
A PRACTICAL ASSESSMENT OF THE IMPACT OF USING MULTIVARIATE
STATISTICAL MODELS IN THE DESIGN OF COASTAL INFRASTRUCTURE
Master Thesis

HYDRAULIC ENGINEERING
Specialization Flood Risk



A PRACTICAL ASSESSMENT OF THE IMPACT OF USING MULTIVARIATE
STATISTICAL MODELS IN THE DESIGN OF COASTAL INFRASTRUCTURE

Master Thesis

HYDRAULIC ENGINEERING
Specialization Flood Risk

Author:

M. S. Timmermans

Studentnumber:

4220226

Thesis committee:

Dr. ir. O. Morales Napoles

Dr. ir. R.C. Lanzafame

Dr. ir. A. Antonini

Ir. A. Lioutas

Ir. M. Benit

Organisation:

TU Delft

TU Delft

TU Delft

Van Oord

Arcadis

Delft University of Technology
Faculty of Civil Engineering & Geosciences



June 11, 2021

ABSTRACT

Offshore and coastal infrastructure must be designed to withstand loading conditions that among others arise from extreme environmental conditions. Physical processes such as storm surges, tides, currents, and waves play an important role in the design of these structures. Several variables characterize the relevant physical processes (e.g. significant wave height, mean wave period, water level, wind speed) and a thorough analysis of these variables is required when dealing with offshore and coastal dynamics, durability, and reliability assessment.

Traditionally, a critical loading condition is defined by characteristic values of environmental variables that are determined based on the highest loads previously experienced. Modern design methods seek to derive loads that correspond to specified reliability by considering the frequency of a specific loading magnitude. Traditional design approaches do not take into account the interrelations and dependencies among the variables of interest. Hence, wrong representations of the physical processes and unnecessary conservative representations of the design loads might occur. This may severely limit their effectiveness and can lead to expensive and inappropriate decisions. Multivariate frequency analysis approaches currently receive much attention within the academic community, however, advanced statistical concepts such as regular vine copula are slow in being taken up by engineering practice.

This thesis presents a practical assessment and further development of a vine-based methodology, used for the derivation of design values, in continuation of the work performed by [Sellés Valls \(2019\)](#). Regular vine copulae are advanced statistical models for high dimensional distributions using (conditional) bivariate copulae as building blocks. This study contributes to bridging the gap between the academic community and engineering practice on one hand, and on the other hand, contributes to a better understanding of the potential added value of incorporating dependence information in the design process of coastal and offshore infrastructure. It has a conceptual point of view where the concept of using dependence information by applying advanced statistical techniques is explored and the required adaptations throughout the entire design process are evaluated.

In this research, it is found that the multivariate vine-based methodology can be successfully incorporated in the design process of a breakwater structure, and on average results in minimal required dimensions of elements of the cross-sectional design that turn out to be smaller and the corresponding costs up to 25% lower compared to the univariate traditional approach. This is realized by adapting the framework enabling an offshore-nearshore transformation of the wave conditions using SWAN software. Furthermore, the theoretical framework is extended by introducing Kendall's measure providing a suitable definition of the critical region from which the critical loading conditions can be obtained. It is concluded that the vine-based approach could act as a tool providing extra information about the behavior of the system and insights on the degree of conservatism of the traditional approach. The considered role of the vine-based methodology in the design process of a breakwater structure (or coastal infrastructure in general) is to provide the practitioner with additional insights supporting the traditional design approach and possibly optimizing the design.

PREFACE

This master thesis marks the end of my time as a student at Delft University of Technology. It fulfills the final requirement in order to obtain the degree of Master of Science in Hydraulic Engineering, with a specialization in Flood Risk and Ports & Waterways at the Delft University of Technology.

I am grateful for the opportunity to perform my graduation in close cooperation with people from both Arcadis and Van Oord. I really appreciate that I was allowed to join their weekly meetings to get a sense of their working environment and getting inspired by the interesting projects they had going on.

I would like to acknowledge several people for their contribution to this thesis. First off, I would like to thank my committee members. I want to thank my daily supervisor dr. ir. Robert Lanzafame for the frequent meetings during which we had many meaningful conversations about the thesis subject as well as non-related topics which I really valued. I want to thank ir. Matthijs Benit for his endless support, always constructive and extensive feedback on the report, and lastly, his not to be underestimated entertaining input during "het koffiemoment". I want to thank dr. ir. Alessandro Antonini for joining the graduation committee during the last phase of the project and for providing me with useful information regarding the breakwater design. I would like to thank ir. Anestis Lioutas for his great advice and for sharing his experience linking the theory to practice. And lastly, my expressions of gratitude go out to dr. ir. Oswaldo Morales Napoles for chairing my MSc. thesis committee and providing me with in-depth feedback and constructive arguments during the meetings which definitely improved the quality of my work.

Finally, I want to thank my parents and my brother for supporting me throughout the years. Especially, my parents, you have always motivated me to achieve to the best of my abilities and provided me with every opportunity not only during my time in Delft but throughout my life. I realize that not everyone is privileged to obtain a second master's degree and without you, it would not have been possible.

Contents

1	Introduction	1
1.1	Context of the Problem	1
1.2	Objectives & Research Question	3
1.3	Thesis Outline	4
2	Theoretical Background	6
2.1	Reliability Analysis	6
2.2	Extreme Value Theory	10
2.3	Copulae	13
2.4	Vines	17
2.5	Return Period & Design Values	22
3	Case Study: Breakwater Design	26
3.1	Background	26
3.2	Breakwater Design Process	28
3.3	Data Collection	29
4	Conventional Approach	34
4.1	Extreme Value Selection	35
4.2	Distribution Modeling	37
4.3	Design Values	39
4.4	Final Design	41
5	Vine-based Approach	42
5.1	Joint Extreme Selection	42
5.2	Joint Distribution Modelling	44
5.3	Design Values	46
5.4	Final Design	52
6	Results	54
6.1	Comparison of the Conventional & Vine-based Approach	54
6.2	Experiments	59
6.3	Concluding Remarks	69
7	Sensitivity Analysis	70
7.1	Data Space	71
7.2	Model Space	72
7.3	Parameter Space	75

8 Discussion	79
8.1 Relevance	79
8.2 Case Study	81
8.3 Limitations & Challenges	81
8.4 Applicability	83
8.5 Results	83
9 Conclusions & Recommendations	85
9.1 Conclusions	85
9.2 Recommendations	87
References	89
A Basics of Probability Theory & Statistics	95
B Failure Mechanisms	99
C Design Input	103
D Peak Over Threshold Procedure	108
E Offshore Conditions Experiment 1.0	112
F Nearshore Conditions Experiment 1.0	113
G Results Experiment 1.0	115
H Bivariate Scatterplots	117
I Offshore Conditions All Experiments	129
J Nearshore Conditions All Experiments	131
K Water Level Time Series	133

1 Introduction

1.1 Context of the Problem

Offshore and coastal infrastructure such as breakwaters, jetties, pipelines, or berthing dolphins must be designed to withstand loading conditions that among others arise from extreme environmental conditions. Physical processes such as storm surges, tides, currents, and waves play an important role in the design of hydraulic structures. In general, several variables characterize the relevant physical processes: for instance, the significant wave height, mean wave period, water level, and wind speed are considered key variables when dealing with offshore and coastal dynamics, durability, and reliability assessment. The design of this type of infrastructure requires a thorough analysis of these environmental variables, often referred to as *design variables*.

The design of offshore and coastal structures is based on the evaluation of failure mechanisms such as overtopping, geotechnical instability, and wave attack among others, which are driven by certain physical processes. A combined action of these physical processes results in critical load conditions which are referred to as *design events* or *storm events* since these extreme events typically occur during a storm. The design event representing a specific critical load combination is used to determine the required geometry, size, and material of each element of the structural design.

Traditionally, a design event consists of characteristic values of environmental variables that are determined based on the highest loads previously experienced. Modern design methods seek to derive loads that correspond to specified reliability by considering the frequency of a specific loading magnitude. The design of objects in the coastal zone is usually done by means of the limit state design method, where the reliability of the design is assessed by the evaluation of the loads and resistance for different modes of failure (failure mechanisms). A so-called limit state function, Z , can be formulated in order to determine whether a failure occurs. The limit state function is defined as:

$$Z = R - S \tag{1.1}$$

where R is the load resistance and S is the load. In case $Z < 0$ a failure occurs. The probability of failure P_f is defined as $P_f = P(Z < 0)$. The reliability calculation requires the quantification of design loads which is commonly done by readily available methods using univariate probabilistic approaches neglecting possible interdependence. From now on, these approaches are referred to as the *traditional* or *conventional approach*. The conventional approach does not take into account the interrelations and dependencies among the variables of interest, which means a loss of potentially useful information regarding the joint occurrence of critical combinations of design variables that may coexist during a single storm. Hence, wrong representations of the physical processes and unnecessary conservative representations of the design loads might occur. The lack of knowledge concerning their joint statistics may severely limit the effectiveness of offshore and coastal structures and can lead to expensive and inappropriate decisions (Li et al., 2008).

HYDRA-NL is a software program developed and extensively used in the Netherlands for the safety assessment of dikes (and flood defences in general) that is based on a probabilistic method considering joint statistics of relevant hydraulic loads (Hydra-NL, 2020). It performs a full probabilistic analysis of the limit state function deriving the failure probability of the system. The usage of joint statistics in the design of flood defences in the Netherlands is considered to be unique compared to other countries.

Over the past decade, the applicability of copulae has been exploited for modelling the joint occurrence of combined conditions: Salvadori and Michele (2004), Grimaldi and Serinaldi (2006), De Michele et al. (2007), and Salvadori and Michele (2010). A copula is a function which couples multivariate distribution functions to their one dimensional marginal distribution functions (Nelsen, 2007), which is a useful method for characterization of the dependence and accuracy of probabilistic models. In particular, bivariate copulae became very popular because they have better capabilities of modeling different dependence structures compared to copulae with higher dimensionality. In contrast to bivariate copulae, multivariate copulae are often limited in the variety of dependence structures they can reassemble. In turn, regular vine copula is a method to model high dimensional distributions using (conditional) bivariate copulae as building blocks and have proven to be a flexible tool in high-dimensional dependence modeling. This methodology originates from Joe (1996) and is investigated and extended in more detail by Bedford and Cooke (2001; 2002).

The concept of return period in civil engineering is important since it is used for designing and sizing reliable offshore and coastal structures, deriving critical (dangerous) events, risk assessment, and for rational decision making (Singh et al., 2007). Univariate frequency analysis used to determine the failure region given a return period, is rather straightforward and has been carried out both in coastal and offshore applications (Haver, 1985; Krogstad, 1985; Kuwashima and Hogben, 1986; Goda, 1988; Ferreira and Guedes Soares, 2000). Multivariate frequency analysis became more and more widespread over the past decade and multiple studies have been devoted to generalization of the univariate case: see Gräler et al. (2013; 2016), and references therein for a thorough review and comparison of procedures. The problem of identifying design events in a multivariate context is troublesome, but at the same time, of fundamental importance in the successful application of a multivariate approach in the design process. In other words, given a prescribed reliability, what joint realizations of the design variables (significant wave height, wind speed, etc.) are considered to be a design event? To this end, a key point of focus is the shortcomings of the return period that emerge in a multivariate context (Serinaldi, 2015).

Although multivariate frequency analysis currently receives much attention within the academic community, advanced statistical techniques such as regular vine copulae are slow in being taken up in engineering practice. Reasons being the costs of obtaining sufficient data for more complex methods; the costs of staff training for advanced techniques that are not common practice, and lastly, the widespread and straightforward application of simple univariate methods combined with the engineering judgment (Li et al., 2008). Nevertheless, there is a growing need amongst engineers, researchers, and practitioners, to be able to quantify the uncertainty associated with multivariate design conditions. The underlying data structure is of great value when deriving design events for the design of offshore and coastal structures because it provides more knowledge about and confidence in the system behavior, and can save money by avoiding an overly conservative design.

The research presented in this thesis, therefore, aims to provide a practical assessment and further development of a vine-based methodology, used for the derivation of design values, in continuation of the work performed by [Sellés Valls \(2019\)](#). The impact and potential added value of the multivariate vine-based approach are evaluated by comparison with the traditional approach for a specific case study.

1.2 Objectives & Research Question

The knowledge gaps presented in the previous section, motivate the initiation of this research, which is covered by the following main research objective:

Perform a practical assessment to investigate how the vine-based methodology can be incorporated in the design of coastal infrastructure and how the usage of dependence information can contribute to additional insights in comparison to the traditional univariate approach assuming independence.

In order to achieve this research objective, the following main research question is formulated:

How can the vine-based methodology using dependence information be incorporated in the design process of a coastal engineering application and how can it provide extra information in contrast to the traditional univariate approach assuming independence?

To further explicate this main research question, the main research question is split and two research sub-questions are formulated. These are stated below, accompanied by their research objective(s).

RQ1. *How can the vine-based methodology using dependence information be incorporated in the design process of a coastal engineering application?*

- (a) Select, modify and perform a case study that is considered representative with respect to relevant physical processes in the field of offshore and coastal engineering,
- (b) Study relevant failure mechanisms of chosen structure and relevant environmental variables,
- (c) Implement and validate the entire design procedure including both the traditional and vine-based approach as described by [Sellés Valls \(2019\)](#),

RQ2. *How can the vine-based methodology incorporated in the design process of a coastal engineering application provide extra information in contrast to the traditional univariate approach?*

- (a) Apply both the traditional and multivariate design approach for the selected case study, and evaluate the differences in the final design, in terms of design dimensions and costs, to quantify the impact of incorporating dependence information.
- (b) Evaluate the possibilities based on existing literature and find a way to cope with the problem of design event identification and deal with the shortcomings of the return period in the multivariate context of the vine-based approach,
- (c) Define and perform computational experiments to provide insights and recommendations about the influence of varying vine copula model configurations on the final results,
- (d) Perform a sensitivity analysis to identify aspects of the design procedure that may influence the reliability of the vine-based methodology.

1.3 Thesis Outline

This thesis is structured in different parts that are graphically presented by Figure 1.1. The left-hand side depicts a flow diagram consisting of steps in a generic design process of an offshore/coastal structure. These steps are elaborated in different chapters of the research which are linked on the right-hand side of the figure.

The focus of this thesis is on a multivariate vine-based approach for the derivation of design values (yellow section). The new approach compared to the traditional approach is illustrated in Chapter 4 and 5 respectively. Prior to this, Chapter 2 elaborates on the relevant theoretical concepts and Chapter 3 introduces the case study. Note, Chapter 3 does not present a completely worked out case resulting in a final cross-sectional design, but rather all required information to evaluate the case study except derivation of design values. The final results for both design approaches are compared in Chapter 6 including several computational experiments on the model performance for different model configurations. The sensitivity of the multivariate design approach is discussed in Chapter 7, followed by the discussion in Chapter 8, and finally, the conclusions and recommendations are given in Chapter 9.

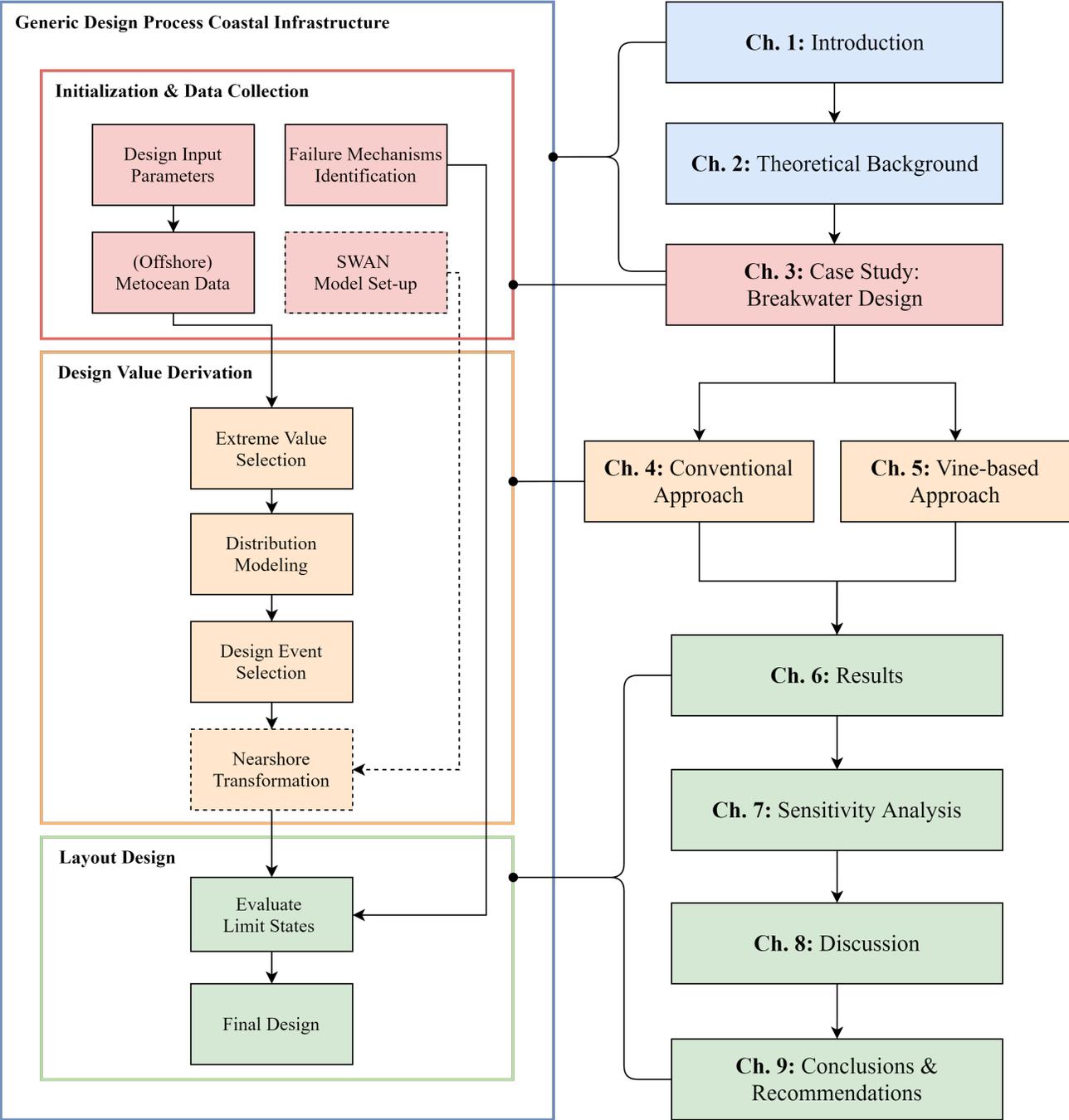


Figure 1.1: The outline of this thesis.

2 Theoretical Background

In this chapter, relevant literature on different subjects related to the research question will be discussed and evaluated. In Section 2.2, a branch of statistics is evaluated regarding the appropriate methods for selection and modeling of extreme events. In Section 2.3, literature is reviewed concerning the concept of copulae. Relevant studies concerning the graphical structures called vines are outlined in Section 2.4. Finally, the concept of return period and in particular the multivariate case is evaluated in Section 2.5. It is assumed that basic concepts from the field of probability theory and statistics are familiar for the reader otherwise it is recommended to read the contents of Appendix A.

2.1 Reliability Analysis

A structure should be designed such that it fulfills the functional requirements. When designing a structure, one aims to come up with a safe and reliable design. The reliability of a structure indicates the likelihood of a system failure during its lifetime. There is a possibility that during its lifetime the structure fails to fulfill the functional requirements due to certain circumstances. A reliability analysis requires an evaluation of the system and all possible failure modes as well as the determination of the acceptable safety level.

2.1.1 Basic Principles

The reliability depends on the margin between the resistance of a structure and the load applied to the structure. The reliability function (or limit state function) Z describes this relationship between the resistance R and the load S :

$$Z = R - S$$

where both the resistance and load are dependent on stochastic variables. Z represents the state of the structure which can have three different states:

$$\text{state} = \begin{cases} \text{failure} & \text{if } Z < 0 \\ \text{limit state} & \text{if } Z = 0 \\ \text{no failure} & \text{if } Z > 0 \end{cases}$$

The reliability can be expressed in terms of the probability of failure P_f :

$$P(Z \geq 0) = 1 - P_f = 1 - P(Z < 0).$$

The design of a structure should guarantee a certain level of safety. Risk reduction and increasing reliability can improve the safety of the design. Note, that it does not hold that a reliable structure also means a safe structure. According to [Mulazzani \(1985\)](#) the definition of reliability and safety is given by:

Definition 1 (Reliability)

The reliability of a system is the probability of accomplishment of a function under specified environmental conditions and over a specified time.

Definition 2 (Safety)

Safety is the probability, that no catastrophic accidents will occur during system operation, over a specified period of time.

As can be concluded from the definitions, reliability is orientated more towards the function of the system, and safety looks at the consequences and possible accidents. The failure of a breakwater that ensures the maneuverability inside a harbour is different with respect to safety compared to the failure of a dike protecting a city from flooding. In general, the consequences of 'failure' for breakwaters are mainly economic and are not as extreme as for example the consequences of the dike breach. One might therefore expect that the required reliability differs among these structures.

Absolute safety is not realistic as this is often only achieved by eliminating all risk which requires a large investment. In order to determine how safe a structure or system should be, an acceptable level of risk needs to be defined. For some systems regulations have been derived, however, it can be a hard question to answer in case there are no regulations available.

2.1.2 Risk Acceptance

The quantification of risks is an important step in the design of a structure. The subsequent step is the evaluation of risk. To what extent should the risk be reduced or what level of risk acceptable for that specific structure and situation? The acceptable risk level is often hard to determine and subjective factors such as political processes and risk perception play an important role. In order to find a substantiated answer to this question, a general framework has been developed by [Vrijling \(2001\)](#). The framework is based on the assumption that accident statistics are the result of societal optimization and reflect what is considered acceptable by society. The framework evaluates the risk on three criteria; individual, societal, and economical risk. The societal risk is concerned with the population wide effects whereas the individual risk is related to the distribution of effects within the population. The economical perspective concentrates on the balance between the potential damage and investment costs. The so-called economical optimization has been developed and applied by [van Dantzig \(1956\)](#) after the 1953 storm surge disaster. The idea is to select the most critical risk level from these criteria to use for the remainder of the reliability analysis.

2.1.3 System Reliability

When assessing the reliability of a structure, the structure is considered as a system consisting of multiple subsystems or elements which can fail in many different ways called failure mechanisms or failure modes [CUR \(1997\)](#). The system can be in a specific state of stability described by a load and resistance parameter. In case the load is equal to the resistance, the stability of the system reached a limit state and the system failed in case the limit state is exceeded. A fault tree is a common way of describing the system and its subsystems in a structured way and it helps with defining the overall failure probability of the system. The probability of failure for each limit state function can be determined if it is possible to formulate the

statistical distributions of the load and resistance parameter. The design parameters can be adjusted such that the determined failure probability (sub)system meets the target reliability. Currently, there are several probabilistic design methods commonly used to calculate the failure probability. These methods will be explained in Section 2.1.5.

A different approach is to define limit state functions by looking at the consequences of a failure instead of the actual failure itself. It is common to differentiate between two limit states; the ultimate limit state (ULS) and the serviceability state (SLS) (Jonkman et al., 2015). The first limit state refers to the complete loss of the system whereas the latter state relates more to the unavailability or hindrance of the system. A specific design criterion is considered for the evaluation of a certain limit state and failure mode. The design criterion is a quantification of the acceptable state of the system. The selection of the design criteria or target reliability is explained in the next section.

2.1.4 Safety Standards & Guidelines

In order to limit the risks, safety standards can be used. A limit value for societal risk, for example, can be used to avoid disproportionate exposures to risks for the entire society. At the start of a reliability assessment for any structure, a target reliability should be derived. These values are generally expressed by means of failure probability or reliability index and depending on the reference period to which they are applied. Design guidelines and codes provide information and recommendations regarding different aspects of the structural design and often include information concerning the relevant safety standards.

In the past decade, different safety standards and guidelines have been conducted for different fields of application. Design principles for the structural design of steel, timber, and concrete structures are provided by the Eurocode. Hydraulic loads on structures are not covered by the Eurocode at the moment of writing this research. Guidelines on the design of hydraulic structures can be found in the PIANC reports or the Rock Manual for example.

Besides the application, the leading guidelines and safety standards also depend on the country. The Eurocode is a European guideline that contains additional appendices with national regulations. Some countries have their own codes and guidelines such as the British Standards and the Coastal Engineering Manual developed by respectively the United Kingdom and the United States of America.

Sometimes the required safety standard and the corresponding target reliability is provided by the client. A general model to derive the target reliability is proposed by Vrijling et al. (1998) and can be applied in case there are no available guidelines and the client did not prescribe a certain safety standard either. The model is an integration of two philosophies regarding acceptable risk defined by VROM (1988) and the Technical Advisory Committee on Water Retaining Structures (TAW) (1984). The model results in three acceptable failure probabilities. The most critical (smallest) of the three criteria can be chosen to determine the acceptable probability of failure of the system and to make sure that all three conditions are fulfilled. The generic model has been used in 2014 in the derivation of new safety standards for the flood defenses in the Netherlands Kuijken (2015).

2.1.5 Reliability Calculation Methods

Once the target reliability is determined, the reliability of the system should be assessed. The system consists of multiple elements which should be analyzed when considering the possible failure mechanisms. The acceptable failure mechanism should be distributed over all failure mechanisms and the design of the system should be accordingly. There are various approaches available for the design of a structure. The approaches each describe a different way of modeling the design variables when evaluating the reliability. The overall objective of the design approach is to design and maintain systems based on an acceptable risk level. The approaches can be categorized as deterministic or (semi-)probabilistic. The level IV method is left out of consideration as this method focuses on the consequences of the failure and the resulting risks.

- *Level 0 - Deterministic*

This is the most commonly applied approach in which only the load is considered to be uncertain. The required target reliability is translated to the design load which is used for the design calculations. So for example, an appropriate return period is selected, and based on that the corresponding wave conditions are used in the design calculations. Nominal values are used for the other variables and an overall safety factor is included. The safety factor makes the design safer, however, it is difficult to quantify by how much.

$$R_{\text{nom}} \geq \gamma S_{\text{nom}}$$

An appropriate value for the design parameter of interest is calculated using the limit state function. The value of the design parameter should be such that $Z \geq 0$ and the design will not fail. Note, that no probability of failure is calculated.

- *Level I - Semi-Probabilistic*

The level I method describes a semi-probabilistic approach in which characteristic values for the load and resistance parameters are used based on respectively high and low percentiles of the cumulative probability distribution. Also, partial safety factors are included in the derivation of design values. The safety factors are based on level II calculations or from relevant guidelines such as the [PIANC \(1992\)](#) or the Eurocode.

$$\frac{R_{\text{rep}}}{\gamma_R} > \gamma_S S_{\text{rep}}$$

The value of the design parameters of interest is calculated in a similar way compared to the level 0 method.

- *Level II - Probabilistic*

The level II approach consists of several methods that take into account the mean, standard deviation, and the correlations of the stochastic variables in the determination of the failure probability. The limit state function is analyzed in order to find the design point which corresponds to the condition in which the design is most probable to fail. This analysis is straightforward in the case of a linear

function and independent and normally distributed variables. However, in reality, this is often not the case and linearization of the limit state function and transformation of the variables is required. Several calculation methods exist such as the first and second-order reliability methods (FORM and SORM respectively). An advantage of these iterative methods is that they do not require much computation time and they provide a physical interpretation.

- *Level III - Probabilistic*

The fundamental idea of the level III approach is a calculation of the failure probability by integration of the joint density function of the limit state function. Often it is not possible to solve the integral analytically and techniques like numerical integration methods or Monte Carlo simulation are used. The level III method is known to be accurate as it makes use of the full probabilistic information. However, in most cases, the joint density distribution is not known and there is some degree of dependence. The joint density can be simplified in case independence is assumed for the stochastic variables in the limit state function.

2.2 Extreme Value Theory

Extreme value theory or extreme value analysis (EVA) is a branch of statistics focusing on the modeling of rare events. Rare or extreme events are events that deviate from the average observed events. EVA plays an important role in many applications where the most interesting values of a large group of random values are either the largest or smallest values. The probability density distribution of these minima or maxima can be characterized as the extreme value distributions. Gumbel was responsible for a significant contribution to the extreme value theory with his book *Statistics of Extremes* (Gumbel, 1958). He especially emphasizes the application of extreme value theory by engineers and statisticians. The first type of problems treated this way consisted of civil engineering related problems such as annual floods flow and precipitation maxima.

2.2.1 Extreme Value Selection Methods

In practice quite often the extreme events must be obtained from time series data using a specific selection method before performing an EVA. The selection of extreme events is an important practicality. The goal is to select only the extreme events corresponding to the tail of the underlying distribution which is not interrelated. Typically, two approaches are used for selection is the block maxima or minima (BM) and the Peak Over Threshold (POT).

Block Maxima

The BM approach separates the whole time series into a number of predefined blocks and samples the extreme event from each block. The length of the block depends on the type of application and the characteristics of available data. The disadvantage of the BM approach is the potential loss of useful information (Lang et al., 1999). For example, a certain block might contain multiple extreme values, however, only the extremest value gets selected. On the other hand, the approach always samples a value from each block even when the block actually does not contain any values which can be characterized as extreme values.

Peak Over Threshold

The POT approach samples all the extreme values that lie above a certain truncation level or threshold (Leadbetter, 1991). This approach has shown to be useful for selection of extreme events in many applications such as for example flood frequency analysis (Bezak et al., 2014) and pitting corrosion (Rivas et al., 2008). The POT approach provides the ability to control the number of extreme events included and therefore allows to capture more information concerning the whole stochastic process compared to the BM approach. However, some difficulties are associated with the application of the POT method (Lang et al., 1999).

A declustering time lag δ is defined to make sure the selected extremes are sufficiently apart avoiding the selection of multiple extremes from the same cluster. Figure 4.1 illustrates the POT procedure applied on a fictive timeseries of variable X . The orange dots indicate the selected extremes which exceed threshold u . The red circle indicates a peak that has not been selected since the inter cluster time C_2 is smaller than the declustering time lag δ . Extremes S_1 and S_2 belong to the same cluster and therefore only the largest extreme is selected.

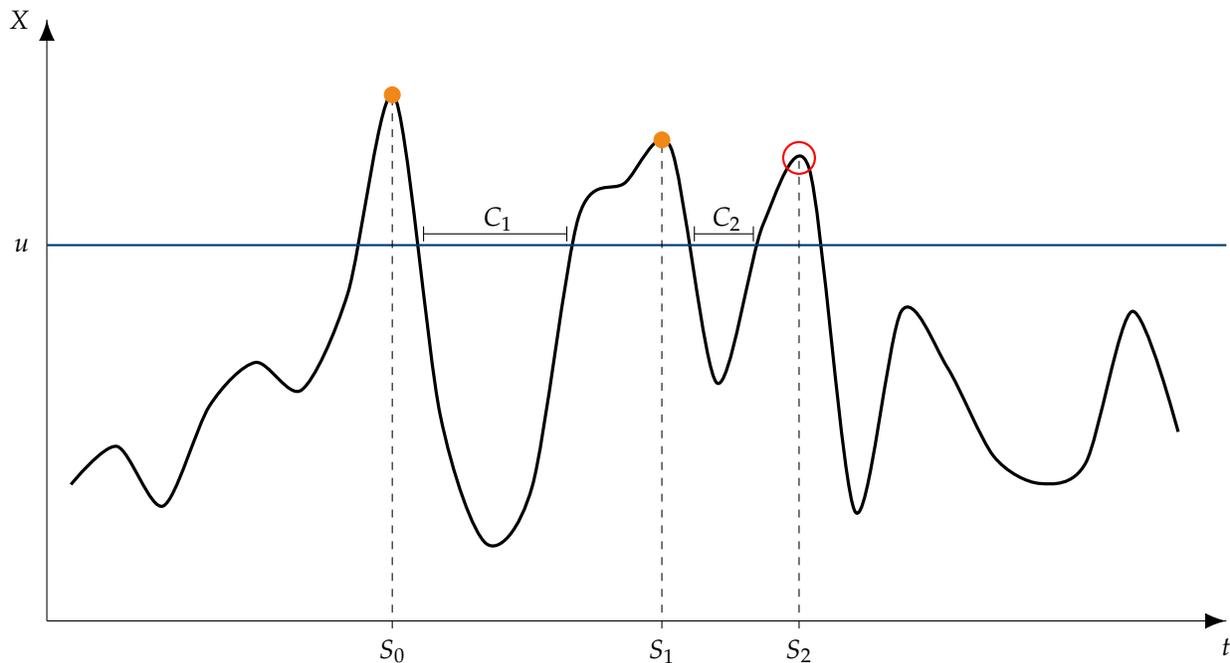


Figure 2.1: An example of the Peak Over Threshold procedure. It can be observed that only two out of three peaks are selected. The peak corresponding to S_3 has not been selected, because the inter cluster time is smaller than the declustering time lag, $C_2 < \delta$ and therefore the largest peak is selected $X[S_1] > X[S_2]$.

According to extreme value theory, the annual number of extremes in excess of high threshold u , should be Poisson distributed and the threshold excesses should be identically, independently distributed (i.i.d.) following a Generalized Pareto distribution (Pickands, 1975; Balkema and De Haan, 1974). It strongly depends

on the choice of threshold level u and declustering time lag δ whether these assumptions are satisfied, but the problem is that this choice is rather subjective since no standardized method is available. Multiple tests have been developed to help to select an appropriate threshold level u and declustering time lag δ .

Cunnane (1979) introduced the dispersion index to evaluate the validity of the Poisson assumption. Let X be a Poisson distributed random variable with parameter λ . Then the distribution function is given by:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}.$$

A property of the Poisson distribution is that $E[X] = \text{VAR}[X]$. The Dispersion Index DI uses this property and is defined by:

$$DI = \frac{\sigma^2}{\mu}, \quad \text{for } \mu \neq 0, \quad (2.1)$$

where μ is the mean and σ^2 is the variance of the Poisson process. The sample of extremes is considered to follow a Poisson process if the hypothesis that DI is significantly close to 1, cannot be rejected. The threshold u should be chosen in accordance with the dispersion index.

The assumption of sampling independently distributed extremes can be evaluated using the Extremal Index EI proposed by Ferro and Segers (2003). It evaluates the distribution of interexceedance times of the sampled extremes and provides a measure of independence. The extremal index is defined by:

$$EI = \begin{cases} 0 & \text{if } \max_i \{T_i\} \leq 1 \\ \min\{1, 2 \left(\frac{\sum_i T_i}{n} \right)^2 / \left(\frac{\sum_i T_i^2}{n} \right)\} & \text{if } 2 \leq \max_i \{T_i\} > 1 \\ \min\{1, 2 \left(\frac{\sum_i (T_i-1)}{n} \right)^2 / \frac{\sum_i (T_i-1)(T_i-2)}{n}\} & \text{otherwise.} \end{cases} \quad (2.2)$$

where T_i is the interexceedance time defined as $T_i = S_i - S_{i-1}$ and S_i is the index of peak i . The EI is equal to 1 in the case of independence and less than 1 if there is some dependence. The declustering time lag δ should prevent the selection of correlated extremes and could be chosen based on the extremal index.

Davison and Smith (1990) proposed to choose the threshold value based on the mean exceedance above threshold. The selected threshold value should be on the domain where the mean exceedance above the threshold is a linear function of the threshold. Lastly, the stability of the Generalized Pareto distribution parameter estimates could provide an indication for an appropriate threshold value.

2.2.2 Extreme Value Distributions

EVA deals with the stochastic behavior of maxima and minima of identical independent distributed (i.i.d.) random variables. The distributional properties of these extremes are determined by the tails of the underlying distribution. It is aimed to select an appropriate probability distribution to best model the sampled extremes. The lack of data concerning the observed extremes and the need for extrapolation is a fundamental difficulty when determining the underlying distributions accurately. Several distribution functions can

be used for modeling purposes of the extremes. The generalized Pareto distribution is a common assumption for the distribution of the POT sample (Pickands, 1975; Balkema and De Haan, 1974).

In case of the BM sample, three widely used families of extreme value distributions are the so-called Type I, Type II and Type III (Kotz and Nadarajah, 2000). The Type I distribution is known as the Gumbel distribution and is also the most commonly referred to in literature concerning extreme value theory. A generic expression of the Type I distribution, the so-called generalized extreme value distribution (GEV), is given by:

$$P(X \leq x) = \left[1 + \zeta \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\zeta}, \quad 1 + \zeta \left(\frac{x - \mu}{\sigma} \right) > 0, \quad -\infty < \zeta < \infty, \quad \sigma > 0,$$

where μ is the location parameter ($-\infty \leq \mu \leq \infty$), ζ is the shape parameter and σ is the scale parameter ($\sigma > 0$). Note, that the type II (Fréchet) and type III (Weibull) are obtained for $\zeta > 0$ and $\zeta < 0$ respectively. The Gumbel family of distributions is obtained for $\zeta = 0$.

The Type II and Type III distribution is also known as the Fréchet and Weibull distribution respectively. Both the Fréchet and Weibull distribution can be transformed to Type I distributions using a relative simple transformation (Kotz and Nadarajah, 2000).

2.2.3 Multivariate Extreme Value Theory

Several issues arise in case more than one variable is considered in the EVA. First of all, there is no distinct definition of a multivariate extreme and moreover, how to sample the multivariate extremes. The study performed by Morton and Bowers (1996) on the influence of extreme offshore conditions upon offshore structures, illustrates the application of different approaches. In this case, the maxima of the considered variables typically do not occur simultaneously: there is a certain delay. The maxima of both variables can be sampled independently or one might define a dominant variable. The idea is to select the maxima of the dominant variable and the associated concomitant variable when sampling the extremes. The results of these different criteria vary and more important, they depend on the type of application.

Secondly, modeling the dependence structures of variables can be challenging and the classical theory of multivariate extreme-values may not be sufficient (Guillou et al., 2018). To solve this issue, different probabilistic models have been developed. An important aspect in model construction is the degree of tail dependence. Several dependence coefficients and tests have been introduced in order to be able to examine and describe the degree of extremal dependence.

2.3 Copulae

A copula is a function which couples multivariate distribution functions to their one dimensional marginal distribution functions (Nelsen, 2007). The word copula was first used in a mathematical sense by Sklar (1959). A steadily growing interest in copulae and their application occurred in the late twentieth and early twenty-first century. According to Fisher (2004), copulae are of interest to statisticians for two main reasons: First, as a way of studying scale-free measures of dependence; and secondly, as a starting point for

constructing families of bivariate distributions, sometimes with a view to simulation.

Copulae are applied in different industries especially applications in finance, insurance, and risk management gained much popularity. In these fields, often situations occur where much information is known about the marginal distribution and relatively less information is available about the joint distribution. Copulae are mathematical objects which help us to understand dependence and can be used to build models.

2.3.1 Definition of Copulae

Sklar's theorem (Sklar, 1959) is the basis of the theory concerning copulae and the foundation of many of its applications in statistics. Sklar's theorem explains the relation between multivariate distribution functions and their univariate margins. Theory on transformations of probability distributions (Bain and Engelhardt, 1987) is given below in order to enhance the interpretation of Sklar's theorem.

Lemma 1 (Probability Transformation)

Let $X \sim F$ and F is a continuous distribution function. Then $F(X) \sim U(0, 1)$.

Proof: $\mathbb{P}(F(X) \leq u) = \mathbb{P}(F^{-1}(F(X)) \leq F^{-1}(u)) = \mathbb{P}(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u, \quad u \in [0, 1]$.

Lemma 2 (Quantile Transformation)

Let $U \sim U(0, 1)$ and F be any distribution function. Then $X = F^{-1}(U) \sim F$.

Proof: $\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x), \quad x \in \mathbb{R}$.

Theorem 1 (Sklar's Theorem)

Let (X_1, X_2, \dots, X_n) be a random vector with joint distribution function H and univariate marginals F_1, F_2, \dots, F_n . Then there exists a copula $C: [0, 1]^n \rightarrow [0, 1]$ such that, for all $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$,

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

C is uniquely determined on $\prod_{j=1}^n \text{Range}(F_j)$. If F_1, F_2, \dots, F_n are continuous, C is unique and can be obtained by means of:

$$C(u_1, u_2, \dots, u_n) = H(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_n^{-1}(u_n)), \quad \mathbf{u} \in \prod_{j=1}^n \text{Range}(F_j)$$

where $\text{Range}(F_j) = \{F_j(x) : x \in \mathbb{R}\}$ denotes the range of F .

Provided that all necessary derivatives exist, the copula density is defined as:

$$c(F_1(x_1), \dots, F_n(x_n)) := \frac{\partial^n C(F_1(x_1), \dots, F_n(x_n))}{\partial F_1(x_1) \dots \partial F_n(x_n)}$$

The formal definition of a copula is given by:

Definition 3 (Copula)

A copula is a distribution function $C : [0, 1]^n \rightarrow [0, 1]$ with uniform margins if and only if:

1. $C(u_1, \dots, u_n) = 0$ if $u_j = 0$ for at least one $j \in \{1, \dots, n\}$
2. $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j \quad \forall u_j \in [0, 1]$ and $j \in \{1, 2, \dots, n\}$
3. C is n -increasing, that is, the C -volume V_C of every n -box $[\mathbf{a}, \mathbf{b}] := [a_1, b_1] \times \dots \times [a_n, b_n]$ contained in $[0, 1]^n$ is non-negative, namely $V_C([\mathbf{a}, \mathbf{b}]) \geq 0$.

An important copula function is the so-called product copula $\Pi(u_1, u_2, \dots, u_n) = \prod_{j=1}^n u_j$. The statistical interpretation of the product copula is that the involved n random variables are independent. Other important copula functions are the Fréchet-Hoeffding upper and lower bound given by $M(\mathbf{u}) = \min(\mathbf{u})$ and $W(\mathbf{u}) = \max(\sum_{j=1}^n u_j + n - 1, 0)$, respectively. The Fréchet-Hoeffding bounds first appeared in an article by Fréchet (Fréchet, 1951) after Hoeffding presented an earlier version (Hoeffding, 1940).

Theorem 2 (Fréchet-Hoeffding Bounds)

For every possible n -copula C and $\mathbf{u} \in \mathbb{I}^n$, we have the bounds

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u}).$$

The bounds can be interpreted as the cases of extreme dependence where the lower bound and upper bound correspond to perfect positive and negative dependence respectively.

2.3.2 Families of Copulae

As mentioned before, copulae couple the marginal distribution functions with a multivariate distribution function. Often the goal is to find a stochastic model which represents the multivariate behavior of a process as close as possible. Therefore, it is very useful to have a variety of copulae at your disposal that can model different types of dependence structures. Over the years, a great number of investigations have been performed in order to define different families of copulae with different properties. Well-known copula families are for example the Elliptical, Archimedean, and Extreme Value copulae. The Elliptical family of copulae contains copulae with elliptical distribution, which have symmetrical tails due to the elliptical form. The Gaussian and Student t copula are well-known copulae from this family. The Archimedean family is widely applied due to the relatively simple nature of the copulae. Additional details concerning these copula families can be found in Durante and Sempi (2010) and Nelsen (2007).

2.3.3 Pair-copulae

In previous subsections, the copulae have been generally introduced for a multivariate case of n variables. Bivariate copulae became very popular since they have better capabilities of modeling different dependence structures compared to copulae with dimension $n > 2$. In this subsection, there will be an emphasis

specifically on bivariate copula and conditional bivariate copula since this is an important theory used in the vine-based methodology.

Multivariate data can be modeled using bivariate copulae. In order to do so, it is shown that the joint density can be decomposed into a number of pair-copula. This will be illustrated by an example for $n = 3$ variables. First of all, using Sklar's theorem (see Theorem 1), the joint density f is defined by:

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \frac{\partial^n C(F_1(x_1), \dots, F_n(x_n))}{\partial x_1 \dots \partial x_n} = \frac{\partial^n C(F_1(x_1), \dots, F_n(x_n))}{\partial F_1(x_1) \dots \partial F_n(x_n)} f_1(x_1) \dots f_n(x_n) \\ &= c(F_1(x_1), \dots, F_n(x_n)) f_1(x_1) \dots f_n(x_n) \end{aligned}$$

The joint density distribution f can also be expressed using a decomposition based on the chain rule and the definition of conditional probabilities:

$$f(x_1, x_2, \dots, x_n) = f_n(x_n) f(x_{n-1}|x_n) f(x_{n-2}|x_{n-1}, x_n) \dots f(x_1|x_2, \dots, x_n)$$

where f_n is the marginal density. According to these definitions, a bivariate and trivariate joint density can be expressed as:

$$\begin{aligned} f(x_1, x_2) &= f_1(x_1) f(x_2|x_1) & (2.3) \\ &= c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2), \end{aligned}$$

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) f(x_2|x_1) f(x_3|x_2, x_1) & (2.4) \\ &= c_{123}(F_1(x_1), F_2(x_2), F_3(x_3)) f_1(x_1) f_2(x_2) f_3(x_3), \end{aligned}$$

where c_{12} and c_{123} are the bivariate and trivariate copula density respectively.

Furthermore, the following bivariate joint density can be expressed as follows:

$$f(x_3, x_2|x_1) = f(x_2|x_1) f(x_3|x_2, x_1) = c_{32|1}(F(x_3|x_1), F(x_2|x_1)) f(x_3|x_1) f(x_2|x_1) \quad (2.5)$$

Both decompositions 2.3 and 2.5 are used to rewrite Equation 2.4 such that the following expression is obtained for the trivariate density composed by bivariate copula densities:

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) c_{32|1}(F(x_3|x_1), F(x_2|x_1)) \times \\ &\quad c_{13}(F_1(x_1), F_3(x_3)) f_3(x_3) & (2.6) \end{aligned}$$

$$= c_{32|1}(F(x_3|x_1), F(x_2|x_1)) c_{12}(F_1(x_1), F_2(x_2)) c_{13}(F_1(x_1), F_3(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \quad (2.7)$$

The 3-dimensional copula in Equation 2.4 is now decomposed in multiple pair-copulae. Note, that the decomposition 2.4 is not unique since the variables can be permuted in 6 different ways. The number of permutations grows fast for increasing dimensionality. A method to describe and categorize these models is established using the so-called regular vines.

2.4 Vines

Vines are graphical structures that represent joint probability structures named after the close resemblance to grapes (Joe and Kurowicka, 2011). The joint distribution of several independent variables is defined by the product of the marginal probability distributions. The presence of some degree of dependence between the variables requires also a certain dependence modeling in the joint distribution. Copulae are popular models to use for this purpose. As explained before, it allows to separate the information from the margins and the dependence. Furthermore, many different copulae have been developed which makes it possible to model a different kind of dependence structures (see Section 2.3.2). In contrast to bivariate copulae, multivariate copulae are often limited in the variety of dependence structures they can reassemble. A popular choice is the normal copula, however, it is not able to model tail dependence. Regular vines are often used to model joint distributions using bivariate copulae as building blocks and have proven to be a flexible tool in high-dimensional dependence modeling. This methodology originates from Joe (1996) and is investigated and extended in more detail by Bedford and Cooke (2001; 2002).

2.4.1 Vine Types

There exist different varieties of graphical models called vines. A regular vine is a special case that is characterized by the number of conditioned variables for each edge. The number of conditioned variables equals 2 for every edge which does not hold for non-regular vines.

Mathematically, a vine \mathcal{V} can be described as a nested set of trees $\{T_1, \dots, T_{n-1}\}$ where the edges of tree T_j are the nodes of tree T_{j+1} for $j = 1, \dots, n-2$. For a regular vine on n variables it holds that two edges in tree j are only joined by an edge in tree $j+1$ if they share a common node, $j = 1, \dots, n-2$. A formal definition is given by (Joe and Kurowicka, 2011)

Definition 4 (Regular Vine)

\mathcal{V} is a regular vine on n elements with $E(\mathcal{V}) = E_1 \cup \dots \cup E_{n-1}$ denoting the set of edges of \mathcal{V}

1. $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$,
2. T_1 is a connected tree with nodes $N_1 = \{1, \dots, n\}$, and edges E_1 ; for $j = 2, \dots, n-1$, T_j is a tree with nodes $N_j = E_{j-1}$,
3. (*proximity condition*) for $j = 2, \dots, n$, $\{a, b\} \in E_j$, $\#(a \Delta b) = 2$ where Δ denotes the symmetric difference operator and $\#$ denotes the cardinality of a set.

Figure 2.2 shows a regular and a non-regular vine. Two nodes in a regular vine can only be connected in case the nodes share a common variable. This holds for the first tree of both vines in Figure 2.2. This does not hold for the second tree of the vine on the right-hand side resulting in a multivariate copula with $n = 4$ variables and therefore, this vine is referred to as a non-regular vine.

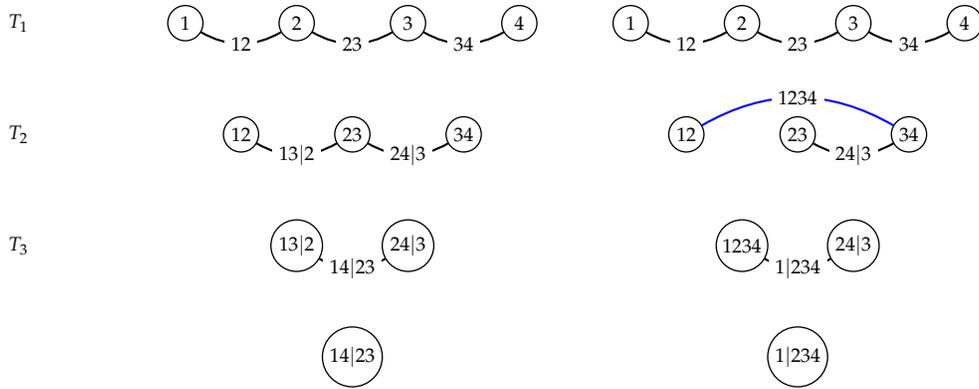


Figure 2.2: A regular vine (left) and a non-regular vine (right) on $n = 4$ variables. The blue edge in the non-regular vine shows a connection between two nodes which do not share common node. Every edge on the left vine represents a (conditional) bivariate copula.

Among the regular vines, one can distinguish two specific types regarding the graphical structure: (canonical) C-vines and (drawable) D-vines. A C-vine is characterized by a unique node in each tree T_i which has the maximum degree of $n - i$. The D-vine is characterized by a degree no higher than 2 for each node in tree T_1 . The degree of a node is the number of edges attaching to the considered node. Although D-vines look less complex, often they are mathematically more comprehensive.

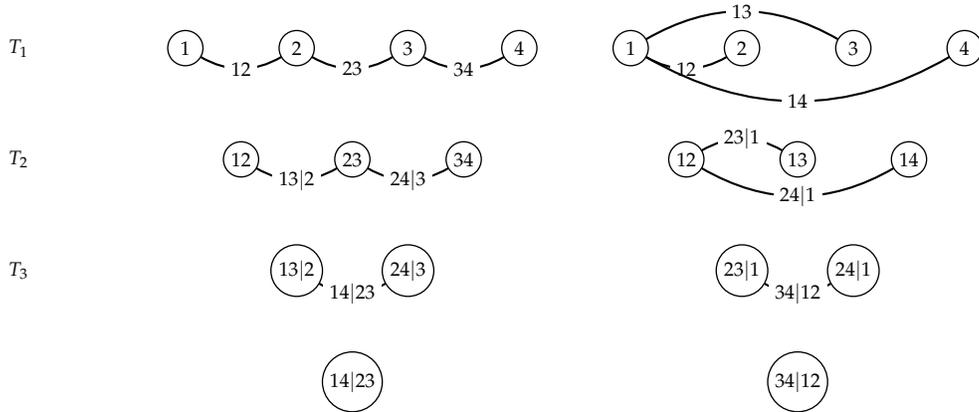


Figure 2.3: A D-vine (left) and a C-vine (right) on $n = 4$ variables. Both vines correspond to the class of regular vines as can be observed from the fact that the conditioned set of variables equals 2 in the total vine.

Two other types of regular vines can be distinguished with respect to their application: vine copulae and partial correlation vine representations.

Partial Correlation Vines

A partial correlation vine is a vine structure that has a partial correlation ρ_e assigned to each edge. The edges in T_1 represent the correlation coefficient whereas the edges in T_2, \dots, T_{n-1} represent the partial correlations. [Bedford and Cooke \(2002\)](#) shows that for any regular vine on n elements, there is a one-to-one correspondence between the set of $n \times n$ positive definite correlation matrices and the set of partial correlation specifications for the vine. The partial correlation vine can be used as a way to selecting partial correlations which uniquely determine the correlation matrix. The partial correlations are algebraically independent and do not need to satisfy any algebraic constraint such as positive definiteness. Another important property is that the product of 1 minus the square of the partial correlations equals the determinant of the correlation matrix. Examples of applications are the generation of correlation matrices and the reparameterization of multivariate normal or elliptical distributions.

Vine Copula

The vine copula is also known as the bivariate copula vine specification. The vine copula is obtained by specifying a bivariate copula C_e for each edge e in the union $E(\mathcal{V}) = E_1 \cup \dots \cup E_{n-1}$ ([Joe and Kurowicka, 2011](#)). Each copula can be chosen independently. Note, that the edges in T_1 represent a marginal bivariate copula C_e and a conditional bivariate copula for edges in T_2, \dots, T_{n-1} . A multivariate Gaussian density distribution parametrized by a partial correlation vine is obtained in case the edges of a vine copula are assigned with bivariate Gaussian copulae. In general, the joint density resulting from a regular vine copula $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$ on n elements is given by:

$$f_{1\dots n} = f_1 \cdots f_n \prod_{e \in E(\mathcal{V})} c_{e_1, e_2 | D_e} (F_{e_1 | D_e}, F_{e_2 | D_e})$$

where $e \in E(\mathcal{V})$ is an edge with conditioned elements e_1, e_2 and conditioning set D_e ([Bedford and Cooke, 2001](#)). Furthermore, $c_{e_1, e_2 | D_e}$ is the conditional copula density.

In this research, the focus will be on regular vine copulae. For detailed information on partial correlation vines, it is referred to [Bedford and Cooke \(2002\)](#) and [Kurowicka and Cooke \(2003\)](#).

2.4.2 Construction

A regular vine can be modeled by an undirected graph composed of a number of trees. Each tree is an acyclic and undirected graph itself. The n nodes in the first tree T_1 each represent an individual variable and the edges E_1 a bivariate copula. The edges will be the nodes of the subsequent tree and are connected only if they share a common node. Therefore, a constraint set, conditioned set, and conditioning set is defined for each edge of a vine. The constraint set consists of all nodes from tree T_1 that is in reach from a specific edge. The conditioned set of an edge is obtained by the symmetric difference of the constraint sets of the joined edges. The intersection of the constraint sets of the joined edges will form the conditioning set. As mentioned before, for regular vines it holds that two edges are only joined if they share a common node such that the cardinality of the conditioned set always equals two and the information can be represented by a bivariate copula. A vine usually is graphically visualized by a range of trees. This is not a very compact way of visualization in case n is large and alternatively, a matrix representation is used in order to store the information and the structure of a vine ([Morales Napoles et al., 2010](#)).

2.4.3 Vine Optimization

Morales Napoles et al. (2010) counted the number of vines and regular vines considering labeled trees on n nodes. It was calculated that the number of labeled regular vines for n variables grows rapidly with $N_{vines} = \binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$ such that for $n = 6$, this already results in 7,776,000 labeled regular vines. The goal is to construct a vine model that represents the dependence structure as good as possible. The brute force approach evaluating all possible 7,776,000 vines in order to select the best vine is computationally not attractive. A good selection method is required for the trees, the pair-copula families and the parameter values.

Little is published on good selection methods for the regular vine structures. Dissmann et al. (2013) developed a sequential approach that used a maximum spanning tree (MST) algorithm based on Kendall's τ to find the best tree starting at T_1 . Other measures instead of Kendall's τ could be used such as tail dependence. Gruber and Czado (2015) introduced a Bayesian approach for the vine selection. Another approach to find the best vine is proposed by Kurowicka (2010) and uses the partial correlations to build the vine from the top node and progress to lower trees while ensuring the regularity condition.

A copula family should be selected for each pair of variables. Different selection criteria can be chosen such as a copula goodness-of-fit test, information criterion, or a selection based on data characteristics. The latter is often used in case a goodness-of-fit test is computationally too demanding. The best fit can be determined using an information criterion such as the AIC or BIC. The AIC performs quite well especially if the sample size and dependence increases (Hans, 2007). Alternatively, a copula goodness-fit test can be applied. According to (Genest et al., 2009), these copula GOF can be divided into three subgroups: (1) copula family specific test, (2) general test including parametric strategy choices, and (3) general applicable tests without any preliminary strategic choices. The latter is the so-called blanket tests which are based on the construction of pseudo observations and the empirical copula. A certain statistic such as the Kolmogorov-Smirnov or Crámer-von Mises statistic is used to assess the performance of the estimated copula compared to the empirical copula (Brechmann, 2010).

Lastly, several estimation methods exist for the determination of the parameter values of the bivariate copulae. Maximum likelihood estimation is a well-known method and widely applied.

2.4.4 Tree-equivalent Classes

The copulae in a vine copula can be chosen independently from each other. However, the copulae specified in one tree will affect the copulae in the subsequent trees. The ordering of the variables in the first tree given a specific tree structure also influences the final result. As explained before, the number of regular vines grows rapidly ($\binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$) for an increasing number of variables n . In order to decide which vine can be appropriate for a certain applications, it is useful to understand how different vines compare.

Morales Napoles et al. (2010) performed a study on the enumeration of all trees, regular vines and tree equivalence classes for different number of variables n . The tree-equivalence classification of vines is used to define sub-classes of vines. The classification is based on unlabeled trees on each level of the vine. The following definitions are used:

Definition 5 (Graph Isomorphism)

Two labeled graphs $G_i = (N_i, E_i)$ and $G_j = (N_j, E_j)$ are isomorphic if there is a bijection $\phi : N_i \rightarrow N_j$ such that for all pairs $(a, b) \in E_i \iff (\phi(a), \phi(b)) \in E_j$. If two graphs are isomorphic they are the same unlabeled graph.

Consider the trees $T_i = (N_i, E_i)$ and $T_j = (N_j, E_j)$. A tree is an undirected acyclic labeled graph with a set of nodes N and edges E consisting of a subset of pairs of N .

Definition 6 (Tree-equivalent Vine)

Two vines are considered to be of the same tree-equivalent class if for each tree T_i in vine $V_k(n)$ a bijection (as defined in definition on Graph Isomorphism) can be found such that $T_i \in V_1(n)$. Furthermore, they are considered to be of the same tree-equivalent class of regular vines if the proximity condition holds.

This information is of importance when constructing a vine. It can be an advantage to know exactly how many regular vines can still be constructed after the construction of the first tree for example. Furthermore, one might want to add extra variables to an existing vine. There can be different ways of linking the variables to the existing vine resulting in potentially different equivalence classes.

Figure 2.4 shows an unlabeled regular vine similar to the vine structures shown by Figure 2.3. First of all, the regular vines from Figure 2.3, excluding the first tree, belong to the same tree equivalent class. Now, the first tree can be constructed by adding a node to the second tree, see T_1 in Figure 2.4. The resulting regular vine belongs to a certain tree equivalent class depending on how the extra variable is linked to the existing tree. The possible links are depicted by the dashed lines and the filled circles. As one can observe, 2 out of 3 linking options result in a D-vine (see nodes indicated with a 'D'), and the other option results in the C-vine (indicated by a 'C') as depicted in Figure 2.3. In fact, two TEC can be distinguished for $n = 4$ variables: D-vines and C-vines. Other TEC besides the C and D-vines have been described for vines on $n = 5$ variables. More details regarding the concept of TEC can be found in [Morales Napoles et al. \(2010\)](#).

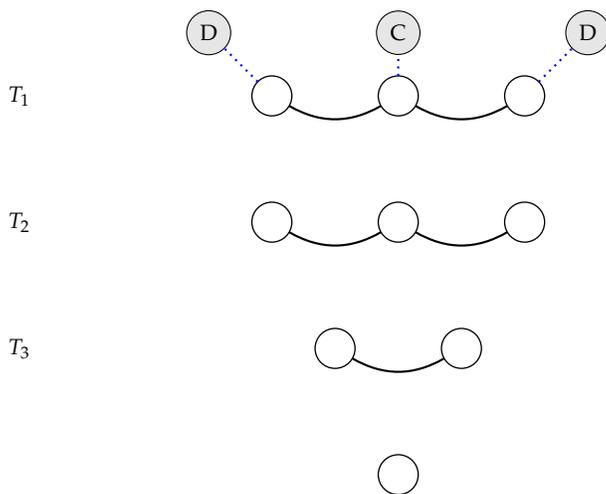


Figure 2.4: The unlabeled vine structure shows the possible extensions to $n = 4$ variables given a vine on $n = 3$ variables (tree T_2 and T_3). The possibilities of linking the 4th variable are depicted by the dashed lines and result in either a C-vine or a D-vine.

2.4.5 Goodness-of-Fit

The selection of a specific regular vine copula out of all the possible configurations including labels can be quite challenging. One aims to select a vine copula that is most appropriate for modeling the multivariate dependence structure. Therefore, a so-called goodness-of-fit (GOF) test should be performed. Several GOF tests for bivariate copula selection have been developed and described over the years (Genest et al., 2009). It is technically more complicated to perform such tests in the multivariate case and there is no widely accepted technique available. There also has been a lack of suitable GOF tests for regular vine models. Aas et al. (2009) proposed a first test based on the probability integral transformation (PIT) of Rosenblatt (1952). This idea has been worked out by different authors including an aggregation function resulting in different tests. Common univariate GOF tests such as the Cramér-von Mises or Kolmogorov-Smirnov tests, are applied after the aggregation step. Different varieties of the PIT based approach have been developed (Berg and Bakken, 2007; Breymann et al., 2003). Another approach is based on the empirical copula process Aas and Berg (2009). Two different tests developed and tested by Schepsmeier (2013b) which are based on the information matrix equality and the specification test introduced by White (1982). Schepsmeier (2013a) provides a comparison of 15 different regular vine copula tests. The results were promising for the approaches based on the White test and information matrix introduced by Schepsmeier (2013b; 2013a) as well as for the test based on the empirical copula process in combination with the PIT (Genest et al., 2009). Poor results have been obtained for the general GOF tests based on the PIT.

2.4.6 Sampling

After a regular vine is constructed modeling a certain multivariate dependence structure, one might want to sample from the vine assessing probabilities and evaluating risks. There are two approaches for sampling vine distributions which are called the cumulative and density approach respectively. The latter is based on a situation in which the distribution of a vine copula is completely specified and thus can be used for sampling. The cumulative approach is based on the inverse cumulative distribution functions. A sample of n independent uniformly distributed variables is used together with the inverse cumulative distributions to determine realizations of the variables of interest. Further details on the sampling approaches can be found in Bedford and Cooke (2002).

2.5 Return Period & Design Values

The concept of return period is a common criterion used for structural design purposes and moreover, in risk analysis and risk assessment. The importance of the concept of the return period, especially in the field of civil engineering, is well known (Singh et al., 2007). First, in Section 2.5.1, the formal definition in the univariate case will be given. Thereafter, the focus will be on the notion of the multivariate return period (MRP) in Section 2.5.2. The construction of MRP is rather complicated and multiple approaches will be outlined.

2.5.1 Univariate Case

Let p be the probability of the occurrence of an event in a certain period of time (normally one year). Assume that the occurrences of such events in different periods of time are independent. The time measured from the first occurrence of the event can be modeled by a geometric random variable with a mean value $\frac{1}{p}$. The formal definition of the return period is now defined as (Castillo, 2012):

Definition 7 (Return Period)

Let E be an event and T the random time between consecutive occurrences of event E . The mean value, τ , of the random variable T is called the return period of the event E .

Event E typically is the exceedance of a certain critical value by a random variable X , $P(X > x_{crit})$. Mathematically, this can be expressed using the CDF F_X . The probability of exceedance of a certain maximum x_{crit} each year is expressed by $P(X > x_{crit}) = 1 - F_X(x_{crit}) = Q$. The corresponding return period of that random variable to exceed the value is defined by $T = \frac{1}{QN_s}$ years. The probability that an event with probability p occurs at least once in a time frame T_l is geometrically distributed:

$$F(T_l) = 1 - (1 - p)^{T_l}$$

where $F(T) = 1 - (1 - p)^{\frac{1}{p}} \approx 1 - \frac{1}{e} = 0.632$ for $T_l = T$ and $T \rightarrow \infty$.

The concept of return period is used in engineering practice to determine design values for relevant variables in the design of a structure. The choice of T depends upon the importance of the structure and the consequences of its failure. A design value x_d corresponding to a certain return period T for variable X can be determined using the CDF F_X as follows (in case of a yearly maximum):

$$T = \frac{1}{QN_s} = \frac{1}{(1 - F_X(x_d))N_s}$$

$$x_d = F_X^{-1}\left(1 - \frac{1}{TN_s}\right)$$

Definition 8 (Design Values)

A value x_d assigned to a design variable associated with a prescribed probability of not being violated by unfavorable values during some reference period.

The design values for relevant design variables are used in the calculation of both the load and resistance when evaluating the limit state function of an object such that the design resistance is equal or larger than the design load, $R_d \geq S_d$.

2.5.2 Multivariate Case

While in the univariate case it is quite straightforward to derive design values based on a return period, in the multivariate case it more difficult. The identification problem of design events is of fundamental importance in the multivariate context. In the univariate case, a one-to-one relationship exists between the return

period and the design variable which disappears in case the number of considered variables is more than 1. Several definitions for the multivariate return period (MRP) have been proposed in literature (Gräler et al., 2013) and (Gräler et al., 2016).

Let's consider a bivariate case where variables X_1 and X_2 are studied. A widely used approach to calculate the bivariate return period by exploiting the bivariate CDF $F_{X_1 X_2}(x_1, x_2)$ which is often done by means of a copula $C_{UV}(u, v)$. One of the definitions is based on the probability $P(X_1 > x_1 \vee X_2 > x_2)$ and is referred to as the *OR* case. The return period expressed by:

$$T_{OR} = \frac{\mu}{1 - C_{UV}(u, v)}$$

where μ is interarrival time in years (typically equals 1 year). Another common return period is the *AND* case and is expressed by:

$$T_{AND} = \frac{\mu}{1 - u - v + C_{UV}(u, v)}.$$

All combinations (u, v) that have the same probability level $t = C_{UV}(u, v)$ also correspond to the same bivariate return period. In the bivariate case, these combinations can be found on an isoline on the copula which means infinite possibilities given that probability distributions are continuous. The one-to-one relationship is no longer present and the choice of design values is not obvious. The design values can be determined using various approaches once the joint density along this curve is defined. A commonly used approach is the most likely event calculated by:

$$(u, v) = \underset{C_{UV}(u, v) = t}{\operatorname{argmax}} f_{XY} \left(F_X^{-1}(u), F_Y^{-1}(v) \right)$$

The corresponding design values are easily calculated through the inverse of the marginal distributions: $x_D = F_X^{-1}(u)$ and $y_D = F_Y^{-1}(v)$.

Another definition of the bivariate return period is the so-called Kendall return period (KRP) and is extended to a multivariate setting by Salvadori et al. (2011). The idea is that the Kendall return period corresponds to the more critical events than the design event. It results in a partitioning of the super-critical and non-critical events. The Kendall distribution function K_C has the ability to project a multivariate distribution to a univariate one. It allows to calculate the probability that the copula value of a certain point (u, v) is smaller (or larger) than a given critical probability level t : $K_C(t) = P[C(u, v) \leq t]$.

The KRP has a disadvantage which is the unboundedness. Therefore, Salvadori et al. (2013) introduced the survival Kendall return period (SKRP) for which the critical layer is bounded. The SKRP is mathematically defined as

$$T_{SKRP} = \frac{\mu}{1 - \bar{K}(t)}$$

with \bar{K} the survival Kendall distribution function given by

$$\bar{K}(t) = P(\hat{C}(\bar{F}_1(x_1), \bar{F}_2(x_2)) \leq t)$$

where \hat{C} is the survival copula and \bar{F}_i the marginal survival distribution functions. The probability level t can be derived for a corresponding design return period T_{SKRP} by means of the inverse survival Kendall distribution $\bar{K}(t)$.

$$t = \bar{K}^{-1}\left(1 - \frac{\mu}{T_{\text{SKRP}}}\right)$$

The probability level can be evaluated analytically or numerically in case there is no expression available for $\bar{K}(t)$ (Salvadori et al., 2011). Once the probability level t is found, the design event can be derived in an analogous way as described for the *OR* and *AND* case.

The probability level is calculated similarly as illustrated above in case a third variable x_3 is introduced. In the 3-dimensional case this results in an isosurface instead of an isoline. All the points (u,v,w) on this surface have the same copula value t . The design event can be selected based on the point on the surface t with the largest joint likelihood, for example. For an n -dimensional situation, the probability level t yields an isohypersurface of dimension $n - 1$. The points on the isohypersurface all have the same probability level t and a design event can be selected based on a selection rule.

Ensemble of design events

The derivation of a design event based on a certain return period is characterized by selecting a single event out of a range of events. Thereby, potentially important information is not used. An ensemble based approach is introduced by (Volpi and Fiori, 2012) that selects a subset of events such that the variability within the set can be examined. This is done by sampling from a probability density function derived from the function over the isoline t (or isosurface in the case of 3 variables). More information is lost for higher dimensional situations. The ensemble approach becomes more advantageous, but also computationally more expensive for increasing dimensionality.

3 Case Study: Breakwater Design

First of all, some background information on the selected breakwater case is provided in Section 3.1. Section 3.2 elaborates on the basics of a breakwater design such as functional requirements, reliability and, the calculation methods for the relevant failure mechanisms. Lastly, the available data and information needed in the design process are presented in Section 3.3. The data is used to derive design values for relevant design variables required for the evaluation of the design formulas. Chapter 4 and 5 elaborate on the derivation approaches applied in this research.

3.1 Background

Around 2010, the Romanian government started a project that is focused on the protection and rehabilitation of the southern part of the Romanian Black Sea coastline. Constanta county is an important region as it contains Romania's largest seaport and the many beaches attract tourism. The purpose of the project is to provide a coastal protection system that reduces the risks of erosion and potential associated flooding.

The project consists of several works such as rehabilitation or extension of existing groynes/breakwaters, construction of offshore breakwaters, and extension of beaches with sand nourishments. The object that is considered in this case study is the extension of a breakwater near Tomis (north of Constanta). Van Oord got the assignment for the work (design and construction) and Arcadis derived the hydraulic boundary conditions. The companies were able to provide useful information on the design of the breakwater as well as on the extreme value analysis. The availability of information and the rather straightforward design of the classical breakwater structure makes it a suitable case for this research.

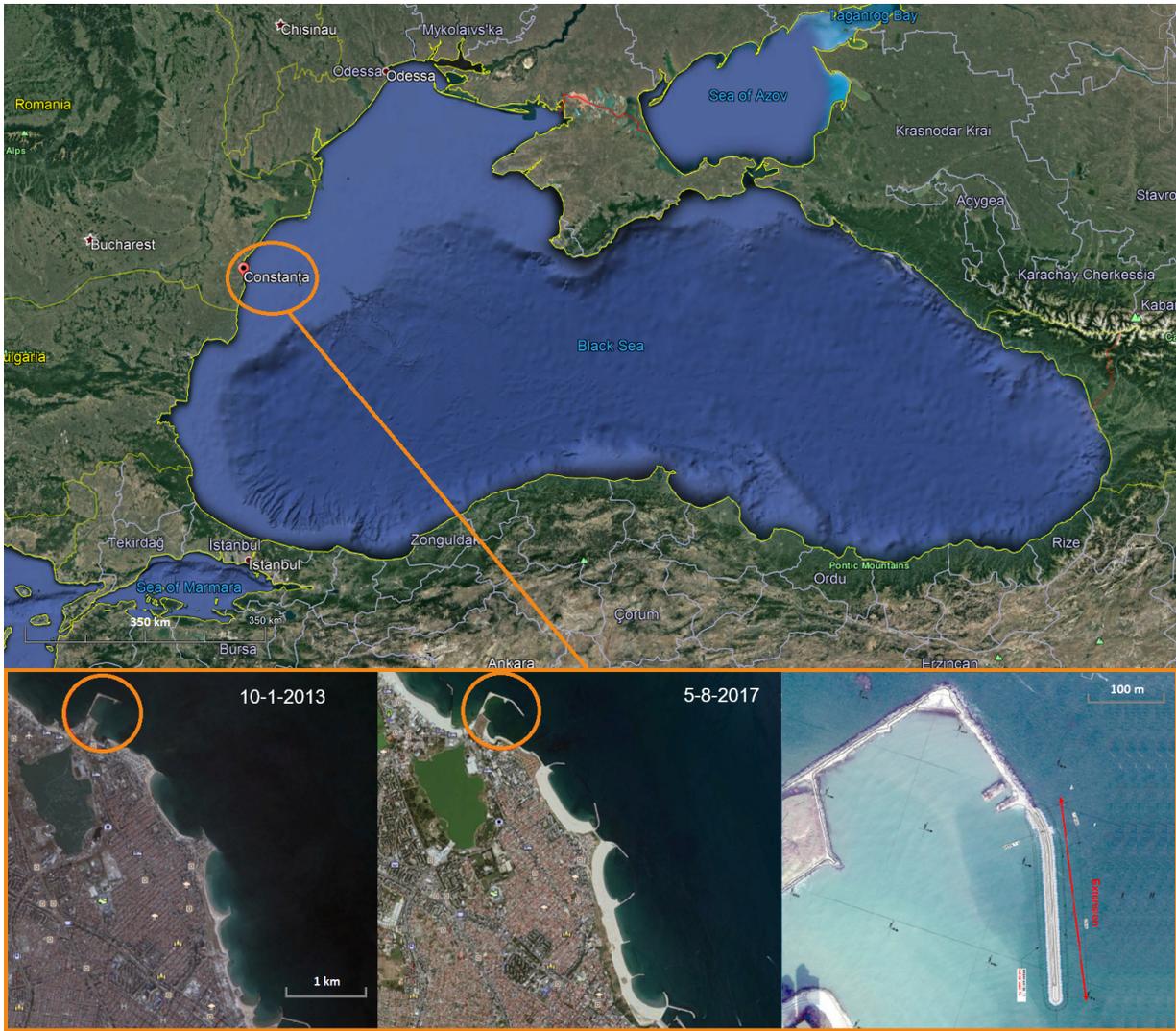


Figure 3.1: Upper: aerial image of Black Sea and the location of Constanta, Romania. Lower: the coastal region near Tomis, north of Constanta. The images clearly illustrate the realization of the coastal rehabilitation and protection objects over the specified period of time. The orange circle indicates the extension of an existing breakwater considered in this case study. Source: Google Earth and Van Oord

Breakwaters are common structures in coastal regions throughout the world that are primarily designed to protect coastal areas/infrastructure. Often breakwaters are designed specifically to protect port facilities or vessels harbored in ports from excessive vessel motions and wave disturbance. In the case of Constanta, the main function of the breakwater is to protect beaches from erosion caused by direct wave impact and currents. Waves cause longshore transport and cross-shore transport of sediment (Bosboom and Stive, 2012), which often result in unwanted erosion of sandy shores. Breakwaters can help to decrease wave heights in the breaker zone and mitigate beach erosion.

3.2 Breakwater Design Process

A typical design process consists of different phases moving from a conceptual design to a detailed design. Along with these design phases, the system-level changes as well: from the role of the breakwater in the entire port system towards the role of a structural element in the cross-sectional design. In this case study, the focus will be the preliminary design of the breakwater's cross-section and the interaction between its structural elements.

The objective of the design process is to find a safe and reliable design that meets the functional requirements and will not fail during its lifetime. Generally speaking, no structure is designed in such a way that it can be guaranteed that it will never fail. From a financial point of view, it will not be attractive (nor feasible) to ensure all risks are mitigated and the failure probability equals zero. Therefore in practice, a design is based on an acceptable failure probability that is determined based on the financial aspects as well as the potential loss of life and ecological damage.

The reliability of a structure indicates the likelihood of a system failure during its lifetime. There is a possibility that during its lifetime the structure fails to fulfill the functional requirements due to certain circumstances. In order to get a proper insight into the structural behavior and reliability of the breakwater design, it is needed to have an idea of potential failure modes. The failure modes considered in this case study are *overtopping* and *armour stability* of the armour on the *sea- and shore-side* slope as well as the *toe armour stability* (see Figure 3.2). The generic form of the design formula for each failure mode in accordance with the Rock Manual (2007) and the EurOtop wave overtopping manual (2018) is illustrated in Appendix B. The filter stability and geotechnical stability are not considered in this case.

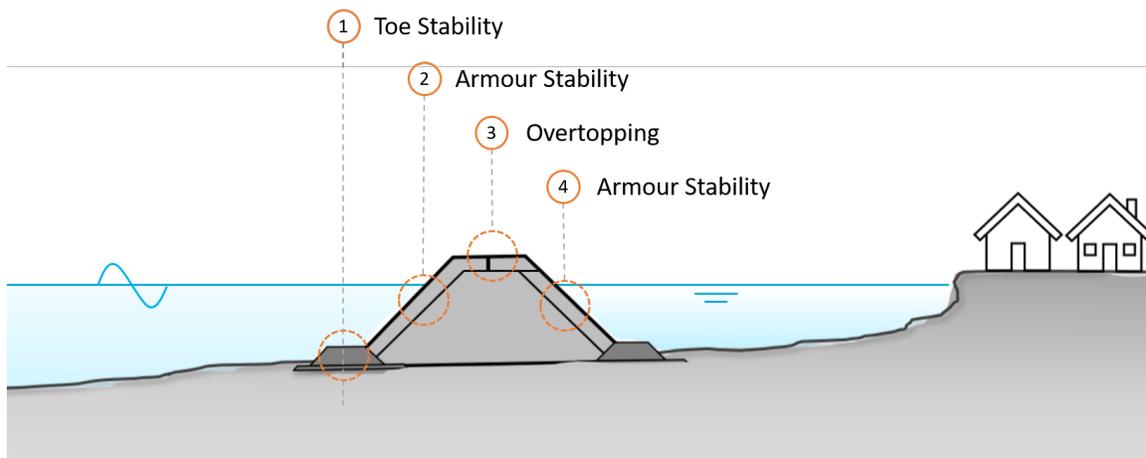


Figure 3.2: Schematic overview of a typical breakwater design situation.

In order to evaluate the reliability of the design in a quantitative way, the breakwater is considered to be a system consisting of multiple (structural) elements. The system can be in a state of stability described by a load S and resistance R parameter. The breakwater is in a limit state when $S = R$ and fails when $S > R$. Note, that the definition 'failure' in case of a breakwater relates to a certain amount of damage which is

specified by the damage level S_d and depends on the considered limit state. Often national regulations and guidelines prescribe scenarios for which the limit state of the system should be evaluated. The following limit states are defined in the Employer' requirements in this case study:

- Operational Limit State (OLS): due to 'typical' conditions normally up to (and including) an event with an annual probability of occurrence (1 year RP).
- Serviceability Limit State (SLS): due to extreme conditions normally up to (and including) an event with an annual probability of occurrence of once every 100 years (100 year RP). The structure should undergo minimal and readily repairable damage.
- Ultimate Limit State (ULS): due to extreme conditions in excess of the serviceability limit state (20% increase) to provide an indication as to the nature of any damage or failure that could result.

According to the employer's requirements, the design lifetime of the breakwater should be 50 years. In this case study, the design event corresponding to the serviceability limit state (SLS) with an annual probability of occurring once every 100 years is used. In other words, the acceptable annual failure probability is 0.01 and 0.39 for the entire design lifetime. A more detailed derivation of the failure probabilities is provided in Section 4.3.

There is no well-defined method for the evaluation of the limit state and derivation of a design that meets the target reliability. The method applied by Van Oord during this project is the so-called *Level 0* method, also the most commonly applied method (see Section 2.1.5). It assigns all uncertainty to the loads and the probability of failure of the structure is assumed to be equal to the probability that this critical load value is exceeded. For each failure mechanism, the limit states should be evaluated to determine the required cross-sectional design that is able to accommodate critical circumstances during its design lifetime. For example, given a critical wave height that is derived based on the prescribed target reliability, the minimum required crest level h_c can be determined satisfying the overtopping design criterion. In a similar fashion, the minimal required nominal armour diameter D_{n50} of the armour layers can be determined.

In this way, the target reliability is implicitly taken into account. The conventional method to determine these critical load values independently is illustrated in Chapter 4. The newly proposed method to determine the critical load values using the dependence between load variables is illustrated in Chapter 5.

3.3 Data Collection

In this section, the available data and information required for the breakwater design are presented. First of all, the deterministic parameters needed for the evaluation of the selected failure modes are given in Section 3.3.1. In addition, there are several parameters that require statistical analysis to determine appropriate design values, for example, environmental conditions such as wave height, wave period, and water level. The available metocean data used in the statistical analysis is discussed in Section 3.3.2 (note: the statistical approaches themselves are discussed in Chapter 4 and 5). The available metocean data is obtained from an offshore location and therefore, an offshore-nearshore transformation is needed to obtain information near the breakwater, discussed in Section 3.3.3.

3.3.1 Design Input

Relevant design input is obtained from the preliminary design documents prepared by Van Oord. The documents describe the preliminary design of the protection and rehabilitation works for the coastal protection of Tomis North. Most design inputs are constant (fixed) values like the water mass density ρ_w , but some input parameters are associated with the acceptable state of the 'breakwater system'. The design criterion like the acceptable mean overtopping discharge q , is chosen based on the considered limit state (SLS). Other examples of design criteria are the acceptable level of damage S_d and N_{od} . A complete overview of the relevant design input can be found in Appendix C.

3.3.2 Offshore Metocean Data

ARCADIS provided the offshore metocean data that originally is conducted from Fugro OCEANOR database. In total 20 years of offshore wind and wave data retrieved from a location (44°00N 29°30E) approximately 70 km offshore. The data are derived from the European Centre for Medium-range Weather Forecasts (ECMWF) operational and hindcast models are calibrated by Fugro OCEANOR against satellite data, and where available in-situ buoy data to ensure that the data are as high quality as possible. The data has been obtained over a period 01-01-1993 - 31-12-2012 on a 6-hour interval. The parameters included:

- H_{m0} : Significant wave height [m]
- $T_{m-1,0}$: Mean wave period [s]
- T_p : Peak wave period [s]
- θ_h : Mean wave direction [degrees]
- U_{10} : Wind speed at 10 meter height [m/s]
- θ_U : Wind direction at 10 meter height [degrees]

The wave parameters H_{m0} , $T_{m-1,0}$ and θ_h are derived based on the entire wave energy spectrum. The dataset also contains information on the wave systems primary swell and wind separately for variables H_{m0} , $T_{m-1,0}$ and θ_h . Generally speaking, there is not a swell component present in the Black Sea however, there are waves with relatively low wave steepness that are classified as a swell system in this case.

All the directions are incoming and measured clockwise from the north. It can be observed that most waves come from directions between 0°N and 120°N and between 180°N and 210°N. The most extreme waves (> 2.5m) come from directions 30°N and 60°N. The wind is distributed more evenly among all directions and the most extreme wind speeds (> 14m/s) come from the northern directions 0°N and 60°N.

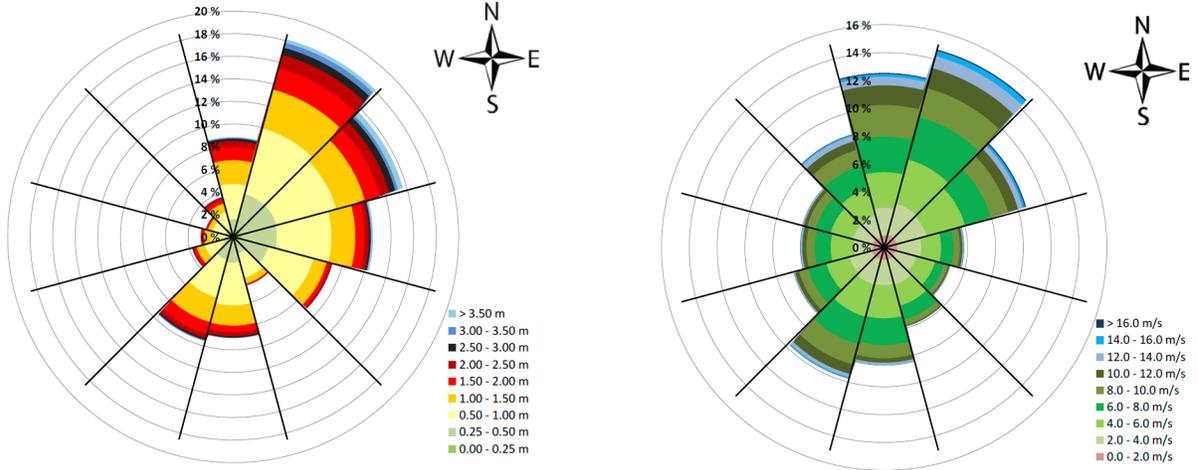


Figure 3.3: The directional distribution of the waves (left) and the wind (right). *Source: Arcadis.*

The dataset is filtered based on the direction corresponding to the significant wave height. The most extreme waves are coming from $15^{\circ}\text{N} < \theta_h < 75^{\circ}\text{N}$ and furthermore, this sector is seen as the most important sector considering the planned position and orientation of the breakwater. Table 3.1 contains some general statistics of the filtered data. The percentages of missing values are minor and therefore no measures are taken to treat them in a sufficient way. Besides the directional parameters and the peak period T_p , all parameters are considered to be continuous. The peak period T_p seems to be discontinuous which is probably an artifact of the data generation. It is stressed that a preliminary survey of the database should always be carried out, in order to fix possible anomalies due to software bugs and/or hardware/instrumental limitations.

	H_{m0} [m]	$T_{m-1,0}$ [s]	T_p [s]	$H_{m0,wind}$ [m]	$T_{m-1,0,wind}$ [s]	$H_{m0,swell}$ [m]	$T_{m-1,0,swell}$ [s]	U_{10} [m/s]
Mean	1.121	4.521	5.163	0.775	2.509	0.676	5.180	6.171
Standard deviation	0.865	1.107	1.372	0.889	2.426	0.396	1.390	3.549
Minimum	0.050	1.000	0.010	0.000	2.000	0.050	2.010	0.020
Maximum	7.410	9.870	9.820	7.110	9.490	3.470	12.790	20.960
Missing values %	0	0	0	0	9.2	0	0.02	0

Table 3.1: General statistics on the relevant wind and wave data.

The Fugor OCEANOR database does not contain any data regarding the water levels in the Black Sea. In the metocean study executed by ARCADIS, the design water levels have been determined based on the available data published in the literature. The water level variation is explained by several phenomena:

- Sea level rise,
- Storm surge,
- Tidal variation,
- Barometric pressure and resulting seiches,
- Seasonal variation,
- Mean sea level.

Based on these sources of information, a water level time-series is generated matching the same period (01-01-1993 - 31-12-2012) and accuracy (6-hour interval). The generation process and validation can be found in Appendix K.

Data is collected on the above described variables (water level variation, wave parameters, and the wind speed) because they serve as input for the offshore-nearshore transformation and/or they play an important role in the evaluation of the relevant failure mechanisms. The wind speed is only used in the transformation whereas the wave data and water level variations are transformed and subsequently used in the evaluation of the failure mechanisms.

3.3.3 Offshore-Nearshore Transformation

A detailed bathymetry dataset is available for the Constanta coastal region. The bathymetry data is required for the determination of the nearshore environmental conditions. As explained in the previous section, the available environmental data is retrieved from a location approximately 70 kilometers offshore. Waves propagating towards the shore transform and change directions as a result of the shallow water and the interaction with the seabed. The conditions at the toe of the breakwater are required for evaluation of the design. In order to determine the nearshore wave conditions, a transformation is performed using the 2D SWAN model (SWAN, 2020). SWAN (Simulating WAVes Nearshore) is a numerical wave model that can be used to obtain realistic estimates of wave parameters in coastal areas, given certain input parameters regarding the wind, bottom, and current conditions. The SWAN model defined by ARCADIS is kept the same, only the input parameters are changed in this case study. The overall model starts approximately at the point of data extraction 70 km offshore of Constanta. The SWAN model uses nested grids to increase spatial grid resolution. At the start, boundary conditions are imposed for the largest grid resulting in new boundary conditions for a more detailed grid. This is repeated a couple of times until sufficient detail is obtained and the nearshore conditions can be retrieved from output locations of interest (see Figure 3.4). The input parameters at the offshore boundary can differ from the output parameters nearshore provided by SWAN. For example, the neutral wind speed at 10-meter elevation is required as input, because the wind influences the locally generated waves, but is not of interest at the nearshore output location.

The output locations of interest, at the toe of the breakwater, are referred to as *BT140* and *BT157*. The bathymetry is based on a survey that has been carried out by Van Oord. Figure 3.5 shows a selection of the bathymetry survey focusing on the extension of the breakwater near Tomis.

name	x [m UTM]	y [m UTM]	depth [m MN75]
BT140	631844.4	4897607.6	-4.6
BT157	632200.7	4897337.3	-5.5

Table 3.2: Details on the output locations near the structure of interest. The SWAN model generates results (nearshore conditions) at these pre-specified locations.

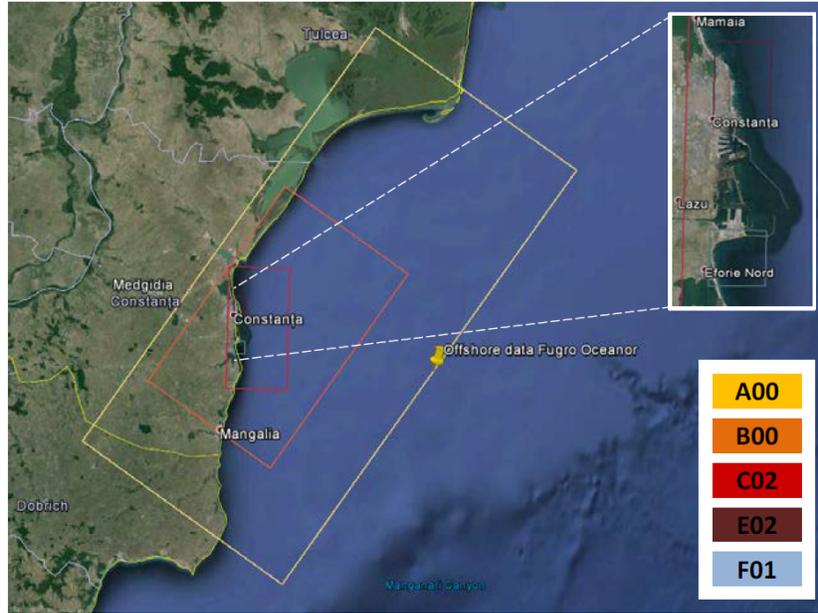


Figure 3.4: The SWAN grids and the Fugro OCEANOR output location. Source: Arcadis.

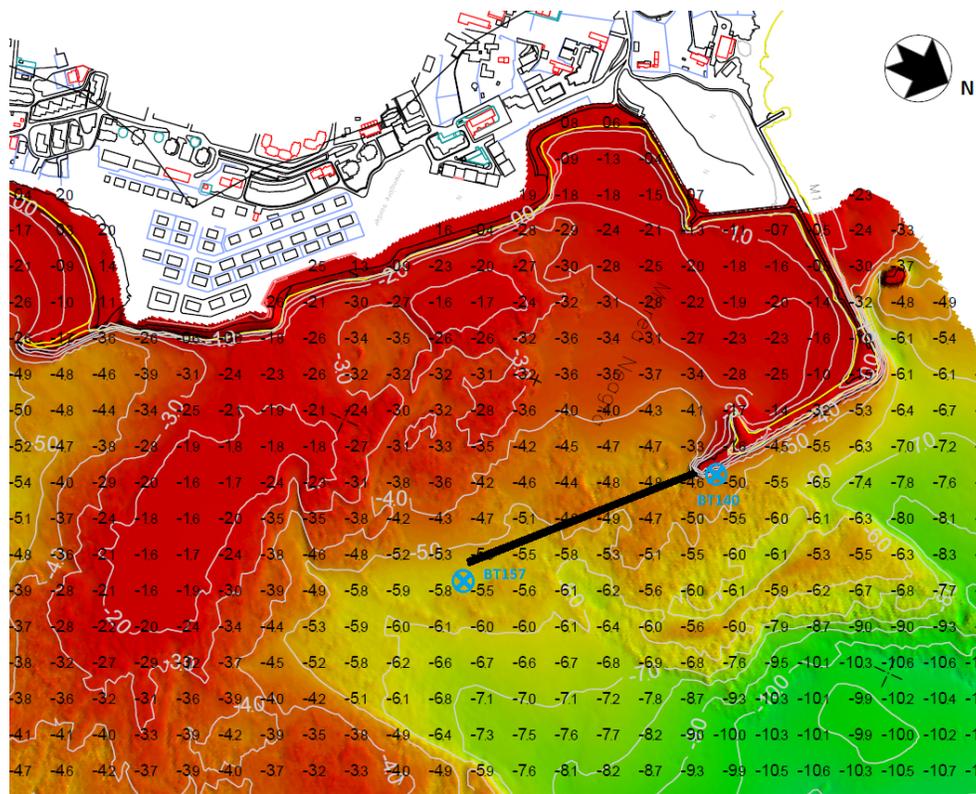
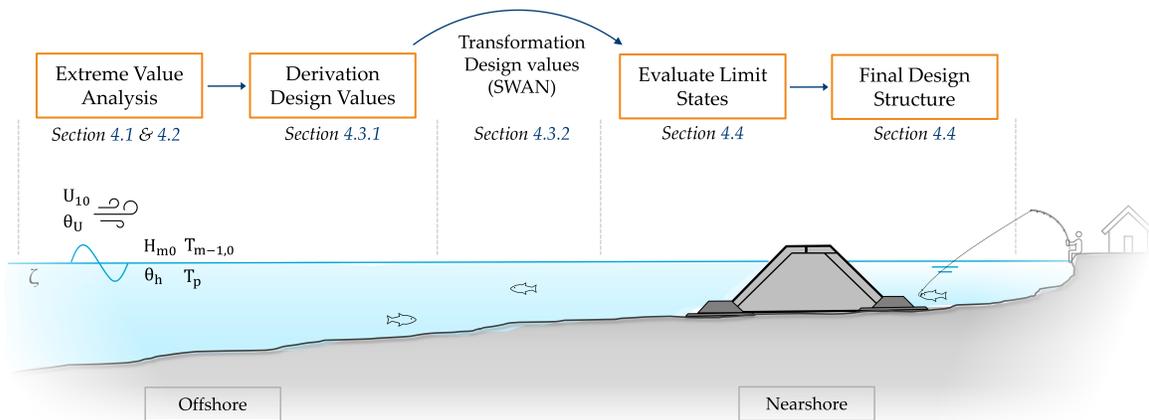


Figure 3.5: The extension of the breakwater considered in this case study is indicated by the black line. The output locations BT140 and BT157 are indicated by the blue crosshairs. Source: Van Oord.

4 Conventional Approach

This chapter illustrates the conventional approach, that is, the most commonly applied approach in current industry practice, to derive design values/loads for use in the evaluation of limit states. Section 4.1 illustrates the process of selecting the extremes from the metocean dataset, which are modeled by univariate statistical distributions (see Section 4.2) to derive design values (see Section 4.3). Lastly, the final design using the conventional approach is presented in Section 4.4.



For the evaluation of the limit states of a breakwater system, it is required to know the nearshore loads consisting of design values for variables like the significant wave height, mean wave period, water level, etc. Therefore, the hydraulic and meteorological conditions at the site should be studied, because often design values are not readily available. In the remainder of this chapter, the procedure to derive these design values using offshore metocean data is illustrated according to the conventional approach assuming independence between the design variables. In other words, it is a univariate approach that does not take into account any dependence information among the design variables. An extreme value analysis is applied to the offshore data and, subsequently the nearshore loads are determined using the 2D SWAN model, see Section 3.3.3. The derivation process is illustrated for a single variable throughout this chapter, but the same process applies to all variables of interest.

4.1 Extreme Value Selection

The metocean timeseries consist of observations measured every 6 hours for a duration of 20 years. For the derivation of design values, the behavior of the extremes is of interest. Therefore, only the extreme observations should be retrieved from the dataset and studied thoroughly for each variable separately. The method applied to select the extreme observations for each variable of interest is the Peak Over Threshold (POT) method. This method selects the height of all peaks that exceed a specific threshold value and are sufficiently separated (see Section 2.2). The threshold value u and declustering time lag δ are carefully selected based on several indicators to make sure the assumptions of the POT method are satisfied. The process of selecting appropriate POT threshold values with the use of these indicators is shown in Appendix D.

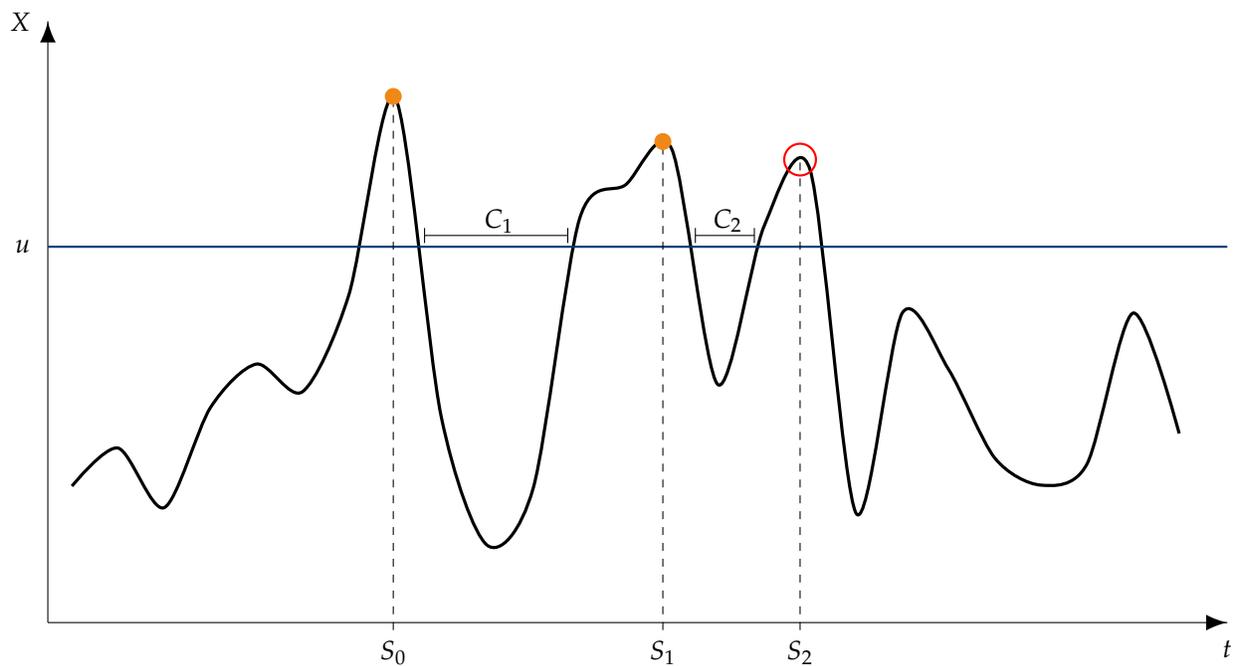


Figure 4.1: An example of the Peak Over Threshold procedure. It can be observed that only two out of three peaks are selected. The peak corresponding to S_3 has not been selected, because the inter cluster time is smaller than the declustering time lag, $C_2 < \delta$ and therefore the largest peak is selected $X[S_1] > X[S_2]$.

Figure 4.2 presents the selected extremes obtained with a POT procedure for variable H_{m0} with threshold value $u = 2.8$ meter and declustering time lag $\delta = 60$ hours. Note, that the final choice of parameter values still is somewhat subjective and other combinations of parameter values do not necessarily need to be incorrect. This procedure is repeated for all variables and the results summarized in Table 4.1.

	H_{m0}	$T_{m-1,0}$	T_p	U_{10}	ζ	ζ_{min}
u	2.8m	6.1s	7.8m	15.6m/s	MN75+0.45m	MN75-0.125
δ	60H	72H	72H	60H	72H	72H
n	130	204	96	43	54	59
N_s	6.5	10.2	4.8	2.15	2.7	2.95

Table 4.1: The threshold value u and declustering time lag δ are chosen based on the evaluation of multiple indicators as shown in this section. The parameter values are required for the application of a POT procedure during an extreme value analysis. The resulting number of extremes n and the characteristic number of extremes a year N_s is given.

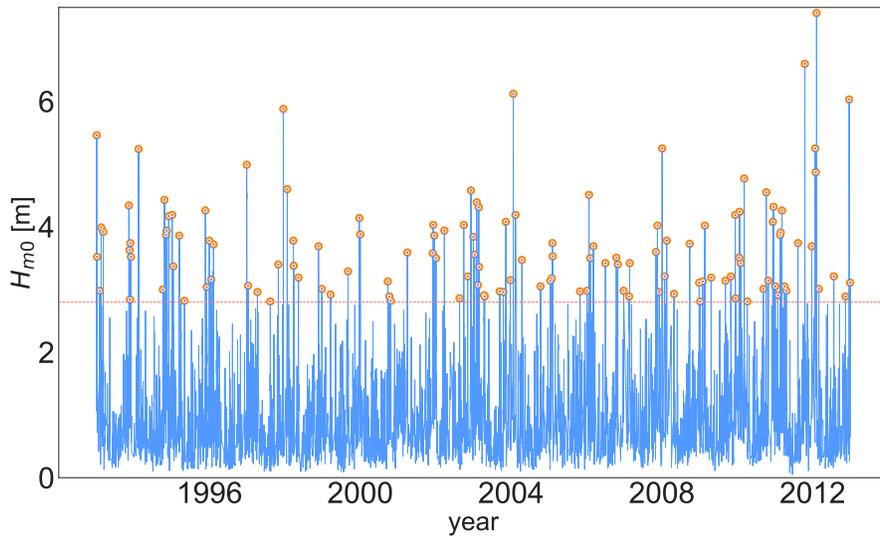


Figure 4.2: The selected extreme wave heights are indicated by the orange dots and have been obtained with the POT procedure with threshold $u = 2.8$ and declustering time lag $\delta = 60H$.

4.2 Distribution Modeling

As mentioned before, the excess above the threshold for the selected extremes with a POT procedure should be distributed according to a Generalized Pareto distribution. The distribution is fitted to the extremes (see Figure 4.3) and the distribution parameters are estimated using a maximum likelihood estimation.

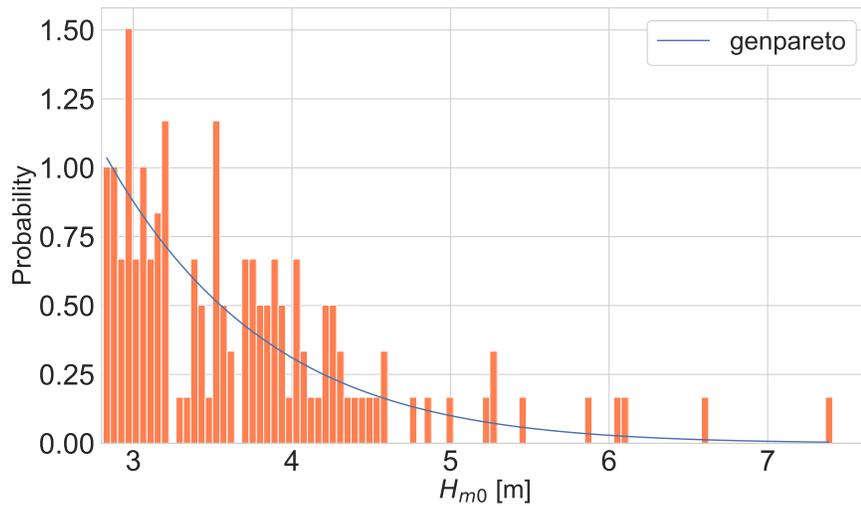


Figure 4.3: The probability density plot of the Generalized Pareto distribution with shape parameter -0.082 and scale parameter 0.964. The distribution parameters are estimated using a Maximum Likelihood estimation.

The so-called quantile-quantile (Q-Q) plot is a graphical method to compare two distributions by plotting their quantiles against each other. The points will be approximately on the line $y = x$ in case both distributions are the same. Figure 4.4 gives the Q-Q plot for the variable H_{m0} modeled by a Generalized Pareto distribution with parameters given in Table 4.2.

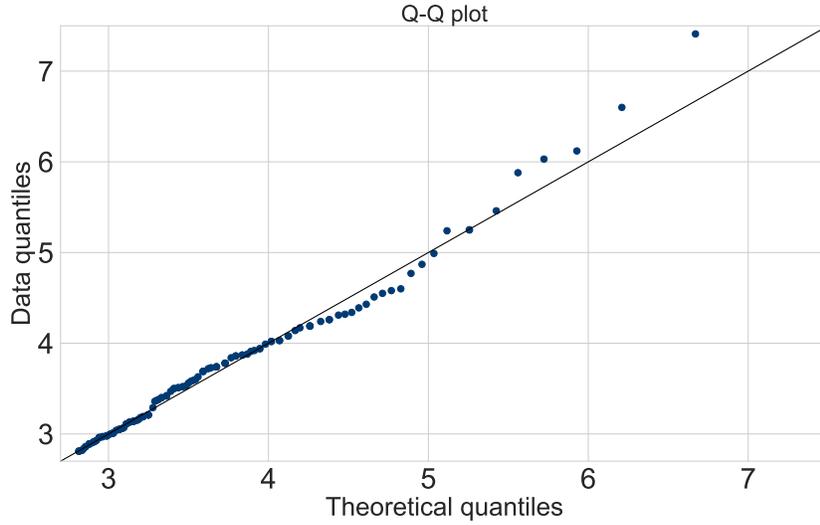


Figure 4.4: The quantile-quantile (Q-Q) plot for extreme wave height modeled by the Generalized Pareto distribution. The theoretical quantiles distribution looks similar to the data quantiles which is confirmed by the Pearson’s correlation statistic.

One can observe that the points approximately follow the line $y = x$ for the lower quantiles and start deviating from the line for higher quantiles, but overall it could be concluded that the data distribution looks similar to the theoretical distribution. This can also be statically assessed by means of Pearson’s correlation coefficient. The correlation coefficient R^2 is defined on $[-1, 1]$ where 0 means no correlation and 1 (-1) means perfect positive (negative) linear correlation. The p-value indicates the probability that an uncorrelated system could produce datasets that have a correlation coefficient at least as extreme as the computed R^2 . As can be observed from Table 4.2, the R^2 is close to 1 for all variables and the p-values are small (p-value $< 10^{-10}$) which indicates that the extremes are properly modeled by the Generalized Pareto distribution for a 5% significance level.

	H_{m0}	$T_{m-1,0}$	T_p	U_{10}	ζ	ζ_{min}
shape mean (standard deviation)	-0.082 (0.078)	-0.192 (0.047)	-0.591 (0.072)	-0.155 (0.156)	-0.058 (0.165)	-0.180 (0.125)
scale mean (standard deviation)	0.964 (0.113)	0.942 (0.078)	1.219 (0.133)	1.595 (0.346)	0.153 (0.033)	0.048 (0.009)
R^2	0.99	0.99	0.93	0.99	0.99	0.83
p-value	≈ 0					

Table 4.2: The distribution parameter estimates for the Generalized Pareto distribution for each variable of interest. The distribution fit is statistically assessed by means of Pearson’s correlation coefficient. The corresponding p-values indicate that the Generalized Pareto distribution is an appropriate distribution to model the extremes.

4.3 Design Values

4.3.1 Offshore Design Values

The next step is to derive the design values (extreme conditions) based upon an annual probability of occurrence once every 100 years (SLS) using the statistical distributions derived in the previous section. The probability of exceedance p of the SLS design event can be expressed in terms of return period T using the following relation:

$$T = \frac{1}{pN_s} \rightarrow p = \frac{1}{TN_s} = 0.0015$$

where N_s is the average number of storms a year. The design value x_d for variable $H_{m0,wind}$ is determined by evaluating the inverse cumulative density function F_X^{-1} for the probability of non-exceedance $1 - p$ as follows:

$$x_d = F_X^{-1}(1 - p) = F_X^{-1}\left(1 - \frac{1}{TN_s}\right). \quad (4.1)$$

$$(4.2)$$

The graphical interpretation of the design value derivation is given by Figure 4.5. The resulting design value (extreme condition) is determined for all relevant variables (see Table 4.3). In order to evaluate the limit state of the structure, an offshore-nearshore transformation is required to obtain the nearshore design values.

	H_{m0}	$T_{m-1,0}$	T_p	U_{10}	ζ	ζ_{min}
N_s	6.5	10.2	4.8	2.15	2.7	2.95
x_d	7.64m	9.71s	9.81s	21.42m/s	MN75+1.18	MN75-0.27

Table 4.3: The offshore design values derived based on a 100 year return period and the statistical distributions given by Table 4.2.

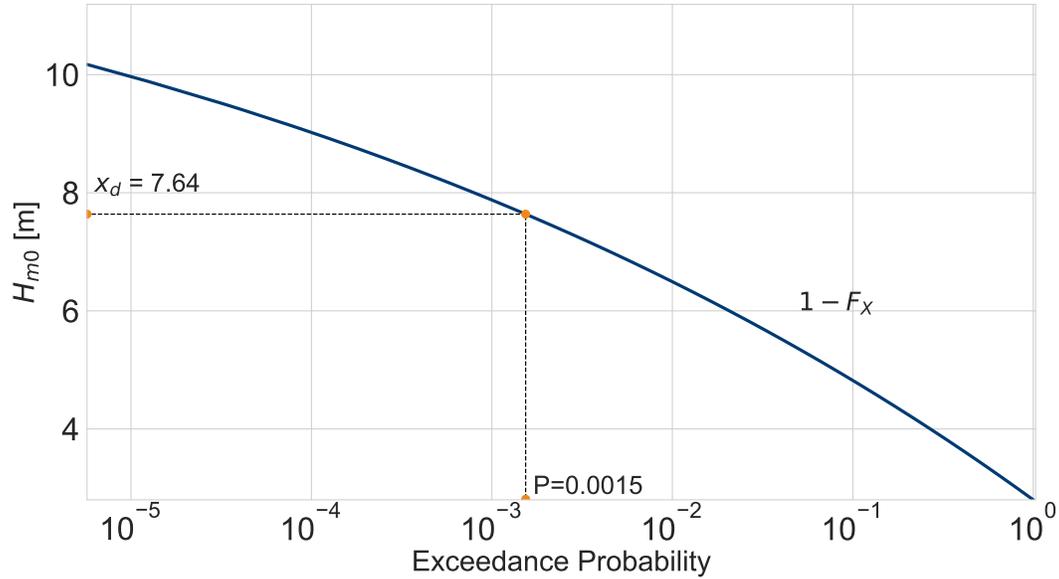


Figure 4.5: The Generalized Pareto cumulative density distribution for variable H_{m0} and the selected design value corresponding to a 100 year return period.

4.3.2 Nearshore Design Values

The offshore-nearshore transformation is performed by the numerical wave model SWAN of the offshore design conditions presented in Table 4.3 and the resulting nearshore conditions are shown in Table 4.4. The results are obtained for a fixed incoming offshore wave direction of 45 degrees (average of the considered directional sector). The offshore-nearshore transformation is performed for both the situation in which extremely high and extremely low water levels are considered. The propagating waves towards the shore clearly transformed when comparing Table 4.3 and Table 4.4.

	H_{m0}^{tot}	$T_{m-1,0}^{\text{tot}}$	T_p	ζ
$x_{d,nearshore}$ (max water level)	2.96m	11.17s	12.52s	MN75+1.28
$x_{d,nearshore}$ (min water level)	2.38m	11.15s	12.51s	MN75-0.17

Table 4.4: The nearshore design values for output location BT140 derived using SWAN model and the offshore design values presented in Table 4.3.

4.4 Final Design

Lastly, the design dimensions of the cross-sectional design are derived by evaluating the limit state for the failure modes described in Appendix B and using the nearshore design conditions presented in Table 4.4. The design dimensions of interest are depicted in Figure 4.6 and the final results are presented in Table 4.5.

	Crest level	Nominal diameter	Nominal diameter	Nominal diameter
Failure mode	Overtopping	Armour Stability (front)	Armour Stability (rear)	Armour Stability (toe)
Design dimension	$h_c = \text{MN75} + 4.07\text{m}$	$D_{n50} = 0.81\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 1.03\text{m}$ (HM 1000-4000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)

Table 4.5: The design dimensions are obtained for the conventional approach using a 100 year return period. The results should be interpreted as the minimal required crest level or nominal diameter in order to withstand the prescribed extreme storm conditions with an annual probability of occurrence once every 100 years.

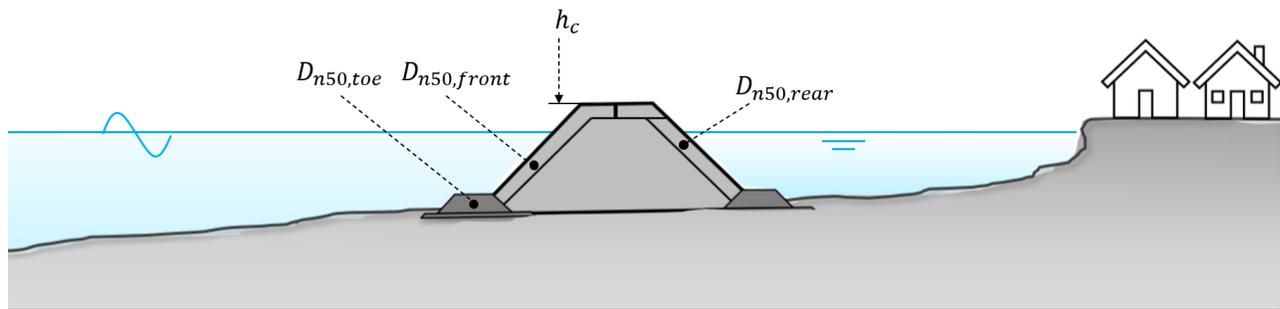
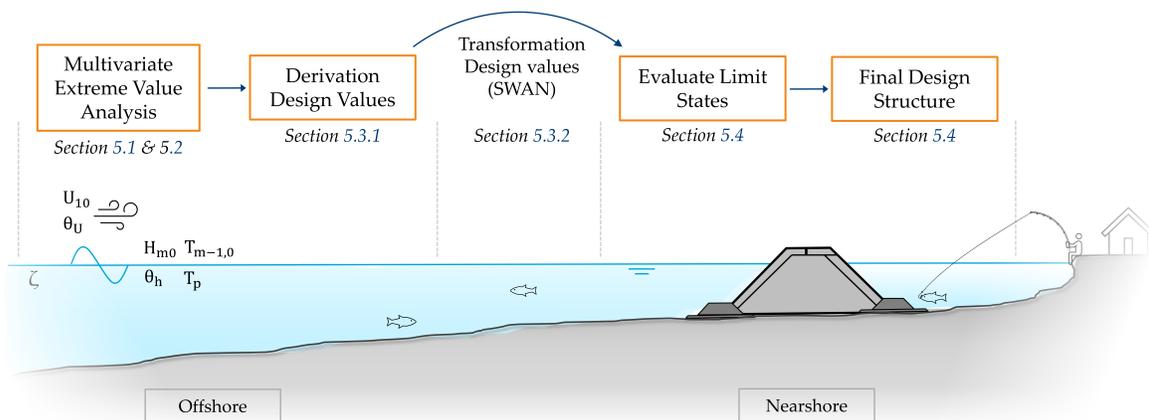


Figure 4.6: Schematic overview of a typical breakwater design situation and design dimensions of interest.

5 Vine-based Approach

This chapter illustrates the vine-based approach, that is, the newly proposed multivariate approach using dependence information, to derive design values/loads that are used in the evaluation of limit states. The aim of this multivariate approach is to derive a set of design values that is more realistic than simply combining design values based on univariate analysis. The general outline of the derivation is similar to the conventional approach presented in the previous chapter, but the emphasis will be on the aspects where both approaches differ. It should be noted that the methodology discussed in this chapter is used in a multivariate context (analyzing the behavior of 7 variables), however, the explanation of certain aspects of the methodology is substantiated by means of figures illustrating the *bivariate case* since higher dimensionality is difficult to visualize.



5.1 Joint Extreme Selection

Instead of selecting the extremes for each variable independently, the joint extremes should be selected together in order to capture the dependence between the 7 variables and to study the influence on the final design. A joint extreme is defined as a set of extreme values selected for a specified time interval Δt . A single *dominant* variable is chosen and the other variables of interest are labeled as *concomitant* variables. The Peak Over Threshold (POT) method is applied as illustrated in Section 4.1 to select the extreme observations, but only for the dominant variable (Zachary et al., 1998). For each extreme value of the dominant variable, the associated extreme values of the concomitant variables are retrieved from a pre-specified and fixed time interval. Figure 5.1 illustrates the joint extreme value selection process.

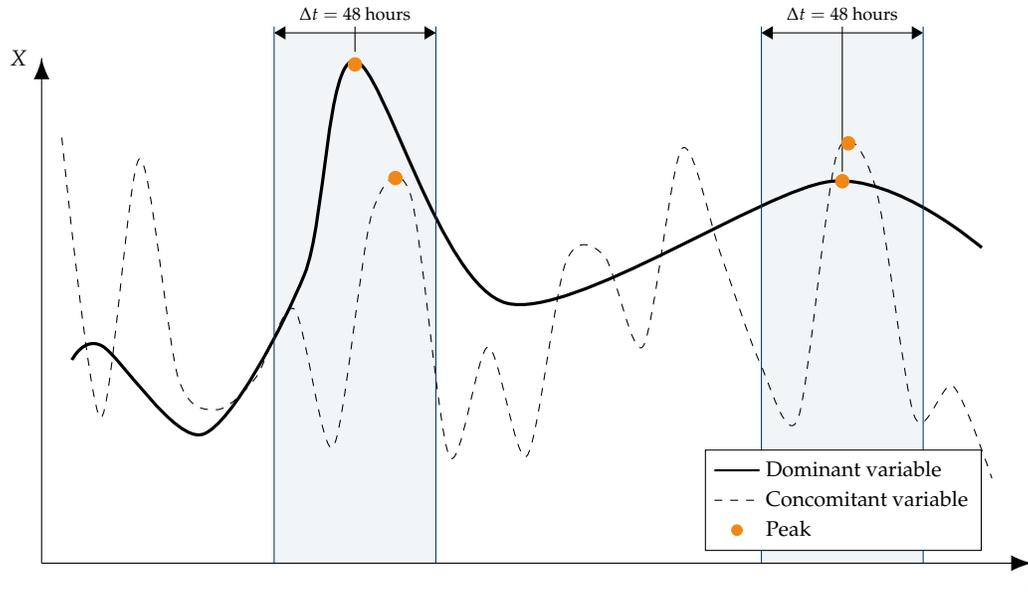


Figure 5.1: The peaks corresponding to the dominant variable are selected using a POT procedure. For each selected peak of the dominant variable, the corresponding peak of the concomitant variable is determined by looking for the extreme value within a time interval Δt . Note, the figure illustrates the joint extreme value selection for two variables, but the method can be applied for multiple concomitant variables.

The significant wave height (wind system) $H_{m0,wind}$ is chosen as the dominant variable because it is determined to be the most relevant variable for the breakwater system and the considered failure mechanisms. The dominant extremes have been obtained with the POT procedure with a threshold value and declustering time lag are identical as described in Section 4.1. The time interval in which the extremes of the concomitant variables are selected is chosen to be 48 hours. In this case, the extremes of the considered variables typically do not occur simultaneously.

5.2 Joint Distribution Modelling

The joint extremes are modeled by a multivariate statistical model called regular vine copula which models the joint occurrence of the 7 variables of interest. The selection of a specific regular vine copula out of all the possible configurations including labels can be quite challenging. One aims to select a vine copula that is most appropriate for modeling the multivariate dependence structure. The regular vine copula fit is optimized using maximum likelihood estimation based on the Akaike criterion. The *VineCopula* package (R programming language) provides a built-in function *RVineStructureSelect* which can be applied to select a reasonable candidate model with only a few seconds of computation time. *RVineStructureSelect* makes use of a maximum spanning tree algorithm and allows for optimization of the pair copulae specification (i.e. family type and parameters) for the edges of each tree. A graphical representation of the resulting vine copula is given in Figure 5.2.

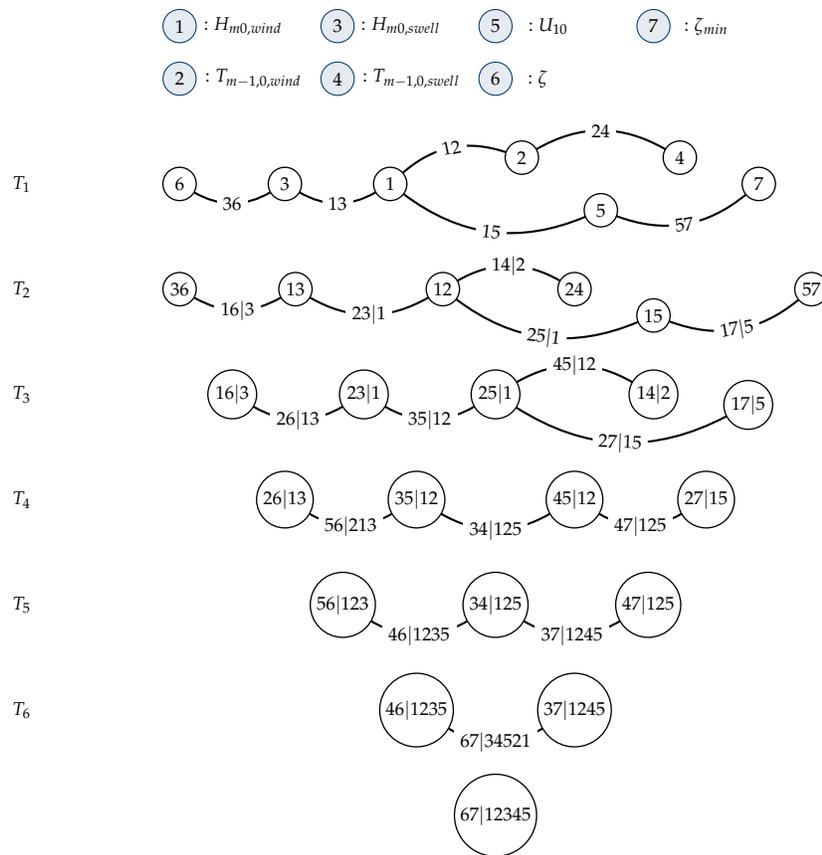


Figure 5.2: The graphical interpretation of the regular vine copula used to model the joint extremes of 7 variables of interest. Each edge in the regular vine copula represents a bivariate (conditional) copula and the edges in tree T_i will be the nodes in tree T_{i+1} . The first tree T_1 consists of 7 nodes representing the variables that are connected by edges which represent the bivariate copulae. The edge between node (6) and (3) reassembles the bivariate copula of the swell wave height $H_{m0,swell}$ and water level ζ .

It is difficult to evaluate the goodness of fit of a regular vine copula. The model fit is optimized by a maximum likelihood estimation procedure, however, this does not guarantee to result in the best available model. The quality of the model fit is further discussed in Chapter 7.

Besides modeling the joint occurrences, also information on the marginal distribution of the variables is required. A similar procedure is performed as described in Section 4.2, except that a suitable model is selected out of several statistical models such as the normal, lognormal, gamma, and Weibull for example (see Figure 5.3). The best model is selected based on the Akaike criterion and the goodness of fit is evaluated by means of a Q-Q plot, Pearson's coefficient, and the Kolmogorov Smirnov test. The results are presented in Table 5.1.

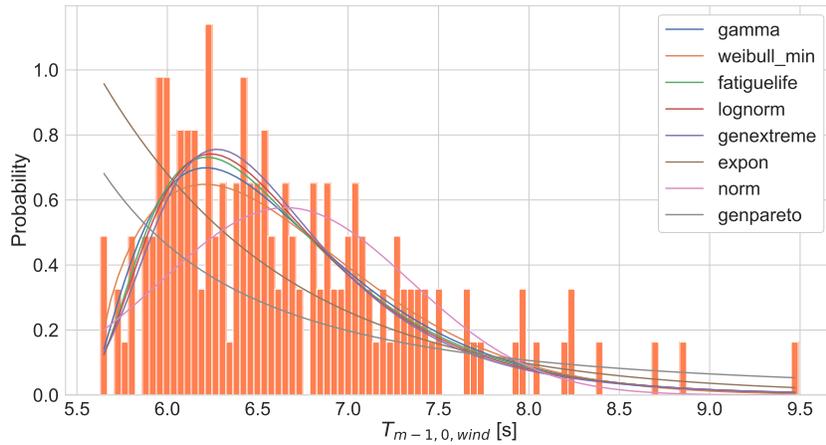


Figure 5.3: A histogram plot for the selected extremes of variable $T_{m-1,0,wind}$ together with the distribution fit of several statistical models. The distribution parameters are estimated using a Maximum Likelihood estimation and the gamma distribution is the most appropriate model according to the Akaike criterion.

	$H_{m0,wind}$	$T_{m-1,0,wind}$	$H_{m0,swell}$	$T_{m-1,0,swell}$	U_{10}	ζ	ζ_{min}
Model	Generalized Pareto	Gamma	Log-normal	Log-normal	Gamma	Log-normal	Weibull Minimum
Parameters:							
<i>shape</i>	-0.0845	2.493	0.133	0.296	3.507	0.123	7.378
<i>loc</i>	2.4	5.553	-1.967	5.305	10.982	-1.312	-0.0488
<i>scale</i>	0.990	0.442	3.620	2.922	1.004	1.660	1.010
R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99
p-value	0.94	0.99	0.49	0.98	0.83	0.65	0.96

Table 5.1: The distribution parameter estimates for the selected models. The distribution fit is statistically assessed by means of Pearson's correlation coefficient. The p-values correspond to the Kolmogorov Smirnov test and indicate that the selected models are suitable for modeling the extremes based on a 5% significance level.

5.3 Design Values

5.3.1 Offshore Design Values

Similar to the conventional approach, the next step is to derive the design values (extreme conditions) based upon an annual probability of occurrence once every 100 years (SLS) using the statistical distributions derived in the previous section. However, the derivation of design values is not as straightforward for the vine-based approach and the differences between both approaches will be addressed before evaluating the results. The keystone of the univariate derivation (Equation 4.2) is given below and certain aspects are marked to point out the differences with respect to the multivariate approach described in this chapter.

$$x_d = F_X^{-1}(1 - p) = F_X^{-1}\left(1 - \frac{1}{T N_s}\right).$$

Subsequently, for each marker, the difference is explained and the multivariate implementation is given. Note, for illustration purposes a bivariate example is used to illustrate the multivariate case, but an extension to higher dimensionality is rather straightforward.

1 One-to-one relation

In the conventional approach, univariate distributions are considered for which a one-to-one relation holds between the return period and a corresponding realization of the variable of interest (see Figure 4.5), however, this one-to-one relation no longer holds in the multivariate context. For illustrative purposes, the bivariate cumulative distribution $F_{x_1 x_2}$ is a surface as shown by the left figure in Figure 5.4, and the right figure shows the probability curves (isolines). Given that $x_1, x_2 \in \mathbb{R}$, an isoline, for example the 0.9 isoline, represents infinite combinations of (x_1, x_2) associated with a similar probability of (non-)exceedance. In other words, there is not a unique design event and a selection strategy is required to determine a single design event out of an ensemble of design events for the final design.

2 Joint Cumulative Density Distribution

In order to derive the set of design events, the cumulative probability function should be available. However, except for simple multivariate distribution models (e.g., Gaussian), analytic integration is very challenging and can be computationally expensive. The joint CDF of a regular vine copula should be approximated empirically, after which the design events corresponding to a specific safety level (return period) can be determined.

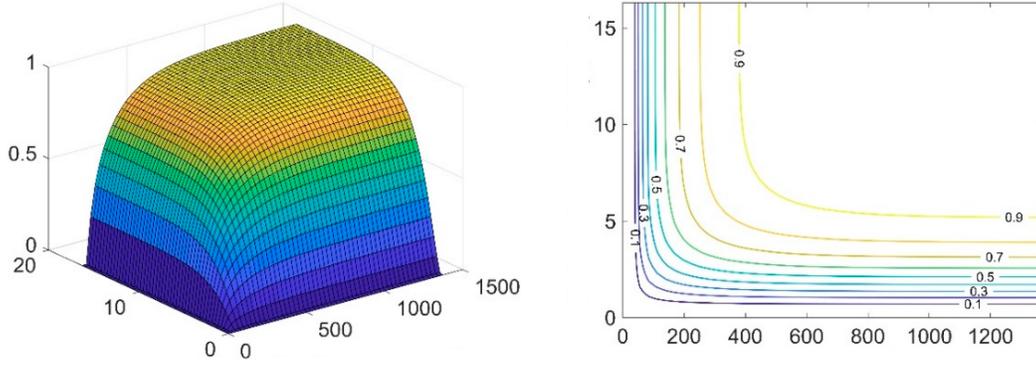


Figure 5.4: Bivariate cumulative probability distribution (left) and the corresponding probability isolines (right) unrelated to the considered case study and used for illustrative purposes only. *Note, the figure illustrates the bivariate cumulative distribution, but in fact the joint behavior of 7 variables is analyzed for the considered case.* Source: Zhong et al. (2020).

3 Multivariate Return Period

The return period is an appealing quantity as it provides an intuitive way to express the reliability of a structure. The statement: "This event E is expected to happen on average once each T years" is formally correct but also possibly misleading because the underlying probability p actually says that there is a probability p to observe the so-called T -year event each year. In the univariate context for independent and identically distributed variables, this often goes without any problems, but the shortcomings of the return period emerge in case a multivariate context is considered (Serinaldi, 2015). The following example illustrates this shortcoming and why it is preferred to reason in terms of probability p instead of return period T .

In the univariate case, the probability to end up in the critical region is uniquely defined as a random variable X exceeding a specific threshold x . This random variable X can be the significant wave height for example and the waves are considered to be critical in case they exceed a specific threshold. The probability is defined as:

$$\mathbb{P}(X > x) = 1 - F_X(x) = p, \quad (5.1)$$

where F_X is the CDF of X . In the conventional approach, this expression is used to derive the design value x for a specified acceptable failure probability p_f . In a multivariate case, there are multiple choices for the definition of the critical region such as the region defined by a simultaneous exceedance of the thresholds (AND scenario) or the region defined by the exceedance of at least one of the thresholds (OR scenario). Mathematically this is expressed by:

$$\begin{aligned} \mathbb{P}(X_1 > x_1 \cap X_2 > x_2 \cap \dots \cap X_m > x_m) &= p_{AND} \rightarrow T_{AND} \\ \mathbb{P}(X_1 > x_1 \cup X_2 > x_2 \cup \dots \cup X_m > x_m) &= p_{OR} \rightarrow T_{OR} \end{aligned}$$

where x_i is a specific threshold value for each of the m variables. The probabilities corresponding to the *AND* and *OR* scenario describe a different critical region indicated by the shaded area in Figure 5.5 for the *bivariate* case. The probabilities can be expressed in terms of the return period, however, reasoning in terms of T year return period easily leads to missing the meaning of the underlying probability scenario and making comparisons of values that seem to have a similar measurement unit, but actually describe incomparable mechanisms of failure. Therefore, it is preferred to reason in terms of probabilities since it allows us to bear in mind the underlying scenarios. The scenario used in the probability calculation should be based upon the considered system and its failure mechanisms.

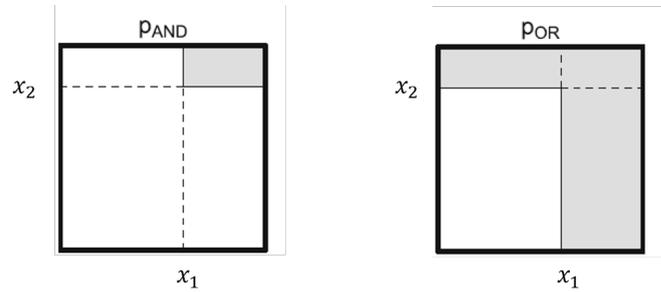


Figure 5.5: Domains and regions of the *AND* and *OR* probability for the bivariate situation. The bold black lines are the domain for which the probability is computed, whereas the shaded areas define the critical region satisfying the condition related to the type of probability. *Note, the figure illustrates the critical region described in the bivariate situation, but in fact the critical region determined by $m = 7$ variables resulting in a difficult to visualize hyperspace instead of a surface.* Source: Serinaldi (2015).

4 Definition Critical Region

The definition that should be used in the breakwater case study only depends on which one better describes the design requirements and mechanisms of failure. In terms of critical regions, *AND* and *OR* describe the probabilities associated with critical regions defined by a unique set of values (x_1, x_2, \dots, x_m) which is considered not to be applicable to a breakwater system. For example, a situation with a low water level and large wave height can be equally dangerous for the breakwater as a situation in which a high water level occurs together with a small wave height. Therefore, in the case of a breakwater system, a probability definition is required where many different joint events (instead of a unique set of values) relate to the same critical region. We implicitly deal with a system that is sensitive to and can fail due to a set of joint events characterized by the same joint probability of exceedance.

The probability definition satisfying this requirements is based on the so-called Kendall's measure described and illustrated by Salvadori et al. (2011, 2013, 2014, 2015). This is a definition where the probability is associated with the critical region defined by an infinite set of points lying on a t -level curve, see Figure 5.6. Kendall's probability is mathematically denoted as:

$$p_K = P(P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_m \leq x_m) > t).$$

Textually, the Kendall's probability gives the likelihood that a joint event occurs which belongs to the critical region defined by the quantile curve t .

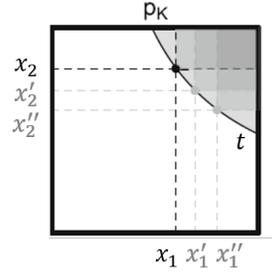


Figure 5.6: Domain and region of the probability based on Kendall's measure for the bivariate situation. The bold black lines are the domain for which the probability is computed, whereas the shaded areas define the critical region satisfying the condition related to the type of probability. *Source: Serinaldi (2015).*

Concluding remarks:

- Instead of reasoning in terms of the return period, it is preferred to reason in terms of probabilities, since it requires consideration of the underlying system and its failure mechanisms,
- The analytical expression for the joint cumulative distribution function is not available and thus an empirical approximation is obtained,
- In the multivariate context, there is no longer a one-to-one relation between the design value and a specific probability level, meaning there are infinite joint design values corresponding to the same probability level, instead of a unique design value.
- The definition used to describe the critical region in the breakwater case should only depend on which definition better describes the design requirements and mechanism of failure. The definition based on Kendall's measure is considered to be a suitable definition in this case study.

Multivariate Implementation

Given the acceptable annual probability of failure $P_{f,y}$, it is possible to derive an ensemble of joint events that have a similar joint exceedance probability for which we can evaluate the final design. A joint event is considered to be a set of realizations of load variables (wave height, period, wind speed, etc.). The set of design events corresponding to the probability quantile curve t is mathematically defined as:

$$\vec{x}_d = \{x \in \mathbb{R}^m : F_{X_1 X_2 \dots X_m}(x) = t\}. \quad (5.2)$$

The Kendall's quantile t is unknown and should be estimated in order to derive \vec{x}_d , which can be done using an algorithm referred to as algorithm 2, proposed by [Salvadori and Michele \(2010\)](#).

Algorithm 1 is used to find an estimate for Kendall's quantile t denoted by q_p such that the critical region (see the shaded area in Figure 5.6) equals the maximum acceptable annual failure probability:

$$P_{f,y} = \frac{1}{N_S T} = P(P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_m \leq x_m) > t),$$

Algorithm 1 Calculation of q_p

Input : Sample size n , probability level p , and number of simulations N .

Output : Quantile estimate q_p .

1: **Step 1 :** Initialization

2: Set critical index $k = \lfloor np \rfloor$ and model copula.

3:

4: **Step 2 :** Perform simulation

5: **for** $i = 1 : N$ **do**

6: $S \rightarrow C$

▷ Draw a sample S with size n from copula C

7: $u = C[S]$

▷ Obtain probabilities

8: $\hat{u} = \text{sort}(u)$

▷ Sort in ascending order

9: $E[i] = \hat{u}[k]$

▷ Store new quantile estimate

10:

11: **Step 3 :** Calculate final estimate

12: $q_p = \text{mean}(E)$

where N_ζ is the yearly average number of extremes. The estimate q_p is used to evaluate Equation 5.2 and find the corresponding design events. However, the analytical expression of the joint cumulative distribution $F_{X_1 X_2 \dots X_m}(x)$ is not available, and therefore, it is substituted by its empirical approximation $\hat{F}_{X_1 X_2 \dots X_m}(x)$. The design events corresponding to quantile curve estimate q_p are now obtained by means of numerical evaluation of the following expression:

$$\vec{x}_d = \{x \in \mathbb{R}^m : q_p - \epsilon \leq \hat{F}_{X_1 X_2 \dots X_m}(x) \leq q_p + \epsilon\},$$

where ϵ is a small constant. The expression is evaluated for a large ($N = 3 \cdot 10^5$) set of joint events randomly sampled from the regular vine copula, resulting in 44 joint design events described by Table 5.2.

	$H_{m0,wind}$ [m]	$T_{m-1,0,wind}$ [s]	$H_{m0,swell}$ [m]	$T_{m-1,0,swell}$ [s]	U_{10} [m/s]	ζ [+MN75]	ζ_{min} [+MN75]
Mean	5.96	8.69	2.76	10.80	20.04	0.75	-0.14
Standard deviation	0.64	0.55	0.31	0.70	2.02	0.13	0.05
Minimum	4.97	7.83	2.27	9.82	16.98	0.60	-0.29
Maximum	8.30	10.38	3.57	12.77	27.36	1.23	-0.07

Table 5.2: General statistics on the offshore conditions of the 44 design events. The offshore conditions for each design event specifically can be found in Appendix E.1.

The entire sample is presented by means of a bivariate scatterplot in Figure 5.7 and the black squares indicate the joint events that belong to the set of design events \vec{x}_d .

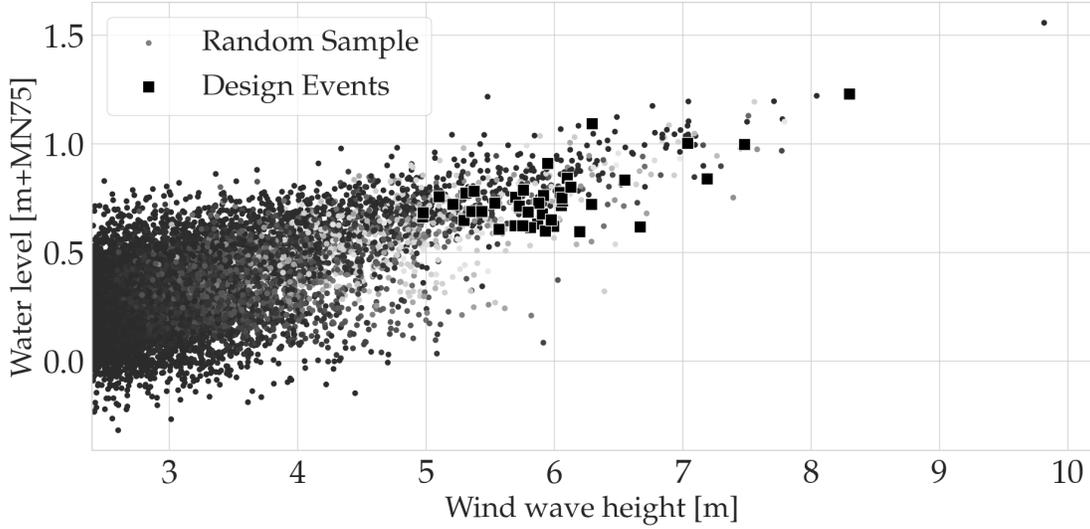


Figure 5.7: The joint occurrences of the significant wave height (wind system) $H_{s,wind}$ and the water level ζ . The black squares indicate the joint events that belong to the set of design events. *Note, the figure visualizes the sample by means of a scatterplot for 2 of the in total 7 dimensions for illustrative purposes.*

5.3.2 Nearshore Design Values

The swell and wind wave systems are combined before the offshore-nearshore transformation is performed by the numerical wave model SWAN. The systems are combined into a single system according to the following equations:

$$H_{m0}^{tot} = 4\sqrt{m_0^{tot}} = 4\sqrt{m_0^{wind} + m_0^{swell}} \quad (5.3)$$

$$T_{m-1,0}^{tot} = \frac{m_{-1}^{tot}}{m_0^{tot}} = \frac{m_{-1}^{wind} + m_{-1}^{swell}}{m_0^{wind} + m_0^{swell}} \quad (5.4)$$

$$T_p = 1.11 \cdot T_{m-1,0} \quad (5.5)$$

The offshore-nearshore transformation is performed for both the situation in which extremely high and extremely low water levels are considered. The resulting nearshore conditions are described in Table 5.3 and Figure 5.8.

	maximum water levels				minimum water levels			
	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]
Mean	2.36	9.41	10.07	0.85	2.17	9.35	9.99	-0.04
Standard deviation	0.17	0.53	0.59	0.13	0.10	0.55	0.60	0.05
Minimum	2.08	8.67	9.24	0.70	1.93	8.38	9.13	-0.19
Maximum	2.87	11.18	11.95	1.33	2.46	11.17	11.81	0.03

Table 5.3: General statistics on the nearshore design values derived using SWAN model and the design events described in Table 5.2. The nearshore conditions for each design event specifically can be found in Appendix F.1.

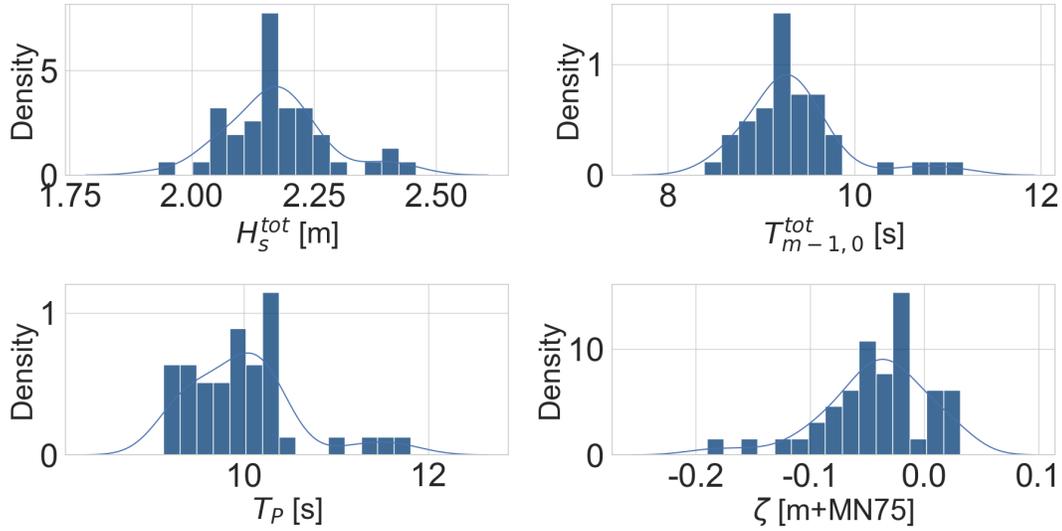


Figure 5.8: The distribution of the nearshore design values for the situation of extremely low water levels at output location *BT140*.

5.4 Final Design

The final step is to derive the design dimensions of the cross-sectional design by evaluating the limit state for the failure modes described in Appendix B with the nearshore conditions derived in Section 5.3.2. The resulting design dimensions are described by Table 5.4 and Figure 5.9.

Design dimension	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]
Mean	2.77	0.64	0.76	0.48
Standard deviation	0.33	0.05	0.08	0.02
Minimum	2.40	0.57	0.65	0.42
Maximum	4.07	0.78	1.02	0.53

Table 5.4: General statistics of the design dimensions derived based on the vine-based approach for output location *BT140*. The results should be interpreted as the minimal required crest level or nominal diameter in order to withstand the prescribed extreme storm conditions with a annual probability of occurrence once every 100 years. The final design corresponding to each design event specifically can be found in Appendix 6.7.

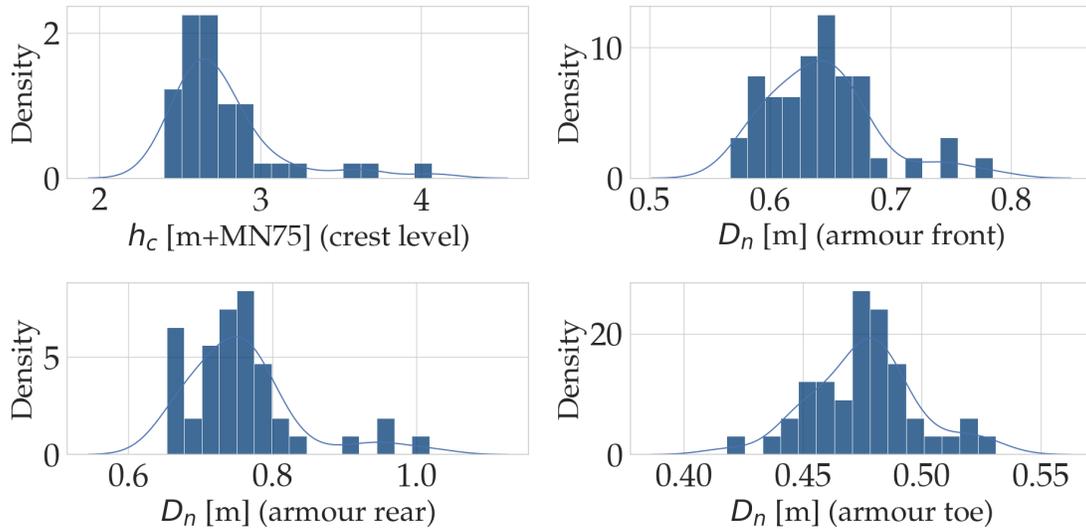


Figure 5.9: The distribution of the minimal required crest level and minimal nominal armour diameters derived according to the vine-based approach.

In contrast to the conventional approach, multiple cross-sectional designs of the breakwater are derived for the selected design event which all have a similar safety level. As can be observed, the obtained solutions differ and the challenge is to select a final design. There is no clearly defined method for the selection of a final design, but the engineer could use different selection strategies depending on the application:

- *Maximum design*
Select the most robust design based on all design events. Robust in the sense of selecting the largest design dimensions.
- *Most likely design*
Choose the final design based on the design event with the largest joint probability of occurrences.
- *Average design*
Choose the final design by averaging over all of the obtained designs.
- *Weight function*
A weight function could be used to aggregate the different designs into a final design.
- *Economic optimization*
The final design can be based on the economic optimum.

6 Results

This chapter considers a quantitative and qualitative comparison between the results obtained by the conventional approach and the vine-based approach. The design approaches consist of several procedures that have been implemented in Python version 3.8 and R programming language version 4.0.2. An HP ZBook Studio G5 with a 2.21 GHZ i7-8750H GPU processor and 16 GB RAM is used to perform the experiments. Section 6.1 contains a more extensive evaluation of the results derived in Chapter 4 and 5. Subsequently, Section 6.2 consists of a side-by-side comparison of the results for several experiments representing varying model configurations. Note, details on the general procedure of both approaches are described in Chapter 4 and 5.

6.1 Comparison of the Conventional & Vine-based Approach

In Chapter 4 and 5, the cross-sectional design of the breakwater is derived and the results are shortly addressed following both the conventional and vine-based approach. This section provides a side-by-side comparison and a more thorough analysis of the results. Table 6.1 presents the results consisting of a single cross-sectional design derived using the conventional approach and general statistics on the set of designs derived based on the vine-based approach.

	BT140				BT157			
Design dimension	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]
	Conventional Approach							
	4.07	0.81	1.03	0.53	3.66	0.67	0.90	0.34
	Vine-Based Approach (n=44)							
Mean	2.77	0.64	0.76	0.48	2.40	0.51	0.64	0.29
Standard deviation	0.33	0.05	0.08	0.02	0.29	0.03	0.06	0.01
Minimum	2.40	0.57	0.65	0.42	2.07	0.45	0.56	0.27
Maximum	4.07	0.78	1.02	0.53	3.54	0.62	0.85	0.33

Table 6.1: The breakwater cross-sectional design obtained according to the conventional approach together with some general statistics on the design dimensions obtained for the vine-based approach. The final design corresponding to each design event specifically can be found in Appendix 6.7.

It can be observed that the required design dimensions of the breakwater cross-section are larger for the output location *BT140* compared to output location *BT157* which is explained by the difference in water depth (0.9 meters). The wave height increases due to the shallower foreshore at location *BT140*. Furthermore, the required design dimensions according to the conventional approach, the conventional design, is larger than the maximum design dimensions derived according to the vine-based approach. Figure 6.1 shows a visualization of the results for output location *BT140* given by Table 6.1. It is observed that the conventional design ends up just outside of the domain of the 44 design events considered in the vine-based approach.

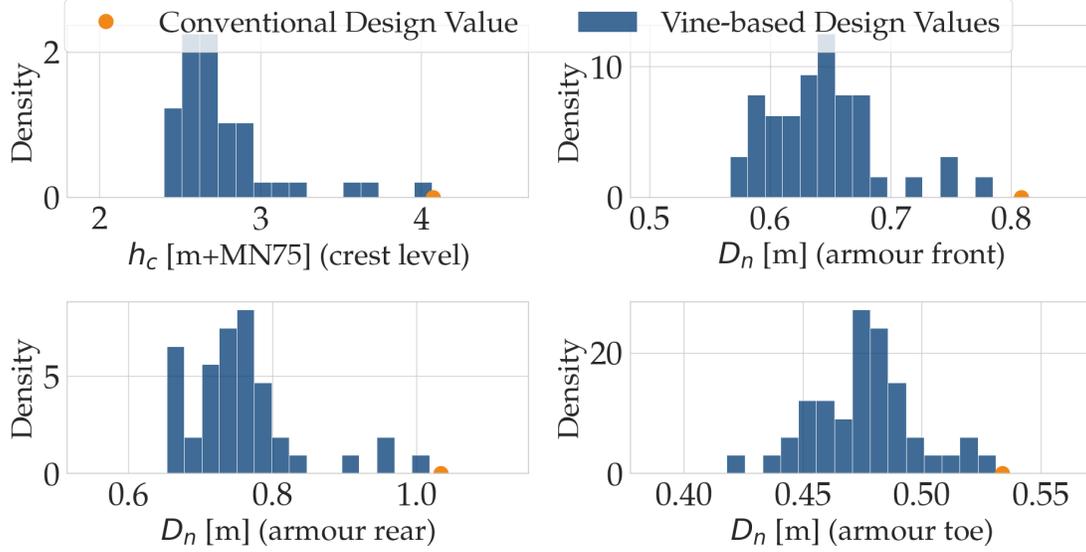


Figure 6.1: The distribution of the minimal required crest level and minimal nominal armour diameters derived according to the vine-based approach for output location *BT140* and an annual probability of occurrence one in 100. The orange dot indicates the design value derived according to the conventional approach.

The behavior of the system and a risk profile of the potential hazardous scenarios are evaluated by means of 2D scatterplots. Note, the considered system relates to the joint behavior of 7 variables, but 2D scatterplots are used here for their straight forward interpretation. These figures not only give an idea of the joint behavior of the variables but also illustrate the potential variability among the design events and the corresponding design dimensions. A complete collection of figures enables the identification of variables that play a dominant role in the final design of the breakwater. In this section, two figures have been included (see Figure 6.2 and 6.3) and the insights gained from these figures are described below. The remaining 2D scatterplots for other combinations of variables and different design dimensions are analyzed as well and can be found in the Appendix H.

Figure 6.2 shows the joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the water level ζ . The squares indicate the joint events that belong to the set of design events corresponding to the same safety level (see Section 5). The color gradient indicates the minimal required crest level for the corresponding design event. First of all, it can be observed that the considered variables are positively correlated which is explained by the fact that both variables are driven by the wind system. The water level and wave height will increase for increasing wind speed. Subsequently, the minimal required crest level for each design event is sorted out evenly (low crest level lower left corner, high crest level upper right corner) which indicates that the considered variables play a dominant role in the determination of the required crest level of the breakwater.

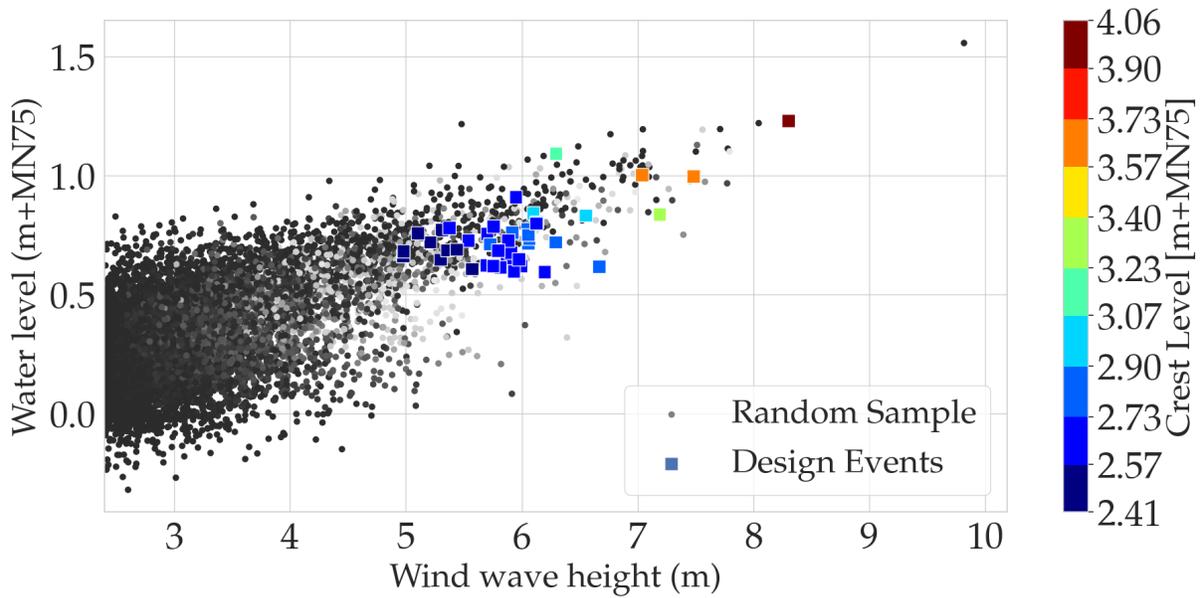


Figure 6.2: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the water level ζ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event. *Note, the figure visualizes the sample by means of a scatterplot for 2 of the total 7 dimensions for illustrative purposes.*

Figure 6.3 shows the joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the wind speed U_{10} . These two variables are by definition not correlated which is confirmed by the randomly scattered joint observations as depicted in Figure 6.3. Also, the required crest level is not sorted out as observed in Figure 6.2. Two design events can be distinguished above the 25 m/s wind speed mark (with a corresponding swell wave height of approximately 3m) of which the design event with lower wind speed appears to relate to a higher crest level. The design events are referred to as design event 12 and 36 (see Table E.1 in Appendix E) and the corresponding offshore and nearshore conditions are given by Table 6.2 and 6.3 respectively. It appears that the water level, wind wave height, and wave periods are in fact larger for design event 12. The combination of the actual wind speed and the consistency/duration of the wind direction is important for the resulting offshore/nearshore conditions. In this case, the situation with lower but more consistent wind speed results in more critical circumstances.

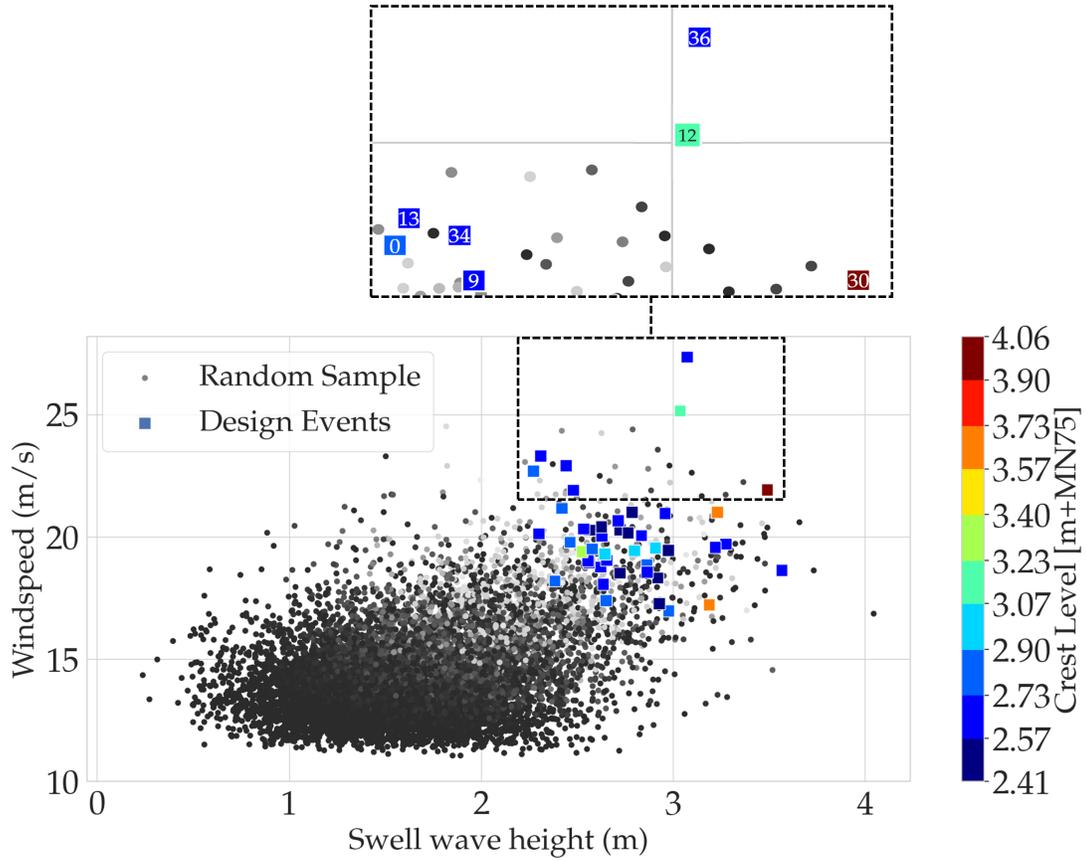


Figure 6.3: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event. Note, the figure visualizes the sample by means of a scatterplot for 2 of the total 7 dimensions for illustrative purposes.

Event	$H_{m0,wind}$ [m]	$T_{m-1,0,wind}$ [s]	$H_{m0,swell}$ [m]	$T_{m-1,0,swell}$ [s]	U_{10} [m/s]	ζ [+MN75]	ζ_{min} [+MN75]
12	6.29	9.03	3.04	11.34	25.15	1.09	-0.07
36	5.95	8.67	3.07	10.84	27.36	0.91	-0.08

Table 6.2: Information on the offshore conditions for design event 12 and 36.

Event	maximum water levels				minimum water levels			
	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]
12	2.31	9.58	10.31	1.19	2.11	9.70	10.31	0.03
36	2.08	9.12	10.18	1.01	1.93	9.30	10.18	0.02

Table 6.3: Information on the nearshore conditions for design event 12 and 36. Note, the situation with the minimum water levels is used for the design calculation of the minimal required diameter of the toe armour.

Some selection strategies have been applied in order to concretize the final design derived according to the vine-based approach. Additionally, appropriate armour classes are selected and subsequently, the costs are calculated for each design which includes the material, transportation, and installation costs. Table 6.4 gives an overview of the results focusing on the BT140 output location only since it is considered to be governing.

		BT140				
Design dimension	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Costs [€/m]	
Conventional Approach						
-	4.07m + MN75	$D_{n50} = 0.81\text{m}$ ($V = 1\text{ m}^3$)	$D_{n50} = 1.03\text{m}$ (HM 1000-4000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)	10300 €/m	
Vine-based Approach						
<i>Average design</i>	2.77m + MN75	$D_{n50} = 0.64\text{m}$ ($V = 1\text{ m}^3$)	$D_{n50} = 0.76\text{m}$ (HM 500-2000)	$D_{n50} = 0.48\text{m}$ (HM 300-600)	-17%	
<i>Maximum design</i>	4.07m + MN75	$D_{n50} = 0.78\text{m}$ ($V = 1\text{ m}^3$)	$D_{n50} = 1.02\text{m}$ (HM 1000-4000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)	0%	
<i>Min cost design</i>	2.40m + MN75	$D_{n50} = 0.59\text{m}$ ($V = 1\text{ m}^3$)	$D_{n50} = 0.67\text{m}$ (HM 500-2000)	$D_{n50} = 0.45\text{m}$ (LM 10-500)	-21%	

Table 6.4: The required cross-sectional design of the breakwater according to different design approaches based on prescribed extreme storm conditions with an annual probability of occurrence once every 100 years.

The Accropode II concrete armour units are applied for front armour of the breakwater based on the available armour stone in the region and economic considerations. The Accropode model with $D_{n50} = 1\text{m}$ ($V = 1\text{m}^3$) is applied in all final designs since it is the smallest available size. The range in the required nominal diameter of the rear armour for different selection strategies is 0.59 - 0.78m. The *maximum* selection strategy results in a HM 1000-4000 grading class while the other selection strategies result in two grading classes smaller at HM 500-2000. The latter is a sufficient grading class for the rear armour based on an average design event (up to a nominal diameter of 0.80m). The required nominal diameter of the *maximum* selection strategy just falls within HM 1000-4000 grading class (starting at 0.98m). The required nominal diameter of the toe armour stone is within the range of 0.45 - 0.53m and the selected grading class is HM 300-600 except for the *min cost* strategy.

The required crest level and rear armour size are significantly reduced for the *average* and *min-cost* selection strategy resulting in a cost reduction of 17% and 21% respectively. The total costs is mainly influenced by the total amount of used material which is directly determined by the crest level rather than the application of different grading classes because the cost prices of these different grading classes are in the same range (see Appendix C). In absolute sense, the *average* design strategy is closer to the *min cost* design compared to the *maximum* design and the *maximum* design is close to the conventional design. Applying the *maximum* selection strategy results in comparable cross-sectional design dimensions and a cost reduction of 0%.

6.2 Experiments

Since the results presented in Section 6.1 are based on a specific model configuration illustrated in Chapter 4 and 5, the question is raised whether the distinction between both approaches becomes more/less obvious when using varying model configurations. To this end, several experiments are formulated aiming to test two hypotheses, which are elaborated below:

1. *The conventional approach results in a conservative design compared to the vine-based approach for the break-water case.*
2. *The decisions related to the model configuration have a significant effect on the results of the vine-based approach.*

6.2.1 Experiment Set-up

In practice, an engineer is required to make a number of decisions during the design process which can be related to for example the data preparation, data modeling, or design calculations. To be more specific, this can be the selection of models used to model the selected extremes or setting specific parameter values for the design calculations. Until this point, the design process is illustrated according to a specific configuration based on how the project has been executed by Arcadis & Van Oord. Relevant decisions along the design process have been elaborated on in previous chapters.

One of the goals of this research is to evaluate the impact of the vine-based approach. It is argued to evaluate multiple configurations to obtain a more underpinned comparison with the conventional approach. Before it is possible to present and analyze the results of the experiments, it is necessary to elaborate on the considered configurations. Therefore, the relevant decision topics are identified for each component in the design process, see Figure 6.4. The following aspects are included in the experiments:

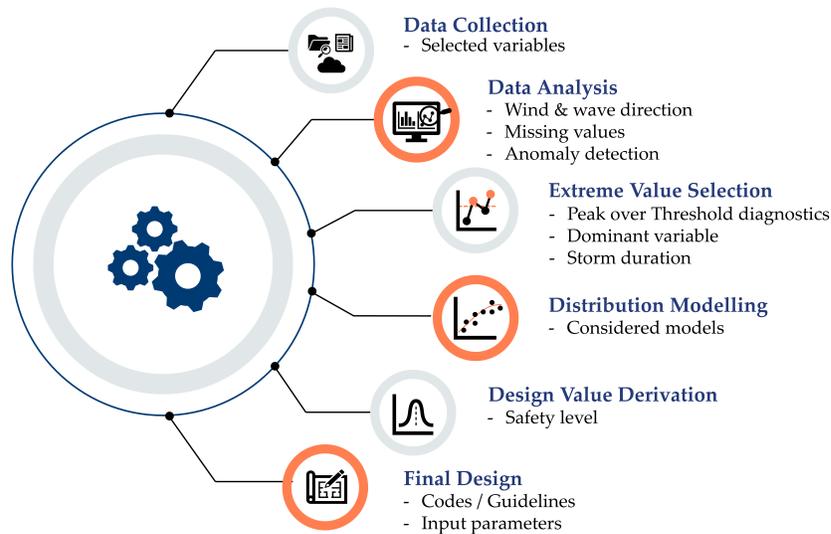


Figure 6.4: A schematic overview of important steps in the design process for the considered case study. For each step, relevant decision topics are identified that could influence the impact of the vine-based approach.

- **Dominant Variable** (*experiment 1.1*)

The dominant variable is used in the joint extreme sampling procedure and should be chosen based on the considered system. In this case, the wave height is seen as the most important load variable for the considered structure. The effect is evaluated when choosing the dominant variable to be the wave height corresponding to the swell sea state instead of the (more extreme in this case) locally generated waves i.e. the wind sea state.

- **Storm Duration** (*experiment 1.2*)

The storm duration is specified at the beginning of the joint extreme sampling procedure (see Figure 5.1) and represents the interval around an extreme value of the dominant variable in which the extreme values for the concomitant variables should be selected. A typical storm duration differs depending on the location and application. The experiment compares the result using two reasonable quantities for the storm duration parameter (48 hours and 72 hours).

- **Safety Level** (*experiment 1.3*)

The required safety level is expressed in terms of the return period and is often prescribed by the client or codes and guidelines. It is of interest to see how the vine-based approach performs for more extreme design scenarios.

- **Selected Variables** (*experiment 2.0 & 2.1*)

At the start of the analysis, the availability and quality of the required data should be judged by the engineer. Limited data is available on the water level variations for the considered case study, and an attempt has been done to generate a timeseries based on the available data instead of using fixed critical water levels.

- **Swell & wind wave components** (*experiment 3.0*)

In most applications, no distinction is made between different components of the wave spectrum i.e. wind and swell waves. It is most convenient to use the resultant wave parameters that are based on both components as the conventional approach does. The vine-based approach will no longer distinguish between different wave components as for this experiment only the resultant wave parameters are considered. This makes it possible to illustrate the effect including the dependence information by using vine copulae when comparing the results to the conventional design. Furthermore, it aims to highlight the effect of distinguishing between different wave components by comparing the results to the results obtained for experiment 1.0.

Experiment 1.0 serves as the starting point and is used as a baseline in comparison to other experiments. The baseline refers to the model configuration used to outline both approaches in previous chapters and the comparison of results in Section 6.1. Table 6.5 presents the experiments and the information on relevant decision topics.

	Dominant Variable	Storm Duration	Return Period	Excluded Variables
Experiment 1.0	$H_{m0,wind}$	48 hours	100 year	-
Experiment 1.1	$H_{m0,swell}$	48 hours	100 year	-
Experiment 1.2	$H_{m0,wind}$	72 hours	100 year	-
Experiment 1.3	$H_{m0,wind}$	48 hours	150 year	-
Experiment 2.0	$H_{m0,wind}$	48 hours	100 year	ζ & ζ_{min}
Experiment 2.1	$H_{m0,wind}$	48 hours	100 year	ζ_{min}
Experiment 3.0	H_{m0}	48 hours	100 year	<i>swell & wind components</i>

Table 6.5: The choices made for the selected decision topics for different experiments.

6.2.2 Experiment Results

The following section consists of an evaluation of the results in which the results of several experiments will be evaluated in comparison with the baseline (experiment 1.0). Table 6.6 presents general statistics of the design dimensions obtained for the vine-based approach for different experiments and output locations together with the design dimensions derived according to the conventional approach. Note, in addition to the design dimensions, the corresponding offshore and nearshore conditions can be found in Appendix I & J. From Table 6.6, it can be observed that the minimum required design dimensions are smaller for the *BT157* output location compared to the *BT140* output location. This holds for all experiments and can be explained by the difference in water depth (see the previous section).

Design dimension	BT140				BT157			
	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]
n = 44	Experiment 1.0				$(H_{m0,wind}, 48 \text{ hours}, 100 \text{ year}, -)$			
Conventional Approach	4.07	0.81	1.03	0.53	3.66	0.67	0.90	0.34
Mean	2.77	0.64	0.76	0.48	2.40	0.51	0.64	0.29
Standard deviation	0.33	0.05	0.08	0.02	0.29	0.03	0.06	0.01
Minimum	2.40	0.57	0.65	0.42	2.07	0.45	0.56	0.27
Maximum	4.07	0.78	1.02	0.53	3.54	0.62	0.85	0.33
n = 41	Experiment 1.1				$(H_{m0,swell}, 48 \text{ hours}, 100 \text{ year}, -)$			
Conventional Approach	4.07	0.81	1.03	0.53	3.66	0.67	0.90	0.34
Mean	1.95	0.51	0.55	0.40	1.65	0.40	0.46	0.25
Standard deviation	0.22	0.05	0.07	0.03	0.20	0.03	0.06	0.02
Minimum	1.55	0.42	0.42	0.34	1.28	0.33	0.35	0.21
Maximum	2.61	0.65	0.78	0.48	2.17	0.49	0.63	0.30
n = 39	Experiment 1.2				$(H_{m0,wind}, 72 \text{ hours}, 100 \text{ year}, -)$			
Conventional Approach	4.07	0.81	1.03	0.53	3.66	0.67	0.90	0.34
Mean	3.08	0.68	0.82	0.49	2.68	0.53	0.69	0.30
Standard deviation	0.35	0.05	0.08	0.02	0.31	0.03	0.07	0.01
Minimum	2.55	0.57	0.68	0.46	2.22	0.48	0.56	0.28
Maximum	4.07	0.80	1.01	0.53	3.61	0.63	0.86	0.33
n = 34	Experiment 1.3				$(H_{m0,wind}, 48 \text{ hours}, 150 \text{ year}, -)$			
Conventional Approach	4.21	0.82	1.06	0.54	3.79	0.68	0.93	0.34
Mean	2.80	0.65	0.77	0.48	2.43	0.52	0.65	0.30
Standard deviation	0.26	0.05	0.07	0.02	0.22	0.03	0.06	0.01
Minimum	2.18	0.51	0.58	0.42	1.91	0.44	0.49	0.27
Maximum	3.50	0.74	0.93	0.52	3.02	0.57	0.77	0.32
n = 34	Experiment 2.0				$(H_{m0,wind}, 48 \text{ hours}, 100 \text{ year}, \zeta \ \& \ \zeta_{min})$			
Conventional Approach	3.59	0.77	1.00	0.53	3.11	0.62	0.85	0.35
Mean	3.20	0.71	0.88	0.51	2.72	0.55	0.73	0.32
Standard deviation	0.22	0.03	0.07	0.02	0.16	0.02	0.05	0.01
Minimum	2.84	0.63	0.77	0.47	2.40	0.50	0.62	0.29
Maximum	3.78	0.78	1.06	0.54	3.14	0.58	0.86	0.33
n = 37	Experiment 2.1				$(H_{m0,wind}, 48 \text{ hours}, 100 \text{ year}, \zeta_{min})$			
Conventional Approach	4.07	0.81	1.03	0.53	3.66	0.67	0.90	0.35
Mean	3.48	0.72	0.90	0.51	3.03	0.57	0.75	0.32
Standard deviation	0.30	0.05	0.08	0.02	0.26	0.03	0.07	0.01
Minimum	2.98	0.63	0.75	0.48	2.56	0.52	0.65	0.30
Maximum	4.13	0.80	1.05	0.54	3.61	0.62	0.88	0.33
n = 57	Experiment 3.0				$(H_{m0}, 48 \text{ hours}, 100 \text{ year}, \text{combined})$			
Conventional Approach	4.07	0.81	1.03	0.53	3.66	0.67	0.90	0.34
Mean	2.83	0.66	0.77	0.48	2.48	0.53	0.66	0.30
Standard deviation	0.34	0.06	0.10	0.03	0.30	0.05	0.08	0.02
Minimum	2.32	0.56	0.65	0.44	2.01	0.45	0.54	0.26
Maximum	3.83	0.79	1.06	0.54	3.34	0.64	0.90	0.35

Table 6.6: General statistics of the design dimensions obtained for the vine-based approach for different experiments and output locations together with the design dimensions derived according to the conventional approach. The experiment number is given followed by the relevant information on the applied configuration (*Dominant Variable, Storm Duration, Return Period, and Excluded Variable(s)*).

Experiment 1.1

A significant decrease in the minimal required design dimensions is observed when comparing the results with experiment 1.0. This is explained by the fact that the swell system is not strongly correlated with the locally generated waves and wind speed ($\rho = 0.47$ and $\rho = 0.41$ respectively). The swell component is on average smaller than the locally generated waves and the influence of the wind speed on the nearshore conditions should not be underestimated. In other words, the multivariate sampling procedure relies on a dominant variable that is not considered to be the most important load variable resulting in a smaller cross sectional design that is in line with the expectations.

Experiment 1.2

It can be observed that the minimum required design dimensions are slightly increased in case a larger storm duration is used in the multivariate sampling procedure. This is in line with expectations because enlarging the interval from which the extremes of the concomitant variables are selected, will result at least in the same extremes or more extreme observations (see Section 5.1). The nominal diameters of the armour stone only increased by a few centimeters on average whereas the minimal required crest level increases by approximately 30 centimeters on average.

Experiment 1.3

The increase in safety level using a return period of 150 years instead of 100 years, results in 22 cm higher wave conditions offshore for the univariate case and on average 15 cm higher wave conditions for the multivariate case (see Appendix I). However, after offshore-nearshore propagation of the waves, only a small difference is visible in resulting nearshore wave conditions (see Appendix J) and as a consequence, the resulting design dimensions illustrate the same effect. Remarkably enough, the maximum observed design dimensions are in fact smaller than the maximum design dimensions observed for experiment 1.0. This phenomenon is explained later in the next section.

Experiment 2.0 & 2.1

Experiments 2.0 & 2.1 are conducted to evaluate the effect of the water level time series when modeling the joint behavior of the entire system. The 1.x experiments reassemble stochastic modeling on all 7 variables whereas the 2.x experiments exclude water level variables from the multivariate analysis and a deterministic approach is used instead. The fixed values used for the maximum water level and the minimum water level are $0.8m + MN75$ and $0.0m + MN75$ respectively based on the design report provided by Van Oord.

A significant increase of the required crest level and nominal armour diameter for the front and rear side of the breakwater is observed for both experiments. The offshore waves are on average 1 meter higher compared to the offshore conditions derived for experiment 1.0. The observed difference is explained by the low correlation of these variables with respect to the other variables and in fact, the minimum water level is correlated the least with the other variables (see Figure 6.5). Two bivariate quantile plots are used to illustrate the consequences of excluding the water level minima, see Figure 6.6 and 6.7.

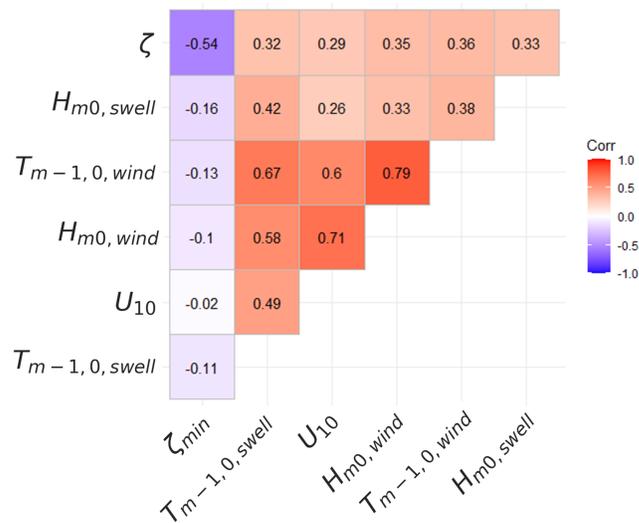


Figure 6.5: A correlation plot for the extremes obtained following the multivariate sampling procedure for the model configuration corresponding to experiment 1.0 and 1.3.

First of all, Figure 6.6 shows the joint distribution of the significant wave height (locally generated waves) and water level minima. The variables have a weak correlation and the observations appear to be randomly distributed. The quantiles indicate that the extremely low water levels do not necessarily correspond to extreme wave heights whereas the contrary is observed for the correlated variables significant wave height and mean wave period, see Figure 6.7. Let's assume that the highest quantile (yellow) corresponds to the same probability of non-exceedance for both examples (100 year return period). The bivariate extremes sampled from these quantiles show different domains for the extreme significant wave height. Figure 6.6 shows that for the highest quantile, the significant wave height is distributed on a relatively large domain [4, 8] containing moderate wave heights as well. The domain of the significant wave height is much more limited towards the extreme situations in the case of the stronger correlated variables (see Figure 6.7). This is an illustration in which two variables are modeled, however, it can be generalized to a multivariate context.

This case perfectly illustrates the extra information obtained when applying the vine-based approach and modeling the entire system by considering the joint occurrences of relevant variables. The dependence structure of the variables which can include various levels of correlations is modeled using vine copula instead of combining the individual extremes and assuming perfect (linear) correlation.

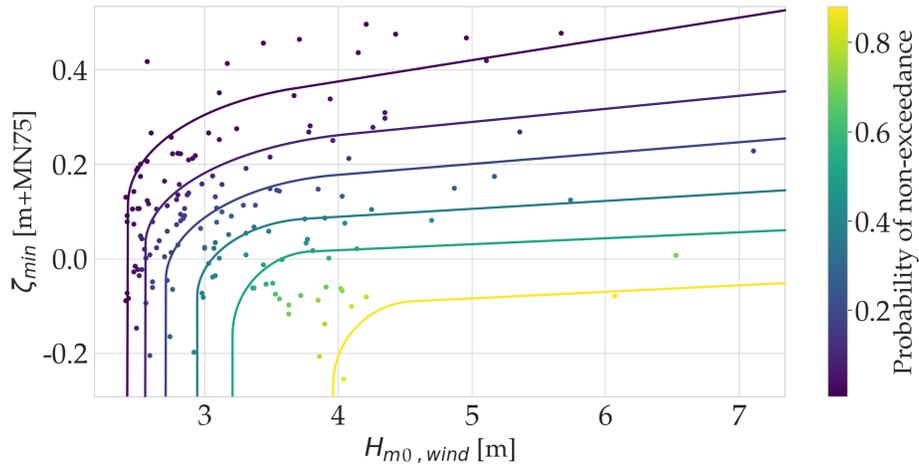


Figure 6.6: The joint extreme value distribution including probability quantiles for variables $H_{m0,wind}$ and ζ_{min} obtained following the multivariate sampling procedure illustrating weak correlation.

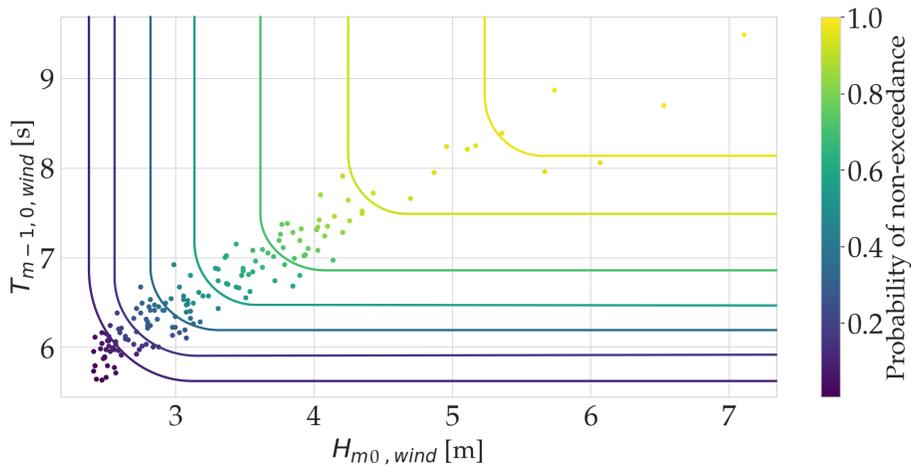


Figure 6.7: The joint extreme value distribution including probability quantiles for variables $H_{m0,wind}$ and $T_{m-1,0,wind}$ obtained following the multivariate sampling procedure illustrating strong correlation.

Experiment 3.0

An increase of the design dimensions by a few centimeters is observed compared to the results of experiment 1.0. The decision to distinguish different wave components does not result in a significant difference in the final design for the considered case study. Therefore, it can be concluded that the difference between the traditional and vine-based approaches can be assigned mainly to the usage of dependence information by incorporating vine copulae in the design process.

6.2.3 Final Design

A final cross-sectional design for the vine-based approach is derived using three different selection strategies. In contrast to the conventional approach, multiple cross-sectional designs of the breakwater are derived for the selected design event which all have a similar safety level. Table 6.7 presents the final cross-sectional design derived for different experiments including the required armour classes and corresponding costs. Experiment 1.0 refers to the model configuration discussed in Section 6.1 and is included here for comparison purposes.

The results presented by Table 6.7 are for the most part in line with the results presented by Table 6.6 however, additional insights are obtained. For example, it appears that the largest cost reduction is obtained for model configuration applied in experiment 1.1 in which the $H_{m0,swell}$ variable is selected as the dominant variable. The waves corresponding to the swell component do not relate to the most critical sea state for the considered application and therefore, the resulting cross-sectional design is smaller (less robust) compared to experiment 1.0. However, it does highlight the importance of selecting an appropriate dominant variable.

A remarkable observation is the increased cost reduction for experiment 1.3 using a return period of 150 years with respect to a return period of 100 years in experiment 1.0. The *maximum* and *min-cost* selection strategy both result in a less robust design for experiment 1.3 compared to the result derived for these strategies in experiment 1.0 which is counter-intuitive since a higher safety level (larger return period) is used. This is expected to be an effect of the small sample size for both experiments ($n = 34$ and $n = 44$ respectively).

When comparing the results of experiments 2.0 and 2.1, a similar observation is made. In general, it can be stated that excluding the water level variables results in a less distinct difference between both the conventional and vine-based design. Furthermore, the required design dimensions turn out larger compared to the required design obtained in experiment 1.0. A maximum cost reduction is obtained for the *min-cost* selection strategy. Applying the *maximum* selection strategy results in a more expensive and robust design in comparison to the conventional approach.

Lastly, the results obtained for experiment 3.0 are comparable to the results obtained for experiment 1.0 in both the obtained design and corresponding cost reduction. It can be concluded that the difference in the final design resulting from both the conventional and vine-based approach is not explained by differentiating between different wave systems.

BT140					
Design dimension	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Costs [€/m]
n = 44	Experiment 1.0 ($H_{m0,wind}$, 48 hours, 100 year, -)				
Conventional Approach	MN75 + 4.07m	$D_{n50} = 0.81m$ ($V = 1 m^3$)	$D_{n50} = 1.03m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	10300 €/m
Average design	MN75 + 2.77m	$D_{n50} = 0.64m$ ($V = 1 m^3$)	$D_{n50} = 0.76m$ (HM 500-2000)	$D_{n50} = 0.48m$ (HM 300-600)	-17%
Maximum design	MN75 + 4.07m	$D_{n50} = 0.78m$ ($V = 1 m^3$)	$D_{n50} = 1.02m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	0%
Min cost design	MN75 + 2.4m	$D_{n50} = 0.59m$ ($V = 1 m^3$)	$D_{n50} = 0.67m$ (HM 500-2000)	$D_{n50} = 0.45m$ (LM 10-500)	-21%
n = 41	Experiment 1.1 ($H_{m0,swell}$, 48 hours, 100 year, -)				
Conventional Approach	MN75 + 4.07m	$D_{n50} = 0.81m$ ($V = 1 m^3$)	$D_{n50} = 1.03m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	10300 €/m
Average design	MN75 + 1.95m	$D_{n50} = 0.51m$ ($V = 1 m^3$)	$D_{n50} = 0.55m$ (HM 300-600)	$D_{n50} = 0.4m$ (LM 10-500)	-26%
Maximum design	MN75 + 2.61m	$D_{n50} = 0.65m$ ($V = 1 m^3$)	$D_{n50} = 0.78m$ (HM 500-2000)	$D_{n50} = 0.48m$ (HM 300-600)	-18%
Min cost design	MN75 + 1.55m	$D_{n50} = 0.42m$ ($V = 1 m^3$)	$D_{n50} = 0.42m$ (LM 10-500)	$D_{n50} = 0.34m$ (LM 10-500)	-31%
n = 39	Experiment 1.2 ($H_{m0,wind}$, 72 hours, 100 year, -)				
Conventional Approach	MN75 + 4.07m	$D_{n50} = 0.81m$ ($V = 1 m^3$)	$D_{n50} = 1.03m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	10300 €/m
Average design	MN75 + 3.08m	$D_{n50} = 0.68m$ ($V = 1 m^3$)	$D_{n50} = 0.82m$ (HMA 1000-3000)	$D_{n50} = 0.49m$ (HM 300-600)	-13%
Maximum design	MN75 + 4.07m	$D_{n50} = 0.8m$ ($V = 1 m^3$)	$D_{n50} = 1.01m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	0%
Min cost design	MN75 + 2.55m	$D_{n50} = 0.57m$ ($V = 1 m^3$)	$D_{n50} = 0.68m$ (HM 500-2000)	$D_{n50} = 0.46m$ (LM 10-500)	-19%
n = 34	Experiment 1.3 ($H_{m0,wind}$, 48 hours, 150 year, -)				
Conventional Approach	MN75 + 4.21m	$D_{n50} = 0.82m$ ($V = 1 m^3$)	$D_{n50} = 1.06m$ (HM 2000-4000)	$D_{n50} = 0.54m$ (HM 300-600)	10500 €/m
Average design	MN75 + 2.8m	$D_{n50} = 0.65m$ ($V = 1 m^3$)	$D_{n50} = 0.77m$ (HM 500-2000)	$D_{n50} = 0.48m$ (HM 300-600)	-18%
Maximum design	MN75 + 3.5m	$D_{n50} = 0.74m$ ($V = 1 m^3$)	$D_{n50} = 0.93m$ (HMA 1000-3000)	$D_{n50} = 0.52m$ (HM 300-600)	-9%
Min cost design	MN75 + 2.18m	$D_{n50} = 0.51m$ ($V = 1 m^3$)	$D_{n50} = 0.58m$ (HM 300-600)	$D_{n50} = 0.42m$ (LM 10-500)	-25%

BT140					
Design dimension	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Costs [€/m]
n = 34	Experiment 2.0 ($H_{m0,wind}$, 48 hours, 100 year, ζ & ζ_{min})				
Conventional Approach	MN75 + 3.59m	$D_{n50} = 0.77m$ ($V = 1 m^3$)	$D_{n50} = 1.0m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	9700 €/m
Average design	MN75 + 3.2m	$D_{n50} = 0.71m$ ($V = 1 m^3$)	$D_{n50} = 0.88m$ (HMA 1000-3000)	$D_{n50} = 0.51m$ (HM 300-600)	-5%
Maximum design	MN75 + 3.78m	$D_{n50} = 0.78m$ ($V = 1 m^3$)	$D_{n50} = 1.06m$ (HM 2000-4000)	$D_{n50} = 0.54m$ (HM 300-600)	+3%
Min cost design	MN75 + 2.84m	$D_{n50} = 0.63m$ ($V = 1 m^3$)	$D_{n50} = 0.77m$ (HM 500-2000)	$D_{n50} = 0.47m$ (LM 10-500)	-11%
n = 37	Experiment 2.1 ($H_{m0,wind}$, 48 hours, 100 year, ζ_{min})				
Conventional Approach	MN75 + 4.07m	$D_{n50} = 0.81m$ ($V = 1 m^3$)	$D_{n50} = 1.03m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	10300 €/m
Average design	MN75 + 3.48m	$D_{n50} = 0.72m$ ($V = 1 m^3$)	$D_{n50} = 0.9m$ (HMA 1000-3000)	$D_{n50} = 0.51m$ (HM 300-600)	-7%
Maximum design	MN75 + 4.13m	$D_{n50} = 0.8m$ ($V = 1 m^3$)	$D_{n50} = 1.05m$ (HM 2000-4000)	$D_{n50} = 0.54m$ (HM 300-600)	+1%
Min cost design	MN75 + 2.98m	$D_{n50} = 0.64m$ ($V = 1 m^3$)	$D_{n50} = 0.79m$ (HM 500-2000)	$D_{n50} = 0.48m$ (HM 300-600)	-14%
n = 57	Experiment 3.0 (H_{m0} , 48 hours, 100 year, combined systems)				
Conventional Approach	MN75 + 4.07m	$D_{n50} = 0.81m$ ($V = 1 m^3$)	$D_{n50} = 1.03m$ (HM 1000-4000)	$D_{n50} = 0.53m$ (HM 300-600)	10300 €/m
Average design	MN75 + 2.83m	$D_{n50} = 0.66m$ ($V = 1 m^3$)	$D_{n50} = 0.77m$ (HM 500-2000)	$D_{n50} = 0.48m$ (HM 300-600)	-16%
Maximum design	MN75 + 3.83m	$D_{n50} = 0.79m$ ($V = 1 m^3$)	$D_{n50} = 1.06m$ (HM 2000-4000)	$D_{n50} = 0.54m$ (HM 300-600)	-3%
Min cost design	MN75 + 2.32m	$D_{n50} = 0.57m$ ($V = 1 m^3$)	$D_{n50} = 0.65m$ (HM 500-2000)	$D_{n50} = 0.44m$ (LM 10-500)	-22%

Table 6.7: The results conducted for multiple experiments evaluating several model configurations of the vine-based approach. The design dimensions of the cross-sectional design are given according to 3 selection strategies and are compared to the conventional design. The results should be interpreted as the minimal required crest level or nominal diameter in order to withstand the prescribed extreme storm conditions with an annual probability of occurrence once every 100 years.

6.3 Concluding Remarks

Two design approaches applied in the design process of a breakwater have been compared in terms of the resulting cross-sectional design. It was found that in general larger dimensions are required for the design near output location *BT140* compared to output location *BT157* which is explained by the difference in water depth (0.9 meters). Based on several experiments representing different model configurations, it can be concluded that the vine-based design approach on average results in minimal required dimensions of elements of the cross-sectional design that turn out to be smaller compared to the results obtained with the traditional approach. In some cases, a breakwater design is obtained with a cost reduction in the order of 25%. Note, that an even larger cost reduction of 31% is obtained for experiment 1.1, however selecting $H_{m0,swell}$ as the dominant variable is not seen as realistic in this case.

In order to evaluate the effect of varying model configurations, several experiments have been formulated aiming to test two hypotheses, which are elaborated below:

1. *The conventional approach results in a conservative design compared to the vine-based approach for the breakwater case.*

The validity of this hypothesis is tested by comparing the results presented in Table 6.1. A design with corresponding reliability that exceeds the target reliability to a great extent can be referred to as a conservative design. Given that the designs presented in Table 6.1 have similar reliability, it can be concluded that the vine-based approach in general results in a less conservative design. Perhaps more important, it should be stressed here that the obtained results are based on more realistic load scenarios since they are based on the interrelations of the variables given that the vine-based approach is properly applied. Thus besides the insights regarding the potential conservatism of the traditional approach, it also provides extra information about the behavior of the system and contributes to a clearer understanding of hazardous scenarios.

2. *The decisions related to the model configuration have a significant effect on the results of the vine-based approach.*

The experiments stand for different model configurations where for each experiment only a single component has changed to observe an isolated effect of the changed parameter. The hypothesis that the decisions related to the model configuration have a significant effect on the result of the vine-based approach is substantiated for all experiments, for the considered breakwater case.

7 Sensitivity Analysis

In this chapter, an exploratory study is performed on the uncertainty of the vine-based approach in the derivation of design values. It is important to be aware of uncertainties in the application of the design approach and perhaps more important, in the evaluation and interpretation of the results. Especially in the early stages of development, the uncertainty can impose a limit on the confidence of the output of the model. A qualitative evaluation of the confidence in the model is provided here. Potential sources of uncertainty will be addressed, and for a selection of sources, recalculations of the model are performed under alternative assumptions to assess the impact.

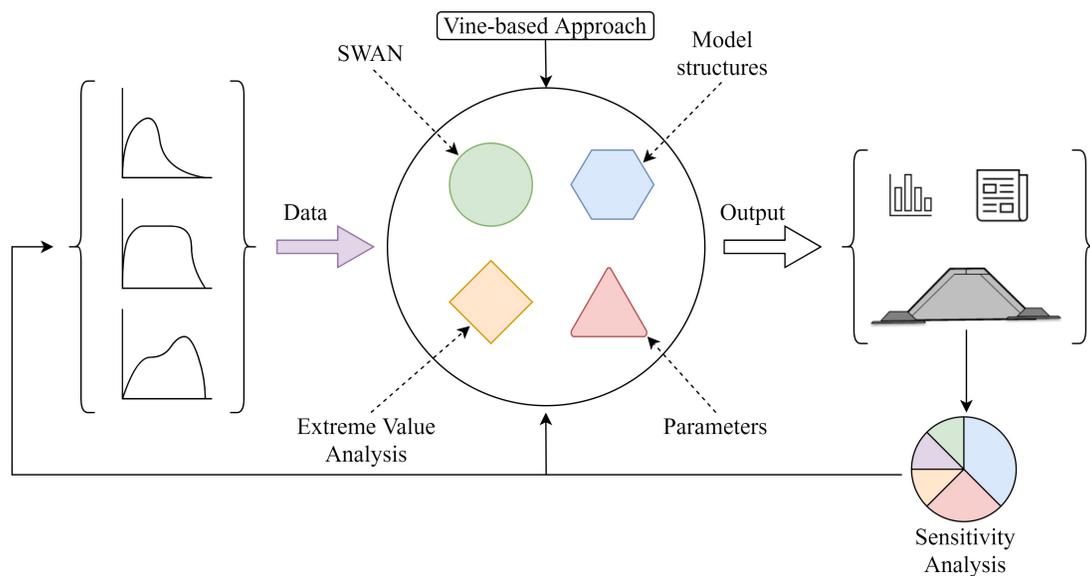


Figure 7.1: The most important sources of uncertainty are indicated in an abstract outline of the applied design procedure. The goal of a sensitivity analysis is to identify and quantify potential sources of uncertainty and the impact on the final solution.

The goal of a sensitivity analysis is to identify and quantify potential sources of uncertainty such as data quality, alternative model structures, and parameter estimation procedures. Figure 7.1 illustrates a possible scheme for the sensitivity analysis for the considered research. Although a thorough sensitivity analysis was not within the scope of this research, the urge was to make the practitioner aware of potential sources of uncertainty and elaborate on a few specific aspects in more detail. Therefore, the uncertainty is divided and allocated at different scale levels, zooming in from *data* towards the *model* and lastly the *parameter* space (see Figure 7.2).

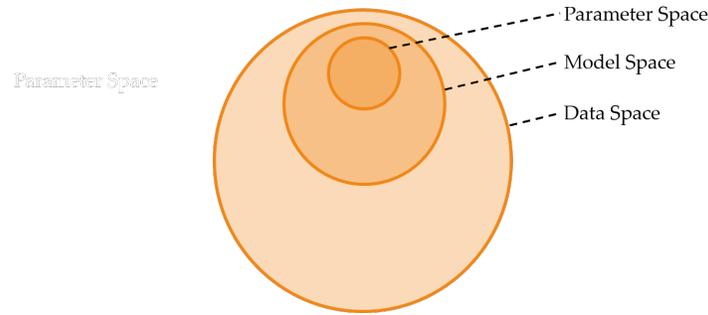


Figure 7.2: Multiple scale levels of uncertainty.

7.1 Data Space

The obtained results are always conditional on the available data. The availability of sufficient and good quality data is necessary to properly model relevant physical processes such as water level variation or wave conditions, and ultimately determine characteristic values used in the design process. Good quality data with high accuracy is often not readily available and might severely restrict the application of procedures such as Peak Over Threshold. A data time series over an extensive period of time is preferred since the extreme behavior of the variables is of interest. In case only a time series of a couple of years is available, it will limit the number of selected extremes which might result in a poor fit of the applied univariate and multivariate models.

The question remains what is sufficient and good quality data? When does the application of the vine-based approach become cumbersome and when does the practitioner obtain poor output because of poor input? This is an important topic in the applicability of the vine-based methodology, however, these questions are hard to answer in a general sense.

An experiment has been conducted where the impact of the length of the time series is evaluated. A 20-year long time series of meteocean data is used for the design of the Constanza coastal protection project. The available time series is conducted over a period 01-01-1993 - 31-12-2012 on a 6-hourly interval. An experiment has been performed assuming a situation in which only a time series is available for a period 01-01-2003 - 31-12-2012. The results are presented in Table 7.1. It can be observed that in general, the cross-sectional design is somewhat smaller for all dimensions when only considering data conducted in the period 01-01-2003 - 31-12-2012. This is in line with the expectations namely, a limited period of time probably results in capturing less severe extremes upon which the final design is based, and therefore, the minimal required cross-sectional dimensions turn out to be somewhat smaller.

		BT140				
Design dimension	Crest level h_c [m + MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Costs [€/m]	
n = 44		Data period: 01-01-1993 - 31-12-2012 (<i>Experiment 1.0</i>)				
Conventional Approach	4.07m + MN75	$D_{n50} = 0.81\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 1.03\text{m}$ (HM 1000-4000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)	10300 €/m	
<i>Average design</i>	2.77m + MN75	$D_{n50} = 0.64\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 0.76\text{m}$ (HM 500-2000)	$D_{n50} = 0.48\text{m}$ (HM 300-600)	-17%	
<i>Maximum design</i>	4.07m + MN75	$D_{n50} = 0.78\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 1.02\text{m}$ (HM 1000-4000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)	0%	
<i>Min cost design</i>	MN75 + 2.4m	$D_{n50} = 0.59\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 0.67\text{m}$ (HM 500-2000)	$D_{n50} = 0.45\text{m}$ (LM 10-500)	-21%	
n = 40		Data period: 01-01-2003 - 31-12-2012				
Conventional Approach	3.9m + MN75	$D_{n50} = 0.76\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 0.97\text{m}$ (HMA 1000-3000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)	10100 €/m	
<i>Average design</i>	2.7m + MN75	$D_{n50} = 0.63\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 0.72\text{m}$ (HM 500-2000)	$D_{n50} = 0.47\text{m}$ (LM 10-500)	-16%	
<i>Maximum design</i>	3.8m + MN75	$D_{n50} = 0.76\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 0.95\text{m}$ (HMA 1000-3000)	$D_{n50} = 0.53\text{m}$ (HM 300-600)	-1%	
<i>Min cost design</i>	2.29m + MN75	$D_{n50} = 0.57\text{m}$ ($V = 1 \text{ m}^3$)	$D_{n50} = 0.63\text{m}$ (HM 500-2000)	$D_{n50} = 0.45\text{m}$ (LM 10-500)	-21%	

Table 7.1: The results obtained using the entire time series of 20 years and only the first 10 years to investigating the influence of the amount of data on the final cross-sectional design. The design dimensions of the cross-sectional design are given according to 3 selection strategies and are presented together with the conventional design. The results should be interpreted as the minimal required crest level or nominal diameter in order to withstand the prescribed extreme storm conditions with an annual probability of occurrence once every 100 years.

7.2 Model Space

First, to prevent any misconceptions, the term “model” is used here to refer to the result of modeling the relevant environmental variables by means of mathematical functions defined by specific parameter values. For example, the generalized Pareto distribution used to model a single variable such as the significant wave height or regular vine copula to model the joint occurrence of a selection of variables. The appropriate model parameters are derived by means of optimizing the overall model fit based on the Akaike Information Criterion (AIC). The AIC is an estimator of the relative quality of statistical models for a given set of data. It deals with a trade-off between the goodness of fit for a certain model and its simplicity. The quality of the fit is dependent on the available data and therefore, the model is as good as the quality of the data together with the flexibility of the considered models. In general, it is difficult to find a model that can be proven to be “valid” in any absolute sense and the topic of model testing inevitably involves subjective issues and is often a trade-off between accuracy and costs.

Perfect accuracy is impossible since a model is only an abstraction of the system it represents. While reducing modeling errors is very important, a balance should also be sought between overall accuracy and other factors. These include development time, computation time, and the cost of developing the model in relation to the expected benefits. In this section, the emphasis will be on modeling the multivariate system by means of a regular vine copula. As explained in Section 2.4, the vine copula is a joint probability density

function consisting of (conditional) pair copulae. The model space becomes very large for increasing number of variables with $N_{vines} = \binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$ for $n = 6$, this already results in 7,776,000 models. Choosing the best model by evaluating all possible models is computationally expensive.

The *VineCopula* package (R programming language) provides a built-in function *RVineStructureSelect* which can be applied to select a reasonable candidate model with only a few seconds of computation time. *RVineStructureSelect* makes use of a maximum spanning tree algorithm and allows for optimization of the pair copulae specification (i.e. family type and parameters) for the edges of each tree. However, one is not sure whether the resulting model is the true best model without complete enumeration of all possible models, and one might wonder what influence a sub-optimal model will have on the final results. There is no widely accepted technique for testing the goodness of fit of a multivariate model which is technically more complicated because, for example, the joint distribution might be asymmetric and contains multiple peaks. Two relatively simple and straight forward methods can be applied to evaluate the accuracy and stability of the resulting vine copula:

1. Correlations

Draw a sample from the resulting vine copula, calculate the correlations and compare with the actually observed correlations of the selected extremes. The model is assumed to model the data properly if it captures the most important correlations and no significant difference in correlations are observed between the sample and the actual data. The results are presented in Table 7.2 and it can be concluded that the differences are minor providing an indication that the chosen model captures the correlations correctly.

	$T_{m-1,0,wind}$ [s]	$H_{m0,wind}$ [m]	$T_{m-1,0,swell}$ [s]	U_{10} [m/s]	ζ [m+MN75]	$H_{m0,swell}$ [m]	ζ_{min} [m+MN75]
$T_{m-1,0,wind}$ [s]		-	-	-	-	-	-
$H_{m0,wind}$	2,0		-	-	-	-	-
$T_{m-1,0,swell}$ [s]	2,0	4,2		-	-	-	-
U_{10} [m/s]	5,3	3,3	8,0		-	-	-
ζ [m+MN75]	1,2	2,4	0,2	1,8		-	-
$H_{m0,swell}$ [m]	0,3	0,1	0,7	3,2	0,7		-
ζ_{min} [m+MN75]	1,5	0,8	0,5	0,4	6,2	1,4	

Table 7.2: The absolute difference ($\times 100$) of the Kendall's τ correlation coefficients for the actual data and a sample obtained from the vine copula.

2. Bootstrap

The advantage of bootstrapping is its simplicity and the straightforward way to derive estimates of relevant statistics. Here, it is used to check the stability of the results. From the sample of joint extremes, randomly draw $n = 159$ samples with replacement (case resampling) and apply *RVineStructureSelect* to find the "best" vine copula based on the random sample. Sampling with replacement basically means that a particular joint extreme observation from the original sample might be sampled more than once. Repeat this procedure many times ($N = 1000$) and evaluate the variation in the resulting fit based on the AIC. Hereby, it is assumed that the original sample of joint extremes is a good representation of the underlying population.

Figure 7.3 shows the distribution of the AIC of the bootstrapped multivariate models. The 95% confidence interval is determined and the bounds are indicated by the black dashed lines. Although it is impossible to know the true confidence interval, bootstrapping gives a reasonable estimate and it can be observed that the fit of the used model falls within the upper and lower bound of the interval. Furthermore, the resulting overall model fit based on the AIC looks stable for the performed simulations based on Figure 7.3. However, this is still a subjective observation and it is recommended to evaluate the stability in more detail by evaluating the final results based on the models corresponding to the bounds of the confidence interval.

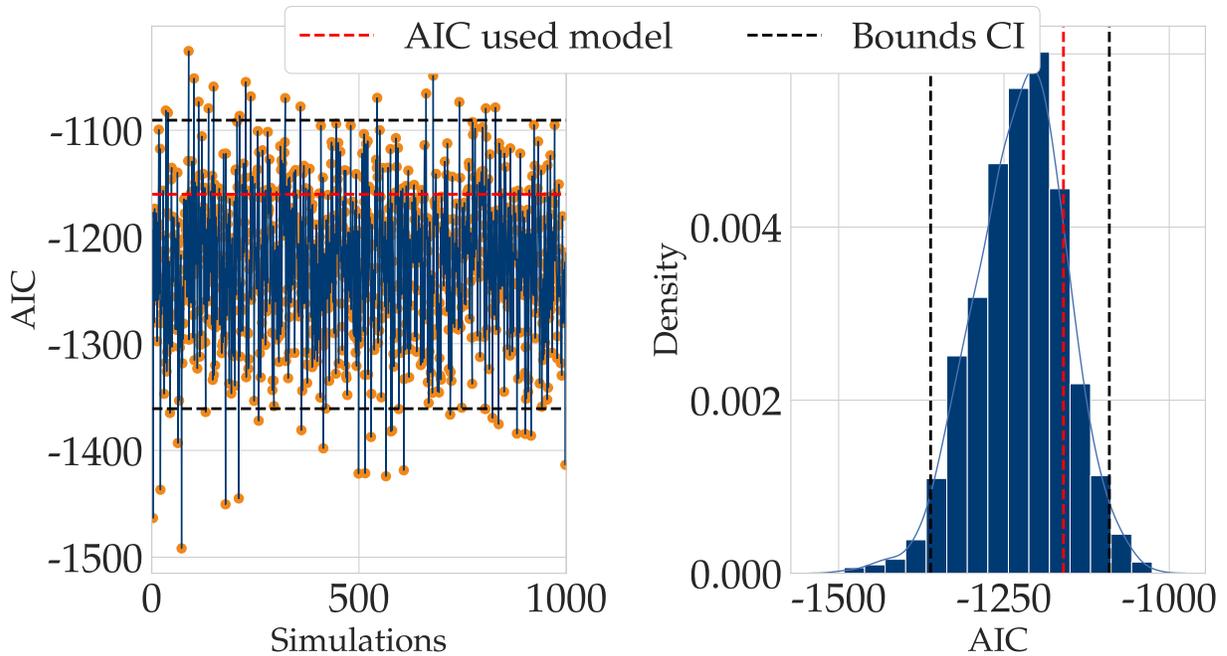


Figure 7.3: Distribution of the bootstrapped multivariate model fits expressed by the AIC. The empirical confidence intervals 95% are indicated by the black dashed lines and the red dashed line indicates the AIC of the actual used model.

7.3 Parameter Space

Three types of parameters are distinguished based on how their parameter values are derived:

- Expert Knowledge,
- Maximum Likelihood Estimates (MLE),
- Simulation.

Expert Knowledge

Certain parameters are not estimated by means of MLE, but are determined by the engineer based on expert knowledge. For example, certain parameters (interval δ , threshold u , etc.) required for the Peak over Threshold (PoT) procedure are based on a number of indicators as illustrated in Appendix D. It is very likely that different parameter values will be obtained in case several engineers follow the exact same procedure independently, because they may interpret the results slightly different based on their personal experience.

Maximum Likelihood Estimation

Parameters of the univariate and multivariate statistical models are determined by MLE which is a method of estimating parameters by maximizing a likelihood function. Given a statistical model, a combination of parameter values is determined for which the observed data is most probable. The likelihood function can be evaluated by different optimization algorithms depending on the applied software which might have an influence on the resulting parameter values. Also, these algorithms have convergence criteria meaning the algorithms terminate in case only minor improvements in solution quality are realized, and therefore it is possible that the solution will be a local optimum instead of the global optimum. In this research, the MLE has been applied using the Nelder-Mead simplex algorithm (Nelder and Mead, 1965) to find the optimum with an acceptable convergence tolerance of 10^{-4} for both the parameter value and likelihood. The uncertainty of the parameter values can be investigated by using different estimation algorithms and/or convergence criteria.

Simulation

As described in Section 5.3, Kendall's measure is used to derive the critical events in a multivariate context for a prescribed safety level. In order to evaluate the Kendall's probability, it is required to estimate the quantile curve t denoted by q_p which is done using Algorithm 2 proposed by Salvadori and Michele (2010) (see Algorithm 1).

The estimation of quantile curve t involves some decisions that might influence the accuracy of the obtained quantile estimate. Therefore, a number of experiments are conducted to evaluate the uncertainty. First of all, the experiments aim to evaluate the influence of parameter values for the number of samples N , and the sample size n . Secondly, for most regular vine copula no analytical expression is unavailable, and instead, an empirical estimation is used in Algorithm 1. Therefore, experiments are conducted using a multivariate model with a known analytical expression (multivariate Gumbel copula) in order to investigate whether the empirical estimation affects the convergence and the final estimate of the quantile curve. Lastly, the influence of the number of included variables is investigated.

N	1000	10000	100	1000	10000	100
n	1.0E+04	1.0E+04	1.0E+05	1.0E+05	1.0E+05	1.0E+06
q_p	0.9907	0.9907	0.9915	0.9916	0.9916	0.9917
t	28s	281s	29s	317s	3044s	280s

Table 7.3: The quantile estimate q_p for different combinations of sample size n and number of simulations N . The experiments have been conducted using the analytical copula expression and including the 6 most correlated variables.

Firstly, the influence of the parameter values for n and N are evaluated. The experiment considers the 6 most correlated variables of the original dataset described in Chapter 3 (all variables except water level minima). Algorithm 1 is applied assuming the data can be modeled by a Gumbel copula with known analytical expression. From the results presented in Table 7.3, it can be concluded that the computation time increases linearly for increasing parameter value n and N . Furthermore, a difference in the quantile estimate q_p can be observed when increasing the sample size from $1e4$ to $1e5$. This can be explained by the probability used in this case which is $p = 1 - (1/100)/7.95 = 0.9987421$. In each simulation, the k -th position ($k = \lfloor np \rfloor$) of the ordered vector with non-exceedance probabilities is selected resulting in $\hat{u}[9987]$ and $\hat{u}[99874]$ for $n = 1e4$ and $n = 1e5$ respectively.

Based on these results presented in Table 7.3, it is preferred to apply Algorithm 1 with $n = 1e5$ and $N = 1e4$. Next, the following experiments are conducted to investigate the effect of using an empirical approximation in case the analytical expression is unknown. It also evaluates the influence of the number of considered variables. Note, that the number of simulations N is set to 100 for experiments 3 and 4 due to computation time.

1. Analytical, 3 most correlated variables, $n = 1e5$, $N = 1e4$, (see Figure 7.4)
2. Analytical, 6 most correlated variables, $n = 1e5$, $N = 1e4$, (see Figure 7.5)
3. Empirical, 3 most correlated variables, $n = 1e4$, $N = 100$, (see Figure 7.6)
4. Empirical, 6 most correlated variables, $n = 1e4$, $N = 100$, (see Figure 7.7)

The following observations are made based on the results presented in Figure 7.4, 7.5, 7.6, and 7.7:

- The observed variability for the experiment with 6 dimensions is larger in contrast to the experiment with 3 dimensions in case the empirical approximation is applied (see Figure 7.6 and 7.7),
- As expected, a non-linear increase of computation time is observed in case the empirical approximation is applied,
- Regarding the convergence and accuracy of the final estimate, it is preferred to apply parameter values $n = 1e5$ and $N = 1e4$ in the quantile estimation. However, it is considered to be computational too expensive for the scope of this project and instead, parameter values $n = 1e4$, $N = 100$ have been applied resulting in a computation time of approximately 15 hours. Based on Figure 7.4 and 7.5, the quantile estimate becomes relatively stable around $N=100$,
- The observed error in the quantile estimate for 3 dimensions and 6 dimensions is 0.06% and 0.12% respectively.

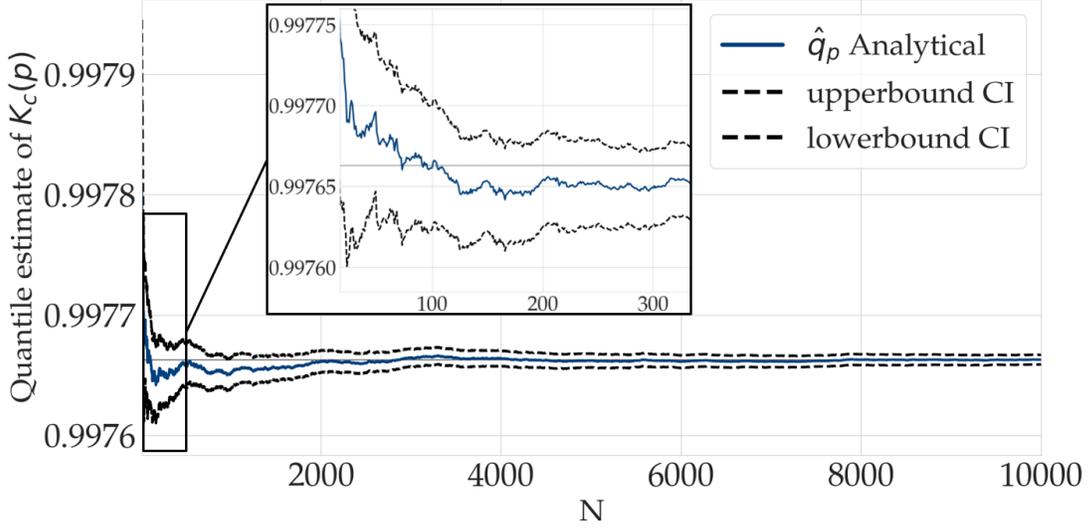


Figure 7.4: The convergence of quantile estimate q_p for increasing number of simulations. The black dashed lines indicate the confidence bounds for 2 times the sample standard deviation. The experiment is conducted using the analytical expression for the Gumbel copula, sample size $n = 1e5$ and, 3 most correlated variables.

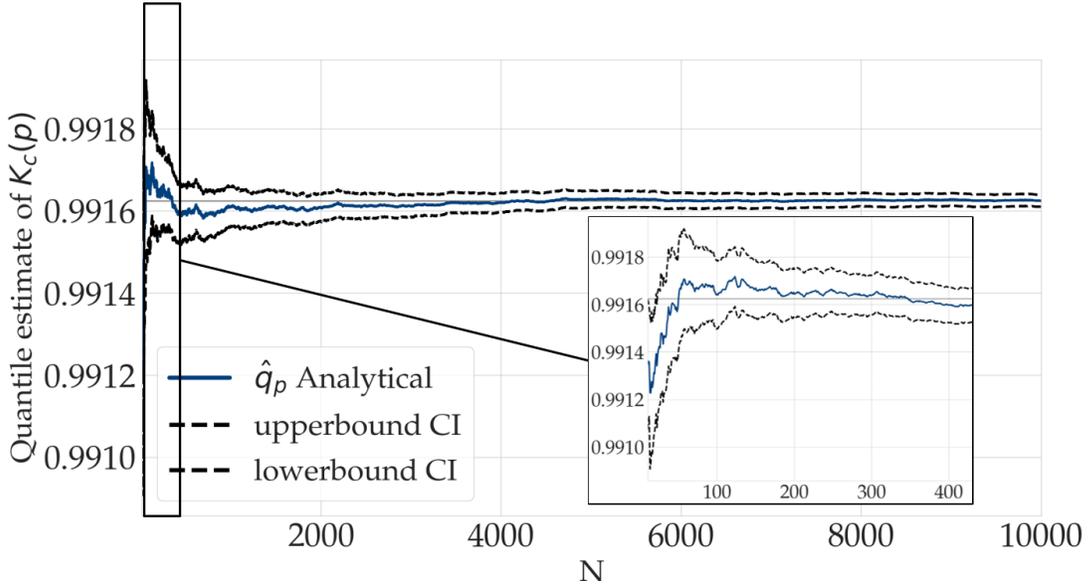


Figure 7.5: The convergence of quantile estimate q_p for increasing number of simulations. The black dashed lines indicate the confidence bounds for 2 times the sample standard deviation. The experiment is conducted using the analytical expression for the Gumbel copula, sample size $n = 1e5$ and, 6 most correlated variables.

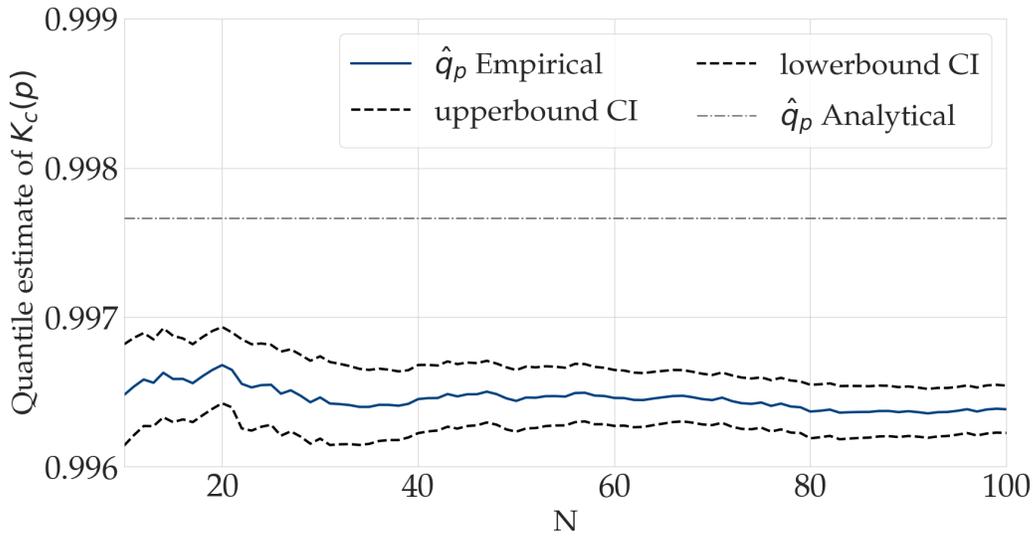


Figure 7.6: The convergence of quantile estimate q_p for increasing number of simulations. The black dashed lines indicate the confidence bounds for 2 times the sample standard deviation whereas the gray dashed line presents the quantile estimate determined by experiment 1. The experiment is conducted using the empirical copula approximation, sample size $n = 1e4$ and, 3 most correlated variables.

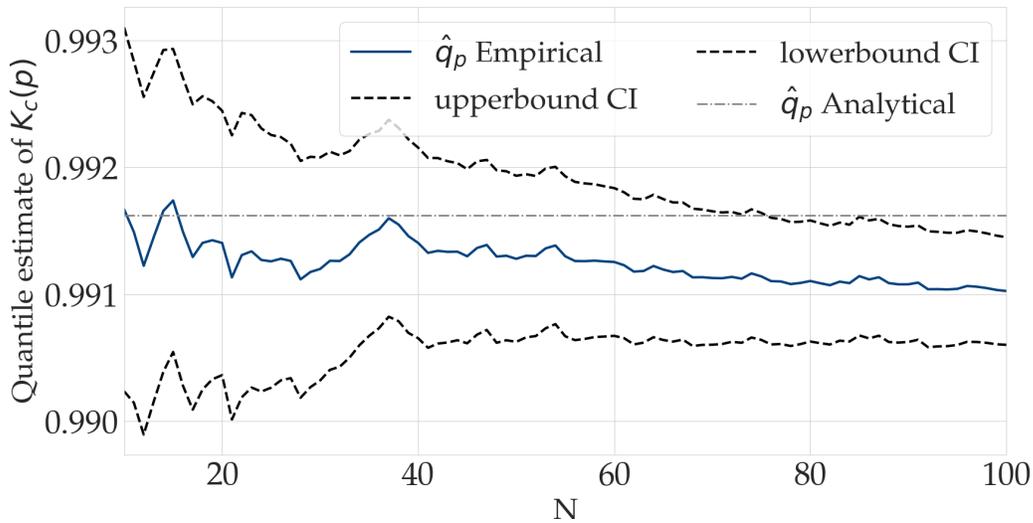


Figure 7.7: The convergence of quantile estimate q_p for increasing number of simulations. The black dashed lines indicate the confidence bounds for 2 times the sample standard deviation whereas the gray dashed line presents the quantile estimate determined by experiment 2. The experiment is conducted using the empirical copula approximation, sample size $n = 1e4$ and, 6 most correlated variables.

8 Discussion

8.1 Relevance

The goal of the thesis is first of all to perform a practical assessment to investigate how the vine-based approach can be incorporated in the design procedure of coastal infrastructure. Secondly, it is aimed to investigate how usage of the dependence structure can contribute to additional insights compared to the traditional approach assuming independence. The vine-based approach has been successfully incorporated and several experiments demonstrate how extra information can be obtained and how it impacts the cross-sectional design of a breakwater. In line with the hypothesis that the traditional approach is expected to result in a conservative design, it can be concluded that generally speaking the vine-based approach leads to a cross-sectional design that is on average smaller compared to the design obtained according to the traditional approach for a similar imposed reliability. In some cases, a breakwater design is obtained with a cost reduction up to 25%.

It is expected that if the results of the vine-based approach are presented to the client, he/she would most probably go for the most conservative design. Still, that means a less conservative design is obtained compared to the traditional approach for most of the experiments conducted in this research. But in this case less conservative means that the minimal required nominal armour diameters and crest level are reduced by only a few centimeters. For example, in case of experiment 1.0 using the most conservative selection strategy *maximum design*, it is determined that the required crest level is reduced by 0 cm, the front armour diameter by 3 cm, the rear armour diameter by 1 cm, and the toe armour diameter by 0 cm compared to the conventional design. In the end, for the design of a breakwater, coastal infrastructure in general, it is not so much about optimizing the design for a couple of centimeters because an appropriate armour class should be selected and the diameter of subsequent armour classes can differ a couple of decimeters. The final design of experiment 1.0 in terms of selected armour grading classes and crest level is the same as the conventional design and the corresponding cost reduction is determined to be 0.

One might wonder whether it is worth the effort to apply the vine-based approach in this case. Engineers confirmed that even if the design is not directly optimized, it provides a good insight into the behavior of the system and a risk profile of the potential hazardous scenarios. Also, it enables to assess the degree of conservatism of the conventional design because currently, this is unknown. Furthermore, engineers acknowledge it is perhaps too early for the whole industry to move to this approach, but at least you have an idea of the joint behavior and how it compares to the conventional approach.

The results of the assessment do not advocate a replacement of the traditional approach by the vine-based approach, but the assessment performed in this research rather provides an insight into how to use the underlying data structure in gaining more knowledge about and confidence in the system behavior. This analysis can be performed in combination with the traditional approach when deriving the required design.

Although the assessment is really valuable for the understanding and successful application of the vine-based approach, it is difficult to generalize the results because it is expected that the results are very much dependent on the application. For example, the relevant failure mechanisms vary for different types of structures which may be sensitive to other load conditions. Neither is it possible to generalize the obtained results for breakwater structures itself because another important aspect is the location and the corresponding characteristic environmental conditions. For example, consider the same breakwater structure located at the Atlantic Ocean with multiple swell systems compared to the Black Sea with its locally generated (wind-driven) waves.

At the beginning of this research, a literature study has been performed to investigate the existing knowledge on the topic of multivariate frequency analysis in the design process of coastal infrastructure. Multiple studies show the possibilities and availability of multivariate methods. Although multivariate frequency analysis currently receives much attention within the academic community, advanced statistical techniques such as regular vine copulae are slow in being taken up in engineering practice. In order to bridge this gap, this study provides an illustration of applying advanced statistical techniques from a conceptual level. It is decided not to focus specifically on the mathematical or practical aspect, but rather using a conceptual point of view combining all these aspects showing the engineering community how and why they should apply these techniques and pointing out the benefits and/or drawbacks of the theoretical concepts to the academic community.

In conclusion, the contributions of this research to the existing knowledge are as follows:

- An illustration of how regular vine copula can be used in the design process of a breakwater and the identification of advantages and disadvantages initiating future research topics,
- Providing a quantitative and qualitative comparison between the traditional approach and the multivariate approach,
- Illustrating how to deal with the problem of identifying design events in a multivariate context.

Also, this research can contribute to the development of new codes and guidelines for the design of coastal infrastructure. At the time of writing this thesis, an updated version of the Eurocode is under development which is going to include the design of coastal infrastructure. The actual design calculations will be comparable to the Rock Manual, but so far during the development stage of the new Eurocode, there appears to be a lot of discussions especially on the topic of return period and wave conditions in the context of applying joint statistics. Practical assessments of multivariate statistical models like the assessment performed in this research are able to shed light on complicated aspects and can enhance the development of these guidelines and codes.

8.2 Case Study

During the case study a couple of assumptions and simplifications have been made to reduce complexity:

- In our application, the principal direction of the waves has been determined and only observations that fall within a directional sector (principal direction ± 30 degrees) are considered in the frequency analysis. Another option is to split the wind and wave data in multiple directional sectors and perform a frequency analysis per sector.
- The offshore data time series does not provide any information on the water level variation. Multiple other sources have been used to generate a water level time series that matches the recording period and frequency of the offshore metocean time series. It is assumed that the simulated water level time series is a sufficient representation of the actual unknown water level variation. However, it is advised being careful with the usage of simulated data as the water level variation does affect the results as shown by experiments 2.1 and 2.2.
- Typically different limit states will be considered in the design of (coastal) infrastructure such as the Operational Limit State (OLS), Serviceability Limit State (SLS), and the Ultimate Limit State (ULS). In this research, only the SLS is considered for sake of simplicity. The other limit states can be evaluated in a similar fashion, but different values are assigned to certain parameters used in the evaluation of the limit states.
- The design decisions made by Van Oord are used as a guide for the case study applied in this research. For example, multiple failure mechanisms exist for a breakwater, however, in line with the design document only the following four failure modes are evaluated in this case study; *overtopping*, *armour stability (front and rear)*, and *toe stability*. Furthermore, it is decided by Van Oord that the Accropode II concrete armour units are applied on the sea side of the breakwater instead of regular armour stone. The minimal available size of these concrete units is 1 m^3 which is already over-dimensioned and as a result, little variation in the required minimal armour size is observed for the conducted experiments.

8.3 Limitations & Challenges

Joint Extremes

The extension from univariate to multivariate approach is not straightforward. First of all, the definition of a joint extreme is a difficult issue since there is no natural way of ordering a set of real-valued numbers corresponding to relevant variables such as significant wave height, mean wave period, and wind speed. In line with the work performed by Sellés Valls (2019), the implemented joint extreme selection approach is based on the joint extreme definition proposed by Zachary et al. (1998). The extremes of the dominant variable are selected by means of the Peak Over Threshold method. For each extreme value of the dominant variable, the associated extreme values of the concomitant variables are retrieved from a pre-specified and fixed time interval. This approach is explained in more detail in Section 5.1. A different approach would be to select the extremes independently for each variable, however, this will result in a loss of the dependence structure since the sampled extremes probably do not occur at the same time.

Offshore-Nearshore Transformation

The methodology choices were constrained by the fact that solely a metocean data time series was available which is conducted from a measurement buoy located 70 km offshore. The conditions at the toe of the breakwater are required for the evaluation of the design. Waves propagating towards the shore transform and change directions as a result of shallow water and seabed interaction. In order to determine the nearshore wave conditions, a transformation of the offshore conditions is performed using 2D SWAN model (SWAN, 2020). Prior to this, it is attempted to execute the design procedure by simply assuming the offshore data time series to be obtained at the toe of the breakwater and another attempt was the transformation of the entire time series. However, this resulted in unrealistic and inaccurate nearshore conditions.

The availability of a time series at the toe of the structure not only makes the offshore-nearshore transformation superfluous, it also enables modeling the system nearshore by means of a vine copula and directly assessing the multivariate density distribution enabling a Monte Carlo type reliability assessment. Recently, Felix Orlin (2020) illustrated such a reliability assessment for the design of fender systems. The required SWAN transformation makes this approach computationally too expensive and instead this research demonstrates the implementation of a multivariate approach when an offshore-nearshore transformation is required.

One-to-one Relation

In the traditional univariate approach, critical loading scenarios are determined by the derivation of design loads for each design variable individually. In a multivariate context, there is no longer a one-to-one relation between the characteristic design load and a specific probability level, meaning there are infinite joint design values corresponding to the same probability level. Consequently, the evaluation of the final cross-sectional design in a multivariate context is less obvious than the traditional approach.

Quantile Curve Estimation

The derivation of critical events in a multivariate context consists of the estimation of the quantile curve t , see Algorithm 1. The number of simulations N and sample size n have an effect on the accuracy of the obtained estimate. It is preferred to apply parameter values $n = 1e5$ and $N = 1e4$ to minimize the error (see Section 7.3). However, it is considered to be computationally too expensive for the scope of this research. Instead, parameter values $n=1e4$ and $N=100$ are used accepting an error of approximately 0.1% in the final quantile estimate based on the performed sensitivity analysis.

Goodness of Fit

There is no widely accepted technique for testing the goodness of fit of a multivariate model. This assessment uses a built-in function *RVineStructureSelect* from the *VineCopula* package (R programming language) to select a reasonable candidate model. However, information on the true best model or the optimality gap is not available for the practitioner. Perfect accuracy is impossible since a model is only an abstraction of the system it represents, but reducing modeling errors is very important.

8.4 Applicability

Kendall's Measure

In order to be able to derive critical loading scenarios in a multivariate context for the evaluation of the required cross-sectional design of the breakwater, it is required to be aware of the underlying failure scenarios. For example, is the structure going to fail for the exceedance of a critical threshold for certain variables (significant wave height, wind speed, etc) individually or is it only going to fail for a combination of exceedances? These critical scenarios belong to the so-called critical region and its definition should only depend on which one better describes the design requirements and mechanisms of failure. For the case study considered in this research, it is chosen to use the Kendall's measure since it is believed to be a suitable definition for the breakwater design procedure based on literature (Salvadori et al.; 2011, 2013, 2014, 2015). In the end, the practitioner is asked to decide upon a suitable definition based on the underlying failure scenarios. The usage of Kendall's measure might not be the best option for a different type of coastal infrastructure which can make the applicability of the multivariate approach more tedious and less attractive.

Multivariate Return Period

For a given probability level, the critical events are derived based on the definition of the critical region. In Section 5.3, it is stressed that one should be careful when expressing the probability level in terms of return period since it can be very misleading in a multivariate context. The definition of a return period in a univariate context is appealing as it provides an intuitive way to express the reliability of a structure, but in a multivariate context, it is preferred to stick to the probabilities to avoid misconceptions. This might be an obstacle in the acceptance of the multivariate approach by the coastal engineering community since the return period is a widely accepted and used principle.

8.5 Results

The critical events are derived by estimating the quantile curve that describes the critical region. For each experiment conducted in this research, approximately 35-55 critical events have been sampled from the quantile curve for which the required design is evaluated after offshore-nearshore transformation. The obtained cross-sectional designs are determined by evaluating the design calculations for each of the sampled critical events that all relate to the same prescribed safety level. Multiple selection strategies exist to choose a final design such as averaging overall designs or selecting the most robust design present in the sample. A trade-off is made between the number of critical events sampled and the computation time of the offshore-nearshore transformation. It can be argued that considering 35-55 critical events is a rather small sample and might cause an error in the results. The reliability of the results can be improved by increasing the number of critical events used for the evaluation of the design in case the practitioner thinks a more precise final design is worth the additional computation costs.

Theoretically infinite critical design events can be found for a given safety level because the one-to-one relationship is lost in a multivariate context. It can be considered a drawback of the multivariate approach that it does not directly lead to a final design, but instead, it demands the engineer to select a final design that is consultation with the other parties involved such as the client and contractor. It should be stressed that this final step is rather of importance even though it is left out of consideration in this study.

In this research, a case study has been performed in which two different design approaches have been applied for the design of a breakwater. In the end, there has not been a definite choice on the selection of a final design of the breakwater because the goal of this study was not so much to come up with a final design, but more to showcase the multivariate approach. However, if we were to select a final design, it would not be a straightforward decision as it is dependent on the application. Typically, the profit for the contractor depends on whether it manages to finalize the project according to all requirements for less money than the prearranged price with the client. Meaning that from a client perspective, it is expected that the most robust design (*maximum* selection strategy) is preferred because financially it does not have a direct effect. While on the other hand, from an engineering firm/contractor perspective it is aimed to maximize the profit, and therefore, the *min cost* design satisfying all requirements is preferred. Note, that this is certainly not always the case since it strongly depends on where the liability is put. Some types of contracts for these large engineering projects also require that the contractor/engineering firm is responsible for the functioning of the structure during its entire lifetime. In that case, it is likely that the most robust design out of the set of designs corresponding to a prescribed safety level is preferred by the contractor/engineering firm.

9 Conclusions & Recommendations

This chapter covers the key findings of this research and the recommendations derived from these findings, where it is aimed to answer the formulated research questions.

9.1 Conclusions

The lack of knowledge concerning the joint statistics of relevant variables in the design of offshore and coastal structures may severely limit their effectiveness and can lead to expensive and inappropriate decisions. Multivariate frequency analysis approaches currently receive much attention within the academic community and is expected to be adopted by codes and guidelines in the near future. However, advanced statistical concepts such as vine copula are slow in being taken up by engineering practice. With the objective to *'perform a practical assessment to investigate how the vine-based approach can be incorporated in the design of coastal infrastructure, and how the usage of dependence information can contribute to additional insights in comparison to the traditional univariate approach assuming independence.'*, this research set out to present of a practical assessment of the vine-based approach in continuation of the work performed by [Sellés Valls \(2019\)](#). By doing so, this study contributes to bridging the gap between the academic community and engineering practice on one hand, and on the other hand, contributes to a better understanding of the potential added value of incorporating dependence information in the design process of coastal and offshore infrastructure. This study is performed from a conceptual point of view where the concept of using dependence information by applying advanced statistical techniques is explored and the required adaptations throughout the entire design process are evaluated.

The key findings of this study are summarised by answering the sub-research questions and, subsequently, the overall conclusion is drawn based on the main research question.

1. *How can the vine-based methodology using dependence information be incorporated in the design process of a coastal engineering application?*

This research shows how the vine-based methodology can be incorporated into the design process of a rubble mound breakwater. Not only does it show how it can be incorporated, but it also shows how it can be adapted in case only offshore data time series is available and an offshore-nearshore transformation is required to determine the nearshore wave conditions. Along the same line, the framework is extended by introducing Kendall's measure to cope with the troublesome notion of the return period in a multivariate context. It enables the derivation of critical events based on a critical region defined by the design requirements and mechanisms of failure. It is found that given a safety level, the vine-based approach does not directly lead to a unique final design, but instead multiple final designs can be obtained. As a result, it demands the practitioner to use a selection strategy which might be experienced as a drawback. Three design selection strategies have been introduced

and illustrated namely: *maximum*, *average*, and *minimum costs*. In general, it is found that the results obtained with the *maximum* selection strategy were close to the results obtained according to the traditional approach.

2. *How can the vine-based methodology incorporated in the design process of a coastal engineering application provide extra information in contrast to the traditional univariate approach?*

In this research, experiments are conducted aiming to compare the vine-based approach with the traditional approach based on the resulting dimensions of the cross-sectional design of the breakwater and the corresponding costs. Based on several experiments representing different model configurations, it can be concluded that the vine-based design approach on average results in minimal required dimensions of elements of the cross-sectional design that turn out to be smaller compared to the results obtained with the traditional approach. In some cases, a breakwater design is obtained with a cost reduction in the order of 25%.

Based on the experiments conducted in this research, it is found that the vine-based methodology can be of added value in the design process of a coastal engineering application in particular for the understanding of the system behavior and the identification of hazardous loading conditions. Instead of assuming independence and simply combining extreme design loads in the traditional approach, a set of realistic design events can be derived representing the actual underlying data structure. Considering the fact that the traditional approach is well embedded in current industry practices and has proven to perform well over the years, it is not to be expected that the vine-based approach will be used *instead* of the traditional approach anytime soon, but rather in *conjunction* with, based on the practical assessment illustrated in this research. The role of the vine-based methodology in the design process of a breakwater structure (or coastal infrastructure in general) is to provide the practitioner with additional insights supporting the traditional design or give a reason to reevaluate or optimize the design. This line of thought concerning the role of the vine-based methodology has been confirmed by several coastal engineers over the course of this research.

Although the relevant physical processes that played a role in the considered case study also are relevant to other types of coastal infrastructure, it is difficult to generalize the results to coastal infrastructure in general because the results might be dependent on the application. For example, the relevant failure mechanisms vary for different types of structures which may be sensitive to other load conditions. Also, the location's specific physical processes and environmental conditions might influence the impact of the vine-based approach.

Finally, the main research question can be answered now that the answer to the sub-research question is formulated. For convenience, it is repeated here:

How can the vine-based methodology using dependence information be incorporated in the design process of a coastal engineering application and how can it provide extra information in contrast to the traditional univariate approach assuming independence?

In this research, it is found that vine-based methodology can be successfully incorporated in the design process of a breakwater structure, and on average results in minimal required dimensions of elements of the cross-sectional design that turn out to be smaller and the corresponding costs up to 25% lower compared to the traditional approach. This is realized by adapting the framework enabling an offshore-nearshore transformation of the wave conditions. Furthermore, it is found from the literature that the identification of critical events in a multivariate approach is not as straightforward but of crucial importance for the design process. In this research, the theoretical framework is extended by introducing Kendall's measure providing a suitable definition of the critical region from which the critical events can be obtained.

Additionally, computational experiments and sensitivity analysis are conducted aiming to provide additional insights enabling the practitioner to make educated choices and further enhance a successful application of the vine-based approach. It is concluded that the vine-based approach could act as a tool providing extra information about the behavior of the system and insights on the degree of conservatism of the traditional approach. The considered role of the vine-based methodology in the design process of a breakwater structure (or coastal infrastructure in general) is to provide the practitioner with additional insights supporting the traditional design approach and possibly optimizing the design.

9.2 Recommendations

A set of recommendations can be made based on the presented conclusions and the discussed limitations. The recommendations are briefly discussed below.

Case Study

- **Application**

First of all, it is recommended to apply the vine-based approach in the design process of a couple of other breakwater projects. In this way, the influence of aspects like the location specific circumstances can be investigated in more detail. Preferably, the different breakwater types are considered, for example, a vertical wall breakwater, to evaluate the effect of different failure modes.

Secondly, it is recommended to apply the vine-based methodology in the design procedure of a different type of coastal infrastructure such as an exposed jetty for similar reasons described above. Also, it may provide additional insights on the possible generalization of the potential of the vine-based approach.

- **Water Level Data**

It is recommended to use the real water level variation time series in the design process. The availability of the water level time series is of importance as shown by experiments 2.1 and 2.2.

Applicability

- **Length of Time Series**

One might wonder whether there is a minimum required time series length in order to properly apply the vine-based methodology. Therefore, it is recommended to perform experiments in which different periods of the considered time series are taken into consideration.

- **Workshop**

Although this research attempts to reduce the gap between the academic community and engineering practice, there is still some work required to bridge the gap entirely. A workshop in which coastal engineers are asked to apply the vine-based methodology to a hypothetical case may provide useful insights that may enhance the applicability.

Results

- **Critical Events**

Increasing the number of critical events sampled from the quantile curve corresponding to a given safety level will increase the accuracy of the final results. The process is straight forward although computationally expensive. This provides an indication of the variability in the final results for small sample sizes.

- **Quantile Curve Estimation**

The sensitivity analysis showed the potential error in the estimation of the quantile curve. Although the error is relatively small, it is recommended to evaluate the effect of this error on the final results.

- **Goodness of Fit for vine copula**

It is recommended to study the fit of the vine copula more thoroughly in future applications as it is important to avoid modeling errors. Note, that there is no widely accepted technique for testing the goodness of fit of a multivariate model. Instead, a combination of alternative testing methods can be added to the standard modeling procedure. For example, the Cramér Von Mises statistic (CvM) indicates how far the empirical copula deviates from the model. Another possibility is the visual comparison of the vine samples and the observations on the unit square or normal space in addition to the comparison in terms of Kendall's τ correlation coefficient.

References

- Aas, K. and Berg, D. (2009). Models for construction of multivariate dependence—a comparison study. *The European Journal of Finance*, 15(7-8):639–659. Publisher: Taylor & Francis.
- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.
- Bain, L. J. and Engelhardt, M. (1987). *Introduction to probability and mathematical statistics*. Brooks/Cole.
- Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, pages 792–804. Publisher: JSTOR.
- Battjes, J. A. and Groenendijk, H. W. (2000). Wave height distributions on shallow foreshores. *Coastal Engineering*, 40(3):161–182.
- Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268. Publisher: Springer.
- Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068. Publisher: JSTOR.
- Bentley, J. L. (1980). Multidimensional divide-and-conquer. *Communications of the ACM*, 23(4):214–229.
- Berg, D. and Bakken, H. (2007). A copula goodness-of-fit approach based on the conditional probability integral transformation. <http://www.danielberg.no/publications/Btest.pdf> (2008-11-01).
- Bezak, N., Brilly, M., and Šraj, M. (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, 59(5):959–977. Publisher: Taylor & Francis. eprint: <https://doi.org/10.1080/02626667.2013.831174>.
- Bondar, C. (2007). The Black Sea level variations and the river-sea interactions. page 8.
- Bosboom, J. and Stive, M. J. F. (2012). *Coastal dynamics I*.
- Brechmann, E. (2010). *Truncated and simplified regular vines and their applications*. PhD thesis, Technische Universität München.
- Breymann, W., Dias, A., and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*. Publisher: Taylor & Francis.
- Castillo, E. (2012). *Extreme Value Theory in Engineering*. Elsevier. Google-Books-ID: CyFd42CkOmMC.
- CIRIA, C. (2007). *The Rock Manual. The Use of Rock in Hydraulic Engineering*.
- Cunnane, C. (1979). A note on the Poisson assumption in partial duration series models. *Water Resources Research*, 15(2):489–494. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/WR015i002p00489>.

- CUR (1997). Kansen in de civiele techniek. D. 1: Probabilistisch ontwerpen in theorie. Technical report, CUR, Gouda. OCLC: 245808238.
- Davison, A. C. and Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1990.tb01796.x>.
- De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R. (2007). A multivariate model of sea storms using copulas. *Coastal Engineering*, 54(10):734–751.
- Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., and Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Science & Business Media. Google-Books-ID: odn7_auSanEC.
- Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69. Publisher: Elsevier.
- Durante, F., Fernández-Sánchez, J., and Sempì, C. (2013). How to Prove Sklar’s Theorem. In Bustince, H., Fernandez, J., Mesiar, R., and Calvo, T., editors, *Aggregation Functions in Theory and in Practise*, Advances in Intelligent Systems and Computing, pages 85–90, Berlin, Heidelberg. Springer.
- Durante, F. and Sempì, C. (2010). Copula Theory: An Introduction. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, Lecture Notes in Statistics, pages 3–31, Berlin, Heidelberg. Springer.
- Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*, 14. Publisher: Citeseer.
- Felix Orlin, F. (2020). Reliability-based Assessment for Fender Systems.
- Ferreira, J. A. and Guedes Soares, C. (2000). Modelling distributions of significant wave height. *Coastal Engineering*, 40(4):361–374.
- Ferro, C. A. T. and Segers, J. (2003). Inference for Clusters of Extreme Values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2):545–556. Publisher: [Royal Statistical Society, Wiley].
- Fisher, N. I. (2004). Copulas. *Encyclopedia of statistical sciences*. Publisher: Wiley Online Library.
- Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3^e série, Sciences, Sect. A*, 14:53–77.
- Genest, C., Remillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199–213. Publisher: Elsevier.
- Gent, M. and Pozueta, B. (2004). REAR-SIDE STABILITY OF RUBBLE MOUND STRUCTURES.
- Goda, Y. (1988). ON THE METHODOLOGY OF SELECTING DESIGN WAVE HEIGHT.
- Grimaldi, S. and Serinaldi, F. (2006). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8):1155–1167.

- Gruber, L. and Czado, C. (2015). Sequential bayesian model selection of regular vine copulas. *Bayesian Analysis*, 10(4):937–963. Publisher: International Society for Bayesian Analysis.
- Gräler, B., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. (2016). An update on multivariate return periods in hydrology. In *Proceedings of the International Association of Hydrological Sciences (IAHS)*, volume 373, pages 175–178. Copernicus. ISSN: 2199-899X.
- Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296.
- Guillou, A., Padoan, S. A., and Rizzelli, S. (2018). Inference for asymptotically independent samples of extremes. *Journal of Multivariate Analysis*, 167:114–135.
- Gulvanessian, H. (2001). EN1990 Eurocode—Basis of structural design. *Proceedings of the Institution of Civil Engineers - Civil Engineering*, 144(6):8–13. Publisher: ICE Publishing.
- Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia Univ. Press, New York, 201.
- Hames, D. P., Gouldby, B. P., and Hawkes, P. J. (2019). Evolution of joint probability methods in coastal engineering practice in the UK. *Proceedings of the Institution of Civil Engineers - Maritime Engineering*, 172(2):45–54. Publisher: ICE Publishing.
- Hans, M. (2007). Estimation and Model Selection of Copulas with an Application to Exchange Rates. Technical Report 056, Maastricht University, Maastricht Research School of Economics of Technology and Organization (METEOR). Publication Title: Research Memorandum.
- Haver, S. (1985). Wave climate off northern Norway. *Applied Ocean Research*, 7(2):85–92.
- Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts für Angewandte Mathematik der Universität Berlin*.
- Hydra-NL (2020). Hydra-NL. Last Modified: 2020-06-08.
- Joe, H. (1996). *Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters*. Institute of Mathematical Statistics. Pages: 120-141 Publication Title: Distributions with fixed marginals and related topics.
- Joe, H. and Kurowicka, D. (2011). *Dependence Modeling: Vine Copula Handbook*. World Scientific.
- Jonkman, S. N., Steenbergen, R., Morales-Nápoles, O., Vrouwenvelder, A., and Vrijling, J. (2015). Probabilistic design: Risk and reliability analysis in civil engineering. *Lecture Notes CIE4130. Delft University of Technology*.
- Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific. Google-Books-ID: GwBqDQAAQBAJ.
- Krogstad, H. E. (1985). Height and period distributions of extreme waves. *Applied Ocean Research*, 7(3):158–165.
- Kuijken, W. (2015). Deltaprogramma 2015. *Deltacommissaris*. Publisher: Deltacommissaris.

- Kurowicka, D. (2010). Optimal truncation of vines. In *Dependence Modeling: Vine Copula Handbook*, pages 233–247. World Scientific.
- Kurowicka, D. and Cooke, R. (2003). A parameterization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications*, 372:225–251. Publisher: Elsevier.
- Kuwashima, S. and Hogben, N. (1986). The estimation of wave height and wind speed persistence statistics from cumulative probability distributions. *Coastal Engineering*, 9(6):563–590.
- Lang, M., Ouarda, T. B. M. J., and Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225(3):103–117.
- Langrené, N. and Warin, X. (2020). Fast multivariate empirical cumulative distribution function with connection to kernel density estimation. *arXiv:2005.03246 [cs, stat]*. arXiv: 2005.03246.
- Leadbetter, M. R. (1991). On a basis for ‘Peaks over Threshold’ modeling. *Statistics & Probability Letters*, 12(4):357–362.
- Li, Y., Simmonds, D., and Reeve, D. (2008). Quantifying uncertainty in extreme values of design parameters with resampling techniques. *Ocean Engineering*, 35(10):1029–1038.
- Morales Napoles, O., Cooke, R. M., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes. Technical report, Delft University of Technology.
- Morton, I. D. and Bowers, J. (1996). Extreme value analysis in a multivariate offshore environment. *Applied Ocean Research*, 18(6):303–317.
- Mulazzani, M. (1985). Reliability Versus Safety. *IFAC Proceedings Volumes*, 18(12):141–146.
- Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.
- Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science & Business Media. Google-Books-ID: yexFAAAAQBAJ.
- PIANC (1992). MarCom WG 12: Analysis of Rubble Mound Breakwaters (1992-1994). text/html, PIANC. Archive Location: <https://www.pianc.org/> Publisher: PIANC.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131. Publisher: Institute of Mathematical Statistics.
- Pullen, T., Allsop, W., Bruce, T., Kortenhuis, A., Schüttrumpf, H., and van der Meer, J. (2018). *EurOtop II Wave Overtopping of Sea Defences and Related Structures: Assessment Manual*.
- Rivas, D., Caleyó, F., Valor, A., and Hallen, J. M. (2008). Extreme value analysis applied to pitting corrosion experiments in low carbon steel: Comparison of block maxima and peak over threshold approaches. *Corrosion Science*, 50(11):3193–3204.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472. Publisher: JSTOR.

- Salvadori, G., De Michele, C., and Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11):3293–3305.
- Salvadori, G., Durante, F., and De Michele, C. (2013). Multivariate return period calculation via survival functions. *Water Resources Research*, 49(4):2308–2311. Publisher: Wiley Online Library.
- Salvadori, G., Durante, F., Tomasicchio, G. R., and D’Alessandro, F. (2015). Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering. *Coastal Engineering*, 95:77–83.
- Salvadori, G. and Michele, C. D. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(12). eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004WR003133>.
- Salvadori, G. and Michele, C. D. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resources Research*, 46(10). eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR009040>.
- Salvadori, G., Tomasicchio, G. R., and D’Alessandro, F. (2014). Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. *Coastal Engineering*, 88:1–14.
- Schepsmeier, U. (2013a). Efficient goodness-of-fit tests in multi-dimensional vine copula models. *arXiv:1309.5808 [stat]*. arXiv: 1309.5808.
- Schepsmeier, U. (2013b). A goodness-of-fit test for regular vine copula models. *arXiv:1306.0818 [math, stat]*. arXiv: 1306.0818.
- Sellés Valls, S. (2019). *A vine-based approach for defining critical infrastructure loads: Designing a breakwater in Galveston Bay, Texas*. PhD thesis, Delft University of Technology, Delft.
- Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29(4):1179–1189.
- Singh, V. P., Jain, S. K., and Tyagi, A. (2007). *Risk and Reliability Analysis*. American Society of Civil Engineers.
- Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *L’Institut de Statistique de L’Universite de Paris*.
- SWAN, T. S. t. (2020). SWAN.
- TAW, T. A. C. o. W. (1984). Some Considerations on Acceptable Risk in the Netherlands. Technical report, Dienst Weg- en Waterbouwkunde, Delft. Publisher: Dienst Weg-en Waterbouwkunde Delft.
- van Dantzig, D. (1956). Economic Decision Problems for Flood Prevention. *Econometrica*, 24(3):276–287. Publisher: [Wiley, Econometric Society].
- Volpi, E. and Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. *Hydrological sciences journal*, 57(8):1506–1515. Publisher: Taylor & Francis.
- Vrijling, J. (2001). Probabilistic design of water defense systems in The Netherlands. *Reliability Engineering & System Safety*, 74(3):337–344.

- Vrijling, J. K., van Hengel, W., and Houben, R. J. (1998). Acceptable risk as a basis for design. *Reliability Engineering & System Safety*, 59(1):141–150.
- VROM (1988). The Dutch National Environmental Policy Plan. *The Netherlands Journal of Housing and Environmental Research*, 4(4):337–345.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25. Publisher: JSTOR.
- Zachary, S., Feld, G., Ward, G., and Wolfram, J. (1998). Multivariate extrapolation in the offshore environment. *Applied Ocean Research*, 20(5):273–295.
- Zhong, M., Wang, J., Jiang, T., Huang, Z., Chen, X., and Hong, Y. (2020). Using the Apriori Algorithm and Copula Function for the Bivariate Analysis of Flash Flood Risk. *Water*, 12(8):2223. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Appendices

A. Basics of Probability Theory & Statistics

Probability Density Functions

Probability theory is a branch of mathematics concerned with the description and analysis of random phenomena. Statistics focuses on the analysis of data using probability theory. The realization of a random variable X is dependent on a random process and therefore the outcome is subject to some variation. All possible outcomes of a random variable is described by the sample space Ω . A probability distribution assigns real values to elements from the sample space and contain all the probabilistic information about X (Dekking et al., 2005).

The random variables can be classified as a discrete or continuous random variable. A discrete random variable is characterized by the occurrence of discrete events. For example, the outcome of the random process of throwing a dice is an integer between 1 and 6. The probability of throwing a certain number x is mathematically expressed as follows:

$$f(x) = P(X = x) \quad \text{where } x \in \Omega = \{1, 2, 3, 4, 5, 6\} \quad (\text{A.1})$$

where $f(x)$ is the probability density function (PDF) and has the following properties:

$$\begin{aligned} 0 \leq f(x) \leq 1 \quad \forall x \\ \sum_{x \in \Omega} f(x) = 1 \end{aligned}$$

The cumulative distribution function (CDF) describes the probability that the random variable X will be less or equal than a certain realization x . The cumulative probability is expressed as:

$$F(x) = P(X \leq x) = \sum_{x_i \in \bar{\Omega}} f(x) \quad \text{where } \bar{\Omega} = \{x_i \in \Omega | x_i \leq x\} \quad \text{and } \bar{\Omega} \subseteq \Omega$$

The variables of interest in the field of civil engineering more often belong to the group of continuous variables. Continuous variables typically have an infinite sample space in contrast to the discrete random variables. The PDF generally is used to determine the probability that a random variable X will fall within a certain interval $[a, b]$,

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

The function f has to satisfy $f(x) \geq 0$ for all x and when the total sample space is integrated, it should hold that $\int_{-\infty}^{+\infty} f(x) dx = 1$. Hence, if f is the PDF, F is the CDF and the random variable X is defined on the domain $[-\infty, \infty]$, then:

$$P(X \leq a) = \int_{-\infty}^a f(x)dx = F(a) \quad (\text{A.2})$$

The probability that the random variable takes on a specific value is zero since there are infinite many possibilities in the continuous case, $P(X = x) = 0$. Parametric distributions are used to adequately model a data sample assuming that it comes from a specific population. Well known discrete parametric distribution are for example the binomial, geometric and poisson distribution. Parametric distributions commonly used for continuous random variables are normal, exponential and uniform distribution for example.

Joint Probability Distributions

The joint probability distribution is a probability density function for 2 or more random variables. Generally speaking, the joint probability distribution of 2 random variables is called a bivariate distribution and a multivariate probability distribution is a n -dimensional distribution with $n > 2$. The probability that a realization (X, Y) will be within a certain rectangle $[a_1, b_1] \times [a_2, b_2]$ is then determined by integrating the bivariate probability density function $f(x, y)$,

$$P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y) dx dy \quad (\text{A.3})$$

Properties of a bivariate probability density function are:

$$\begin{aligned} f(x, y) &\geq 0 \quad \forall \quad x, y \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy &= 1 \end{aligned} \quad (\text{A.4})$$

Note, it holds that $f(x, y) = f(x)f(y)$ in case of independent variables X and Y .

In case of a multivariate probability distribution with $n > 2$ continuous variables, the joint probability distribution F is specified by:

$$F(a_1, a_2, \dots, a_n) = P(X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n) \quad \text{for} \quad -\infty < a_1, a_2, \dots, a_n < \infty \quad (\text{A.5})$$

where $f(x_1, x_2, \dots, x_n) = \frac{\delta^n}{\delta x_1 \delta x_2 \dots \delta x_n} F(x_1, x_2, \dots, x_n)$.

Conditional Probability

Another important concept in probability theory is the conditional probability. The probability that event A occurs given the occurrence of event B is called a conditional probability. It basically means which fraction of the probability of B is also in the event A . The formal definition is given as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{for} \quad P(B) > 0$$

where $P(A \cap B)$ is the intersection of event A and B .

The dependence of variables has an influence in the probability calculations. In case event A and B are independent, it holds that $P(A \cap B) = P(A)P(B)$. One can now observe that conditionalizing on event B does not influence the probability of the occurrence of event A as $P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$. The concept of conditional probability makes it possible to use available information in the probability calculation and furthermore, makes it possible to update probability distributions.

If two continuous variables X_1 and X_2 are considered to have a joint distribution $f(x_1, x_2)$ then the conditional probability density function of X_1 given $X_2 = x_2$, is defined by:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)} \quad \text{for } f(x_2) > 0$$

where $f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$ is the marginal distribution of x_2 . By interchanging x_1 and x_2 , the conditional probability density for x_2 given x_1 is obtained:

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} \quad \text{for } f(x_1) > 0$$

Combining both expressions result in Bayes theorem which is important in the field of Bayesian statistics.

$$f(x_1|x_2)f(x_2) = f(x_2|x_1)f(x_1)$$

Dependence Measures

Pearson's Correlation Coefficient

The covariance of two random variables is a statistical measure indicating the amount of linear dependence. The mathematical definition is the expected value of the product of the deviations of the random variables with respect to their mean:

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] \quad (\text{A.6})$$

The Pearson's product moment correlation coefficient ρ_{XY} is a concept which is directly related to the covariance. The definition is given by:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (\text{A.7})$$

where the variance of X is defined as $\sigma_X = \sqrt{E[(X - E[X])^2]}$. The Pearson's coefficient ρ_{XY} is equal to -1 / 1 in case of fully positive/negative dependence between random variables X and Y . Furthermore, $\rho_{XY} = 0$ in case of independence. However, it does not mean X and Y are independent if $\rho_{XY} = 0$ because, the correlation coefficient is solely a measure of linear dependence and there may be some non-linear dependencies. The linear correlation coefficient might be often misleading and therefore, two alternative measures of dependence are concisely presented below (Embrechts et al., 2001).

Spearman's Rank Correlation Coefficient

Another statistical measure for correlation is the Spearman's Rank coefficient $r_{X,Y}$ based on the ranks of variables X and Y . It explores the so-called monotonic dependence rather than the linear dependence and typically is more robust, more sensitive to non-linear dependencies.

$$r_{X,Y} = \rho(F_X(X), F_Y(Y)) \quad (\text{A.8})$$

where F_X and F_Y are cumulative probability distribution which may be a parametric or an empirical estimate.

Kendall's τ correlation coefficient

A third correlation coefficient is the Kendall's τ coefficient which is also a rank correlation based coefficient. The correlation is based on concordance and discordance of pairs of observations and reflects the consistency of similar ordering of both variables. Various possibilities exist to calculate the Kendall's coefficient depending on the occurrence of tied observations and the difference in scaling. A pair of observations is said to be tied in case one of the variables is neither concordant or discordant.

Tail Dependence

Kendall's τ and Spearman's ρ measure the dependence on the whole space $[0, 1]^2$. Tail dependence is a measure of dependence focusing on the lower left and upper right quadrants of the space $[0, 1]^2$. The lower tail dependence and upper tail dependence of random variables X_1 and X_2 are defined by λ^{lower} and λ^{upper} respectively.

$$\lambda^{lower} = \lim_{t \rightarrow 0^+} P\left(X_2 \leq F_2^{-1}(t) | X_1 \leq F_1^{-1}(t)\right),$$

$$\lambda^{upper} = \lim_{t \rightarrow 1^-} P\left(X_2 > F_2^{-1}(t) | X_1 > F_1^{-1}(t)\right)$$

In other words, tail dependence of a bivariate distribution can be described as the probability that random variable X_1 exceeds a quantile q given that the other random variable X_2 also exceeds the q -quantile. The Student t , Clayton and the Gumbel copulae are examples of distributions that are able to model dependence in the tails ([Aas and Berg, 2009](#)).

B. Failure Mechanisms

This section contains the calculation methods for potential failure mechanisms considered in the breakwater case study.

Overtopping

The required crest level of the structure is determined by the overtopping criterion and can be calculated as given by the EurOtop overtopping manual (2018) (eq. 5.12 & 5.13).

$$\frac{q}{\sqrt{g \cdot H_s^3}} = \frac{0.026}{\sqrt{\tan \alpha}} \gamma_b \cdot \xi_{m-1,0} \cdot \exp \left[- \left(2.5 \frac{R_c}{\xi_{m-1,0} \cdot H_s \cdot \gamma_b \cdot \gamma_f \cdot \gamma_\beta \cdot \gamma_v} \right)^{1.3} \right] \quad (\text{B.1})$$

with a maximum of:

$$\frac{q}{\sqrt{g \cdot H_s^3}} = 0.1035 \cdot \exp \left[- \left(1.35 \frac{R_c}{H_s \cdot \gamma_f \cdot \gamma_\beta \cdot \gamma^*} \right)^{1.3} \right] \quad (\text{B.2})$$

where:

R_c :	freeboard [m]	γ_b :	influence factor for berm [-]
q :	maximum overtopping discharge [l/s/m]	γ^* :	combined influence factor [-]
H_s :	significant wave height [m]	γ_f :	influence factor for slope roughness [-]
g :	gravitational constant [m/s^2]	γ_β :	influence factor for oblique wave attack [-]
α :	slope of sea side armour layer [rad]	γ_v :	influence factor for wall at end of slope [-]
$\xi_{m-1,0}$:	surf similarity parameter [-]		

The surf similarity or breaker parameter ζ is used to describe the wave action on a slope and is defined by:

$$\zeta = \frac{\tan \alpha}{\sqrt{s_o}} \quad (\text{B.3})$$

where the fictitious wave steepness s_o is expressed as:

$$s_o = \frac{H}{L_o} = \frac{2\pi H}{gT^2} \quad (\text{B.4})$$

Stability Concrete Armour Units

$$\frac{H_s}{\Delta D_n} = 2.8 \quad (\text{B.5})$$

where

H_s :	significant wave height [m]	N_{od} :	Damage level [-]
Δ :	$\rho_s / \rho_w - 1$ relative buoyant density [-]		

Armour Stability Sea Side

Furthermore, interlocking armour units are designed in accordance with the specific design manual for the selected interlocking units. The required armour stone diameter for the sea side armour layer is calculated using the Van Der Meer equations as illustrated in the Rock Manual (2007) (eq. 5.139 & 5.140).

For shallow water ($h < 3H_s$):

$$\frac{H_s}{\Delta D_{n50}} = c_{pl,shallow} P^{0.18} \left(\frac{S_d}{\sqrt{N}} \right)^{0.2} \left(\frac{H_s}{H_{2\%}} \right) (\zeta_{s-1,0})^{-0.5} \quad \text{plunging waves } (\zeta_{s-1,0} < \zeta_{cr}) \quad (\text{B.6})$$

$$\frac{H_s}{\Delta D_{n50}} = c_{s,shallow} P^{-0.13} \left(\frac{S_d}{\sqrt{N}} \right)^{0.2} \left(\frac{H_s}{H_{2\%}} \right) \sqrt{\cot \alpha} (\zeta_{s-1,0})^P \quad \text{surging waves } (\zeta_{s-1,0} \geq \zeta_{cr}) \quad (\text{B.7})$$

where

D_{n50} :	nominal diameter [m]	P :	notional permeability parameter [-]
H_s :	significant wave height [m]	N :	number of waves [-]
$H_{2\%}$:	wave height exceeded by 2% of the waves [m]	S_d :	damage level parameter [-]
$\zeta_{s-1,0}$:	surf similarity parameter [-]	$c_{pl,shallow}$:	coefficient [-]
$T_{m-1,0}$:	spectral mean energy wave period [s]	$c_{s,shallow}$:	coefficient [-]
Δ :	$\rho_s / \rho_w - 1$ relative buoyant density [-]	ρ_s, ρ_w :	mass density (rock & water) [kg/m^3]
ζ_{cr} :	$\left[\frac{c_{pl}}{c_s} P^{0.31} \sqrt{\tan \alpha} \right]^{\frac{1}{P+0.5}}$, critical value for ζ [-]	α :	slope of sea side armour layer [rad]

The Van Der Meer formulae (see Rock Manual 2007 eq. 5.136 & 5.137) for deep water ($h \geq 3H_s$):

$$\frac{H_s}{\Delta D_{n50}} = c_{pl,deep} P^{0.18} \left(\frac{S_d}{\sqrt{N}} \right)^{0.2} \zeta_m^{-0.5} \quad \text{plunging waves } (\zeta_m < \zeta_{cr}) \quad (\text{B.8})$$

$$\frac{H_s}{\Delta D_{n50}} = c_{s,deep} P^{-0.13} \left(\frac{S_d}{\sqrt{N}} \right)^{0.2} \sqrt{\cot \alpha} \zeta_m^P \quad \text{surging waves } (\zeta_m \geq \zeta_{cr}) \quad (\text{B.9})$$

where

D_{n50} :	nominal diameter [m]	P :	notional permeability parameter [-]
H_s :	significant wave height [m]	N :	number of waves [-]
$H_{2\%}$:	wave height exceeded by 2% of the waves [m]	S_d :	damage level parameter [-]
ζ_m :	surf similarity parameter [-]	$c_{pl,deep}$:	coefficient [-]
$T_{m-1,0}$:	spectral mean energy wave period [s]	$c_{s,deep}$:	coefficient [-]
Δ :	$\rho_s / \rho_w - 1$ relative buoyant density [-]	ρ_s, ρ_w :	mass density (rock & water) [kg/m^3]
ζ_{cr} :	$\left[\frac{c_{pl}}{c_s} P^{0.31} \sqrt{\tan \alpha} \right]^{\frac{1}{P+0.5}}$, critical value for ζ [-]	α :	slope of sea side armour layer [rad]

The method introduced by Battjes and Groenendijk (2000) can be used to obtain estimates for the wave height exceeded by 2% of the waves $H_{2\%}$. An illustration of the method is given by the Rock Manual (Box 4.4). Battjes and Groenendijk (2000) proposed a composite Weibull distribution to describe the cumulative distribution of the wave height in shallow water and the breaker zone. The composite Weibull distribution is given by:

$$P(H) = P(\underline{H} < H) = \begin{cases} 1 - \exp\left(- (H/H_{rms})^2\right) & \text{for } H < H_{tr} \\ 1 - \exp\left(- (H/H_{rms})^{3.6}\right) & \text{for } H \geq H_{tr} \end{cases} \quad (\text{B.10})$$

where H_{tr} is the transitional wave height, indicating the transition between the shallow water and breaker zone. H_{rms} is the so called root mean squared wave height. The following expressions are used to calculate these wave heights:

$$H_{tr} = (0.35 + 5.8 \tan(\alpha_{bed}))h \quad (\text{B.11})$$

$$H_{rms} = (0.6725 + 0.2025 (H_s/h))H_s \quad (\text{B.12})$$

where

h : the local water depth [m] α_{bed} : bed slope in front of structure [rad]

Armour Stability Shore Side

The required stone size D_{n50} of the rear side of the breakwater can be calculated with the following equation derived by Gent and Pozueta (2004):

$$D_{n50} = 0.008 \left(\frac{S_d}{\sqrt{N}} \right)^{-1/6} \left(\frac{u_{1\%} T_{m-1,0}}{\sqrt{\Delta}} \right) (\cot \alpha_{rear})^{-2.5/6} (1 + 10 \exp(-R_c/H_s))^{1/6} \quad (\text{B.13})$$

where

α_{rear} : slope of shore side armour layer [rad] R_c : freeboard [m]
 H_s : significant wave height [m] N : number of waves [-]
 $T_{m-1,0}$: spectral mean energy wave period [s] S_d : damage level parameter [-]
 Δ : $\rho_s/\rho_w - 1$ relative buoyant density [-] ρ_s, ρ_w : mass density (rock & water) [kg/m^3]

and $u_{1\%}$ is the maximum (depth-averaged) velocity at the rear side of the crest (shore side) during a wave overtopping event, exceeded by 1% of the incident waves. The velocity can be calculated by:

$$u_{1\%} = 1.7 (g\gamma_{f-c})^{0.5} \left(\frac{R_{u1\%} - R_c}{\gamma_f} \right)^{0.5} / \left(1 + 0.1 \frac{B}{H_s} \right) \quad (\text{B.14})$$

where:

B : crest width [m] R_c : freeboard [m]
 γ_f : roughness factor of sea side armour layer [-] γ_{f-c} : roughness at the crest [-]

and $R_{u1\%}$ is the fictitious run-up level exceeded by 1% of the incident waves. The fictitious run-up level is defined by:

$$\frac{R_{u1\%}}{\gamma H_s} = \begin{cases} c_0 \zeta_{s-1,0} & \text{if } \zeta_{s-1,0} \leq p \\ c_1 - c_2 / \zeta_{s-1,0} & \text{otherwise.} \end{cases} \quad (\text{B.15})$$

where:

c_0, c_1 :	coefficients [-]	γ :	reduction factor [-]
p :	$0.5 c_1 / c_0$ [-]	$\zeta_{s-1,0}$:	surf similarity parameter [-]
c_2 :	$c_1^2 / 4c_0$ [-]		

The reduction factor γ takes into account the angular wave γ_β effects and the roughness γ_f .

Toe Stability

The stability of the toe of the breakwater can be evaluated using the following equation:

$$\frac{H_s}{\Delta D_{n50}} = \left(2 + 6.2 \left(\frac{h_t}{h} \right)^{2.7} \right) N_{od}^{0.15} \quad (\text{B.16})$$

where:

H_s :	significant wave height [m]	h_t :	water depth at the toe [m]
h :	water depth in front of the structure [m]	N_{od} :	damage number [-]
Δ :	$\rho_s / \rho_w - 1$ relative buoyant density [-]	ρ_s, ρ_w :	mass density (rock & water) [kg/m^3]

C. Design Input

This section contains relevant design input obtained from the preliminary design documents prepared by Van Oord and provides useful information for the evaluation of the failure modes. Most design inputs are constant (fixed) values like the water mass density ρ_w , but some input parameters are associated with the acceptable state of the 'breakwater system' such as the acceptable mean overtopping discharge q . A few parameters have not been specified since a statistical analysis is required for the derivation of the parameter values which is described in Chapter 4 and 5.

General Information

- The Black Sea is brackish, a sea water density of 1018 kg/m^3 shall be applied in the design.
- The mass density of the rock armour used for design of rubble mound structure armour layers is assumed $2,350 \text{ kg/m}^3$.
- The gravitational constant $g = 9.81 \text{ m/s}^2$.
- All elevations are expressed in meters above Mean Sea Level of the Black Sea (MN75)
- All horizontal coordinates shall be expressed in meters according to Universal Transverse Mercator zone 35 (UTM35).

Overtopping

Table C.1 gives an overview of the design parameter values used for the overtopping failure mechanism. The influence factors γ_b , γ^* and γ_v are set equal to 1 since they do not play a role in the design. The influence factor for the oblique wave attack γ_β does play a role, however, the most important wave direction is assumed to be (on average) perpendicular to the breakwater extension.

Parameter	Description	Unit	Value
H_s	significant wave height	[m]	<i>see Chapter 4 & 5</i>
q	maximum overtopping discharge	[l/s/m]	50
$\tan(\alpha)$	slope of sea side armour layer	[-]	1 : 1.5
γ_f	influence factor for slope roughness	[-]	0.46
γ_β	influence factor for oblique wave attack	[-]	1
γ_v	influence factor for wall at end of slope	[-]	1
γ^*	combined influence factor	[-]	1
γ_b	influence factor for berm	[-]	1

Table C.1: Design parameters for the overtopping failure mode.

Armour Stability Sea Side

Table C.2 gives an overview of the design parameter values required for the evaluation of the armour stability on the seaside of the structure. The value for the damage level S_d is based on the SLS design event. The bed slope is based on the measurements station which is located 70 kilometers offshore. The total water depth at this location is 60 meters.

Parameter	Description	Unit	Value
H_s	significant wave height	[m]	<i>see Chapter 4 & 5</i>
$T_{m-1,0}$	mean wave period	[s]	<i>see Chapter 4 & 5</i>
S_d	damage level parameter	[-]	2
P	notional permeability parameter	[-]	0.4
N	number of waves	[-]	3000
$\tan(\alpha)$	slope of sea side armour layer	[-]	1 : 1.5
$c_{s,deep}$	coefficient	[-]	1
$c_{pl,deep}$	coefficient	[-]	6.2
$c_{s,shallow}$	coefficient	[-]	1.3
$c_{pl,shallow}$	coefficient	[-]	8.4
$\tan(\alpha_{bed})$	bed slope in front of structure	[-]	60 : 70000
h	water depth in front of structure	[m]	<i>see Chapter 4 & 5</i>

Table C.2: Design parameters for the armour stability failure mode.

The water depth is calculated by $h = \zeta - z_{bottom}$ where z_{bottom} is constant and depends on the output location (see Section 3.3.3). The water level ζ is a hydraulic boundary condition for which critical values are derived in Chapter 4 & 5.

In case the concrete armour units Accropode II have been applied, the following information is required:

Parameter	Description	Unit	Value
H_s	significant wave height	[m]	<i>see Chapter 4 & 5</i>
N_{od}	damage number	[-]	0

Table C.3: Design parameters for the armour stability failure mode.

Armour Stability Shore Side

Table C.4 presents the values for the design parameters required for the evaluation of the rear side armour stability. The value for the damage level S_d is based on the SLS design event.

Parameter	Description	Unit	Value
H_s	significant wave height	[m]	see Chapter 4 & 5
$T_{m-1,0}$	mean wave period	[s]	see Chapter 4 & 5
S_d	influence factor for oblique wave attack	[-]	2
N	number of waves	[-]	3000
$\tan(\alpha_{rear})$	slope of shore side armour layer	[-]	1 : 2
B	crest width	[m]	12
γ_f	influence factor for slope roughness	[-]	0.46
γ_β	influence factor for oblique wave attack	[-]	1
γ_{f-c}	roughness at the crest	[-]	0.69
c_0	coefficient	[-]	1.45
c_1	coefficient	[-]	5.1
c_2	coefficient	[-]	4.484
p	coefficient	[-]	1.759
h	water depth in front of structure	[m]	see Chapter 4 & 5

Table C.4: Design parameters for the armour stability failure mode.

The water depth is calculated by $h = \zeta - z_{bottom}$ where z_{bottom} is constant and depends on the output location (see Section 3.3.3). The water level ζ is a hydraulic boundary condition for which critical values are derived in Chapter 4 & 5.

Toe Stability

Table C.5 presents the relevant design parameters for the evaluation of the toe stability. The water depth is calculated by $h = \zeta - z_{bottom}$ where z_{bottom} is constant and depends on the output location (see Section 3.3.3). The water level ζ is a hydraulic boundary condition for which critical values are derived in Chapter 4 & 5.

Parameter	Description	Unit	Value
h	water depth in front of structure	[m]	see Chapter 4 & 5
h_t	water depth at the toe	[m]	h - 2
N_{od}	damage number	[-]	0.5

Table C.5: Design parameters for the toe stability failure mode.

Rock Grading

A breakwater structure typically is made out of rock / concrete units of different sizes. The stability of the structure highly depends on the different rock gradings and the way they are applied. One can imagine that in case large rock sizes are applied on top of a layer of course graded rock, there is a chance that the smaller rocks are flushed out due to wave interaction leading to instabilities. The rock sizes and porosity are important parameters when designing such a structure. Regulations have been made on rock gradings to provide designers a more systematic approach in the design process.

The rock gradings should be according to the standard classes described in NEN-EN 13383-1. Any non-standard grading should satisfy the requirements set out in The Rock Manual 2007. Often the designer is limited by the available quarry at the site since importing rock from is too expensive. The standard and non-standard gradings which are available at the site for our case study are presented in Table C.6.

Grading name	M_{50} [kg]	ELL <5% kg	NLL <10% kg	NUL >70% kg	EUL >97% kg
CP 45/125 mm	N/A	22.4 mm	45 mm	125 mm	180 mm
LM 10-500 kg	90-260 kg	1 kg	10 kg	500 kg	1000 kg
HMA 1000-3000 kg	1868-2308 kg	650 kg	1000 kg	3000 kg	4500 kg
HM 300-600 kg	380-550 kg	80 kg	300 kg	600 kg	1150 kg
HM 500-2000 kg	800-1300 kg	300 kg	500 kg	2000 kg	2500 kg
HM 1000-4000 kg	2200-2800 kg	700 kg	1000 kg	4000 kg	5000 kg
HM 2000-4000 kg	2750-3300 kg	1000 kg	2000	4000 kg	6000 kg
HM 3000-9000 kg	5500-6500 kg	2500 kg	3000 kg	9000 kg	11000 kg

Table C.6: The rock grading classes according to NEN-EN 13383-1 available at the site.

A grading name, *LM 10-500 kg*, for example, is a combination of the type of grading *LM* and the nominal lower and upper limit of the grading class *10-500 kg*. Possible types of gradings are:

- "Heavy grading" (HM) contains larger sizes appropriate for armour layers,
- "Light grading" (LM) appropriate for armour layers, underlayers, and filter layers,
- "Course grading" (CP) often used for filter layers.

Limits on the armour stone sizes are introduced to specify the characteristics of a grading class. A set of nominal limits corresponds to the target size of the armour stone and a set of extreme limits corresponds to the tolerances.

- ELL (extreme lower limit): the mass below which no more than 5% passing by mass is permitted,
- NLL (nominal lower limit): the mass below which no more than 10% passing by mass is permitted,
- NUL (nominal upper limit): the mass below which no less than 70% passing by mass is permitted,
- EUL (extreme upper limit): the mass below which no less than 97% passing by mass is permitted.

For the cost estimation of the design, the unit cost prices presented in Table C.7. The unit costs include the material, transportation, and installation costs calibrated for this specific case study.

Section	core	filter layer	underlayer	armour layer
Cost [EUR/ m^3]	30	30	43	35

Table C.7: Unit prices for different sections of the breakwater.

D. Peak Over Threshold Procedure

Mean Residual Life

A suitable first attempt to find appropriate threshold values is by means of a so-called mean residual life plot. The mean excess above the threshold value u is calculated for different threshold values and is plotted in combination with a confidence interval of 2 times the sample standard deviation. According to [Davison and Smith \(1990\)](#), the selected threshold value should be on the domain where the mean excess is a linear function of the threshold value. The domain of interest is approximately $[1, 3.8]$ based on Figure D.1.

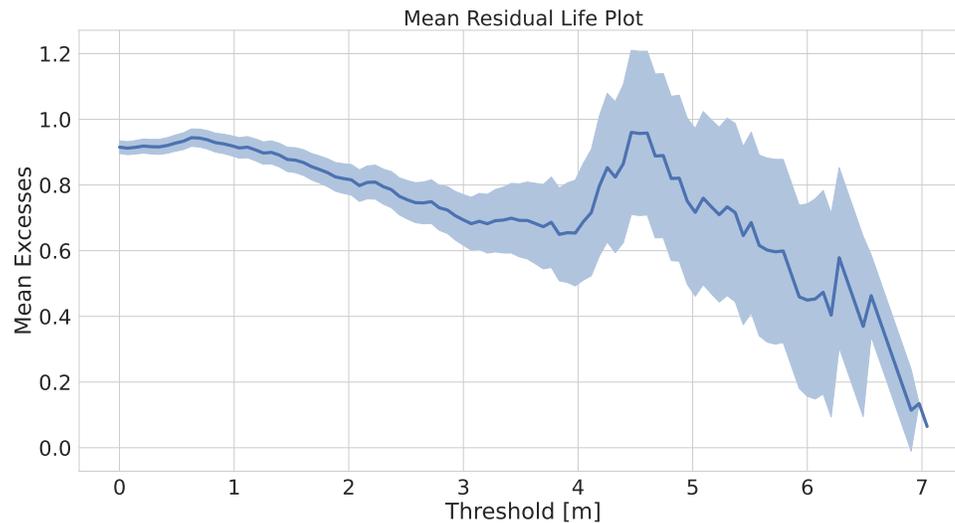


Figure D.1: The mean excess is considered to be a linear function of the threshold value u on the domain $[1, 3.8]$. This gives the first indication of potential appropriate threshold values.

Extremal Index

The POT procedure assumes that the extremes are independent. The extremal index (EI) can be used to validate whether a particular threshold value results in a set of independently distributed extremes. The EI is defined on $[0,1]$ such that the set of extremes is independent in case $EI = 1$ and some degree of dependence is present for $EI < 1$. The EI is calculated for different combinations of threshold values and interval parameters. The selected threshold values are obtained from the domain of interest $[1, 3.8]$ and the results are depicted in Figure D.2. It can be observed that less dependence is present among the selected extremes obtained with lower threshold values. Based on this figure, threshold values 1.0 and 2.1 are preferred as they exceed $EI = 0.9$ for a smaller declustering time lag (48 hours). However, one might want to evaluate the EI for additional threshold values on the domain $[2.1, 3.1]$ since the progression of the lines differs quite substantially.

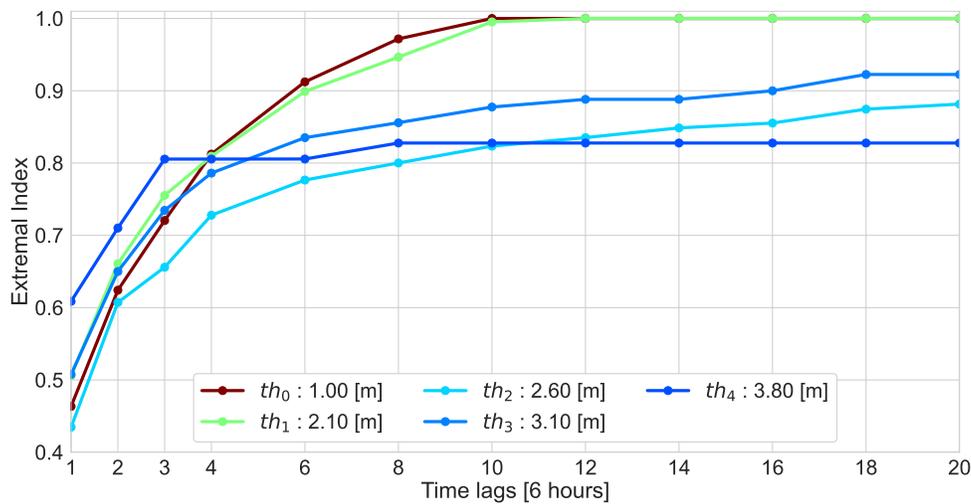


Figure D.2: Each line presents the extremal index (EI) as a function of the declustering time lag. It can be observed that for a time lag of 48 hours, the independence assumption is already satisfied for low threshold values as $EI > 0.9$.

Dispersion Index

The dispersion index (DI) is used to verify the assumption that the annual number of extremes given threshold value u follows a Poisson process. In case the DI is significantly close to 1, the Poisson Process assumption seems reasonable. Figure D.3 presents the dispersion index calculated for a declustering time lag of 48 hours in the top figure. The confidence bounds indicate the domain for which the Chi-squared test with null hypothesis $H_0 : DI = 1$ cannot be rejected. The confidence interval is relatively large since the time series only consist of 20 years of data (small sample size results in uncertainty). The lower figure shows the average number of extremes a year. The trade-off is to keep as many extremes (information) but to make sure the extremes are sufficiently dispersed ($DI = 1$). The thresholds 2.1 and 2.6 are preferred since the corresponding DI is close to 1 and N_s is relatively large.

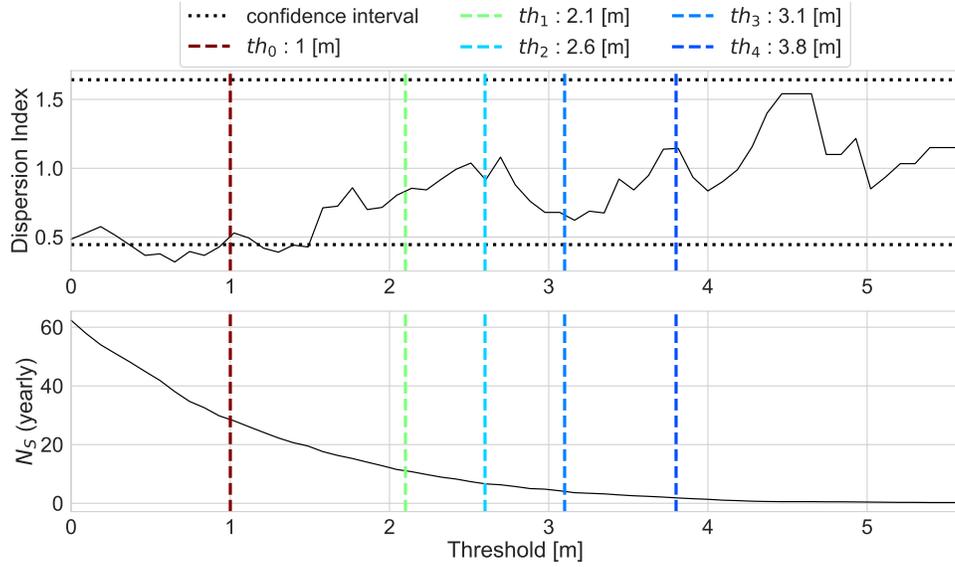


Figure D.3: The top figure contains the dispersion index for a minimal declustering time lag of 48 hours and the figure below shows the effect on the yearly average number of extremes for increasing threshold value. The colored vertical lines relate to the selected threshold values of interest.

Based on these indicators, it is decided to narrow down the threshold domain of interest from $[1, 3.8]$ to $[2.1, 2.6]$ and recalculate the extremal index plot, see Figure D.4. Based on this figure, a 72 hours declustering time lag seems reasonable and is used to calculate the corresponding dispersion index, see Figure D.5. It can be observed that all thresholds applied to satisfy the dispersion criterion.

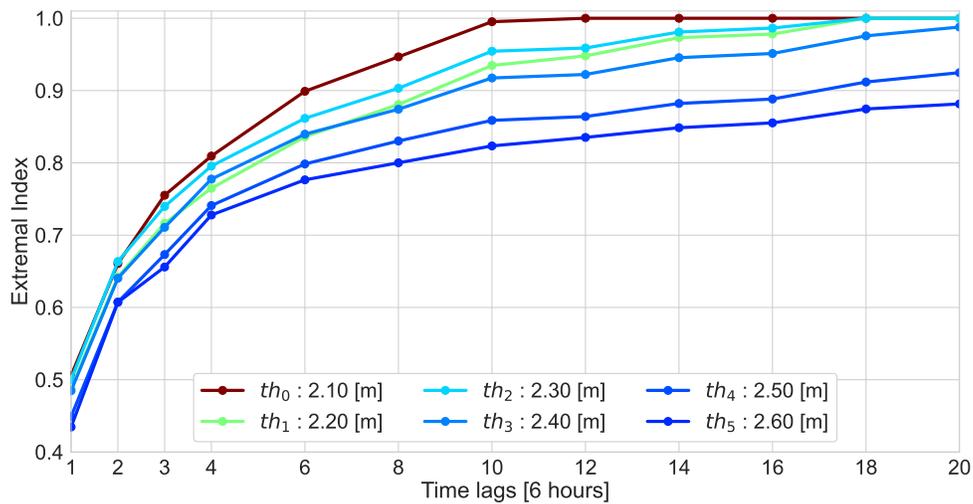


Figure D.4: Each line presents the extremal index (EI) as a function of the declustering time lag. It can be observed that large thresholds th_4 and th_5 only reach $EI > 0.9$ for large time lags. The other investigated threshold values satisfy the independence assumption for a declustering time lags around 60 or 72 hours.

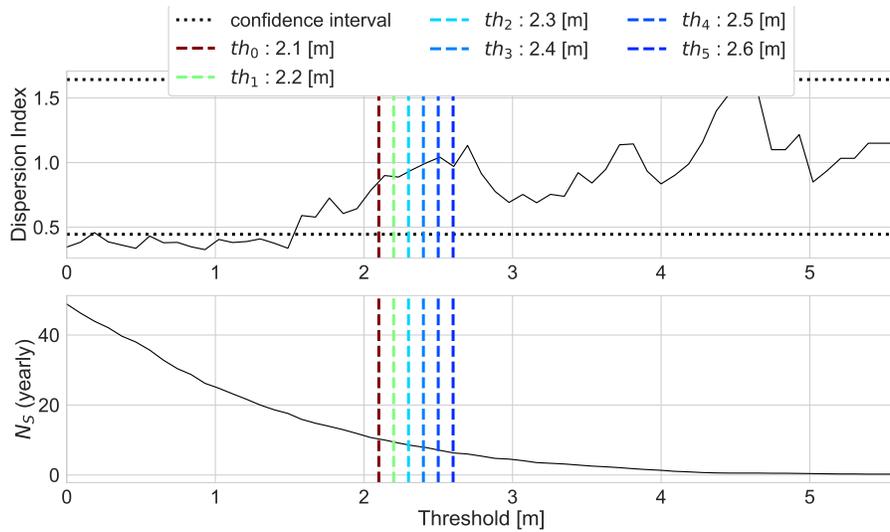


Figure D.5: The top figure contains the dispersion index for a minimal declustering time lag of 72 hours and the figure below shows the effect on the yearly average number of extremes for increasing threshold value. The colored vertical lines relate to the selected threshold values of interest.

Stability Distribution Parameter Estimates

According to the extreme value theory, the excess above the threshold should be distributed according to a Generalized Pareto distribution. The distribution is fitted to the extremes and the distribution parameters are estimated using a maximum likelihood estimation. Figure D.6 presents the resulting parameter estimates and their stability for different threshold values and a declustering time lag of 72 hours. It can be observed that the parameter estimates become unstable for threshold values $u > 3.1$. The distribution parameter estimates are considered stable for the applied threshold values on domain $[2.1, 2.6]$.

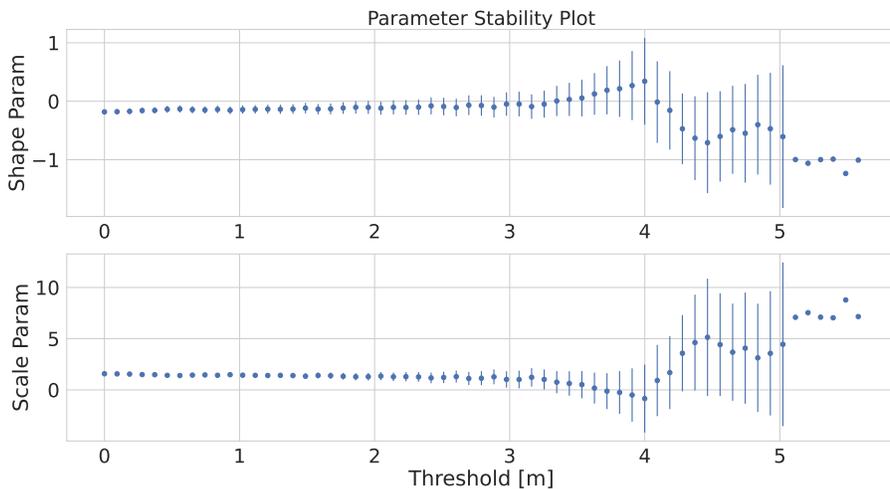


Figure D.6: The Generalized Pareto distribution parameter estimates and the corresponding uncertainty for different threshold values and a declustering time lag of 72 hours. It can be observed that the parameter estimates start to become unstable and less accurate for threshold values $u > 3.1$.

E. Offshore Conditions Experiment 1.0

Event	$H_{m0,wind}$ [m]	$T_{m-1,0,wind}$ [s]	$H_{m0,swell}$ [m]	$T_{m-1,0,swell}$ [s]	U_{10} [m/s]	ζ [m+MN75]	ζ_{min} [m+MN75]
0	5.92	9.07	2.27	11.19	22.70	0.76	-0.19
1	5.99	8.57	2.62	10.41	18.77	0.62	-0.19
2	5.87	8.38	2.56	11.12	18.97	0.63	-0.17
3	7.48	9.99	3.23	12.78	21.01	1.00	-0.07
4	5.57	8.49	2.78	11.09	21.02	0.61	-0.19
5	5.30	8.22	2.72	10.52	20.29	0.65	-0.14
6	6.05	8.56	2.38	10.59	18.21	0.78	-0.16
7	6.05	8.84	2.86	10.24	18.86	0.72	-0.11
8	5.31	8.05	2.56	10.08	20.29	0.77	-0.12
9	5.70	8.20	2.48	9.92	21.92	0.76	-0.14
10	5.21	7.91	2.72	9.82	18.53	0.72	-0.15
11	7.04	10.18	3.19	12.50	17.24	1.00	-0.12
12	6.29	9.03	3.04	11.34	25.15	1.09	-0.07
13	6.06	8.27	2.31	10.03	23.31	0.74	-0.19
14	6.06	8.51	2.42	10.00	21.18	0.74	-0.15
15	6.10	8.56	2.80	10.50	19.45	0.84	-0.09
16	5.69	8.57	3.27	10.81	19.72	0.62	-0.15
17	6.13	8.98	2.91	11.11	19.55	0.81	-0.09
18	5.82	8.59	2.56	10.62	19.04	0.62	-0.22
19	4.97	8.61	2.97	10.09	19.46	0.66	-0.14
20	5.73	8.47	2.65	11.34	19.05	0.71	-0.12
21	6.55	9.23	2.64	11.41	19.31	0.83	-0.10
22	5.91	8.50	2.63	9.92	20.04	0.68	-0.14
23	7.19	9.64	2.53	11.85	19.41	0.84	-0.10
24	5.36	8.28	2.63	10.16	20.42	0.69	-0.13
25	5.54	8.45	2.64	10.28	18.08	0.73	-0.13
26	6.29	9.11	2.58	11.31	19.52	0.72	-0.11
27	5.44	8.23	2.92	10.89	18.34	0.69	-0.12
28	5.75	8.68	2.96	10.80	20.97	0.62	-0.16
29	6.20	8.92	2.71	11.07	20.66	0.60	-0.26
30	8.30	10.38	3.49	12.56	21.92	1.23	-0.07
31	6.06	9.07	2.98	11.18	16.98	0.75	-0.15
32	6.67	9.09	2.46	10.83	19.79	0.62	-0.29
33	5.76	8.42	2.30	10.67	20.14	0.79	-0.18
34	6.13	8.30	2.44	9.83	22.93	0.80	-0.15
35	5.93	8.60	2.83	10.91	20.06	0.60	-0.21
36	5.95	8.67	3.07	10.84	27.36	0.91	-0.08
37	5.88	8.46	2.53	10.37	20.34	0.73	-0.12
38	5.97	8.43	2.86	10.42	18.57	0.65	-0.14
39	5.72	8.99	2.65	10.99	17.40	0.71	-0.15
40	5.37	8.47	3.22	11.04	19.58	0.78	-0.09
41	5.10	7.83	2.76	10.32	20.17	0.76	-0.12
42	4.98	8.10	2.93	10.47	17.27	0.68	-0.16
43	5.79	8.60	3.57	11.06	18.65	0.69	-0.12

Table E.1: Information on the offshore conditions for the selected design events corresponding to experiment 1.0.

F. Nearshore Conditions Experiment 1.0

Event	maximum water levels				minimum water levels			
	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]
0	2.31	9.61	10.27	0.86	2.14	9.59	10.25	-0.09
1	2.41	9.35	10.10	0.72	2.20	9.29	9.99	-0.09
2	2.38	9.20	9.72	0.73	2.17	9.15	9.61	-0.07
3	2.74	10.73	11.45	1.10	2.40	10.70	11.42	0.03
4	2.22	9.19	9.71	0.71	2.10	9.11	9.58	-0.09
5	2.16	8.82	9.33	0.75	2.06	8.77	9.30	-0.04
6	2.47	9.41	10.19	0.88	2.22	9.28	10.01	-0.06
7	2.42	9.46	10.22	0.82	2.24	9.36	10.15	-0.01
8	2.18	8.72	9.27	0.87	2.04	8.73	9.25	-0.02
9	2.25	9.02	9.37	0.86	2.07	8.96	9.33	-0.04
10	2.22	8.74	9.30	0.82	2.09	8.38	9.13	-0.05
11	2.74	10.81	11.55	1.10	2.40	10.82	11.53	-0.02
12	2.31	9.58	10.31	1.19	2.11	9.70	10.31	0.03
13	2.17	9.19	10.12	0.84	2.15	9.39	10.04	-0.09
14	2.41	9.46	10.14	0.84	2.17	9.32	9.92	-0.05
15	2.49	9.49	10.22	0.94	2.23	9.36	10.08	0.01
16	2.32	9.32	10.08	0.72	2.17	9.23	9.97	-0.06
17	2.43	9.62	10.29	0.90	2.25	9.53	10.25	0.01
18	2.34	9.26	9.97	0.72	2.16	9.16	9.74	-0.12
19	2.13	9.12	9.86	0.76	2.04	9.09	9.75	-0.04
20	2.34	9.18	9.74	0.81	2.17	9.09	9.58	-0.02
21	2.53	9.84	10.42	0.93	2.30	9.78	10.38	0.00
22	2.38	9.28	9.86	0.78	2.17	9.23	9.72	-0.04
23	2.66	10.33	11.21	0.94	2.37	10.26	11.07	-0.01
24	2.18	8.84	9.34	0.79	2.06	8.83	9.32	-0.03
25	2.29	9.07	9.58	0.83	2.15	8.92	9.45	-0.03
26	2.45	9.70	10.33	0.82	2.26	9.65	10.31	-0.01
27	2.26	8.94	9.42	0.79	2.13	8.80	9.36	-0.02
28	2.28	9.37	10.11	0.72	2.15	9.27	10.00	-0.06
29	2.38	9.58	10.26	0.70	2.19	9.50	10.21	-0.16
30	2.87	11.18	11.95	1.33	2.46	11.17	11.81	0.03
31	2.45	9.67	10.35	0.85	2.25	9.60	10.32	-0.05
32	2.50	9.72	10.35	0.72	2.23	9.69	10.32	-0.19
33	2.33	9.13	9.53	0.89	2.13	9.09	9.47	-0.08
34	2.22	9.29	10.19	0.90	2.19	9.51	10.32	-0.05
35	2.36	9.37	10.10	0.70	2.17	9.30	9.97	-0.11
36	2.08	9.12	10.18	1.01	1.93	9.30	10.18	0.02
37	2.37	9.23	9.72	0.83	2.17	9.20	9.63	-0.02
38	2.44	9.36	10.13	0.75	2.21	9.22	9.89	-0.04
39	2.36	9.53	10.27	0.81	2.20	9.47	10.24	-0.05
40	2.25	9.23	9.98	0.88	2.13	9.15	9.83	0.01
41	2.14	8.66	9.24	0.86	2.01	8.65	9.20	-0.02
42	2.15	8.72	9.32	0.78	2.05	8.63	9.29	-0.06
43	2.38	9.44	10.20	0.79	2.22	9.34	10.14	-0.02

Table F.1: Information on the nearshore conditions ζ for the selected design events corresponding to experiment 1.0.

G. Results Experiment 1.0

BT140								
Event	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Costs [€/m]			
Conventional Approach								
-	4.07	0.81 $V = 1 m^3$	1.03 HM 1000-4000	0.53 HM 300-600	10300 €/m			
Vine-based Approach								
0	2.79	0.63 $V = 1 m^3$	0.76 HM 500-2000	0.47 HM 300-600	-16%			
1	2.65	0.66 $V = 1 m^3$	0.76 HM 500-2000	0.48 HM 300-600	-18%			
2	2.61	0.65 $V = 1 m^3$	0.74 HM 500-2000	0.48 HM 300-600	-18%			
3	3.60	0.75 $V = 1 m^3$	0.95 HMA 1000-3000	0.52 HM 300-600	-6%			
4	2.48	0.61 $V = 1 m^3$	0.71 HM 500-2000	0.46 LM 10-500	-20%			
5	2.40	0.59 $V = 1 m^3$	0.67 HM 500-2000	0.45 LM 10-500	-21%			
6	2.86	0.67 $V = 1 m^3$	0.78 HM 500-2000	0.49 HM 300-600	-15%			
7	2.78	0.66 $V = 1 m^3$	0.77 HM 500-2000	0.49 HM 300-600	-16%			
8	2.55	0.60 $V = 1 m^3$	0.67 HM 500-2000	0.45 LM 10-500	-19%			
9	2.60	0.61 $V = 1 m^3$	0.70 HM 500-2000	0.45 LM 10-500	-19%			
10	2.53	0.61 $V = 1 m^3$	0.67 HM 500-2000	0.46 LM 10-500	-19%			
11	3.64	0.75 $V = 1 m^3$	0.96 HMA 1000-3000	0.52 HM 300-600	-5%			
12	3.12	0.63 $V = 1 m^3$	0.76 HM 500-2000	0.46 LM 10-500	-12%			
13	2.58	0.59 $V = 1 m^3$	0.70 HM 500-2000	0.47 HM 300-600	-19%			
14	2.79	0.66 $V = 1 m^3$	0.77 HM 500-2000	0.48 HM 300-600	-16%			
15	2.95	0.68 $V = 1 m^3$	0.79 HM 500-2000	0.48 HM 300-600	-14%			
16	2.59	0.63 $V = 1 m^3$	0.74 HM 500-2000	0.48 HM 300-600	-19%			
17	2.91	0.66 $V = 1 m^3$	0.79 HM 500-2000	0.49 HM 300-600	-15%			
18	2.58	0.64 $V = 1 m^3$	0.74 HM 500-2000	0.48 HM 300-600	-19%			
19	2.47	0.58 $V = 1 m^3$	0.69 HM 500-2000	0.45 LM 10-500	-20%			
20	2.65	0.64 $V = 1 m^3$	0.73 HM 500-2000	0.47 HM 300-600	-18%			
21	3.06	0.69 $V = 1 m^3$	0.83 HMA 1000-3000	0.50 HM 300-600	-13%			
22	2.67	0.65 $V = 1 m^3$	0.75 HM 500-2000	0.48 HM 300-600	-18%			
23	3.28	0.73 $V = 1 m^3$	0.90 HMA 1000-3000	0.52 HM 300-600	-10%			
24	2.46	0.60 $V = 1 m^3$	0.68 HM 500-2000	0.45 LM 10-500	-20%			
25	2.62	0.63 $V = 1 m^3$	0.72 HM 500-2000	0.47 LM 10-500	-18%			
26	2.86	0.67 $V = 1 m^3$	0.80 HM 500-2000	0.49 HM 300-600	-15%			
27	2.55	0.62 $V = 1 m^3$	0.70 HM 500-2000	0.47 LM 10-500	-19%			
28	2.58	0.62 $V = 1 m^3$	0.74 HM 500-2000	0.47 HM 300-600	-19%			
29	2.67	0.65 $V = 1 m^3$	0.78 HM 500-2000	0.49 HM 300-600	-18%			
30	4.07	0.78 $V = 1 m^3$	1.02 HM 1000-4000	0.53 HM 300-600	-0%			
31	2.89	0.67 $V = 1 m^3$	0.80 HM 500-2000	0.49 HM 300-600	-15%			

Event	Crest level h_c [m+MN75]	Armour front D_{n50} [m]	Armour rear D_{n50} [m]	Toe armour D_{n50} [m]	Costs [€/m]
32	2.79	0.68 $V = 1 m^3$	0.81 HMA 1000-3000	0.50 HM 300-600	-16%
33	2.71	0.64 $V = 1 m^3$	0.73 HM 500-2000	0.47 LM 10-500	-17%
34	2.70	0.61 $V = 1 m^3$	0.72 HM 500-2000	0.48 HM 300-600	-17%
35	2.60	0.64 $V = 1 m^3$	0.75 HM 500-2000	0.48 HM 300-600	-19%
36	2.68	0.57 $V = 1 m^3$	0.68 HM 500-2000	0.42 LM 10-500	-18%
37	2.70	0.65 $V = 1 m^3$	0.74 HM 500-2000	0.47 LM 10-500	-17%
38	2.69	0.66 $V = 1 m^3$	0.77 HM 500-2000	0.48 HM 300-600	-17%
39	2.75	0.64 $V = 1 m^3$	0.77 HM 500-2000	0.48 HM 300-600	-17%
40	2.68	0.61 $V = 1 m^3$	0.72 HM 500-2000	0.46 LM 10-500	-18%
41	2.49	0.58 $V = 1 m^3$	0.65 HM 500-2000	0.44 LM 10-500	-20%
42	2.42	0.59 $V = 1 m^3$	0.66 HM 500-2000	0.45 LM 10-500	-21%
43	2.72	0.65 $V = 1 m^3$	0.76 HM 500-2000	0.48 HM 300-600	-17%

Table G.1: The required cross-sectional design of the breakwater for all design events based on prescribed extreme storm conditions with an annual probability of occurrence once every 100 years.

H. Bivariate Scatterplots

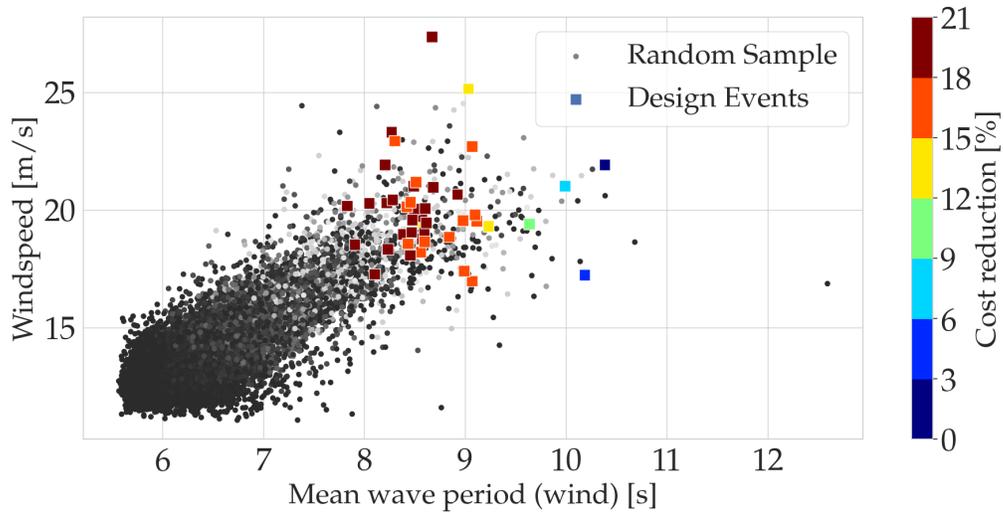


Figure H.1: The joint occurrences of the wind speed U_{10} and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the obtained cost reduction with respect to the conventional design.

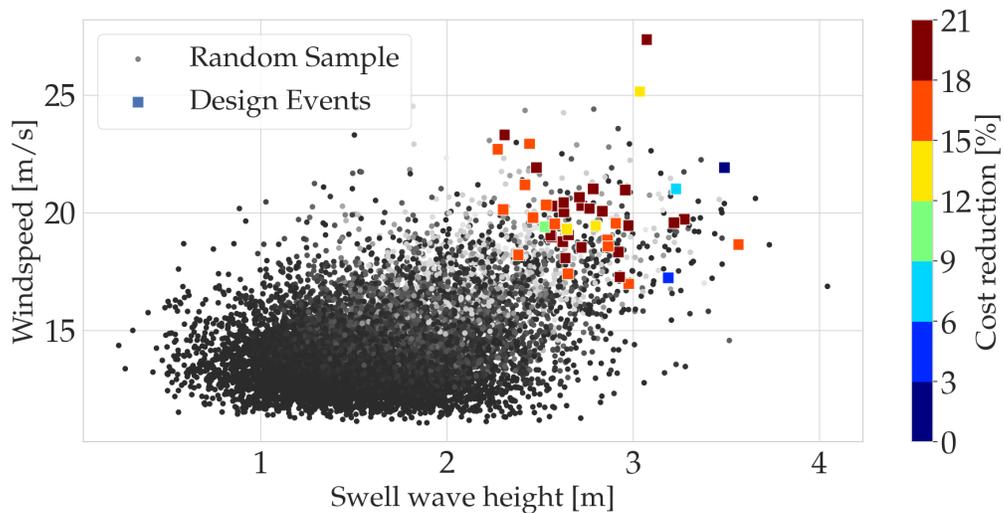


Figure H.2: The joint occurrences of the significant wave height (wind system) $H_{m0,swell}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the obtained cost reduction with respect to the conventional design.

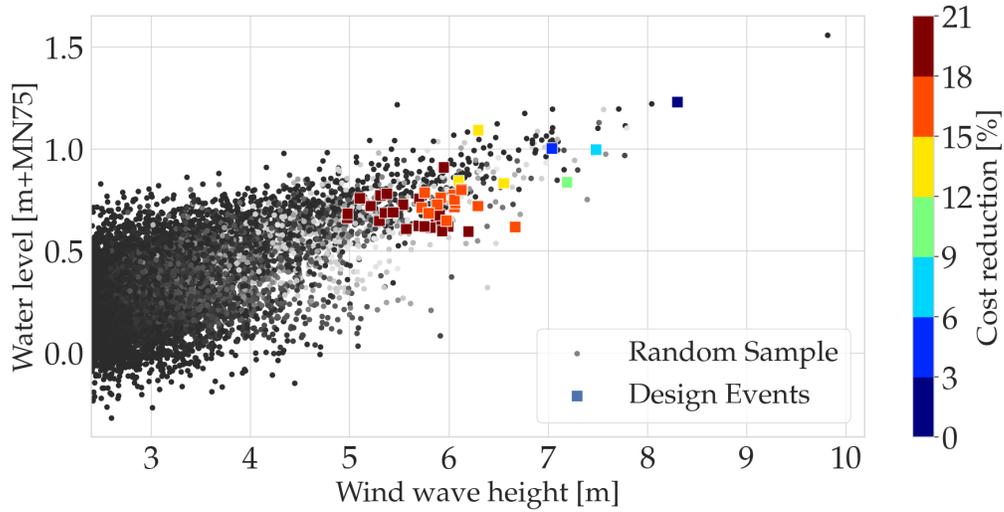


Figure H.3: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the water level ζ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the obtained cost reduction with respect to the conventional design..

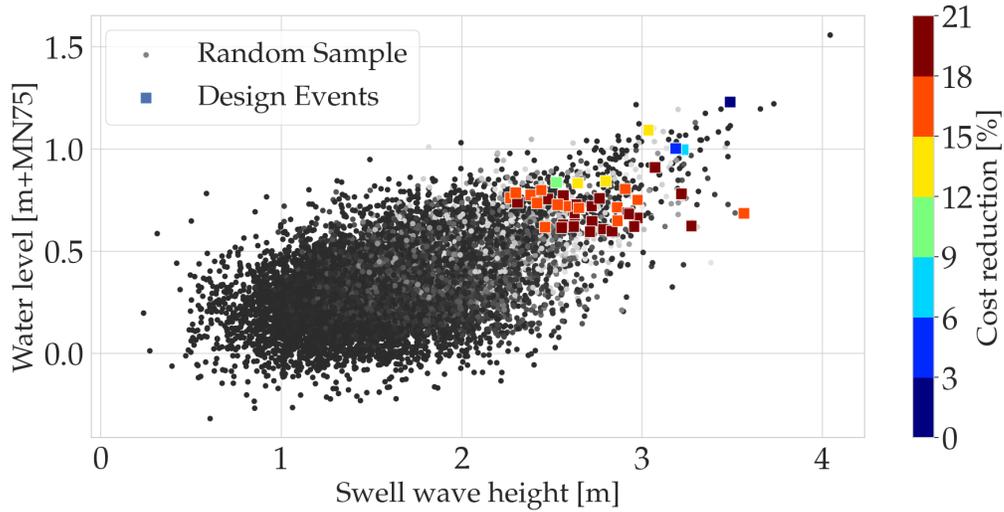


Figure H.4: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the water level ζ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the obtained cost reduction with respect to the conventional design.

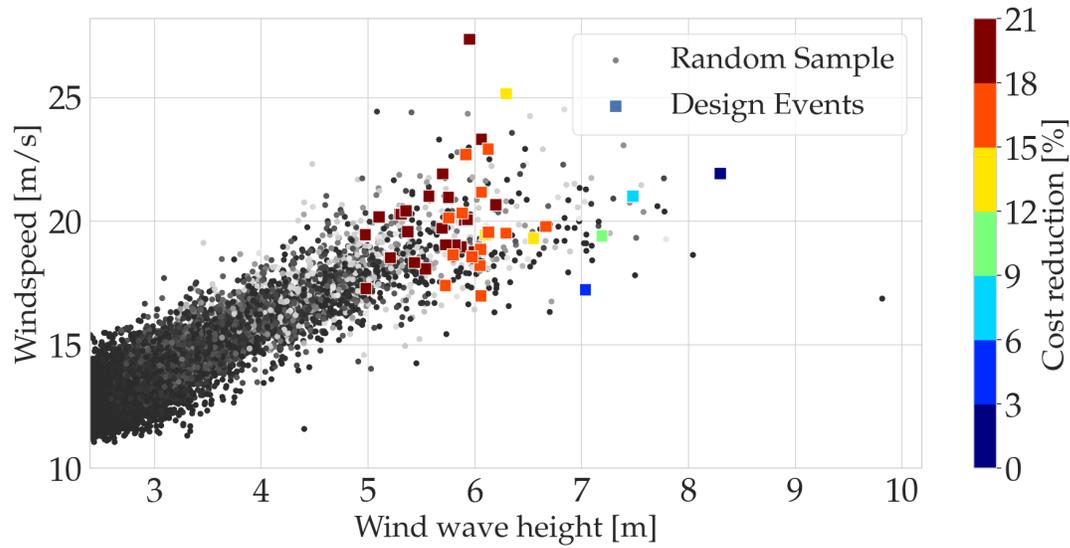


Figure H.5: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the obtained cost reduction with respect to the conventional design.

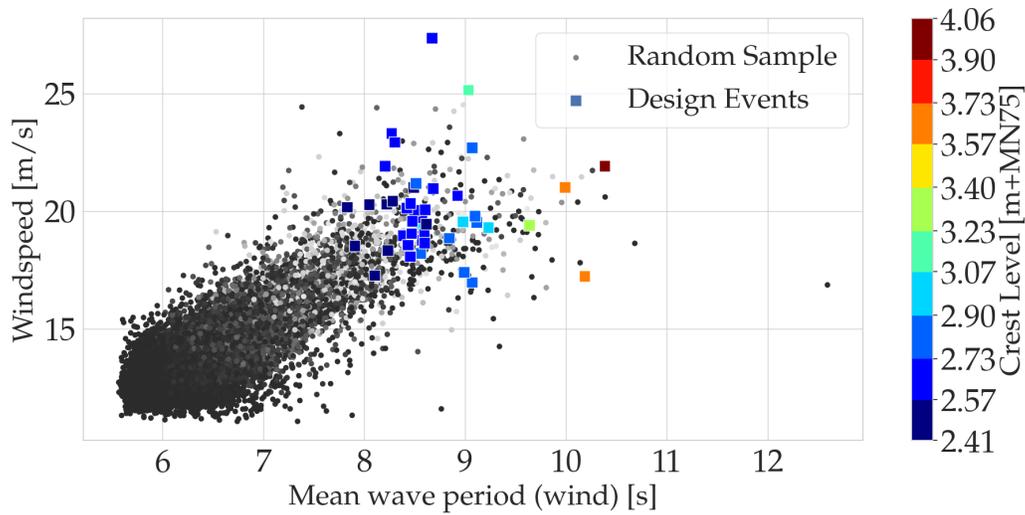


Figure H.6: The joint occurrences of the wind speed U_{10} and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event.

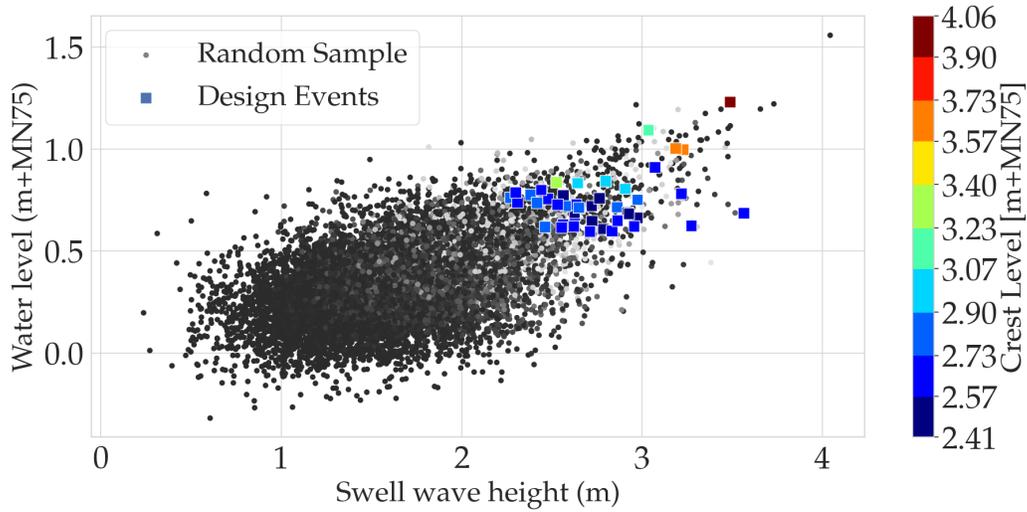


Figure H.7: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the water level ζ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event.

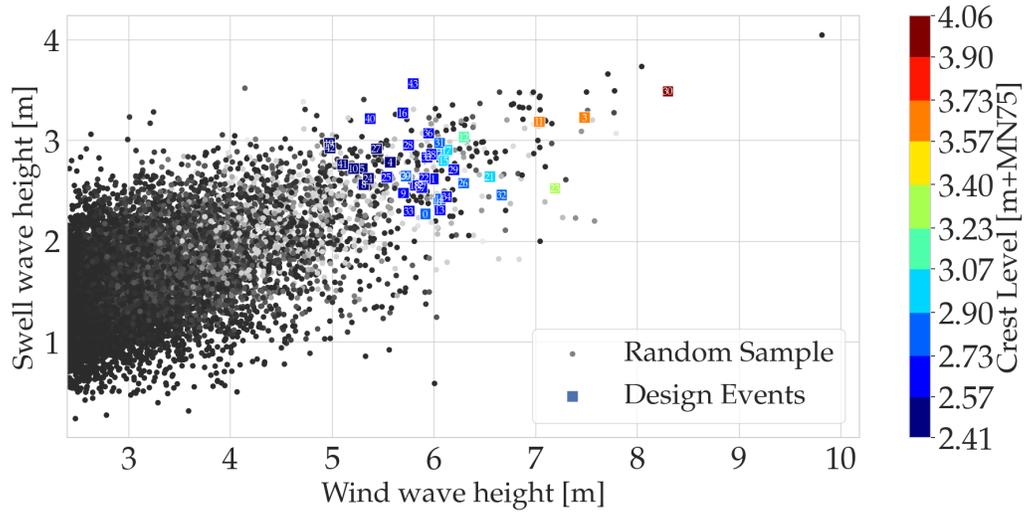


Figure H.8: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the significant wave height (swell system) $H_{m0,swell}$ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event.

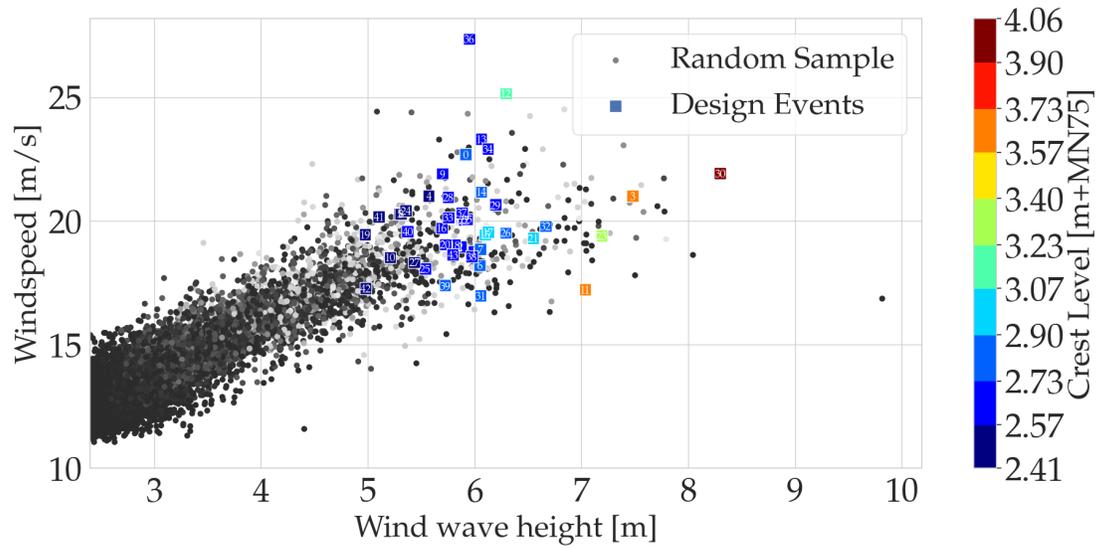


Figure H.9: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event.

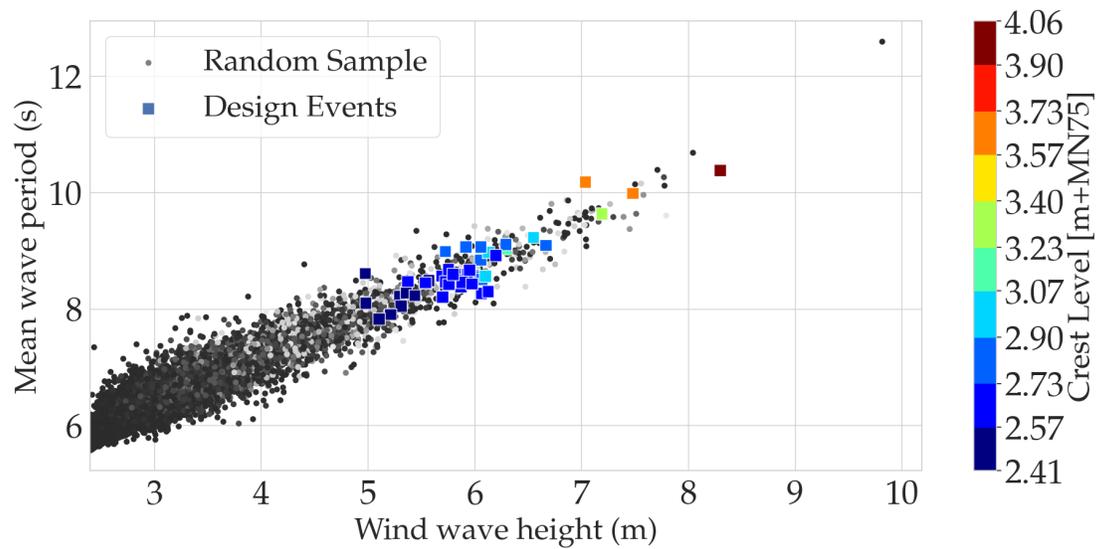


Figure H.10: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares indicate the joint events that belong to the set of design events. The color gradient indicates the minimal required crest level for the corresponding design event.

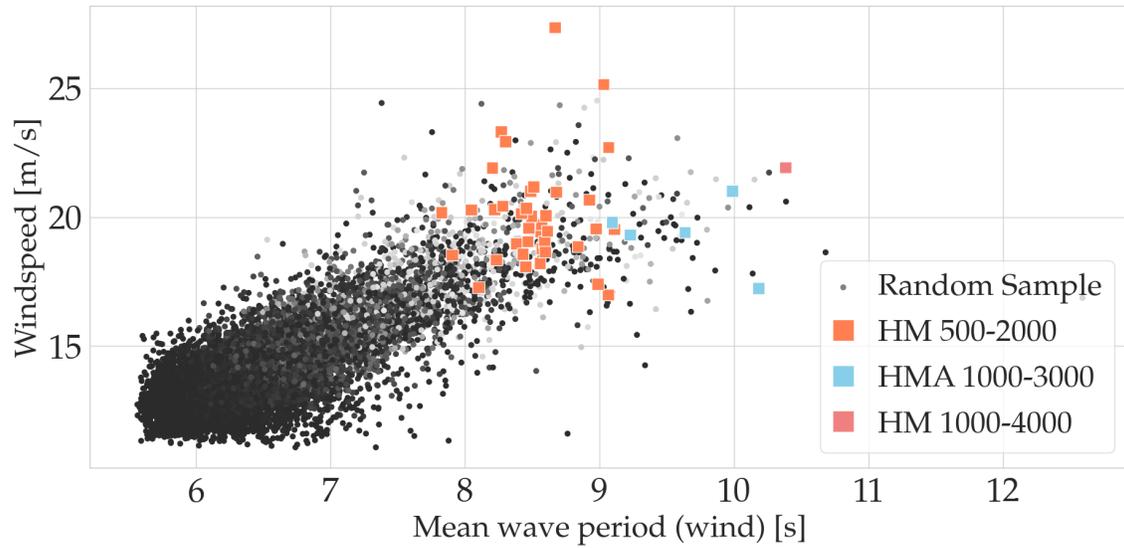


Figure H.11: The joint occurrences of the wind speed U_{10} and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

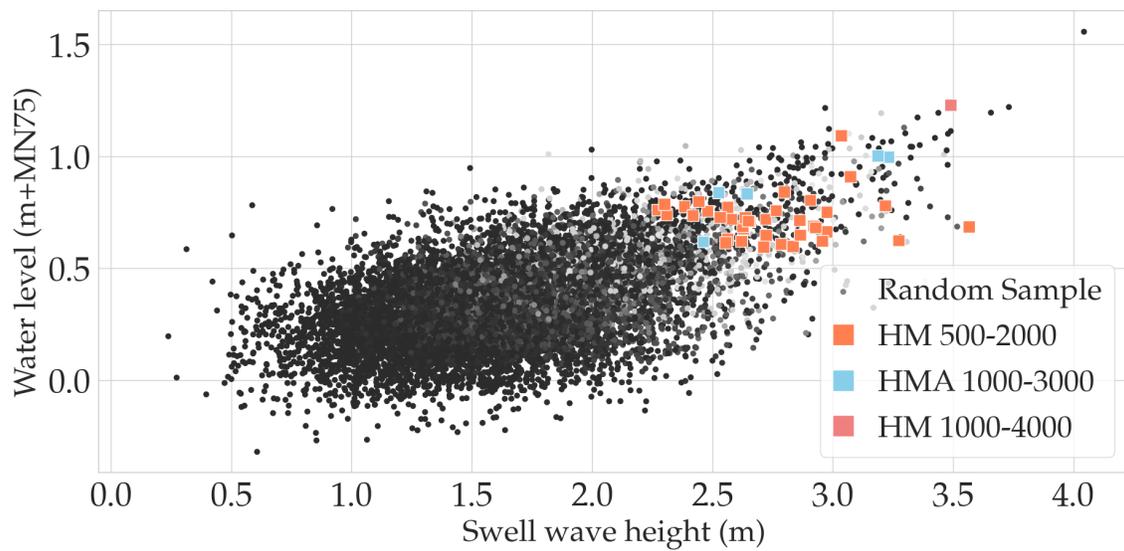


Figure H.12: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the water level ζ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

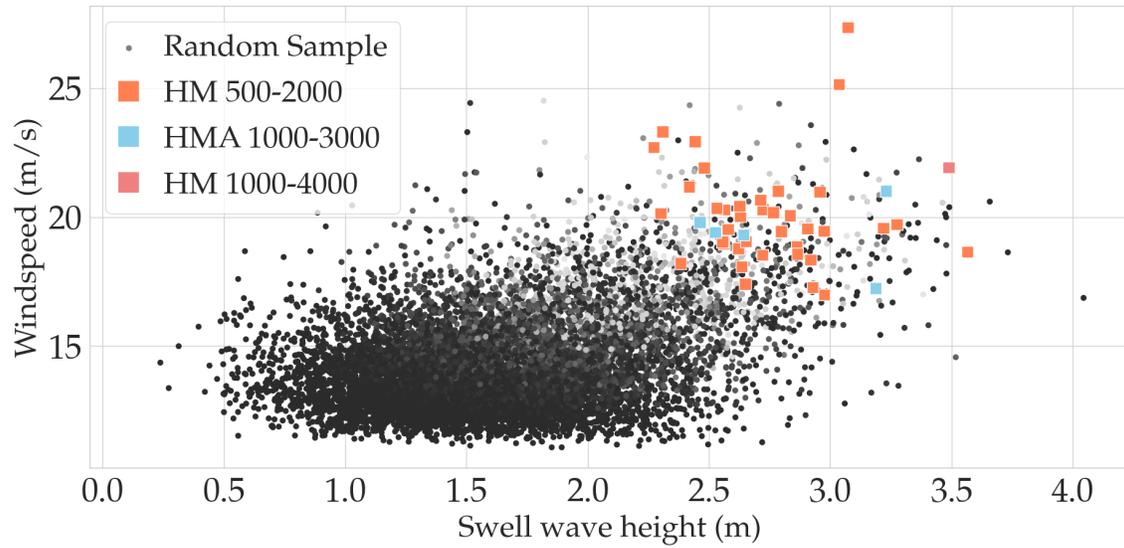


Figure H.13: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

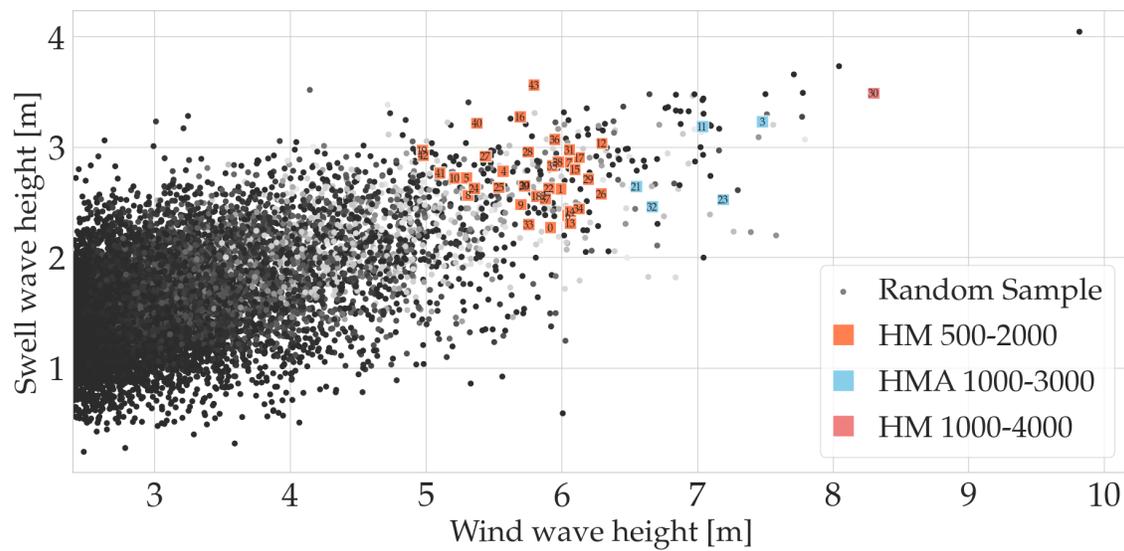


Figure H.14: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the significant wave height (swell system) $H_{m0,swell}$ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

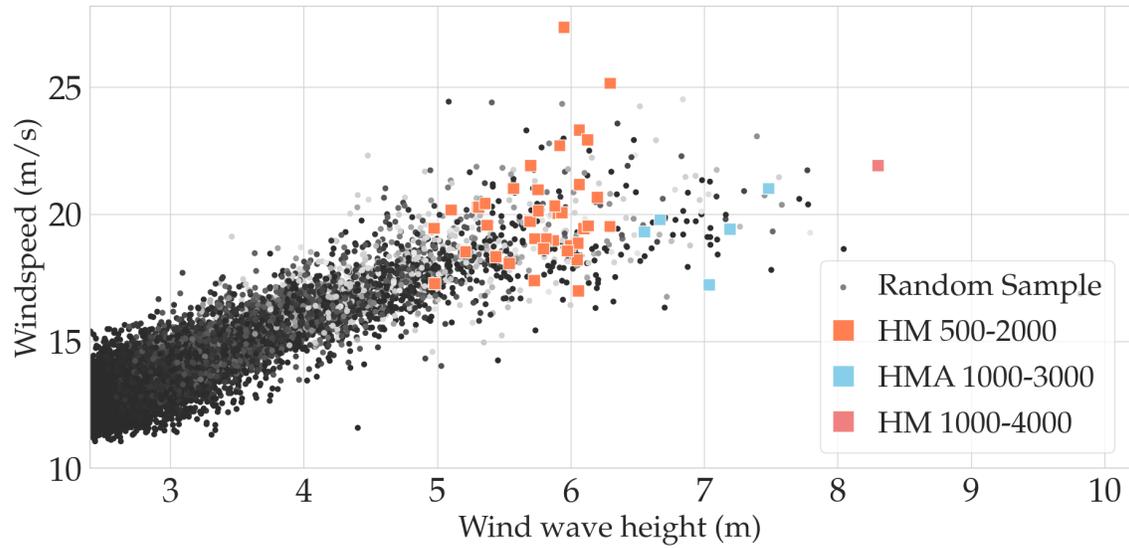


Figure H.15: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

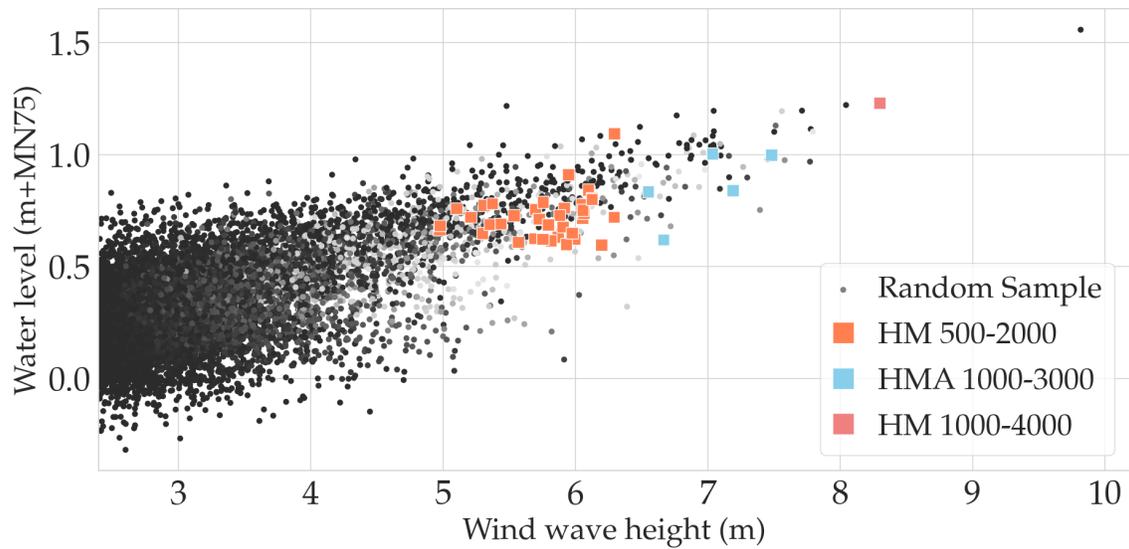


Figure H.16: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the water level ζ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

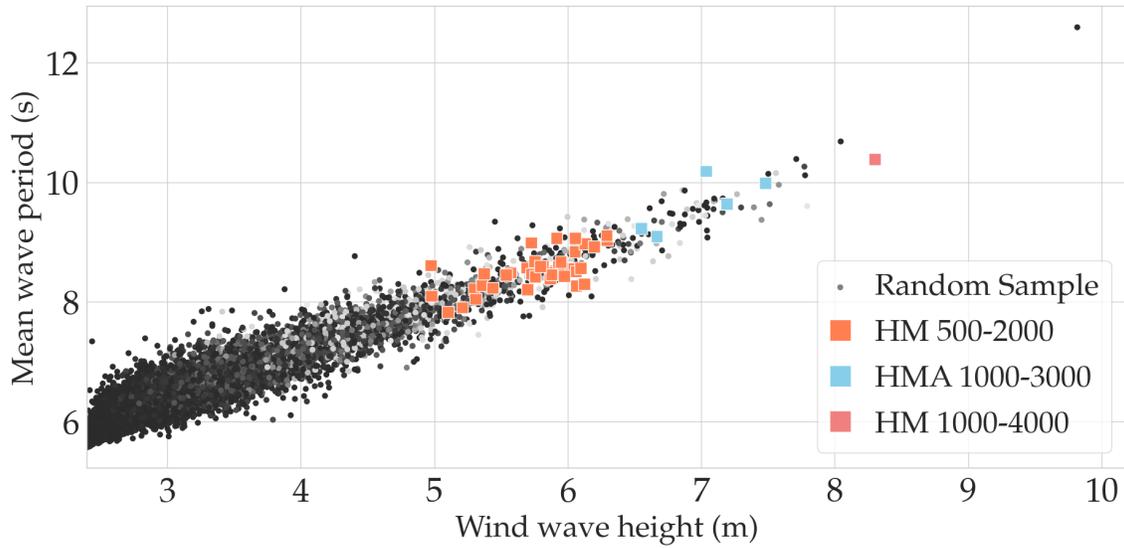


Figure H.17: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required rear armour class.

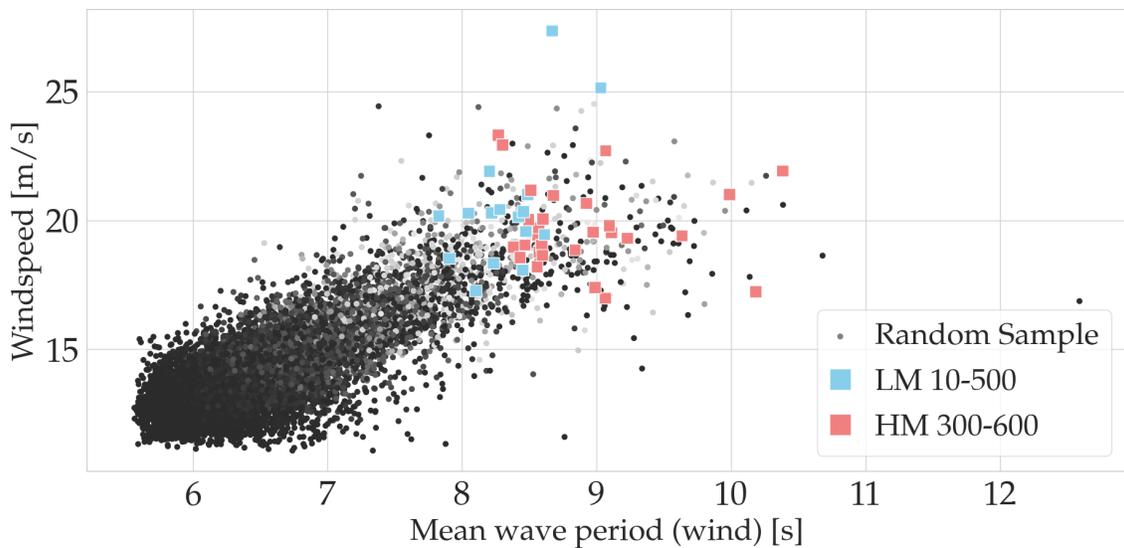


Figure H.18: The joint occurrences of the wind speed U_{10} and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

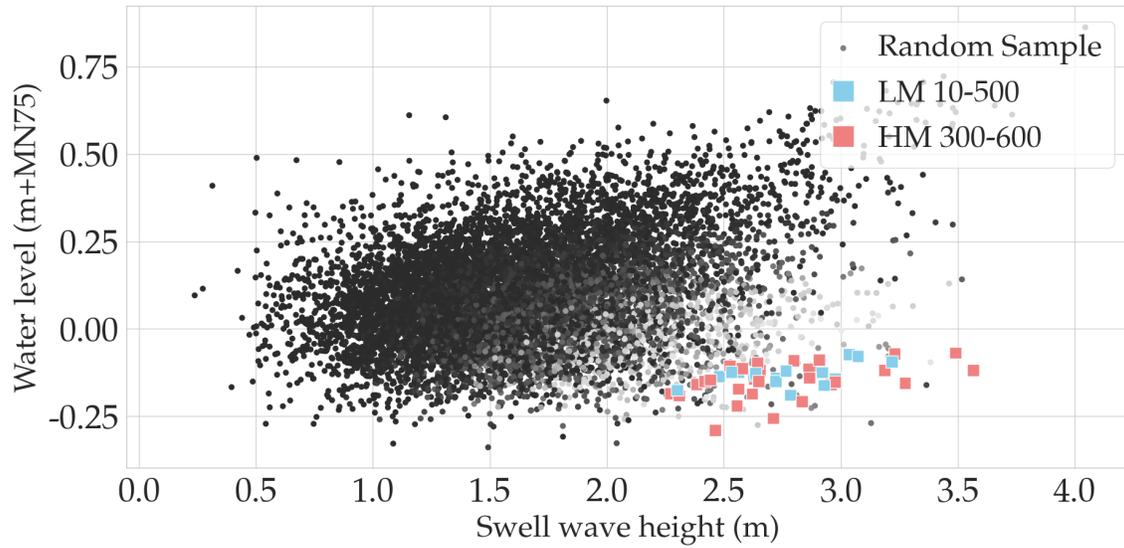


Figure H.19: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the minimum water level ζ_{min} representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

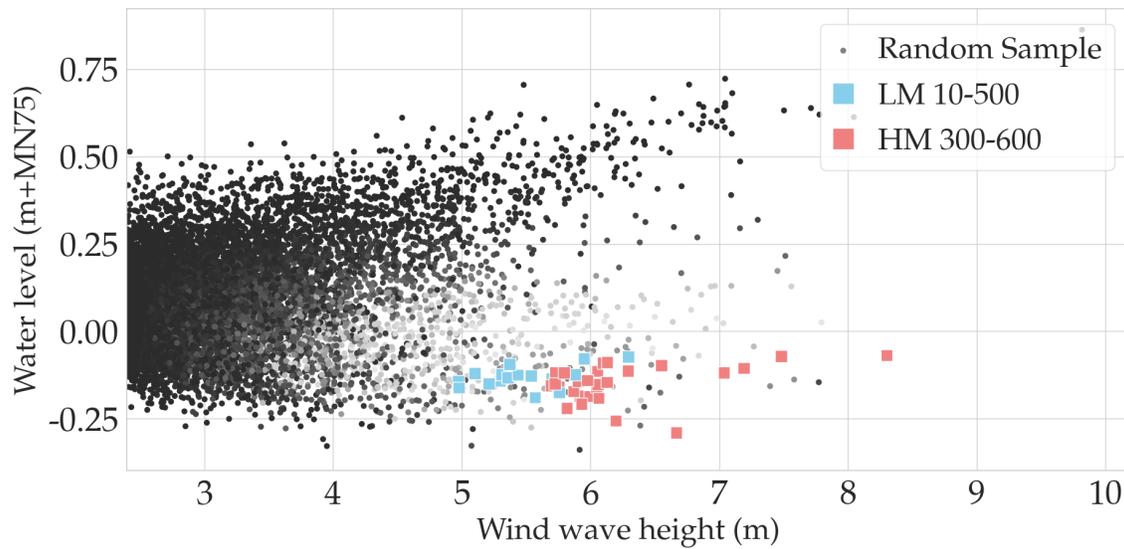


Figure H.20: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the minimum water level ζ_{min} representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

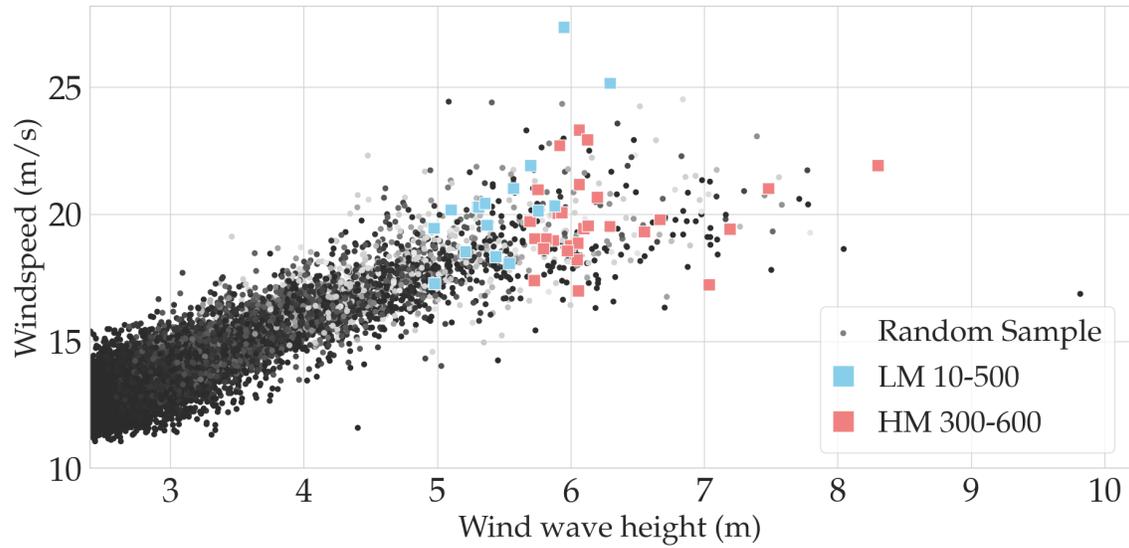


Figure H.21: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

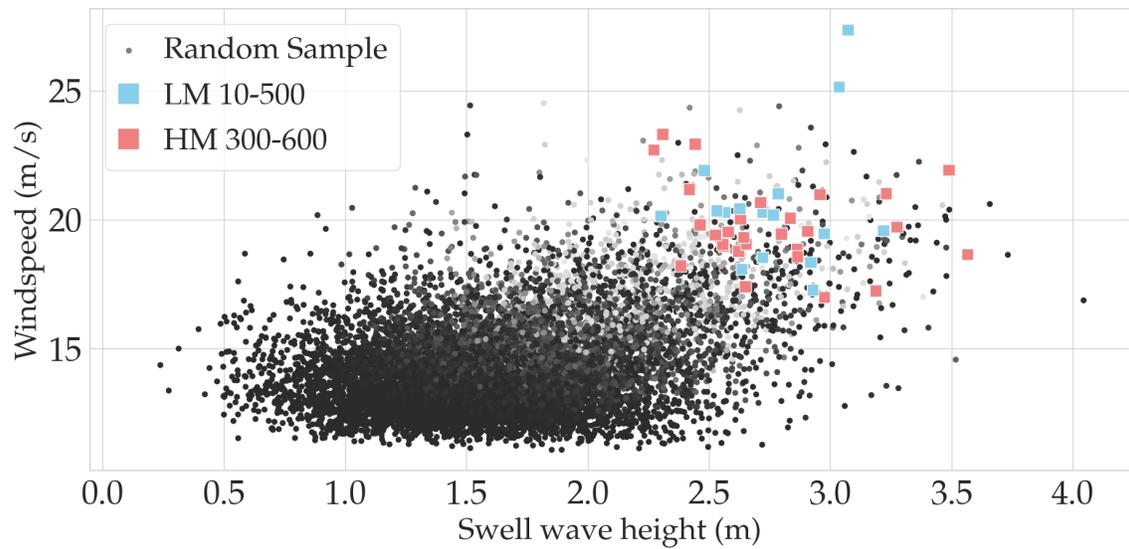


Figure H.22: The joint occurrences of the significant wave height (swell system) $H_{m0,swell}$ and the wind speed U_{10} representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

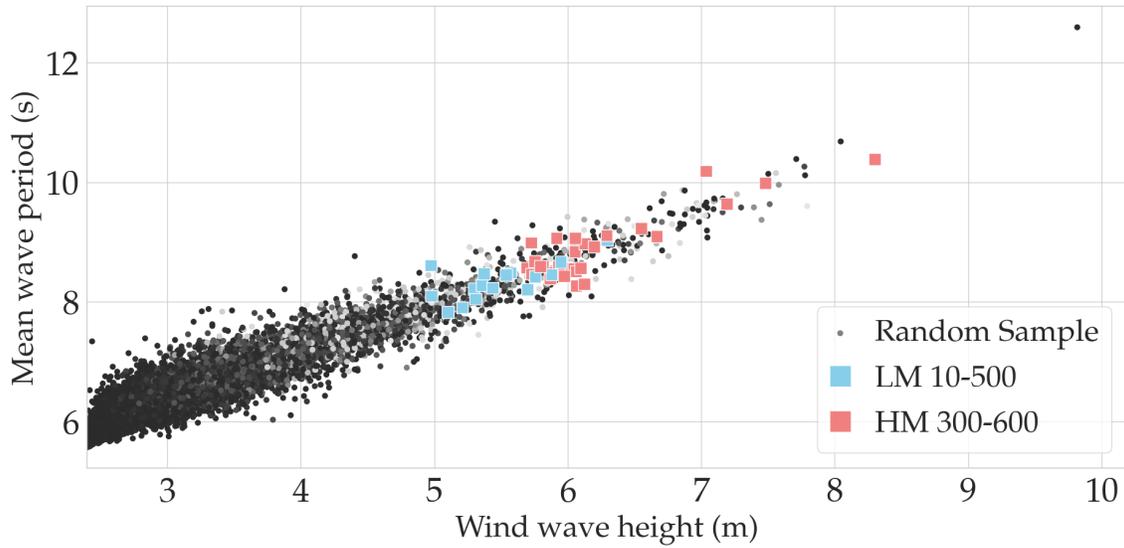


Figure H.23: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the mean wave period (wind system) $T_{m-1,0,wind}$ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

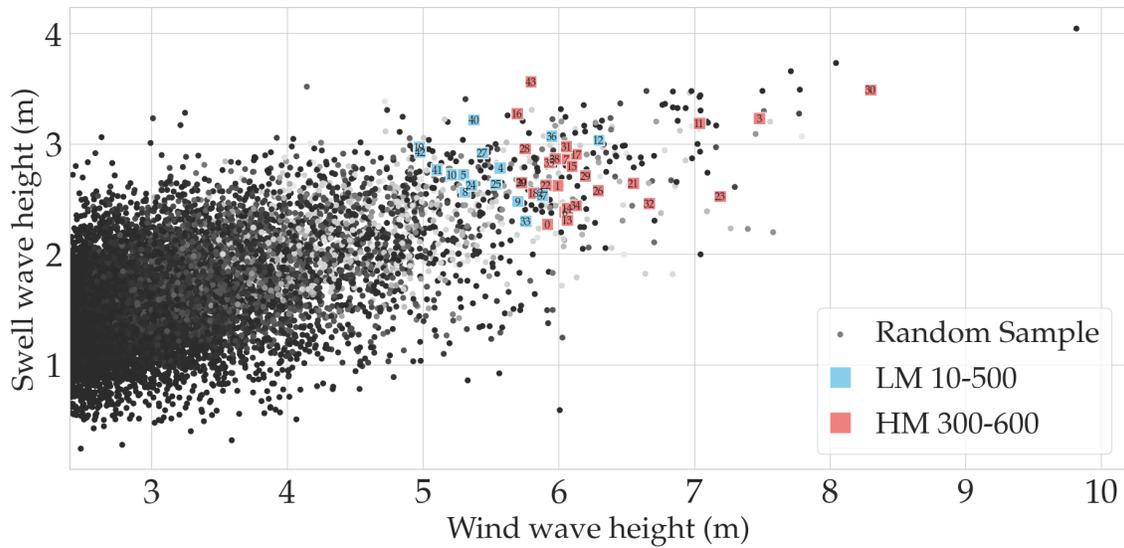


Figure H.24: The joint occurrences of the significant wave height (wind system) $H_{m0,wind}$ and the significant wave height (swell system) $H_{m0,swell}$ representing the random sample obtained from the vine copula. The squares represent the joint events that belong to the set of design events and the color of each square indicates the minimal required toe armour class.

I. Offshore Conditions All Experiments

	H_{m0} [m]	T_p [s]	ζ [m+MN75]	ζ_{min} [m+MN75]	U_{10} [m/s]
n = 44	Experiment 1.0 ($H_{m0,wind}$, 48 hours, 100 year, -)				
Conventional	7.64	9.81	1.18	-0.27	21.42
Mean	6.03	9.77	0.75	-0.14	20.04
Standard deviation	0.62	0.58	0.13	0.05	1.99
Minimum	5.12	8.91	0.60	-0.29	16.98
Maximum	8.36	11.60	1.23	-0.07	27.36
n = 41	Experiment 1.1 ($H_{m0,swell}$, 48 hours, 100 year, -)				
Conventional	7.64	9.81	1.18	-0.27	21.42
Mean	4.34	8.62	0.49	-0.03	16.58
Standard deviation	0.53	0.46	0.08	0.05	1.41
Minimum	3.39	7.84	0.36	-0.18	14.09
Maximum	6.18	10.06	0.72	0.03	19.81
n = 39	Experiment 1.2 ($H_{m0,wind}$, 72 hours, 100 year, -)				
Conventional	7.64	9.81	1.18	-0.27	21.42
Mean	6.51	10.28	0.87	-0.11	19.81
Standard deviation	0.71	0.56	0.12	0.05	1.69
Minimum	5.49	9.49	0.72	-0.21	16.95
Maximum	8.08	11.51	1.27	-0.04	24.38
n = 34	Experiment 1.3 ($H_{m0,wind}$, 48 hours, 150 year, -)				
Conventional	7.86	9.82	1.23	-0.27	21.69
Mean	6.18	9.87	0.74	-0.15	20.18
Standard deviation	0.51	0.52	0.09	0.05	1.72
Minimum	5.18	8.96	0.61	-0.28	17.64
Maximum	7.46	11.00	0.95	-0.09	24.32
n = 34	Experiment 2.0 ($H_{m0,wind}$, 48 hours, 100 year, ζ & ζ_{min})				
Conventional	7.64	9.81	0.80	0.00	21.42
Mean	7.11	10.74	0.80	0.00	20.66
Standard deviation	0.58	0.57	0.00	0.00	1.00
Minimum	6.03	9.97	0.80	0.00	19.59
Maximum	8.51	12.26	0.80	0.00	23.82
n = 37	Experiment 2.1 ($H_{m0,wind}$, 48 hours, 100 year, ζ_{min})				
Conventional	7.64	9.81	1.18	0.00	21.42
Mean	7.24	10.94	1.04	0.00	20.64
Standard deviation	0.79	0.85	0.14	0.00	1.22
Minimum	6.24	9.94	0.86	0.00	19.43
Maximum	10.32	14.89	1.68	0.00	24.10
n = 57	Experiment 3.0 (H_{m0} , 48 hours, 100 year, combined)				
Conventional	7.64	9.81	1.18	-0.27	21.42
Mean	6.09	9.49	0.75	-0.16	19.03
Standard deviation	0.73	0.35	0.09	0.07	1.46
Minimum	4.94	9.00	0.63	-0.35	17.18
Maximum	7.89	10.41	1.06	-0.06	25.40

Table I.1: General statistics on the set of design events (offshore conditions) for different experiments. The wind and swell sea state have been combined and the peak wave period has been determined which is required as input for the SWAN model performing the offshore-nearshore transformation.

J. Nearshore Conditions All Experiments

	maximum water levels				minimum water levels			
	H_{m0}^{tot} [m]	$T_{m-1.0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]	H_{m0}^{tot} [m]	$T_{m-1.0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]
n = 44	Experiment 1.0 ($H_{m0,wind}$, 48 hours, 100 year, -)							
Conventional	2.96	11.17	12.52	1.28	2.38	11.15	12.51	-0.17
Mean	2.36	9.41	10.07	0.85	2.17	9.35	9.99	-0.04
Standard deviation	0.17	0.53	0.59	0.13	0.10	0.55	0.60	0.05
Minimum	2.08	8.67	9.24	0.70	1.93	8.38	9.13	-0.19
Maximum	2.87	11.18	11.95	1.33	2.46	11.17	11.81	0.03
n = 41	Experiment 1.1 ($H_{m0,swell}$, 48 hours, 100 year, -)							
Conventional	2.96	11.17	12.52	1.28	2.38	11.16	12.51	-0.17
Mean	1.87	8.01	8.67	0.59	1.86	7.90	8.62	0.07
Standard deviation	0.17	0.50	0.51	0.08	0.15	0.50	0.51	0.05
Minimum	1.54	7.08	7.73	0.46	1.54	7.03	7.67	-0.08
Maximum	2.37	9.60	10.28	0.82	2.28	9.53	10.25	0.13
n = 39	Experiment 1.2 ($H_{m0,wind}$, 72 hours, 100 year, -)							
Conventional	2.96	11.17	12.52	1.28	2.38	11.15	12.51	-0.17
Mean	2.48	9.83	10.54	0.97	2.27	9.78	10.45	-0.01
Standard deviation	0.18	0.54	0.55	0.12	0.10	0.56	0.58	0.05
Minimum	2.08	9.11	9.61	0.82	2.09	9.02	9.52	-0.11
Maximum	2.92	11.03	11.71	1.37	2.48	11.01	11.65	0.06
n = 34	Experiment 1.3 ($H_{m0,wind}$, 48 hours, 150 year, -)							
Conventional	3.01	11.38	12.55	1.33	2.39	11.37	12.54	-0.18
Mean	2.39	9.51	10.20	0.84	2.20	9.45	10.11	-0.05
Standard deviation	0.17	0.46	0.52	0.09	0.10	0.47	0.53	0.05
Minimum	1.86	8.46	9.25	0.71	1.91	8.57	9.20	-0.18
Maximum	2.72	10.56	11.38	1.05	2.40	10.50	11.34	0.01
n = 34	Experiment 2.0 ($H_{m0,wind}$, 48 hours, 100 year, ζ & ζ_{min})							
Conventional	2.82	11.16	12.52	0.90	2.50	11.15	12.51	0.10
Mean	2.59	10.32	11.02	0.90	2.38	10.27	10.96	0.10
Standard deviation	0.12	0.52	0.59	0.00	0.08	0.53	0.59	0.00
Minimum	2.30	9.59	10.26	0.90	2.18	9.55	10.23	0.10
Maximum	2.84	11.72	12.62	0.90	2.54	11.72	12.60	0.10
n = 37	Experiment 2.1 ($H_{m0,wind}$, 48 hours, 100 year, ζ_{min})							
Conventional	2.96	11.17	12.52	1.28	2.50	11.15	12.51	0.10
Mean	2.64	10.39	11.10	1.12	2.38	10.35	11.04	0.10
Standard deviation	0.17	0.53	0.60	0.10	0.08	0.54	0.59	0.00
Minimum	2.31	9.48	10.26	0.96	2.26	9.50	10.21	0.10
Maximum	2.94	11.36	12.26	1.31	2.51	11.35	12.19	0.10
n = 57	Experiment 3.0 (H_{m0} , 48 hours, 100 year, combined)							
Conventional	2.96	11.17	12.52	1.28	2.38	11.15	12.51	-0.17
Mean	2.41	9.47	10.20	0.85	2.20	9.39	10.13	-0.06
Standard deviation	0.20	0.67	0.87	0.09	0.12	0.71	0.90	0.07
Minimum	2.07	8.58	9.24	0.73	1.98	8.43	9.18	-0.25
Maximum	2.90	11.56	12.59	1.16	2.51	11.55	12.58	0.03

Table J.1: General statistics on the design events (nearshore conditions) derived using SWAN model for the output location BT140

	maximum water levels				minimum water levels			
	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]	H_{m0}^{tot} [m]	$T_{m-1,0}^{tot}$ [s]	T_p [s]	ζ [m+MN75]
n = 44	Experiment 1.0 ($H_{m0,wind}$, 48 hours, 100 year, -)							
Conventional	2.46	10.86	12.56	1.30	1.74	10.26	12.53	-0.15
Mean	1.87	9.17	10.89	0.87	1.54	8.68	10.24	-0.02
Standard deviation	0.11	0.54	0.94	0.13	0.07	0.54	0.61	0.05
Minimum	1.67	8.20	9.46	0.71	1.42	7.74	9.25	-0.17
Maximum	2.27	10.84	12.45	1.35	1.76	10.32	11.92	0.05
n = 41	Experiment 1.1 ($H_{m0,swell}$, 48 hours, 100 year, -)							
Conventional	2.46	10.86	12.56	1.30	1.74	10.28	12.54	-0.15
Mean	1.47	7.83	8.99	0.61	1.33	7.41	8.79	0.09
Standard deviation	0.12	0.50	0.67	0.08	0.10	0.49	0.55	0.05
Minimum	1.23	6.84	7.92	0.48	1.11	6.47	7.79	-0.06
Maximum	1.80	9.28	11.14	0.84	1.62	8.93	10.34	0.15
n = 39	Experiment 1.2 ($H_{m0,wind}$, 72 hours, 100 year, -)							
Conventional	2.46	10.86	12.56	1.30	1.74	10.26	12.53	-0.15
Mean	1.96	9.61	10.83	0.99	1.60	9.11	10.59	0.01
Standard deviation	0.12	0.55	0.63	0.12	0.08	0.57	0.54	0.05
Minimum	1.76	8.49	10.00	0.84	1.47	8.24	9.72	-0.09
Maximum	2.30	10.86	12.44	1.39	1.76	10.37	11.73	0.08
n = 34	Experiment 1.3 ($H_{m0,wind}$, 48 hours, 150 year, -)							
Conventional	2.51	11.06	12.59	1.35	1.75	10.44	12.56	-0.16
Mean	1.89	9.26	10.92	0.86	1.56	8.77	10.29	-0.03
Standard deviation	0.10	0.53	0.79	0.09	0.06	0.50	0.53	0.05
Minimum	1.62	7.79	9.40	0.73	1.42	7.57	9.29	-0.16
Maximum	2.10	10.34	12.46	1.07	1.69	9.85	11.42	0.03
n = 34	Experiment 2.0 ($H_{m0,wind}$, 48 hours, 100 year, ζ & ζ_{min})							
Conventional	2.28	10.74	12.55	0.92	1.89	10.38	12.54	0.12
Mean	2.01	10.01	11.24	0.92	1.70	9.59	11.07	0.12
Standard deviation	0.07	0.54	0.61	0.00	0.05	0.58	0.59	0.00
Minimum	1.85	9.00	10.36	0.92	1.58	8.59	10.31	0.12
Maximum	2.14	11.39	12.69	0.92	1.79	11.07	12.66	0.12
n = 37	Experiment 2.1 ($H_{m0,wind}$, 48 hours, 100 year, ζ_{min})							
Conventional	2.46	10.86	12.56	1.30	1.89	10.38	12.54	0.12
Mean	2.08	10.12	11.27	1.14	1.70	9.66	11.16	0.12
Standard deviation	0.10	0.57	0.60	0.10	0.05	0.58	0.59	0.00
Minimum	1.90	9.18	10.34	0.98	1.62	8.80	10.31	0.12
Maximum	2.29	11.16	12.42	1.33	1.79	10.73	12.32	0.12
n = 57	Experiment 3.0 (H_{m0} , 48 hours, 100 year, combined)							
Conventional	2.46	10.86	12.56	1.30	1.74	10.26	12.53	-0.15
Mean	1.93	9.29	11.02	0.87	1.57	8.80	10.39	-0.04
Standard deviation	0.17	0.62	1.05	0.09	0.12	0.68	1.00	0.07
Minimum	1.64	8.42	9.40	0.74	1.37	7.86	9.29	-0.23
Maximum	2.36	11.21	12.62	1.18	1.87	10.85	12.60	0.05

Table J.2: General statistics on the design events (nearshore conditions) derived using SWAN model for the output location BT157

K. Water Level Time Series

The Fugor OCEANOR database does not contain any data regarding the water levels in the Black Sea. In the metocean study executed by ARCADIS, the design water levels have been determined based on the available data published in the literature [Bondar \(2007\)](#). The water level variation in the Black Sea is explained by several phenomena:

- Sea level rise,
- Storm surge,
- Tidal variation,
- Barometric pressure and resulting seiches,
- Seasonal variation,
- Mean sea level.

Two different data sources are used to generate a 6 hourly water level time series for the same period as the data obtained from the Fugor OCEANOR database. Details on the two data sources are presented in Table K.1.

Name	lat	lon	period	sampling rate
T1	44.147532	28.672009	2015-2020	20 sec - 8 min
T4	44.1875	28.6875	1992-2019	24H

Table K.1: Details on data sources corresponding to water level data.

Data source *T4* contains daily averaged water levels and the sampling period coincides with the Fugor OCEANOR database (1993 - 2012). However, it is daily averaged water levels that will properly represent long term water level variations such as sea level rise and seasonal fluctuations but do not properly represent short term water levels variations such as tides and storm surges. The data source *T1* is obtained with a higher sampling rate capable of capturing short term water level variations.

Generation

The water level variation for the period of interest on a 6-hourly interval is composed according to the following steps:

1. The average daily water levels are retrieved from data source *T4* for the period of interest. Interpolation is performed in order to obtain a 6-hourly time series from daily averaged data points.
2. [Bondar \(2007\)](#) provides information on the tidal component of the water level variation. The tide is described as semi-diurnal with a frequency of 12 hours and 25 minutes. The tidal component is modeled as a harmonic component with an amplitude of 5 centimeters. The phase ρ_1 is determined based on the data source *T1*.

$$\zeta_1(t) = A_1 \cos(\omega_1 t + \rho_1)$$

where

$$\begin{aligned} A_1 &= 0.05 \\ \omega_1 &= \frac{360}{T_1} \frac{\pi}{180}, \quad T_1 = 12 \text{ uur } 25 \text{ min} \\ \rho_1 &= \frac{1}{5}\pi \end{aligned}$$

3. Also, it provides information on seiches which is in fact a long wave caused by sudden changes in the atmospheric pressure. The wave runs back and forth through the Black Sea in 4.5 hours. This component is modeled by a harmonic component with damping and an amplitude of 10 centimeters. The occurrence of a seiche is random and on average once a month.

$$\zeta_2(t) = A_2 \cos(\omega_2 t + \rho_2)$$

where

$$\begin{aligned} A_2 &= \frac{e^{-t}}{10} \\ \omega_2 &= \frac{360}{T_2} \frac{\pi}{180} \quad T_2 = 4.5 \text{ uur} \\ \rho_2 &= 0 \end{aligned}$$

4. The storm surge component is determined by the evaluating the wind induced setup calculation from the Rock Manual:

$$\eta = \frac{1}{2} \frac{\rho_{\text{air}}}{\rho_w} C_D \frac{U_{10}^2}{gh} F$$

where

$$\begin{aligned} \rho_{\text{air}} &= 1.225 \text{ kg/m}^3 \\ \rho_w &= 1018 \text{ kg/m}^3 \\ C_D &= \text{drag coefficient} \\ U_{10} &= \text{normalized wind speed at 10 meter height} \\ h &= 4.8 \text{ m} \\ F &= 500 \text{ km} \end{aligned}$$

The consistency of the direction the wind is coming from influences the wind induced setup. The maximum setup recorded in the Black Sea is 0.80 meter and the water level can rise up to 0.10 - 0.20 meter after sudden directional changes within a period of 1 - 5 hours according to Bondar (2007). Therefore, the storm surge component is generated taking into consideration the wind induced setup, direction, and duration. This is determined as follows:

$$\zeta_3(t) = \sum_{i=0}^{n=3} \eta_i \cos\left(\frac{\theta_{t-i}}{180} \pi - \frac{1}{2\pi}\right) \omega_i$$

where

$$\omega = \{0.4, 0.3, 0.2, 0.1\}$$

The four components described above are combined to obtain the simulated water level time series. Figure K.1 presents a part of the simulated water level time series (gray) together with the average daily water level (blue). It can be observed that the simulated water level time series fluctuates more compared to the blue line which is due to the addition of short term water level variations.

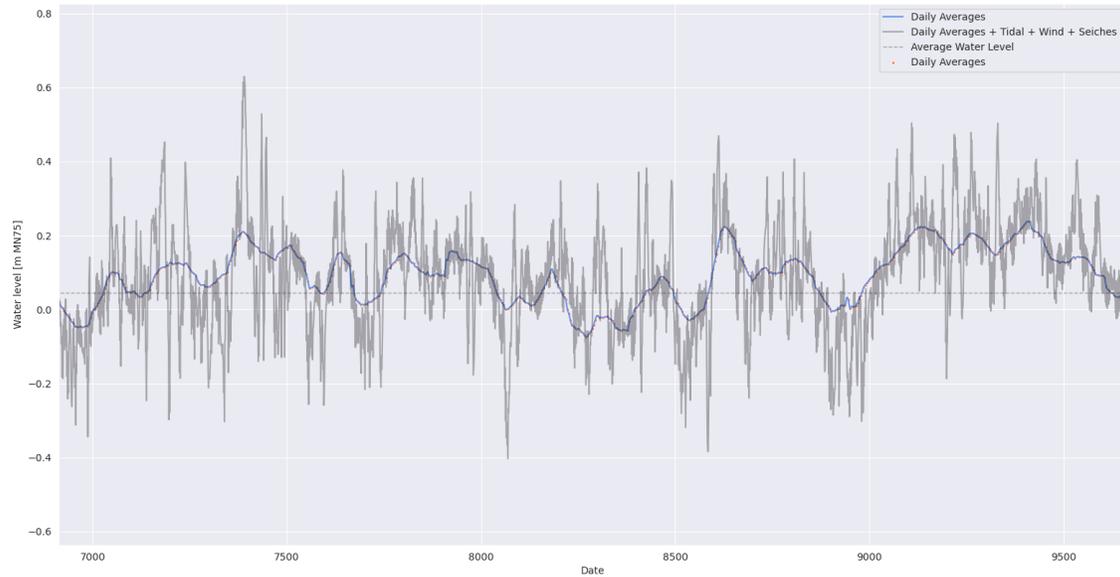


Figure K.1: A part of the resulting simulated water level time series.

Validation

In order to validate the generated water level time series and to get an idea of whether it is a proper representation of the real situation, the generated time series is analyzed by means of a fast Fourier transformation (FFT). The water level variation can be described by a summation of multiple harmonic components. The FFT decomposes the signal and provides the relevant frequencies of the components present in the signal. The resulting frequencies are compared to the FFT applied on data source *T4*, see Figure K.3. It can be concluded that no information is lost by interpolation because the relevant low frequencies are also observed in the simulated time series. The FFT on the daily averages does not give any information on the high frequencies (short term variations) which is due to the Nyquist frequency (sampling rate is every 6 hours).

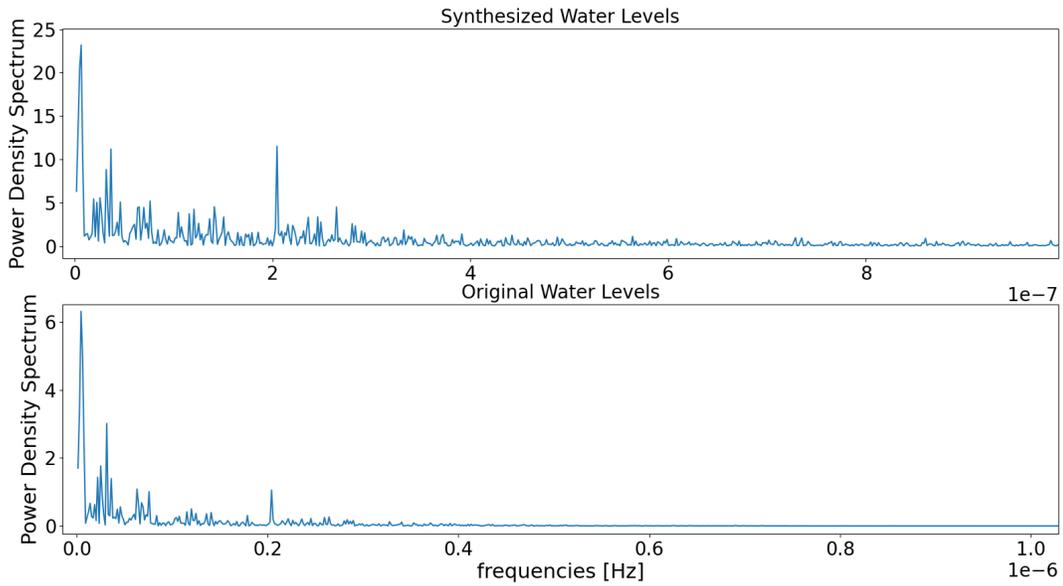


Figure K.2: The results of the fast fourier transformation on the original (upper) and synthesized (lower) water level time series.

Subsequently, a Non-Uniform Fast Fourier Transformation (NUFFT) is applied on the data source *T1* (high sampling rate) to check for the high frequencies. The results show a lot of noise and no relevant frequencies have been selected for the simulated time series. Based on the validation results, the simulated time series is considered to be a proper representation of the real situation and contains all relevant frequencies.

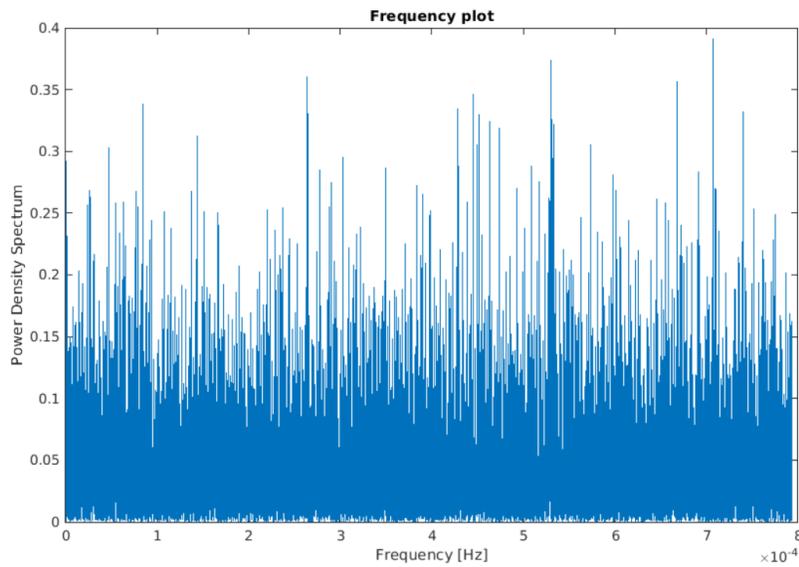


Figure K.3: The results of the non-uniform fast fourier transformation on the data source *T1*.