Delft University of Technology
Hydraulic Engineering

Master Thesis

Extreme value analysis of complex wave systems

Len van der Kooij

January 27, 2020

Graduation committee: — *Organisation:*
Dr. ir. Morales Nápoles O. — TU Delft
Dr. ir. Lanzafame R.C. — TU Delft
Dr. ir. Antonini A. — TU Delft
Dr. ir. Caires S. — Deltares
Ir. van der Lem J.C. — Royal HaskoningDHV

The work of this document is supported by Deltares and Royal HaskoningDHV. Their help is hereby gratefully acknowledged.

Copyright ©2020 department of Hydraulic Engineering

All rights reserved

Abstract

The design of offshore and coastal hydraulic structures is very much dependent on the hydraulic boundary conditions, such as significant wave height, mean wave period and wave direction, among other parameters. A proper design value of these parameters is required during the design process, based on the corresponding safety philosophy and the lifetime of the structure. In this context, an extreme event is characterised by a combination of unfavourable parameters. However, the interdependencies between the parameters are not always accounted for in the design process, despite the fact that some parameters are clearly related. This potentially leads to an overly conservative or optimistic design.

To complicate matters further, a given sea state might consist of a combination of wind wave and swell systems, sometimes coming from different directions and with different spectral shapes. Different combinations of crossing wave systems might lead to the same total significant wave height, mean wave period and mean wave direction. Only analysing the total wave parameters might oversimplify the situation in the presence of combined wave systems. In this thesis a methodology has been developed to establish extreme offshore wave conditions given the presence of these combined wave systems.

A time series that partitions the total wave into a wind wave- and swell component is used as input for the analysis. The location of interest being off the coast of southern Brazil, where combined sea states are observed regularly. The main objective is to compute design values for all wave parameters of interest. With these design values a number of extreme offshore sea states are described in terms of a single total wave system and equivalent combinations of two wave systems. The former resulting in a single-peaked wave spectrum and the latter in an equivalent double-peaked wave spectrum. The extreme offshore sea states are transformed to the nearshore and compared. The single-peaked and equivalent double-peaked wave spectra may result in very similar values for the wave energy nearshore, albeit with different spectral shapes and directions. For the investigated directional combination, this means that the more elaborate approach with two wave systems potentially affects the design of coastal infrastructure if it is sensitive to spectral shape and direction, although the uncertainty of the result is not quantified. Equivalent wave systems could be compared for other directional combinations in future research to investigate if the more elaborate approach results in a more cost-effective design of coastal infrastructure.

The quality of the multivariate vine copula model, used to compute the set of design values for the wave parameters of interest, is assessed in multiple ways. It is recommended not to pick a single set of design values at a point of high joint probability density. Instead it is suggested to use conditionalised samples from the vine copula model to determine the most unfavourable combination of load parameters, which has to be evaluated case-by-case.

Preface

Before you lies a document resulting from a nine month project, in partial fulfilment to obtain the degree of Master of Science in Hydraulic engineering at the Delft University of Technology.

In reference to the contents of this thesis, the topic of multivariate statistics is an exciting one, where multiple interesting discussions are currently taking place within the scientific community. I think that going forward, this topic shall remain very relevant for the design of coastal infrastructure and beyond. I would therefore like to thank Oswaldo Morales Nápoles for introducing me to the the world of statistics. I really appreciate the time you took to help me with my issues and enjoyed our theoretical discussions. I would like to express my gratitude to Robert Lanzafame, Sofia Caires and Cock van der Lem and to Deltares and Royal HaskoningDHV for providing me with comments, tools, advice and useful insights that helped me to complete this project. Many thanks as well to Alessandro Antonini for joining the graduation committee during the last phase of the project. Furthermore, I would like to thank Susana Sellés Valls, Emiel Moerman, Olaf Scholl and Stef Boersen who dedicated a lot of time to help me during the past nine months.

To Gerbant van Vledder, your unwavering knowledge of wave modelling and statistics have been of great help to me. It is truly a shame that you are not able to see the final result, I think you would have liked it. My thoughts go out to your family and friends.

As this project concludes my studies in Delft, I realise how great the last years of my life have been. I am grateful for all the friends I have met during this time. You have always been there for me in good and in bad times.

Special thanks to my family for supporting me throughout the years. Especially my parents, you have always given me every opportunity not only during my time in Delft, but throughout my life in general. This is more than someone could ever ask for.

Len van der Kooij
Rotterdam, 2020

Contents

Glossary

AIC	Akaike Information Criterion
AM	Annual Maxima
BB1	Short form of the Clayton-Gumbel bivariate copula
BB6	Short form of the Joe-Gumbel bivariate copula
BB7	Short form of the Joe-Clayton bivariate copula
BB8	Short form of the Joe-Frank bivariate copula
CDF	Cumulative Distribution Function
CI	Confidence Interval
CSM	Cumulative Steepness Method
CvM	Cramér von Mises statistic
DNV	Det Norske Veritas
DSPR	Directional SPReading
ECMWF	European Centre for Medium-range Weather Forecasting
ERA5	ECMWF ReAnalysis 5th generation data set
EVA	Extreme Value Analysis
GEV	Generalised Extreme Value distribution
GPD	Generalised Pareto Distribution
IID	Independent and Identically Distributed
JONSWAP	JOint North Sea WAve Project
MLE	Maximum Likelihood Estimate
MWD	Mean Wave Direction
NPBN	Non-Parametric Bayesian Network
ORCA	metOcean data tRansformation Classification and Analysis
PCC	Pair Copula Construction (Vine Copula)
PDF	Probability Density Function
POT	Peaks Over Threshold
PWM	Probability Weighted Moments
RP	Return Period
ST1	Survival Tawn type 1 bivariate copula
SWAN	Simulating WAves Nearshore
2D	Two dimensional (wave spectrum)

1 Introduction

1.1 General background

Coastal defence structures like sea walls, breakwaters and dikes are found all over the world. Sea levels are expected to rise up to a meter or even more in the coming century, according to some projections (Church et al., 2013). Coastal defence structures are therefore expected to become more relevant to support human activities in deltas and other low lying areas.

The design of a hydraulic structure is very much dependent on the hydraulic boundary conditions, such as significant wave height, mean wave period (or peak period, depending on the method used) and wave direction, among other parameters. A proper design value of these parameters is required during the design process based on the corresponding safety philosophy and the lifetime of the structure. The choice of a safety philosophy depends on the importance of the structure and the consequences of failure of the structure. If the consequences of failure of a certain structure are limited, a higher probability of failure of the structure might be acceptable in the design. Additionally, the structure might have a service lifetime of 50 years, 100 years or even longer, depending on the functional requirements of the structure. Data sets of that length are often not available, so statistical methods are used to obtain the design values.

The appropriate safety philosophy and required lifetime of the structure are then determined. The next step is to determine the extreme conditions that the structure might face during its lifetime. If for instance a port is constructed, a breakwater might be part of the design. The primary function of such a breakwater is to protect vessels that are moored inside the port from waves coming from offshore. The parameters that determine the rock size of the armour layer in a breakwater include, but are not limited to, the incoming significant wave height, mean wave period, rock density and the slope of the structure (Schiereck, 2001). In a stability calculation the significant wave height and wave period are load parameters and the other parameters are strength parameters. Besides the wave height and wave period, the direction of the waves is also relevant, as oblique waves require a smaller armour layer for stability than waves directed perpendicularly to the breakwater. Because of these complexities, an extreme event is characterised by a combination of unfavourable significant wave height, wave period and wave direction. The interdependency of the parameters should also be accounted for in the design process, as some parameters are clearly related. A clear relation between significant wave height and wave period of extreme wind wave events has been observed, meaning that wind waves with a very high significant wave height also tend to have longer periods. Assuming that significant wave height and wave period are independent for wind waves leads to a design that might be too conservative.

Furthermore, both wind waves and swell from different directions may be present in extreme situations, leading to a combined sea state (sometimes called cross sea in literature (Toffoli et al., 2005)). This might be more relevant depending on the location of interest. Some offshore locations may even experience extreme events that consist of multiple *wave systems* at once: a wind wave component and multiple swell components, for example. These combinations complicate the statistical analyses required to obtain the design values. Different combinations of individual wave system significant wave height, wave period and wave direction may correspond to the same total significant wave height, wave period and wave direction. This leads to questions on how to deal appropriately with these combined wave systems.

1.2 Problem statement and objectives

As stated previously, the engineering practice requires a value of the incoming significant wave height, wave period and wave direction with a return period of a certain number of years. This return period is dependent on the intended lifetime and safety policy of the structure. Univariate extreme value analysis is often performed to obtain the design values of these parameters for the return periods of interest separately for each variable. For two variables, two univariate distributions are often combined to determine the bivariate distribution, assuming either independence or full dependence. A possible reason for engineers to assume independence or full dependence might be that it saves a lot of time in calculation, considering that a joint probability calculation is not a straightforward task.

However, if two variables are considered, the assumption of independence or full dependence might not be valid in some cases where the reality is somewhere in between. In that case, assuming independence might lead to a design that is too conservative, while assuming full dependence could lead to a design that is too optimistic. A more complex bivariate analysis is thus required to account for the appropriate dependence between the two variables. An example of two related variables is the total significant wave height and wave period of oceanic waves, for which a bivariate analysis determines the design values of extreme events.

Moreover, a further distinction can be made between wind waves and swell. A clear relation between the wave height and wave period has been established for wind waves, but less so for swell (Holthuijsen, 2007). A given sea state might consist of a combination of wind wave and swell systems, sometimes coming from different directions and with different spectral shapes. This complicates matters further as different combinations of crossing wave systems might lead to the same total significant wave height, wave period and wave direction. Only analysing the total wave parameters might oversimplify the situation in the presence of combined wave systems. At the moment it is unclear how to deal appropriately with these combined wave systems for the offshore situation. Another problem is that nearshore data coming from reanalysis data sets is less reliable because of shallow water effects and local bathymetry. Only offshore data in deeper water are considered reliable enough for the analyses. Thus, offshore wave conditions have to be transformed to the nearshore situation to analyse how an extreme event consisting of multiple sea states affects nearshore wave conditions.

1.2.1 Primary objective

The general background and problem statement have led to the following primary research objective:

To develop a methodology to establish extreme offshore wave conditions given the presence of combined wave systems.

To achieve this objective, a number of sub questions are formulated:

1. How is a combined sea state defined?
2. How are extreme offshore wave events defined in a statistical sense?
3. What is the dependence between the parameters of an extreme event?
4. What are the extreme wave conditions at the offshore location for large return periods?

A more concrete elaboration of the primary research objective is provided in Chapter 2.

1.2.2 Secondary objective

Besides the primary research objective, a secondary objective is formulated:

- To investigate nearshore wave conditions given the presence of an extreme combined sea state event offshore.

1.3 Outline

The general outline of this thesis is described:

- In **Chapter 2** the methodology to establish extreme wave conditions given the presence of combined wave systems is proposed. The methodology consists of a number of work steps, where each step of the proposed method is explained. Each work step has its own chapter, explaining the specifics and motivation of each action in detail.
- **Chapter 3** consists of a literature study. The reader is provided with theoretical concepts necessary fulfil the primary and secondary research objectives. The first part of the literature study describes concepts regarding waves and wave spectra. Next, univariate extreme value theory is described, after which bivariate analysis using copulas is described. The literature study concludes with theory on multivariate structures based on bivariate copulas and some information about reanalysis data sets.
- In **Chapter 4** the data collection process is described. The data are collected from a reanalysis data set and the variables of interest are described. The geographical location of the data set is described and analysed.
- **Chapter 5** explains the univariate extreme value analysis and describes which variable is recognised as extreme and which variables are considered concomitant of the extreme event. This results in a data set of extreme events, out of which a number of additional parameters are derived.
- In **Chapter 6** a bivariate analysis is performed. The chapter begins with bivariate analysis of different parameters using an adapted copula based approach. Additionally a copula based approach is compared to a non-parametric bivariate method.
- In **Chapter 7** a multivariate analysis is performed. The problem is simplified to limit the number of variables after which a multivariate vine copula is fitted and the design values for the parameters of interest are calculated.
- **Chapter 8** focusses on the secondary objective about offshore to nearshore transformation. The design realisations resulting from the bivariate and multivariate analyses are used to generate a number of 2D wave spectra. These spectra are run through a model in order to determine nearshore conditions.
- In **Chapter 9** the most important conclusions of this thesis are summarised, followed by a discussion. The main part of the report ends with a number of recommendations for further research in **Chapter 10**.

2 Methodology

With the primary objective stated in Chapter 1, a methodology to establish extreme wave conditions given the presence of combined wave systems is proposed. The goal of this chapter is to provide a more concrete definition of what this goals entails and how it must be achieved. A background to the proposed methodology is provided in Chapter 3.

In order to answer the primary research question wave parameters from a reanalysis time series are used. A reanalysis time series is produced by combining historical observations and model data using an assimilation algorithm into one data set. This results in a time series of hourly values of a large number of metocean parameters with a length of multiple decades. The idea is to compare wave parameters of the **total sea state** to parameters belonging to so-called **partitions**. The reanalysis data set used in this thesis contains parameters of the total sea state, wind wave- and swell partitions. Both the total wave and partitioned parameters are derived from the same 2D wave spectrum, while the total amount of energy in each 2D wave spectrum is conserved by the partitioning algorithm. This means that adding up the energy of all partitions always results in the energy of the total sea state. Using a number of statistical methods the return values of these spectral parameters are determined for a number of return periods, all conditional on a return value of a single dominant variable. The total significant wave height is chosen as the dominant variable in this research.

The result of the statistical analyses is a table with values for all the relevant wave parameters of the total sea state and partitions, all values being conditional on a return value of the dominant variable. With these parameter values it is then possible to generate 2D wave spectra of extreme sea states, with corresponding return periods much larger than the length of the original time series, thus completing the primary research objective.

With the return values of the total sea state parameters a single-peaked 2D wave spectrum is generated, while the return values belonging to the wind wave and swell partitions allow for a double-peaked wave spectrum to be generated that is equivalent to the single-peaked wave spectrum. The generated single-peaked and double-peaked wave spectra can be compared in a number of different applications. One of these applications is to transform the offshore wave conditions to nearshore, which is the secondary objective of this thesis.

The methodology consists of a number of work steps. Each step of the proposed method is explained in detail. The first steps marked in blue in Figure 2.1 describe the methodology to achieve the primary objective. The work step of the secondary objective is marked in green in Figure 2.1 and is described after the work steps of the primary objective.

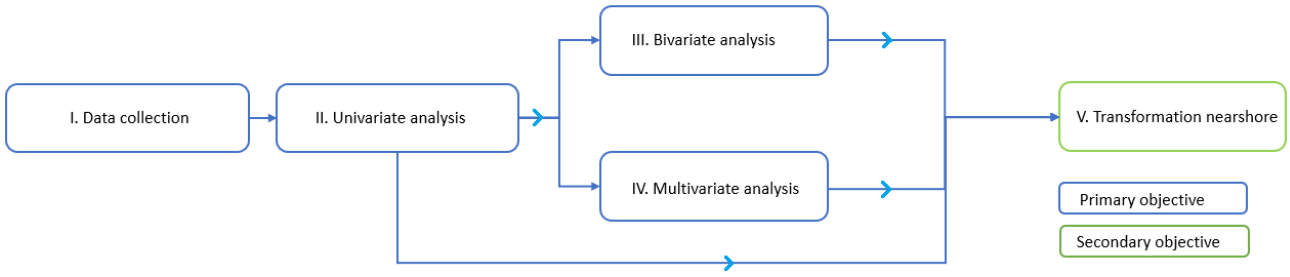


Figure 2.1: Work plan consisting of a number of work steps.

2.1 Work steps

Each consecutive work step of the methodology is described. The required input of the work step is mentioned, followed by the actions performed in the work step. The output of each work step is described as well.

I **Data collection**
The first step is to collect the necessary data for the analysis. The output of this step is a data set with the

hourly values of the variables of interest at the location of interest. The length of the time series may vary depending on the reanalysis data set. In the case of this analysis the data set consists of hourly values from 1979 to 2018. The first step consists of the following actions:

- Selecting location of interest.
- Determining variables of interest.
- Selecting a data set.
- Plot and inspect data.

The required data are collected at a location of interest. The location that is chosen in this research is an offshore location near the southern coast of Brazil. There two main reasons for choosing this location, the first reason has to do with the wave climate in the area. Because of its geographic location, the location of interest is subjected to high wind sea mainly from the north, as well as high swell events originating in the turbulent region south of the location of interest called the *roaring forties*, named after the 40th southern latitude known for its frequent storms. Combined sea states are thus observed regularly in this general location, which makes the location a suitable case to answer the primary research question. The second reason for choosing the location of interest is that a bivariate analysis studying the dependence between wave parameters has been performed in close proximity to the location of interest by Sellés Valls (2019). The results of this thesis could be compared to the other study at a later point.

Next, the variables of interest are determined. In order to achieve the primary research objective, the significant wave height, wave period and wave direction are required of the total wave, wind wave- and swell partition. A more elaborate description of the location and parameters of interest is provided in Chapter 4.

There are multiple options to obtain the required wave data at the location of interest. One could install measuring devices on a buoy that measure the wave parameters of interest. Another option to obtain the data is to use a so-called reanalysis data set, which combines vast amounts of historical observations into a global model using data assimilation and advanced modelling (ECMWF, 2018). An advantage of using a reanalysis data set compared to wave buoy data, is that a large number of wave parameters can be obtained on a global grid with a time scale of multiple decades. Furthermore, the resolution of the grid and the length of the time series of these reanalysis data sets are both expected to increase. This means that a longer time series can be obtained on a finer grid in the future. Using a reanalysis data set is therefore considered a more powerful and flexible method to obtain the required data compared to using a wave buoy.

After all the hourly values of the parameters of interest are obtained at the location of interest, the data set is plotted and inspected to see if it is complete. Given that the primary research objective is about the presence of combined wave systems, the occurrence of combined wave systems is checked in the data set.

II **Univariate analysis**

Univariate Extreme Value Analysis (EVA) is the second step in the methodology of this thesis. This step takes the results of step I (data collection) as input, a data set for a (set of) location(s) of hourly values of offshore wave variables. The length of the data set may vary, data sets of multiple decades are common in practice. The marginal distribution for each variable given a certain combination of wind wave and swell direction is determined, all based on a dominant variable. Step II consists of the following actions:

- Selecting the wave direction sectors of interest.
- Selecting a dominant variable and performing EVA on it.
- Fitting distributions to the concomitant variables.
- Deriving additional parameters out of the resulting extreme data set.

The fact that mean wave direction is a circular quantity poses difficulties when performing a multivariate extreme value analysis. Due to the scope of this thesis, a pragmatic choice is made with regard to the mean wave direction. The data set is conditionalised based on a combination of wind wave and swell direction.

This means that only data points with a concomitant wind wave direction between a certain interval and a swell direction between another interval are considered in the analysis. The decision of the specific directional combinations is made by the engineer performing the analysis. Different directional combinations might be of interest from the engineering point of view, which could be analysed at a later point.

Next, the dominant variable (sometimes called extreme variable) of the data set is defined. The choice of a dominant variable depends on the design of the structure, as some structures might be more sensitive to incoming wave height while others might be more sensitive to mean wave period. The choice of a dominant variable may therefore differ between studies.

Because of the relatively short length of the available time series a peak-over-threshold approach is preferred as the method of extreme value analysis, this is explained in the literature study. A peak-over-threshold analysis is performed on the dominant variable. If the dominant variable exceeds a certain threshold, the event is considered extreme. The values of other variables during the extreme event are associated- or concomitant values. Concomitant variables do not necessarily have to be extreme values themselves, but are the other values corresponding to an extreme event. The variables of swell and wind waves are considered in this step, given an extreme event. This results in a fitted Generalised Pareto Distribution (GPD) for the dominant variable and different marginal distributions for the concomitant variables. Next, a number of interesting (from the engineering perspective) return periods for the dominant variable are chosen. The fitted GPD of the dominant variable is extrapolated to these return periods of interest, as will be explained in Chapter 5.

Additionally, the directional spreading and peak enhancement factor needed to generate 2D wave spectra are derived out of the extreme data set or assumed for the total wave, wind wave- and swell partitions.

III **Bivariate analysis**

The output of step II is a data set of a combination of directions of interest consisting of extreme events based on a dominant variable together with the fitted marginal distributions. The output of step II is used as input for this step. The goal of the bivariate analysis in this thesis is to obtain associated values of the total mean wave period conditional on univariate return values of the total significant wave height for a number of return periods. By separating the dependence structure from the marginal information the former can be modelled separately. Such bivariate dependence structures are called copulas and are used in this work step. A copula-based approach is applied together with truncation to come to a result for the total mean wave period. The copula-based approach consists of the following actions:

- A number of copulas are fitted to the data.
- The model adequacy of each copula is determined and a choice for a copula is made.
- A very large sample is obtained from the copula, converted to physical units and truncated to account for the limiting steepness.
- The associated values for the mean wave period are determined conditional on the univariate return values of the total significant wave height.

It is noted that the choice for a copula-based approach for the bivariate analysis is not an essential part of the overall methodology. Copulas are preferred because of their flexibility in use and are simply a means to obtain the associated values for the total mean wave period. Besides this point, a copula-based approach is also used later on in the multivariate analysis. To someone who is not familiar with copulas it might be beneficial to first understand how copulas work for two variables before advancing to higher dimensions.

IV **Multivariate analysis**

The output of step II is used as input for this step. A multivariate analysis is performed on the extreme data set using a pair copula construction (also known as a vine copula in literature). The multivariate analysis results in design values for the parameters of interest, that are used to build a double-peaked 2D wave spectra for a number of return periods. The multivariate analysis consists of a number of actions:

- A number of simplifications are made in order to reduce the number of variables in the multivariate analysis as much as possible.
- A vine copula model is fitted to model the dependence structure of the extreme data set. The choice of this model is based on a number of model adequacy tests.
- The design values are calculated using the model. The joint density is used to determine the design values for the parameters of interest.
- The robustness of the model is tested in a number of ways in order to investigate the quality of the outcome.

V **Secondary objective: Determine how this might affect the nearshore situation**
Once the variance density spectra are known offshore, a transformation to the nearshore is performed by using a SWAN model. For engineers constructing coastal infrastructure, it would be useful to investigate if the presence of a combined sea state offshore affects nearshore wave conditions as well. The propagation of wave energy to the nearshore is investigated, thus wind is not included in the analysis.

2.2 Methodology in the design phase

The design of a breakwater is used as an example in this section in order to put the developed methodology of this thesis in the appropriate context. For the construction and use of infrastructure in general, a number of phases are defined by Hertogh and Bosch-Rekveldt (2015):

- Initial phase
- Design phase
- Realisation phase
- Use phase
- Recycling or demolition phase

The steps described in this thesis are all in the design phase. Besides the steps described in the Section 2.1, there are other steps required during the design phase before a final design is achieved. A simplified flowchart of the design phase is used as an example in Figure 2.2.

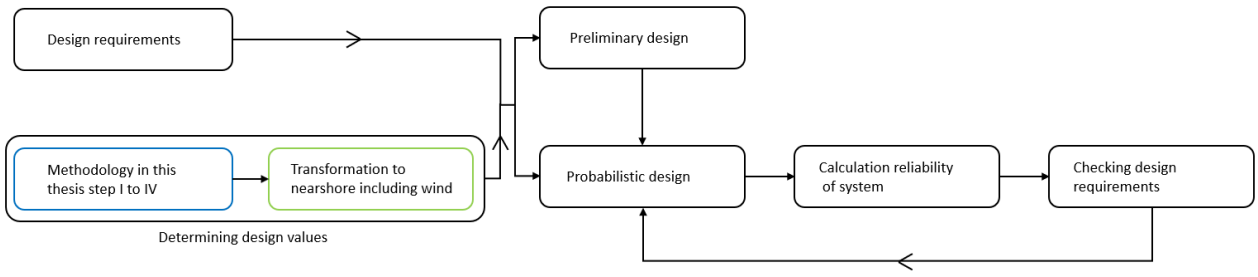


Figure 2.2: Simplified flowchart of the design phase. Adapted from Everts (2016).

It is noted that Step V of the methodology of this thesis does not lead to the design values, as the transformation to the nearshore is performed without the inclusion of wind. If wind is included in the transformation to the nearshore, the outcomes may be used in the design. The other steps in the design phase, shown in black in Figure 2.2, are not performed in this thesis.

3 Literature study

This chapter gives an overview of several theoretical concepts, that are used in this research. As such, this chapter functions as a background to Chapter 2. The concepts are applied to the problem to help answer the research questions. The Literature study begins by describing different concepts regarding waves and wave spectra in 3.1. Next, extreme value theory is described in 3.2. Concepts related to bivariate analysis are described in 3.3 to 3.5, followed by theory related to multivariate analysis in 3.6 and 3.7.

3.1 Wave signal and spectrum

The concepts in this section are largely based on the book by Holthuijsen (2007). Before a wave period is defined, a precise definition of what actually accounts as a wave is required, because not every surface elevation change accounts as a wave. If the surface elevation is recorded over time ($\eta(t)$), the mean of the surface elevation is considered the reference point for all waves. The surface elevation $\eta(t)$ is considered a Gaussian process. The wave height (H) is defined as the difference between the maximum and minimum elevation of a wave and is approximately Rayleigh distributed in deep water. The wave period (T) is defined as the time between one zero crossing and the next. A random surface elevation record with the definition of wave height and wave period is shown in Figure 3.1. The time record with a total duration D is used to explain a number of concepts in the next sections.

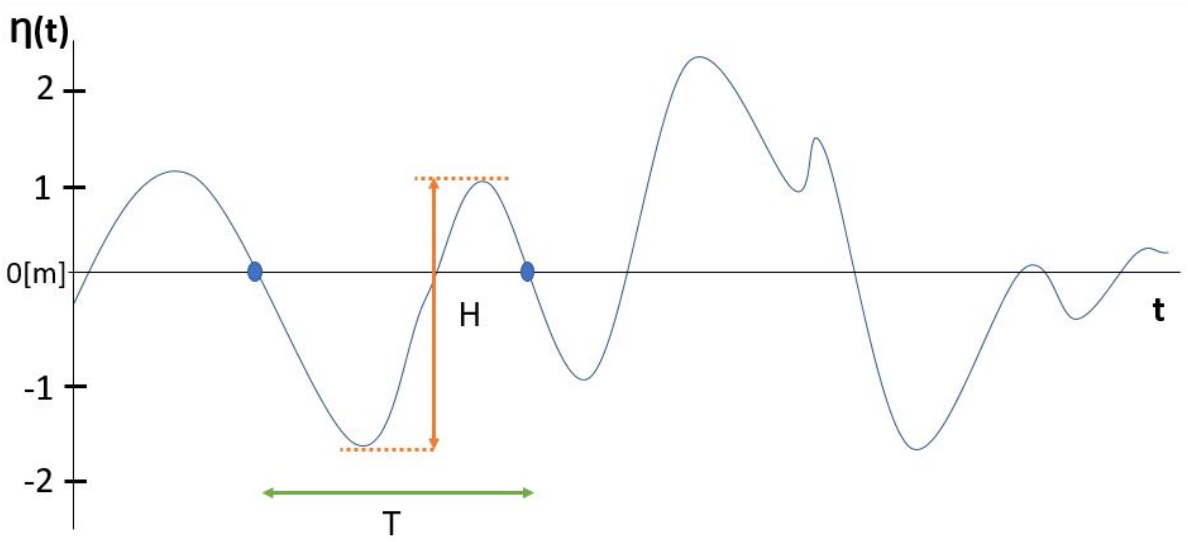


Figure 3.1: Definition of the wave height H and wave period T shown in a surface elevation time record, the record has a total duration D.

3.1.1 1D variance density spectrum

A variance density spectrum is widely used in various engineering practices to describe the characteristics of the ocean surface. The aim of describing waves as a variance density spectrum is not to describe any single detailed events, but rather the ocean surface as a stochastic process. A random phase/ amplitude model is introduced in order to convert a wave signal into a variance density spectrum. The transformation of a wave signal, as is shown in Figure 3.1, into a variance density spectrum consists of a number of steps.

1. The time record is transformed into the sum of harmonic components using a Fourier transform

$$\eta(t) = \sum_{i=1}^{N} a_i \cdot cos\left(2\pi f_i t + \alpha_i\right) \qquad \text{(Eq. 3.1)}$$

In which:

$\eta(t)$	[m]	surface elevation as a function of time t
N	[-]	number of harmonic components
a_i	[m]	amplitude of the components
f_i	[1/s]	$= i/D$ frequencies of the components
α_i	[rad]	phase of the components

The most important result of this first step is the set of amplitudes and frequencies of the harmonic components.

2. The amplitude of each component is converted to an expected value of the variance, a more relevant parameter statistically speaking:

$$\text{expected value of the variance} = E\{\frac{1}{2}\underline{\bar{a}}_i^2\} \quad \text{(Eq. 3.2)}$$

3. The discrete variance spectrum is converted into a continuous variance density spectrum, because at sea all frequencies are observed, not just at discrete frequencies of the previous steps. A continuous variance density is obtained by letting the width of the frequency bands approach zero.

$$E(f) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} E\{\frac{1}{2}\underline{a}^2\} \quad \text{(Eq. 3.3)}$$

The an example of a typical variance density spectrum is shown in Figure 3.2.

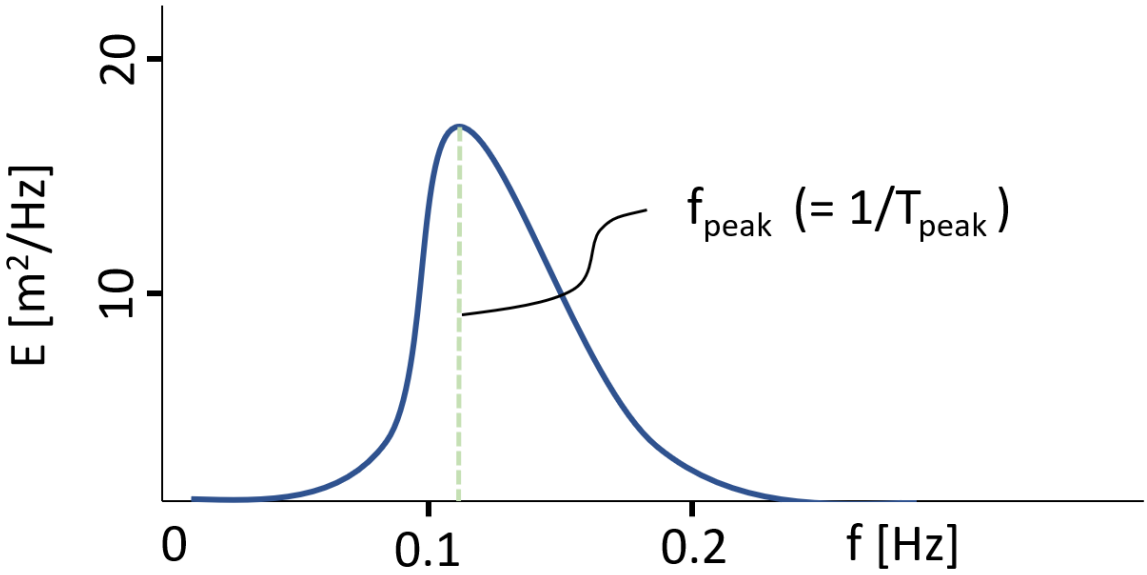


Figure 3.2: Example of a typical single-peaked 1D wave spectrum. The peak frequency (easily transformed into peak period) is where most of the spectral energy is concentrated.

A single-peaked spectrum can sometimes be represented by an analytical expression. For a young sea state in deep water and with a limited fetch, engineers often use a JONSWAP spectrum for their wave analyses, JONSWAP standing for JOint North Sea WAve Project (Hasselmann et al., 1973). A Pierson-Moskowitz spectrum is used for fully developed sea states with long fetches and has a lower peak frequency compared to a young sea state (Pierson and Moskowitz, 1964). The JONSWAP variance density as a function of the frequency $E(f)$ is defined in Eq. 3.4.

$$E(f) = \frac{\alpha g^2}{16\pi^4} f^{-5} \exp\left[-\frac{5}{4}\left(\frac{f}{f_\mathrm{m}}\right)^{-4}\right] \gamma^b$$
$$b = \exp\left[-\frac{1}{2\sigma^2}\left(\frac{f}{f_\mathrm{m}} - 1\right)^2\right] \qquad \text{(Eq. 3.4)}$$
$$\sigma = \begin{cases} 0.07 & \text{for } f \leq f_\mathrm{m} \\ 0.09 & \text{for } f > f_\mathrm{m} \end{cases}$$

In which:

α	[-]	energy scale parameter equal to 0.0081
f_m	$[Hz]$	peak frequency (sometimes noted as f_p)
γ	[-]	peak enhancement factor, usually $1 \leq \gamma \leq 10$
g	$[m/s^2]$	gravitational constant

For $\gamma = 1$ Eq. 3.4 reduces to a Pierson-Moskowitz spectrum. Expressions of γ are described in terms of H_s and $T_{m0,2}$ by Isherwood (1987). In terms of interpretation, an important definition is that the total variance is by definition the integrated variance density spectrum.

$$\text{total variance} = \bar{\eta^2} = \int_0^\infty E(f)df \qquad \text{(Eq. 3.5)}$$

The total variance is proportional to the total energy with a factor ρg, in which ρ is the density of the water and g is the gravitational acceleration. Another important variable is the *nth-order moment* of the variance density spectrum, used for certain wave characteristics and defined as:

$$m_n = \int_0^\infty f^n E(f)df \qquad \text{for} \quad n = ..., -2, -1, 0, 1, 2, ... \qquad \text{(Eq. 3.6)}$$

Note that Eq. 3.6 reduces to Eq. 3.5 for $n = 0$, thus the zeroth-order moment is equal to the total variance.

3.1.2 Wave period and wave height

The T_m is defined as the mean period of a number of waves. T_m can be calculated from a time series analysis directly or from different order spectral moments. Three types of mean periods are defined based on the spectral moments:

$$T_{m-1,0} = m_{-1}/m_0 \qquad \text{(Eq. 3.7)}$$

$$T_{m0,1} = m_0/m_1 \qquad \text{(Eq. 3.8)}$$

$$T_{m0,2} = \sqrt{m_0/m_2} \qquad \text{(Eq. 3.9)}$$

The mean wave period based on the moment of order -1 ($T_{m-1,0}$), is known as the energy period. The energy period can be used together with the significant wave height H_{m0} to determine the the wave energy flux per wave crest length (hence the name). $T_{m0,1}$ is the mean wave period based on the first order spectral moment. $T_{m0,1}$ is the reciprocal of the mean frequency and is useful when constructing wave spectra if the peak period is not available. $T_{m0,2}$ is known as the zero-crossing mean wave period, because it corresponds to the mean period acquired using the zero-crossing method. A downside of using $T_{m0,2}$ is that it is sensitive to noise in the high frequency range, because it uses the second order spectral moment. The peak period is defined as the wave period where the maximum wave energy is concentrated in a wave spectrum. The peak period is empirically linked to the mean wave period, depending on the spectral shape (Verhagen et al., 2009).

$$T_m \approx (0.75 \quad \text{to} \quad 0.85) \cdot T_{peak} \qquad \text{(Eq. 3.10)}$$

The significant wave height is defined as the average of the highest one third of waves in a certain wave record. Given that the wave height is Rayleigh distributed, the significant wave height can also be expressed in terms of the zeroth-order moment (m_0) of the variance density spectrum:

$$H_{m0} \approx 4 \cdot \sqrt{m_0} \qquad \text{in deep water} \tag{Eq. 3.11}$$

The mean wave period and significant wave height that are used for this research are both based on the spectral moments of the spectrum. Different notations of the significant wave height are found in literature. The notations H_{m0}, $H_{1/3}$, H_s or H_{visual} all represent the significant wave height and are effectively equal for deep water conditions, but derived in different ways. It is important to keep in mind that there are other definitions of these parameters as well, provided in Holthuijsen (2007).

3.1.3 2D frequency-direction spectrum

The variance density spectrum, often called a 1D spectrum, gives information of the frequencies present in a given sea state. However, the 1D variance density spectrum gives no information about the direction of the different components. The 1D spectrum is expanded to a 2D spectrum by adding the directional information, resulting in a frequency-direction spectrum. The surface elevation is again described by a number of harmonic components, similar to Eq. 3.1, but this time the surface elevation is also a function of the x,y-space as a sum of propagating harmonic components:

$$\eta(x,y,t) = \sum_{i=1}^{N}\sum_{j=1}^{M} \underline{a}_{i,j} cos(\omega_i t - k_i x \cdot cos(\theta_j) - k_i y \cdot sin(\theta_j) + \alpha_{i,j}) \tag{Eq. 3.12}$$

In which:

ω_i	[rad/s]	radian frequency ($= 2\pi/T$)
k_i	[1/m]	wave number ($= 2\pi/L$)
θ_j	[rad]	direction of wave propagation
a_i	[m]	amplitude of the components
α_i	[rad]	phase of the components
N, M	[-]	number of components in x,y-direction

It is noted that radian frequency and wave number are related in the dispersion relation, thus reducing the number of indices in Eq. 3.12. The dispersion relation for linear wave theory is defined as with water depth d as:

$$\omega^2 = gk \cdot tanh(kd) \tag{Eq. 3.13}$$

In a similar procedure as with the 1D spectrum is performed to obtain a continuous 2D spectrum:

$$E(f,\theta) = \lim_{\Delta f \to 0} \lim_{\Delta \theta \to 0} \frac{1}{\Delta f \Delta \theta} E\{\frac{1}{2}\underline{a}^2\} \tag{Eq. 3.14}$$

The 1D variance density spectrum is easily obtained from the 2D spectrum by integrating all the directions:

$$E(f) = \int_0^{2\pi} E(f,\theta) d\theta \tag{Eq. 3.15}$$

Again the total variance is the integral of the spectrum, but this time also integrated over the directions:

$$\bar{\eta^2} = \int_0^{\infty}\int_0^{2\pi} E(f,\theta) d\theta df \tag{Eq. 3.16}$$

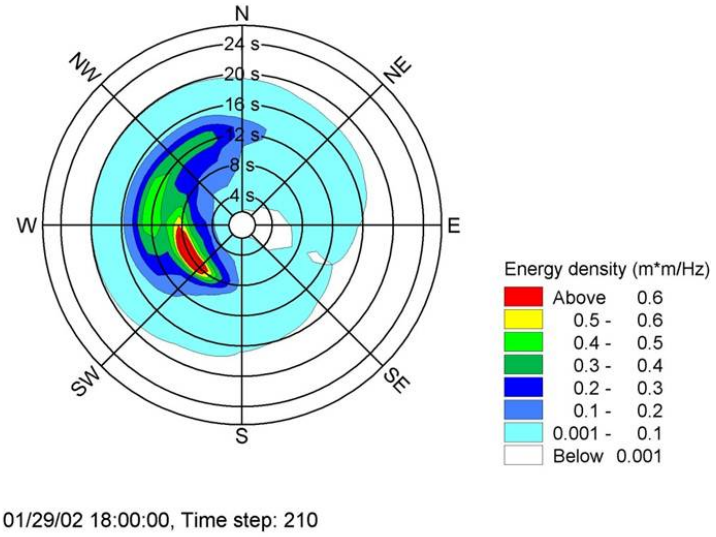


Figure 3.3: An example of a (radial) 2D wave spectrum at an offshore location. The energy density (proportional to the variance density) is represented by a colour scale. The figure shows that wave energy originates from different directions, with the peak direction being SSW and peak period of about eight seconds. Source: `http://dhiuk-demos.blogspot.com/2012/07/`

The directional spreading per frequency $\sigma(f)$ in radians is defined as:

$$\sigma(f) = \sqrt{2(1-\sqrt{a^2+b^2})}$$
$$a(f) = \frac{\int_0^{2\pi} E(f,\theta)\cos\theta d\theta}{E(f)} \qquad \text{(Eq. 3.17)}$$
$$b(f) = \frac{\int_0^{2\pi} E(f,\theta)\sin\theta d\theta}{E(f)}$$

The directional spreading in a 2D spectrum is frequency dependent, with a different value of the directional spreading is determined for each frequency. A median value of the directional spreading is often used to represent the sea state. This results in only one value for σ_θ per sea state, often converted to degrees. The maximum value of the directional spreading is $\sqrt{2}$ rad, or about 81°. A method to obtain the median value of the directional spreading of a sea state is described in Kuik et al. (1988). Unlike the 1D spectrum of Eq. 3.4, a 2D (JONSWAP) spectrum requires five parameters as input:

Variable	Unit	Description
H_{m0}	m	significant wave height
θ	°	mean wave direction
T_p	s	peak wave period
σ_θ	°	directional spreading
γ	-	wave spectral peakedness

Table 3.1: Variables needed to generate a 2D JONSWAP spectrum.

A SWAN model (Booij et al., 1999) also allows for $T_{m0,1}$ to be used as input for the wave period instead of T_p. A 2D JONSWAP shape often assumed for wind waves. Other shapes are sometimes used for swell, such as a Gaussian model. A recent study by Lucas and Guedes Soares (2015) shows that a JONSWAP might be a better model for swell as well.

3.2 Univariate extreme value analysis

For many engineering problems, the statistical distribution of the extreme value of a certain parameter is relevant. If a data set is limited in time, but extreme values with higher return periods are required, Extreme Value

Analysis (EVA from now on) is often used to calculate return periods of the variable of interest. Assume that a 50 year long data set is available for a certain location. However, the required return periods could be in the order of 500 or even 1000 years for the case of a breakwater, as demonstrated in Appendix A.1. EVA is used to estimate the return values for these longer return periods. In order to evoke extreme value theory, a number of fundamental conditions are to be fulfilled. The first one is that the values must be statistically **independent** from each other. The second condition is that the values are **identically distributed**. These two assumptions are referred to together as the **i.i.d. assumption**. The first assumption might pose some problems if hourly values for significant wave heights are analysed, because extreme wave heights are not independent. An hourly value for a large significant wave height is likely to be preceded by another. To overcome this, only extremes that are spaced out long enough in time are to be considered in EVA, for extreme wave events this minimum spacing is typically in the order of a few days (Holthuijsen, 2007).

Two types of EVA are described next, the Peak-Over-Threshold (POT) method and the Annual Maxima (AM) method. The basic steps for both methods are the same:

1. Define the extreme values in the data set (define the size of the blocks for AM or the limiting threshold for POT).
2. Fit a trend line distribution to these extreme values.
3. Define the extreme value for a certain return periods based on the (extrapolated) fitted trend line.

3.2.1 Peak-over-threshold approach

$X_1, X_2, ..., X_n$ is defined as a sequence of independent random variables with a common unknown distribution function, u is defined as a threshold and F_u is defined as a distribution of the excess values above the threshold. A POT approach is applied on swell wave height in Figure 3.4.

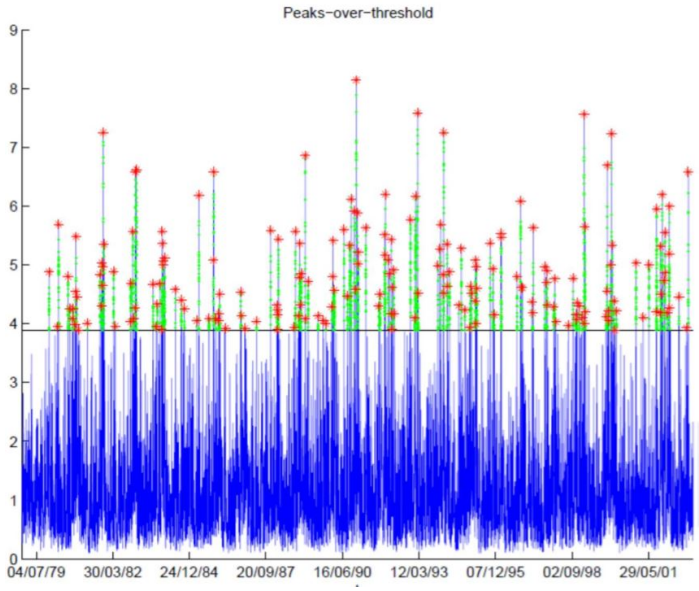


Figure 3.4: POT approach: a reasonable threshold u is set such that F_u is a distribution of excess values. From Caires and van Os (2012).

The Pickands-Balkema-de Haan theorem states that for a large class of underlying distribution functions and a large threshold u, F_u is approximated by the Generalised Pareto Distribution (GPD) (Pickands, 1975). The GPD distribution is given by:

$$F_u(y) = 1 - \left(1 + \xi \frac{y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \qquad \text{where} \quad 0 < y < \infty, \quad \tilde{\sigma} > 0 \quad \text{and} - \infty < \xi < \infty \qquad \text{(Eq. 3.18)}$$

In which:

$\tilde{\sigma}$ scale parameter
ξ shape parameter
For $\xi = 0$ the function becomes the exponential function with a mean $\tilde{\sigma}$ (Type I)
For $\xi > 0$ the function becomes the Pareto distribution (Type II)
For $\xi < 0$ the function becomes the bèta distribution (Type III)

The choice of a threshold is a trade off between bias and variance. A too high threshold might generate too little data excesses, leading to a high variance. A threshold that is too low might violate the asymptotic behaviour of the model, which results in bias. The threshold stability property is obtained if for u_0 the GPD is a reasonable model and for a higher threshold u the GPD is a reasonable model as well. The GPDs for both thresholds should have identical shape parameters and the scale parameters should be related by (Caires and van Os, 2012):

$$\tilde{\sigma_u} = \tilde{\sigma_{u_0}} + \xi(u - u_0) \qquad \text{(Eq. 3.19)}$$

The advantage of using the POT approach is that all large values are retained if the threshold is set correctly. As discussed, a disadvantage of the POT approach is that it might be difficult to set this threshold correctly as well as setting a correct minimum spacing between peaks to guarantee independence. The GPD can be fitted with different methods, such as the Maximum Likelihood Estimate (MLE) or by Probability Weighted Moments (PWM).

3.2.2 Annual maxima approach

Besides the POT approach, the AM approach is frequently used in extreme value analysis. Again, the same data set of $X_1, X_2, ..., X_n$ is defined, but instead of a threshold, the maximum value in a certain block is added to the extreme data set. To avoid any effects of seasonality, and thus violating the i.i.d. assumption, the size of the blocks is often chosen to be one year. The AM approach is applied on the same wave data as in the previous section, as shown in Figure 3.5.

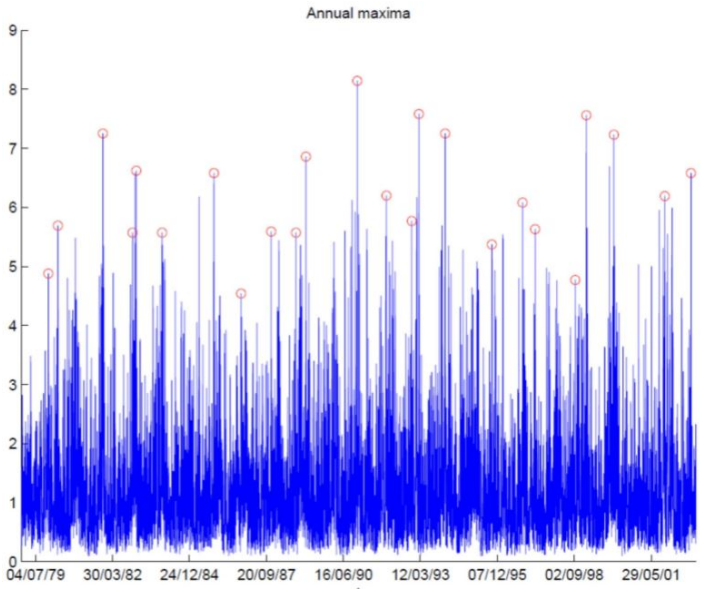


Figure 3.5: The extreme value data set (red dots) consists of the highest yearly value when using the AM approach. From Caires and van Os (2012).

According to the Fisher-Tippett theorem the extreme data set of block maxima $M_n = max\{X_1, ..., X_n\}$ converges to Generalised Extreme Value (GEV) distribution for $n \to \infty$ (Coles, 2001). The non-degenerate cumulative GEV is defined as:

$$F(s, \xi) = \begin{cases} \exp(-1(1 + \xi s)^{-1/\xi}) & \xi \neq 0 \\ \exp(-\exp(-s)) & \xi = 0 \end{cases} \qquad \text{(Eq. 3.20)}$$

In which:

- s normalised variable $s = (x - \mu)/\sigma$
- μ location parameter, $\mu \in \mathbb{R}$
- σ scale parameter, $\sigma > 0$
- ξ shape parameter
  For $\xi = 0$ the function becomes the Gumbel distribution (Type I)
  For $\xi > 0$ the function becomes the Fréchet distribution (Type II)
  For $\xi < 0$ the function becomes the Weibull distribution (Type III)

An advantage of using the AM approach over the POT approach is that the observations in the extreme data set are independent, because consecutive yearly maxima are independent. A disadvantage is that some extreme observations are excluded, because only one event per year is allowed in the extreme data set. Different fitting methods exist in order to fit a trend line through the GEV, such as the Maximum Likelihood function or the Method of Moments.

If the POT approach and the AM approach are compared, the AM approach is less suited to deal with data sets smaller than 100 years. This is because the limited amount of available data points to do the analysis. The block size could be made smaller to increase the number of extreme data points, a disadvantage of this is the violation of the i.i.d. assumption because of seasonal effects. The POT approach has more data points in the extreme data set, if a reasonable threshold value is chosen. **For a data set of less than 100 years a POT approach is generally preferred over the AM approach.**

3.3 Dependence

In statistics two events are considered dependent if the occurrence of one event influences the probability of occurrence of the other event. Independence of two events may be described in a similar way. If two events are independent the occurrence of one event does not influence the probability of occurrence of the other. A more formal definition of independence is stated in Eq. 3.21. If the joint probability distribution of two marginal distributions x_1 and x_2 can be defined as:

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \qquad \text{(Eq. 3.21)}$$

For all x_1 and x_2, the two random variables X_1 and X_2 are independent.

3.3.1 Correlation coefficients

One way to describe dependence between two random variables is by correlation. However, it is important not to confuse correlation and dependence. This is because **two variables might be uncorrelated, but this does not mean that they are independent**. At the same time, the reverse is always true: two independent variables always have zero correlation. Two correlation coefficients are discussed in this section, Pearson's product moment correlation coefficient is defined as:

$$\rho_{X_1,X_2} = \frac{\text{cov}(X_1X_2)}{\sigma(X_1)\sigma(X_2)} = E\left\{\frac{X_1 - E(X_1)}{\sigma(X_1)} \cdot \frac{X_2 - E(X_2)}{\sigma(X_2)}\right\} \qquad \text{with} \quad -1 \leq \rho_{X_1,X_2} \leq 1 \qquad \text{(Eq. 3.22)}$$

In which:

- $\text{cov}(X_1X_2)$ covariance between X_1 and X_2
- $\sigma(X_1)$ standard deviation of X_1, in turn defined as $\sqrt{E\left[(X_1 - E[X_1])^2\right]}$
- $\sigma(X_2)$ standard deviation of X_2
- $\text{E}(X_1)$ expected value of X_1

An alternative to Pearson's coefficient is Spearman's rank correlation, which assesses monotonic relations. If Spearman's rank correlation is equal to 1, all data points above a certain x-value all have a greater value of y. Spearman's rank correlation coefficient is defined as:

$$r_{X_1,X_2} = \rho\left(F_{X_1}(X_1), F_{X_2}(X_2)\right) \qquad \text{with} \quad -1 \leq r_{X_1,X_2} \leq 1 \tag{Eq. 3.23}$$

In which:

$F_{X_1}(X_1)$	ranked variable of X_1
$F_{X_2}(X_2)$	ranked variable of X_2

Alternatively Kendall's τ correlation coefficient is defined, which has more attractive properties over Spearman's r for ranked data sets, as argued by Gibbons (1993). Moreover, Kendall's τ seems to be better suited for smaller data sets compared to Spearman's r. Kendall's τ is defined as:

$$\tau = \frac{\mathrm{C} - \mathrm{D}}{\binom{n}{2}} \qquad \text{with} \quad -1 \leq \tau \leq 1 \tag{Eq. 3.24}$$

In which:

C	number of concordant pairs in a set of observations
D	number of disconcordant pairs in a set of observations
n	number of observations

It is noted that some functions in the `R` programming language may only take Kendall's τ as an input parameter and not Spearman's r, underscoring the importance of knowing the definition of Kendall's τ.

3.4 Bivariate joint probability analysis

There are multiple ways to model bivariate joint probability of two dependent variables. Repko et al. (2004) describes five methods that model the relation between significant wave height (H_s) and peak period (T_p) and compares them. A study by Sellés Valls (2019) summarised all five bivariate models and concluded that each of the five bivariate models has limitations, namely that the different random variables have different parent distributions and that the two variables cannot be assumed independent, as some methods propose. The methods also have limitations when two different variables are used, other than H_s and T_p. Additionally, Zachary et al. (1998) describes a non-parametric approach that can be used for metocean parameters. This approach is adapted with some additional steps by Caires and Gent (2008). A different way to approach bivariate analysis is by using parametric copulas. A copula allows the marginal information to be modelled separately from the dependence structure, thus resulting in a more flexible method. The advantages of copulas over other bivariate methods are described in many articles, such as articles of Michele et al. (2006), Genest and Favre (2007), Li et al. (2014) and Jane et al. (2016) to name just a few. A copula based approach is thus used in this thesis for bivariate and later multivariate analysis.

3.4.1 Copulas and Sklar's theorem

A copula is used in statistics to model bivariate dependence. An elaborate description of copulas and their mathematical properties are found in the book by Nelson (2006), while a more concise guide is provided by Schmidt (2007). The theory about copulas described in this section is largely based on these sources. When a copula is used to model bivariate dependence, different dependence patterns can be applied while the marginal distributions are kept the same. This makes using copulas very flexible compared to other bivariate dependence methods. A copula is a bivariate distribution in the unit square: $[0,1] \quad [0,1]$ with uniform marginal distributions. Two random variables X_1 and X_2 are joined by a copula if the joint distribution can be written as:

$$F_{X_1X_2} = C\left(F_{X_1}(x_1), F_{X_2}(x_2)\right) \tag{Eq. 3.25}$$

By Sklar's theorem, the probability density function f of F satisfies:

$$f(x_1, ..., x_d) = c\left(F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right) \prod_{j=1}^{d} f_j\left(x_j\right) \quad \text{(Eq. 3.26)}$$

In which:

$j \in 1, ..., d$ where d is the dimensionality (equal to the number of variables)

The copula density c is obtained from the copula C for any dimension d by:

$$c(F_1(x_1), ..., F_d(x_d)) = \frac{\partial^d}{\partial F_1(x_1) ... \partial F_d(x_d)} \cdot C(F_1(x_1), ..., F_d(x_d)) \quad \text{(Eq. 3.27)}$$

For the bivariate case $F_1(x_1)$ and $F_2(x_2)$ are often noted as u and v respectively in literature, to shorten the notation:

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) \quad \text{(Eq. 3.28)}$$

For each given continuous joint distribution there is a unique copula. The empirical copula is obtained by ranking the data and re-scaling it into the unit square:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right) \qquad \text{with:} \quad 0 \leq u, v \leq 1 \quad \text{(Eq. 3.29)}$$

In which:

R_i	ranked data of random variable X_1
S_i	ranked data of random variable X_2
n	number of data points in the data set

The empirical copula (representing the data) is then usually compared to a number of theoretical copulas. These theoretical copulas are classified into two main classes: Archimedean and Elliptical copulas, as shown in Figure 3.6. The most common Archimedean copulas are the Gumbel, Clayton and Frank copula. An advantage of the Archimedean copulas is that they are described by an explicit formula with only one parameter. The most common Elliptical copulas are the Gaussian (sometimes called 'Normal') copula and the Student-t copula.

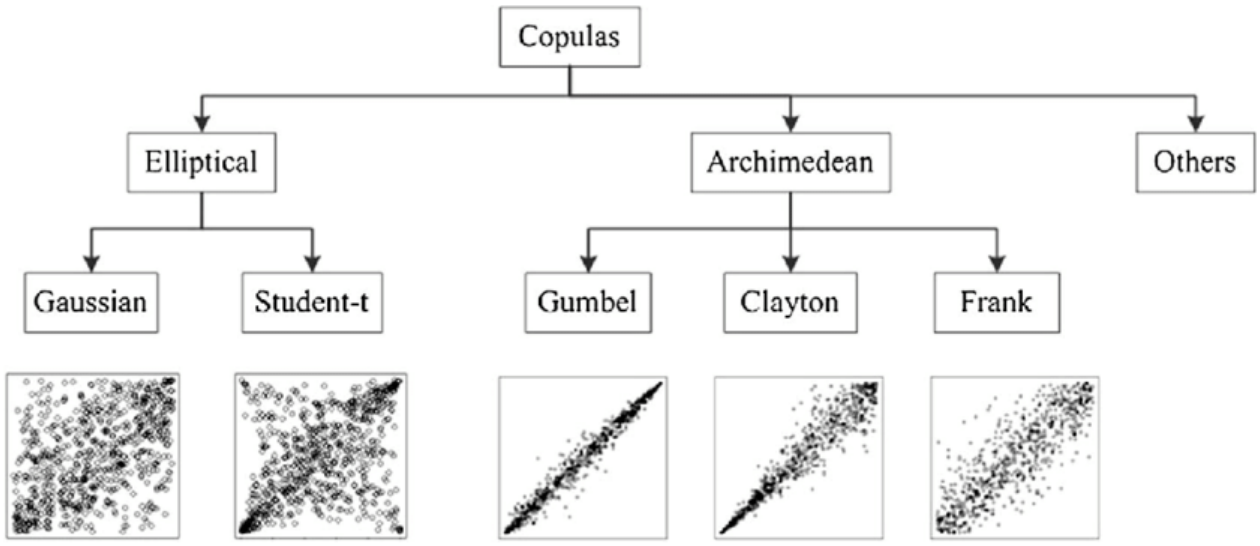


Figure 3.6: Copulas are classified into two main classes: Archimedean and elliptical. Copulas belonging to the other branch are not discussed in this review. Note that the marginal distributions of the copulas are the same (standard uniform), even though their dependence patterns are different. From (Li et al., 2015).

It is noted that not all copulas allow for negative dependence, such as the Gumbel and Clayton copulas. Such copulas are rotated to fit the negative dependence pattern. The mathematical formulation of each copula is shown in Table 3.2.

Name	**Definition**	**Parameter(s)**
Independence	$C(u,v) = uv$	-
Gaussian	$C_\theta(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{\frac{1}{2}}} \exp\left\{-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right\} dxdy$	$-1 \leq \theta \leq 1$
Student-t	$C_{\theta,\delta}(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{\frac{1}{2}}} \left\{1 + \frac{x^2-2\theta xy+y^2}{v(1-\theta^2)}\right\}^{-(\delta+2)/2}$	$-1 \leq \theta \leq 1, \delta \geq 2$
Clayton	$C_\theta(u,v) = \max\left(\left[u^{-\theta} + v^{-\theta} - 1\right]^{\frac{-1}{\theta}}, 0\right)$	$\theta \in [-1,\infty)\backslash\{0\}$
Gumbel	$C_\theta(u,v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right\}$	$\theta \in [1,\infty)$
Frank	$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$	$\theta \in (-\infty,\infty)\backslash\{0\}$
Joe	$C_\theta(u,v) = 1 - \left((u)^{-\theta} + (v)^{-\theta} - u^{-\theta}v^{-\theta}\right)^{1/\theta}$	$\theta > 1$
BB1 (Cla-Gum)	$C_{\theta,\delta}(u,v) = \left(1 + \left[\left(u^{-\theta}-1\right)^\delta + \left(v^{-\theta}-1\right)^\delta\right]^{1/\delta}\right)^{-1/\theta}$	$\theta > 0, \delta \geq 1$
BB6 (Joe-Gum)	$1 - \left(1 - \exp\left[-\left(-\ln\left(1-u\right)^\theta\right)\right)\right)^\delta + \left(-\ln\left(1-\left(1-v\right)^\theta\right)\right)^\delta]^{1/\delta})^{1/\theta}$	$\theta \geq 1, \delta \geq 1$
BB7 (Joe-Cla)	$C_{\theta,\delta}(u,v) = 1 - \left[1 - \left(\left(1-u^\theta\right)^{-\delta} + \left(1-v\right)^{-\delta} - 1\right)^{-1/\delta}\right]^{1/\theta}$	$\theta \geq 1, \delta > 0$
BB8 (Joe-Fra)	$C_{\theta,\delta}(u,v) = \frac{1}{\theta}\left(1 - \left[1 - \frac{1}{1-(1-\delta)^\theta}\left(1-\left(1-\delta u\right)^\theta\right)\left(1-\delta v\right)^\theta\right)\right]^{\frac{1}{\delta}})$	$\theta \geq 1, 0 < \delta \leq 1$
Tawn	$C(u,v) = \exp\left(ln(uv)\cdot A(t)\cdot\frac{ln(v)}{ln(uv)}\right)$	$\theta \in [-1,\infty)$, $t \in [0,1]$
	$A(t) = (1-\psi_2)(1-t) + (1-\psi_1)t + \left[\left(\psi_1(1-t)\right)^\theta + \left(\psi_2 t\right)^\theta\right]^{1/\theta}$	$0 \leq \psi_1, \psi_2 \leq 1$

Table 3.2: A number of copula definitions and parameter limitations .

The copulas shown in Table 3.2 have either one or two parameters, except for the Tawn copula (Tawn, 1988) which has additional parameters (ψ_1, ψ_2 and t) to model asymmetry. The BB1, BB6, BB7 and BB8 are so-called mixed copulas that have two parameters.

The Gaussian- and t-copulas allow for negative dependence. Other copulas can be rotated 90 or 270 degrees to fit a negative dependence pattern. A copula that is rotated 180 degrees is often referred to as a *survival* copula in literature. A copula is rotated using Eq. 3.30, Eq. 3.31 and Eq. 3.32.

$$C_{90}(u,v) = v - C(1-u,v) \quad \text{(Eq. 3.30)}$$

$$C_{180}(u,v) = u + v - 1 + C(1-u,1-v) \quad \text{(Eq. 3.31)}$$

$$C_{270}(u,v) = u - C(u,1-v) \quad \text{(Eq. 3.32)}$$

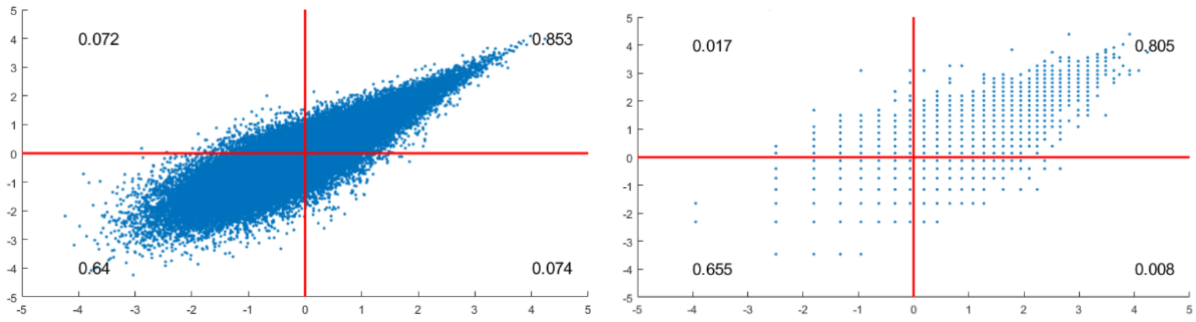
3.5 Goodness-of-fit tests

Several goodness-of-fit tests exist to test if a model is an adequate representation for the data. For copulas and multivariate copula-based models semi-correlations, the Cramér von Mises statistic and the Akaike Information Criterium are often used.

3.5.1 Semi-correlations

Semi-correlations are a goodness-of-fit measure to test if the empirical copula fits the theoretical copula well in a certain quadrant of interest. In order to perform a semi-correlation test the theoretical copula and the empirical

data are transformed from the unit square into the standard normal space. **Pearson's** correlation coefficient of each quadrant is calculated and compared to the correlation of the transformed theoretical copula. If the correlations are similar, the theoretical copula might be a good model for the empirical data in the quadrant of interest, as shown for the north-east quadrant in Figure 3.7.



(a) Data generated with a theoretical copula (standard normal)

(b) Empirical data

Figure 3.7: The transformed theoretical copula (Gumbel in this case) is compared to the the empirical data in each quadrant. Pearson's correlation coefficient is shown in each respective quadrant. Source: author.

3.5.2 Cramér von Mises statistic

Another way to quantify the goodness-of-fit of a copula is the Cramér von Mises statistic (CvM). The empirical copula of Eq. 3.29 is subtracted by the theoretical copula, the result is summed, squared and multiplied by the number of points (n). If a number of theoretical copulas are compared to the empirical data, the lowest CvM is the best fit for that data. The Cramér von Mises statistic is defined as:

$$\mathrm{CvM} = n \sum_{i=1}^{n} \left\{ C_n \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - C_{\theta n} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right\}^2 \qquad \text{(Eq. 3.33)}$$

Most definitions of the Cramér von Mises statistic include a multiplication with the number of points n, although this multiplication is left out in some literature (Genest et al., 2009). The Cramér von Mises statistic is often used in combination with semi-correlations to determine the best fit for the data.

3.5.3 Akaike Information Criterium (AIC)

A way to compare different models of to a data set is by means of the Akaike Information Criterion (AIC from now on). The method, first described by Akaike (1973), balances the goodness-of-fit (by means of a likelihood function) with the number of parameters. A penalty is assigned to using too many parameters in the model, thus discouraging overfitting. For bivariate copulas the number of parameters is one or two depending on the copula of choice. The lowest AIC score is assumed to be the best model for the data. The AIC is defined as:

$$\mathrm{AIC} = 2k - 2log(L) \qquad \text{(Eq. 3.34)}$$

In which:

- k number of parameters
- L likelihood function estimate

It is noted that the AIC scores mean nothing on their own, but may only be used to compare models. A clear explanation and example using the AIC is provided by Snipes and Taylor (2014).

3.6 Multivariate probability models

As stated, a copula based approach is used to model the dependence between two variables, such as a significant wave height and mean wave period of a single sea state. However, given that the primary research objective is to develop a methodology to analyse *multiple* sea states at once, the bivariate analysis has to be expanded to more

than two dimensions. Performing a multivariate analysis requires high degrees of statistical understanding with multiple options available in literature. A number of options to model multivariate dependence are discussed briefly, before moving on to the multivariate method of choice. The models that are discussed in the following sections are all copula based.

3.6.1 Non-parametric Bayesian Network

A Non-Parametric Bayesian Networks (NPBN) is a directed acyclic graph where the nodes represent the variables and the arcs are represented by (conditional) rank correlations, first described by Kurowicka and Cooke (2004). The absence of an arc between two nodes is interpreted as the two nodes being independent. The visual representation of the variables into nodes and arcs allows for a complete and relatively simple overview of the model. A comprehensive study on Bayesian Networks is provided by Weber et al. (2012). The arcs between the nodes are represented as (conditional) one-parameter copulas. Furthermore, the copulas used must exhibit the *zero independence* property. This property entails that a copula is independent when a correlation of zero is applied, thus excluding the use of the t-copula which does not have this property. Because the arcs in the NPBN are directed, a choice has to be made which variables direct to which other variables. This is referred to in literature as a *parent* and *child* node respectively. This decision is up to the modeller, as no unique procedure of the qualitative part of the model exists. It is noted that NPBNs allow for modelling of dependent extreme variables (Hanea et al., 2015), but due to the scope of this thesis, NPBNs are not further explored.

3.6.2 Multivariate copulas

Copulas provide a flexible dependence modelling tool in the bivariate case, even allowing expansions to the trivariate case using conditional laws (Chakak and Koehler, 1995). Problems arise when extending copulas to higher dimensions. Building multivariate copulas that have a closed form remains an open issue. Estimating the required parameters has proven to be a challenging task for multidimensional copulas. Multivariate copulas are thus not used in this thesis.

An extensive review of multivariate copulas is provided in Chapter 2 of the book of Kurowicka and Joe (2011), with many further references on the topic of multivariate copulas therein.

3.7 Vines

Given that problems arise when one tries to model dependence structures using multivariate copulas, a different approach is considered by using so-called *vines*. First developed by Cooke (1997), a vine $\mathcal{V}$ is a nested set of trees $\mathcal{V} = \{T_1, \ldots, T_{n-1}\}$, where n is the number of variables, also referred to in literature as *nodes*. The edges of each tree j are the nodes of the next tree $j+1$, where $j = 1, ..., n-1$. Regular and irregular vines are defined. In a regular vine with n variables, one node from a certain tree T_j connects to two edges from the next tree T_{j+1}, whereas irregular vines do not have this property. An example of a regular and irregular vine is shown in Figure 3.8.

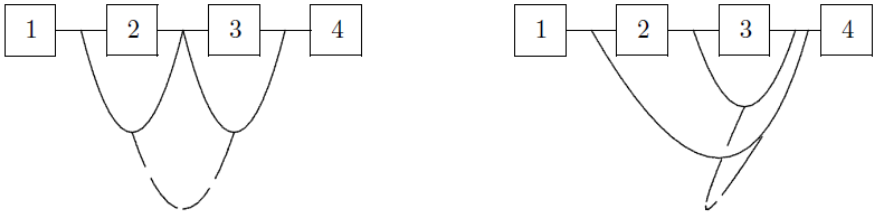


Figure 3.8: Two graphs showing a regular vine (left) and an irregular vine (right). The vines have four numbered variables or nodes. The arcs between the nodes represent the edges between two variables. From Kurowicka and Joe (2011).

A regular vine on n nodes has $n(n-1)/2$ edges, which are represented by arcs between the nodes. A regular vine on four nodes thus has three trees and six edges. Sampling theory and estimation are well developed in the case of regular vines. Furthermore, regular vines have proven to be more useful than irregular vines in many applications.

Thus irregular vines are not further explored in this research. Regular vines can themselves be divided into two types, partial correlation vines and pair-copula constructions (PCCs). In case of a partial correlation vine, the dependence between the edges is represented by a correlation matrix. The edges of PCCs are represented by bivariate copulas. Partial correlation vines are very similar to NPBNs, with one notable exception being that NPBNs are *directed* acyclic graphs while partial correlation vines are undirected. An elaborate comparison between partial correlation vines, PCCs and NPBNs is provided in Chapter 14 of Kurowicka and Joe (2011) with many further references therein.

3.7.1 Pair copula constructions (vine copulas)

Pair copula constructions are often referred to as vine copulas in literature and will be named so from now on. Within the category of regular vine copulas, different structures are defined. A regular vine copula on more than three nodes has at least two different classes: canonical vines and drawable vines, shown in Figure 3.9. These vine copulas are often abbreviated as C- and D-vines respectively.

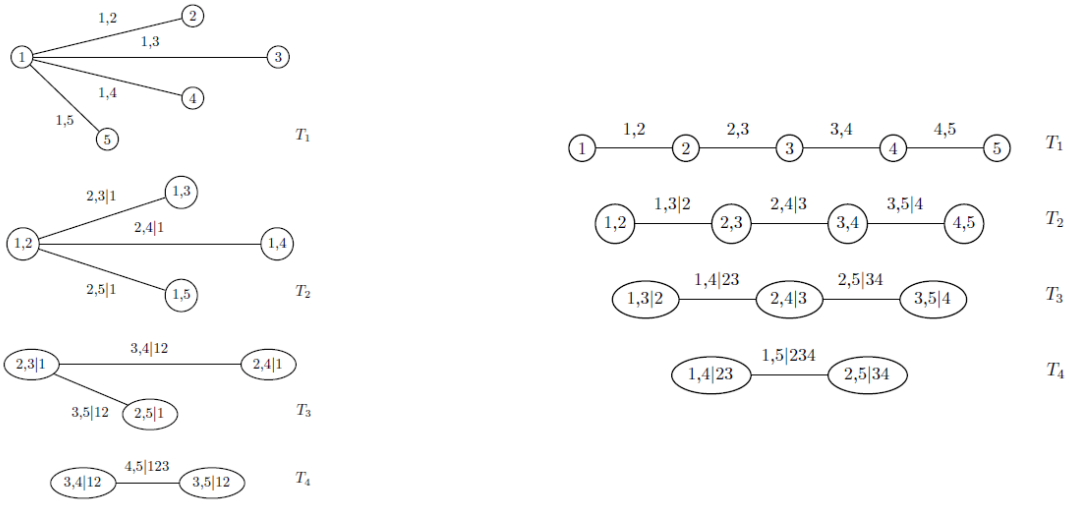


Figure 3.9: Two graphs showing a C-vine copula (left) and a D-vine copula (right). The vine copulas have five numbered nodes, four trees and ten edges each. From Brechmann and Schepsmeier (2013).

It is noted that the regular vine left in Figure 3.8 is also a D-vine. The difference with the right vine in Figure 3.9 (besides the number of nodes) is that the edges and trees are labelled. Moreover, the trees in Figure 3.9 are drawn separately. The labels on the edges of the vine copulas show which nodes the edges connect to and on which node the edges are conditioned. As stated in Aas et al. (2009), **each edge represents a bivariate copula**, with the conditioned variables left of the |-sign and the conditioning variables on the right of the sign. The conditional bivariate distribution $h_\theta(u, v)$ is obtained by Eq. 3.35, where v is the conditioning variable.

$$h_\theta(u, v) = \frac{\partial C_\theta(u, v)}{\partial v} \qquad \text{(Eq. 3.35)}$$

The parameter θ is the copula parameter, which is different depending on the copula of choice, as shown in Table 3.2. Fitting a vine copula is performed in a number of steps. First, the vine structure is selected, which means that the nodes are ordered in a certain way. It may not be known beforehand which node structure is appropriate for the problem at hand. In this case all known regular vines are fitted and the structure is determined that best fits the data. It is noted that the number of regular vines increases rapidly with the number of nodes, as demonstrated by Morales-Nápoles et al. (2010). The number of regular vines on n nodes is calculated by Eq. 3.36.

$$\text{number of regular vines} = \begin{pmatrix} n \\ 2 \end{pmatrix} \times (n-2)! \times 2^{(n-2)(n-3)/2} \qquad \text{(Eq. 3.36)}$$

The number of regular vines rapidly increases with n:

n	number of regular vines
3	3
4	24
5	480
6	23040
7	2580480
8	660602880

It should be no surprise that the required computational time to fit all possible regular vines increases with the number of nodes. After the vine structure is selected, the adequate bivariate copulas have to be selected and estimated for each (conditional) bivariate distribution, represented by the edges in the graph. All fitted vine copulas are compared to each other by using the AIC of Eq. 3.34. Two important results can be extracted once the best vine copula has been fitted:

- A multidimensional sample in the unit hypercube can be extracted from the vine copula.
- The vine copula density at any point in the unit hypercube can be calculated.

Aas et al. (2009) describes an algorithm that samples from C- and D-vines, resulting in a n-dimensional sample in the unit hypercube. The second result is the density that can be obtained for any point in the unit hypercube. The four dimensional density $f(x_1, x_2, x_3, x_4)$ is expressed for a C-vine in Eq. 3.37 and for a D-vine in Eq. 3.38. As stated before, a vine copula on three nodes only has one class. The density of a vine on three nodes $f(x_1, x_2, x_3)$ is defined in Eq. 3.39.

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = {} & f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4) \\ & \cdot c_{12}\{F_1(x_1), F_2(x_2)\}\cdot c_{13}\{F_1(x_1), F_3(x_3)\}\cdot c_{14}\{F_1(x_1), F_4(x_4)\} \\ & \cdot c_{23|1}\{F(x_2|x_1), F(x_3|x_1)\}\cdot c_{24|1}\{F(x_2|x_1), F(x_4|x_1)\} \\ & \cdot c_{34|12}\{F(x_3|x_1,x_2), F(x_4|x_1,x_2)\} \end{aligned} \quad \text{(Eq. 3.37)}$$

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = {} & f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4) \\ & \cdot c_{12}\{F_1(x_1), F_2(x_2)\}\cdot c_{23}\{F_2(x_2), F_3(x_3)\}\cdot c_{34}\{F_3(x_3), F_4(x_4)\} \\ & \cdot c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\}\cdot c_{24|3}\{F(x_2|x_3), F(x_4|x_3)\} \\ & \cdot c_{14|23}\{F(x_1|x_2,x_3), F(x_4|x_2,x_3)\} \end{aligned} \quad \text{(Eq. 3.38)}$$

$$\begin{aligned} f(x_1,x_2,x_3) = {} & f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3) \\ & \cdot c_{12}\{F_1(x_1), F_2(x_2)\}\cdot c_{23}\{F_2(x_2), F_3(x_3)\} \\ & \cdot c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \end{aligned} \quad \text{(Eq. 3.39)}$$

In which:

$f_1(x_1)...f_4(x_4)$	marginal densities of each parameter
c	bivariate (conditional) copula density
F	(conditional) cumulative probability

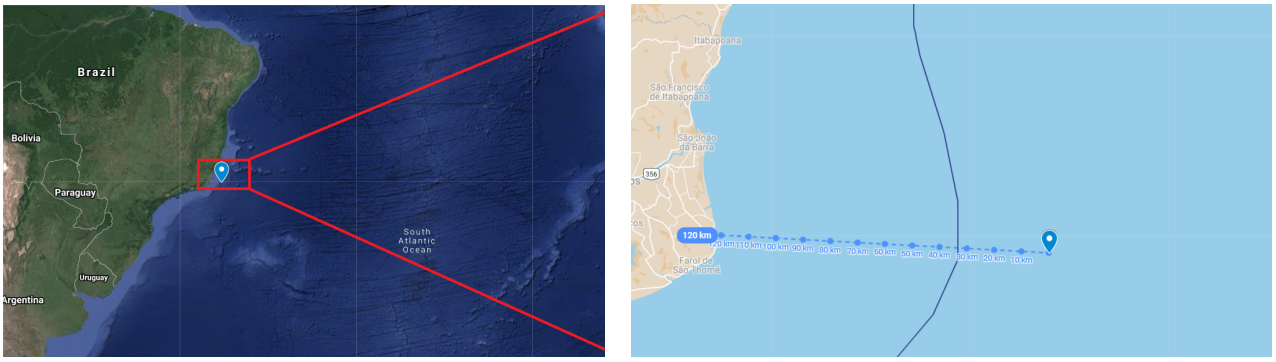
A closed form of the multivariate cumulative probability is not available for vine copulas.

4 Data collection

This chapter describes the process of collecting the necessary wave data for the analysis. First, the location of interest is described. The reanalysis data set is selected and the hourly values of the variables of interest are acquired for the selected time period. Next, an overview is given of the data. The result of the data collection process is a data set of hourly values of the variables of interest from 1979 to 2018.

4.1 Location of interest

Wave data are collected at the location of interest. The location of interest is situated at the following coordinates: 22S, 39.8W, shown on a map in Figure 4.1. This corresponds to the Southern Atlantic ocean, 120 km off the coast of Brazil. The water depth at the location is about 1100 m. The large depth ensures that all waves at the location of interest can be considered deep water waves, avoiding any shallow water and local bathymetry effects in the analysis. Wave data acquired from reanalysis data sets at locations with limited depth are considered less reliable. Reliable long term data sets nearshore are often not available for the design of coastal structures. If one is interested in extreme wave conditions nearshore, the offshore data have to be transformed to the nearshore situation to analyse how the presence of combined wave systems affects the nearshore design values for the variables of interest. Moreover, even if reliable long term nearshore data would be available, it is still recommended to start with offshore data. This is because in the case of a breakwater, the nearshore wave climate might change because of the construction of the breakwater itself. It is assumed that the offshore wave climate is not affected by such a coastal structure.



(a) The location of interest is located about 120 km offshore of the coast. The blue marker represents the location of interest.

(b) The solid blue line represents the edge of the continental shelf, east of which the depth increases quickly to more than 1000 m.

Figure 4.1: The location of interest in the southern Atlantic ocean. Source: Google maps.

The wind climate at the location of interest is relatively mild, with trade winds blowing from the north-east most of the time, with the wind direction shifting to east to south-west occasionally. In the period from May to August (winter on the southern hemisphere), strong winds from the south occur regularly. Wind coming from land (north-west) almost never occurs throughout the year. Wind data from the ERA5 reanalysis data set are plotted in a rose in Figure 4.2 in order to illustrate the local wind climate. Besides this, wind data are not further used in this research.

Given the wind climate at the location of interest, it is no surprise that the area is subjected to wind waves coming mainly from the north (Figure 4.3c). High swell events from the east and south are observed as well, although less frequent in the period from December to February. Swell events originating from the south are generated partly in the turbulent region called the *roaring forties*, named after the 40th southern latitude known for its frequent storms. Combined sea states with wind waves from the north and swell from the east or south are observed regularly at the location of interest, making it a suitable case to answer the primary research question.

4.2 Reanalysis data set

The data considered in this research consist of hourly values downloaded from the ERA5 reanalysis data set. ERA5 is co-funded by the European Centre for Medium-Range Weather Forecasts and provides open source meteorological data. ERA5 combines model data with satellite observations into a global reanalysis data set that provides a wave parameter time series from 1979 until present day (ECMWF, 2018). The documentation of the ocean parameters is found in Bidlot (2016). The hourly values of each parameter provided in the data set that is used in this research are from the year 1979 to 2018, meaning that the data set of each parameter contains 350640 data points. The resolution of the reanalysis data grid is 0.25°x0.25°. At each grid point a 2D variance density spectrum is calculated, which describes how the wave energy is distributed over the frequencies and directions. The model derives the other wave parameters based on this calculated 2D spectrum. The total significant wave height is defined using the spectral zeroth order spectral moment H_{m0} from Eq. 3.11, for instance.

There are multiple mean wave periods available for use in the reanalysis data set of ERA5, $T_{m-1,0}$, $T_{m0,1}$ and $T_{m0,2}$ (defined in Eq. 3.7, Eq. 3.8 and Eq. 3.9 respectively). The mean wave period based on the first order spectral moment ($T_{m0,1}$) can be used directly in a SWAN model (Booij et al., 1999), used to transform offshore events to nearshore. The peak period can also be used in SWAN for the total wave, wind wave and swell partitions. However, the peak period is only available for the total sea state, but not for the wind wave and swell partitions.

4.2.1 Partition in swell and wind waves

The ERA5 data set contains wind waves and swell wave parameters. Wind waves are defined as spectral components that are subjected to wind forcing. The other components are classified as swell waves. The partition criterion is given by Bidlot (2016):

$$1.2 \times 28\,(u_*/c)\cos(\theta - \phi) > 1 \qquad \text{(Eq. 4.1)}$$

Where u_* is the friction velocity is a function of the air density and the atmospheric surface stress. The parameter c is the phase speed of the waves and ϕ is the wind direction. The total significant height of the total, wind wave partition and total swell partition are related by Eq. 4.2.

$$\text{swh} = \sqrt{\text{shww}^2 + \text{shts}^2} \qquad \text{(Eq. 4.2)}$$

In which:

swh	[m]	total significant wave height
shww	[m]	wind wave partition partition significant wave height
shts	[m]	swell partition significant wave height

4.2.2 Variables of interest

The variables of interest are obtained from the ERA5 reanalysis data set. As stated, the ERA5 reanalysis data set contains wind wave and total swell parameters. Moreover, it is possible to partition the total swell further in up to three swell partitions by means of a modified partitioning algorithm of Hanson and Phillips (2001). However, due to the scope of this thesis, only a wind wave and total swell partition are considered. The variables of interest are listed in Table 4.1. The wind variables (*wind* and *dwi*) are only used to illustrate the local wind climate and are not further analysed in this thesis.

ERA5	Literature	Unit	Description
swh	H_{m0}	m	total significant wave height
mwd	θ	°	total mean wave direction (clockwise from the north)
mp1	$T_{m0,1}$	s	total mean wave period based on the first order spectral moment
pp1d	T_p	s	total peak wave period based on a 1D spectrum
wdw	σ_θ or DSPR	rad	total wave directional width or directional spreading
wsp	γ	-	total wave spectral peakedness or peak enhancement factor
shww	H_{m0}	m	wind wave partition significant wave height
mdww	θ	°	wind wave partition mean wave direction (clockwise from the north)
p1ww	$T_{m0,1}$	s	wind wave partition mean wave period based on first spectral moment
p2ww	$T_{m0,2}$	s	wind wave partition mean wave period based on second spectral moment
dwww	σ_θ or DSPR	rad	wind wave partition directional width or directional spreading
shts	H_{m0}	m	swell partition significant wave height
mdts	θ	°	swell partition mean wave direction (clockwise from the north)
p1ps	$T_{m0,1}$	s	swell partition mean wave period based on first spectral moment
p2ps	$T_{m0,2}$	s	swell partition mean wave period based on second spectral moment
dwps	σ_θ or DSPR	rad	swell partition directional width or directional spreading
wind	U_{10}	m/s	Wind speed 10 m above the surface
dwi	θ_w	°	Wind direction (clockwise from the north) 10 m above the surface

Table 4.1: Variables of interest available in the ERA5 reanalysis data set, including their more common literature notation, unit and description.

The directional spreading is only available as an hourly value and not per frequency. Remember that in order to generate 2D JONSWAP spectra, five parameters are needed, as listed in Table 3.1. Notably absent in the data set are the T_p and γ of the wind wave- and swell partitions. The spectral peakedness of wind wave and swell partition is derived with H_{m0} and $T_{m0,2}$ later on. As stated in Section 4.2, $T_{m0,1}$ is used for the other analyses.

As noted previously, up to three *sub* partitions of swell are available in the data set. However, for these sub partitions only H_{m0}, $T_{m0,1}$ and θ are available, lacking directional spreading and wave spectral peakedness needed for the creation of a 2D JONSWAP spectrum. The length of the time series is from January 1st 1979 to December 31st 2018, resulting in a time series of 350640 hourly values for each parameter. Variables of interest will mostly be referred to by their ERA5 abbreviation from now on.

4.3 Data overview

The data set is visualised using roses and density scatters. The linear correlation between the parameters is calculated to quantify the dependence between parameters of interest. The goal of the data visualisation is to get a better understanding of the local wave climate and the relations between the different parameters. The parameters $pp1d$, wdw, wsp, $p2ww$, $dwww$, $p2ps$ and $dwps$ of Table 4.1 are not shown in this section, as they are treated in later chapters.

4.3.1 Wind and wave roses

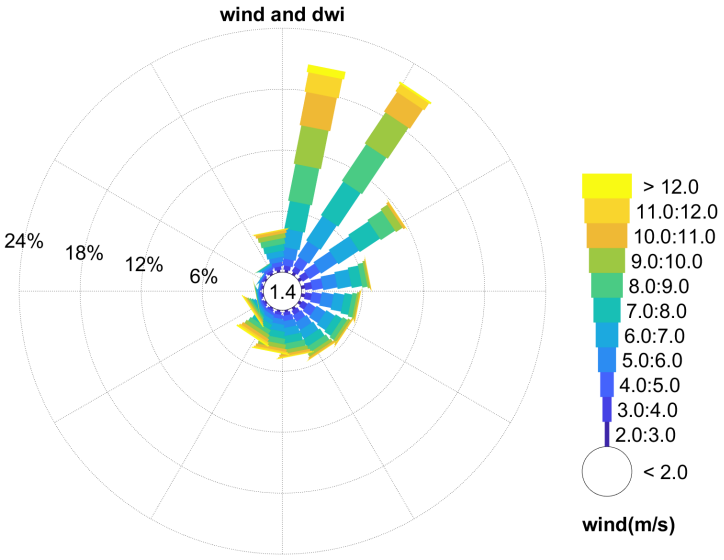
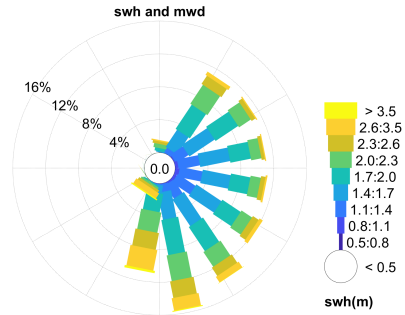
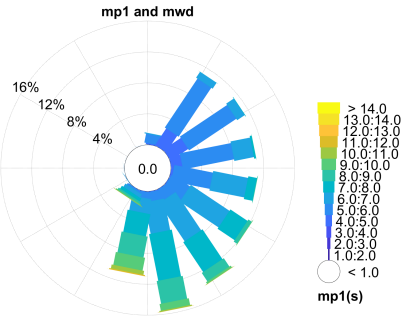


Figure 4.2: Wind climate represented by a rose of the wind speed 10 meter above the surface as a percentage of the total occurrence.

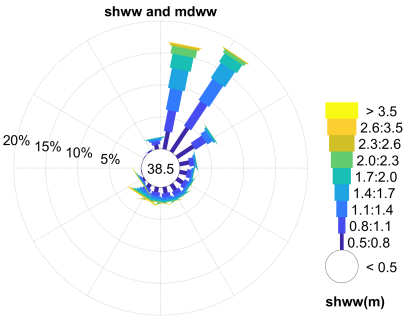
Besides the wind rose shown in Figure 4.2, wave roses of the total, wind wave partition and total swell partition are shown in Figure 4.3. The wave roses show the direction of the total and each partition as a percentage of the total occurrence. The number in the circle in the middle of each wave rose is the percentage of waves that are under a certain threshold, which is shown in each respective legend. Each rose has sixteen directional bins, with north being at 0 or 360 degrees.



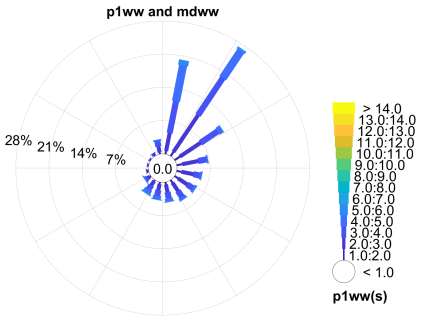
(a) Total significant wave height.



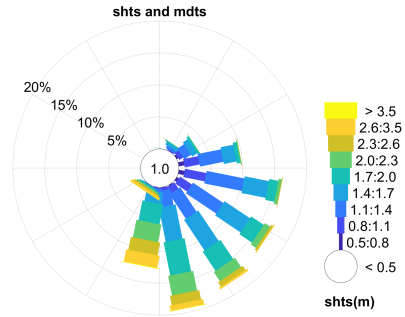
(b) Total mean wave period.



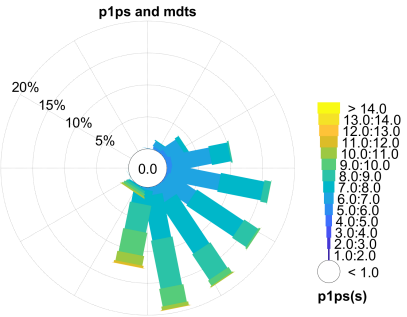
(c) Significant wave height wind waves.



(d) Mean wave period wind waves.



(e) Significant wave height swell.



(f) Mean wave period swell.

Figure 4.3: Roses of the total wave, wind wave partition and swell partition for the significant wave height and mean wave period. The data show that the wind wave and swell partitions have a different direction most of the time.

The data show that most wind waves originate from the north to east, with swell mainly coming from the east and south, directions that have long fetches for swell to develop. Note that a large percentage of the wind waves is actually under the threshold of 0.5 m (Figure 4.3c), which is not true for swell. Extreme sea states are barely visible in the wave roses, as they (per definition) only account for a very small percentage, thus *density scatter* plots are used to visualise the data in a different manner.

4.3.2 Directional density scatters

Density scatter plots are a useful method to visualise large data sets. Density scatters work by first plotting all data points as individual dots. If data points are in close proximity to each other the point density in the area increases, indicated by a brighter colour. Areas with relatively low point density have a darker colour. Contours are added that indicate lines of equal point density. A density scatter plot showing the direction of the wind wave partition (*mdww*) versus the direction of the total swell partition (*mdts*) for all 350640 data points is shown in Figure 4.4.

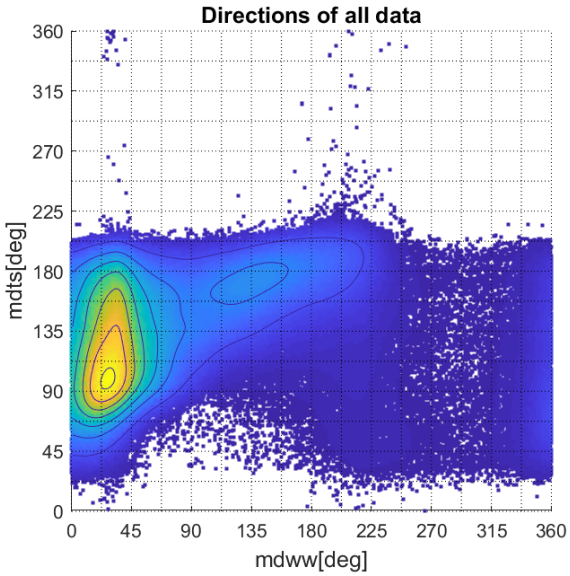
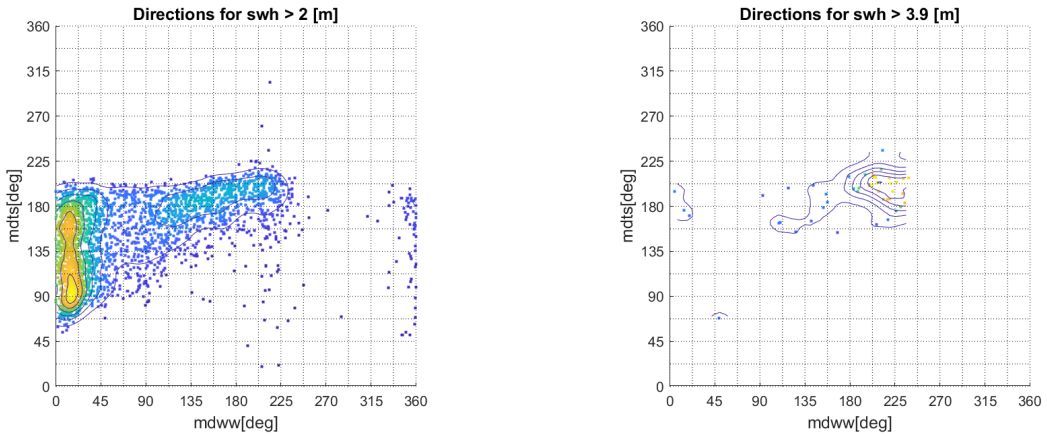


Figure 4.4: A density scatter plot showing the direction of the wind wave partition (*mdww*) versus the direction of the total swell partition (*mdts*) for all 350640 data points (40 years). A higher point density is indicated with a brighter colour. The contours indicate lines of equal point density.

Figure 4.4 shows a high density of points between 0 and 45 degrees for wind wave direction and 60 and 180 degrees for the swell direction. This observation is in agreement with the wave roses of Figure 4.3, which show the data in a different way.

Given that the directions of the normal conditions are within a certain interval, one could assume that the directions of the partitions are within this same interval for all extreme events. However, this assumption is not valid for this data set. To show that this is the case, the data are conditioned on the total significant wave height (*swh*). Figure 4.5a shows the direction of the wind wave and swell partition for a total significant wave height higher than two meters. The density scatter plot shows a high concentration of points in a similar directional interval as in Figure 4.4.



(a) (b)

Figure 4.5: Direction of wind waves and swell for different values of *swh*.

The directions of events with *swh* above 3.9 m are shown in Figure 4.5b. This figure shows that for the highest *swh* the point density is highest for the a directional interval between 160 and 230 degrees approximately of the wind sea and swell partition. It is thus noted that the directional interval with the highest point density is different for normal conditions and extreme conditions.

4.3.3 Significant wave height and mean wave period

The dependence between the significant wave height and mean wave period ($T_{m0,1}$) in the data set is expressed by means of a correlation coefficient for the total wave, wind wave- and swell partition. Generally Kendall's τ is preferred over Spearman's r, because it is slightly more robust (Croux and Dehon, 2010). However, because Kendall's τ takes very long to compute for large data sets, Spearman's r is shown in Figure 4.6 to express the dependence. For the extreme data set, which is considerably smaller in size, Kendall's τ is used. Definitions of the different correlation coefficients and further references are described in Section 3.3 in the literature study.

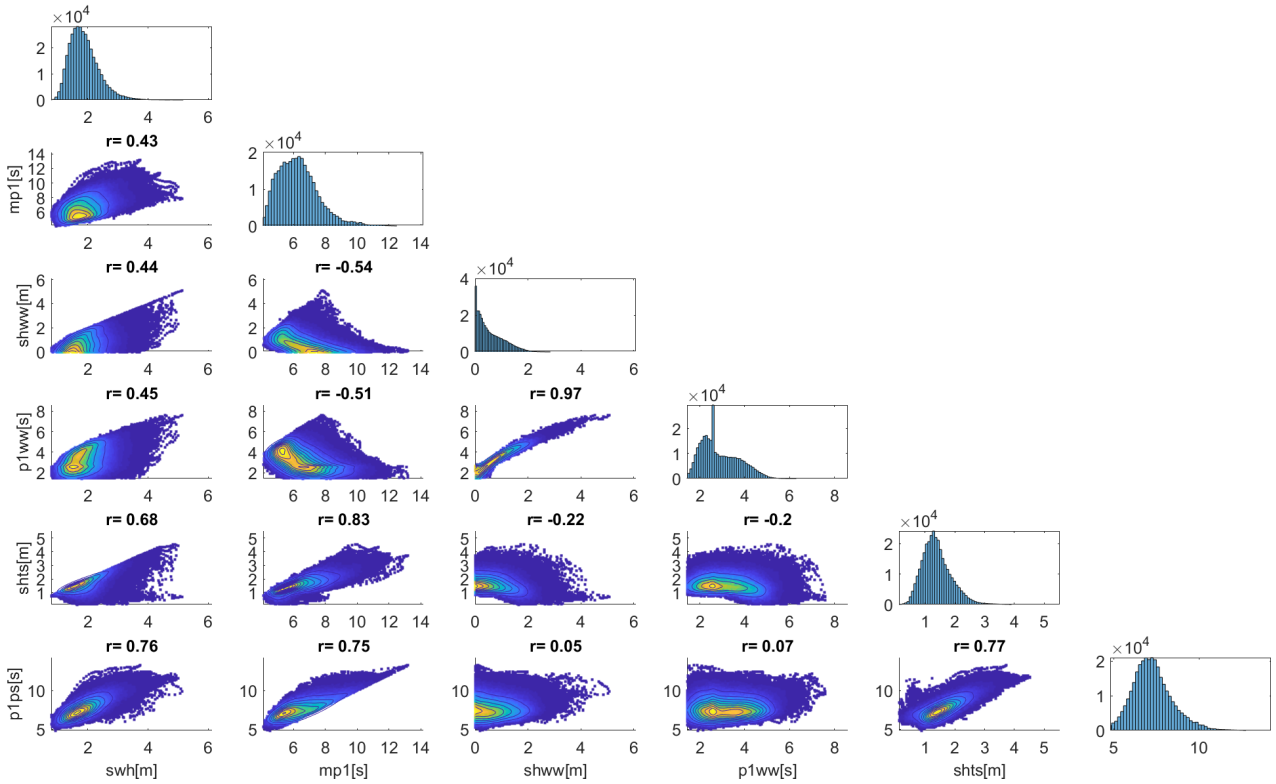


Figure 4.6: Density scatter plots of significant wave height and mean wave period of the total, wind wave partition and swell partition for the entire data set (40 years). Spearman's r is calculated for each combination of variables and shown in the title of the respective sub figures. The occurrence histograms of each variable are plotted in the diagonal.

The histograms shown in Figure 4.6 show the number of occurrences within a certain bin for each parameter. The data show that for *shww* the most probable value is close to zero, while for *shts* the most probable value is between 1 and 2 m. This indicates that the presence of swell is more constant and almost always higher than zero for the location of interest, while wind waves are not always present. The occurrence histogram of $p1ww$ shows a spike around 2.5 s, which mostly coincides with a very low significant wave height. This spike in occurrences stands out in comparison with the rest of the data and is investigated further.

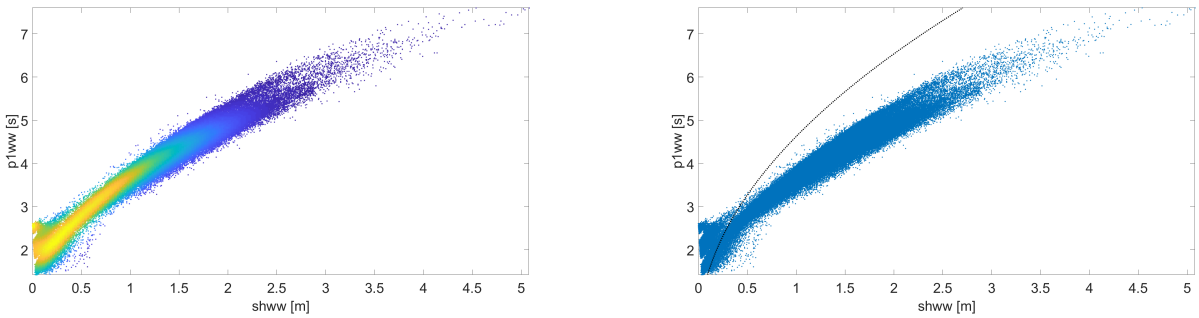


Figure 4.7: Density scatter plot of the significant wave height of wind waves ($shww$) versus the mean wave period of wind waves ($p1ww$) shown in (a) and the same data plotted together with a steepness line shown in (b).

Figure 4.7 shows that many points in the lower tail of the $shww$-$p1ww$ data set seem to have a steepness that is different compared to data points in the upper tail. The black line in Figure 4.7b represents a constant steepness of $s = 0.03[-]$, with s being defined in Eq. 4.3.

$$s = \frac{2\pi \cdot shww}{g \cdot p1ww^2} \qquad \text{(Eq. 4.3)}$$

The gravitational constant g of equation Eq. 4.3 is assumed as 9.81 $[m/s^2]$. It is assumed that data points with $s < 0.03[-]$ are not wind waves. Next, these data points are censored and the occurrence histograms are plotted again, as shown in Figure 4.8.

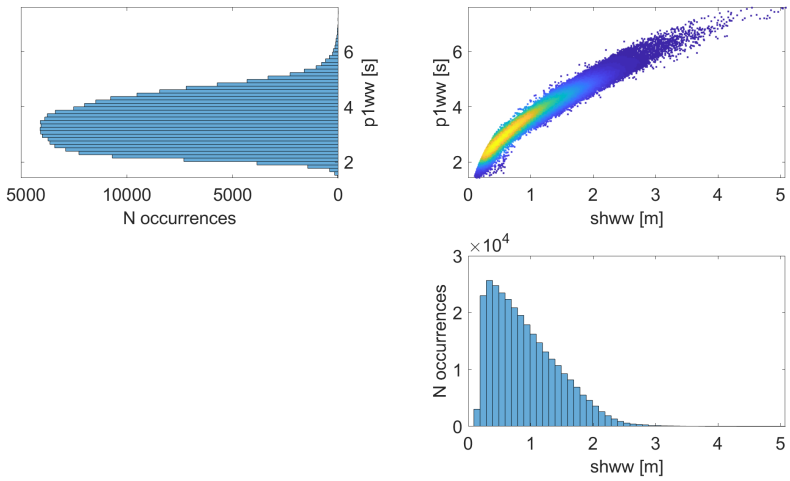


Figure 4.8: Density scatter plot of the significant wave height of wind waves ($shww$) versus the mean wave period of wind waves ($p1ww$) together with the occurrence histograms. Unlike Figure 4.6, data that have a steepness of $s < 0.03[-]$ are not plotted in this figure.

The occurrence histograms in Figure 4.8 do not show the spike that is shown in the occurrence histograms of Figure 4.6. Wind wave data that are outside a certain steepness are considered unrealistic. It is therefore assumed that the spike is the result of quirks in the partitioning algorithm, which apparently assigns some swell energy to the wind wave partition. As will be shown later in Figure 5.4 and 7.1, the values of $shww$ and $p1ww$ in the extreme data set all have a steepness $s > 0.03[-]$. Therefore, the spike shown in Figure 4.6 is ignored in the rest of the analyses.

The density scatter plot of the wind wave partition ($shww$ versus $p1ww$) is compared to the total swell partition ($shts$ versus $p1ps$). Figure 4.6 shows that on average the significant wave height of the total swell partition

is higher than the wind wave partition. However, the highest value of *shww* is higher than the highest value of *shts*. Furthermore, *shww* and *p1ww* are almost completely dependent, while *shts* and *p1ps* are less dependent, as expressed by a lower correlation coefficient in the latter. This observation is in concurrence with general assumptions about wind waves and swell.

As stated before, two independent variables have a correlation coefficient that is equal to zero, while the reverse is not always true. Although arbitrary, in practice two variables with $-0.15 < r < 0.15$ are considered independent in most cases. A weak negative correlation ($r = -0.22$) is observed between *shww* and *shts*, which could be interpreted as the wind wave and swell being not totally independent. Generally speaking, however, wind waves and swell are generated in different locations so should always be independent. The observed negative correlation could be the result of interaction between wind waves and swell or be due to effects of the partitioning algorithm.

4.4 Concluding remarks

- Considering the available wave parameters in the ERA5 reanalysis data set, the absence is noted of the spectral peakedness (γ) of the wind wave- swell partition, needed to generate 2D JONSWAP spectra. This parameter will be derived from the other parameters, as explained later on.
- Wave roses show that most of the wind waves originate from the north, while most of the swell originates from the south and east. However, it is concluded from Figure 4.5b that the largest waves originate from approximately 160 to 230 degrees. From this conclusion, it follows that the highest total significant wave height is most likely **not** the result from a crossing sea state.
- From the density scatters of the significant wave height and mean wave period of Figure 4.6 it is concluded that given normal conditions, the sea state is mostly dominated by swell. Mean wave period and the significant wave height for the wind waves partition are on average lower than for the swell partition. However, for extreme significant wave heights, the wind waves are likely to be higher.
- The wind wave parameters *shww* and *p1ww* are almost fully correlated, while the swell parameters *shts* and *p1ps* are less correlated. For the total wave the dependence is even less, especially in the upper tail. Some extremes of *swh* have either a high wind wave or swell component or a combination of both wind waves and swell. Small correlations are observed between wind wave and swell parameters, which are mostly negligible, although the negative correlation between *shww* and *shts* seems to be significant. Generally speaking, however, wind waves and swell are generated in different locations so should always be independent. The observed negative correlation could be the result of interaction between wind waves and swell, be due to effects of the partitioning algorithm in the reanalysis data set, or be caused by something else entirely.

5 Univariate analysis

The data collection described in Chapter 4 results in a multivariate data set of hourly values of variables from 1979 to 2018. In this chapter the extreme data set is defined. First the directional intervals of interest are chosen. The dominant variable is selected out of the variables of interest. A Peak-Over-Threshold (POT) approach is applied, resulting in a Generalised Pareto Distribution (GPD) for the dominant variable. Univariate extreme value analysis is also applied to other variables, in order to compare these to the design values later. The concomitant variables are fitted to different distributions, which do not necessarily have to be extreme value distributions. Next, the dependencies between variables of the extreme data set are examined and associated values are determined for the mean wave direction, directional spreading and peak enhancement factor.

5.1 Selecting directions of interest

The directions of interest are selected, which means that the data set is conditioned on the mean wave direction of wind waves ($mdww$) and swell ($mdts$). From now on all data points considered are between:

$$0° < mdww < 45° \text{ and } 67.5° < mdts < 112.5° \qquad \text{(Eq. 5.1)}$$

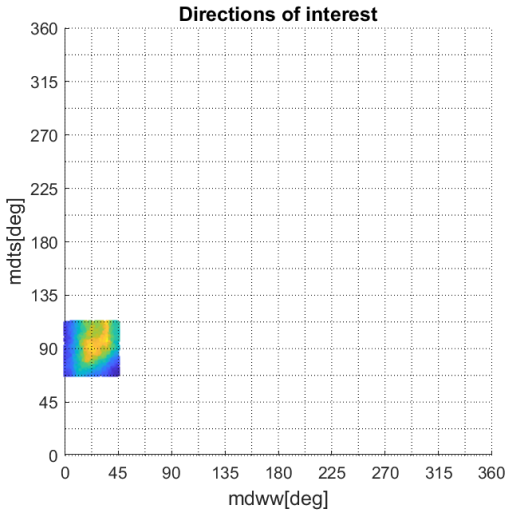


Figure 5.1: Density scatter for the directions of interest. Unlike Figure 4.4, the iso-lines are not shown in this figure.

Because of the scope of this thesis, it is assumed that the directional interval of Eq. 5.1 is the most interesting combination of directions from an engineering point of view. However, it is stressed that **the choice of the directions of interest is only for illustrative purposes in this thesis**, as the data show that the highest values for swh are not within this specific directional interval. Restricting the directions has the advantage that the directions can be assumed as some deterministic value between the directions of interest instead of a circular quantity, simplifying the analysis later on.

A different combination of directions could be chosen if the engineer performing the analysis decides that it is more relevant to the problem at hand. Moreover, if the entire analysis is automated, the procedure could be repeated for all combinations of directions with relative ease. But for now, the analysis is restricted to one combination of directional intervals. The total amount of data points within this interval is 62333, which is a lot less than the 350640 data points (40 years) of total data set.

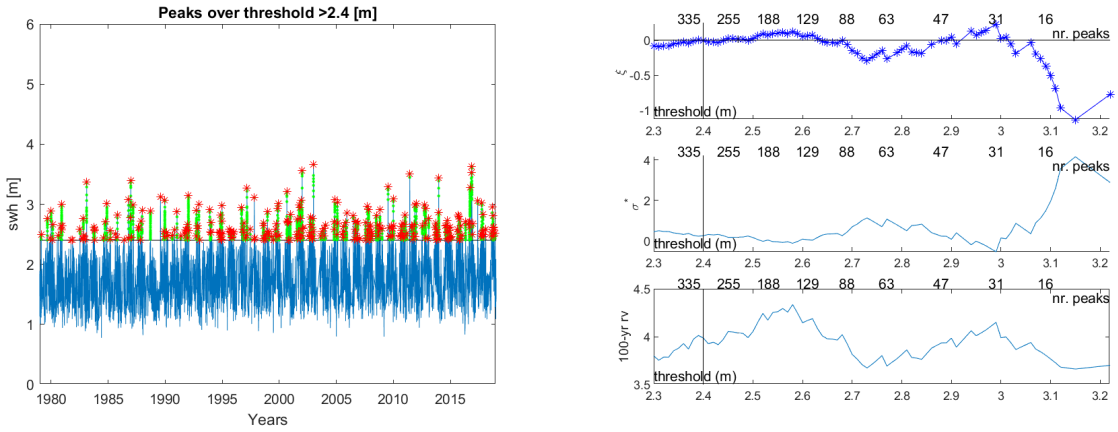
5.2 Univariate extreme value analysis

A dominant variable is chosen out of the variables of interest. Univariate extreme value analysis is performed on the dominant variable as well on some other variables.

5.2.1 Dominant variable

For most engineering applications the significant wave height is a more important load parameter than the mean wave period. The data set contains three significant wave height parameters, *swh*, *shww* and *shts* for the total wave, wind wave partition and total swell partition respectively. The data set shows that *swh* extremes are higher than the extremes of *shts* and *shww*, which makes sense considering their relation described in Eq. 4.2. Thus ***swh* is chosen as the dominant variable in this analysis**. The variable *shww* can be calculated with the same equation if *shts* and *swh* are known. It is noted that choosing a different dominant variable is not necessarily wrong, but that the choice depends on the application. For instance, if an offshore vessel is sensitive to very large mean wave periods, but not so much the significant wave height, it might be better to pick the mean wave period as the dominant variable and determine the corresponding associated values for the other parameters.

A univariate extreme value analysis is performed on *swh*. Because of the length of the data set, a peak-over-threshold (POT) approach is preferred over a annual maxima approach. As stated in the Literature study, choosing a threshold is not a trivial step. The choice of a threshold is a trade off between bias and variance. A threshold that is too high might generate too little data excesses, leading to a high variance. A threshold that is too low might violate the asymptotic behaviour of the model, which results in bias. As a general rule of thumb, the extreme data set should have about 5-10 peaks per year, which results in an extreme data set of about 200 to 400 peaks for this analysis. The extremes are at least two days apart to ensure independence. Furthermore, the resulting shape and scale parameters should be more or less stable if a slightly different threshold is chosen. With the exception of the most volatile locations, offshore significant wave height tends to show either type I or type III behaviour. This means that the shape parameter ξ is smaller or equal to zero.



(a) (b)

Figure 5.2: Peaks over threshold shown as red stars and threshold stability graphs of shape parameter (ξ), scale parameter (σ) and 100 year return value for *swh*.

The POT analysis is performed using the ORCA toolbox (Caires and van Os, 2012) in Matlab. Considering the previous statements, a threshold is of 2.4 m is chosen for *swh* (Figure 5.2b). This results in 310 peaks, shown as red stars in Figure 5.2a. A GPD is fitted through the extreme values using a probability weighted moments (PWM) method, which is recommended over Maximum Likelihood Estimate (MLE) for smaller data sets (Hosking and Wallis, 1987). This fitted GPD is extrapolated to a return period of 500 years in Figure 5.3. The 95% confidence interval bounds, represented by dashed lines in the same figure, are calculated using a bootstrap method.

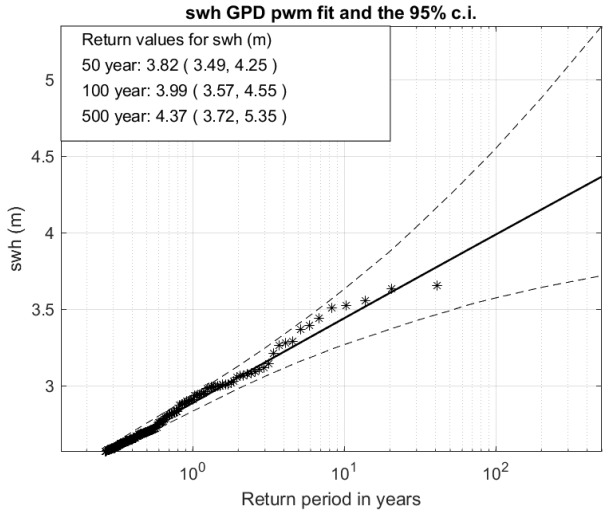


Figure 5.3: Extrapolated fitted GPD together with the 95% confidence interval for swh.

5.2.2 Other variables

Even though swh is chosen as the dominant variable, univariate extreme value analysis is also performed on other wave parameters, in order to compare this with the results from the bivariate and multivariate analyses later on. The return values for return periods of 50, 100 and 500 years are calculated for the variables in the same way as for swh and the results are shown in Table 5.1. Plots of the peaks, threshold stability and return values for these variables are shown in Appendix B.

Variable	**Unit**	**Return values**		
		50 yrs	100 yrs	500 yrs
swh	m	3.82	3.99	4.37
$mp1$	s	9.15	9.40	9.96
$shww$	m	3.60	3.75	4.09
$p1ww$	s	6.60	6.79	7.22
$shts$	m	2.97	3.11	3.42
$p1ps$	s	9.61	9.79	10.20

Table 5.1: Extrapolated GPDs give the return values for univariate POT analysis.

5.3 Extreme data set

At each peak value for the dominant variable *swh*, the values of all other variables at the time of the sampled peaks are stored. An alternative way of stating this is to say that the other variables are *concomitant* to the dominant variable. This results in an extreme data set consisting of 310 values for all parameters listed in Table 4.1. First, the dependencies between variables of the extreme data set are examined. Next, deterministic values for the mean wave direction, directional spreading and peak enhancement factor are derived from the extreme data set or assumed.

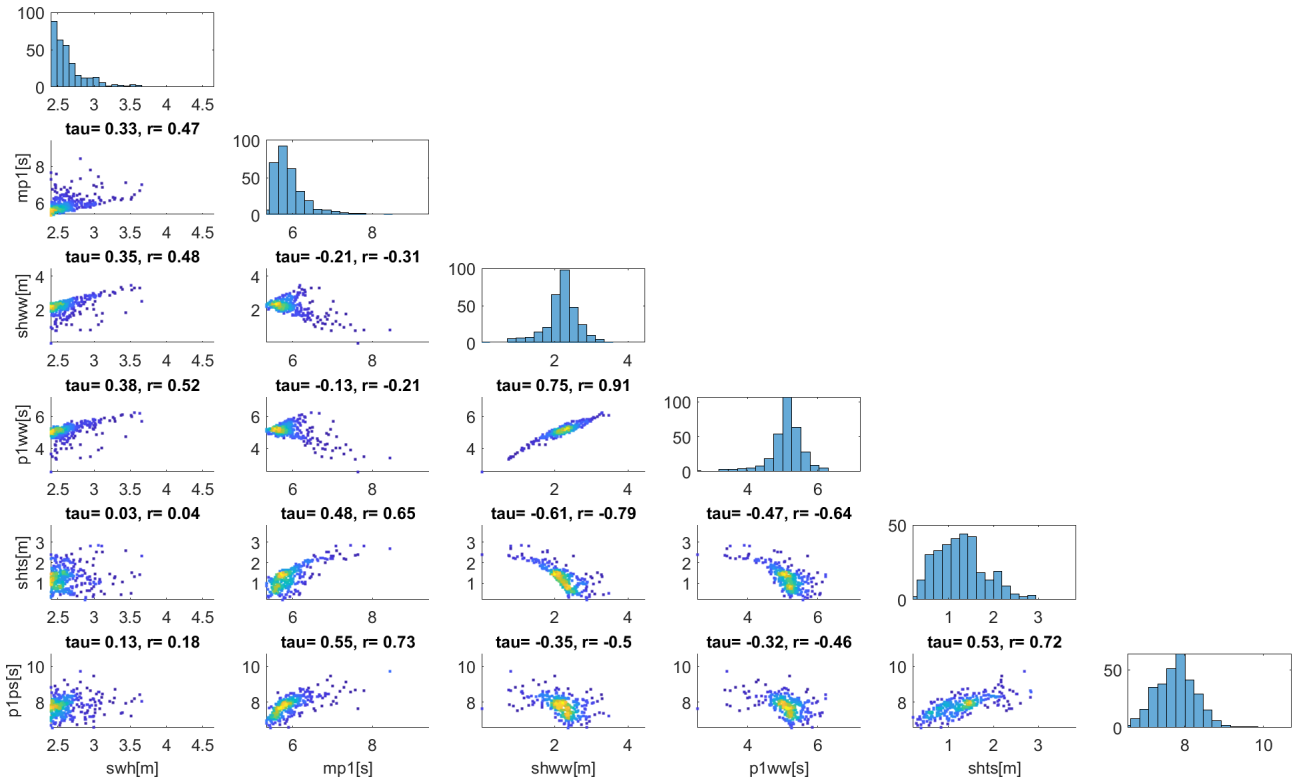


Figure 5.4: Density scatter plots and occurrence histograms of the extreme data set for the significant wave height and mean wave period of the total, wind wave partition and swell partition. Wind wave and swell are within the directions of interest shown in Figure 5.1. Kendall's τ and Spearman's r are calculated for each combination of variables and shown in the title of the respective sub figures.

The correlations of the entire data set of Figure 4.6 are compared to the extreme data set of Figure 5.4. It is noted that under normal conditions the total wave has a high swell component. However, extreme values in the considered directional sectors of interest for wind waves and swell (Figure 5.1) of *swh* seem to be more dependent on wind waves than on swell.

Different correlations are observed between the variables in the two data sets. Although slightly less correlated compared to the entire data set, *shww* and *p1ww* are still strongly correlated in the extreme data set. A steepness curve is used to approximate the relation between *shww* and *p1ww*, thus assuming the two variables are fully dependent. By assuming full dependence, the multivariate analysis is simplified, as will be explained in later chapters.

Next, the wind wave and swell parameters are considered. A general assumption is that wind waves and swell are independent as they are generated in different locations. As stated previously, under normal conditions the significant wave height of wind waves (*shww*) and swell (*shts*) show a weak negative correlation of $r = -0.22$, which is almost negligible. However, for the considered extreme events a strong negative correlation ($r = -0.79$) between *shww* and *shts* is observed. This negative correlation seems to be in disagreement with the general assumption of independence between wind waves and swell and thus seems counter-intuitive at first glance.

However, it is important to understand the connection between the total wave, wind waves and swell as described by Eq. 4.2. This relation together with the choice of the dominant variable results in a strong negative correlation between *shww* and *shts* for extreme events. Figure 5.5 shows *shww* versus *shts* under normal conditions and for the extremes.

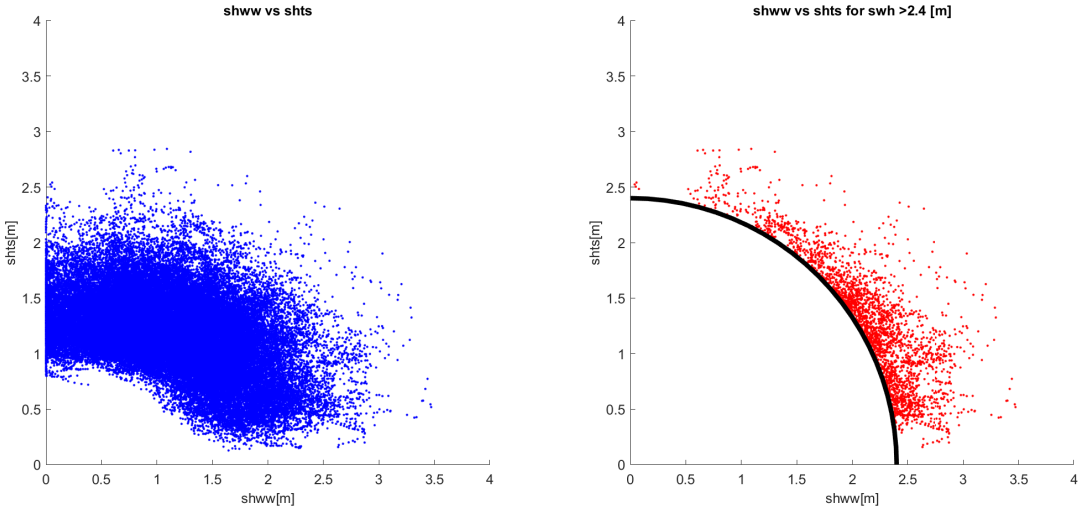


Figure 5.5: Significant wave height of wind waves (*shww*) versus swell (*shts*). The entire time series is shown as blue dots and the extremes are shown as red dots in the right figure. The black circle in the right figure represents a constant value of $swh = 2.4m$.

5.3.1 Fitting marginal distributions

Each peak event in the extreme data set consists of one hourly value of the dominant variable and associated values for the non-dominant variables. These non-dominant variables are often referred to as concomitant variables. Distributions are fitted to the concomitant variables $mp1$, *shww*, $p1ww$, *shts* and $p1ps$ in order to model the data. Distributions are fitted that most accurately represent the data in the upper tails of the distributions. The fitting process and considerations are explained in Appendix B. The result of the fitting process is shown in Figure 5.6.

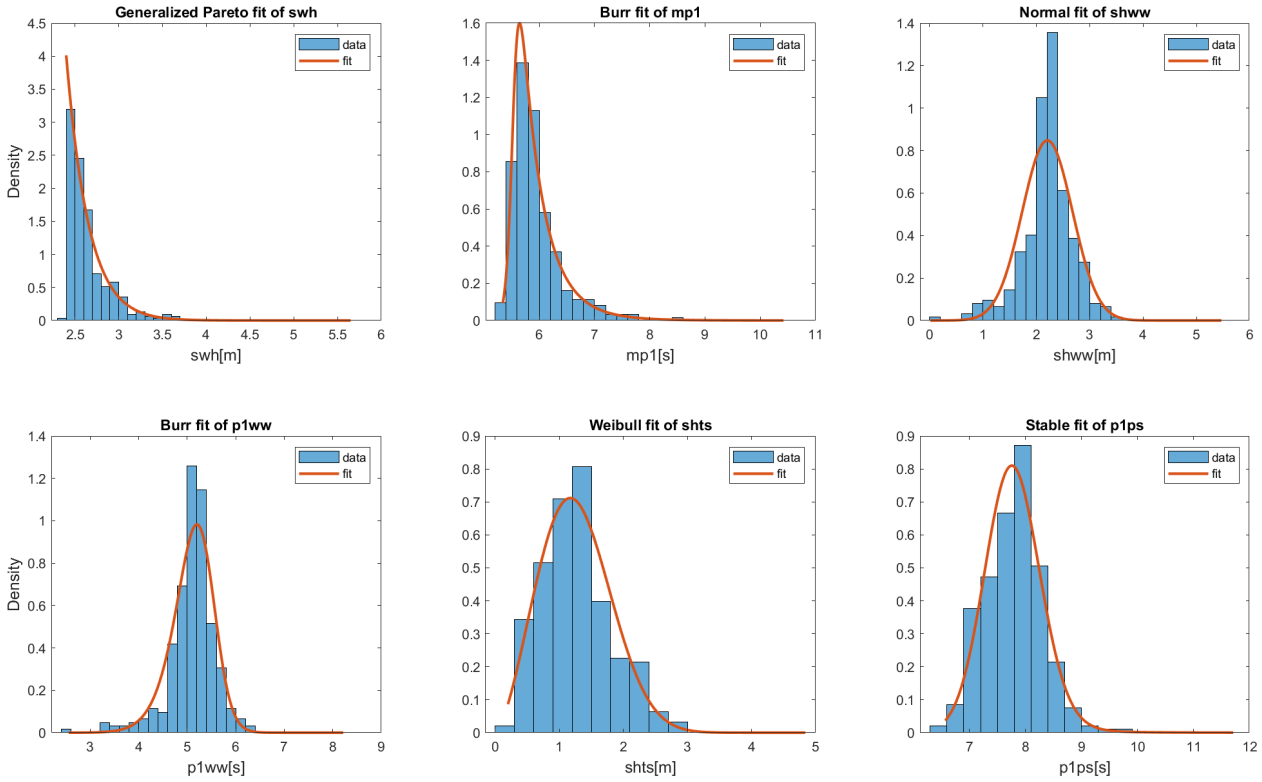


Figure 5.6: The result of the fitting process is a suitable probability density function for the concomitant variables. Note that the GP fit of the variable *swh* (top left figure) is the same as in Figure 5.3, except that the function is shown as a PDF.

5.3.2 Mean wave direction

Next, the mean wave direction of the total wave, wind wave partition and swell partition is assumed for higher return periods using the average directions in the extreme data set, shown in Figure 5.7.

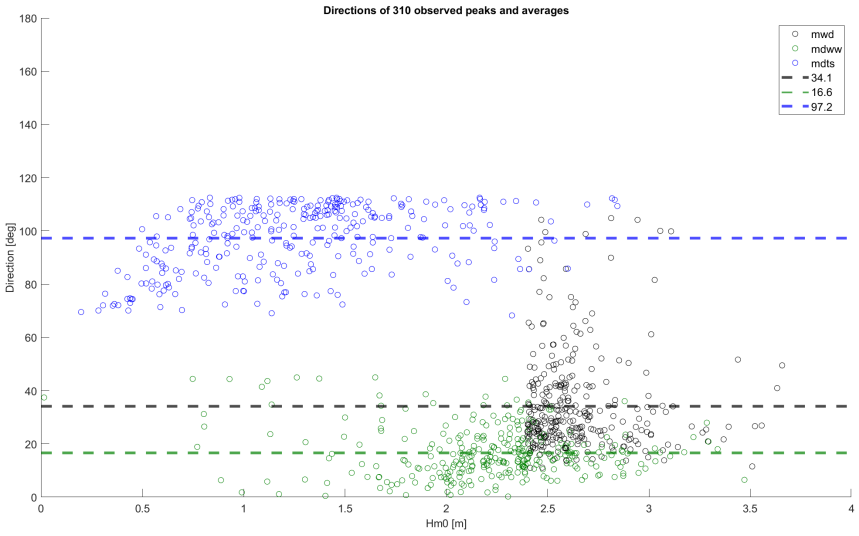
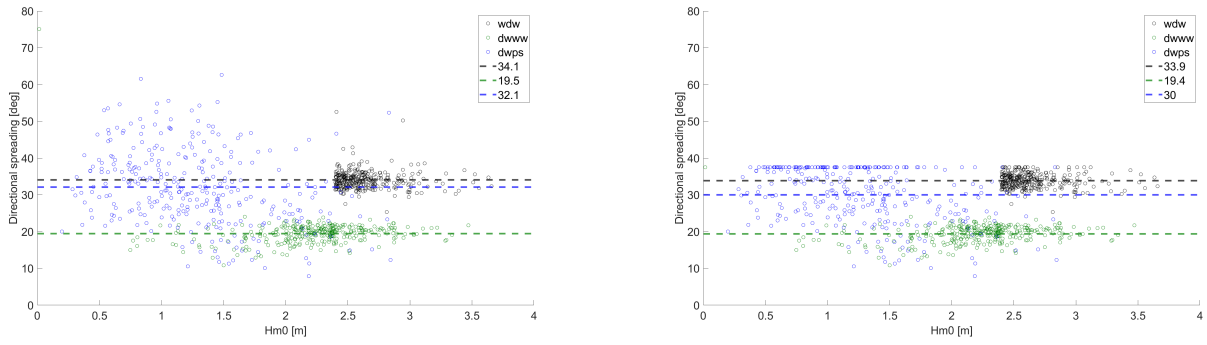


Figure 5.7: The significant wave height versus the mean wave direction of the total wave (*mwd*), wind wave partition (*mdww*) and swell partition (*mdts*) of the extreme data set. The average directions are represented by dashed lines with the average values shown in degrees in the legend.

It is assumed that the direction of the total wave, wind wave- and swell partition for higher return periods is equal to the average directions of those partitions of the extreme data set. This assumption reduces the number of variables by three by assuming deterministic values for the directions. The assumed direction remains the same for all higher return periods.

5.3.3 Directional spreading

Hourly values of the directional spreading as defined by Kuik et al. (1988) are available in the data set for the total wave, wind wave- and swell partition named *wdw*, *dwww* and *dwps* respectively. Because SWAN can only take in a directional spreading up to 37.5° as input, all values are capped to this value, which is shown in 5.8. It is assumed that capping the data has a negligible effect on the overall outcome.



(a) (b)

Figure 5.8: Directional spreading of the total wave (*wdw*), wind wave partition (*dwww*) and swell partition (*dwps*) of the extreme data set are shown in (a). Figure (b) shows the same data, except that all values are capped at 37.5°. The average directional spreading values are represented by dashed lines and shown in degrees in the legend.

In order to reduce the number of variables in later analyses, a deterministic value is assumed for the directional spreading of the total wave. The directional spreading is assumed as the average directional spreading of in the (capped) extreme data set. This means that the directional spreading of the total wave is assumed to be 33.9°, even for higher return periods. Note that this approach is similar to the assumptions made with respect to the mean wave directions.

The same could be done for the wind wave- and swell partitions. However, making a similar assumption in this regard would be problematic. This is because in general swell systems have a lower value for the directional spreading than wind wave systems, usually in the order of 20°. A study by Ewans (2002) shows that for a different data set the directional spreading of swell tends towards this value for extreme events. Moreover, a recent paper by Smit and Janssen (2019) states that the directional spreading of swell is a balance between geometrical and diffusive spreading, resulting in a stable spectral shape and a corresponding directional spreading of about 15°. Because a directional spreading 30° is considered unrealistic for swell, a value of 15° is assumed from now on. A possible explanation for the lack of convergence for extreme events in the ERA5 data set is that the total swell data consists of multiple crossing swell systems, that result in higher values for the directional spreading.

For the wind wave partition the capped data seems to suggest a value of 19.4° is appropriate for the directional spreading. However, this value is considered too low for wind waves. A widely used value for the directional spreading of wind waves is 30°, which is used instead of 19.4 °. This is also in agreement with observations in a different data set, as shown in Figure 5.9. The unrealistic values for the directional spreading of the wind wave partition in the ERA5 data set may again be an artefact of the partitioning algorithm.

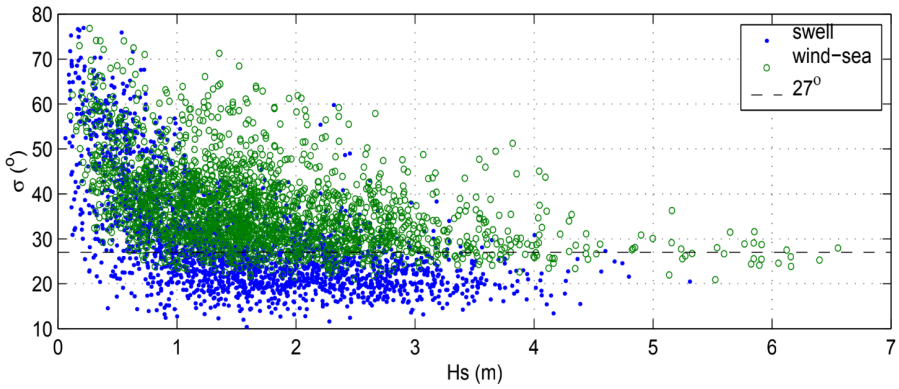


Figure 5.9: Significant wave height versus directional spreading of a different data set (Maui platform off the coast of New Zealand). From Ewans (2002).

5.3.4 Peak enhancement factor

The peak enhancement factor (γ) is determined for the total wave, wind wave- and swell partition. In the ERA5 reanalysis data set a variable is available called the *wave spectral peakedness* or *wsp*, a dimensionless variable between 1 and 10. It assumed that this is the peak enhancement factor γ for the total wave.

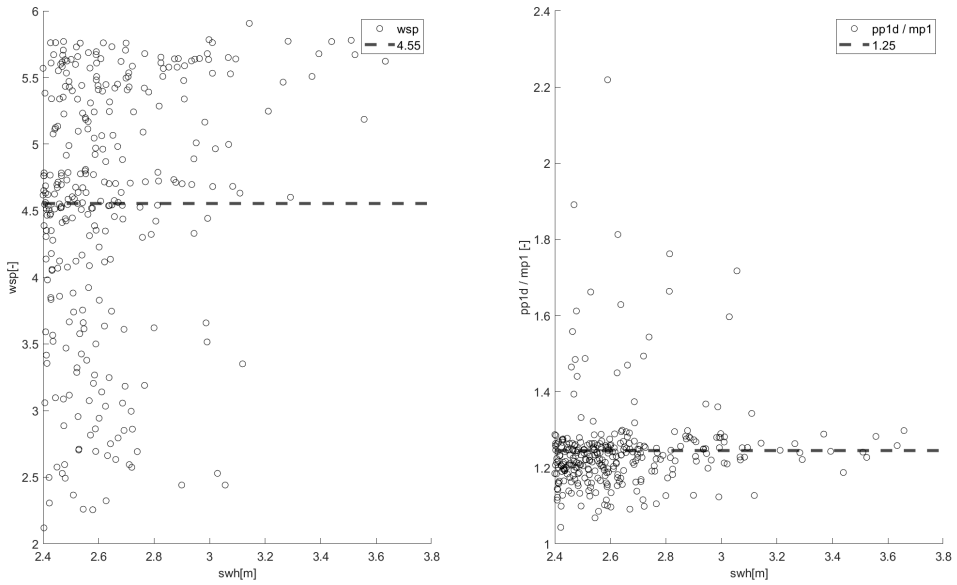


Figure 5.10: The total significant wave height (*swh*) versus the peak enhancement factor of the total wave (*wsp*) of the extreme data set plotted in the left figure. The right figure shows *swh* versus the proportion of the peak period and mean wave period of the total wave.

The average value of *wsp* in the extreme data set is 4.55, which is higher than one might expect for a fully developed sea state. It is unclear what causes the high values for *wsp* in the data set. The result of *wsp* is compared to the right of Figure 5.10, which shows the proportion $pp1d/mp1$. A proportion of 1.25 means that $mp1$ / $pp1d = 0.8$. An empirical relation between this fraction and γ is provided by DNV (2012) and shown in Eq. 5.2.

$$\frac{T_{m0,1}}{T_p} = 0.7303 + 0.04936\gamma - 0.006556\gamma^2 + 0.0003610\gamma^3 \qquad \text{(Eq. 5.2)}$$

This results in $\gamma = 1.8[-]$ for the total wave, which seems to be a more realistic value than the 4.55 that results from the average *wsp*. The former value is thus assumed to be the same for all higher return periods.

Next the peak enhancement factor is determined for the wind wave and swell partition. Unfortunately no value of T_p is available in the ERA5 data set for the wind wave and swell partition, so γ has to be derived with other parameters. The empirical relation proposed by Isherwood (1987) can be used to find a value for the peak enhancement factor. However, the results of this analysis are invalid for the swell partition and unrealistic for the wind wave partition. The results are therefore not used in this thesis, as is explained in Appendix B.4.

To find satisfactory values, Appendix G of Dingemans (1987) is used instead. Similarly to the directional spreading, the peak enhancement factor is assumed to be equal to the average of the extreme observations. This results in $\gamma_{ww} = 1.0$ and $\gamma_s = 1.7$, as shown in Figure 5.11.

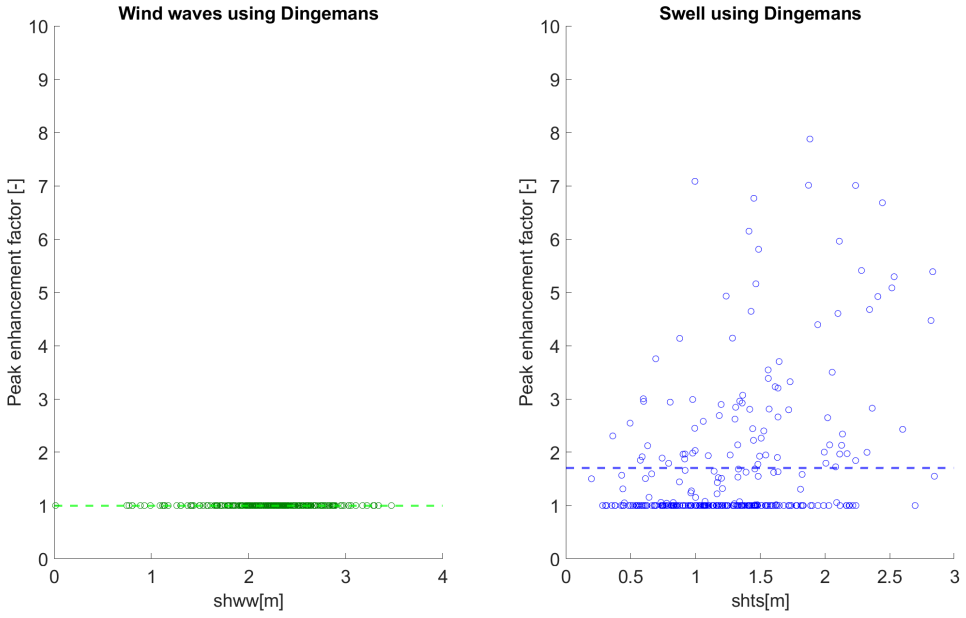


Figure 5.11: The peak enhancement factor of the wind wave (left) and swell (right) partitions. The average peak enhancement factor of the extreme data set is indicated with a horizontal dashed line.

The resulting values of Figure 5.11 are within an acceptable range, although the peak enhancement factor for swell is rather low. It is suspected that low values for the peak enhancement factor for the total swell are the result of multiple (crossing) swell systems. If two or more swell systems are treated as a single system, the value of the peak enhancement factor of these systems combined is lower than for the individual systems. The results for the peak enhancement factor of the wind waves are all the same value ($\gamma_{ww} = 1.0$). This is because the empirical relation is applied to the extreme data set. Every extreme data point has a very high wind wave component, resulting in fully developed sea states for the wind wave partition and thus corresponding values for γ_{ww} equal to 1.0.

5.4 Concluding remarks

- A combination of wind wave and swell directions is chosen. The combination was chosen in line with the goals of this research and does not contain the highest total wave events. The analysis could be repeated for all possible combinations if the procedure is automated in the future.
- The total significant wave height is chosen as the dominant variable. A POT analysis is performed on the dominant variable. Extrapolating the fitted GPD results in return values for the extreme variable for a number of return periods. It is noted that choosing a different variable as extreme is not necessarily wrong,

but that the choice depends on the application. Distributions are fitted to the dominant variable and to the concomitant variables $mp1$, $shww$, $p1ww$, $shts$ and $p1ps$.

- The mean wave direction for the total wave, wind wave- and swell partition is assumed as a deterministic value by calculating the average direction in the extreme data set for each partition.
- For the total wave the directional spreading is assumed as the average value in the (capped) extreme data set. The same approach yields unsatisfactory results for the wind wave- and swell partitions. Instead, more realistic values for directional spreading of the partitions are assumed. A possible explanation for the unrealistic result for swell might be that the parameters in the extreme data set belong to the *total* swell partition. Using the total swell partition in the analysis disregards the presence of multiple crossing swell systems during certain extreme events.
- The peak enhancement factor for the total wave is available in the ERA5 time series. However, the extreme data set does not seem to converge to realistic values. Therefore, the peak enhancement factor for the total wave is calculated for each extreme event using an empirical relation DNV (2012). For the single-peaked wave system the peak enhancement factor is assumed as a deterministic value. This assumption is within the theoretical limits and is assumed to be a the value for the peak enhancement factor for higher return values.

  The peak enhancement factor is not directly available in the ERA5 time series for the wind wave- and swell partitions. Moreover, the peak wave period (T_p) is also not available, so a different approach is required for the partitions. Peak enhancement factors are calculated using the steepness relation of Isherwood (1987) and Appendix G in the report of Dingemans (1987). It is concluded the empirical relation of Isherwood does not provide satisfactory values for the peak enhancement factor for the wind wave- swell partition. Therefore the resulting values from the report of Dingemans are used.
- As stated, the main goal of this thesis is to build a single-peaked 2D wave spectrum based on the total wave parameters and compare it an equivalent double-peaked spectrum generated with parameters from the wind wave and swell partitions. Not all values for the parameters have been determined yet, as they are the result of the bivariate or multivariate analyses of later chapters. The results thus far are summarised in Table 5.2 and 5.3.

ERA5	**Symbol**	**Unit**	**Design realisation**		
swh	H_{m0}	m	cond. swh=3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
mwd	θ	°	34.1	34.1	34.1
$mp1$	$T_{m0,1}$	s	-Chapter 6-	-Chapter 6-	-Chapter 6-
wdw	σ_θ	°	33.9	33.9	33.9
-	γ	-	1.8	1.8	1.8

Table 5.2: Single peaked 2D spectrum parameters so far. Note that 50, 100 and 500 years is only the univariate return period of swh and not the joint return period.

ERA5	**Symbol**	**Unit**	**Design realisation**		
swh	H_{m0}	m	cond. swh=3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
shww	H_{m0}	m	-Chapter 7-	-Chapter 7-	-Chapter 7-
mdww	θ	°	16.6	16.6	16.6
p1ww	$T_{m0,1}$	s	-Chapter 7-	-Chapter 7-	-Chapter 7-
dwww	σ_θ	°	30	30	30
-	γ_{ww}	-	1.0	1.0	1.0
shts	H_{m0}	m	-Chapter 7-	-Chapter 7-	-Chapter 7-
mdts	θ	°	97.2	97.2	97.2
p1ps	$T_{m0,1}$	s	-Chapter 7-	-Chapter 7-	-Chapter 7-
dwps	σ_θ	°	15	15	15
-	γ_s	-	1.7	1.7	1.7

Table 5.3: Double peaked 2D spectrum parameters so far. Note that 50, 100 and 500 years is only the univariate return period of *swh* and not the joint return period.

6 Bivariate analysis

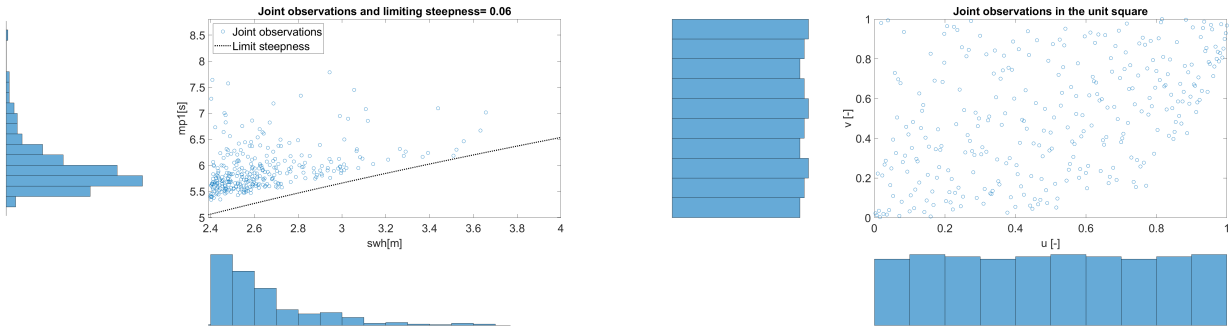
The goal of this chapter is to obtain the associated values for the total mean wave period conditional on the return values of the total significant wave height. The associated return values for the total mean wave period are determined using a copula-based approach, which is adapted to account for the limiting steepness. It is important to underscore that the bivariate analysis is performed separately from the multivariate analysis and only requires input from Chapter 5. For a complete oversight, including the other steps in the analysis, the reader is referred to Figure 2.1.

6.1 Bivariate analysis using a copula

Data show that joint observations of the significant wave height versus the mean wave period do not occur outside the limiting steepness, an observation which is in agreement with theory (Holthuijsen, 2007). This is accounted for in the approach by truncating simulations that are outside the limit steepness. The steps in the applied method are described:

1. Determine which copula best models the dependence between the two variables. From here there are multiple methods to find design values. Because of the assumption that the dominant variable *swh* is more important for the final design than the concomitant variable *mp*1, the so-called *conditional design realisation* is preferred over the *most-likely design realisation*, as described by Salvadori et al. (2014). It is stressed that a different design realisation might be preferred depending on the application. Selecting the appropriate design realisation is a choice, as there is no unique way to go from risk to design values.
2. Simulate a large number of samples from the copula of choice. A sample in the order of $1 \cdot 10^8$ simulations seems to be appropriate for the investigated univariate return values.
3. Convert the sample from the unit square to physical units using the inverse cumulative distribution of the two fitted marginal distributions.
4. Truncate simulations that are outside the limit steepness.
5. Round all simulations to two decimals, this ensures that simulations can have the exact value of the univariate return value of the dominant variable.
6. Count the number of simulations of the concomitant variable conditional to a univariate return value of the dominant variable.
7. Decide which value to choose as the associated value for the concomitant variable from resulting occurrence histogram. In this approach the mode of the resulting histogram is chosen.

The first step in the approach is to determine the appropriate bivariate copula. The joint observations are transformed to the unit square (Figure 6.1b) and a number of copulas are fitted. The goal in this step is to determine which copula best describes the dependence pattern of the data, hence the conversion to the unit square. The selection is based on three model adequacy or *goodness-of-fit* tests, described in Section 3.5 of the literature study.



(a) (b)

Figure 6.1: Joint observations with marginal distributions and in the unit square

The results of the model adequacy tests are shown in Table 6.1. Plots with the semi-correlations and the Cramér von Mises statistic are shown in Appendix C.

Copula family	**Parameter(s)**	**Av. in Matlab**	**Model adequacy**		
			ρ_{NE}	CvM	AIC
Gaussian	$\theta = 0.47$	Yes	0.11	0.335	-79.73
t	$\theta = 0.47$, $\delta = 2$	Yes	0.47	0.335	-82.65
Clayton	$\theta = 0.97$	Yes	0.15	0.555	-73.89
Gumbel	$\theta = 1.49$	Yes	0.54	0.446	-80.66
Frank	$\theta = 3.17$	Yes	0.28	**0.328**	-74.61
Survival Tawn type 1	$\theta = 2.03$, $\delta = 0.46$	No	**0.20**	-	**-91.13**

Table 6.1: Three model adequacy tests for six different copula families. The semi-correlation is shown for the north east quadrant, which is $\rho_{NE} = 0.22$ for the joint observations. The best model according to each test is shown in bold. The Cramér von Mises (CvM) statistic is not available for the Survival Tawn type 1 copula.

Out of the copula families that were examined, it seems that the Survival Tawn Type 1 copula is best suited to model the dependence for the joint observations. Next, a (small) sample is obtained from the copula and compared to the joint observations.

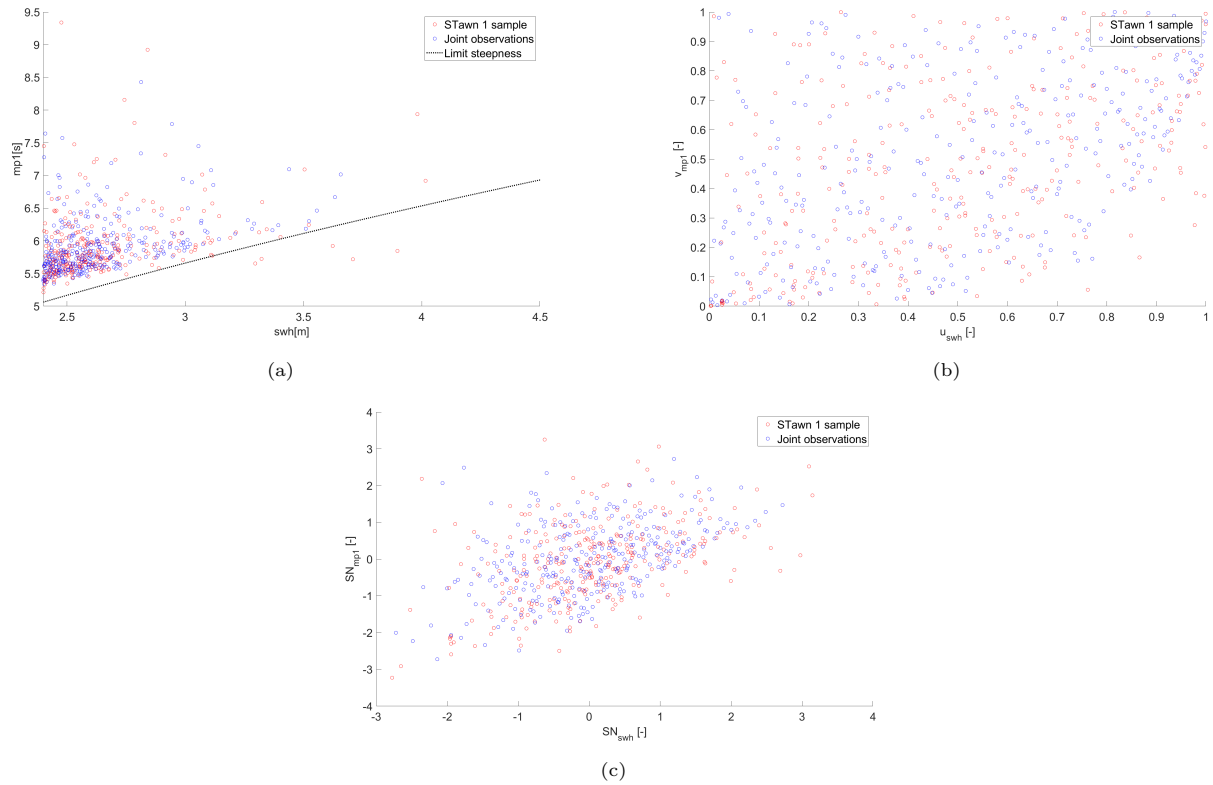


Figure 6.2: Joint observations and a sample from the survival Tawn type 1 copula transformed to physical units (a), the unit square (b) and in the standard normal space (c).

Although the model adequacy tests suggest that the survival Tawn type 1 copula is the best model for the data out of the copulas that were examined, it seems that the copula does not preserve the limiting steepness well enough, as some simulations of the sample are outside the limit steepness and are considered physically impossible. This is shown in Figure 6.2. To overcome this problem, simulations that are outside the limiting steepness are truncated. The limit steepness truncation discards about 2% of the generated samples. Next, in order to determine the associated values for the mean wave period a very large sample ($1 \cdot 10^8$ simulations) is generated and truncated accordingly.

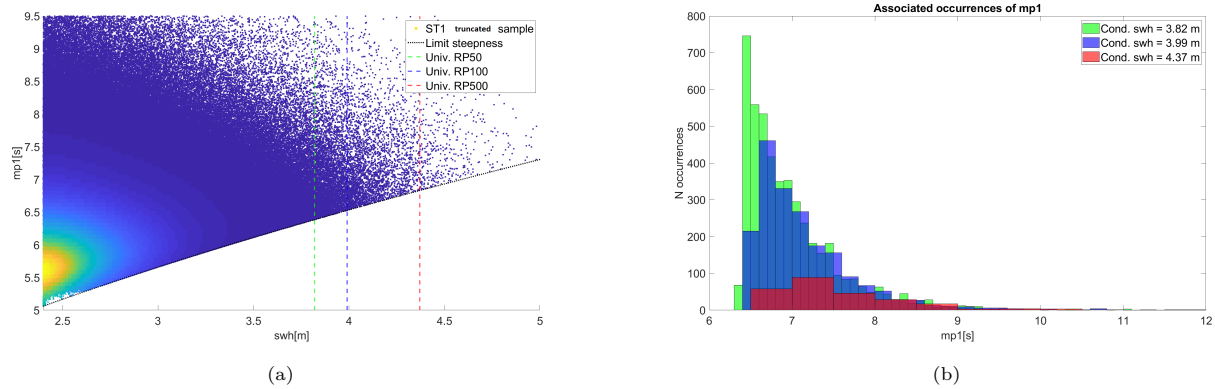


Figure 6.3: A density scatter plot of a very large truncated sample with the univariate return values for the significant wave height (a). The occurrence histograms with the associated values of the mean wave period are shown in (b) conditional on three return values of the significant wave height.

The resulting associated occurrence histograms of Figure 6.3 are asymmetric, the shape entails that the median value is slightly higher than the mode (the point with the highest number of occurrences). It is assumed that the

point with the highest number of occurrences is also the point with the highest probability density. These values are shown, together with their corresponding steepness ($s = \frac{2\pi H_{m0}}{g \cdot {T_{m0,1}}^2}$), in Table 6.2.

ERA5	**Symbol**	**Unit**	**Design realisations**		
swh	H_{m0}	m	cond. swh=3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
*mp*1	$T_{m0,1}$	s	6.45	6.70	7.25
	s	-	0.059	0.057	0.053

Table 6.2: Associated values of *mp*1 conditional on three univariate return values of *swh*, together with the corresponding steepness s. Note that 50, 100 and 500 years is only the univariate return period of *swh* and not the joint return period.

Table 6.2 shows that all design realisations are within the steepness limit as all values of s are below the limit of $s = 0.06$. Furthermore the resulting steepness is different for each design realisation.

An alternative approach where the associated values of *mp*1 are determined by means of the copula density $c(u, v)$ is provided in Appendix C.2. This alternative approach provides a motivation for truncating the simulations outside the limit steepness, as the the alternative approach results in associated values of the mean wave period that are physically impossible. Additionally, a non-parametric method (with some additional steps referred to as the ORCA approach) is applied on the time series for comparison with the copula-based approach. Compared to the copula-based truncation approach the ORCA approach results in higher associated values for the mean wave period. However, because of the visual mismatch between its joint PDF and joint observations it is concluded that the design realisations provided by the ORCA approach are too conservative (e.g. the values for the mean wave period are too high). The design realisation provided by the ORCA approach are therefore not used in this thesis. Further differences between these two methods are discussed in Appendix C.4.

6.2 Concluding remarks

- Simulated extremes that are outside the realistic steepness limit are truncated in the bivariate copula-based approach. This means that strictly speaking the method is no longer fully parametric, but could be considered semi-parametric. However the resulting associated values for the total mean wave period seem to be more realistic when the truncation is applied to preserve the limiting steepness. A different approach to solve the problem with the limiting steepness is by applying a copula that models the steepness even more sharply than the Tawn copula, such as the *skew-t* copula, which was applied, among other models, by Jaeger and Morales-Nápoles (2017). A downside of using the skew-t copula is that a closed form of the copula is not available and has to be approximated numerically.

- A downside of the truncation approach applied in this chapter is that it only provides the associated value for the mean wave period given a univariate return value of the total significant wave height (referred to as the *conditional design realisation*). A joint return period cannot be extracted directly and has to be approximated through sampling, which is shown in Appendix C.3. This is because the approach is no longer fully parametric. This could be problematic for applications where the *most-likely design realisation* is preferred and joint exceedance probability is required exactly.

  An alternative approach to find the appropriate combination might be to select a number of combinations with equal joint return period and run all of these through the same model to find out which combination has the most detrimental effect. For the bivariate example of this chapter this would mean finding a number of combinations (say a couple dozen) of H_{m0} and $T_{m0,1}$ all with a same joint return period of say 100 years. These points are thus all located on the same iso-line. Next one could investigate which combination is least favourable to the stability of a structure. It is noted that some combinations have a higher probability density than others, which could to be taken into account by means of some sort of weighing factor.

7 Multivariate analysis

The values needed to construct a single-peaked 2D wave spectrum are now determined. The goal of this chapter is to find the associated values for significant wave height and mean wave period of wind waves and swell conditional on univariate return values of the total significant wave height. These values could be used to generate an equivalent double-peaked 2D wave spectrum, which could be compared to the single-peaked spectrum. It is important to underscore that the multivariate analysis is performed separately from the bivariate analysis and only requires input from Chapter 5. For a complete oversight, including the other steps in the analysis, the reader is referred to Figure 2.1.

7.1 Multivariate analysis using vine copulas

The multivariate analysis consists of a number of steps, which are described next.

1. Two simplifications are applied in the multivariate analysis. First, the mean wave period and significant wave height of the wind wave partition are assumed to be fully dependent. In reality the dependence is very high, but the two parameters are not fully dependent. The advantage of this assumption is that one variable does not have to be considered as it calculated using the other. A one-to-one steepness relation is used in the model.

2. The second simplification results from the energy relation of Eq. 7.2, relating the significant wave height of the total, wind wave- and swell partition. If two significant wave heights are known, the third is calculated using this relation. This results in two options where either the significant wave height of the wind waves or swell drops out of the analysis. The total significant wave height does not drop out, as it is the dominant variable.

3. For the two options of the previous step, the structure of all possible regular vines is determined. The number of regular vines on three nodes is equal to three, thus six regular vines are considered in this analysis. Irregular vines are not considered.

4. A vine copula model is chosen for each option based on a number of model adequacy tests.

5. The design realisations of the two models of choice are computed. For three dimensions, the calculation of the design realisation is analogous to the bivariate analysis. With the vine copula and the marginal distributions, a joint PDF on three variables is constructed. This joint PDF is conditionalised on four univariate return values of the dominant variable. Conditionalising on a number of return values of the dominant variable reduces the dimension of the joint PDF by one, thus resulting in a joint PDF of two (concomitant) variables. The point with the highest joint density is chosen as the design realisation of the two concomitant variables.

6. Using the values obtained in the design realisation, the values of the other parameters are calculated, by using the relations of step 1 and 2. The output of both vine models is compared in a number of ways.

7.2 Model simplifications

Two model simplifications are applied in order to reduce the number of variables from five to three. First, the relation between wind wave parameters $shww$ and $p1ww$ is approximated using a steepness relation of Eq. 7.1, shown as a black dotted line in Figure 7.1.

$$\mathrm{p1ww} = \sqrt{2\pi \cdot \mathrm{shww}/(sg)} \qquad \text{(Eq. 7.1)}$$

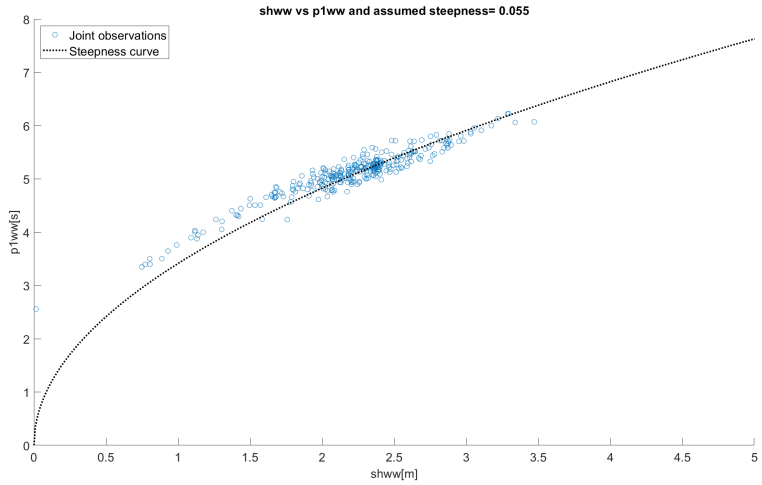


Figure 7.1: In the extreme data set the correlation between $shww$ and $p1ww$ is $r = 0.91$ or $\tau = 0.75$. To simplify the analysis, it is assumed that $shww$ and $p1ww$ are fully dependent with $r = 1$ and that all points are on the steepness line plotted as a black line. A steepness value of $s = 0.055$ is chosen.

It is noted that because of the steepness assumption made in Figure 7.1, every point of the wind wave partition satisfies the steepness limit of $s < 0.06[-]$. The swell- and total wave observations also satisfy this physical limit, as is demonstrated in Appendix B.2.

The second simplification uses the energy relation of the different partitions. Recall that the total energy of the partitions is conserved, meaning that adding up the energy of the wind wave- and swell partition results in the total wave energy of each sea state. The significant wave height squared is linearly related to the wave energy, thus Eq. 7.2 is used to calculate the significant wave height of the wind wave partition ($shww$) and swell ($shts$) if either is known.

$$\text{shww} = \sqrt{\text{swh}^2 - \text{shts}^2} \qquad \text{or} \qquad \text{shts} = \sqrt{\text{swh}^2 - \text{shww}^2} \tag{Eq. 7.2}$$

Using the energy relation of Eq. 7.2 and assuming that the significant wave height and mean wave period of wind waves are fully dependent, the analysis is reduced to three variables. These are either:

- Option 1: swh, $shts$ and $p1ps$.

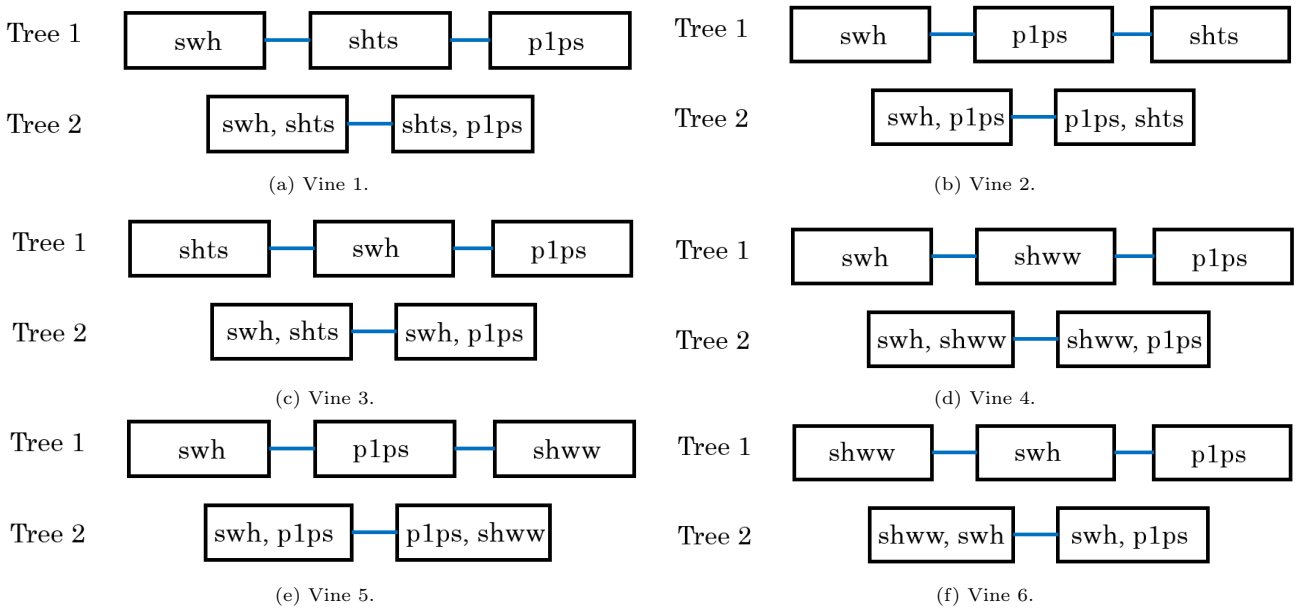
Or

- Option 2: swh, $shww$ and $p1ps$

A third option could be $shww$, $shts$ and $p1ps$. However, this option is not considered as it does not directly include the dominant variable swh. It is not known beforehand if the outcomes of option 1 and 2 are in agreement with each other, therefore both options are modelled. Using a vine copula on three nodes, the values for $shts$ and $p1ps$ conditional on swh are determined for option 1. Using a different vine copula on three nodes, the values for $shww$ and $p1ps$ conditional on swh are determined for option 2.

7.3 Possible tree structures

Both option 1 and 2 are modelled by a vine copula on three nodes. Thus, six regular vine copulas (three for each option) are fitted. The tree structures of these six models are shown in Figure 7.2.



(a) Vine 1. (b) Vine 2. (c) Vine 3. (d) Vine 4. (e) Vine 5. (f) Vine 6.

Figure 7.2: For three nodes, three regular vines are defined. Replacing node *shts* by *shww* results in an additional three regular vines for a total of six. The edges of each tree are marked in blue and represent a bivariate copula.

7.4 Choosing a model

Next, all vine copula models are fitted using the `VineCopula` package in `R`. For a detailed documentation of all functions used during the fitting process, the reader is referred to Schepsmeier et al. (2012). Recall that the edges of each tree are represented by (conditional) bivariate copulas. The bivariate copulas used (excluding rotated copulas) for the two vine copula models are shown in Table 3.2. After all regular vine copulas are fitted, a vine copula model is chosen for option 1 and 2 that best represents the dependence structure of the data. The model choice is based on the AIC and the CvM statistic, which are calculated for each fitted vine.

Vine #	**CvM**	**AIC**
Vine 1	0.5629	-244.89
Vine 2	0.5644	-240.02
Vine 3	0.7356	-238.83
Vine 4	1.4509	-385.02
Vine 5	1.0990	-346.73
Vine 6	1.6213	-381.50

Table 7.1: Model adequacy of the vines is determined by the AIC and the Cramér von Mises statistic.

Vine 1, 2 and 3 (belonging to option 1) are considered first. Based on the AIC and the CvM statistic, vine 1 best represents the dependence structure of the data. A summary of vine 1 is shown in Table 7.2.

Tree	**Edge**	**Copula**	**parameter 1 (θ)**	**parameter 2 (δ)**
1	shts, p1ps	Gaussian	0.72	-
	shts, swh	Tawn type 2	1.52	0.12
2	swh, p1ps; shts	Survival BB7	1.10	0.24

Table 7.2: Summary of vine 1.

The vine of choice according to the CvM statistic and the AIC for vine 4, 5 and 6 (belonging to option 2) are not in agreement, as is shown in Table 7.1. The CvM statistic suggests that vine 5 is the best model, while the

AIC suggests that vine 4 is to be preferred. A closer examination of the CvM statistic of vine 4 suggests that the error between the model and the data is lower for very large values of *swh* and *shww*, while for less extreme values the error is larger. Because the main goal of the model is to calculate very extreme design realisations, it is decided that vine 4 is to be preferred over vine 5 and vine 6. A summary of vine 4 is shown in Table 7.3. The different outcomes for the CvM statistic as well as additional model adequacy tests are shown in Appendix D.1.

Tree	**Edge**	**Copula**	**parameter 1 (θ)**	**parameter 2 (δ)**
1	shww, p1ps	BB8 rotated 90°	-2.76	-0.85
	shww, swh	Tawn type 2	2.62	0.55
2	swh, p1ps; shww	BB8	3.94	0.71

Table 7.3: Summary of vine 4.

If both vine 1 and vine 4 would perfectly fit the dependence structure of the data, the resulting design realisation would be the same. Next, the design realisations of the two vines are determined and compared to find out if this is the case.

7.5 Design realisations

Recall from Section 3.7 that two results can be extracted once a vine copula on three nodes has been fitted:

- A multidimensional sample in the unit cube (or hypercube for more than three dimensions) can be extracted from the vine copula.
- The vine copula density at any point in the unit cube can be calculated. These are the last three terms of Eq. 7.3.

The latter of these two results is used to together with the fitted marginal distributions shown in Figure 5.6 to construct the three-dimensional PDF $f(x_1, x_2, x_3)$:

$$\begin{aligned} f(x_1, x_2, x_3) = & f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ & \cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \\ & \cdot c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \end{aligned} \quad \text{(Eq. 7.3)}$$

In which:

$f_1(x_1), ..., f_3(x_3)$	marginal densities of each variable
c	bivariate (conditional) copula density, corresponding to the edges of the vine.
F	(conditional) cumulative probability of each variable

The method to obtain the design realisation for three dimensions is analogous the bivariate analysis of Chapter 6, namely the conditional design realisation described for the bivariate case by Salvadori et al. (2014). The design realisations of option 1 and 2 are determined. With the results from vine 1 and the marginal distributions $f(swh)$, $f(shts)$ and $f(p1ps)$ a joint PDF on three variables $f(swh, shts, p1ps)$ is constructed for option 1. Similarly for option 2, a joint PDF of three variables is constructed using the marginal distributions $f(swh)$, $f(shww)$ and $f(p1ps)$ and the (vine copula) density of vine 4.

The joint PDFs for option 1 and 2 are conditionalised on four univariate return values of the dominant variable (swh). These values correspond to univariate return periods of 10, 50, 100 and 500 years. Conditionalising on a number of return values of the dominant variable reduces the dimension of the joint PDF by one, thus resulting in a joint PDF of two (concomitant) variables. The point with the highest joint density is chosen as the design realisation of the two concomitant variables.

7.5.1 Option 1

Vine 1 results in a joint PDF of *swh*, *shts* and *p1ps*. Conditionalising on four univariate return values of *swh* results in four joint PDFs for *shts* versus *p1ps*, which are shown in Figure 7.3. Additionally, the (conditional) marginal densities are shown for each of the four PDFs. Of these marginal densities, the values of the 5th, 50th and 95th percentiles are indicated with dashed or dotted lines. The values of these percentiles are shown in Table D3 in Appendix D.

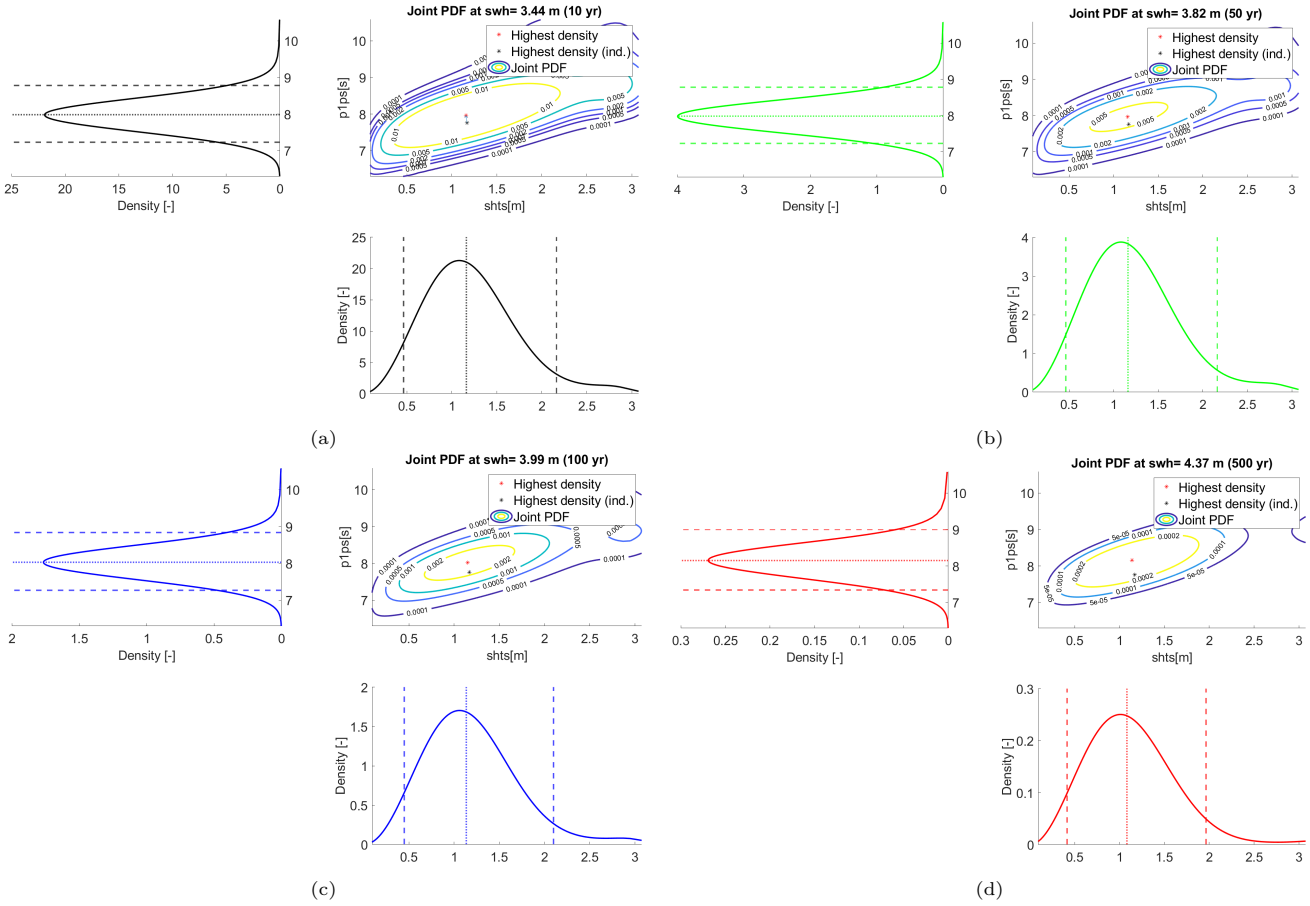
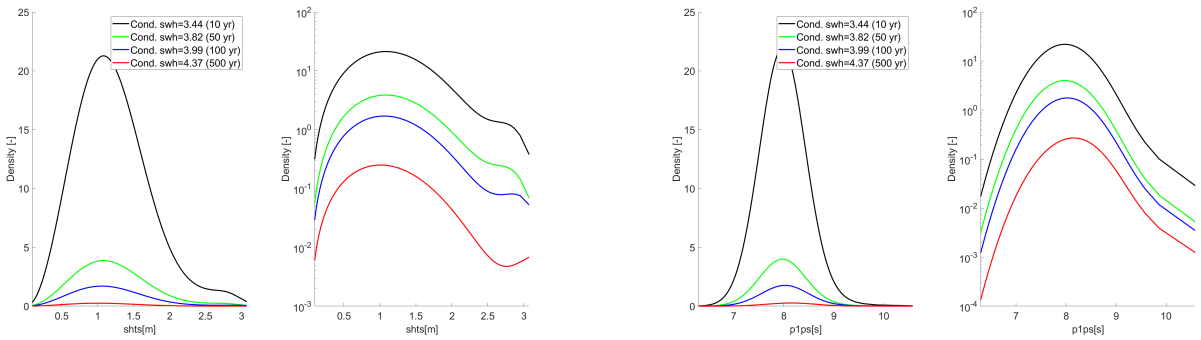


Figure 7.3: Results of vine 1. The 5th and 95th percentiles are marked with dashed lines in the marginal PDFs of each sub-figure, while the 50th percentile is marked with a dotted line.

The resulting joint PDFs of Figure 7.3 show that the point of highest density (represented by a red star) does not change very much for higher conditional values of the dominant variable. Furthermore, the point of highest density is also plotted in case *swh*, *shts* and *p1ps* would be independent. This point represented by a black star. The joint PDF in case of independence is merely a multiplication of the three marginal distributions; $f(swh)$, $f(shts)$ and $f(p1ps)$. Considering the proximity of the red star to the black star, it is concluded that assuming that *swh shts* and *p1ps* are all independent has a very similar result when only looking at the point of highest density. The biggest difference with the independent case is in the shape of the joint PDF in the *shts*, *p1p1*-plane. The result from vine 1 shows that the joint PDF has an oval type shape, indicating dependence between *shts* and *p1ps*. This shape is different from the independent case which has a circular shape in said plane.

Next, the marginal (conditional) densities of Figure 7.3 are considered, not te be confused with the fitted marginal distributions of Figure 5.6. The four marginal densities of *shts* are plotted in the same figure in order to compare the shapes. The same is done for four marginal densities of *p1ps*, as is shown in Figure 7.4. Contrary to Figure 7.3, the 5th, 50th and 95th percentiles are not shown in Figure 7.4.



(a) (b)

Figure 7.4: Vine 1 margins of *shts* (a) and *p1ps* (b) plotted in the same graph. Additionally the same margins are plotted using a log scale for the vertical axis to compare the shape of the curves.

The mode of the graphs (the point with the highest density) is practically the same for all four graphs of *shts* (Figure 7.4a) and increases slightly for *p1ps* (Figure 7.4b) given higher return values of *swh*. This should not come as a surprise given the low value of the correlation coefficient in the extreme data set between the pair *swh-shts* ($\tau = 0.03$) and the pair *swh-shts* ($\tau = 0.13$). The correlation coefficients of all pairs of variables in the extreme data set are shown in Figure 5.4.

The shape of the conditional densities of *shts* is slightly asymmetrical, while the conditional densities of *p1ps* are almost completely symmetrical. The former indicates that the considered extreme values of *swh* are mostly due to high wind waves and that the significant wave height of swell stays more or less constant for more extreme events. However, there is a small probability that extreme values of *swh* consists of a combination of a large swell and wind waves.

7.5.2 Option 2

Vine 4 results in a joint PDF of *swh*, *shts* and *p1ps*. Conditionalising on four univariate return values of *swh* results in four joint PDFs for *shww* versus *p1ps*, which are shown in Figure 7.5. Additionally, the (conditional) marginal densities are shown for each of the four PDFs. Of these marginal densities, the values of the 5th, 50th and 95th percentiles are indicated with dashed or dotted lines. The values of these percentiles are shown in Table D3 in Appendix D.

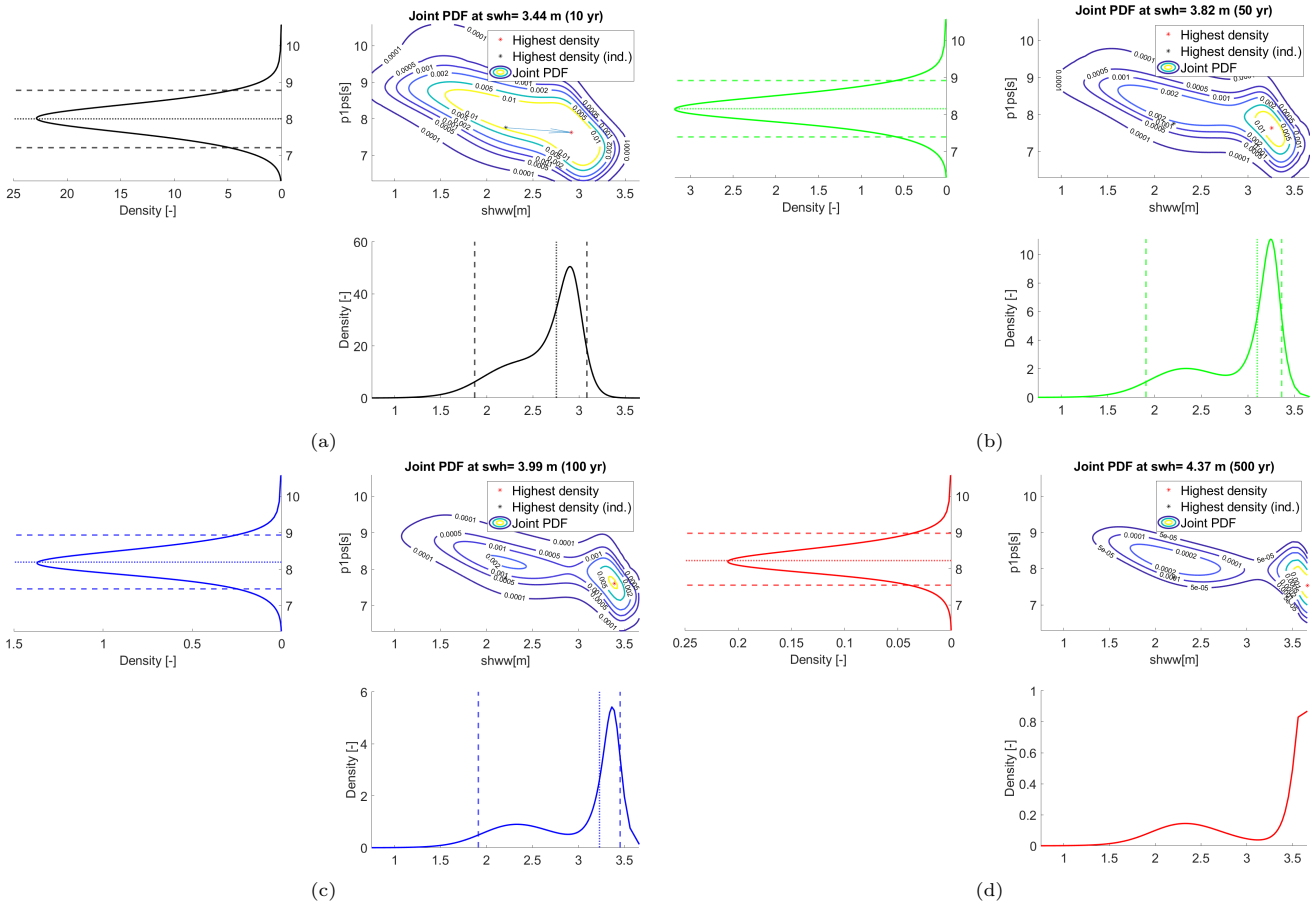


Figure 7.5: Results of vine 4. The 5th and 95th percentiles are marked with dashed lines in the marginal PDFs of each sub-figure, while the 50th percentile is marked with a dotted line.

The resulting joint PDFs of Figure 7.5 show that the point of highest density (represented by a red star) shifts to the right for higher conditional values of the dominant variable. Similarly to option 1, the conditional marginal densities are plotted in the same graph in Figure 7.6.

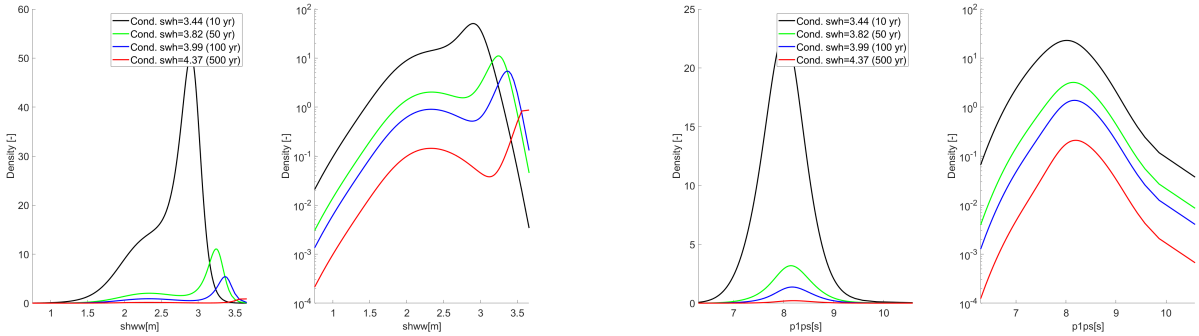


Figure 7.6: Vine 4 margins of *shww* (a) and *p1ps* (b) plotted in the same graph. Additionally the same margins are plotted using a log scale for the vertical axis to compare the shape of the curves.

The shape of the marginal distributions of *p1ps* is very similar compared the same graphs of option 1, shown in Figure 7.4. When examining the marginal distributions of *shww*, Figure 7.6a shows that the mode of *shww* increases for higher return values of *swh*. This is in agreement with the positive correlation observed between the pair *swh*-*shww* in the extreme data set ($\tau = 0.35$). The same graphs also show a secondary peak that indicates

that there is a small probability that an extreme value of swh is due to a smaller value of $shww$ (and thus a higher value of $shts$), which means the extreme event is caused by a combination of wind wave and swell energy. The secondary peak of $shww$ is higher than the secondary peak of $shts$ compared to the density at their respective highest modes.

7.5.3 Comparing the points of highest density

The points of highest density are compared for option 1 (vine 1) and 2 (vine 4). The missing values of $shww$ (or $shts$) and $p1ww$ are calculated using the relations described earlier (Eq. 7.1 and Eq. 7.2). The results are shown in Table 7.4.

Variable (unit)				
$swh(m)$	cond. swh=3.44 (10 yr)	3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
Vine 1: $shww(m)$	3.24	3.64	3.82	4.22
Vine 4: $shww(m)$	2.92	3.25	3.39	3.66
Vine 1: $p1ww(s)$	6.14	6.52	6.67	7.02
Vine 4: $p1ww(s)$	5.83	6.15	6.28	6.53
Vine 1: $shts(m)$	1.16	1.16	1.15	1.14
Vine 4: $shts(m)$	1.82	2.01	2.10	2.39
Vine 1: $p1ps(s)$	7.96	7.96	8.02	8.15
Vine 4: $p1ps(s)$	7.63	7.64	7.61	7.53

Table 7.4: Points of highest density of wind wave and swell parameters of Vine 1 and Vine 4, conditional on the return values of swh. Note that 10, 50, 100 and 500 years is only the univariate return period of swh and not the joint return period.

First the similarities of the outcomes of vine 1 and 4 are discussed. For the point with the highest density:

- The value of the significant height of the wind waves ($shww$) is higher than the value of the significant wave height of swell ($shts$). Wind waves are dominant given an extreme value of the total significant wave height (swh).
- The period of the wind waves ($p1ww$) increases for higher values of the total significant wave height, which makes sense given the one-to-one relation with the significant height of the wind waves.
- The period of the swell ($p1ps$) shows little change for more extreme values of the total significant wave height.

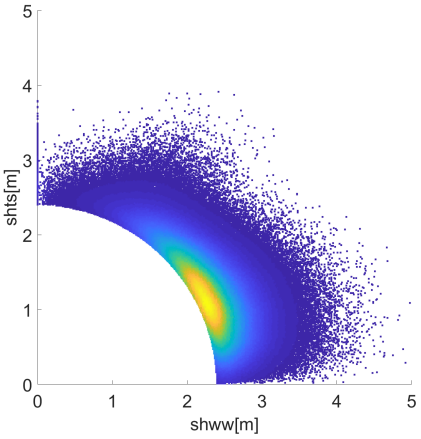
However the two outcomes also appear to show some differences:

- The value of the significant height of the swell is about two times higher for vine 4 than for vine 1. Furthermore, the former increases for more extreme events while the latter does not.
- The period of the swell increases slightly for vine 1, opposed to decreasing for vine 4. This seems off as the period and significant wave height of swell are positively dependent.

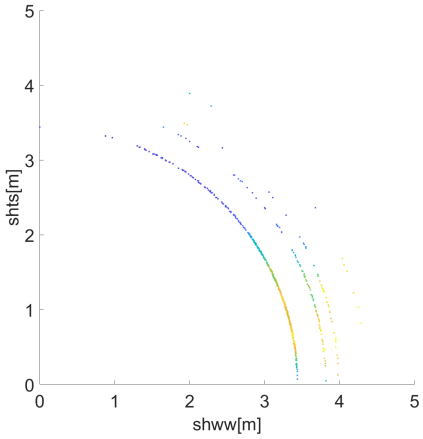
The differences between the two models seems off, as both models are based on the same extreme data set, the only difference being the significant wave height of wind waves is used instead of swell. In theory both models should provide the same result.

7.5.4 Comparing a large sample converted to physical units

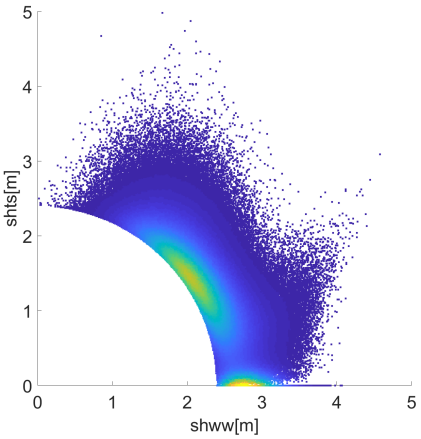
The differences between vine 1 and 4 are investigated further. To this end a large sample is extracted from both vines (10^6 realisations in each sample). These realisations are converted from the unit square to physical units using the marginal distributions of Figure 5.6. Next the significant wave height of wind waves ($shww$) is plotted against the significant wave height of swell ($shts$).



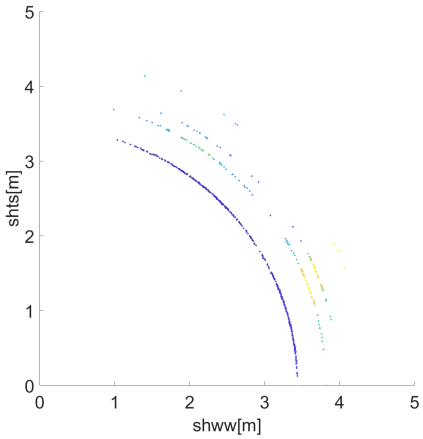
(a) Large sample of vine 1 converted from the unit square to physical units.



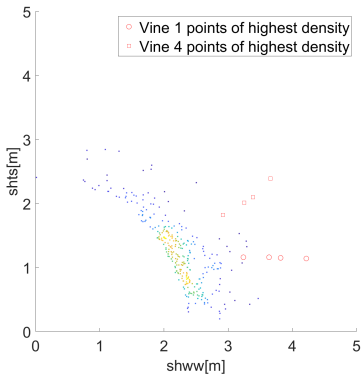
(b) Large sample of vine 1 conditionalised on four return values of swh.



(c) Large sample of vine 4 converted from the unit square to physical units.



(d) Large sample of vine 4 conditionalised on four return values of swh.



(e) Joint extreme observations and the points of highest density, determined previously.

Figure 7.7: Density scatter plots of a large sample converted to physical units of vine 1 (a) and vine 4 (c). The vine sample is conditionalised on four return values of the total significant wave height (swh) shown in (c) and (d) respectively. The joint observations of the extreme data set are shown, together with the points of highest density determined previously in (e).

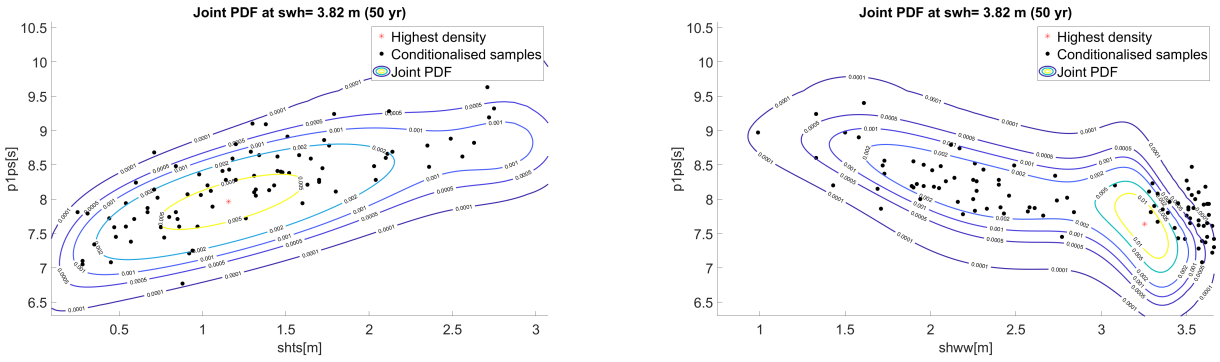
Figure 7.7 visualises the difference between vine 1 and 4 and compares it to the joint observations in the extreme data set. Additionally the points of highest density of Table 7.4 are plotted in Figure 7.7e. Next, a possible cause for the observed differences between the models is discussed, together with some concluding remarks about the multivariate analysis.

7.5.5 Comparing conditionalised samples to conditional densities

Recall that two types of results can be extracted once the vine copula model for three variables has been fitted:

- A multidimensional sample in the unit cube can be extracted from the vine copula.
- The vine copula density at any point in the unit cube can be calculated.

Both of these results are converted from the unit cube to physical units using the marginal distributions and compared. The multidimensional samples come from the vine copula itself, while the vine copula density is determined using kernel density estimation. The two results are compared by conditionalising on the 50 year return value of the dominant variable. This is done for option 1 (vine 1) and option 2 (vine 4) to illustrate the difference, the results are shown in Figure 7.8.



(a) Option 1 (vine 1)

(b) Option 2 (vine 4)

Figure 7.8: The conditional density is compared to a conditionalised sample for option 1 (vine 1) and option 2 (vine 4). Both the samples and the joint densities are conditionalised on the same return value of the dominant variable ($swh = 3.82$ m) to illustrate the difference.

The results of option 1 of Figure 7.8a are examined first. The conditionalised samples seem to match the conditional joint density well, although the samples seem to point to a slightly higher point of highest density for the variable $p1ps$. The secondary mode indicated by the conditional density (located at $shts = 2.6$ m approximately) also seems to be indicated by the conditionalised samples, although the number of samples in this location is low.

Next option 2 of Figure 7.8b is examined. The conditionalised samples and the conditional density are not in agreement considering the point of highest density. This could be because a kernel density estimation is used for the conditional density, possibly causing numerical inaccuracies in the result. The conditionalised samples come from the vine copula itself and are not affected by this. Apparently the results of option 1 are to a lesser extent affected by these numerical inaccuracies compared to the results of option 2. It is not immediately clear why this is. The secondary mode of the conditional density of Figure 7.8b (located at $shww = 2.2$ m approximately) seems to be in agreement with the conditionalised samples. The comparison shown in Figure 7.8 underscores that in terms of fitting vine 1 may perform better than vine 4, although this does not mean that the outcome of vine 4 must be disregarded in a design. Samples of vine 1 and vine 4 are compared to the joint observations in Appendix D.

7.6 Concluding remarks

As shown in Table 7.4 and Figure 7.7 vine 1 and 4 show some differences when considering the point with the highest joint density and a large sample. Three possible conclusions are discussed next.

- *Statement: both models are completely incorrect and a different approach is required.*

  A sample of both vine 1 and 4 is obtained and compared to the joint observations in the extreme data set in order to find out if this is the case. If both models are indeed completely incorrect, the vine samples would deviate from the joint observations. However, the joint observations seem to match fairly well with the generated extremes when compared, including when both data sets are converted from the unit square

to physical units or the standard normal space. Moreover, the correlation coefficients (Kendall's τ) of the samples are close to the correlation coefficients of the joint observations. The joint observations are compared to the vine samples in Appendix D.1.2. It is therefore concluded that both vine 1 and 4 are at least not completely incorrect.

- *Statement: one model clearly outperforms the other in terms of fitting. One should investigate which model is better, after which the other model can be disregarded entirely.*

  A number of model adequacy tests are performed in order to investigate this statement. The Cramér von Mises statistic (CvM) of the two different models is compared, as shown in Table 7.1. At first glance, the CvM statistic does seem to indicate that vine 1 (CvM=0.5629) is a better model than vine 4 (CvM=1.4509), thus supporting the statement above. However, the CvM statistic for the vine copula model is a summation of the CvM statistic of three unconditional bivariate edges, namely all bivariate combinations possible between the parameters, as is shown in Appendix D.1. Out of these three bivariate combinations, vine 1 and vine 4 only have one in common, *swh-p1ps*. For this combination vine 1 seems to be a slightly better fit for the data than vine 4, thus supporting this concluding statement.

  However, both vine 1 and 4 result in very similar values for *p1ps*, so making a decision based solely on this outcome would seem insufficient. The biggest difference in outcome between both vines is for *shts* conditional on *swh*. The CvM statistic does not provide an answer to which vine performs better in terms of fitting for these two variables.

  Again considering Table 7.1, only this time the values for the AIC, it becomes clear that the AIC of vine 4 (-385.02) is lower than the AIC of vine 1 (-244.89). But similarly to the CvM statistic, the AIC may only be used to compare models with the same variables. The AIC may therefore only be used to compare vine 1, 2 and 3 or vine 4, 5 and 6. A comparison between vine 1 and 4 based on the AIC is not valid. It is concluded that neither model can be disregarded entirely based on the CvM statistic and the AIC.

- *Statement: both models are fairly correct, even though the point of highest density shows a different result. One model performs slightly better than the other in terms of fitting.*

  For vine 1 and 4, the percentiles of the conditional marginal densities are compared in Appendix D.1.4. The percentiles show that for the wave periods of wind waves (*p1ww*) and swell (*p1ps*), the design realisations are within the 90% confidence interval. However, for the significant wave height of wind waves (*shww*) and swell (*shts*) this is not the case. This indicates that for these two variables at least, the outcomes of vine 1 and 4 are not in agreement.

  Next, two joint observations are compared to the outcomes of vine 1 and 4. Given that the total length of the time series is 40 years, the fourth highest event should be comparable to the design realisation conditional on the 10 year return value of *swh*. This comparison shows that the value of *shts* is somewhere in between vine 1 and 4, while the value of *shww* is closer to vine 1 than vine 4. Additionally, the highest event in the time series is compared to the design realisation conditional on the 50 year return value of *swh*. The comparison is shown in Appendix D.1.3. The highest event is more similar to vine 4 than vine 1, contradicting to the fourth highest event. It is concluded that the joint observations check shown in Appendix D.1.3 does not clearly indicate which model is to be preferred. The conclusion, that both models might be fairly accurate, is therefore not rejected.

  Finally, large samples of vine 1 and 4 are compared and shown in Figure 7.7. If the shape of the scatter density plots of both samples is compared to the joint distributions, vine 1 seems to represent the data slightly better than vine 4. This is also confirmed when calculating the correlation between *shww* and *shts* for the samples, which is $\tau_{V1} = -0.57$ for vine 1 and $\tau_{V4} = -0.72$ for vine 4, while the correlation of the joint observations $\tau_{ex} = -0.61$. It is concluded that both models are indeed fairly correct, even though the points of highest density are different. However, vine 1 seems to perform slightly better than vine 4. It is stressed that the 'performance' is only in terms of fitting. For a real design it is not recommended to disregard vine 4, but to consider both models.

8 Transformation to nearshore

The design realisations corresponding to extreme offshore wave spectra have been determined in previous chapters. In this chapter the nearshore propagation of these wave spectra is calculated. The main goal is to compare nearshore wave conditions for a number of single-peaked wave spectra with equivalent double-peaked wave spectra consisting of a wind wave- and swell component. SWAN (version 41.31) (Booij et al., 1999) is used to calculate the nearshore wave conditions for a number of offshore sea states. The outcome of the model runs are not the design values, because wind input is not added in the model.

8.1 Bathymetry

The bathymetry near the location of interest has been provided by Royal HaskoningDHV. The bathymetry is shown together with the data gathering point in Figure 8.1.

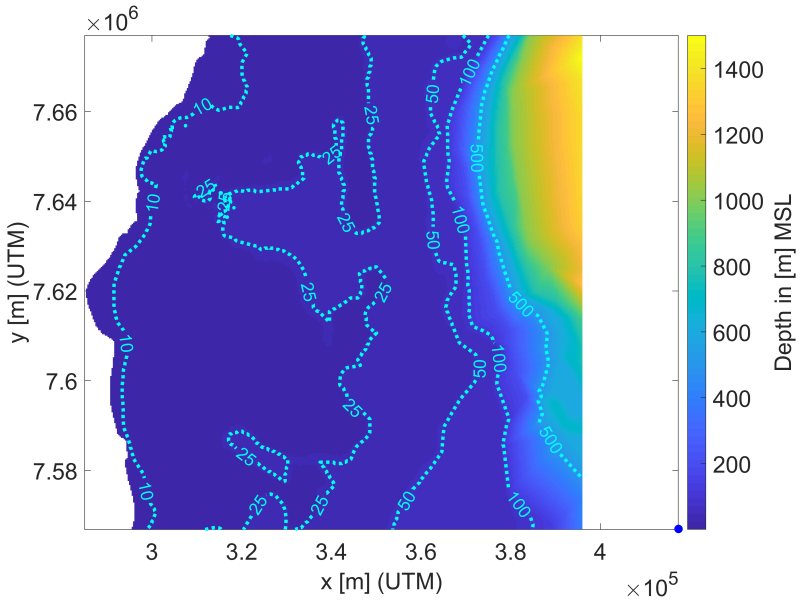


Figure 8.1: Local bathymetry together with the data gathering point plotted as a blue dot in the lower right corner. Dotted depth iso-lines are added in light blue for 10, 25, 50, 100 and 500 m.

The location of interest of Chapter 4 has the coordinates: 22S, 39.8W. This corresponds to the data gathering point of Figure 8.1. North is up in all spatial figures of this chapter.

8.2 Computational grid

Besides the local bathymetry a SWAN model requires a computational grid. A two-dimensional rectangular grid is used to compute the wave field. The grid has a resolution of 300 x 300 m and spans 111 x 110 km in total.

The full 2D action density spectrum is discretely calculated at each grid point. No nested computational grids are used in this model. Besides the discretisation in space, a discretisation in frequency and direction is required when calculating the full 2D spectrum for certain locations in the model. The directional spectrum is divided into 144 bins, which means that the density is calculated every 2.5 °. The frequency is discretised into 46 bins from 0.03 s^{-1} to 0.8 s^{-1}.

8.3 Boundary conditions

The boundary conditions of the model correspond the design realisations calculated in previous chapters. For the single-peaked spectra the parameters are shown in Table 8.1. The parameters for the equivalent double-peaked spectra are shown in Table 8.2.

ERA5	**Symbol**	**Unit**	R1	R4	R7
swh	H_{m0}	m	cond. swh=3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
mwd	θ	°	34.1	34.1	34.1
*mp*1	$T_{m0,1}$	s	6.45	6.70	7.25
wdw	σ_θ	°	33.9	33.9	33.9
-	γ	-	1.8	1.8	1.8

Table 8.1: Single peaked 2D spectrum parameters. Note that 50, 100 and 500 years is only the univariate return period of *swh* and not the joint return period.

ERA5	**Symbol**	**Unit**	R2	R3	R5	R6	R8	R9
			Vine 1	Vine 4	Vine 1	Vine 4	Vine 1	Vine 4
swh	H_{m0}	m	cond. swh=3.82	3.82	3.99	3.99	4.37	4.37
shww	H_{m0}	m	3.64	3.25	3.82	3.39	4.22	3.66
mdww	θ	°	16.6	16.6	16.6	16.6	16.6	16.6
*p*1*ww*	$T_{m0,1}$	s	6.52	6.15	6.67	6.28	7.02	6.53
dwww	σ_θ	°	30	30	30	30	30	30
-	γ_{ww}	-	1.0	1.0	1.0	1.0	1.0	1.0
shts	H_{m0}	m	1.16	2.01	1.15	2.10	1.14	2.39
mdts	θ	°	97.2	97.2	97.2	97.2	97.2	97.2
*p*1*ps*	$T_{m0,1}$	s	7.96	7.64	8.02	7.61	8.15	7.53
dwps	σ_θ	°	15	15	15	15	15	15
-	γ_s	-	1.7	1.7	1.7	1.7	1.7	1.7

Table 8.2: Double peaked 2D spectrum parameters, conditional on three univariate return values of *swh* (3.82 m, 3.99 m and 4.37 m). Note that R2 and R3 are equivalent to R1, R5 and R6 to R4 and R8 and R9 to R7 of Table 8.1.

Table 8.2 shows parameters of double-peaked spectra resulting from the point with highest density of vine 1 and vine 4, which are equivalent to the design realisations of Table 8.2, where each column represents a design realisation (labelled R1 to R9). The nearshore conditions resulting from the transformation of R1 is compared to R2 and R3. Similarly, R4 is compared to R5 and R6. Lastly, R7 is compared R8 and R9. The equivalent double-peaked design realisations shown in Table 8.2 show that the realisations belonging to vine 1 (R2, R5 and R7) all have a lower swell energy compared to vine 4 (R3, R6 and R9).

Because the goal in this chapter is to calculate the propagation of the single-peaked and double-peaked spectra, local wind is turned off in the model. The boundary conditions are defined at the north, east and south boundary of the model.

8.4 Model settings

Several physical processes are considered in the stationary model, which are summarised in Table 8.3.

Model parameter	Setting
Generation	3rd generation, which includes quadruplet interactions
White capping	Based on the Cumulative Steepness Method (CSM)
Wave breaking	Depth induced with default parameters
Bottom friction	JONSWAP friction is used ($c_{f,JON} = 0.038$)
Wind forcing	off
Triad interactions	off
Required accuracy	99% of previous iteration, with as many iterations as necessary for convergence

Table 8.3: Settings of the physics of the SWAN model.

For a more elaborate documentation of the physical processes used in the SWAN model, the user is referred to the SWAN user manual (SWAN, 2006) and technical documentation (SWAN, 2018).

8.5 Output

The output of the model is twofold. First, output is generated for the parameters H_{m0}, $T_{m0,1}$, θ (mean wave direction) and other spectral parameters at each grid point of the model. This can be used to generate surface maps of each wave parameter. The surface maps are used for illustration purposes and to validate the model. Surface maps of H_{m0} and $T_{m0,1}$ are shown in Appendix E.

The second type of output of the model consists of a number of so-called output points, where 1D and 2D wave spectra are calculated. The nearshore output point of interest is chosen at a nearshore location with shallow water (depth is 9.1 m). This location is shown as a green dot in Figure 8.2 and is chosen for illustration purposes. Additionally, some output points are chosen near the boundaries and in the centre of the model, again for validation purposes. These output points are plotted as red dots in Figure 8.2.

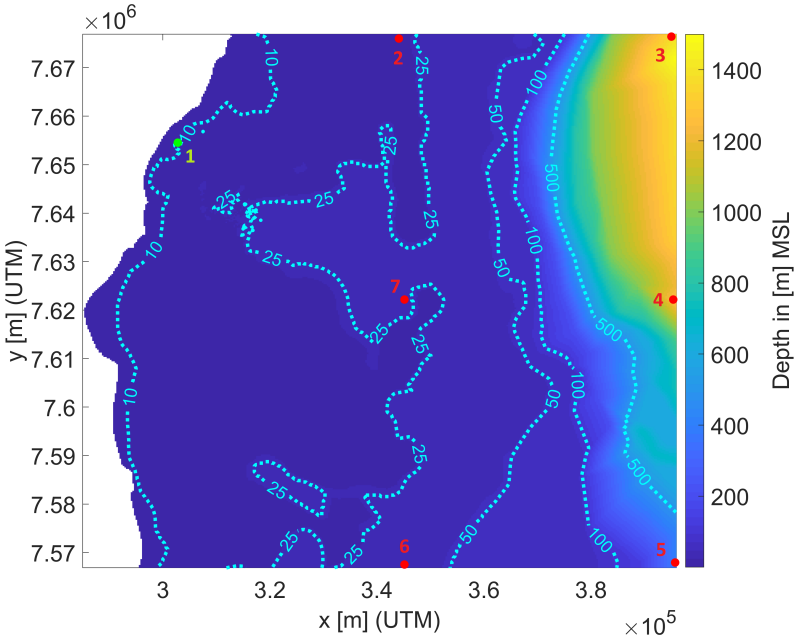
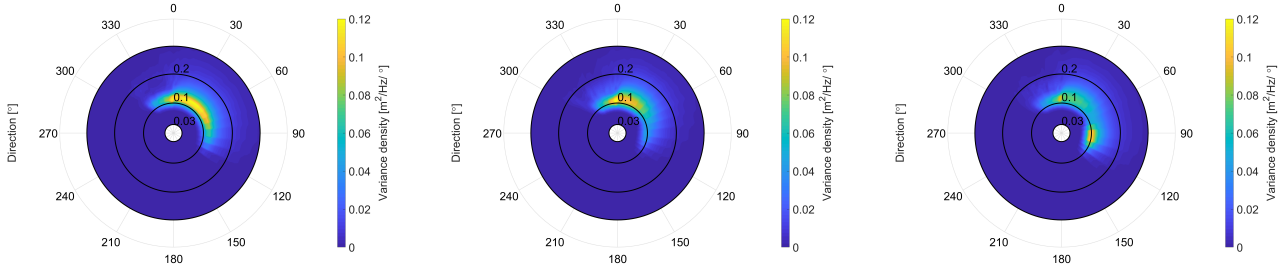


Figure 8.2: Local bathymetry together with output points marked as dots numbered 1 to 7. The green dot (1) is the nearshore location of interest with a local water depth of 9.1 m. The output of the red dots is used to validate the model. Dotted depth iso-lines are added in light blue for 10, 25, 50, 100 and 500 m.

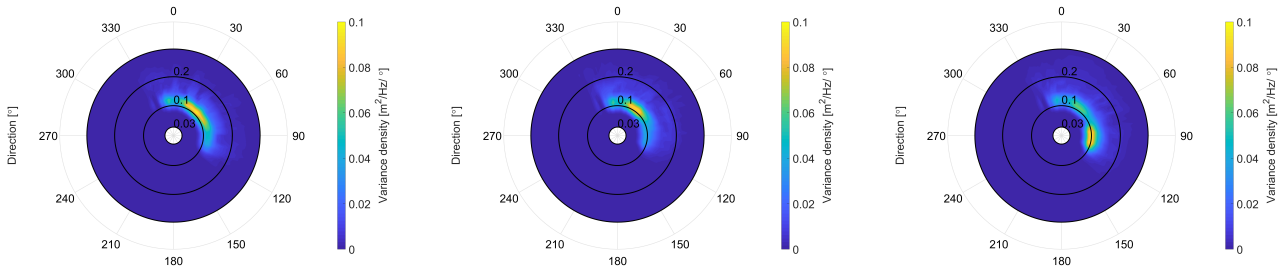
8.5.1 Variance density spectra

The boundary conditions R1, R2 and R3 of Table 8.1 and 8.2 are applied to the model. The 2D variance density spectra are calculated for the output points 4, 7 and 1 of Figure 8.2. Point 4 represents an offshore location near the boundary of the model with a local depth of 1246 m. As stated previously, point 1 is the nearshore location of interest with a local depth of 9.1 m. Finally, point 7 is a located between the offshore and nearshore point and has a local depth of 25.5 m. The (2D) variance density spectra at these points are shown for R1, R2 and R3 in Figure 8.3, 8.4 and 8.5 to illustrate the propagation of the wave energy to the nearshore. The outcome for the other boundary conditions (R4 to R9) is qualitatively similar to the outcome of R1, R2 and R3, only with higher wave energy input. The variance density spectra for R4 to R9 are shown in Appendix E.



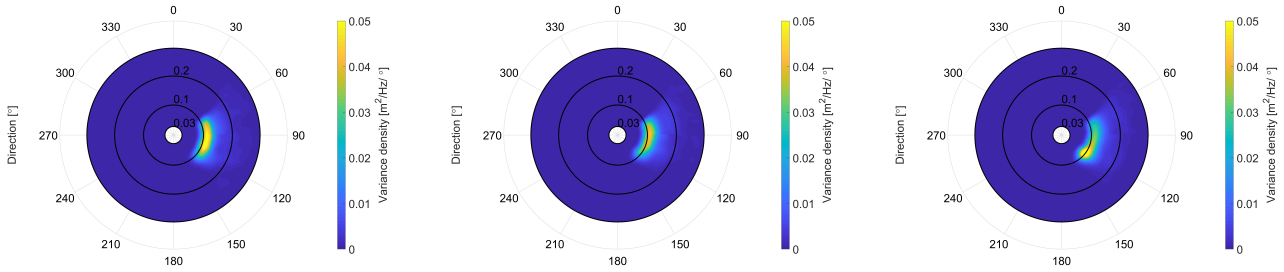
(a) R1 (b) R2 (c) R3

Figure 8.3: The variance density spectra offshore (point 4 of Figure 8.2) for boundary conditions R1 (from Table 8.1), R2 and R3 (from Table 8.2). The frequencies of 0.03, 0.1, 0.2 and 0.3 Hz are shown as black circles.



(a) R1 (b) R2 (c) R3

Figure 8.4: The variance density spectra at a point between the offshore and nearshore location (point 7 of Figure 8.2) for boundary conditions R1 (from Table 8.1), R2 and R3 (from Table 8.2). The frequencies of 0.03, 0.1, 0.2 and 0.3 Hz are shown as black circles.

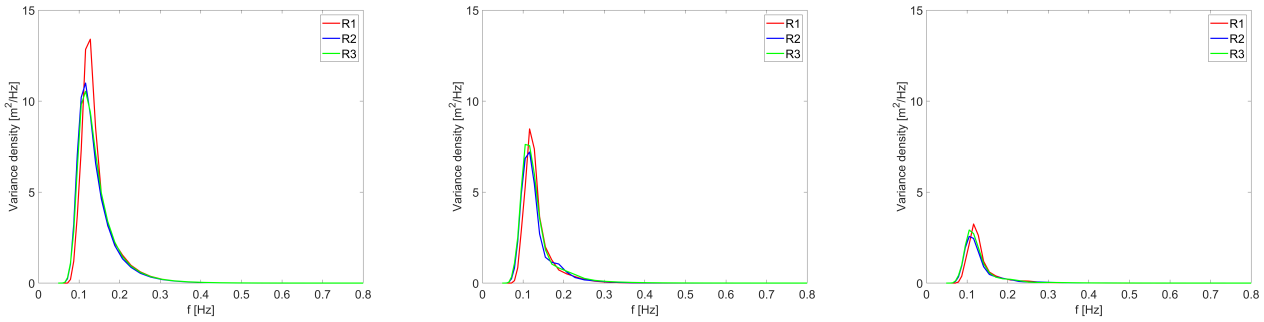


(a) R1 (b) R2 (c) R3

Figure 8.5: The variance density spectra nearshore (point 1 of Figure 8.2) for boundary conditions R1 (from Table 8.1), R2 and R3 (from Table 8.2). The frequencies of 0.03, 0.1, 0.2 and 0.3 Hz are shown as black circles.

The variance density is calculated for 25 frequency bins from $0.03 < f < 0.3\ Hz$, which corresponds to wave periods between 3.33 and 33.33 s. The directions are divided into 144 bins, which corresponds to 2.5 ° per bin. The 1D variance density spectra of Figure 8.6 are obtained by integrating the 2D variance density spectra over

all directions using Eq. 3.15. The figures in this section show that wave energy dissipates as it propagates to the nearshore.



(a) Offshore (point 4) (b) Between offshore and nearshore (point 7) (c) Nearshore (point 1)

Figure 8.6: 1D variance density spectra for the boundary conditions R1, R2 and R3. The 1D spectra are calculated for the same points as in Figure 8.3, 8.4 and 8.5.

8.5.2 Spectral parameters nearshore

The green dot labelled as 1 in Figure 8.2 represents the nearshore point of interest and is investigated further. The values for a number of spectral parameters for the nearshore location of interest are given in Table 8.4. The values are stated per run, labelled R1 to R9. The corresponding boundary conditions of R1 to R9 are stated in Table 8.1 and 8.2. It is again stressed that the resulting values of Table 8.4 are not design values, since wind input is absent in the model.

SWAN	**Parameter**	**Unit**	R1	R2	R3	R4	R5	R6	R7	R8	R9
Hsig	H_{m0}	m	1.58	1.53	1.61	1.65	1.59	1.66	1.83	1.70	1.79
Hswell	$H_{m0,swell}$	m	0.49	0.68	0.71	0.62	0.75	0.75	0.92	0.93	0.88
Dir	θ, MWD	°	97.6	96.7	105.0	98.6	96.9	105.5	101.0	97.7	107.3
PkDir	θ_p	°	106.3	88.8	123.8	106.3	88.8	123.8	108.8	106.3	126.3
Tm01	$T_{m0,1}$	s	7.34	7.48	7.55	7.59	7.62	7.75	8.13	7.98	8.11
Tm02	$T_{m0,2}$	s	6.93	6.72	6.81	7.14	6.88	7.13	7.67	7.26	7.65
TPsmoo	$T_{p,smoothed}$	s	8.47	9.16	9.18	8.66	9.31	9.25	9.40	9.72	9.37
Dspr	σ_θ	°	20.7	21.5	23.4	20.7	21.2	23.0	20.4	20.6	22.4

Table 8.4: SWAN output of a number of spectral parameters of the nearshore location of interest (point 1 of Figure 8.2) for each run, labelled R1 to R9.

8.6 Concluding remarks

The resulting nearshore significant wave height from Table 8.4 (H_{m0}) of R1, R2 and R3 are compared. Based on the results shown in Table 8.4 and the variance density spectra shown in Section 8.5.1, it is concluded that in terms of total wave energy (related to H_{m0}), the resulting nearshore values for the different boundary conditions are very similar. Although the result for H_{m0} for R2 is slightly lower compared to R1 and R3. The results for R4 to R9 are qualitatively very similar to R1, R2 and R3. This means that in terms of total wave energy that arrives nearshore, using a single-peaked spectrum (R1, R4 and R7) as a boundary condition results in a very similar value compared to an equivalent double-peaked spectrum (R2, R3; R5, R6; R8, R9).

Next the shapes of the nearshore 2D variance density spectra of R1, R2 and R3 (Figure 8.5) are compared. It seems that performing the calculation with a single-peaked spectrum (R1) at the boundaries of the model may lead to a slightly different spectral shape nearshore compared to a double-peaked boundary condition (R2 and R3), for the directional combination chosen in this thesis. It seems that the mean wave direction shifts more

towards the south for R2 and R3. This is also indicated by the value for θ in Table 8.4.

It is concluded that in terms of total wave energy that arrives nearshore, using a single-peaked spectrum results in a very similar value compared to an equivalent double-peaked spectrum. However, the shape and direction of the nearshore spectra is somewhat different. Although the uncertainty of the result is not quantified, the differences in peak direction between the runs are sufficiently large to potentially be statistically significant. Furthermore, judging from Figure 8.5, the shape of the 2D spectra seems to be different between the runs, although the differences in shape seem to be less significant than the differences in direction. This means that the more elaborate approach where a double-peaked spectrum is used as boundary condition potentially affects the design of coastal infrastructure if it is sensitive to spectral shape and direction.

It should be stressed that the outcome of the model runs are only for a given combination of wave directions at the boundaries of the model. The conclusions of this section are therefore only applicable for the (illustrative) directional combination chosen in this thesis. Furthermore, the resulting values are not the design values, since no wind input is taken into account in the model.

9 Conclusions

This research aims to develop a methodology to establish extreme offshore wave conditions given the presence of combined wave systems. In this chapter the main results and findings of this thesis are summarised. Additionally, the quality of the methodology is discussed.

9.1 Data collection

A 40 year ERA5 time series at an offshore location near the southern coast of Brazil is considered. The time series, for which wind wave- and swell energy is partitioned, shows that most of the wind waves originate from the north, while most of the swell originates from the south and east. However, for the highest total significant wave height, it becomes clear that the wind sea and swell most likely have a more similar direction (both from approximately 160 to 230 degrees). It is concluded that the highest total significant wave height is most likely **not** the result from a crossing sea state.

It is concluded that given normal conditions, the sea state is mostly dominated by swell. Mean wave period and the significant wave height for the wind waves partition are on average lower than for the swell partition. However, for extreme significant wave heights, the wind waves are likely to be higher.

9.2 Univariate analysis

The time series is conditionalised on a combination of wind wave and swell direction. This combination is chosen for illustrative purposes in this thesis and does not contain the highest total wave events. The analysis could be repeated for all possible combinations if the procedure is automated in the future. The total significant wave height is assumed to be the most important parameter and is thus chosen as the dominant variable. Next a univariate analysis is performed on the dominant variable. It is concluded that the choice of this variable as the dominant one is dependent on the application. The variables of interest are either fitted to various distributions or assumed as deterministic values.

It is concluded that the directional spreading of the wind wave- and swell partition yields unsatisfactory values when averaging their values over the extreme data set, it is not fully clear why this is. A possible explanation for the unrealistic result for swell might be that the parameters in the extreme data set belong to the *total* swell partition. Using the total swell partition in the analysis disregards the presence of multiple crossing swell systems during certain extreme events.

The peak enhancement factor for the total wave is available in the ERA5 time series. However, the extreme data set does not seem to converge to realistic values. Therefore, the peak enhancement factor for the total wave is calculated for each extreme event using an empirical relation, out of which a deterministic value is assumed. The peak enhancement factor for the wind wave- and swell partitions is deterministically assumed in a similar way, only using a different empirical relation.

9.3 Bivariate analysis

In order to generate the single-peaked 2D wave spectrum, design values of the total mean wave period are required. To find the associated values of the total mean wave period conditional on a number of univariate return values of the dominant variable, a copula-based approach is applied. A number of bivariate copulas are fitted. It is concluded that none of the fitted copulas model the steepness relation between the two parameters well enough, resulting in unrealistic design values for the total mean wave period. Therefore, a truncation approach is applied which results in values for the total mean wave period that are more realistic.

A downside of the truncation approach is that it only provides the associated value for the mean wave period given a univariate return value of the total significant wave height. A joint return period cannot be extracted,

because the approach is no longer fully parametric. This is problematic as for some applications the joint exceedance probability is required. Based on other research, it is concluded that the skew-t copula can be used in this case to model the two variables, if the joint exceedance probability is required.

9.4 Multivariate analysis

In order to obtain the design values necessary to generate a double-peaked 2D wave spectrum, the dependence structure of the multivariate extreme data set is modelled using a regular vine copula. The vine structure is selected based on a number of model adequacy tests. With the vine copula, the design values of the concomitant variables are determined conditional on a number of univariate return values of the dominant variable. The values of the concomitant variables are chosen corresponding to the point with the highest conditional density.

The robustness of the model is investigated in various ways. The model is set up in a slightly different manner, and the outcome of the two models is compared. When considering the points of highest density, the two models are not fully in agreement. Although qualitatively the two models show some of the same outcomes, it is concluded that the model is not very robust. However, it seems that vine 1 performs slightly better than vine 4 in terms of fitting. This does not mean that the results of vine 4 should be disregarded in the design.

9.5 Transformation to nearshore

The nearshore conditions are calculated using a SWAN model. The parameters determined in the previous chapters are used to construct a number of single-peaked and equivalent double-peaked wave spectra. These spectra are used as boundary conditions in a SWAN model and transformed to a nearshore point of interest. The nearshore wave conditions of the equivalent spectra are compared. In terms of total wave energy that arrives nearshore, using a single-peaked spectrum results in a very similar value compared to an equivalent double-peaked spectrum, for the directional combination chosen in this thesis. However, the shape and direction of the nearshore spectra seem to be slightly different, although the uncertainty is not quantified in this thesis. This means that the more elaborate double-peaked approach potentially affects the design of coastal infrastructure if it is sensitive to spectral shape and direction.

It should be stressed that the outcome of the model runs are only for a given combination of wave directions at the boundaries of the model. The conclusions of the model are therefore only applicable for the (illustrative) directional combination chosen in this thesis. Furthermore, the resulting values are not the design values, since no wind input is taken into account in the model.

9.6 Discussion

9.6.1 Discussion on the quality of the methodology

The methodology explained in this thesis contains a number of assumptions and sources of uncertainty. Each of these assumptions might contribute some uncertainty to the final result. It is unknown what the uncertainty of the final result of the analysis is and what assumptions or variables are the biggest contributors to the overall uncertainty. Some common sources of uncertainty are discussed:

- Statistical uncertainty
  - A source of statistical uncertainty is the extrapolation that is applied in the methodology. Extrapolation is required as the length of the time series is shorter than the required return periods, which is often the case in the design of structures. As shown for the univariate analysis of the dominant variable in Figure 5.3, the statistical uncertainty increases for higher return periods. For instance, in the univariate case for a return period of 500 years the 95% confidence interval is larger than for smaller return periods. Although not quantified in this thesis, the statistical uncertainty is also relevant for higher dimensions, such as in the bivariate and multivariate analyses.
- Model uncertainty

  - Uncertainty in the spectral parameters obtained from the ERA5 time series.
  - Fitting all marginal distributions, bivariate copulas and vine copulas all include some uncertainty. It is not clear beforehand which distribution best fits which parameter, the final result changes slightly if a different distribution type is used. This uncertainty is not quantified in this thesis.

- Inhomogeneity
  - Inhomogeneity due to seasonality is not directly taken into account in the current analysis. This might be problematic as different sea states might be more frequent in certain seasons. It might be worthwhile to conditionalise the data set further into summer and winter populations, for example. As of yet it is unknown what the effect of including seasonality in the approach is on the final result of the analysis.
  - It is assumed that the time series used in this thesis is stationary, meaning that the statistical properties of the time series do not change over time. Non-stationarity of the time series is a cause of inhomogeneity and thus violates the IID assumption. Any effects of possible climate change are not taken into account. If effects of climate change are known, possible trends (such as more intense storms or sea level rise) could be detected and taken into account for the final result.

- Deterministic assumptions
  - Spectral shape, mean wave direction and directional spreading are assumed as a deterministic values for all extreme events. Although this might be a reasonable assumption for some of these parameters, the spectral shape of more extreme events might not be so constant as is assumed in this thesis. A more elaborate statistical model, which incorporates these parameters is required to validate this assumption.
  - The relation between the mean wave period and the significant wave height of wind waves is assumed deterministically with a steepness relation (Figure 7.1). However, in reality the correlation is high, but the two variables are not fully dependent.

It is noted that the sources of uncertainty discussed in this section are common among similar statistical research.

9.6.2 Discussion on choosing a design realisation

It has been concluded previously that the multivariate model lacks robustness. This entails that choosing a single combination of load parameters as a design realisation might not be advisable in the design process, as was done when choosing the point of highest density in the multivariate analysis. Besides the robustness, some other problems of this approach are discussed:

- The point with the highest density is not necessarily the combination of load parameters that is most detrimental to the stability of the structure in question. In order to clarify this point, an example is given. A joint PDF of swh, $shww$ and $p1ps$ is conditionalised at the univariate 100 year return value of $swh = 3.99m$. The point with the highest density is indicated with a red star in Figure 9.7.

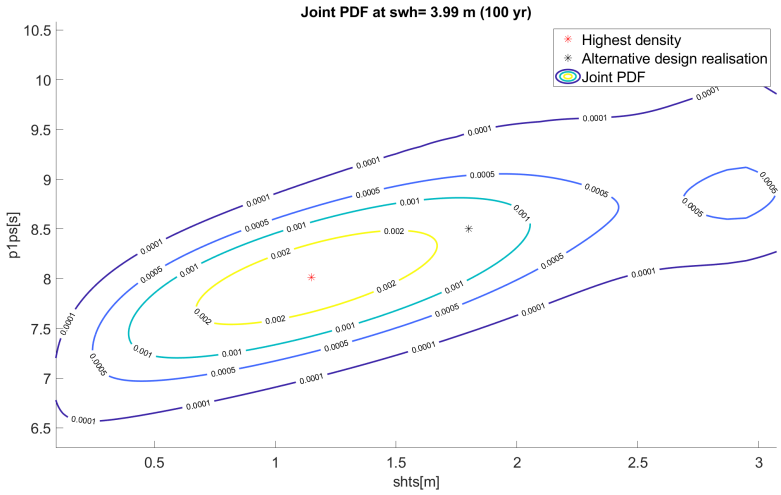


Figure 9.7: Joint density conditionalised at the 100 year return value of the total significant wave height (swh).

Next assume that a breakwater is being designed. During the design process it turns out that that the structure is strong enough to withstand the combination of load parameters corresponding to the point of highest density. However, if a load combination occurs with a higher swell energy and lower wind wave energy, shown by the black star in Figure 9.7, the structure fails. By assuming only one point in the design, other combinations that have a slightly lower density are not taken into account. It is not known beforehand which combination is the most detrimental to the structure as this is dependent on the application.

- Related to this matter, in the current methodology it is assumed that the total significant wave height is the dominant variable. However, the choice of the dominant variable might be different for some applications. If an application is equally sensitive to two or even more load parameters, the choice of a dominant variable is not so straight forward any more.

- A wind wave system assimilates swell if they share the same mean wave direction. This is not taken into account in this research as the time series is conditionalised on wind waves and swell coming from different directions. However, if the procedure is automated in the future, this should be taken into account for combinations where both partitions have a similar direction.

- If the methodology is further developed to include more load parameters, for instance when multiple crossing swell systems are treated separately, choosing the point of highest density could also lead to problems. A joint PDF of six parameters conditionalised on one dominant variable results in a five-dimensional PDF, possibly with multiple points of high density. This problem also increases further as the dimensionality grows.

- Lastly, it should again be mentioned that choosing the point with the highest density as the design realisation is simply a choice. Other design realisations might be favoured, depending on the situation. Similarly to the discussion for the bivariate case (Chapter 6), for the multivariate case there is no unique way to go from risk to design conditions.

10 Recommendations

Following the conclusions and observations from the analyses performed in this thesis, some recommendations are made. The recommendations are made to improve the developed methodology by suggesting future research and by giving some practical recommendations.

10.1 Recommendations for further research

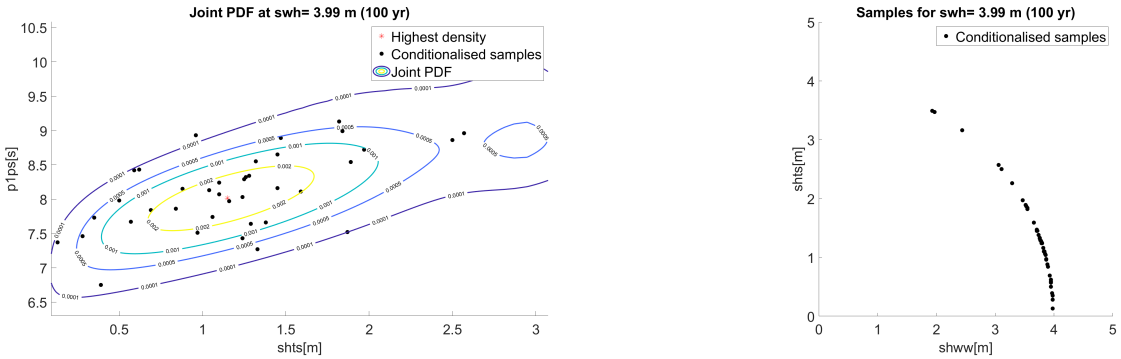
10.1.1 Bivariate analysis

As stated in the conclusions, the truncation approach used in Chapter 6 provides the associated values conditional on the univariate return values of the dominant variable. A major limitation of this approach is that it is no longer fully parametric because of the applied truncation. A consequence of this is that the joint density can only be approximated by taking a very large sample. Similarly, the joint exceedance probability cannot be calculated directly using the truncation approach and also has to be approximated, as is demonstrated in Appendix C.3. Other studies suggest that using a skew-t copula fits the data more accurately than any of the copulas tested in this thesis. It is recommended to investigate if this is the case and possibly use a different approach in the bivariate analysis.

An alternative approach to find the appropriate design combination for the bivariate case might be to select a number of combinations with equal joint exceedance probability and run all of these through the same model to find out which combination has the most unfavourable effect on the application. It is noted that some combinations have a higher probability density than others, which could to be taken into account by means of some sort of weighing factor.

10.1.2 Multivariate analysis

As stated in the discussion, if the design point is chosen as the point of highest density conditional on a univariate return value of the dominant variable, only one combination of load parameters is taken into account. It is not evident beforehand if the point of highest density is the most unfavourable combination of load parameters for a given application. It is therefore recommended to generate a number of samples and conditionalise on a return value of the dominant variable. The next step is to investigate which combination of parameters is most unfavourable for the application in question.



(a) (b)

Figure 10.8: The significant wave height of swell (*shts*) versus the mean wave period of the swell (*p1ps*) is shown in (a) for a number of conditionalised samples. The significant wave height of the wind wave partition (*shww*) is plotted versus the *shts* in (b) for the same samples.

To demonstrate this, a large number (10^6 for example) of storm events are generated and conditionalised on a return value of the dominant variable, as is shown in Figure 10.8. It is recommended to run all of these events

through the same application and determine which combination of load parameters is most unfavourable, instead of assuming the point of highest density as the design realisation. The procedure could be repeated for other return values of the dominant variable and directional combinations of wave systems. If the model is expanded to include more dimensions in the future, it is still possible to conditionalise on a return value of the dominant variable.

For this same application, it is recommended to perform a sensitivity analysis. This is could help to determine which variable is selected as dominant. Furthermore, a sensitivity analysis results in possible insight on which parameters need to be obtained more accurately in order to reduce the overall uncertainty of the outcome. It is therefore also recommended to perform a more elaborate uncertainty analysis in order to quantify the uncertainty of the result of the analysis.

Once the model robustness is increased, the model could be expanded further. Whereas the current methodology treats the swell partition as a single wave system, in reality this swell partition may at certain moments in time consist of multiple crossing swell systems, each with their own spectral parameters. Furthermore, wind input could be added to the model as well. As stated in the discussion, if the number of dimensions in the model increases, the conditional design realisation is no longer a good method of obtaining the design values. The sample-based approach described previously, or a different design realisation all together is recommended for higher dimensions.

Lastly, the southern coast of Brazil is not the only place where the combined wave systems are observed. Once a sound methodology has been established and the model robustness is increased, it is recommended to repeat the analysis with a time series of a different location where an extreme wave event is likely to consist of multiple crossing sea states. By repeating the analysis the model is validated with a different data set.

10.2 Practical recommendations

In this thesis an ERA5 time series is used. The parameters of this time series are based on a full 2D wave spectrum. However, the full 2D spectrum itself is not freely available. Even though the full 2D spectrum could provide some useful insights to the engineer performing the analysis when trying to explain the behaviour of certain parameters. Furthermore, for the sub-partitions of swell, only a limited number of parameters is currently available in the ERA5 time series. This means that if the methodology is expanded to include multiple crossing swell systems in the future, more parameters would have to be derived, for which the full 2D wave spectrum is required. It is therefore recommended to use a time series for which the full 2D spectral information is available.

A second practical recommendation is to automate the procedure where ever possible. If the method proposed in this thesis is applied on real designs in the future, the overall work load of the design process has to be minimised while not loosing sight of the quality of the outcome.

Lastly, in reference to the specific variables used in this thesis: considering the results of Chapter 8, it seems like the wave energy that arrives nearshore (related to H_{m0} of Table 8.4) is quite similar for the single-peaked and equivalent double-peaked offshore wave spectra, although wind was not considered in the transformation. It is recommended to investigate if this is also the case for all other directional combinations (only one directional combination is considered in this thesis). If this is the case and it is decided to build a coastal structure that is mostly sensitive to incoming significant wave height, it might be better not to go through multivariate analysis in the first place, but to perform the analysis only with total wave parameters. This will most likely yield a much faster result compared to performing the multivariate analysis, while having almost the same outcome as the multivariate analysis.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 2nd international symposiium on information theory, bn petrov and f. czáki, eds. *Ciaki eds, Akademiai Kaid5, Budapest.*

Bidlot, J. (2016). *Ocean wave model output parameters.* ECMWF.

Booij, N., Ris, R., and Holthuijsen, L. (1999). A third-generation wave model for coastal regions, part i, model description and validation. *J. Geophys. Res.*, 104:7649–7656.

Brechmann, E. and Schepsmeier, U. (2013). Modeling dependence with c- and d-vine copulas: The r package cdvine. *Journal of Statistical Software*, 52:1–27.

Caires, S. and Gent, M. (2008). Extreme wave loads. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, 2.

Caires, S. and van Os, J. (2012). *Analysis, classification and transformation of metocean data.* Deltares.

Chakak, A. and Koehler, K. J. (1995). A strategy for constructing multivariate distributions. *Communications in Statistics-Simulation and Computation*, 24(3):537–550.

Church, J., Clark, P., Cazenave, A., Gregory, J., Jevrejeva, S., Levermann, A., Merrifield, M., Milne, G., Nerem, R., Nunn, P., Payne, A., Pfeffer, W., Stammer, D., and Unnikrishnan, A. (2013). Sea level change. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, page 1139.

Coles, S. (2001). *An introduction to statistical modelling of extreme values.* Springer-Verlag, London, the UK, 1st edition.

Cooke, R. (1997). Markov and entropy properties of tree- and vine-dependent variables. *American Statistical Association - Proceedings of the Section on Bayesian Statistical Science.*

Croux, C. and Dehon, C. (2010). Influence functions of the spearman and kendall correlation measures. *Statistical Methods & Applications*, 19(4):497–515.

Dingemans, M. (1987). *Verification of numerical wave propagation models with laboratory measurements.* Deltares.

DNV (2012). *Modelling and analysis of marine operations.* Det Norske Veritas.

ECMWF (2018). Era5 hourly data on pressure levels from 1979 to present. `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview`.

Everts, P. (2016). *Probabilistic Design of a Rubble Mound Breakwater.* Master's thesis, Delft University of Technology.

Ewans, K. (2002). Directional spreading in ocean swell. pages 517–529.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering - J HYDROL ENG*, 12.

Genest, C., Rémillard, 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.

Gibbons, J. (1993). *Nonparametric measures of association.* Number nr. 91 in Quantitative applications in the social sciences. Sage Publications.

Hanea, A., Napoles, O. M., and Ababei, D. (2015). Non-parametric bayesian networks: Improving theory and reviewing applications. *Reliability Engineering System Safety*, 144:265 – 284.

Hanson, J. L. and Phillips, O. M. (2001). Automated analysis of ocean surface directional wave spectra. *Journal of Atmospheric and Oceanic Technology*, 18(2):277–293.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Mueller, P., Olbers, D., Richter, K., Sell, W., and Walden, H. (1973). Measurements of wind-wave growth and swell decay during the joint north sea wind project (jonswap). *Deutschen Hydrographischen Zeitschrift*, A8:95.

Hertogh, M. and Bosch-Rekveldt, M. (2015). *Integraal Ontwerp en Beheer*. TU Delft, Delft, the Netherlands.

Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge University Press, Cambridge, the UK, 1st edition.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349.

Isherwood, R. (1987). Technical note: A revised parameterisation of the jonswap spectrum. *Applied Ocean Research*, 9(1):47 – 50.

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

Jane, R., Valle, L. D., Simmonds, D., and Raby, A. (2016). A copula-based approach for the estimation of wave height records through spatial correlation. *Coastal Engineering*, 117:1 – 18.

Kuik, A. J., van Vledder, G. P., and Holthuijsen, L. H. (1988). A method for the routine analysis of pitch-and-roll buoy wave data. *Journal of Physical Oceanography*, 18(7):1020–1034.

Kurowicka, D. and Cooke, R. (2004). Distribution-free continuous bayesian belief nets.

Kurowicka, D. and Joe, H. (2011). *Dependence Modeling Vine Copula Handbook*. World Scientific Publishing, Singapore.

Li, F., van Gelder, P., Ranasinghe, R., Callaghan, D., and Jongejan, R. (2014). Probabilistic modelling of extreme storms along the dutch coast. *Coastal Engineering*, 86:1 – 13.

Li, J., Zhu, X., Lee, C.-F., Wu, D., Feng, J., and Shi, Y. (2015). On the aggregation of credit, market and operational risks. *Review of Quantitative Finance and Accounting*, 44:161–189.

Lucas, C. and Guedes Soares, C. (2015). On the modelling of swell spectra. *Ocean Engineering*, 108:749–759.

Michele, C. D., Salvadori, G., Passoni, G., and Vezzoli, R. (2006). A multivariate model of sea storms using copulas. *Coastal Engineering*, 54(10):734 – 751.

Morales-Nápoles, O., Cooke, R., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes.

Nelson, R. (2006). *An introduction to copulas*. Springer, Portland, USA, 2nd edition.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Pierson, W. and Moskowitz, L. (1964). A proposed spectral form for fully developed wind seas based on similarity theory of s.a. kitaigorodskii. *Journal of Geophysical research*, 69:5181–5190.

Repko, A., Gelder, P. V., Voortman, H., and Vrijling, J. (2004). Bivariate description of offshore wave conditions with physics-based extreme value statistics. *Applied Ocean Research*, 26(3):162 – 170.

Salvadori, G., Durante, F., De Michele, C., Bernardi, M., and Petrella, L. (2016). A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resources Research*, 52(5):3701–3721.

Salvadori, G., Tomasicchio, G., 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., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., et al. (2012). Vinecopula: Statistical inference of vine copulas. *R package version*, 1.

Schiereck, G. (2001). *Introduction to bed, bank and shore protection*. Delft University Press, Delft, the Netherlands, 1st edition.

Schmidt, T. (2007). Coping with copulas. *Copulas - From Theory to Application in Finance*.

Sellés Valls, S. (2019). *Bivariate probability analysis of extreme offshore events*. Van Oord.

Smit, P. B. and Janssen, T. T. (2019). Swell propagation through submesoscale turbulence. *Journal of Physical Oceanography*, 49(10):2615–2630.

Snipes, M. and Taylor, D. (2014). Model selection and akaike information criteria: An example from wine ratings and prices. *Wine Economics and Policy*, 3.

SWAN (2006). *SWAN user manual*. TU Delft.

SWAN (2018). *SWAN scientific and technical documentation*. TU Delft.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3):397–415.

Toffoli, A., Lefèvre, J., Bitner-Gregersen, E., and Monbaliu, J. (2005). Towards the identification of warning criteria: Analysis of a ship accident database. *Applied Ocean Research*, 27(6):281 – 291.

Verhagen, H., d'Angremond, K., and van Roode, F. (2009). *Breakwaters and closure dams*. VSSD, Delft, the Netherlands, 2nd edition.

Weber, P., Medina-Oliva, G., Simon, C., and Iung, B. (2012). Overview on bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4):671 – 682. Special Section: Dependable System Modelling and Analysis.

Zachary, S., Feld, G., Ward, G., and Wolfram, J. (1998). Multivariate extrapolation in the offshore environment. *Applied Ocean Research*, 20(5):273 – 295.

A Additional theory

The theory in this appendix gives provides some additional context for some of the decisions made in the applied methodology. However, the contents of this appendix are not used directly in this research.

A.1 Lifetime

An illustrative calculation for an arbitrary breakwater is performed to underscore the need for extreme value theory. If the design lifetime of the breakwater is assumed to be 50 years, which is a common value for breakwaters, one could assume that the design conditions of such a breakwater would also be 1/50 years. However, this is incorrect, because the probability of failure of the structure is given by the Poisson distribution:

$$P_f = 1 - exp(-f \cdot T_L) \tag{A1}$$

In which:

P_f	[-]	probability of failure of the structure during the period T_L
f	[1/year]	average frequency of the event
T_L	[year]	lifetime of the breakwater

Using Eq. A1 and assuming a design frequency of 1/50 years, the probability of failure of the structure during its lifetime is calculated as $P_f = 1 - exp(-(1/50) \cdot 50) = .632$. This means that the structure has a 63% probability of failure during its lifetime, which is considered too high for a breakwater. A more economical probability of failure is in the order of about 5% for breakwaters (Verhagen et al., 2009). Which means that the design frequency becomes:

$$\begin{aligned} f &= -\frac{1}{T_L} \cdot ln(1 - P_f) \\ &= -\frac{1}{50} \cdot ln(1 - 0.5) \\ &= 0.001[yr^{-1}] \end{aligned}$$

The return period (RP) is 1 divided by the frequency, or $RP = 1000[yr]$. This goes to show what order of magnitude to expect for return periods of interest in case of a breakwater. Furthermore, it underscores the need for extreme value theory, because data records of 500 or 1000 years are (almost) never available in practice.

A.2 Hazard Scenarios

Extensively treated by Salvadori et al. (2016), a method is proposed on how to deal with hazard scenarios containing multiple parameters. S_u is defined the *upper set* of events that are considered detrimental for the structure of interest. Two hazard scenarios are defined that have different upper sets.

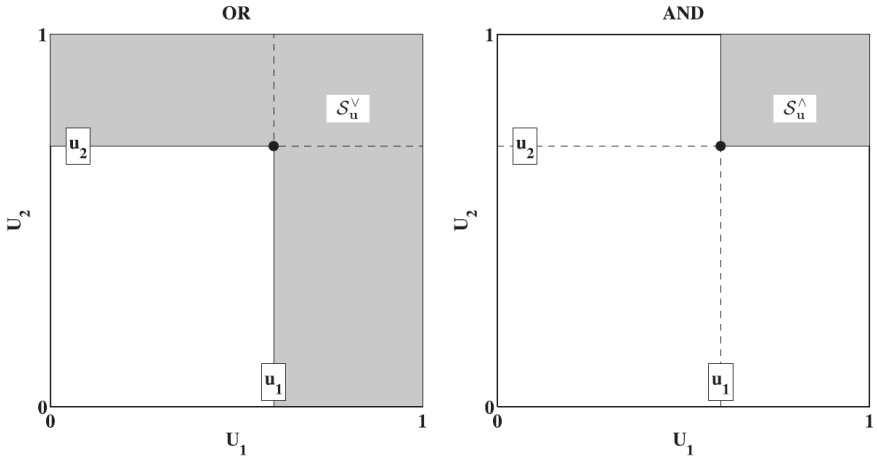


Figure A1: The OR and AND hazard scenario in the case of a bivariate copula. The upper set (S_u) of each hazard scenario is marked in grey. From Salvadori et al. (2016).

A hazard scenario in the context of bivariate analysis is a set of events that can be interpreted as dangerous, as different combinations of (load) parameters could be detrimental when considering a structure of interest. The hazard scenarios OR and AND are defined in the bivariate copula space [0 1] [0 1], as is shown in Figure A1. An example where the OR hazard scenario could be applicable in the analysis is a river with a certain discharge and water level, which show a positive dependence. In this case a hazardous event is characterised by *either* a high water level *or* a high discharge.

Another example is considered to explain the AND hazard scenario. Consider storms which are characterised by a maximum significant wave height in meters and storm duration in minutes. It is assumed that maximum significant wave height and storm duration show positive dependence, meaning that longer lasting storms result in a higher maximum significant wave height. Suppose a breakwater is able to withstand a certain significant wave height without any problems for a arbitrary duration. This same breakwater is only damaged by a storm that produces a very high significant wave height *in combination with* a long duration. Storms that last long, but produce a lower significant wave height are no problem for the breakwater in question. Similarly, storms that produce a high significant wave height, but do not last very long will not damage the structure. Thus the AND hazard scenario is applied.

B Univariate analysis results

B.1 POT results

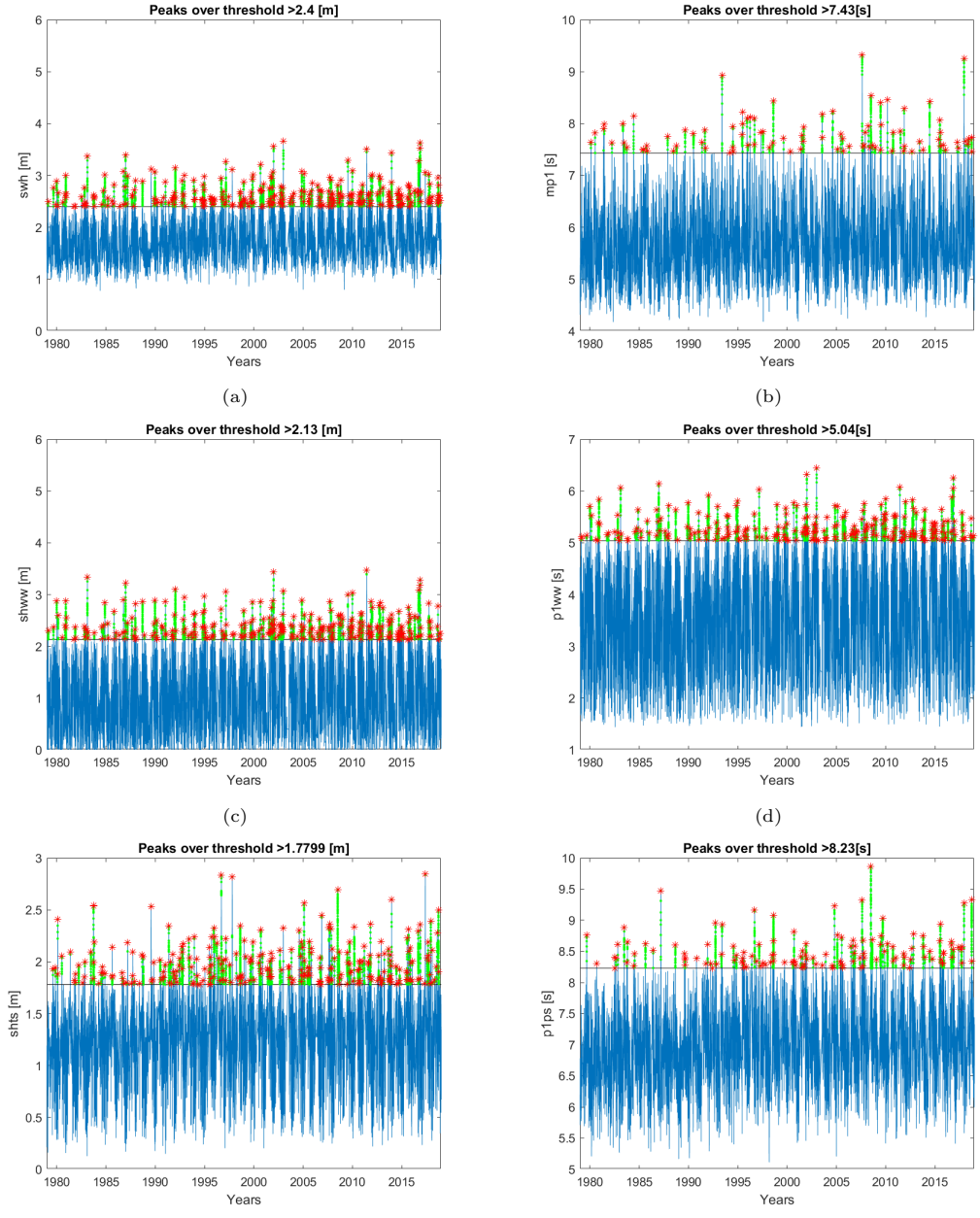
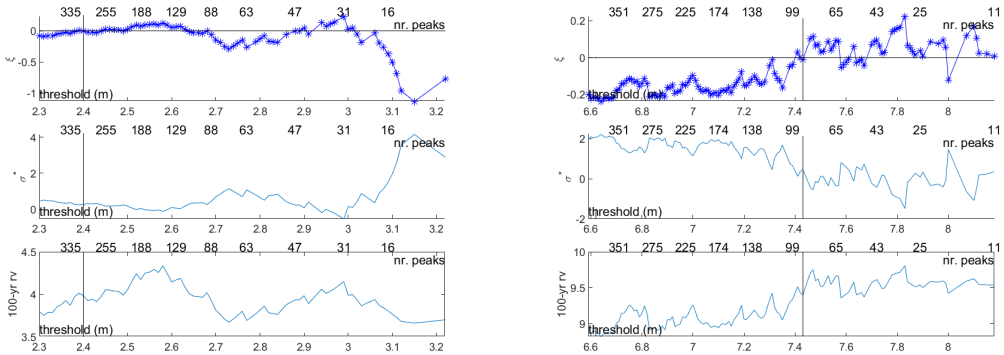
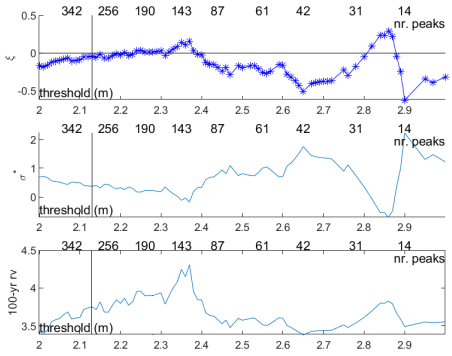


Figure B1: Peaks and threshold of swh, $mp1$, $shww$, $p1ww$, $shts$, $p1ps$

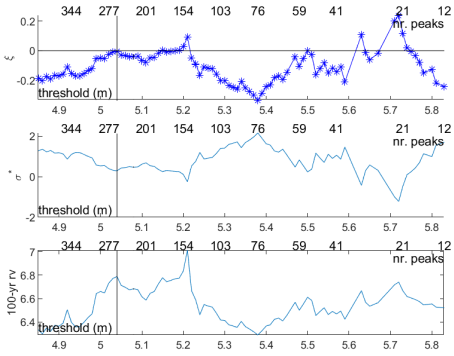


(a) swh

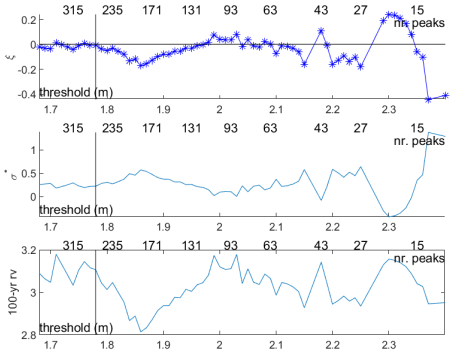
(b) $mp1$



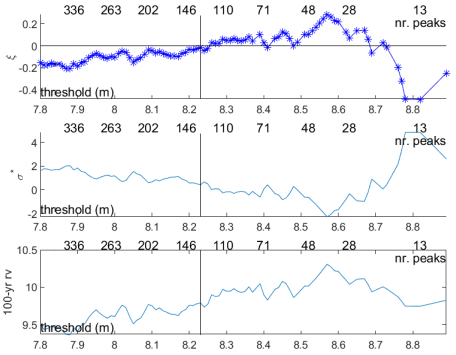
(c) $shww$



(d) $p1ww$



(e) $shts$



(f) $p1ps$

Figure B2: The shape (ξ), scale (σ) and 100 year return value threshold stability for swh, $mp1$, $shww$, $p1ww$, $shts$, $p1ps$. The chosen threshold is also shown in the Figure B1.

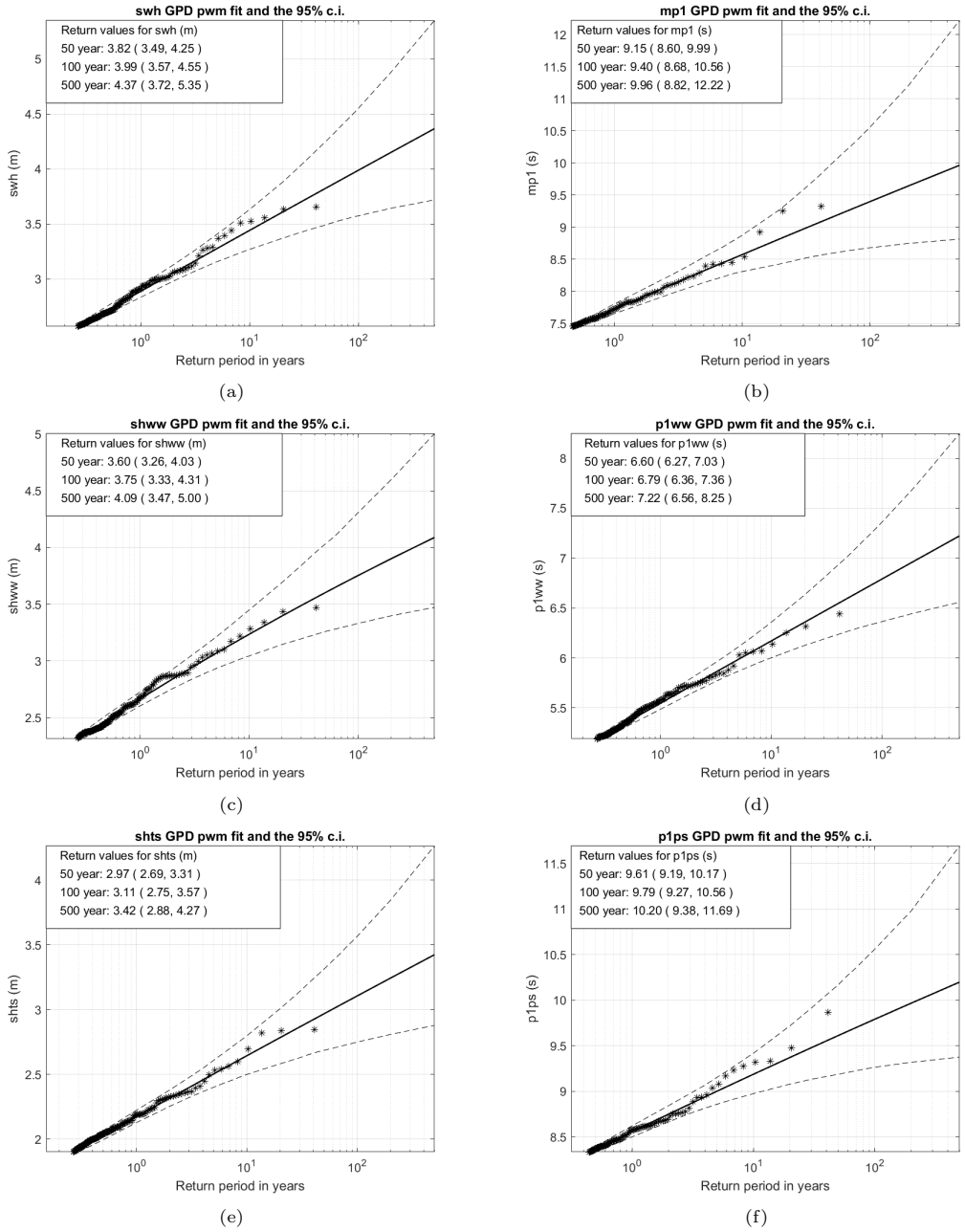


Figure B3: Extrapolated fitted GPDs, together with the 95% CI for swh, $mp1$, $shww$, $p1ww$, $shts$, $p1ps$

B.2 Checking the limit steepness

The wave steepness is defined in Eq. 4.3. The significant wave height (H_{m0}) is plotted against the mean wave period ($T_{m0,1}$) for the total, wind wave- and swell partition of the extreme data set.

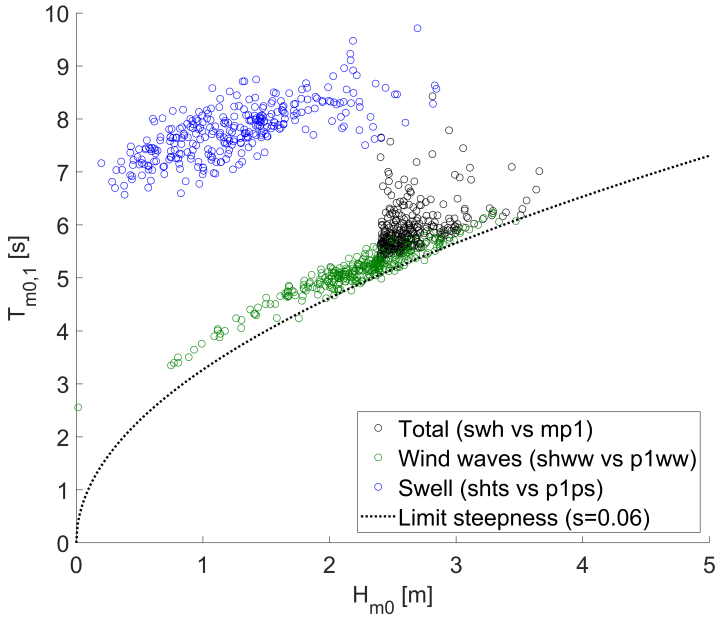
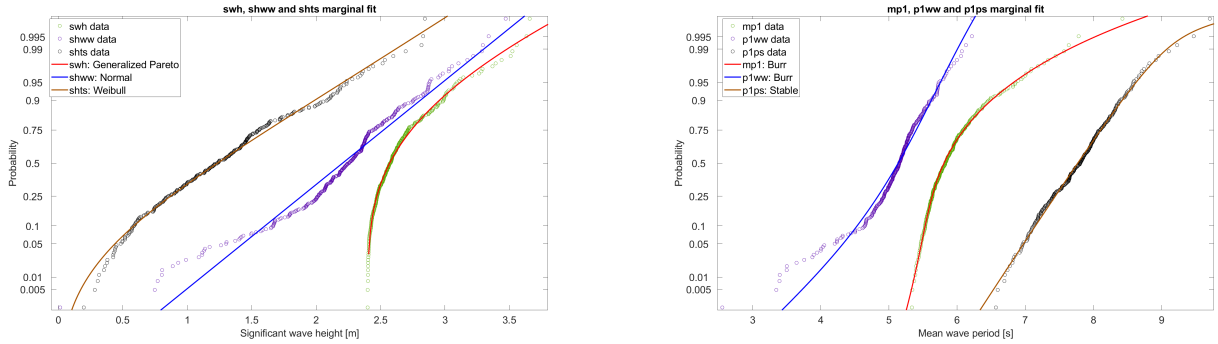


Figure B4: Joint observations of the extreme data set for the total, wind wave- and swell partition.

Figure B4 shows that almost all observations in the extreme data set are above the limit steepness line, which means $s < 0.06[-]$. As explained in Chapter 7, the significant wave height and mean wave period are assumed to be fully dependent, with all points located on a steepness line ($s = 0.055[-]$). From Figure B4 it is concluded that all swell observations satisfy the steepness condition of $s < 0.06[-]$.

B.3 Marginal probability plots

The marginal distributions are fitted using the '*distribution fitter*' application in Matlab, which allows a number of distributions to be fitted with relative ease. Next, using the `fitdist()` function, the data are fitted to the distribution of choice. PWM fitting is not available for the `fitdist()` function, instead the data are fitted using a Maximum Likelihood Estimation (MLE). The probability plots of and the fitted distribution are shown in Figure B5. The parameters of the fitted marginal distributions are shown in Table B1.



(a) (b)

Figure B5: Probability plots.

Variable	Distribution type	Parameters			
swh	Generalised Pareto	$k = -0.0256$	$\sigma = 0.2491$	$\theta = 2.4000$	
*mp*1	Burr	$\alpha = 5.5345$	$c = 91.8157$	$k = 0.1552$	
shww	Normal	$\mu = 2.2062$	$\sigma = 0.4707$		
*p*1*ww*	Burr	$\alpha = 5.7019$	$c = 15.6856$	$k = 3.8921$	
shts	Weibull	$A = 1.4292$	$B = 2.5228$		
*p*1*ps*	Stable	$\alpha = 1.9470$	$\beta = 1.0000$	$\gamma = 0.3483$	$\delta = 7.7603$

Table B1: Parameters of the fitted marginal distributions.

B.4 Peak enhancement factor using the Isherwood relation

In this section, an empirical relation proposed by Isherwood (1987) is used to determine the peak enhancement factor for the wind wave- and swell partition. First the wave steepness (s) is calculated for each peak in the extreme data set using the significant wave height, mean zero-crossing period and gravitational constant as shown in B1.

$$s = \frac{2\pi H_{m0}}{g T_{m0,2}^2} \tag{B1}$$

B1 uses $T_{m0,2}$ to calculate s. The ERA5 data set provides $T_{m0,2}$ for the wind wave- and swell partition, but not for the total wave. The peak enhancement factor for the total wave is therefore not calculated using the Isherwood relation. The empirical relation between γ and s is shown in B2 and is valid for $0.03 < s < 0.15$ and $0.6 < \gamma < 8$. The resulting peak enhancement factors for wind waves and swell are shown in Figure B6.

$$\gamma = \begin{cases} 10.54 - 1.34 s^{-1/2} - \exp\left(-19 + 3.775 s^{-1/2}\right) & s \geqslant 0.037 \\ 0.9 + \exp\left(18.86 - 3.67 s^{-1/2}\right) & s < 0.037 \end{cases} \tag{B2}$$

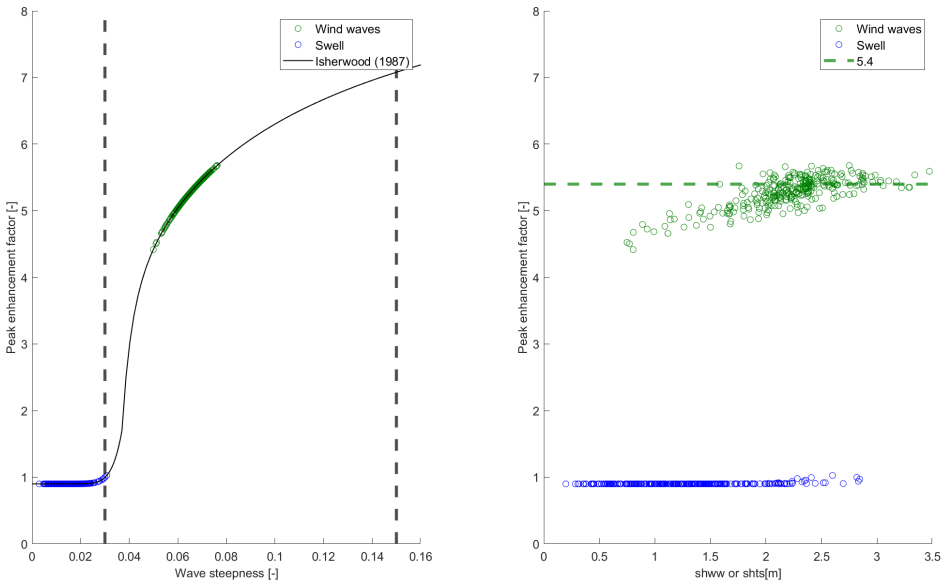


Figure B6: The empirical relation proposed by Isherwood (1987) links the wave steepness to the peak enhancement factor. This relation is used to calculate the peak enhancement factor for the wind wave- and swell partition. The steepness bounds are represented by black dotted lines in the left figure.

From Figure B6 it is concluded that the empirical relation does not provide a valid result for the swell data, as most swells fall outside the steepness bounds. Moreover, swell tends to have a higher value for γ than wind waves, somewhere in the order of 7 or 8, which is definitely not the case for the extreme data set.

One possible explanation for the unsatisfying results for the peak enhancement factor for swell might be that the parameter in the extreme data set belong to the *total* swell partition. The total swell partition might consist of multiple crossing swell systems itself for some points in time. The ERA5 reanalysis data set provides up to three (sub) partitions of the total swell. The parameters H_{m0}, $T_{m0,1}$ and θ are available for these sub partitions. Unfortunately the parameters γ, σ_θ, T_p and $T_{m0,2}$ are not available. If these sub partitions would be taken into account at a later moment in order to generate 2D wave spectra, these parameters would have to be assumed or be derived from the full 2D spectrum if available. Another possible explanation is that the relation proposed by Isherwood uses $T_{m0,2}$, which is more sensitive to errors compared to other definitions of the mean wave period.

For wind waves the empirical relation seems to provide a valid result as all peaks are within the steepness bounds. The right of Figure B6 shows the significant wave height versus the peak enhancement factor. However, a value for $\gamma = 5.4$ seems to be very high for a fully developed wind wave system. The result is possibly due to quirks in the partitioning algorithm, or may be caused by something else entirely. It is concluded the empirical relation of Isherwood does not provide satisfactory values for the peak enhancement factor for the wind wave-swell partition. The results of this method are therefore not used.

C Bivariate analysis results

C.1 Copula model adequacy

The model adequacy of a number of copulas is assessed by determining semi-correlations and the Cramér von Mises statistic.

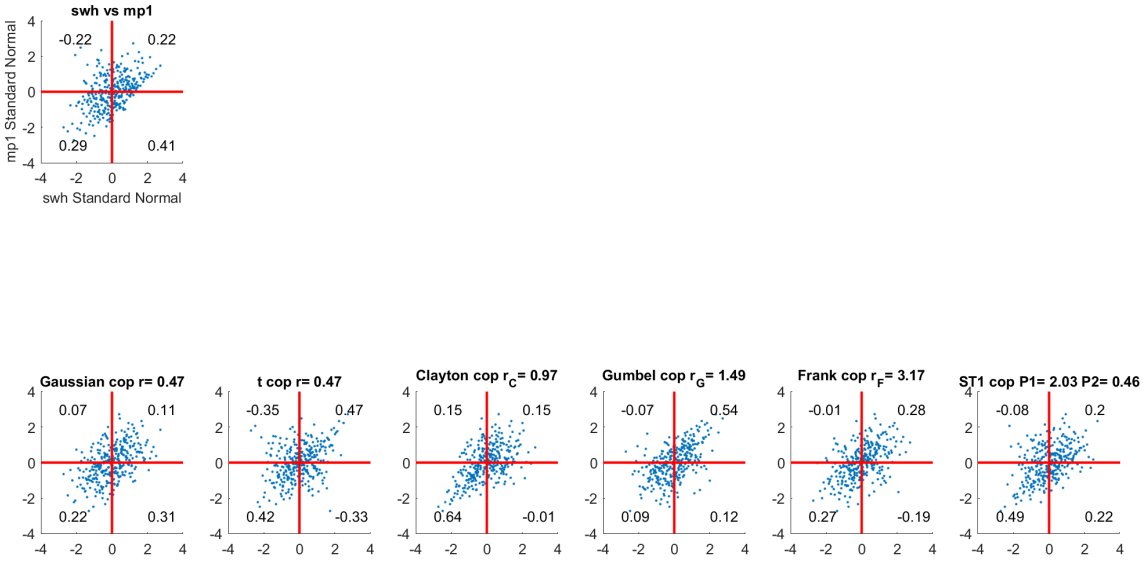


Figure C1: The extreme data of *swh* versus *mp*1 plotted in the standard normal space. Semi-correlations (which is **Pearson's** ρ and not Spearman's r) are plotted in each respective quadrant. Next a sample is taken from six different copulas and the semi-correlations are determined as well.

Another model adequacy test is the Cramér von Mises statistic (CvM), the CvM statistic indicates how far the empirical copula (based on the extreme data set) is off from the model. In general a lower value for the CvM statistic means the model is a better fit for the data. The error between the empirical data and the model is indicated with a lighter colour in Figure C2, whereas a darker colour indicates that the data and the model correspond well. For more information about the CvM statistic the reader is referred to Section 3.5.

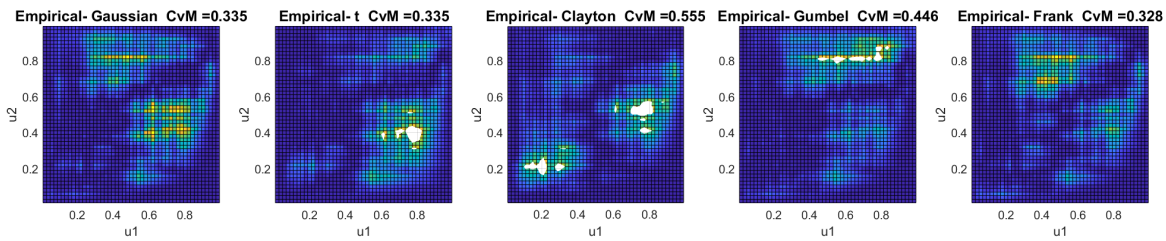


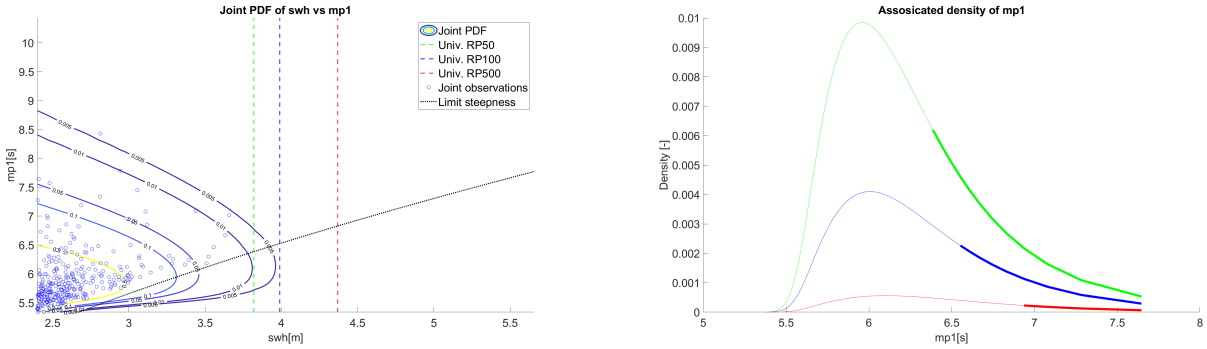
Figure C2: Cramér von Mises statistic for all copulas except the survival Tawn type 1, for which the CvM is not available.

C.2 Copula based approach without truncating

This section demonstrates why a truncation approach is necessary in order to obtain results for the mean wave period that are physically possible. Similarly to the truncation approach described in Section 6.1, the conditional design realisation (Salvadori et al., 2014) is applied in this method as well. It is stressed that a different design realisation might be preferred depending on the application.

The copula density $c(u, v)$ is used together with the marginal PDFs $f(x_1)$ and $f(x_2)$ (as shown in Figure 5.6) to construct a joint PDF for *swh* and *mp*1:

$$f(x_1, x_2) = f(x_1) \cdot f(x_2) \cdot c(u, v) \tag{C1}$$



(a) (b)

Figure C3: Joint PDF is shown as contour lines together with the univariate return values of *swh*, joint observations and the limit steepness are shown in (a). The associated densities, conditional on the univariate return values of *swh* are shown in (b), with the same colour for the respective return period. The associated densities above the limit steepness are marked with a wider line.

The corresponding values for *mp*1 are determined using the joint PDF constructed with the copula. The point with the highest associated density for *mp*1 according to Figure C3b are actually located below a steepness of $s = 0.06$, which is considered to be the limiting steepness by Holthuijsen (2007). The point with the highest associated density for *mp*1 that is physically possible for *mp*1 is thus the intersection of the steepness limit with the univariate return values of *swh*, as shown in Table C1.

ERA5	**Symbol**	**Unit**	**Design realisations**		
swh	H_{m0}	m	cond. swh=3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
*mp*1	$T_{m0,1}$	s	6.39	6.53	6.83
	s	-	0.06	0.06	0.06

Table C1: Associated values of *mp*1 conditional on three univariate return values of *swh*, together with the corresponding steepness s. Note that 50, 100 and 500 years is only the univariate return period of *swh* and not the joint return period.

The results of this approach provide an argument for the use of truncation when dealing with physical limits such as a limiting steepness. Alternatively, a different copula could also be used to model the steepness more accurately, such as the *skew-t* copula. This is also discussed in Chapter 6.

C.3 Joint return period approximation

In this section the joint return period iso-lines are approximated for the bivariate truncation approach used in Chapter 6. A large sample (10^4 realisations) from the ST1 copula is truncated and converted back to the unit square. The approximation is performed by sampling, because the copula is no longer fully parametric. Next, the empirical copula of Eq. 3.29 is used to calculate the joint exceedance probability (defined as the AND hazard scenario of Appendix A.2). The joint exceedance probability is converted back to physical units using the marginal

distributions of the variables *swh* and *mp*1. The resulting (approximated) joint exceedance iso-lines are converted to joint return periods in years and shown in Figure C4.

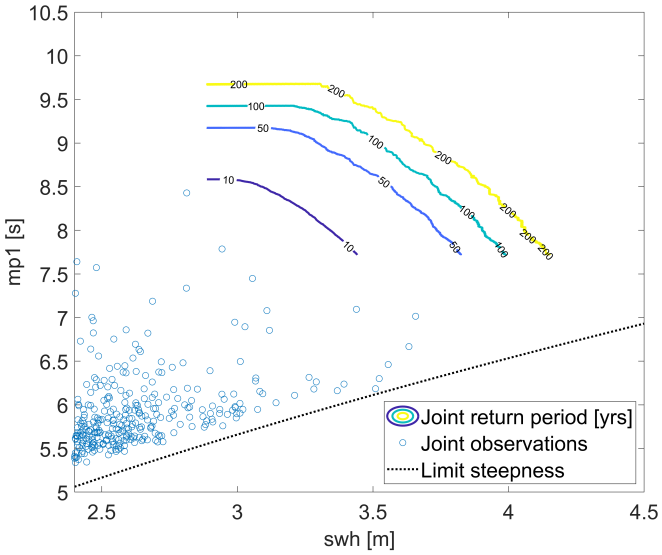


Figure C4: Joint return period iso-lines approximation using the truncated ST1 copula.

If a design realisation is required with a specified return period for exceedance, a point located on the iso-lines of Figure C4 can be chosen. The most-likely design realisation is determined by selecting a point along the iso-line that has the highest joint probability density, thus the joint PDF is required. However, as discussed in Salvadori et al. (2014), which point to choose is currently a discussion point in the scientific community and is also dependent on the application. In reference to the other design realisation chosen in this thesis:

- The conditional design realisation is applied in Chapter 6. This design realisation only uses the conditionalised joint PDF and is often preferred when one variable is considered more important in the design than the other. Choosing this design realisation is simply a choice, because an application is not considered in this thesis and there is no unique way of going from risk to design values.
- The iso-lines shown in Figure C4 can be used in combination with the joint probability density to obtain the most-likely design realisation. The most-likely design realisation is chosen when both variables (*swh* and *mp*1 in this case) are about equally important in the design. Similarly to the discussion in the multivariate context (discussed in Section 9.6.2), choosing a single combination of design values at a point of high probability density for the bivariate case is not necessarily the most unfavourable combination of variables. The same recommendation made in Section 10.1.2 is applied to the bivariate case. It is recommended to pick a number of points on the same iso-line and evaluate which combination of design variables is most unfavourable for the application.

C.4 Non-parametric approach

Additionally to the copula-based approach, a non-parametric method is used to calculate the associated values for the total mean wave period. The bivariate is density is calculated with the method described by Zachary et al. (1998). Furthermore the non-parametric approach was applied on a H_{m0}, $T_{m-1,0}$ time series by Caires and Gent (2008). In the latter paper, the conditional design realisation was used to find the associated values for $T_{m-1,0}$ conditional on univariate return values of H_{m0}. Using scripts from the ORCA toolbox in Matlab, the joint PDF is estimated for a time series. It is important to distinguish between the non-parametric approach and the ORCA approach, as the latter has some additional steps. The ORCA approach is applied to the time series in this section. The steps in the ORCA approach are described.

1. Divide data into blocks of 48 hours and take the maximum of H_{m0} and $T_{m0,1}$ in each block. This is done to ensure that observations are independent.
2. Perform univariate EVA using a POT approach on the two marginal distributions, resulting in two Generalised Pareto distributions $F_{H_{m0}}$ and $F_{T_{m0,1}}$.
3. Transformation of the marginal distributions to the standard Fréchet distribution: $\psi_{H_{m0}} = -log(F_{H_{m0}})^{-1}$ and $\psi_{T_{m0,1}} = -log(F_{T_{m0,1}})^{-1}$.
4. Each observation $(H_{m0,i}, T_{m0,1,i})$ is transformed to pseudo-radial (R) and pseudo-angular (W) components: $R_i = \psi_{H_{m0,i}}(H_{m0,i}) + \psi_{T_{m0,1},i}(T_{m0,1,i})$ and $W_i = \psi_{H_{m0,i}}(H_{m0,i})/R_i$
5. Choice of a threshold r_0 such that the conditional distribution of W for $R > r_0$ is good approximation for the theoretical limit.
6. Estimation of the density function h of all W_i above the threshold $R > r_0$. h is estimated with a kernel density estimator.
7. Finally, the joint density function is obtained by: $f(H_{m0}, T_{m0,1}) = \psi'_{H_{m0}}(H_{m0})\psi'_{T_{m0,1}}(T_{m0,1}) \cdot 2h(w)/r^3$. Where ψ'_x is the derivative of ψ_x.

Step 3 to 7 are in the non-parametric approach, whereas step 1 to step 7 are defined here as the ORCA approach. What is not mentioned in these steps is that the data set is truncated to not include swell events in the ORCA approach. This is done to achieve homogeneity in the data set and thus avoid contamination in the population. The truncation can be performed using a (crude) steepness line with all events above it being categorised as swell and thus truncated. Besides this difference, another difference is observed between the ORCA approach and the copula-based approach is the definition of what accounts as an extreme event. In the copula based approach only one threshold is used for the dominant variable (H_{m0}), the concomitant value of $T_{m0,1}$ is used that occurs during the peak value of H_{m0}. The peaks in the copula based approach are declustered with peaks occurring having to occur at least 48 hours apart in time. This is different from the ORCA approach where a POT is applied to both variables. Furthermore, the maxima of both H_{m0} and $T_{m0,1}$ are used in blocks of 48 hours. Despite these differences, the ORCA approach is applied to the ERA5 time series and the outcomes are compared to the copula based approach.

The ORCA approach results in a bivariate density function, or joint PDF, for the variables H_{m0} (swh) and $T_{m0,1}$ $(mp1)$. This joint PDF is shown in Figure C5 together with the thresholds used for the respective variables and a red arc representing the threshold $R > r_0$ chosen in step 5. swh and $mp1$ will be used from now on instead of H_{m0} and $T_{m0,1}$.

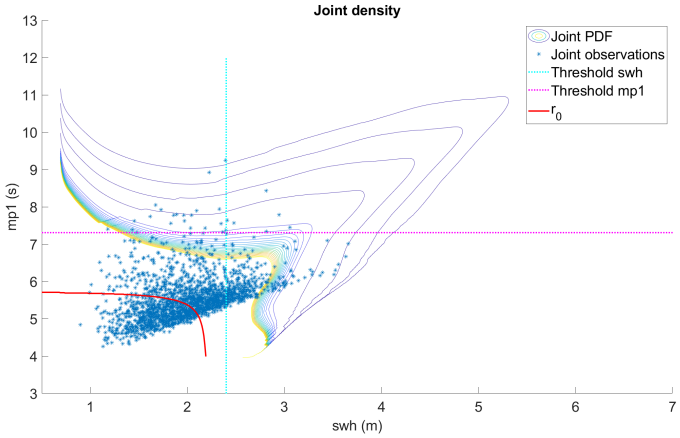
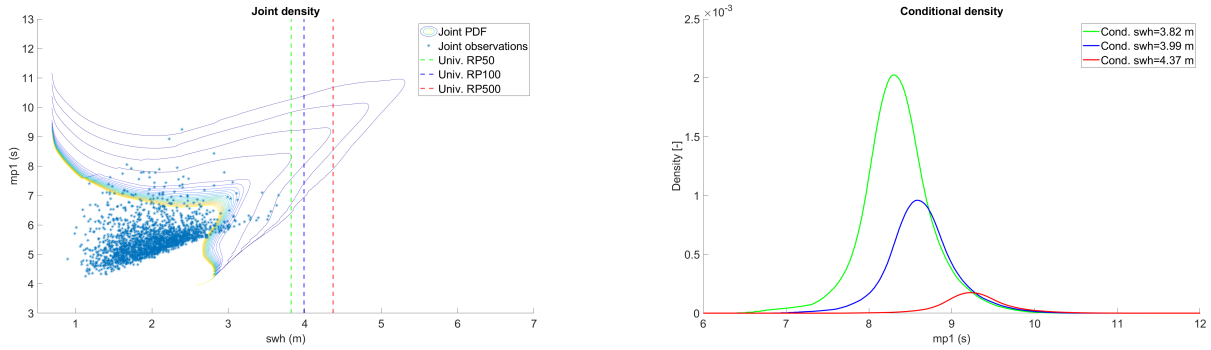


Figure C5: The joint observations plotted with the joint PDF, the univariate thresholds for swh and $mp1$ and the red arc for which $R > r_0$.

Similarly to the copula-based approach, the conditional design realisation is used in the ORCA approach. The goal is to obtain the associated values for $mp1$ given a univariate return value of the dominant variable swh. The joint PDF, together with the univariate return values for swh (dashed lines) are shown in Figure C6.



(a) (b)

Figure C6: A joint PDF of the bivariate data together with the univariate return values of swh (dashed lines) are shown in (a). The conditional densities of the mean wave period are shown in (b) conditional on three return values of the significant wave height.

Using the conditional density graphs of Figure C6 the design realisation is obtained. The value with the highest conditional density is chosen as the associated value for $mp1$ and shown in Table C2.

ERA5	**Symbol**	**Unit**	**Design realisations**		
swh	H_{m0}	m	cond. swh=3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
$mp1$	$T_{m0,1}$	s	8.31	8.59	9.25
	s	-	0.035	0.035	0.033

Table C2: Associated values of $mp1$ conditional on three univariate return values of swh, together with the corresponding steepness s. Note that 50, 100 and 500 years is only the univariate return period of swh and not the joint return period.

C.4.1 Concluding remarks

- The design realisation of the ORCA approach result in a higher value for the mean wave period (Table C2) compared to the copula-based approach with truncation and the copula-based approach without truncation (Table 6.2 and C1). However, the shape of the joint PDF of Figure C6 seems off compared to the joint observations. The joint PDF predicts that the extremes in the upper tail are more correlated, this is known as upper tail dependence. However, the joint observations suggest that there is little to no upper tail dependence between the two variables. The reason for the mismatch between the joint observations and the joint PDF is possibly because of the contamination of the data set with high swell events. The ORCA approach truncates these swell events, the non-parametric approach might therefore be better suited to deal with uncontaminated wind wave data, which often tends to show a higher correlations and upper tail dependence.
- Using the associated value for the mean wave period, the steepness is calculated. It is observed that the steepness is more or less constant for more extreme design realisations. If the same observation is true for other data sets as well, a quicker way to calculate the associated values is by just assuming the steepness is constant and find the associated value for the mean wave period with the return value for the significant wave height and a line of constant steepness. From an engineering perspective this would be a favourable approach, because it saves time in the calculation for practically the same result.
- Compared to the copula-based truncation approach the ORCA approach results in higher associated values for the mean wave period. However, because of the visual mismatch between the joint PDF and the joint observations it is concluded that the design realisations provided by the ORCA approach are too conservative (e.g. the values for the mean wave period are too high). The design realisation provided by the ORCA approach are therefore not used in this thesis.

D Multivariate analysis results

D.1 Vine copula model adequacy

D.1.1 Cramér von Mises statistic (CvM)

The Cramér von Mises statistic (CvM) of vine 1 and 4 is shown in Figure D1 and D2 respectively. The CvM statistic for the vine copula model is a summation of the CvM statistic of three unconditional bivariate edges, namely all bivariate combinations possible between the parameters. For vine 1 these combinations are: *swh-shts*, *swh-p1ps* and *shts-p1ps*. For vine 4 these combinations are: *swh-shww*, *swh-p1ps* and *shww-p1ps*.

The CvM statistic indicates how far the empirical copula (based on the extreme data set) is off from the model. In general a lower value for the CvM statistic means the model is a better fit for the data. The error between the empirical data and the model is indicated with a lighter colour, whereas a darker colour indicates that the data and the model correspond well. For more information about the CvM statistic the reader is referred to Section 3.5.

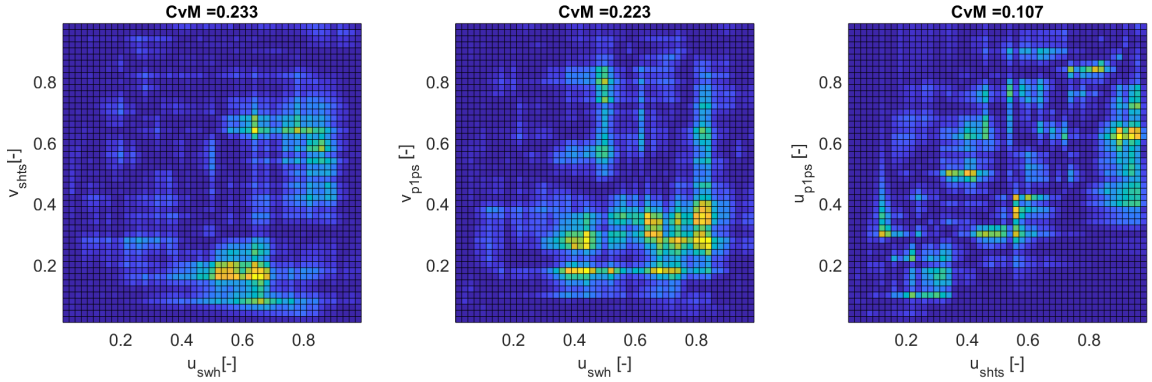


Figure D1: Vine 1 Cramér von Mises statistic: CvM=0.5629 [-].

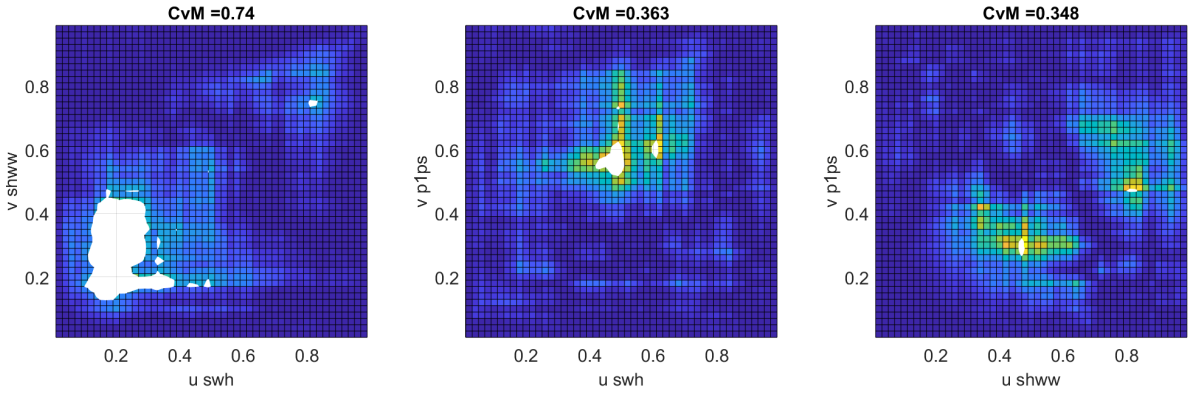


Figure D2: Vine 4 Cramér von Mises statistic: CvM=1.4509 [-].

D.1.2 Comparing vine samples to observations

Vine 1 and vine 4 are sampled and compared to the joint observations in physical units, the unit square and the standard normal space. Kendall's τ is calculated for each pair of variables of the sample (τ_{V1} or τ_{V4}) and the extreme data set (τ_{ex}) and shown in each subplot.

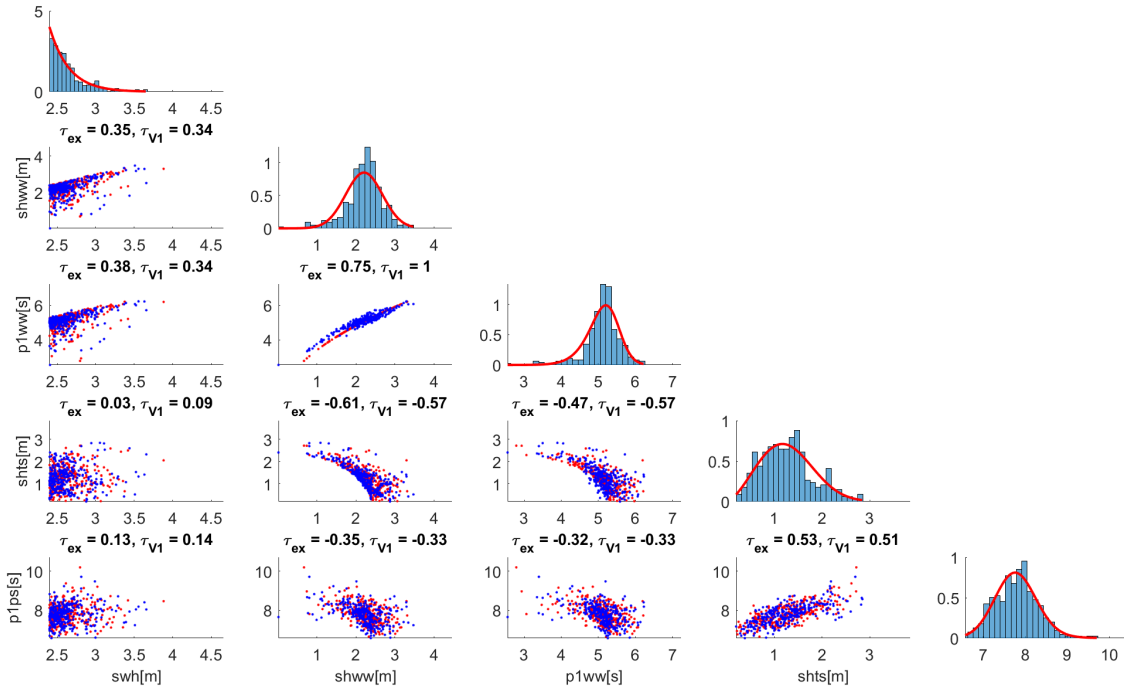


Figure D3: Vine 1 sample (red) and joint observations (blue) in physical units.

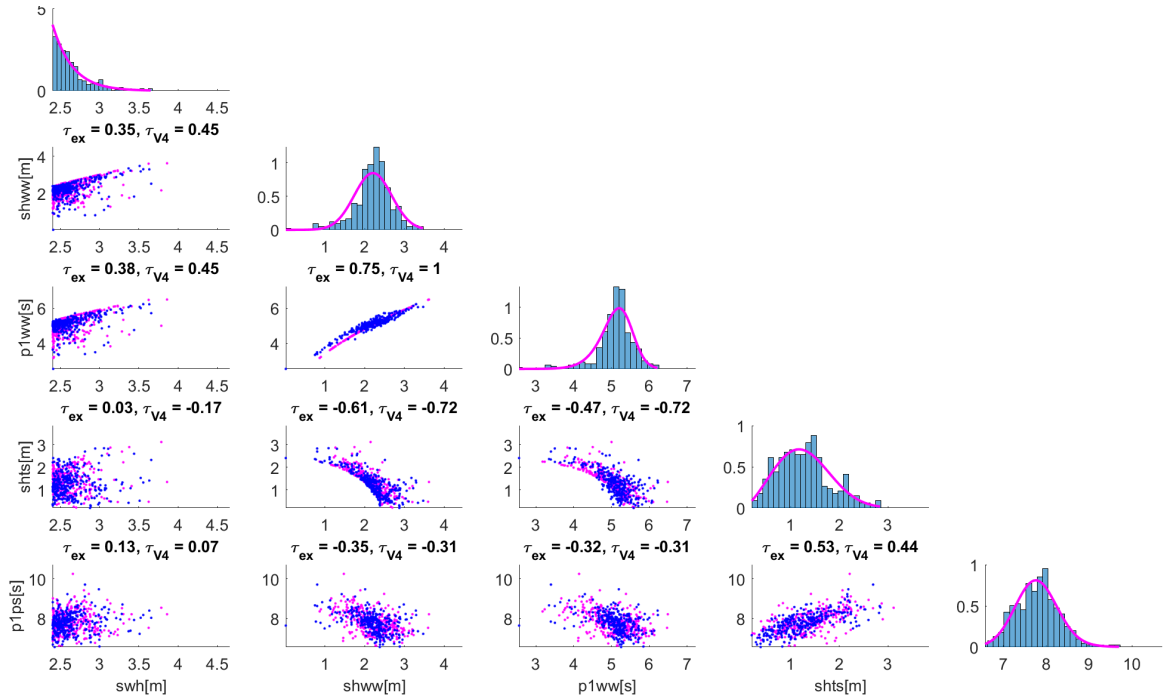


Figure D4: Vine 4 sample (magenta) and joint observations (blue) in physical units.

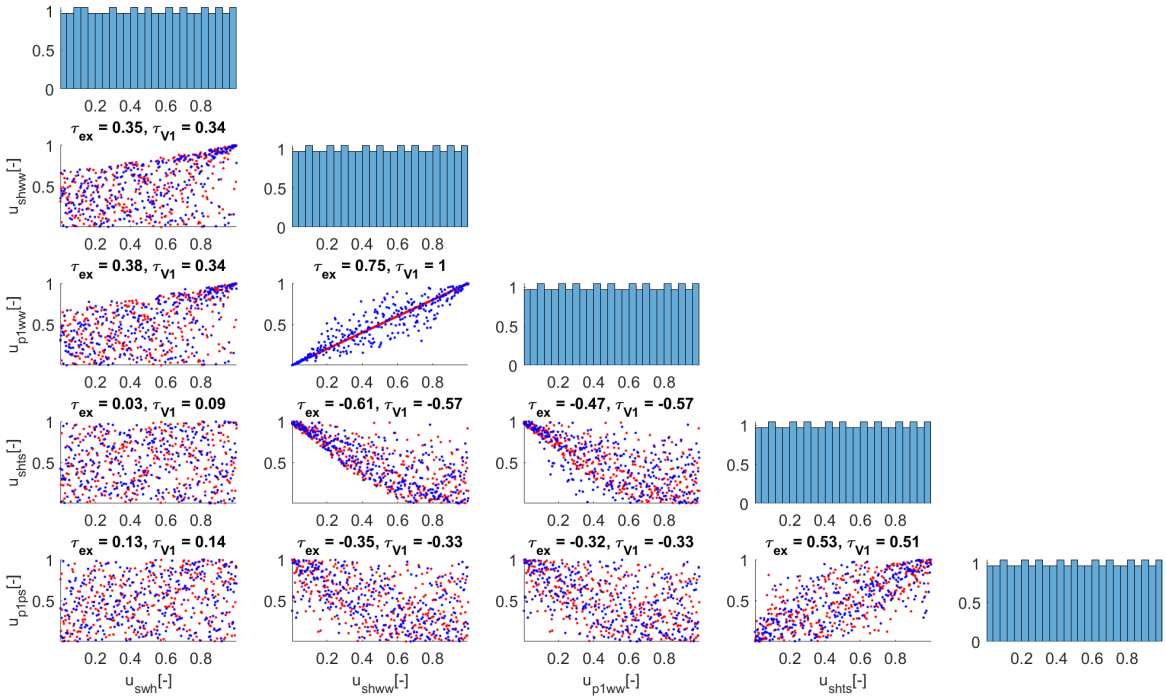


Figure D5: Vine 1 sample (red) and joint observations (blue) in the unit square.

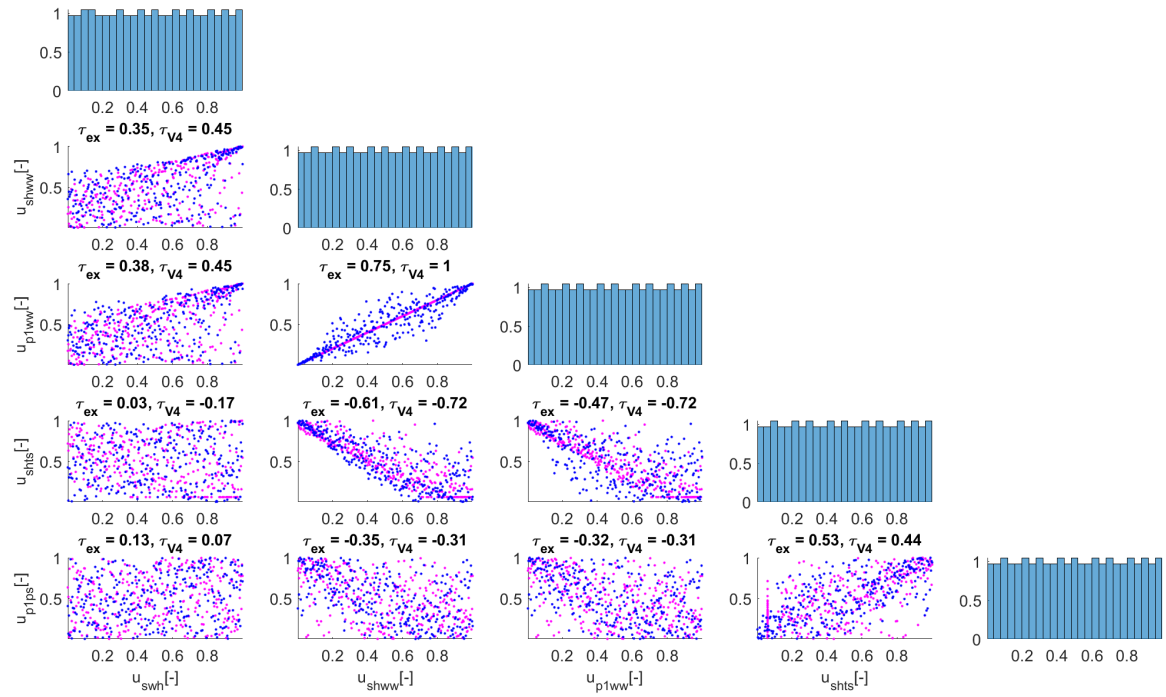


Figure D6: Vine 4 sample (magenta) and joint observations (blue) in the unit square.

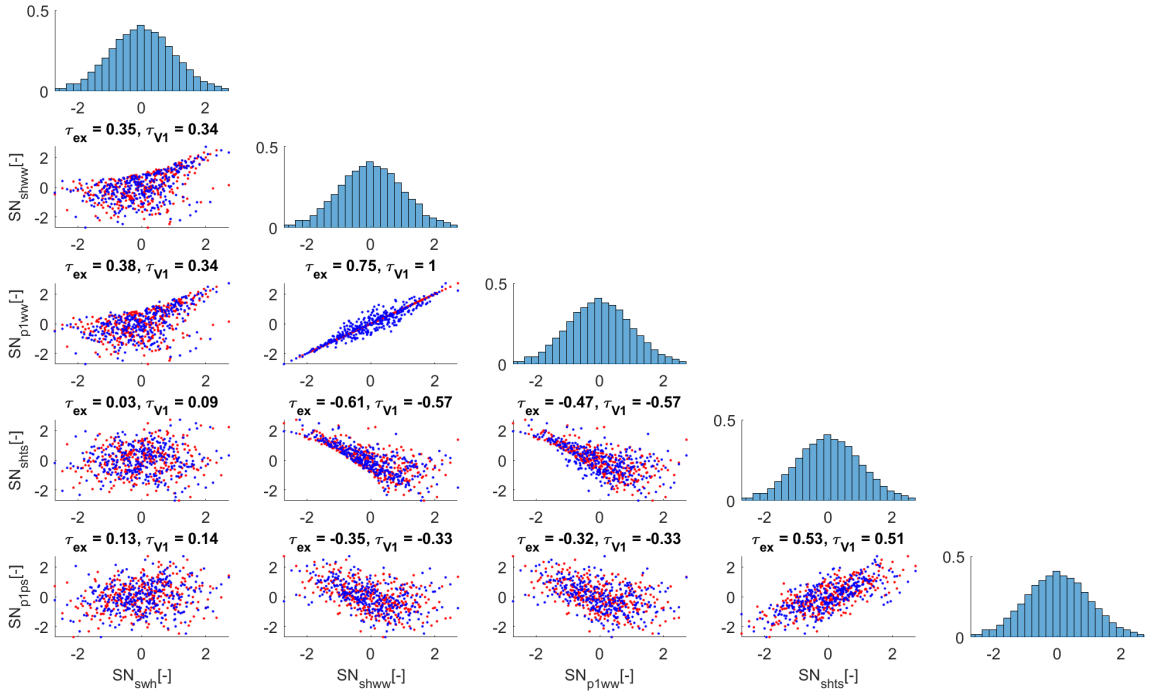


Figure D7: Vine 1 sample (red) and joint observations (blue) in the standard normal space.

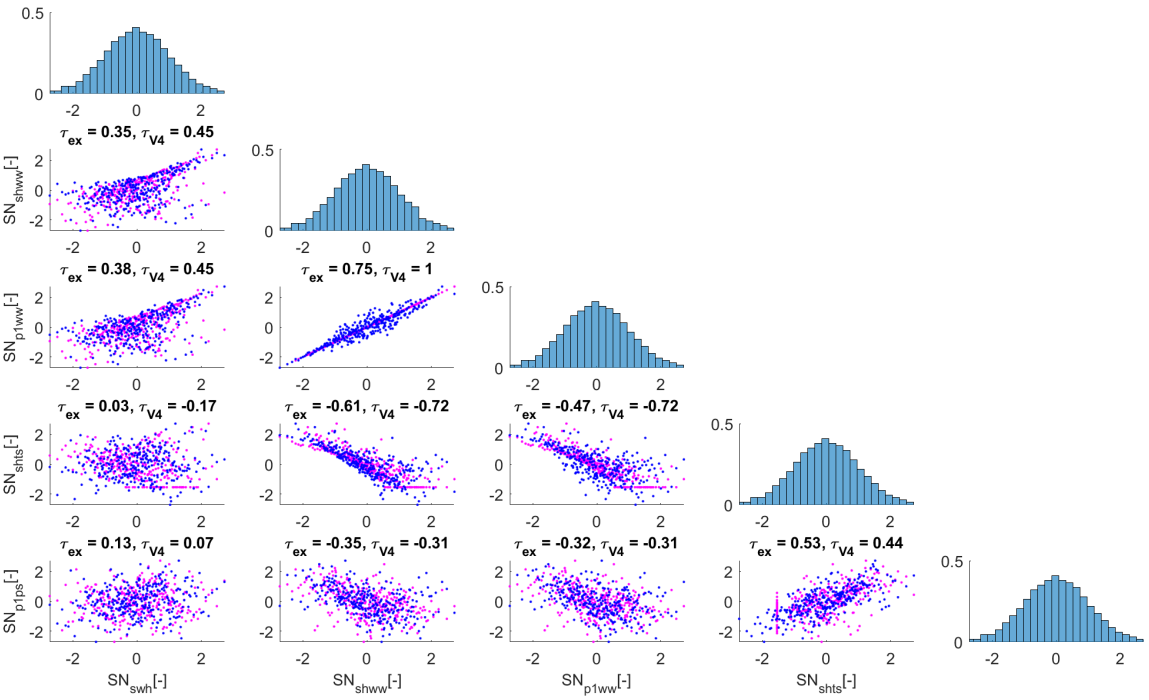


Figure D8: Vine 4 sample (magenta) and joint observations (blue) in the standard normal space.

D.1.3 Joint observations check

Variable (unit)	Vine 1	Vine 4	4th highest event
$swh(m)$	cond. swh=3.44 (10 yr)	cond. swh=3.44 (10 yr)	3.52
$shww(m)$	3.24	2.92	3.17
$p1ww(s)$	6.14	5.83	6.00
$shts(m)$	1.16	1.82	1.53
$p1ps(s)$	7.96	7.63	8.04

Table D1: Points of highest density of wind wave and swell parameters of Vine 1 and Vine 4, conditional on a return values of *swh* compared to the fourth highest observation in the (40 year long) time series. Note that 10 years is only the univariate return period of *swh* and not the joint return period.

Variable (unit)	Vine 1	Vine 4	highest event
$swh(m)$	cond. swh=3.82 (50 yr)	cond. swh=3.82 (50 yr)	3.66
$shww(m)$	3.64	3.25	2.51
$p1ww(s)$	6.51	6.15	5.71
$shts(m)$	1.16	2.01	2.32
$p1ps(s)$	7.96	7.64	8.04

Table D2: Points of highest density of wind wave and swell parameters of Vine 1 and Vine 4, conditional on a return values of *swh* compared to the highest observation in the (40 year long) time series. Note that 50 years is only the univariate return period of *swh* and not the joint return period.

D.1.4 Percentiles

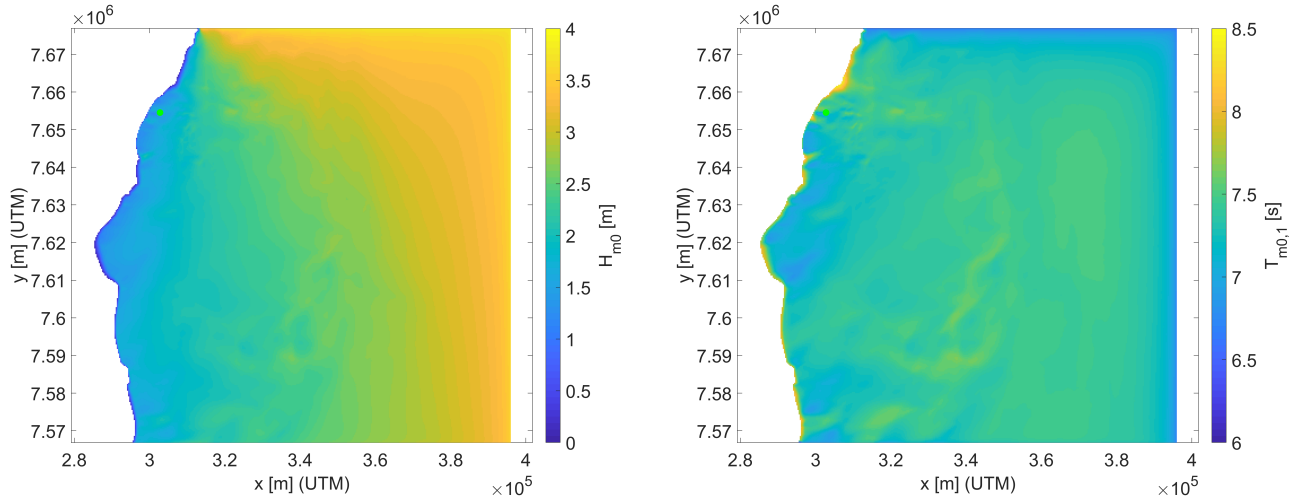
Variable (unit)				
$swh(m)$	cond. swh=3.44 (10 yr)	3.82 (50 yr)	3.99 (100 yr)	4.37 (500 yr)
Vine 1: $shww(m)$	2.67 - **3.24** - 3.41	3.15 - **3.64** - 3.79	3.39 - **3.82** - 3.96	3.90 - **4.23** - 4.35
Vine 4: $shww(m)$	1.87 - **2.75** - 3.08	1.91 - **3.10** - 3.36	1.91 - **3.23** - 3.45	**3.36**
Vine 1: $p1ww(s)$	5.58 - **6.14** - 6.30	6.05 - **6.51** - 6.64	6.28 - **6.67** - 6.80	6.74 - **7.02** - 7.12
Vine 4: $p1ww(s)$	4.66 - **5.66** - 5.99	4.71 - **6.01** - 6.26	4.72 - **6.13** - 6.34	**6.26**
Vine 1: $shts(m)$	0.46 - **1.16** - 2.16	0.46 - **1.16** - 2.16	0.45 - **1.14** - 2.10	0.41 - **1.08** - 1.96
Vine 4: $shts(m)$	1.52 - **2.06** - 2.89	1.81 - **2.23** - 3.31	2.00 - **2.35** - 3.50	**2.79**
Vine 1: $p1ps(s)$	7.21 - **8.00** - 8.78	7.40 - **8.16** - 8.93	7.46 - **8.19** - 8.94	7.54 - **8.23** - 8.98
Vine 4: $p1ps(s)$	7.23 - **7.98** - 8.78	7.23 - **7.98** - 8.78	7.26 - **8.03** - 8.84	7.34 - **8.15** - 9.00

Table D3: Percentiles for wind wave and swell parameters of Vine 1 and Vine 4, conditional on the return values of *swh*. The 5th, **50th** and 95th percentiles are calculated based on the conditional margins of *shww*, *shts* and *p1ps*. Note that 10, 50, 100 and 500 years is only the univariate return period of *swh* and not the joint return period.

E SWAN output

E.1 Surface maps of H_{m0} and $T_{m0,1}$

The surface maps shown in Figure E1 are used for illustration purposes and to validate the model.



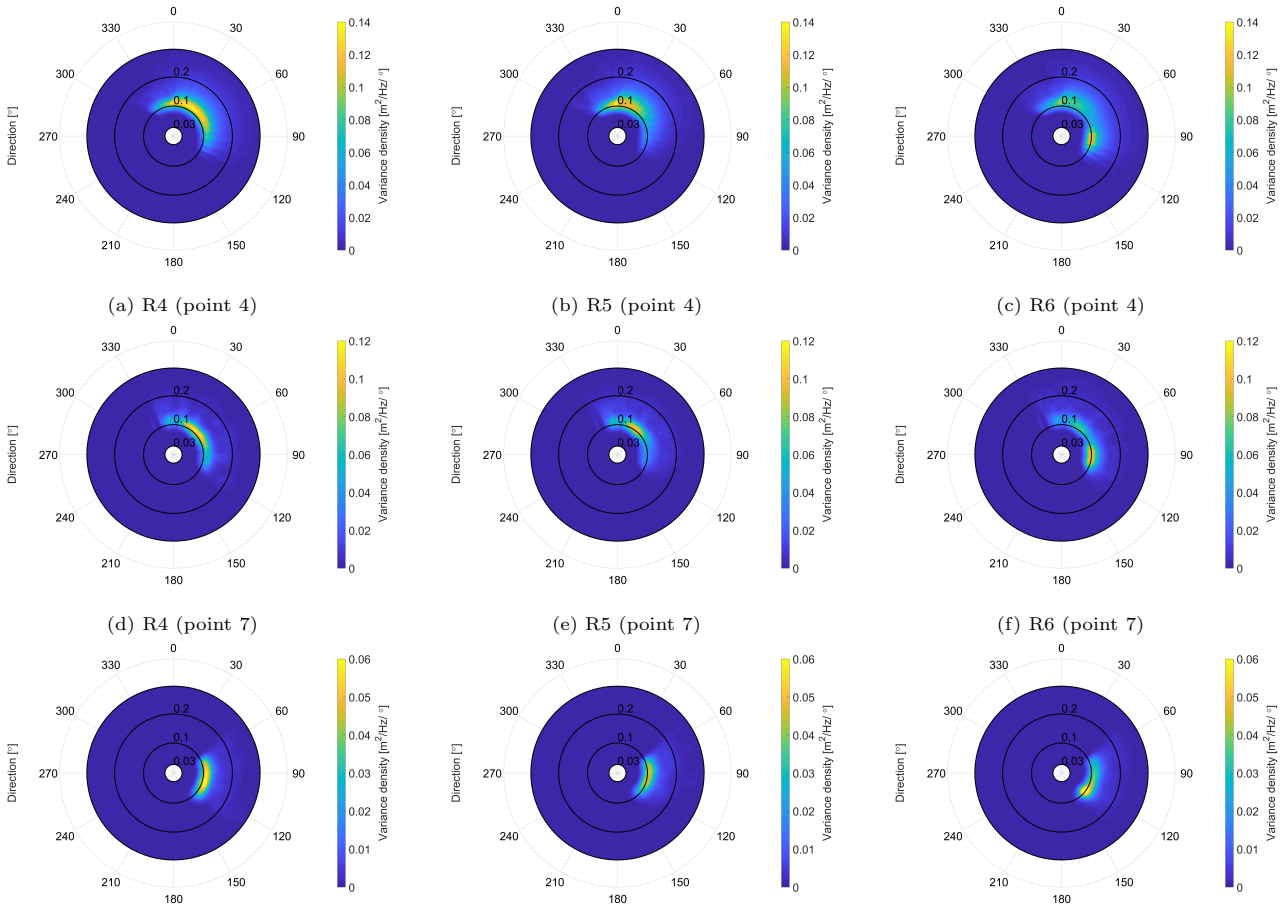
(a) (b)

Figure E1: For the boundary conditions labelled as R1 in Table 8.1 the total significant wave height is shown in (a) and the mean wave period is shown in (b). The nearshore point of interest is again plotted as a green dot.

E.2 Variance density spectra for R4 to R9

The variance density spectra for boundary conditions R4 to R9 of Table 8.1 and 8.2 are shown in this section. The output for boundary conditions R1, R2 and R3 is shown in Chapter 8. The 2D variance density spectra are calculated for the output points 4, 7 and 1 of Figure 8.2. Point 4 represents an offshore location near the boundary of the model with a local depth of 1246 m. As stated previously, point 1 is the nearshore location of interest with a local depth of 9.1 m. Finally, point 7 is a located between the offshore and nearshore point and has a local depth of 25.5 m. The (2D) variance density spectra at these points are shown for R4 to R9 to illustrate the propagation of the wave energy to the nearshore. The variance density is calculated for 25 frequency bins from $0.03 < f < 0.3$ Hz, which corresponds to wave periods between 3.33 and 33.33 s. The directions of the 2D spectra are divided into 144 bins, which corresponds to 2.5 °.

Additionally, the 1D variance density spectra are obtained by integrating the 2D variance density spectra over all directions using Eq. 3.15. The figures in this section show how wave energy dissipates as it propagates to the nearshore.

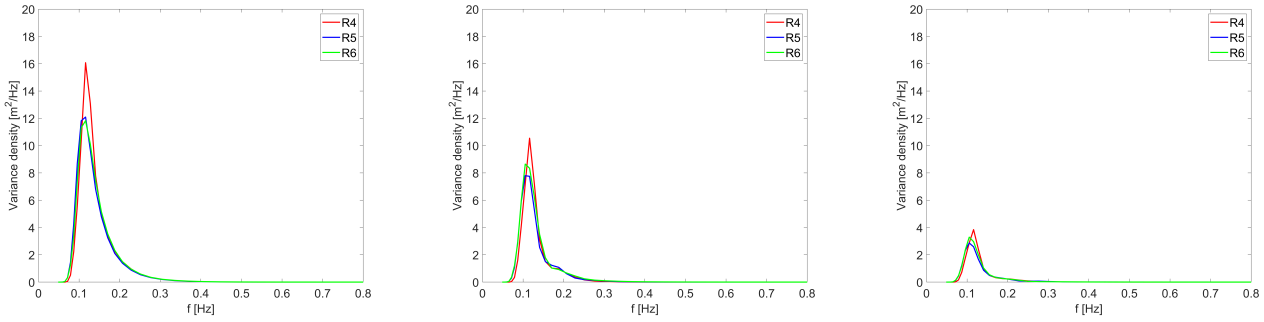


(a) R4 (point 4) (b) R5 (point 4) (c) R6 (point 4)

(d) R4 (point 7) (e) R5 (point 7) (f) R6 (point 7)

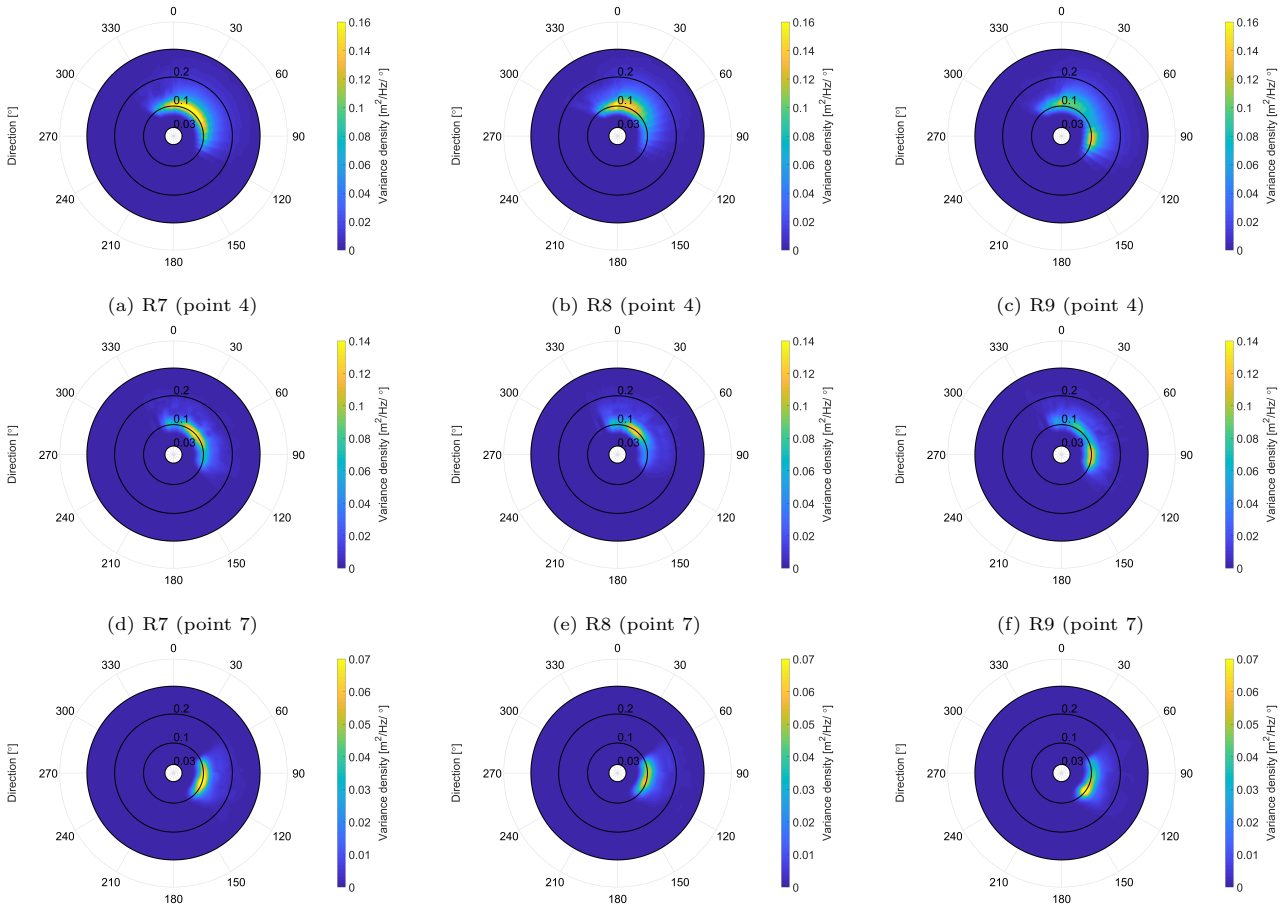
(g) R4 (point 1) (h) R5 (point 1) (i) R6 (point 1)

Figure E2: The variance density spectra given boundary conditions R4, R5 and R6 (from Table 8.1 and 8.2). The spectra are shown for an offshore point (point 4), a point between the offshore and nearshore point (point 7) and a point nearshore (point 1). The locations of point 4, 7 and 1 are shown in Figure 8.2. The frequencies of 0.03, 0.1, 0.2 and 0.3 Hz are shown as black circles.



(a) Offshore (point 4) (b) Between offshore and nearshore (point 7) (c) Nearshore (point 1)

Figure E3: 1D variance density spectra for the boundary conditions R4, R5 and R6 for point 4, 1 and 7 of Figure 8.2.

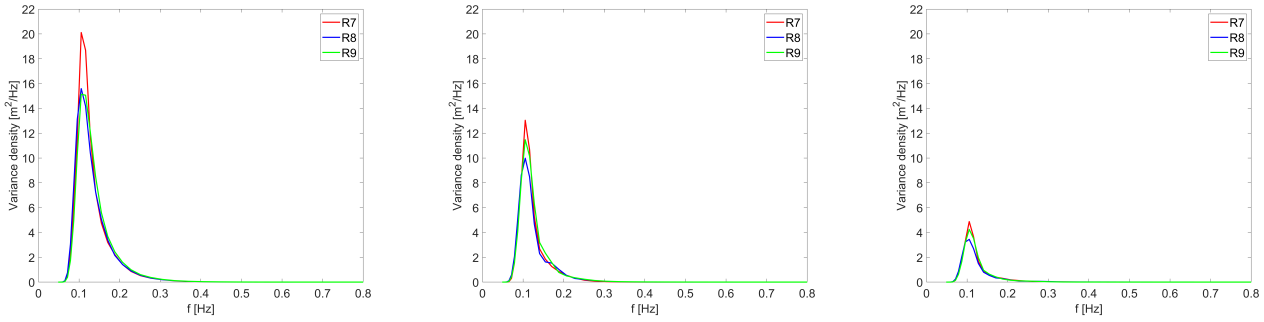


(a) R7 (point 4) (b) R8 (point 4) (c) R9 (point 4)

(d) R7 (point 7) (e) R8 (point 7) (f) R9 (point 7)

(g) R7 (point 1) (h) R8 (point 1) (i) R9 (point 1)

Figure E4: The variance density spectra given boundary conditions R7, R8 and R9 (from Table 8.1 and 8.2). The spectra are shown for an offshore point (point 4), a point between the offshore and nearshore point (point 7) and a point nearshore (point 1). The locations of point 4, 7 and 1 are shown in Figure 8.2. The frequencies of 0.03, 0.1, 0.2 and 0.3 Hz are shown as black circles.



(a) Offshore (point 4) (b) Between offshore and nearshore (point 7) (c) Nearshore (point 1)

Figure E5: 1D variance density spectra for the boundary conditions R7, R8 and R9 for point 4, 1 and 7 of Figure 8.2.

F ERA5 parameters

ERA5 notation	Literature notation	Unit	Description
swh	H_{m0}	m	total significant wave height
mwd	θ	°	total mean wave direction (clockwise from the north)
mp1	$T_{m0,1}$	s	total mean wave period based on the first order spectral moment
pp1d	T_p	s	total peak wave period based on a 1D spectrum
wdw	σ_θ, DSPR	rad, °	total wave directional width or directional spreading
wsp	γ	-	total wave spectral peakedness or peak enhancement factor
shww	H_{m0}	m	wind wave partition significant wave height
mdww	θ	°	wind wave partition mean wave direction (clockwise from the north)
p1ww	$T_{m0,1}$	s	wind wave partition mean wave period based on first spectral moment
p2ww	$T_{m0,2}$	s	wind wave partition mean wave period based on second spectral moment
dwww	σ_θ, DSPR	rad, °	wind wave partition directional width or directional spreading
shts	H_{m0}	m	swell partition significant wave height
mdts	θ	°	swell partition mean wave direction (clockwise from the north)
p1ps	$T_{m0,1}$	s	swell partition mean wave period based on first spectral moment
p2ps	$T_{m0,2}$	s	swell partition mean wave period based on second spectral moment
dwps	σ_θ, DSPR	rad, °	swell partition directional width or directional spreading
wind	U_{10}	m/s	Wind speed 10 m above the surface
dwi	θ_w	°	Wind direction (clockwise from the north) 10 m above the surface