thus, a symmetric model such as the Gaussian copula might fail in describing the dependence structure. The aforementioned comparison is performed using Cramer–von–Mises statistic (Genest et al., 2009). The Cramer–von–Mises statistic (S_{CvM}) evaluates the distance between the empirical and the parametric copula; the perfect fit means $S_{CvM} \to 0$. Therefore, the parametric copula with the lowest value of S_{CvM} provides the best fit.

Following the protocol by Hanea et al. (2006), a rank correlation matrix can be derived from the DAG of the BN. Therefore, validation procedures based on the comparison between rank correlation matrices are used. A rank correlation matrix is a matrix composed of the rank correlations between two of the random variables so the element $e_{i,j}$ is the rank correlation between the random variables i and j. The Spearman's rank correlation coefficient (Spearman, 1904), r, quantifies the strength and direction of the association between two ranked variables. Thus, it assess monotonic relations between two variables. $r \in [-1, 1]$, where $r = 1$ and -1 represent perfect positive and negative monotonic dependence, respectively. r is given by

$$r = \frac{Cov[R(X), R(Y)]}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{6}$$

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

r assesses the strength and direction of association between two ranked variables. This is, it provides a measure of monotonicity of the relation between two variables. $r \in [-1, 1]$, where $r = 1$ and -1 represent perfect positive and negative monotonic dependence, respectively. r is defined as

In this study, three comparisons are performed: the first validates that the joint Gaussian copula adequately represents the underlying multivariate distribution, the second verifies that the defined DAG reasonably represents the observed dependence structure, and the third assesses how far the proposed model is from the best possible model. For the first comparison, the rank correlation matrix obtained from the observations (empirical rank correlation, ER) is compared with that obtained from a saturated BN (SR). A saturated BN is one where all nodes are connected to each other and, thus, represents the best possible model. However, it does not explain the underlying dependence structure and can be seen as an "overfit" of the observations. If SR and ER are "close" to each other, it means that the assumption of the joint Gaussian copula is able to describe the observed dependence. For the second comparison, ER is compared to the rank correlation matrix obtained from the defined DAG, BR. If ER and BR are "close", it means that the main features of the observed dependence structure are properly represented, being thus the DAG a satisfactory model. Finally, for the third comparison, BR is compared with SR. This enables determining how close the defined DAG is to the best possible model, SR.

To perform these comparisons and quantify how different ER, SR and BR are, dissimilarity measures for Gaussian densities are applied. Specifically, Morales-Nápoles et al. (2013, 2014) defined the d-calibration score ($d(\Sigma_1, \Sigma_2)$) as

$$d(\Sigma_1, \Sigma_2) = 1 - \sqrt{1 - \eta(\Sigma_1, \Sigma_2)} \tag{7}$$

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 as

$$\eta(\Sigma_1, \Sigma_2) = \frac{det(\Sigma_1)^{1/4} det(\Sigma_2)^{1/4}}{det\left(\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2\right)^{1/2}} \tag{8}$$

$d(\Sigma_1, \Sigma_2)$ provides a measure of "how distant" the elements of two correlation matrices are. If the score is 1, matrices are equal, and it becomes closer to 0 as matrices differ from each other element-wise.

5. Building the model

This section starts with the development of the proposed BN. Afterwards, the proposed model is validated and, finally, its performance is tested.

5.1. Definition of the model

The empirical rank correlation based on the training subset, ER, is computed as a starting point to define the model (see Section 4.2.1) and shown in Fig. 3(a). The significance of the individual rank correlations, r, was also computed and almost all the observed r were found significant (see Fig. 9 in Appendix A).

In order to define the DAG, first, arcs are added between the variables that present the strongest rank correlations ($|r| > 0.3$). That provides with a first "skeleton" of the model. Afterwards, arcs are added one by one between two random variables following with $|r| > 0.1$ and it was assessed whether that arc made a difference in the $d-cal$ of at least $\Delta(d - cal) > 0.01$. Once adding more arcs did not make a difference $\Delta(d - cal) > 0.01$, arcs were removed one by one from the obtained BN to test their significance in the final model. If removing the arc made a difference $\Delta(d - cal) > 0.01$, the arc stayed. If it did not, it was removed. Fig. 4 shows the final defined DAG.

As described in Section 4, the defined DAG was quantified following the procedure described in Hanea et al. (2006) and the nodes were quantified using empirical distribution functions.

5.2. Validation of the model

As exposed in Section 4.2.1, the assumption of the Gaussian copula is first validated using S_{CvM} to compare the fit of different copula families to the empirical bivariate copulas observed in the training subset. Fig. 10 in Appendix B presents the values of S_{CvM} for each pair of random variables. In 29% of the pairs Gaussian copula is selected as the best fitting model, while in 51% of the pairs a symmetric model (either Gaussian or Frank here) is selected as the best model. Moreover, in 80% of the pairs, either the Gaussian copula provides the best performance or its metric has a difference lower than 15% with the best fitting copula. Therefore, overall, it can be assumed that tail dependence does not seem to be significant in most of the random variable pairs and, thus, Gaussian copula can be a reasonable model.

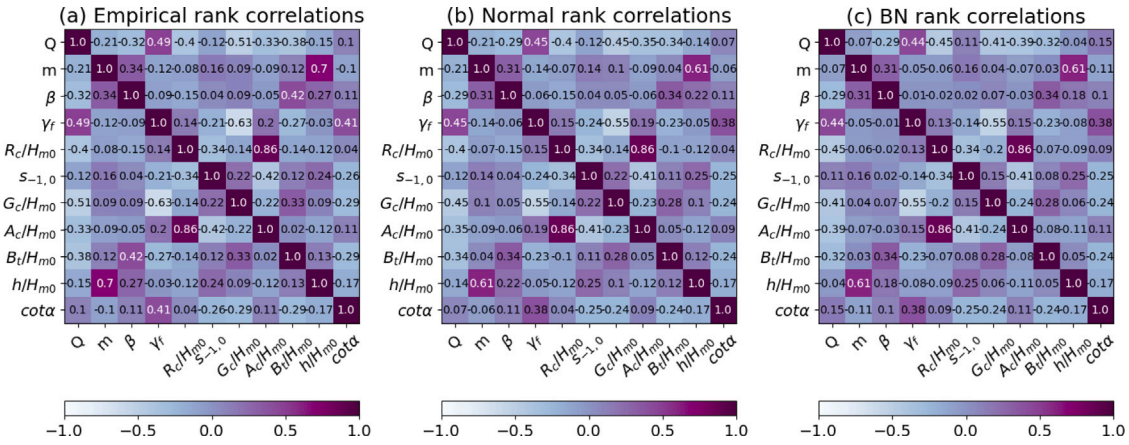

Fig. 3. Rank correlation matrices based on the training subset: (a) empirical, (b) saturated model, and (c) BN model.

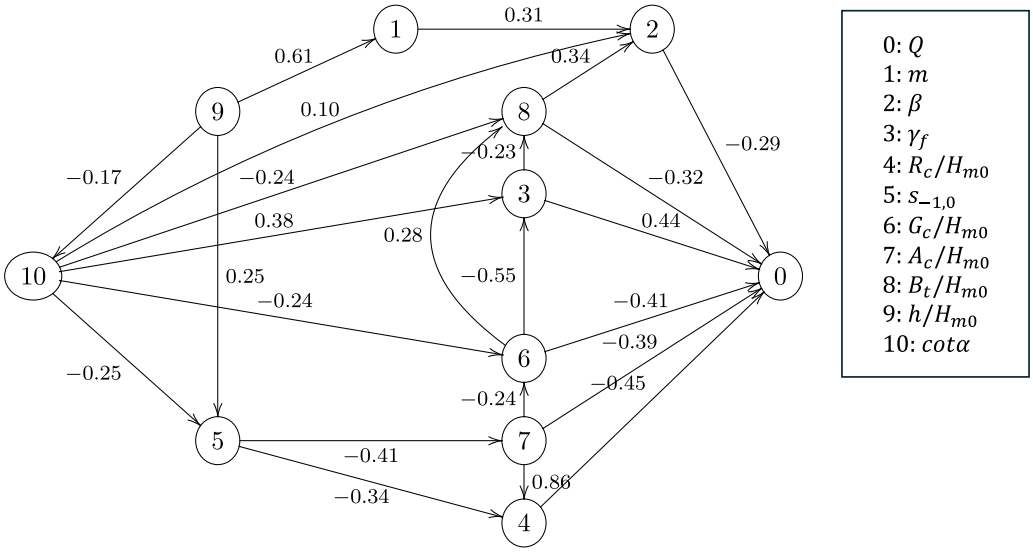

Fig. 4. Proposed BN model. Labels in the nodes represent random variables while labels in the edges are (un)conditional rank correlations.

Further validations are performed through the comparison of rank correlation matrices (see 4.2.1). Fig. 3 presents the empirical rank correlation matrix computed from the training subset (ER), the rank correlation matrix corresponding to the saturated BN (SR) and the rank correlation matrix obtained from the proposed DAG in Fig. 4. The differences between these three matrices are assessed using $d-cal$. $d-cal(ER, SR) = 0.88$ indicates that the empirical observations are well described by the Gaussian copula, using SR as the best possible model. $d-cal(BR, SR) = 0.76$ and $d-cal(ER, BR) = 0.71$ indicate that the defined DAG, although not being saturated, approximates well the best possible model, SR, and the dependence in the observations.

Note that these validation steps are conducted using the data in the training subset and are also part of the development of the model as explained in Section 4.2.1.

5.3. Testing of the model

In this section, the test subset is used to evaluate the model from both the statistical perspective and the performance perspective. First, the proposed model is validated from the statistical perspective. To this end, the empirical rank correlation based on the testing subset, ER_{test}, is computed, as shown in Fig. 5.

ER_{test} is compared to SR and BR using $d-cal$ similar to the previous section. $d-cal(ER_{test}, SR) = 0.88$ and $d-cal(ER_{test}, BR) = 0.71$

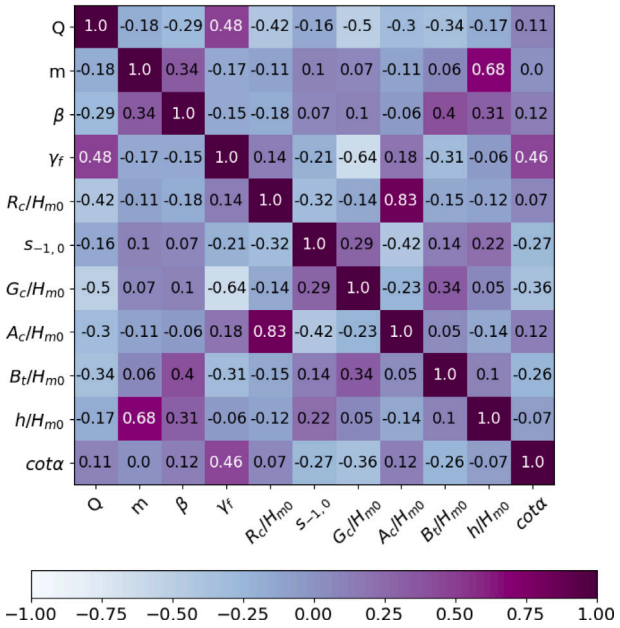

Fig. 5. Rank correlation matrix computed from the observations in the testing subset.

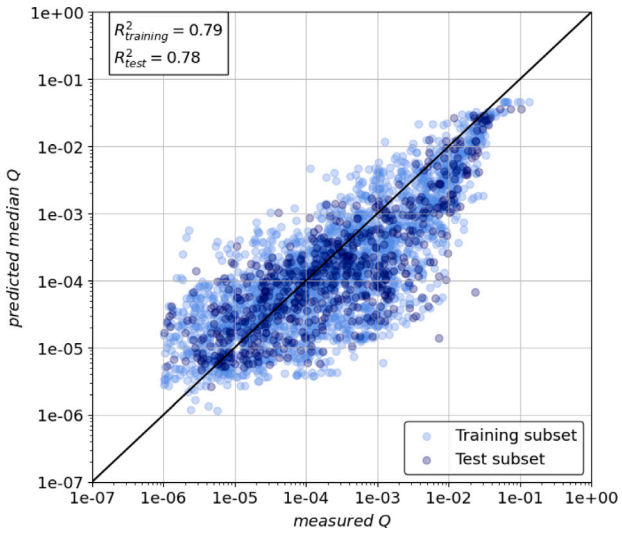

Fig. 6. Comparison between the measured dimensionless mean wave overtopping discharge Q and the predicted median by the BN.

indicate that the proposed model is a reasonable approximation of the observed dependence in the testing dataset and the saturated model.

From the performance perspective, the accuracy of the proposed model is approximated by comparing the observations of Q with the 50% percentile (median) of the predicted distribution of Q given by the model. It should be noted that the aim of the proposed model is not to be used as a deterministic predictor. In order to assess the goodness of fit in a quantitative manner, the coefficient of determination, $0 \leq R^2 \leq 1$, is applied, which is defined as

$$R^2 = 1 - \frac{\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} (o_i - e_i)^2}{\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} (o_i - \overline{o})^2} \tag{9}$$

where N_{obs} is the number of observations, o_i and e_i are the observed and estimated values, respectively, and $\overline{o}$ is the observed mean. Thus, R^2 assesses roughly the percentage of the variance explained by the model.

Fig. 6 presents the comparison between the observed Q and the median of the estimated distribution of Q from the BN for both the training and the test subsets. $R^2 = 0.79$ and 0.78 for the training and test subsets, respectively. Thus, the performance of the BN is similar for both the training and test subsets, indicating that the model does not overfit the training subset. The accuracy of the estimated value of Q is satisfactory with a high value of R^2, considering that the goal of this research is not to develop a deterministic estimator of Q. Appendix C presents a brief comparison of the estimations of deterministic models in the literature in the context the estimated distributions by the developed BN.

Finally, in order to account for the uncertainty of Q, the observations of Q are compared to the 5%, 50% and 95% of the predicted distribution. This comparison is shown for 100 randomly selected observations in Fig. 7. Overall, 85% and 88% of the observations of Q in the training and test subsets fall within the 5% and 95% percentiles, respectively.

6. Example of application

Here, an example is given on how to apply the developed probabilistic model to design the crest level of a mound breakwater. A rock-armored mound breakwater ($\gamma_f = 0.49$ for CLASH NN according to Molines and Medina, 2015) with $cot\alpha = 2.5$ placed on a gentle sea bottom, $m = 50$, faces a storm that can be characterized by $H_{m0} = 5$ m, $T_{-1,0} = 10$ s, and $\beta = 25°$. The structure is placed at $h = 7.5$ m, and presents a $A_c = 7$ m, $G_c = 3$ m, and $B_t = 2$ m. This translates into the following dimensionless explanatory variables: $m = 50$, $\beta = 25°$, $\gamma_f = 0.49$, $s_{-1,0} = 0.032$, $G_c/H_{m0} = 0.6$, $A_c/H_{m0} = 1.4$, $B_t/H_{m0} = 0.4$, $h/H_{m0} = 1.5$ and $cot\alpha = 2.5$. R_c wants to be determined, so different values are considered; $R_c = 9, 10$ and 11 m, leading to $R_c/H_{m0} = 1.8, 2.0$ and 2.2, respectively. The model is conditionalized on the aforementioned values of the explanatory variables, and the conditional distributions of Q, denoted here as $F(Q) = P[Q \leq x|m,\ \beta, \gamma_f,\ R_c/H_{m0}, s_{-1,0},\ G_c/H_{m0}, A_c/H_{m0},\ B_t/H_{m0},\ h/H_{m0}, cot\alpha]$, depicted in Fig. 8, are obtained.

The probability of failure is defined as the probability of exceeding a tolerable Q conditioned to the explanatory variables, $p_f = P[Q > x|m,\ \beta, \gamma_f,\ R_c/H_{m0}, s_{-1,0},\ G_c/H_{m0}, A_c/H_{m0},\ B_t/H_{m0},\ h/H_{m0}, cot\alpha]$. Given a value of the target probability of failure (for illustration purposes, here $p_f = 0.2$ so $F(Q) = 1 - p_f = 0.8$) and a tolerable value of Q (here, $Q < 10^{-4}$), the crest level can be chosen in Fig. 8 to meet the design conditions. In this case, $R_c/H_{m0} = 2.0$ is required. This process can be repeated iteratively changing other design variables. It should also be noted that the model does not need to be conditionalized in all the explanatory variables to obtain a distribution of Q. Thus, if some information on the structure is missing, the model can still provide $F(Q)$, although the predicted uncertainty is expected to be higher.

This section presents an example of the application of the developed model, although there are others. For instance, in the design phase, it can be applied to assess the sensitivity of Q to the geometry of the breakwater given some design constrains (e.g.: design wave loading). That can help to optimize the design and gain insight into the physical interactions between the variables.

7. Discussion and conclusions

In this study, a probabilistic model based on bivariate Gaussian copulas, a Gaussian copula-based Bayesian Network (BN), is proposed to describe the joint probability distribution of Q and a set of explanatory variables on mound breakwaters. The goal of this model is to estimate the distribution of Q conditional to the values of (all or some of) the explanatory variables. In this manner, given a tolerable Q, a probability of failure can be directly computed, and the uncertainties related to the structural response can be incorporated in a probabilistic framework.

To develop the proposed model, a subset of CLASH database is used with a total of 3179 test. 80% of those tests are used for the development of the model while 20% of them are used for testing. Ten dimensionless explanatory variables are selected to describe Q with the following experimental ranges $7.6 \leq m \leq 1000$, $0 \leq \beta \leq 80°$, $0.38 \leq \gamma_f \leq 1.00$, $0 \leq R_c/H_{m0} \leq 4.37$, $1.31 \cdot 10^{-3} \leq s_{-1,0} \leq 0.069$, $0 \leq G_c/H_{m0} \leq 6.67$, $0 \leq A_c/H_{m0} \leq 4.2$, $0 \leq B_t/H_{m0} \leq 15.9$, $1.03 \leq h/H_{m0} \leq 17.6$, and $1.19 \leq cot\alpha \leq 4$. The proposed model has been developed for conventional mound breakwaters and is valid within the experimental ranges of the observations. Therefore, it is encouraged to check its validity outside the experimental ranges of this study. Also, it is recommended to extend the application of this type of probabilistic models to other coastal structures to directly account for the uncertainty of overtopping in the design phase.

The proposed BN is shown in Fig. 4. Empirical cumulative distribution functions are used to quantify the nodes of the BN. One of the main assumptions of the proposed model is the use of Gaussian copula to model the dependence between the random variables. Gaussian copulas are adequate if the shape of the dependence is symmetric, so tail dependence is not significant. Here, the Gaussian copula assumption is validated in statistical terms using Cramer–von–Mises statistic and the training subset. No significant tail dependence was observed for 80% of the variables pairs, so the assumption of the Gaussian copula was deemed reasonable. However, if in light of new data and/or a wider range of application, tail dependence becomes significant in most of the variable pairs, different copula families should be used

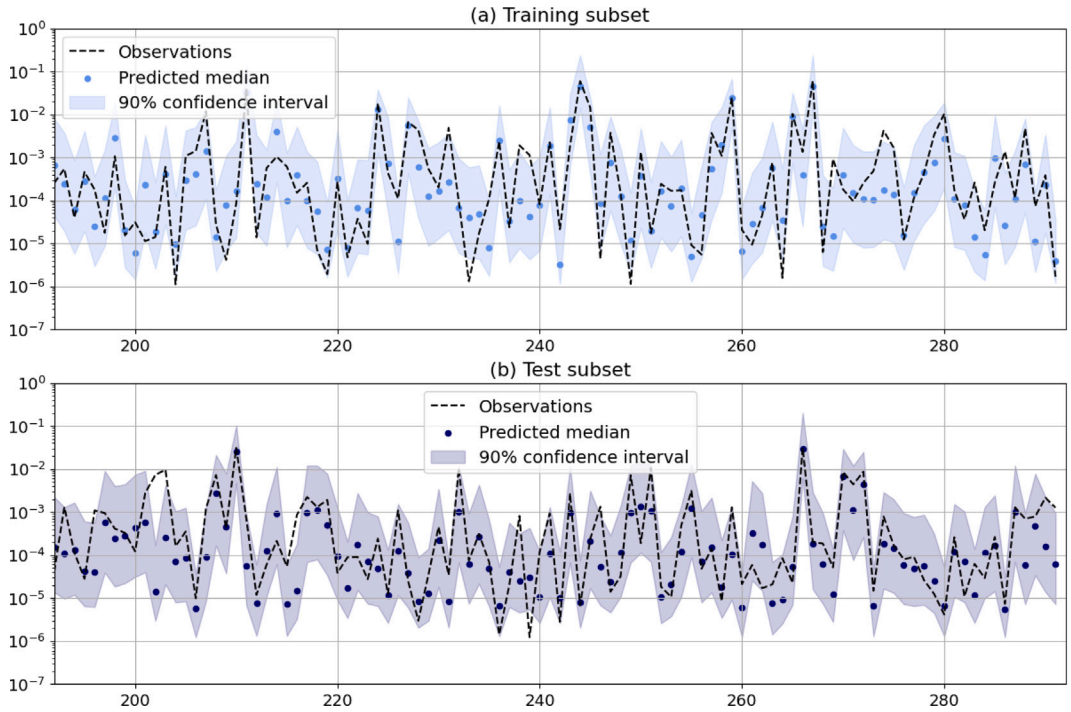

Fig. 7. Comparison of the observed Q and the estimated 5%, 50% and 95% percentiles of Q by the BN.

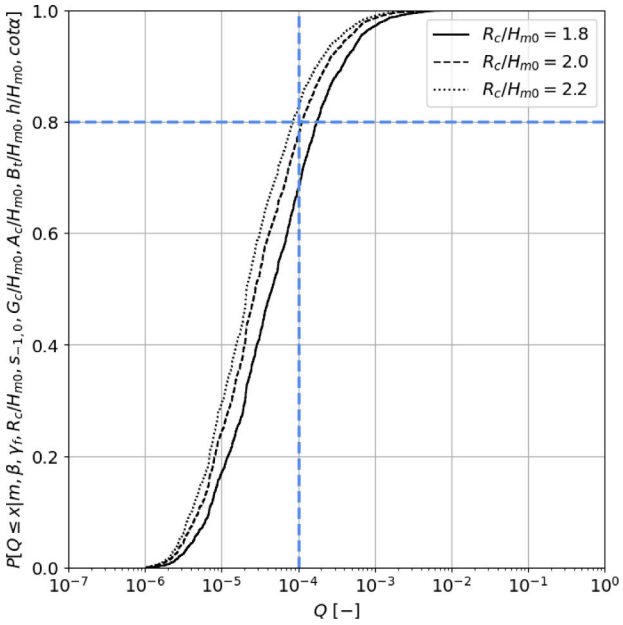

Fig. 8. Example of application of the proposed model to estimate Q.

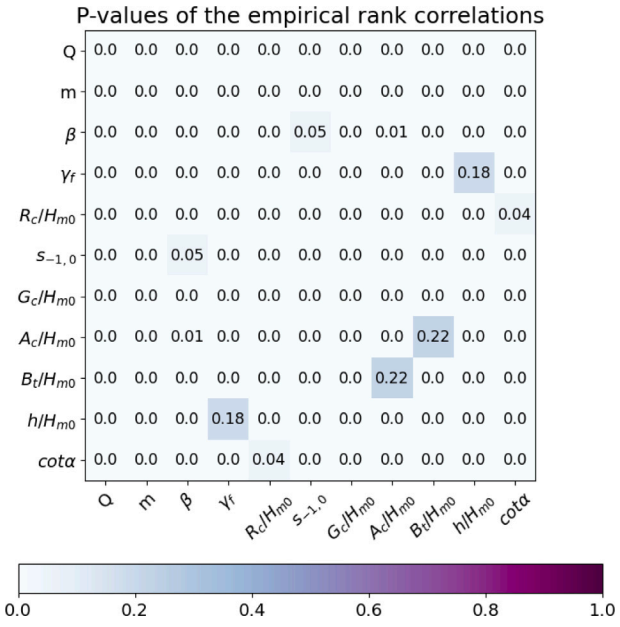

Fig. 9. P-values associated with the empirical rank correlations.

to model the dependence between the variables and more complex models such as vine-copulas can be employed (e.g.: Pouliasis et al., 2021; Mares-Nasarre et al., 2024a).

The proposed model is evaluated using the testing subset in both statistical and performance terms. In statistical terms, the proposed model seems to satisfactorily capture the dependence structure between the studied variables ($d(ER_{test}, BR) = 0.71$, see Section 5.3). In performance terms, the predicted mean of the distribution of Q is compared to the observations obtaining a reasonable performance ($R^2_{test} = 0.78$). Note that the purpose of the proposed model is to estimate the distribution of Q and not to be used as a deterministic estimator. Also, the percentage of the observations that lay within the predicted 90% confidence intervals is quantified, being close to the expected 90%. Overall, the accuracy of the model is adequate, being able to accurately capture the dependence between the studied variables.

Finally, the use of the model for the probabilistic design of the crest elevation of mound breakwaters is also illustrated through one example. It should be noted that the less information provided to the

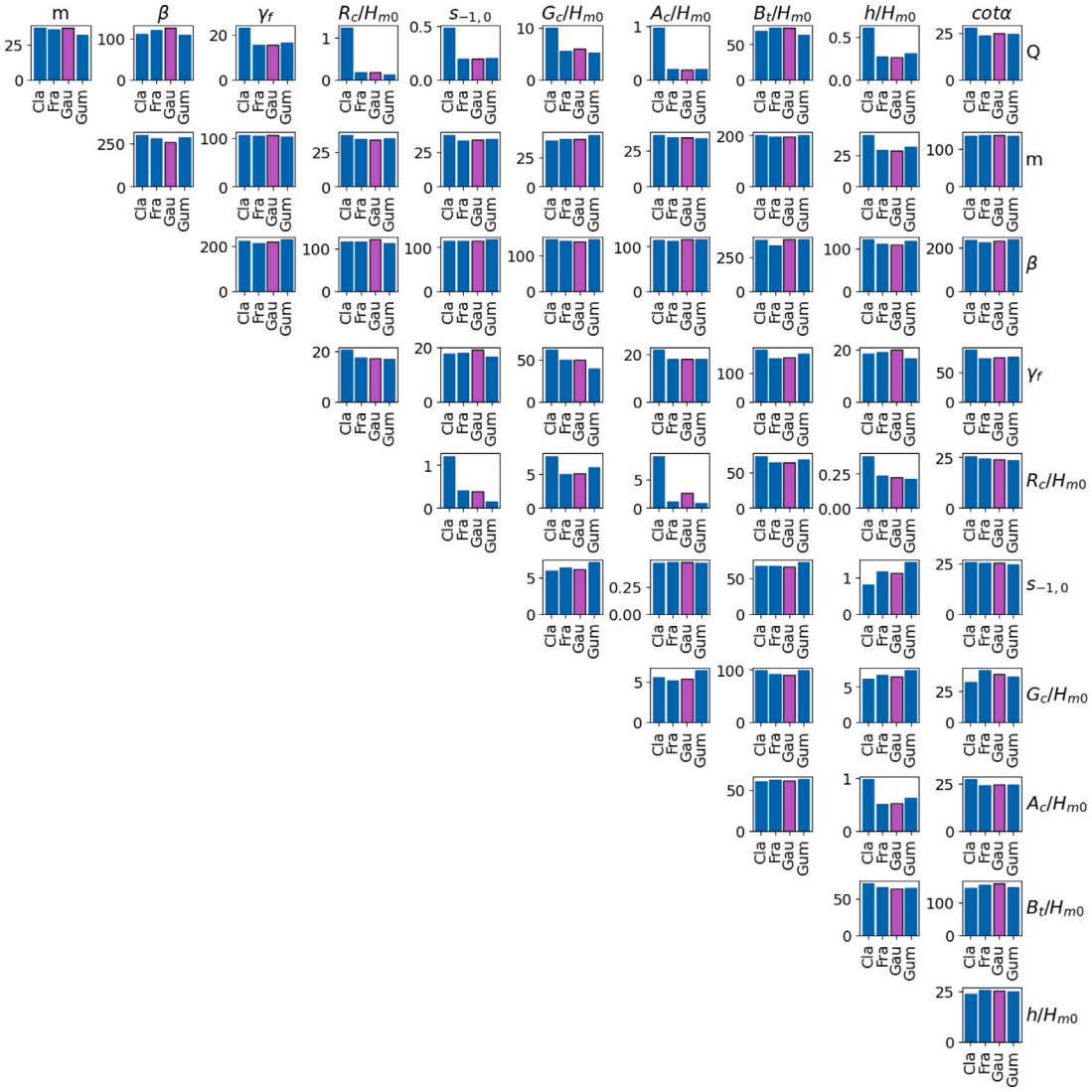

Fig. 10. Cramer von Mises statistic computed between the empirical copula defined with each pair of variables and the Clayton, Frank, Gaussian and Gumbel copula fitted to that pair.

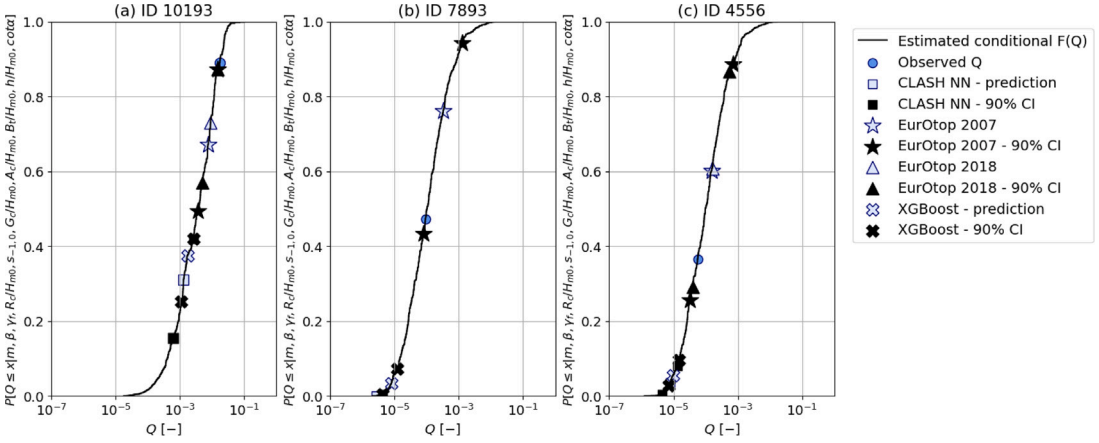

Fig. 11. Conditional distribution of Q for three study cases and the realizations of the estimation and the 90%confidence intervals for a selection of methods in the literature.

model, the wider the estimated distribution of Q as the uncertainty is higher. Consequently, the BN can be used in early stages conditionalizing on a few explanatory variables as an exploration. However, it is recommended to conditionalize as many variables as possible when information is available.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author wants to acknowledge Dr. Oswaldo Morales Nápoles for the conversations about the early stages of this research.

Appendix A. Empirical P-values

Fig. 9 shows the significance of the individual rank correlations computed in the empirical rank correlation ER in Fig. 3. Assuming a significance $\alpha = 0.05$, $P-values < 0.05$ indicate a significant correlation between the two random variables.

Appendix B. Cramer–von–mises statistic

Fig. 10 reports the values of Cramer–von–Mises statistic calculated between the empirical copula defined with each pair of variables and the Clayton, Frank, Gaussian and Gumbel copula fitted to that pair. The lower the value of the statistic, the better model the copula is.

Appendix C. Comparison with methods in the literature

In this section, the predictions of the estimators in the literature for q (EurOtop, 2007; van Gent et al., 2007; EurOtop, 2018; den Bieman et al., 2021) are put into the distribution of q provided by the probabilistic model proposed in this study. To do so, three cases are randomly selected from the testing subset. The three cases are as follows:

- ID 10193: $h = 0.262$ m, $H_{m0} = 0.093$ m, $T_{-1,0} = 2.13$ s, $h_t = 0.262$ m, $B_t = 0$, $h_b = 0$, $B = 0$, $R_c = 0.118$ m, $A_c = 0.083$ m, $G_c = 0$, $cot\alpha_d = cot\alpha_u = 2.7$, $\gamma_f = 1$, $\beta = 0$, $tan\alpha_{bottom} = 0.171$, $Q = 1.78 \cdot 10^{-2}$.
- ID 7893. $h = 0.485$ m, $H_{m0} = 0.112$ m, $T_{-1,0} = 1.228$ s, $h_t = 0.455$ m, $B_t = 0$, $h_b = 0$, $B = 0$, $R_c = 0.13$ m, $A_c = 0.13$ m, $G_c = 0.13$ m, $cot\alpha_d = cot\alpha_u = 2$, $\gamma_f = 0.47$, $\beta = 25°$, $tan\alpha_{bottom} = 0.001$, $Q = 9.3 \cdot 10^{-5}$.
- ID 4556. $h = 0.353$ m, $H_{m0} = 0.145$ m, $T_{-1,0} = 2.19$ s, $h_t = 0.353$ m, $B_t = 0$, $h_b = 0$, $B = 0$, $R_c = 0.4$ m, $A_c = 0.4$ m, $G_c = 0$, $cot\alpha_d = cot\alpha_u = 4$, $\gamma_f = 1$, $\beta = 0$, $tan\alpha_{bottom} = 0.01$, $Q = 5.7 \cdot 10^{-5}$.

Fig. 11 shows the computed conditional distributions of Q for the three selected cases using the proposed model in this study, as well as the predicted Q and confidence intervals given by the aforementioned methods in the literature.

In general, it can be seen that the uncertainty predicted by van Gent et al. (2007) and den Bieman et al. (2021) is much lower than the other methods. As a result, the actual observation of Q can be far from those uncertainty bounds. The confidence interval given by EurOtop (2007) is usually the widest, although still narrower than that provided by the BN. Regarding the uncertainty given by EurOtop (2018), it is usually in between those given by EurOtop (2007) and van Gent et al. (2007) in terms of width. It should be noted that the predictions of EurOtop (2007) are not shown in Fig. 11(b) since they are on the order of 10^{-10}, far from the actual observation.

Data availability

The data and model in this research is publicly available in https://doi.org/10.5281/zenodo.15790655.

References

Altomare, C., Suzuki, T., Chen, X., Verwaest, T., Kortenhaus, A., 2016. Wave overtopping of sea dikes with very shallow foreshores. Coast. Eng. 116, 236–257. http://dx.doi.org/10.1016/j.coastaleng.2016.07.002.

Andersen, T.L., Skals, K., Burcharth, H.F., 2009. Comparison of homogenous and multilayered berm breakwaters with respect to overtopping and front slope stability. In: Proceedings of the 31st International Conference on Coastal Engineering (Hamburg, Germany). World Scientific, pp. 3298–3310. http://dx.doi.org/10.1142/9789814277426_0273.

Antão, E., Guedes Soares, C., 2014. Approximation of bivariate probability density of individual wave steepness and height with copulas. Coast. Eng. 89, 45–52. http://dx.doi.org/10.1016/j.coastaleng.2014.03.009.

Barros, B., Conde, B., Riveiro, B., Morales-Nápoles, O., 2024. Gaussian copula-based Bayesian network approach for characterizing spatial variability in aging steel bridges. Struct. Saf. 106, 102403, doi:j.strusafe.2023.102403.

Besley, P., 1999. Overtopping of Sea-Walls Design and Assessment Manual. Tech. Rep., Environmental Agency, Bristol (UK), Technical Report 178.

Besley, P., Reeves, M., Allsop, N., 1993. Random Wave Physical Model Tests: Overtopping and Reflection Performance. Tech. Rep., Report IT, 384.

Carro, H., Sande, J., Figuero, A., Alvarellos, A., Peña, E., Rabuñal, J., Guerra, A., Pérez, J.D., 2024. Machine learning tool for wave overtopping prediction based on the safety-operability ratio. Ocean Eng. 312, 119006. http://dx.doi.org/10.1016/j.oceaneng.2024.119006.

CLASH database, 2005. Retrieved from https://www.overtopping-manual.com/assets/downloads/Database_20050101.xls.

Clavero, M., Díaz-Carrasco, P., Losada, M.Á., 2020. Bulk wave dissipation in the armor layer of slope rock and cube armored breakwaters. J. Mar. Sci. Eng. 8 (3), http://dx.doi.org/10.3390/jmse8030152.

Czado, C., 2019. Analyzing dependent data with vine copulas. Lect. Notes Stat. Springer 222, http://dx.doi.org/10.1007/978-3-030-13785-4_1.

den Bieman, J.P., van Gent, M.R., van den Boogaard, H.F., 2021. Wave overtopping predictions using an advanced machine learning technique. Coast. Eng. 166, 103830. http://dx.doi.org/10.1016/j.coastaleng.2020.103830.

Díaz-Carrasco, P., 2023. Hydraulic performance analysis for homogeneous mound breakwaters: Application of dimensional analysis and a new experimental technique. Ocean Eng. 286, 115598. http://dx.doi.org/10.1016/j.oceaneng.2023.115598.

Díaz-Carrasco, P., Moragues, M.V., Clavero, M., Losada, M.Á., 2020. 2D water-wave interaction with permeable and impermeable slopes: Dimensional analysis and experimental overview. Coast. Eng. 158, 103682. http://dx.doi.org/10.1016/j.coastaleng.2020.103682.

EurOtop, 2007. European overtopping manual-wave overtopping of sea defenses and related structures: assessment manual.. Retrieved from www.overtopping-manual.com.

EurOtop, 2018. Manual on wave overtopping of sea defences and related structures. An overtopping manual largely based on European research, but for worldwide application.. Retrieved from www.overtopping-manual.com.

Galland, J.-C., 1994. Rubble mound breakwater stability under oblique waves: an experimental study. In: Proceedings of the 24th International Conference on Coastal Engineering, Kobe (Japan). pp. 1061–1074.

Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. Insurance Math. Econom. 44 (2), 199–213. http://dx.doi.org/10.1016/j.insmatheco.2007.10.005.

van Gent, M.R.A., 2001. Wave runup on dikes with shallow foreshores. J. Waterw. Port, Coast. Ocean. Eng. 127 (5), 254–262. http://dx.doi.org/10.1061/(ASCE)0733-950X(2001)127:5(254).

Hanea, A.M., Kurowicka, D., Cooke, R.M., 2006. Hybrid method for quantifying and analyzing Bayesian belief nets. Qual. Reliab. Eng. Int. 22 (6), 709–729.

Hanea, A.M., Kurowicka, D., Cooke, R.M., Ababei, D.A., 2010. Mining and visualising ordinal data with non-parametric continuous BBNs. Comput. Statist. Data Anal. 54 (3), 668–687.

Hanea, A.M., Morales Nápoles, O., Ababei, D.A., 2015. Non-parametric Bayesian networks: Improving theory and reviewing applications. Reliab. Eng. Syst. Saf. 144, 265–284. http://dx.doi.org/10.1016/j.ress.2015.07.027.

Haralambides, H., 2017. Globalization, public sector reform, and the role of ports in international supply chains. Marit. Econ. Logist. 19, 1–51. http://dx.doi.org/10.1057/s41278-017-0068-6.

Irías Mata, M., van Gent, M.R., 2023. Numerical modelling of wave overtopping discharges at rubble mound breakwaters using openfoam®. Coast. Eng. 181, 104274. http://dx.doi.org/10.1016/j.coastaleng.2022.104274.

Jaeger, W.S., Morales-Nápoles, O., 2017. A vine-copula model for time series of significant wave heights and mean zero-crossing periods in the north sea. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A: Civ. Eng. 3, 04017014.

Joe, H., 1997. Multivariate Models and Multivariate Dependence Concepts. CRC Press, http://dx.doi.org/10.1201/9780367803896.

Koosheh, A., Etemad-Shahidi, A., Cartwright, N., Tomlinson, R., van Gent, M.R.A., 2022. Experimental study of wave overtopping at rubble mound seawalls. Coast. Eng. 172, 104062. http://dx.doi.org/10.1016/j.coastaleng.2021.104062, Retrieved from https://www.sciencedirect.com/science/article/pii/S0378383921002003.

Koot, P., Mendoza-Lugo, M.A., Paprotny, D., Morales-Nápoles, O., Ragno, E., Worm, D.T., 2023. Pybanshee version (1.0): A python implementation of the MATLAB toolbox BANSHEE for non-parametric Bayesian networks with updated features. SoftwareX 21, 101279. http://dx.doi.org/10.1016/j.softx.2022.101279, Retrieved from https://www.sciencedirect.com/science/article/pii/S2352711022001972.

Kurowicka, D., Cooke, R., 2004. Non-parametric continuous Bayesian belief nets with expert judgement. In: Probabilistic Safety Assessment and Management: PSAM 7—ESREL'04 June 14–18, 2004, Berlin, Germany, Volume 6. Springer, pp. 2784–2790.

Leontaris, G., Morales-Nápoles, O., Wolfert, A., 2016. Probabilistic scheduling of offshore operations using copula based environmental time series – an application for cable installation management for offshore wind farms. Ocean Eng. 125, 328–341. http://dx.doi.org/10.1016/j.oceaneng.2016.08.029.

Lucio, D., Lara, J.L., Tomás, A., Losada, I.J., 2024. Probabilistic assessment of climate-related impacts and risks in ports. Reliab. Eng. Syst. Saf. 251, 110333. http://dx.doi.org/10.1016/j.ress.2024.110333.

Lucio, D., Tomás, A., Lara, J., Camus, P., Losada, I., 2020. Stochastic modeling of long-term wave climate based on weather patterns for coastal structures applications. Coast. Eng. 161, 103771. http://dx.doi.org/10.1016/j.coastaleng.2020.103771.

Lykke Andersen, T., Burcharth, H., 2009. Three-dimensional investigations of wave overtopping on rubble mound structures. Coast. Eng. 56 (2), 180–189. http://dx.doi.org/10.1016/j.coastaleng.2008.03.007, The CLASH Project.

Mares-Nasarre, P., García-Maribona, J., Mendoza-Lugo, M.A., Morales-Nápoles, O., 2023. A copula-based Bayesian network to model wave climate multivariate uncertainty in the alboran sea. In: 33rd European Safety and Reliability Conference. pp. 1053–1060.

Mares-Nasarre, P., Molines, J., Gómez-Martín, M.E., Medina, J.R., 2020. Individual wave overtopping volumes on mound breakwaters in breaking wave conditions and gentle sea bottoms. Coast. Eng. 159, 103703. http://dx.doi.org/10.1016/j.coastaleng.2020.103703.

Mares-Nasarre, P., Muscalus, A., Haas, K., Morales-Nápoles, O., 2024a. The probabilistic dependence of ship-induced waves is preserved spatially and temporally in the savannah river (USA). Sci. Rep. 14, 28154. http://dx.doi.org/10.1038/s41598-024-78924-z.

Mares-Nasarre, P., van Gent, M.R., Morales-Nápoles, O., 2024b. A copula-based model to describe the uncertainty of overtopping variables on mound breakwaters. Coast. Eng. 189, 104483. http://dx.doi.org/10.1016/j.coastaleng.2024.104483.

Medina, J.R., Molines, J., Gómez-Martín, M.E., 2014. Influence of armour porosity on the hydraulic stability of cube armour layers. Ocean Eng. 88, 289–297. http://dx.doi.org/10.1016/j.oceaneng.2014.06.012.

Mendoza-Lugo, M.A., Morales-Nápoles, O., Delgado-Hernández, D.J., 2022. A non-parametric Bayesian network for multivariate probabilistic modelling of weigh-in-motion system data. Transp. Res. Interdiscip. Perspect. 13, 100552. http://dx.doi.org/10.1016/j.trip.2022.100552.

Molines, J., Bayon, A., Gómez-Martín, M.E., Medina, J.R., 2019. Influence of parapets on wave overtopping on mound breakwaters with crown walls. Sustain. 11 (24), http://dx.doi.org/10.3390/su11247109.

Molines, J., Medina, J.R., 2015. Calibration of overtopping roughness factors for concrete armor units in non-breaking conditions using the CLASH database. Coast. Eng. 96, 62–70. http://dx.doi.org/10.1016/j.coastaleng.2014.11.008.

Molines, J., Medina, J.R., 2016. Explicit wave-overtopping formula for mound breakwaters with crown walls using CLASH neural network–derived data. J. Waterw. Port, Coast. Ocean. Eng. 142 (3), 04015024. http://dx.doi.org/10.1061/(ASCE)WW.1943-5460.0000322.

Morales-Nápoles, O., Delgado-Hernández, D.J., De-León-Escobedo, D., Arteaga-Arcos, J.C., 2014. A continuous Bayesian network for earth dams' risk assessment: methodology and quantification. Struct. Infrastruct. Eng. 10 (5), 589–603. http://dx.doi.org/10.1080/15732479.2012.757789.

Morales-Nápoles, O., Hanea, A.M., Worm, D., 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, http://dx.doi.org/10.1201/b15938-205.

Nelsen, R.B., 2006. An Introduction to Copulas (Springer Series in Statistics). Springer-Verlag, Berlin, Heidelberg, http://dx.doi.org/10.5555/1204326.

Oumeraci, H., Kortenhaus, A., Burg, S., 2007. Investigations of Wave Loading and Overtopping of an Innovative Mobile Flood Defence System: Analysis of Model Tests and Design Formulae. Tech. Rep. No. 949, LeichtweiB-Institut fur Wasserbau, Technische Universitat Braunschweig.

Owen, M.W., 1980. Design of Seawalls Allowing for Wave Overtopping. Tech. Rep., HR-Wallingford. Technical Report EX-924, UK..

Paprotny, D., Morales-Nápoles, O., Worm, D.T., Ragno, E., 2020. BANSHEE–a MATLAB toolbox for non-parametric Bayesian networks. SoftwareX 12, 100588. http://dx.doi.org/10.1016/j.softx.2020.100588, Retrieved from https://www.sciencedirect.com/science/article/pii/S2352711020303010.

Pearl, J., 2013. A constraint propagation approach to probabilistic reasoning. http://dx.doi.org/10.48550/ARXIV.1304.3422, Retrieved from https://arxiv.org/abs/1304.3422.

Pearson, K., Galton, F., 1895. VII. Note on regression and inheritance in the case of two parents. Proc. R. Soc. Lond. 58 (347–352), 240–242. http://dx.doi.org/10.1098/rspl.1895.0041.

Pepi, Y., Romano, A., Franco, L., 2022. Wave overtopping at rubble mound breakwaters: A new method to estimate roughness factor for rock armours under non-breaking waves. Coast. Eng. 178, 104197. http://dx.doi.org/10.1016/j.coastaleng.2022.104197.

Pouliasis, G., Torres-Alves, G.A., Morales-Nápoles, O., 2021. Stochastic modeling of hydroclimatic processes using vine copulas. Water 13 (16), http://dx.doi.org/10.3390/w13162156.

Spanish recommendations for maritime structures. General procedure and requirements in the design of harbor and maritime structures. Part 1. 2001. Retrieved from https://www.puertos.es/sites/default/files/2024-04/ROM%200.0-01%20%28EN%29.pdf.

Spanish recommendations for the project design and construction of breakwaters. Part I: Calculation and project factors. Climate agents. 2009. Retrieved from https://www.puertos.es/sites/default/files/2024-04/ROM%201.0-09%20%28EN%29.pdf.

Romano, A., Bellotti, G., Briganti, R., Franco, L., 2015. Uncertainties in the physical modelling of the wave overtopping over a rubble mound breakwater: The role of the seeding number and of the test duration. Coast. Eng. 103, 15–21. http://dx.doi.org/10.1016/j.coastaleng.2015.05.005.

Sibuya, M., et al., 1960. Bivariate extreme statistics. Ann. Inst. Statist. Math. 11 (2), 195–210. http://dx.doi.org/10.1007/BF01682329.

Sklar, M., 1959. Fonctions de repartition a n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris 8, 229–231, Retrieved from https://cir.nii.ac.jp/crid/1573387449735953792.

Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15 (1), 72–101.

Steendam, G.J., van der Meer, J.W., Verhaeghe, H., Besley, P., Franco, L., van Gent, M.R.A., 2005. The international database on wave overtopping. In: Coastal Engineering 2004: (in 4 Volumes). World Scientific, pp. 4301–4313.

TAW, 2002. Wave Run-Up and Wave Overtopping at Dikes. Tech. Rep., Technical Report.

Torres-Alves, G.A., Morales-Nápoles, O., 2020. Reliability analysis of flood defenses: The case of the nezahualcoyotl dike in the aztec city of tenochtitlan. Reliab. Eng. Syst. Saf. 203, 107057. http://dx.doi.org/10.1016/j.ress.2020.107057.

van der Meer, J.W., Janssen, J.P.F.M., 1994. Wave Run-Up and Wave Overtopping at Dikes. Tech. Rep., Delft Hydraulics. Technical Report no. 485.

Van Doorslaer, K., De Rouck, J., Audenaert, S., Duquet, V., 2015. Crest modifications to reduce wave overtopping of non-breaking waves over a smooth dike slope. Coast. Eng. 101, 69–88. http://dx.doi.org/10.1016/j.coastaleng.2015.02.004.

van Gent, M.R.A., 1999. Physical model investigations on coastal structures with shallow foreshores: 2D model tests with single and double-peaked wave energy spectra. Tech. Rep..

van Gent, M.R.A., 2014. Oblique wave attack on rubble mound breakwaters. Coast. Eng. 88, 43–54. http://dx.doi.org/10.1016/j.coastaleng.2014.02.002.

van Gent, M.R.A., 2022. Wave overtopping at dikes and breakwaters under oblique wave attack. In: Proceedings of the 37th International Conference on Coastal Engineering (Sydney, Australia).

van Gent, M.R.A., van den Boogaard, H.F.P., Pozueta, B., Medina, J.R., 2007. Neural network modelling of wave overtopping at coastal structures. Coast. Eng. 54 (8), 586–593. http://dx.doi.org/10.1016/j.coastaleng.2006.12.001.

van Gent, M.R.A., Wolters, G., Capel, A., 2022. Wave overtopping discharges at rubble mound breakwaters including effects of a crest wall and a berm. Coast. Eng. 176, 104151. http://dx.doi.org/10.1016/j.coastaleng.2022.104151.

Verschuur, J., Koks, E.E., Hall, J.W., 2023. Systemic risks from climate-related disruptions at ports. Nat. Clim. Chang. 13, 804–806. http://dx.doi.org/10.1038/s41558-023-01754-w.

Victor, L., Troch, P., 2012. Wave overtopping at smooth impermeable steep slopes with low crest freeboards. J. Waterw. Port, Coast. Ocean. Eng. 138 (5), 372–385. http://dx.doi.org/10.1061/(ASCE)WW.1943-5460.0000141.

Yu, Y.-X., Liu, S.-X., Zhu, C.-H., 2002. Stability of armour units on rubble mound breakwater under multi-directional waves. Coast. Eng. J. 44 (2), 179–201. http://dx.doi.org/10.1142/S0578563402000482.

Zanuttigh, B., Formentin, S.M., van der Meer, J.W., 2016. Prediction of extreme and tolerable wave overtopping discharges through an advanced neural network. Ocean Eng. 127, 7–22. http://dx.doi.org/10.1016/j.oceaneng.2016.09.032.