Predicting Coastal Hazards with a Bayesian Network

A probabilistic approach to creating an Early Warning System for marine coastal hazards for Praia de Faro, Portugal

MSc. Thesis Laurens Poelhekke

Delft University of Technology
Faculty of Civil Engineering and Geosciences
Department of Hydraulic Engineering

September 2015

Report type
MSc. Thesis Delft University of Technology

Title
Predicting Coastal Hazards with a Bayesian Network

Subtitle
A probabilistic approach to creating an Early Warning System for marine coastal hazards for Praia de Faro, Portugal

Date
September 16th, 2015

Author

Name:	Laurens Poelhekke
Student no:	1382888
Faculty	Civil Engineering and Geosciences
Program/Track	Hydraulic Engineering, Master of Science Civil Engineering
Specialization	Coastal Engineering

Graduation Committee

Chairman	Prof. Dr. Ir. M.J.F. Stive	Coastal Engineering, Delft UT
First supervisor	Dr. Ir. A. van Dongeren	Coastal Morphology, Deltares
Supervisor	Prof. Dr. Ir. O. Ferreira	CIMA, University of the Algarve
Supervisor	Prof. Dr. Ir. A.J.H.M. Reniers	Fluid Mechanics, Delft UT
Supervisor	Dr. Ir. C. den Heijer	Coastal Engineering, Delft UT & Deltares
Supervisor	MSc. W.S. Jäger	Coastal Engineering, Delft UT

In collaboration with
Deltares Deltares Enabling Delta Life University of the Algarve UAlg UNIVERSIDADE DO ALGARVE

ABSTRACT

Recent and historic events have demonstrated the European vulnerability to coastal floods. Larger and more extreme events in Asia and the Americas have shown the devastating effects that these low-frequency high-impact floods can have. In Europe this has given rise to a large integral research project RISC-KIT (Resilience Increasing Strategies for Coastlines toolKIT), that aims to develop a set of tools to reduce the coastal vulnerability. Part of this is an operational Early Warning System (EWS) for hot spot areas to which this thesis is a contribution.

Coastlines with sandy shores (beaches and dunes) have a large response to these low-frequency high impact events (e.g. beach and dune erosion). Houses and infrastructure close to the shoreline are not only vulnerable to flooding but also to damage caused by overwash and coastal erosion. Furthermore, the pathways of floods in sandy areas are subject to morphological changes. Current flood hazard models do not incorporate morphological processes and are therefore not appropriate to model coastal hazards at sandy coastlines. A process based model such as XBeach is capable of modeling the coastal response in two dimensions (2D), however, coming at the cost of much longer computational runtimes making it not a very useful part of an EWS.

A solution is found in utilizing a probabilistic model as a surrogate for a process based model. The probabilistic model is fed with data created with the process based model. The process based model essentially trains the probabilistic model allowing it to take over its function. More specifically a Bayesian Network (BN) has been trained to replace an XBeach model. BNs are generally graphical probabilistic models that have been successfully used in the past for a wide variety of purposes, including coastal modeling. In a BN the offshore boundary conditions, describing a storm (e.g. wave height and water level), and onshore hazard intensities (overwash and erosion) are treated as random variables and are connected through conditional dependencies. This has led to the main research question:

How can Bayesian Networks be used as part of an early warning system for spatially varying coastal hazards at sandy coastlines?

Praia de Faro, located in the south of Portugal, has been chosen as a case study site for this research. This small settlement is situated on top of a sandy barrier and is exposed to yearly recurring overwash events and has experienced damages to houses and infrastructure due to severe erosion of the coastline. In developing the concept for the case study site there are three main issues that had to be overcome: (1) A BN requires a lot of data to be trained well and the higher the complexity of a BN the more training data is required, (2) XBeach is a computationally expensive model posing a barrier to generating lots of training data and (3) there is no existing storm dataset that is large enough to train a BN with to be able to act instead of an XBeach model.

The first issue is solved by setting up a BN with very low complexity. The constructed BN consists of six random variables and predicts overwash and coastal erosion based on inputs for the offshore wave height and water level of a storm and a location at Praia de Faro. The second issue has been solved by reducing the number of grid cells in the XBeach model to a quarter of the original size while still producing good results. The third problem has been solved by creating a synthetic dataset using a local storm dataset existing of recorded wave heights, periods, surge levels and storm durations. Copulas have been used to create bivariate distributions of the variables pairs, which

could then be used to sample a new, synthetic, dataset. The advantage of using the bivariate distributions is that the synthetic dataset mimics the characteristics of the original dataset and therefore includes the natural variability. Three hundred storms, sampled from the synthetic dataset, have been used to force an XBeach model that has been set up for Praia de Faro. Of each storm data is extracted from XBeach creating a dataset of cases. This dataset is fed into the BN, giving it knowledge to base future predictions on.

The BN is tested by training it with 90% of the data and then testing how well it can predict the remaining 10% of the dataset. A number of test results give insight in the performance of the BN. The Mean Absolute Error (MAE) of the mode of the predictions are large for all hazard nodes. For the coastal erosion, overwash depth and overwash velocity the MAEs are respectively: 19 m, 0.4 m and 0.9 m/s, with standard deviations of, respectively, 27.5 m, 1.2m and 1.7 m/s. With respect to predicting coastal erosion and overwash as part of an EWS these errors are too large. However, only the mode of the prediction is considered for the MAE. The predictions are often bi-modal, creating very large errors. The bimodality has the result that the MAE often over or underestimates the real error of a single prediction and is therefore only an indicator of the overall performance. Other results indicate that where the BN is not very good at predicting the magnitude of erosion and overwash, it is good at predicting whether or not the erosion or overwash will affect the houses and infrastructure.

Considering the research question, a BN is able to translate offshore storm boundary conditions to onshore hazard intensities. It can, however, only do this with relatively low complexity. It therefore loses a lot of spatial information that the process based model XBeach gives. The (large) natural variability that is seen in the storm dataset is passed on to the predictions of the BN, giving wide predictions. The BN in its current setup needs to be improved in order to be useful in an operational EWS. It will either need extra training data or increased complexity which implicitly also means more training data is needed. Proposed is to add an extra variable in the BN, the peak period of the waves, as this variable shows the least correlation with the wave height in the actual storm dataset.

ACKNOWLEDGEMENTS

This thesis has been carried out as the final chapter of the Master of Science program Hydraulic Engineering at the Faculty of Civil Engineering and Geosciences at the Delft University of Technology. The research has been hosted and facilitated by Deltares in Delft, the Netherlands, and the University of the Algarve in Faro, Portugal.

I am most grateful to my graduating committee, consisting of Marcel Stive, Oscar Ferreira, Ad Reniers, Ap van Dongeren, Kees den Heijer and Wiebke Jäger. Their expertise, commitment, interest, feedback and critical minds have guided me and contributed to better results and representation of this research.

I would like to express special gratitude towards Ap (Deltares) and Oscar (UAlg) for providing the topic and hosting the research. Additionally I would like to thank Haris Plomaritis for his help during my stay in Portugal. Further I would like to thank all staff of the CIMA research group for making my three months in Portugal unforgettable!

Special thanks go to Ap en Wiebke who supported me during all phases of the process. Thank you Wiebke for patiently guiding me through the mathematics of this research and for all the fruitful discussions we have had.

I would also like to thank Huub van Verseveld who introduced me to this topic and the people and for helping me, in especially the first phases, of the project. Lastly, I would like to thank Joost den Bieman en Robert McCall for always finding the time to answer my questions concerning the XBeach modeling.

TABLE OF CONTENTS

TABLE OF APPENDICES

1 INTRODUCTION

Over the past decades a number of coastal floods have demonstrated the vulnerability of the coastal zone in Europe. Examples are the North Sea Flood of 1953 in the Netherlands, Belgium and the United Kingdom, the Ligurian Flash Floods of 2011 (in Italy), and the flooding of a part of France due to the storm Xynthia in 2010. Larger and more extreme events such as the hurricanes in the United States of America (Katrina 2004, and Sandy 2012) and Typhoons in Asia (Haiyan in 2013 and Nargis in 2008) have shown the devastating effects of these low-frequency, high impact flood events. These recent historic events, climate change and an increase in economic and social activity in coastal and riverine areas have given rise for the European Union (EU) to issue a flood directive and for the United Nations (UN) to formulate the Hyogo Framework of Action (HFA).

As an answer to the EU Flood directive and the HFA a large research project has been set up between universities, research institutions and municipal authorities across different member states of the EU. The project, abbreviated as RISC-KIT, aims to reduce risk and increase resilience of coastal zones against these low frequency, high impact meteorological events (Van Dongeren et al., 2014).

The RISC-KIT project aims to achieve this goal by developing new methods tailored for the European coasts. This thesis contributes to the development of an Early Warning System (EWS) for flood events. The relevance of an EWS is supported by the UN who have identified it as key in reducing casualties and economic losses due to flood events (UNISDR, 2002).

For coastlines with sandy shores (beaches and dunes) the response of the coastline to high impact events is very large (e.g. dune and beach erosion). Current flood hazard models do not include these morphological processes and are therefore not sufficient for these types of coastlines. Process based models such as XBeach are capable of modeling the coastal response but are not very useful as an EWS due to the long duration of the computations.

A solution is found in utilizing a probabilistic model as a surrogate for a process based model. The probabilistic model is fed with data that is created with the process based model. This way the probabilistic model gains the same knowledge about the processes within the bounds that it is trained. More specifically a Bayesian Network (BN) will be created as a surrogate for an XBeach model. A BN is in essence a graphical probabilistic model consisting of random variables and conditional dependencies between these variables. The BN treats the hazard intensities (e.g. water levels and flow velocities) and the forcing (e.g. storm surge, wave heights and periods) as random variables. The conditional dependencies between the random variables can be determined by the BN if it is trained with enough data. The training data is produced with an XBeach model of the coastline. Since a BN is solely based on statistics its runtime is very low and therefore able to function as an EWS if trained well.

As a case study site for the development and initial implementation the beach settlement of Faro, Praia de Faro, has been selected. Faro lies in the South of Portugal and is the regional capital of the Algarve, with its airport situated on the coastal lowlands. The region is a very important tourist destination and essential for the Portuguese economy. Praia de Faro itself is pestered by yearly recurring overwash events and has experienced damages to houses and infrastructure due to severe erosion of the coastline.

1.1 ORGANIZATION OF THE REPORT

This report is structured following the research process. Parts of the report have been moved to the appendix to which is referred as appropriate. The problem definition and research question are defined in chapter 2 and preceded by a section with a section containing relevant background information to the subject. A coastal analysis of the case study site is presented in chapter 3. The development and implementation of the EWS, including the setup of the XBeach model and BN, are extensively described in chapter 4. In chapter 5 the predictive skill of the BN is analyzed and results are presented. The report is concluded with the conclusions and recommendations in chapters 6 and 7.

2 PROBLEM DEFINITION

In this chapter the research problem will be formulated in several steps. First a section with relevant background information is presented, followed by the significance of the research and the problem definition. The research question is delineated by setting a main objective which is hence specified in several sub objectives. Lastly the research approach is presented.

2.1 BACKGROUND INFORMATION

2.1.1 RISC-KIT: RESILIENCE-INCREASING STRATEGIES FOR COASTS – TOOLKIT

The EU-Flood directive of 2007 requires member states to create a flood risk management plan with the focus on prevention, protection and preparedness to be completed before the end of 2015. Within the directive flood risk is defined as "the combination of the probability of a flood event and the adverse consequences for human health, the environment, cultural heritage and economic activity associated with the event." (EU, 2007) The Hyogo Framework for Action 2005-2015 from the UN Office for Disaster Risk Reduction (UNISDR) proposes 5 priority actions with the goal to increase resilience of nations against disasters, including floods.

For the purpose of fulfilling the goals set by the EU Flood Directive and the HFA the project Resilience – Increasing Strategies for Coasts – toolKIT (RISC-KIT), has been set up in cooperation between several research institutions, universities and municipal authorities of different member states of the EU. The main objective of the RISC-KIT project is given below, as stated in Van Dongeren et al. (2014):

> *"The main objective of Resilience-Increasing Strategies for Coasts – toolKIT (RISC-KIT) is to develop methods, tools and management approaches to reduce risk and increase resilience to low-frequency, high-impact hydro-meteorological events in the coastal zone. These products will enhance forecasting, prediction and early warning capabilities, improve the assessment of long-term coastal risk and optimise the mix of prevention, mitigation and preparedness measures."*

This objective is further specified into seven subgoals. Number four and six of these goals govern the subject of the proposed thesis and are given below:

> *"4. Development of an impact-oriented Early Warning and Decision Support System (EWS/DSS) for hot spot areas consisting of: i) a free-ware system to predict hazard intensities using coupled hydro-meteo and morphological models and ii) a Bayesian-based Decision Support System which integrates hazards and socio-economic, cultural and environmental consequences;"*

> *"6. Application of CRAF and EWS/DSS tools at the case study sites to test the DRR plans for a combination of scenarios of climate-related hazard and socio-economic vulnerability change and demonstration of the operational mode;"*

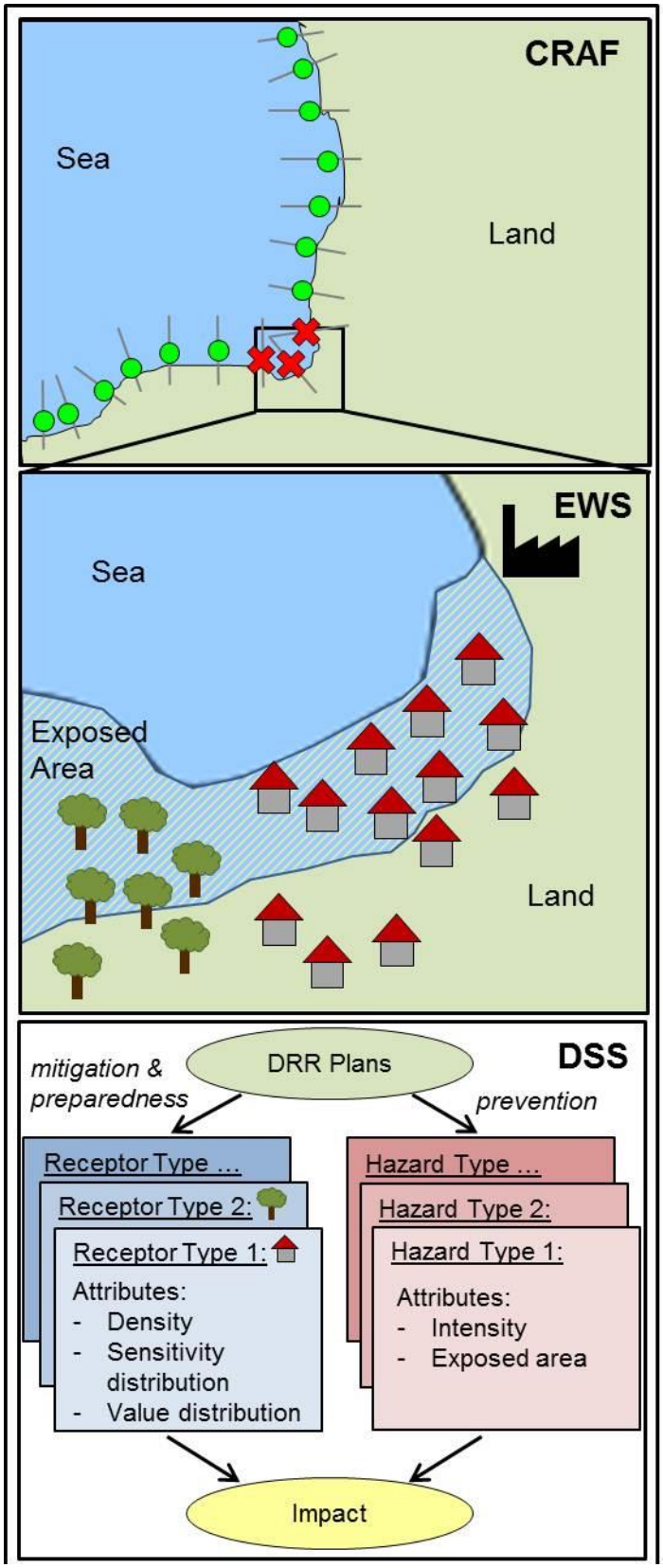

Figure 2-1 Conceptual drawing of the CRAF (top panel), the EWS (middle panel) and the DSS (bottom panel) (Van Dongeren et al., 2014).

The Coastal Risk Assessment Framework (CRAF) and EWS/DSS are to be implemented at different spatial scales, as illustrated in the top panel of Figure 2-1. The CRAF is to assess coastal areas at regional scales, in the order of 100km. The goal is to identify critical coastal areas, or *hotspots*, of high risk of spatial scale in the order of 10km.

For these hotspots an EWS will be developed to deliver short-term forecasts and warnings. This tool will have to be constructed of generic components and must have functionality across European Coasts. For marine hazards, such as hurricanes and storms, it will have to consist of a 2 dimensional model train that can compute from conditions at sea to hazard intensities on shore.

The hazard intensities on shore can be related to the total expected impact when related to the receptor attributes. For this a Decision Support Tool (DST) will be developed, as illustrated in the bottom panel of Figure 2-1.

The goal is to supply decision makers with a tool to gain quick insight into hazard intensities and impacts.

2.1.2 COASTAL HAZARD MODELING

In the modeling of flood hazards several steps have to be taken to get from the external forcing (i.e. a storm surge or high river discharge) to the impact (i.e. casualties and economic loss). In general these steps are divided between several computer models that are then coupled (the output of one, is the input of the other). The reason this type of modeling is not done in one large computational model has to do with the different spatial scales of the information.

In case of a coastal flooding, the external forcing is generally a large storm or a tsunami. These natural phenomena are very large in size (order of magnitude of 200-1000km) and therefore require a model that spans over a large area. In such a model it is impractical to have a high resolution as this leads to a high computational time. Near the shore and on land the behavior of the flow of water is largely influenced by local effects such as topology and roughness and relatively high detail is required in the modeling. It is therefore practical to describe the different aspects in computational models at different scales. Figure 2-2 illustrates the typical different modeling steps at different spatial scales.

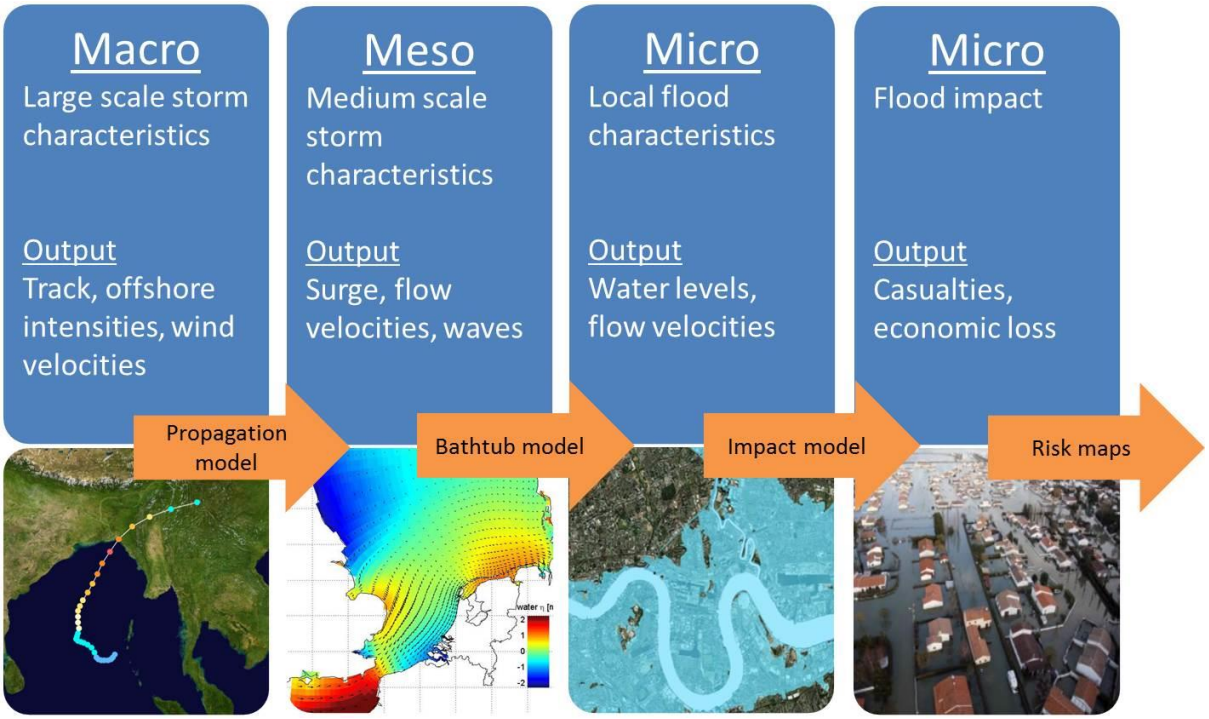

Figure 2-2 Spatial scales and characteristics of coastal hazard modeling, adopted from Van Verseveld (2014).

At larger scales the hazard is described at relatively low resolution and with the shore as a boundary. The results of this (e.g. surge levels, flow velocities and waves) can be used to compute different local flooding scenarios. These local models often impose a breach of a barrier and then use the shallow water equations to calculate the propagation of the water into the hinterland, of which the elevation is known. Different scenarios are computed by imposing breaches at different or several locations. The outputs are hazard intensities such as temporally and spatially varying water levels, flow velocities and water level rise rates. These may then be used to estimate the impact by combining it with geographical oriented data such as population density, buildings and economic

value. Since risk is defined as the product of probability and consequences, risk maps can be drawn up by determining the probability of a flood event.

It should be noted that most flooding models are static in nature and neglect run up from waves and morphology. This assumption does not hold for sandy coasts when dune erosion and/or run-up play an important role as the morphological changes change the pathway of the flood. This is especially true for the first row of houses at the coastline; these are not flooded in a static model but are in fact damaged due to run up and dune erosion. In these cases, at micro scale, a model is required that takes into account flow as well as morphology. The disadvantage is that such a model will require more computation time as more processes are included. Two examples of models capable of this are Delft3D (when D-FLOW and D-MORPH are used) and XBeach. The model XBeach will be used in this research and is described in the next chapter.

2.1.3 XBEACH

XBeach is a two dimensional process based model that incorporates hydrodynamics and morphodynamics to compute the natural coastal response (including dune erosion, overwash and breaching) for hurricane and storm conditions. The model consists of formulations for short wave envelope propagation, non-stationary shallow water equations, sediment transport and bed update (Roelvink et al., 2010).

The model has been developed for the purpose of modeling storm impacts on low-lying sandy coasts with significant alongshore variability. The variability may be caused by anthropogenic obstacles (e.g. inlets, revetments, sea walls) or have natural causes (e.g. shoals, dune height variations, or rip channels). The incorporation of avalanching[1] and wave group forcing for swash motions is especially important for the process of dune erosion. This supplies sediment from the dunes to the swash and surf zone which is then transported seaward (Roelvink et al., 2009).

McCall et al. (2010) performed a simulation of Santa Rosa Island (Florida, USA) under the condition of Hurricane Ivan (2004). By comparing pre- and post-storm data it was shown that XBeach is capable of simulating complex runup and inundation overwash over terrain with significant alongshore variability.

In Vousdoukas et al. (2012) research has been conducted for the performance of XBeach for steeper slopes and consequently higher Iribarren numbers, on the Portuguese coast. The study shows that the default setup of XBeach can overestimate the effect of avalanching as well as the dune and beach-face erosion. It also showed to be most sensitive for the input parameters of the beach slope and the surf similarity parameter.

[1] Avalanching is a process that occurs due to periodic undercutting of the dune profile by swash motions. As the profile steepens the slope of the dune becomes instable and slumps.

2.1.4 BAYESIAN NETWORKS

A BN is a type of statistical model, typically a graphical probabilistic model. It represents a certain number of stochastic variables[2] and the conditional dependencies[3] between these variables. The BN is described in a probabilistic graph, as shown in Figure 2-3, in which the stochastic variables are represented in the nodes (circles) and the conditional dependencies on the edges (arrows). A BN is acyclic, which means that there is no path that starts at a certain node and can end up at the same node again. The relationship between two variables is based on Bayes rule:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{1}$$

In which, $P(A|B)$ is the probability of A given B is true, $P(B|A)$ is the probability of B given A is true and $P(A)$ and $P(B)$ are the individual probabilities of A and B. The nodes in the BN are possible variables within some system and the arrows indicate if the state of a variable has influence on the state of another variable.

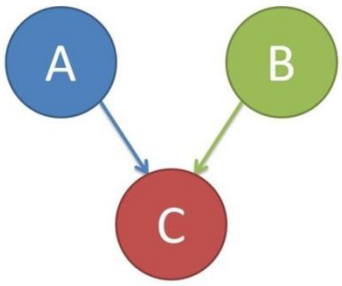

Figure 2-3 Simple Bayesian Network

A good example to demonstrate a BN is found in Den Heijer et al. (2012) and was adopted from Pearl (1988)**.** The network represents the distribution of five variables, P(B, E, A, C, R), as shown in, Figure 2-4. It relates the triggering of a household alarm to either a burglar or an earthquake (when living in an earthquake prone area). When the alarm rings a neighbor may call to inform you. When on your rush home you hear an earthquake report on the radio, the degree of confidence (or belief) that a burglary has triggered the alarm will now have decreased.

[2] A stochastic variable, or random variable is a variable whose possible values are the outcomes of a random experiment (e.g. throwing a dice). The value the random variable takes on pertains to a certain chance, or probability.

[3] Consider Figure 2-3: The variables A and B are conditionally independent. Initially they do not affect each other, but do affect the occurring of event C. If event C occurs, the occurring of A will now affect the probability of B occurring and vice versa.

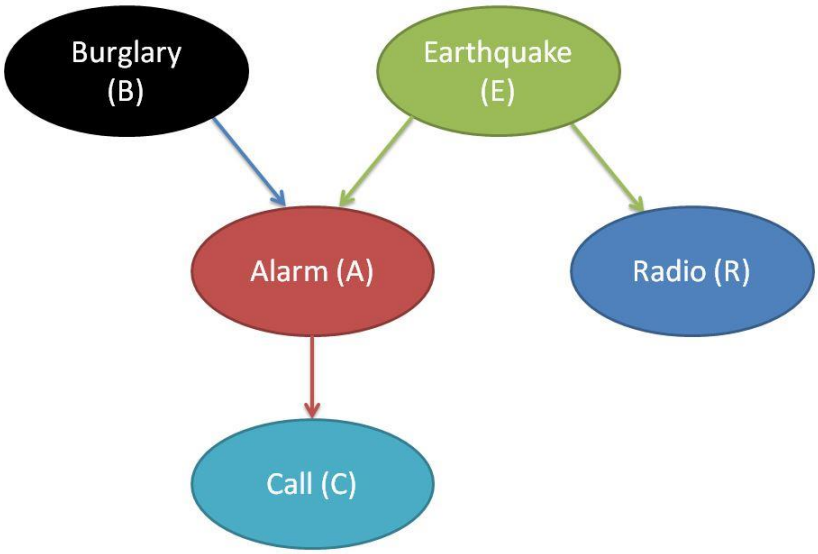

Figure 2-4 Example of a Bayesian Network of an alarm, triggered by either a burglary or earthquake

The BN is a computational tool to calculate probabilities of certain events (of the burglary) given evidence (the call and the radio report). When evidence is supplied, the network updates the probability of occurring of a certain event. The network has to be trained with either statistical data or expert judgment in order to establish the probability functions of the different variables. For the previous example it would need data on the occurrences of burglaries and earthquakes in the area, as well as the reliability of the alarm.

The structure of a BN has to be determined by the user, which may be a strength as non-existent links between variables will not be made. It is also a limitation as the structure of more complex networks may not be so straightforward. Another limitation is that the BN is not capable of making predictions outside the range of the training data. This will be of importance when using a BN in situations where the risk consists of low frequency events with high consequences. Little or no data are available when considering low frequency events such as a 1 in 10,000 year storm.

2.1.4.1 APPLICATON OF BN'S

BNs have a wide range of applicability and are subsequently used in many different fields. In literature many examples can be found ranging from traffic flow forecasting (Sun et al., 2006) to the diagnosis of breast cancer (Kahn et al., 1997). A more recent example is found in the thesis of Jäger (2013) where a BN is used to model the human influence on Safety.

The use of BN's has also proved their use in the field of coastal engineering and geology. Hapke & Plant (2010) show that a Bayesian approach is useful for predicting cliff erosion by linking the forcing variables (e.g. wave conditions) and initial conditions (e.g. cliff geometry) to cliff erosion in a BN. Dune erosion volumes due to storm impact, as predicted by an empirical model, can also be reproduced by a BN as shown in Den Heijer et al. (2012). A BN has also been used to predict coastal vulnerability to sea level rise (Gutierrez et al., 2011) and van Verseveld et al. (2015) relates the onshore hazard intensities of hurricane sandy in a part of New York to damages using a BN.

2.1.5 COPULAS

Hydrological phenomena such as storms can be described by a set of random variables such as the significant wave height, surge level, storm duration and peak wave period. The random variables are all due to the same meteorological system and therefore related in one way or another. The interrelation between these variables has generally been described by classical bivariate distributions, such as the Gaussian, log-normal, gamma and extreme value distributions. This approach limits the individual behavior of the random variables to the same distribution, where this is not always the case. Copulas, however, are able to describe and model the interrelation between several random variables, and in the bivariate case, without the restriction of having to be of the same distribution (Genest & Favre, 2007, Schmidt, 2007). Just like marginal distributions can be fit to a random variable a copula can be 'fit' to describe the dependence between two or more random variables. Several copulas exist and can be tried to fit a dataset and tested on its goodness of fit. Once a copula is known it can be used to model an existing dataset to create a synthetic dataset that mimics the characteristics of the original data.

In this research only the bivariate copulas will be used and the explanation is therefore restricted to this type. The mathematical description of a copula is described by Sklar's theorem (Sklar 1959) and states that every joint cumulative distribution function $H(x, y)$ with marginal distributions $F(\mathrm{x})$ and $G(y)$ can be written as:

$$\mathrm{H}(x, y) = \mathrm{C}\{F(\mathrm{x}), G(\mathrm{y})\}, \quad x, y \in \mathbb{R}$$

Where $C : [0,1]^2 \rightarrow [0,1]$ is a 2-dimensional copula.

The dependence structure between the two variables has to be described isolated from their marginal behavior. In order to do this the random variables first have to be transformed to uniformly distributed random variables. This is possible if the inverse of the cumulative distribution function of the marginal parametric distribution exists.

If Z is the cumulative distribution function of the random variable A then $A = Z^{-1}(U_A)$ and $U_A = Z(\mathrm{A})$, where Z^{-1} denotes the generalized inverse of Z. The copula C now becomes:

$$C(x, y) = H\{\mathrm{F}^{-1}(\mathrm{u}_x), G^{-1}(u_y)\}, \quad \mathrm{u}_x, \mathrm{u}_y \in [0,1]$$

Four copulas will be considered: the Gaussian, Skew-t, Clayton and Frank. The last two belong to the class of Archimedean copulas. Each copula has different characteristics that can be used to better describe some given dataset.

The Gaussian copula is given by:

$$C(\mathrm{u}_1, u_2) = \Phi_\rho\{\Phi^{-1}(\mathrm{u}_1), \Phi^{-1}(\mathrm{u}_2)\}, \quad \mathrm{u}_1, u_2 \in \mathbb{R}$$

In which Φ_ρ is the bivariate normal distribution, with correlation ρ, and Φ^{-1} is the inverse of the standard univariate normal distribution.

The Archimedean copulas are constructed with a generator $\phi:[0,1]\rightarrow[0,\infty]$ and have the form:

$$C(u_1,u_2)=\phi^{-1}[\phi(u_1)+\phi(u_2)]$$

Three copulas belonging to this class are Clayton, Frank and Gumbel, each has their own generator and properties, summarized in Table 2-1.

Table 2-1 Archimedean copula: generator functions and tail dependence.

Copula	Generator for ϕ	Parameter	Tail dependence
Clayton	$\frac{x^{-t}-1}{\delta}$	$\delta \geq -1$	Lower
Frank	$-\log\left(\frac{e^{-\beta x}-1}{e^{-\beta}-1}\right)$	$\beta \in \mathbb{R}$	None
Gumbel	$\lvert\log(x)\rvert^{\alpha}$	$\alpha \geq 1$	Upper

The previously mentioned copulas can simulate tail dependence but are symmetrical in nature. The last copula that will be considered is the skew-t copula and does not have this limitation and thus allows for more heterogeneity in the data to be modeled. A detailed explanation of this copula can be found in Demarta & McNeil (2005) and an application in Jäger & Morales (2014).

2.2 SIGNIFICANCE OF THE RESEARCH

This research will explore the possibility of creating an EWS for marine coastal hazards using a BN approach. The focus lies on populated coastal areas with sandy shores. The need for this research arises from the goals formulated in the HFA and the requirements formulated in the EU Flood Directive for the EU member states. Furthermore the UN have identified that EWS are key in reducing casualties and economic losses due to flood events in UNISDR (2002).

A more general significance, other than a tool within the RISC-KIT project, can be found in the use of a BN as a surrogate for a morphological model. Such a network can be a powerful tool for decision makers and engineers in coastal zone management as it gives quick insight in the spatial variation of possible dangers. The development of this tool for the European coasts does not limit its use for other areas on the globe with similar coastlines. A proof of concept for the European coast allows for further exploration for other vulnerable coastlines.

2.3 PROBLEM DEFINITION

Current coastal hazard models are generally static in nature and use the shallow water equations for calculating the propagation of water into the hinterland. These models do not include the morphologic response of a coastline to high impact events such as large storms and hurricanes. For sandy shores with beaches and dunes the coastal response is non-negligible and influences the

pathway of a flood to the hinterland, changing the flood duration, extend and depth fields. Current models therefore underestimate or wrongly predict the hazard intensities and impact of the flood when used for these coastlines.

A morphological model such as XBeach is capable of modeling the morphological response of the coastline due to storm events, see section 2.1.3. The model is capable of giving temporally and spatially varying output of hazard intensities such as flow velocity, water levels, water level rise rates, sedimentation and erosion. Rather than imposing a breach to model a breach is predicted by the morphological processes, also identifying weak spots. The model does not, however, link these hazard intensities to impact. The consequence of considering more processes (morphology) is an increase in the run time of the model. This poses a problem as weather forecasts are usually not very accurate until two or three days beforehand and are often only really reliable a day in advance. This makes the use of a morphological model as an EWS difficult as the runtime can easily become longer than a day, depending on the size and resolution of the model and the available computational power.

As a solution the development of a BN is proposed as a surrogate that incorporates the knowledge of the morphological model XBeach but has a fraction of the runtime as it is only based on statistics and not on physical processes. One of the strong points of XBeach is that it has 2D functionality, meaning it can give insight in the spatial variability of a flood. Incorporating this in a BN can lead to large complexity (a large number of nodes and edges). This is a problem as the larger the complexity of a BN, the more training data it requires to function properly. Training data is acquired by running scenario's in XBeach and uploading the in- and output into the BN. However, as previously stated running XBeach in 2D is computationally expensive and thereby limits the number of scenarios that can be run.

2.4 RESEARCH QUESTION AND OBJECTIVES

The insights obtained from the background information, significance and the problem definition have led to the following research question for this thesis:

> **How can Bayesian Networks be used as part of an Early Warning System for spatially varying coastal hazards at sandy coastlines?**

This research question is delineated by setting the following main objective, which is then subdivided into 5 different parts:

> **To quantify spatially varying coastal hazard intensities resulting from different forcing scenarios (e.g. storms) using a Bayesian Network.**

More specifically, to develop a BN that connects hydrodynamic near-shore boundary conditions to predict spatially varying overwash levels, flow velocities and coastline retreat at the case study site Praia de Faro, Southern Portugal.

This objective is divided into five sub objectives, namely:

1. Development of a Bayesian Network that can act as a surrogate for a 2D XBeach model.
2. Development of an XBeach model that generates suitable data for the Bayesian Network.

3. Investigate an efficient method for generating sufficient data for the Bayesian Network.
4. Apply the Bayesian Network to the case study site of Praia de Faro on the Portuguese coastline.
5. Evaluate the performance of the Bayesian Network with respect to its use as an EWS.

2.5 RESEARCH APPROACH

To answer the research question and reach the set goals the following steps are taken. (1) A general coastal analysis is performed of the study area with a focus to its response to high energy, low frequency events. A site is then chosen for which the EWS will be developed. (2) At a general level the modeling concept is already known; a BN will act as a surrogate for an XBeach model to be used as an EWS for coastal hazards. This is further developed and specified into a detailed concept in section 4.2. (3) The model concept is then implemented for the case study site by creating a storm dataset and an XBeach which are then used to create another dataset that on which the BN can be trained.

The resulting BN is then extensively analyzed and evaluated with respect to its use as an EWS. The obtained insights are discussed in the conclusion and recommendations are given for further research and improvement of the modeling concept.

3 STUDY SITE ANALYSIS

In order to model the situation at Praia de Faro correctly the characteristics of the coastal area need to be known. This will ensure that the right settings are used as well as enable qualitative and quantitative analysis of the modeling output.

3.1 CLIMATOLOGY

The Ria Formosa is located in the Algarve, in the south of Portugal, with a semi-arid climate. Summers are generally warm with an average maximum temperature of 29°C and winters are mild with an average minimum temperature of 10°C. The precipitation is negligible with around 400 to 600 mm per year, of which most falls in the winter months.

3.2 COASTAL GEOMORPHOLOGY

The southern coast of Portugal can roughly be divided into two sections. The western part ranges from the towns of Sagres to Quarteira and consist mainly of a rocky coastline consisting of cliffs with beaches in between outcrops. The eastern part consists of the Ria Formosa, starting five kilometer east of Quarteira and continues up to the Spanish border. The Ria Formosa is a 50 km wide coastal lagoon enclosed by five sand barriers and two spits, shown in Figure 3-1. The lagoon consists of a complex system of channels, tidal flats and marshes. It has six inlets of which four are natural and two have been fixed with groynes. Illustrating the highly dynamic nature of the system is the recent opening of a seventh inlet due to the breaching of Barreta Island. It is, however, questionable if this inlet will remain open for a prolonged period. (Vousdoukas et al., 2012a)

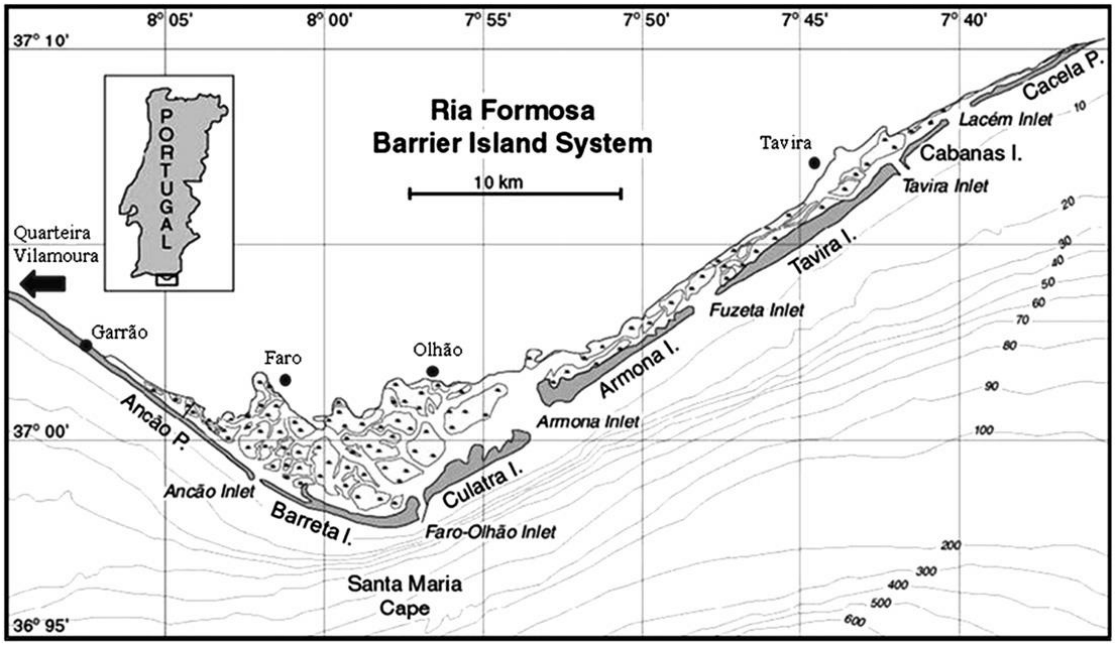

Figure 3-1 Image of the Ria Formosa, showing its location, the barriers, inlets and marshland. (Vousdoukas et al., 2012a)

The location of the barrier system of Ria Formosa adjacent to a cliffed coast of moderate relief is atypical for a barrier system with tidal flats; generally such systems are found on coastal plains (large

areas of low-lying land). The existence of this system is believed to be due to the presence of a shallow platform that is bounded by a relatively steep scarp on the inner continental shelf, visible in the bathymetry in Figure 3-1 (the steep slope indicated by the densely packed lines in front of Barreta and Culatra islands). Previously, in times of lower sea levels, this platform has been emerged. The theory is that as sea levels rose, the escarpment became a headland on which the shoreline began to pivot, allowing for the formation of spits. With further sea level rise the coastline retreated further and the spits became barriers with a tidal lagoon trapping sediment, to form tidal flats and marshland. At this point in time the barrier system is transgressing as a response to sea level rise. The evolution of the coastline is illustrated in Figure 3-2 (Pilkey et al. 1989).

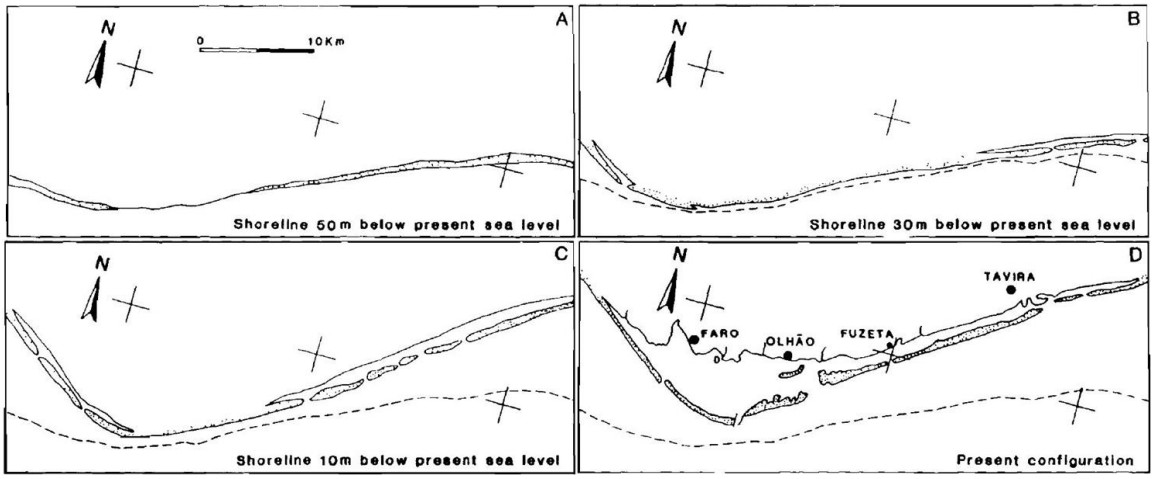

Figure 3-2 Evolution of the coastline in chronological order from A to D. (Pilkey et al., 1989)

Praia de Faro is located on the Ancão Peninsula, in the westernmost part of the system, shown in Figure 3-3. The peninsula is northwest – southeast orientated and is a spit that receives sediment from upstream eroding cliffs and from the continental shelf (Pilkey et al., 1989). It continues until the Ancão inlet and is partially inhibited by the settlement of Praia de Faro.

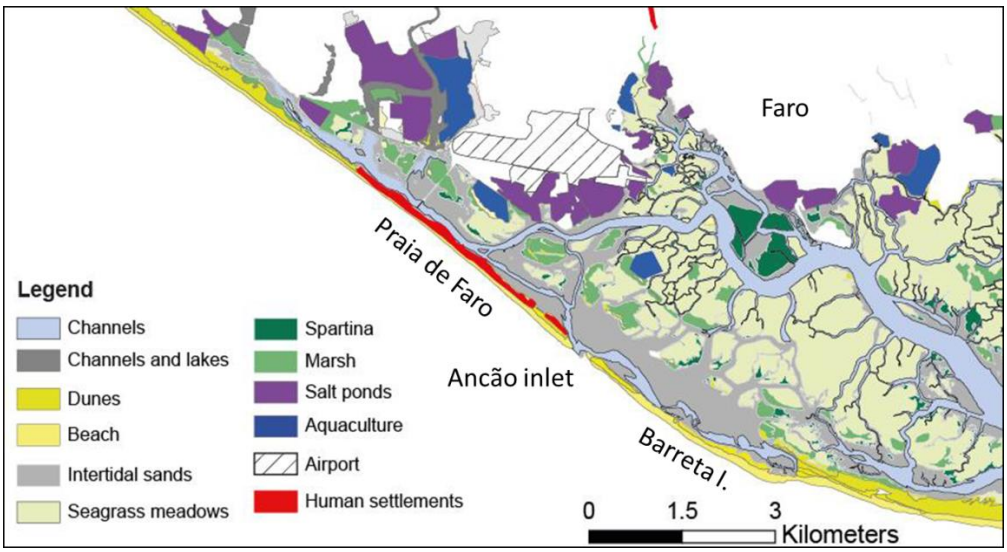

Figure 3-3 Praia de Faro, location and land use (Source: UAlg).

A typical cross-shore profile of a barrier island is shown in Figure 3-4, identifying all the different elements. From the sea side to the land side several areas can be distinguished. The nearshore area contains the surf zone and migrating submerged sand bars. The sea meets the land at the steeper beach face which is followed by the gentler sloped berm. Behind the berm the vegetated coppice mounds and foredune begin, which are in turn followed by the back barrier and bay side consisting of marshland and tidal flats.

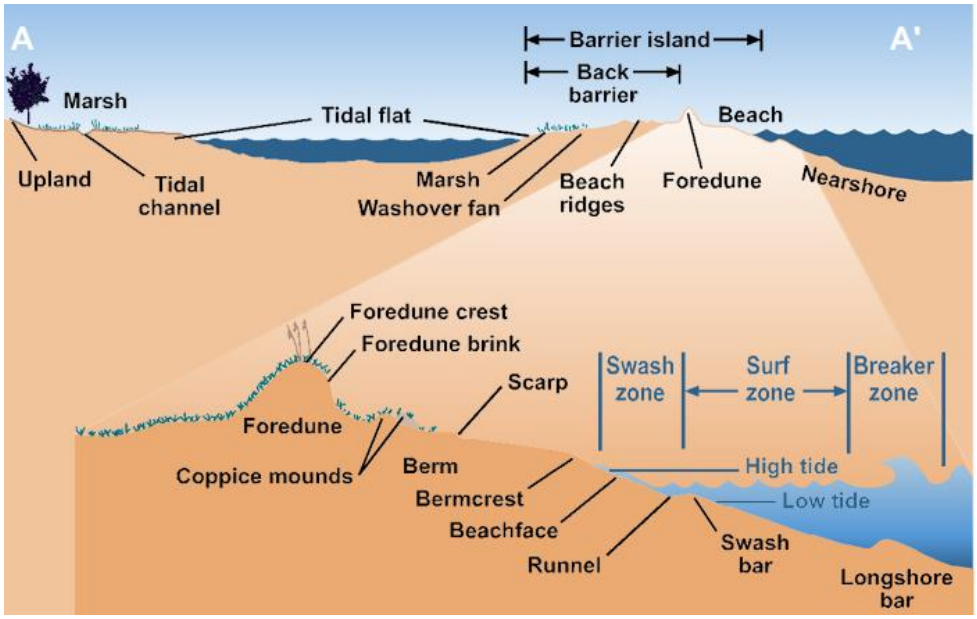

Figure 3-4 Typical cross section of a barrier island (Source: University of Texas)

The barrier of Praia de Faro shows the same features visible in Figure 3-4, where humans have not intervened. A large part of the peninsula is inhabited which has changed the natural configuration. In Figure 3-5 an aerial view of Praia de Faro is showing the barrier, the settlement of Praia de Faro and the access road connecting the barrier to the mainland. Directly east of the access road houses have been built on top of the dunes and directly face the berm of the beach. East of the access road buildings are located slightly less seaward and are protected by a single row of dune. At the access road there is a parking lot that extends from the backside of the barrier to the bermcrest. At the far side of the barrier (west of the access road), houses have been built by the local fisherman community directly behind the dunes. At this part of the barrier coppice mounds are also visible, indicating less human interference. The slope of the beach face at Praia de Faro is steep, typically above 10%, and ranges from 6% to 15%, categorizing the beach as intermediate to reflective (classification according to Wright & Short 1984).

Figure 3-5 Aerial view of Praia de Faro

3.3 WAVE CLIMATE

The wave climate can be observed in Figure 3-6 and Figure 3-7. The wave climate is moderate to high, with an average significant wave height of 0.92 meters and average peak period of 8.2 seconds. The predominant wave direction is west – southwest (70%), with a fraction of the waves originating from the southeast (23%) (Costa et al., 2001).

The wave climate has a strong seasonal variability as can be seen in Figure 3-8. The average monthly significant wave height and peak period follow a similar pattern over the year; the winter months, starting in October and finishing at the end of April, know more energetic sea states as compared to calmer summer periods. The correlation between the wave height and the period is further illustrated in the tables in Figure 3-9. The table shows the probability of the joint occurrence, in any one year, of a certain significant wave height and a certain peak period for the southeast, west and southwest directions. Waves originating from the southeast generally have shorter periods, indicating a predominance of wind waves and little swell waves. The west and southwestern directions show a combination of shorter and longer periods, indicating a mix of both sea and swell waves.

In general the state of the sea at a certain moment in time is described statistically by a wave spectrum. This treats the sea surface as a stochastic process; the waves that propagate over the sea surface are treated as random variables (a variable that cannot be predicted exactly), which together make up the total surface of the sea (the sea state) (Holthuijsen, 2007). The wave spectrum gives insight in the occurrence of certain wave heights and wave periods.

An analysis of deep water wave data from the Portuguese west coast (Figueira da Foz, 90m depth) shows that 30% percent of the wave spectra are double peaked (Pires Silva & Sarmento, 1991). This indicates a sea state existing of both swell and wind waves as is also seen in Figure 3-9. Another analysis of sea spectra for the North Atlantic Ocean shows an occurrence of double peaked spectra

of 40% for lower sea states (Guedes Soares, 1984). Both analyses conclude that for higher sea states (storms) the occurrence of double peaked spectra drops significantly. There has, however, not been any work done regarding the spectral distribution of the south coast of Portugal.

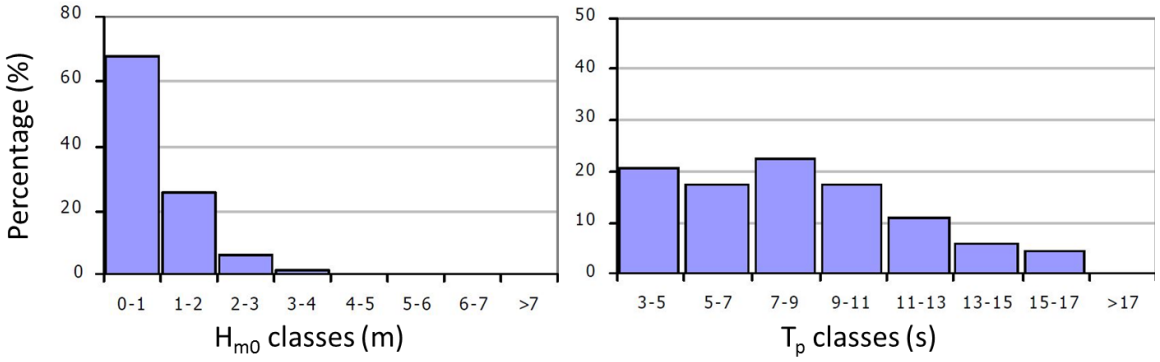

Figure 3-6 Relative distribution of significant wave height (Hmo) and peak period (Tp) (Costa et al., 2001).

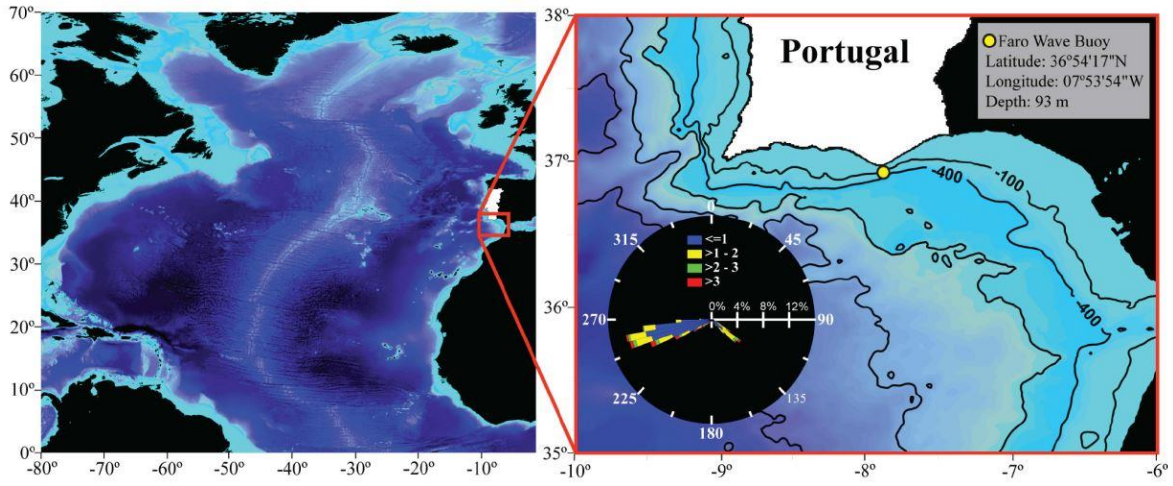

Figure 3-7 Wave rose distribution for measured data of the Faro Buoy (Almeida et al., 2011a).

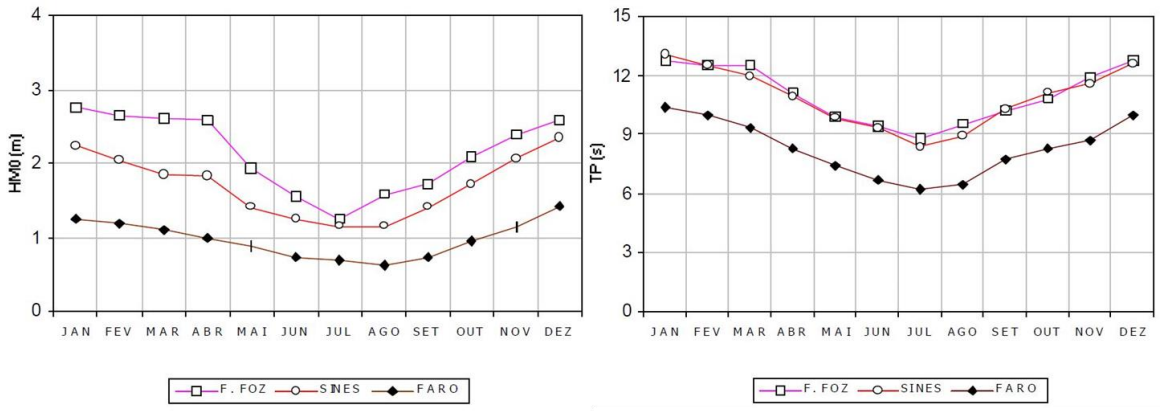

Figure 3-8 Monthly variation of average significant wave height and peak period (Costa et al., 2001).

SE (23,2%)	3-7	7-11	11-15	>15
<1	4.0	5.3		
1-3	14.7	37.3		
3-5		1.6		
>5				

SW (18,3%)	3-7	7-11	11-15	>15
<1	17.3	13.8	19.2	12.8
1-3	5.4	16.5	4.8	6.0
3-5		3.1	0.6	0.1
>5			0.2	

W (52,3%)	3-7	7-11	11-15	>15
<1	25.2	34.6	17.6	1.7
1-3	5.8	8.5	5.3	0.9
3-5		0.2	0.1	
>5				

Figure 3-9 Joint distribution of significant wave height (hm0) and peak period (Tp) per direction at Faro (Costa et al., 2001).

3.4 STORM CLIMATE

The Ria Formosa is prone to storms from the southwest and the southeast. The southwester's are generated by a deep atmospheric low pressure system that follows a path more southerly than normal and normally occur between December and March. The southeaster's are due to the occurrence of a strong *Levante* wind (a warm and strong easterly wind that develops in the West of the Mediterranean Sea) and occur from autumn until spring (October – May). The distribution for the number of storms per direction follows that of the average wave climate: approximately 70% of storms originate from the southwest and 30% from the southeast (Almeida et al., 2011a).

Storms in the area are defined by a minimum significant wave height higher than 3 meters and with a minimum duration of 3 hours. Individual events are separated by a 30 hour interval between measurements of 3 meter significant wave height (Almeida et al., 2011a).

Return periods for wave heights at the Ria Formosa are found in the report of Pires (1998), and are shown in Table 3-1. For the same return period the waves from the southwest are much higher than those originating from the southeast. This is in line with expectations and can be explained by looking at Figure 3-7. The fetch over which waves can be generated for the southeastern direction is limited where it is practically unlimited for the southwest.

Table 3-1 Significant wave height for different return periods for waves per direction based on modelled data west of Ria Formosa (Pires, 1998).

Mean direction (°)	Return period (year)				
	5	**10**	**25**	**50**	**100**
232 (SW)	5.7 m	6.4 m	7.4 m	8.1 m	8.8 m
128 (SE)	4.4 m	4.6 m	4.8 m	5.0 m	5.1 m

A relationship between the significant wave height and the peak period has been estimated by Rodrigues et al. (2012). A linear regression on wave, tide and storm data in the period from 1997 to 2007 for storms originating from the west and southwest, results in: [4]

$$T_p = 0.834 \times H_s - 6.565 \quad (2)$$

In which,

T_p = peak period (s)
H_s = significant wave height (m)

This relation can be used to calculate wave periods accompanying the wave heights from Table 3-1. It can also be used to calculate the deep water wave length, which allows the wave steepness to be derived. The equation for the wave steepness is:

$$S_p = \frac{H_s}{L_p} = \frac{2\pi H_s}{g T_p^{\,2}} \quad (3)$$

In which,

S_p = significant wave steepness (-)
L_p = deep water wave length (m)
H_s = significant wave height (m)
T_p = peak period (s)
g = gravity = 9.81 (m/s^2)

Using equations (2) and (3) the wave heights from Table 3-1, wave periods and wave steepness are derived for the different return periods, shown in Table 3-2.

Table 3-2 Wave steepness and period estimated using the findings of (Rodrigues et al., 2012) **and** (Pires, 1998)**.**

Return period (years)	Hs (m)	Tp (s)	Steepness (-)
5	5,7	11,3	0,0285
10	6,4	11,9	0,0289
25	7,4	12,7	0,0292
50	8,1	13,3	0,0292
100	8,8	13,9	0,0292

3.5 STORM GROUPS

After the occurrence of a storm the beach needs time to recover to its original state after it has been partially eroded. This period of time is referred to as the beach recovery period. If the frequency of storms is high enough it may happen that another storm takes place during the beach recover period. This enhances the impact of the storm and the two events cannot be treated as individual

[4] It should be noted that the relation between the significant wave height and peak period in Rodrigues et al. (2012) shows a large scatter for wave heights < 4,5m, but shows a good fit for waves > 5m.

events, as the first storm influences the impact of the second storm. If one or more storms occur in succession, within the beach recovery period, a storm group is born. The significance of storm groups is accentuated by the fact that two small storms, of relatively short return periods in short succession, can have a much larger effect than one large individual storm, with a high return period. For the Portuguese west coast the erosion due to a storm group with a return period of one year can cause the same amount of erosion as a single storm with a return period of nine years (Ferreira, 2006).

3.6 EXTREME WATER LEVELS

Increased water levels associated with storms are roughly due to two meteorological factors: the increase in water levels due to low atmospheric pressure in the eye of a storm and the setup of water induced by wind blowing the water in the direction of the shore. Another important factor that determines the height of the storm surge is the size of the coastal shelf. The Ria Formosa has a relatively short coastal shelf, as can be seen in Figure 3-7, causing the storm surge level to be relatively low.

Analysis of the joint occurrence of storm wave height and surge level (or residual water level) shows a positive correlation, shown in the left panel of Figure 3-10; higher wave heights are associated with higher water levels. Further analysis includes the North Atlantic Oscillation[5] (NAO), shown in the center and right panels of Figure 3-10. Two observations are made: (1) higher wave heights are associated with a negative NAO index as well as higher surge levels and (2) for a positive NAO index wave height greater than 4 m do not occur and surge levels are similarly low (Plomaritis et al., 2015).

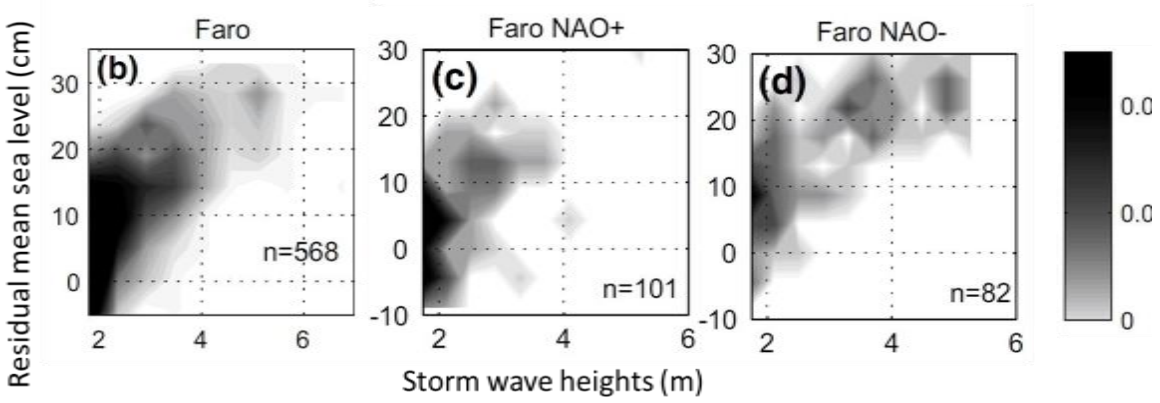

Figure 3-10 Left panel: Joint occurrence of storm wave heights and residual means sea level. Centre and right panel: Joint probability distribution of storm wave heights and residual sea level during NAO > 1.5 and NAO < –1.5. Color scale indicates probability of occurrence (Plomaritis et al., 2015).

A relationship between the storm surge and the significant wave height is determined by Rodrigues et al. (2012) with linear regression on wave, tide and storm data in the period from 1997 to 2007 for storms originating from the west and southwest, resulting in:

[5] The North Atlantic Oscillation is a climatic phenomenon in the North Atlantic Ocean and describes the fluctuation between two semi –permanent pressure zones in the Atlantic Ocean: the Icelandic low and the Azores High. A positive index (NAO+) indicates a high pressure different between the two zones. Conversely a negative index (NAO-) indicates a relatively small pressure difference.

$$S = 0.111 \times H_s - 0.175 \quad (4)$$

In which,

S = storm surge (m)
H_s = significant wave height (m)

In Almeida et al. (2012) a similar analysis has been performed on the same dataset. This analysis, however, includes all directions, resulting in a similar relation between the surge and wave height:

$$S = 0.1029 \times H_s - 0.1544 \quad (5)$$

Rodrigues et al. (2012) uses the relationship between surge levels and wave heights (equation (4)) to estimate return periods for surge levels by coupling them to the return periods found for wave heights in Pires (1998) (Table 3-1). In Table 3-3 both the relations found by Rodrigues et al. (2012) and Almeida et al. (2012) are used to determine surge levels for different return periods. The differences between the estimated surge levels are small and thus in agreement with each other.

Table 3-3 Return periods for surge levels and significant wave height (Hs).(Pires, 1998) **&** (Rodrigues et al., 2012) **&** (Almeida et al., 2012)**.**

Return Period (year)	Hs (m)	Surge (m)	
	Pires	Almeida et al. 2012	Rodrigues et al. 2012
5	5.7	0.43	0.46
10	6.4	0.50	0.54
25	7.4	0.61	0.65
50	8.1	0.68	0.72
100	8.8	0.75	0.80

The increase in water level due to the storm surge is not very high when compared to the magnitude of the tide in the area. The tide at Praia de Faro is semi-diurnal and the average range (with respect to mean sea level) for neap and spring tide is 1.3 and 2.8 meters respectively. However, ranges of 3.5 meters can be reached. When comparing the water levels of the tide to the surge, the tide is the dominant factor. This means that the storm surge is only important if it occurs jointly with a high tidal elevation.

An analysis of extreme water levels, including tide and surge levels, has been performed in Carrasco et al. (2012). A data set from the Huelva tide gauge for the period 1996 – 2011 was used (60 km east of Praia de Faro) to create a probability density function of the extreme water levels, shown in Figure 3-11. Water levels for several return periods have been determined and are shown in Table 3-4.

Table 3-4 Return periods for extreme water levels based on the Huelva tide gauge (Carrasco et al., 2012).

Return period (years)	Water level (m above MSL)
5	2.30
10	2.39
25	2.40
50	2.43
100	2.48

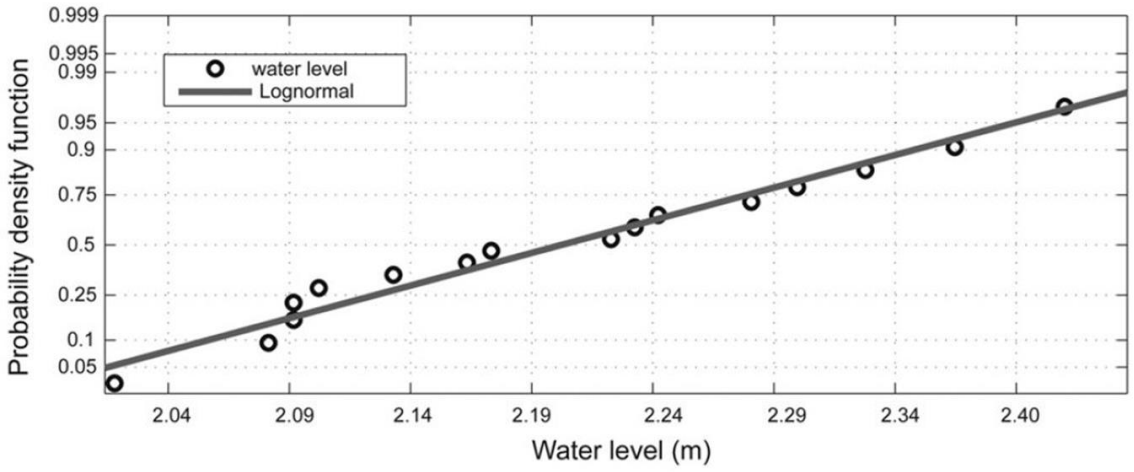

Figure 3-11 Lognormal distribution fitted to annual maximum water levels from the Huelva tide gauge (Carrasco et al., 2012).

Due to the absence of any large rivers and negligible precipitation in the area the total water level can be assumed to consist solely of the tidal elevation and the storm surge. Using the storm surge levels from Table 3-3 and the water levels from Table 3-4 a rough estimate for the extreme tidal elevation can be obtained, shown in Table 3-5. The tidal elevation for increasing return periods shows a decreasing trend. This may be due to the fit of the regression line on the data; a slight tilt in the fit would give a different trend.

Table 3-5 Estimation of the extreme tidal elevation.

Return Period (year)	Hs (m)	Surge, RSL (m)		WL (m above MSL)	Maximum Tidal level (m above MSL)	
	Pires	Almeida	Rodrigues	Carrasco	Carrasco – Almeida	Carrasco – Rodrigues
5	5.7	0.43	0.46	2.30	1.87	1.84
10	6.4	0.50	0.54	2.39	1.89	1.85
25	7.4	0.61	0.65	2.40	1.79	1.75
50	8.1	0.68	0.72	2.43	1.75	1.71
100	8.8	0.75	0.80	2.48	1.73	1.68

3.7 MORPHODYNAMICS

Coastal morphodynamics describes the interaction between the coastal system and the environment. It concerns the external hydrodynamic forcing (waves, currents and tides) drive the processes that cause sediment transport. Sediment transport causes morphological change, which in its turn changes the processes that drive the sediment transport. This feedback loop is the main cause for complexity in coastal evolution (Masselink & Hughes, 2003).

Considering the morphodynamics of a coastal system, time scale is important. When considering the long term, the equilibrium state is governed by the prevailing average environmental conditions (e.g. average wave height and direction, tidal currents and sea level rise). Short term events such as storms generally do not dominate the equilibrium of a coastline, but can have large local effects such as overwash or the breaching of a barrier. On long term scales many storms do have influence on the shaping and reshaping of barrier islands. This research concerns the effects of storms and therefore the long term morphodynamics will only be described concisely.

The main sources of sand for the area are from the continental shelf and cliff erosion (Pilkey et al., 1989). The net long-shore transport in the area is typically from west to east and ranges from 10,000 to 40,000 m^3/year according to Almeida et al., (2011b), however a recent report shows that the long-shore transport should be in the order of 100,000 m^3/year (GTL, 2014). Considering the Ancão peninsula, the urbanized dunes tend to erode at the central and western parts, where the non-urbanized dunes tend to accrete at the eastern part (Vousdoukas et al., 2011). The Ancão inlet channel is migrating eastward, with the direction of the littoral drift, at increasing rates (Vila-Concejo et al., 2002).

The morphologic response of barriers during storms is described by four regimes, defined by (Sallenger, 2000), and as shown in Figure 3-12. In the swash regime wave runup is restricted to the beach. Sand is transported offshore and is stored in offshore bars. This sand is naturally returned to the beach during periods of calmer weather following the storm. In the collision regime part of the dune is eroded and transported offshore. This erosion is considered as a semi-permanent change to the dune. During an overwash event the wave runup exceeds the crest of the dune and sand is transported landwards, contributing to the transgression of the barrier. When the water level reaches a level higher than the crest of the dune the barrier may be inundated, transporting large amounts of sediment landwards (Sallenger, 2000).

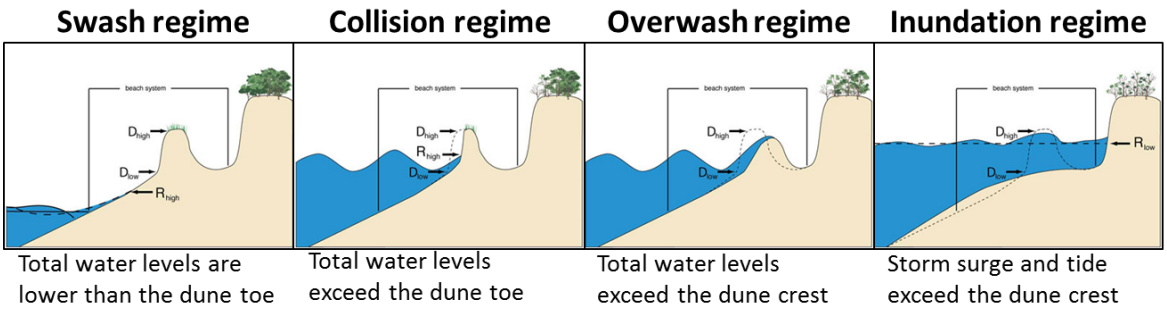

Figure 3-12 Impact scales on barrier according to Sallenger (2000) (source: USGS).

The main coastal hazard for the Ancão Peninsula is that of overwash and dune erosion in the collision regime. In Rodrigues et al. (2012) a vulnerability map of Praia de Faro is presented concerning overwash events for storms with different return periods, shown in Figure 3-13. Especially the areas where urbanization has destroyed the dunes and thereby lowered the elevation are susceptible to overwash events. This will be elaborated further in the next section.

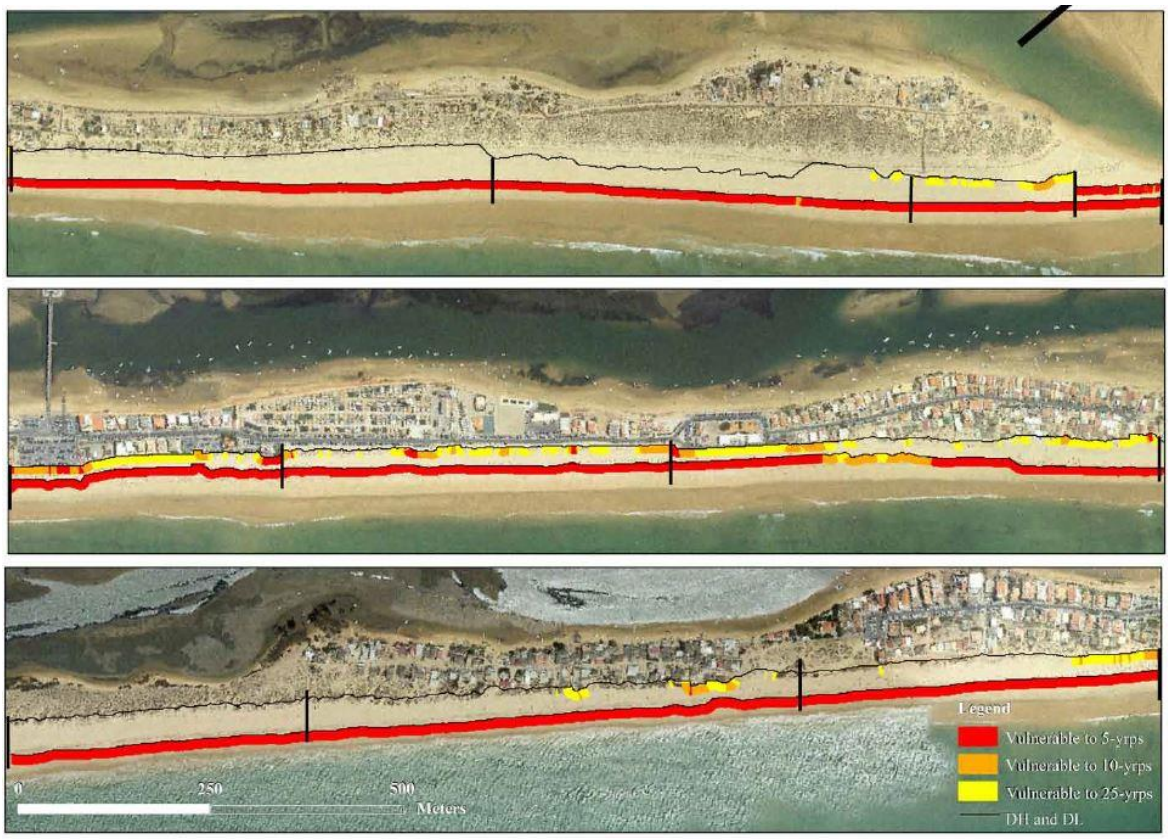

Figure 3-13 Representation of collision and overwash regimes vulnerability for 5, 10 and 25 year return period storms. The seaward line represents collision potential and the landward line the overwash potential. Top: eastern part of the Ancão peninsula. Centre: middle part of Praia de Faro. Bottom: western part of Praia de Faro. Vulnerability is shown in red, orange and yellow:

3.8 COASTAL VULNERABILITY

Continuing on the previous section the main hazards for Praia de Faro are those of erosion in the collision regime and overwash. The frequency of these events is quite high, with the expectation of approximately one overwash event per year (Almeida et al., 2012).

In Almeida et al. (2012) two methods have been used to establish thresholds for storms that cause damages due to overwash and coastline retreat. Thresholds have been identified using (1) historic records of storm impact on the coastline and the associated hydrodynamic conditions and (2) by performing calculations to determine extreme runup for the local bathymetry. With respect to thresholds determined by the first method the events can be divided into several groups: (1) single storms from the southwest, (2) single storms from the southeast, (3) storm groups from the southwest, and (4) storm groups from the southeast. The thresholds found for each category are shown in Table 3-6, as well as their return periods. A large finding is the difference in magnitude of

the storms from the different origins that impact the coastline. A storm from the southeast has to be significantly larger for a similar impact of a smaller storm from the southwest. This is due to the cuspate shape of the Ria Formosa. Waves from the southeast lose energy due to diffraction around the tip of the Ria Formosa and attack the beach at an oblique angle, causing lower runup.

Table 3-6 Thresholds for storms causing damage at Praia de Faro (Almeida et al., 2012).

Description	Threshold			Return period (years)
	Hs	Duration (days)	# of storms	
SW	± 4.7	2 days	-	3.1
SE	± 6	2 days	-	40
SW group	> 3.5	< 2 days	2	1.7
SE group	> 3.9	1 days	3	40

Overwash calculated according to the scale of Sallenger (2000) with use of the equation proposed in Holman (1986) for R_2 (the runup level exceeded 2% of the time) give correct predictions for 91% of the historic cases. For five cross-shore sections (Figure A-1) these calculations have been performed of which the determined thresholds are displayed in Figure A-2 . These calculations show the spatial variability that can also be seen in Figure 3-13, of the previous section. The most vulnerable areas are those where the dune has been destroyed by the settlement and the first barrier after the beach face are the local houses (Almeida et al., 2012).

Having larger consequences but a lower probability, another danger is the breaching of the barrier (or the creation of an inlet). The Ria Formosa has seen several natural inlet formations over the past century. However, for Praia de Faro this is not considered as an imminent threat at this moment.

The most vulnerable part of Praia de Faro is the parking lot at the entrance road of Praia de Faro. The reason for this is threefold: (1) the elevation of the sea wall is one of the lowest points on the barrier, (2) the berm of the beach at this point is very thin because the parking lot is built more seaward than most other structures and (3) under some storm conditions the beach is not scarped, instead a ramp is formed enhancing runup and overwash. An additional danger is that the access road situated at the parking lot is also the only road off the barrier.

4 DEVELOPMENT EARLY WARNING SYSTEM

This chapter describes the development of the EWS. First the general concept of the model will be explained and its different elements. This is followed by the setup and development of each element.

4.1 MODEL TRAIN

A model train is a set of models in which information is passed from one model to another. The different models describe different aspects and operate at different information levels. A model train is often used for coastal hazard modeling as described in section 2.1.2.

For the case study site five different models are used to propagate a storm from the largest scale to the smallest scale at three different spatial scales (Figure 4-1 & Figure 4-2). These models are the basis of the Early Warning System of which the last part is subject to this research. The largest scale is the Gulf of Cadiz, seen in the top left panel of Figure 4-2. For this area two models have been made, a Delft3D[6] model that calculates water levels, and a SWAN[7] model that calculates wave propagation. The large scale models generate the boundary conditions for the models on the intermediate scales of the Ria Formosa and the Ancão Peninsula (the right panel of Figure 4-2). The water levels and wave spectra that are derived from the intermediate scale can be used to force an XBeach model for the smallest scale: Praia de Faro, seen in the bottom left panel. The XBeach model uses both the water level and the wave spectra to calculate morphological changes to the coastline.

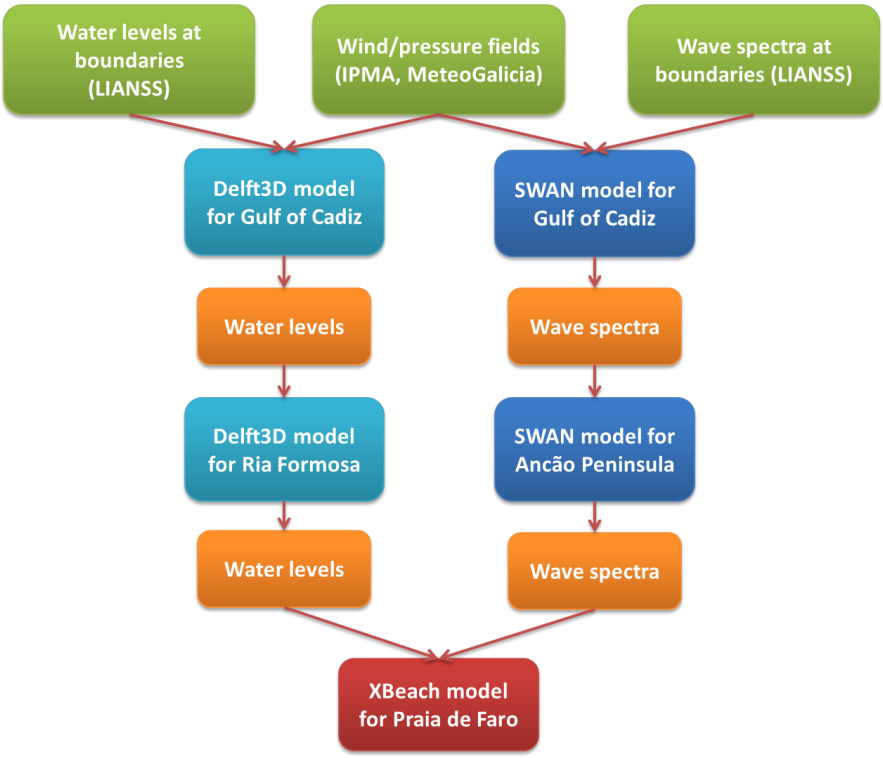

Figure 4-1 Scheme of the model train for Praia de Faro.

[6] http://oss.deltares.nl/web/delft3d
[7] http://swanmodel.sourceforge.net/

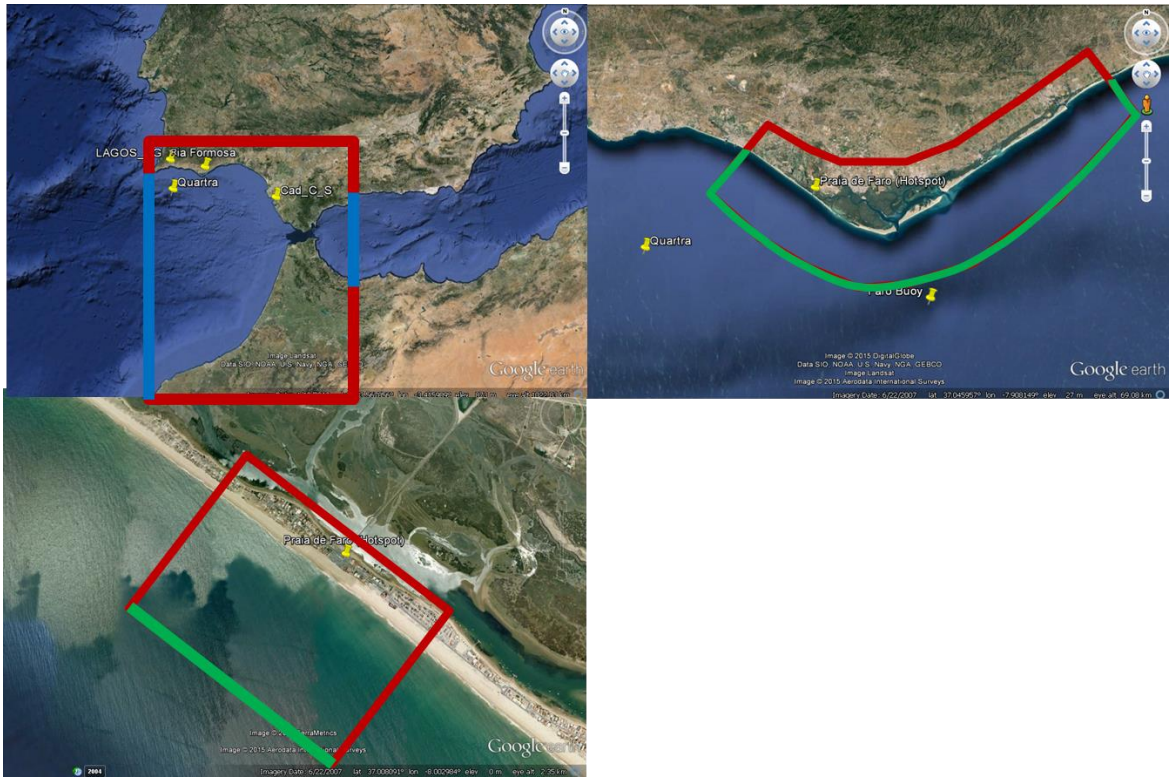

Figure 4-2 Locations of the different scale models. Top left: Gulf of Cadiz, Delft3D & SWAN. Top right: Ria Formosa, Delft3d & SWAN. Bottom left: Praia de Faro, XBeach.

4.2 GENERAL MODEL APPROACH

The general concept has been touched upon in previous chapters. In essence a statistical model will be used to act as a surrogate for a process based coastal morphodynamic model. The reason for this is that the process based model has long run times and requires expert knowledge to set up and interpret, where the statistical model works nearly instantaneous and requires much less expertise to use.

More specifically a BN will be fed by and later substitute an XBeach model; both are described in chapters 2.1.3 and 2.1.4. The XBeach model is able to translate offshore marine boundary conditions (water levels, wave heights, wave periods, etc.) to onshore hydrodynamic conditions and morphological changes on the coastline (the local hazard intensities). As output of the model the near-shore and onshore hazard intensities can be requested. The XBeach model can be forced with a range of storms, described by different offshore boundary conditions, resulting in a range of onshore local hazard intensities. These boundary conditions can be derived from outer models at the -20 meter depth contour line described in the previous section. When multiple runs are done a dataset of *cases* can be created of the in- and output combinations of each individual run. One case then holds the information of one run. This dataset in which each case describes an individual storm can then be used to set up a BN.

The BN consists of nodes and arcs as explained previously. The nodes hold information about the individual variables of the dataset and the arcs indicate dependencies between the individual

variables. When the nodes and arcs are set up the network can be *trained* with the dataset, giving the network knowledge.

As described in chapter 2.3, there are two large issues with this concept: (1) XBeach has long run times making it hard to create a large dataset and (2) a BN generally needs a lot of data to be well trained and increasingly so for higher complexity (more nodes and interrelations). Another issue (3) is that there is no large existing storm dataset that can be used to force XBeach and thereby create a dataset to train the BN. This dataset will thus have to be created. These three issues will be addressed in this chapter.

The first issue is tackled by reducing the number of grid cells of the XBeach model to a minimum and thereby reducing the necessary runtime of an individual storm. The second issue is addressed by reducing the complexity. The solution for this is found in the analysis of the coastal system, where many of the boundary conditions that describe a storm show significant correlation. The correlation is found between the significant wave height, peak period, storm duration and storm surge level. There is also correlation with the direction of the storm but in this research only storms from the southwest are considered.[8] The correlation between the variables makes it possible to group all of them into two nodes in the BN: the significant wave height and the maximum water level. The only independent variables that are left are now the water level variation due to the tide and the significant wave height. The last issue is solved by creating a synthetic dataset based on a smaller existing dataset of historical storms using copulas. The steps that have to be taken in the development are shown in Figure 4-3.

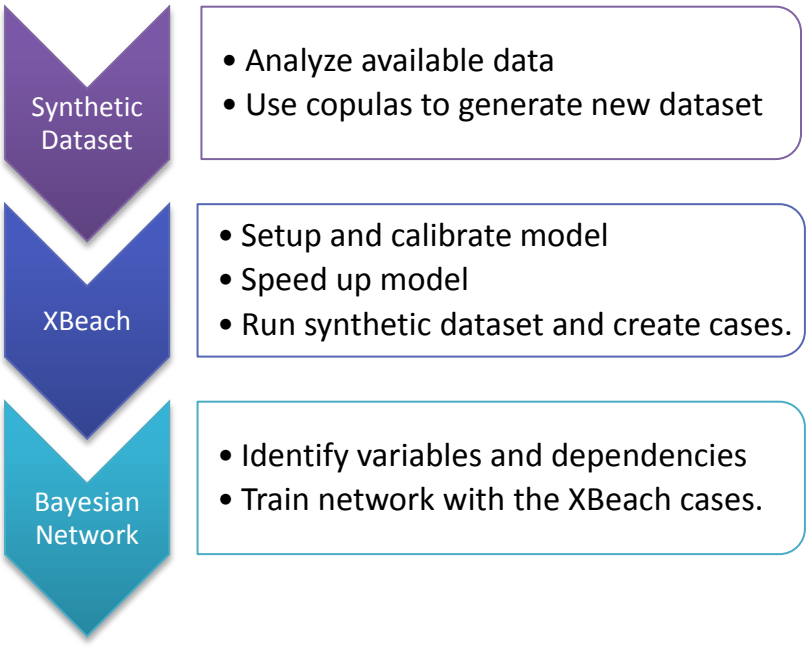

Figure 4-3 General model approach.

[8] Storms from the southeast are smaller in size due to the limited fetch and have already been diffracted around the tip of the Ria Formosa before they reach the case study site making them much less destructive

4.2.1 MODELING OF A STORM

Over the duration of a storm there is always a variance in the significant wave height, period and water levels. The actual course of a storm is unknown and therefore an assumption needs to be made. Generally speaking, a storm increases and decreases in strength over time. To mimic this behavior the course of a storm is modeled in a triangular way; the wave height and surge increase at a constant rate until the peak is reached and then decreases again at a constant rate. This is illustrated in the top panel of Figure 4-4. In this the significant wave height and surge levels are assumed to follow the same pattern without a lag. The peak period is linked to the wave height by assuming a constant steepness over the course of the storm.

The water level during a storm is in Praia de Faro for a large part determined by the tide. Because it cannot be predicted which tidal signal will occur during any given storm in the future several signals will be considered. The tide is modeled as a sinus with constant amplitude and a semi-diurnal period (M2, 12 hours and 25 minutes). Different signals are created by using three different amplitudes, a low, medium and high amplitude of 1, 1.25 and 1.5 meters. As the phase of the tide is also unknown each storm will be run with a tide starting in a random phase (bottom panel of Figure 4-4). The total water level is then made up out of the surge level and the tidal elevation. Because of this the highest water level during the storm does not necessarily occur at the moment the surge reaches its highest level as could also happen during a real storm.

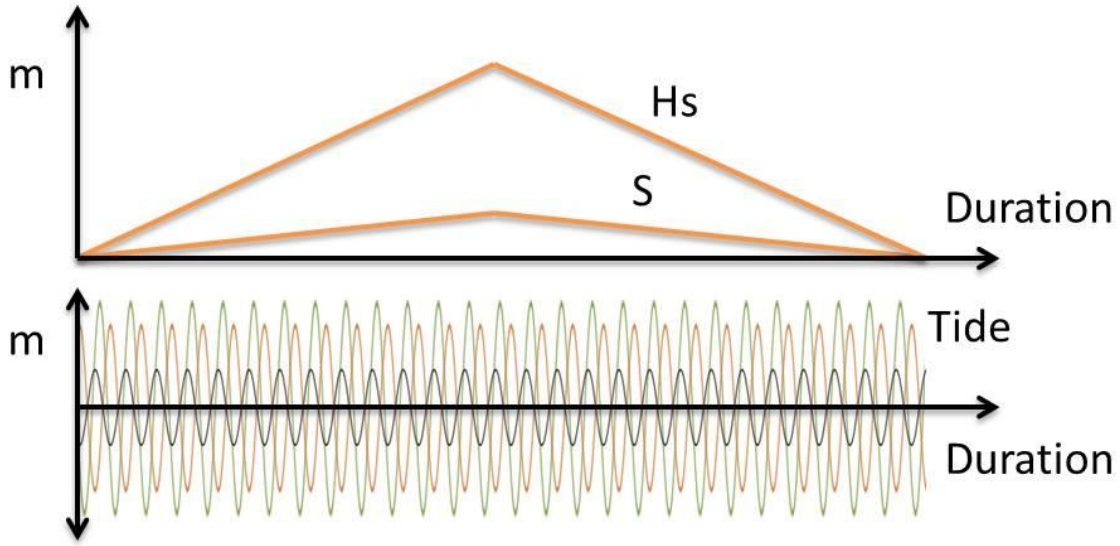

Figure 4-4 Modeling of a storm. Top panel: wave height and surge level. Bottom panel: tidal signal.

4.3 SYNTHETIC DATASET DEVELOPMENT

This chapter will cover the creation of the synthetic storm dataset that is needed to train the BN. The creation of the synthetic storms follows a statistical approach using copulas to generate a dataset that is representative of reality. Another, easier, method of obtaining a synthetic storm dataset is by using linear regression lines of the available datasets. This method excludes the natural variation that is seen in the data and is therefore not preferred.

Actual data of storms that have occurred in the region are used to create copulas between individual variables. These copulas bear the pattern of the actual data and can be used to sample new data points from. These new data points form the set of synthetic storms that will be used to train the BN.

There are three main steps that have to be taken to produce the synthetic dataset that will be treated in the following sections:

1. Fit marginal distributions to the measured data points.
2. Fit different copulas to each pair of variables, using the marginal distributions.
3. Sample from the copula to obtain the synthetic data.

The dataset will be created for the range of storms with a significant wave height of at least 3 meters to a maximum of 8.8 meters. Based on expert knowledge storms with a lower than 3 meter significant wave height have no significant impact. Storms with a significant wave height of 8.8 meters have a return period of once every 100 years, which is the locally the preferred safety level.

4.3.1 AVAILABLE DATA

There are two datasets available for the region concerning storm events, to be named the *surge dataset* and the *duration dataset*. Both datasets have been supplied by the University of the Algarve.

The first dataset, the “surge dataset”, covers a period of ten years and includes wave and surge data from June 1997 until June 2007. The wave data have been recorded by the Instituto Hidrográfico de Portugal with a directional wave-rider buoy, placed offshore at 93 meters depth near Praia de Faro. The wave buoy records 20 min every 3 hours, except during storms when it records every half an hour. Storms have been selected by peak over threshold analysis with a threshold of 3 meters. Only storms originating from the western quadrant (180°-270°) have been included in the dataset. Storm surge levels have been obtained from a nearby Spanish network of tide gauges, some 80km east of the study area. (Rodrigues et al., 2012)

The second dataset, the “duration dataset”, is based on the Faro wave buoy (see Figure B-8) and consists of a 20 year wave record from January 1993 until December 2013. Waves are recorded every 3 hours for 20 minutes to give a significant wave height. A peak over threshold analysis with a threshold of 2.5 meters identifies individual storms. Storms separated with less than 24 hours are grouped together in the dataset and the total duration of individual storms has also been included.

Both datasets contain information about the dependency between the significant wave height and the peak period. The “duration dataset” dataset covers a longer time period, including the period of the “surge dataset”, and will therefore be used to evaluate the dependency between the significant

wave height and the peak period. The “duration dataset” is further used for the relation between the wave height and the storm duration. The “surge dataset” is only used to determine the relation between the wave height and the storm surge levels. The raw data points of the surge and duration dataset are shown in Figure 4-5 and Figure 4-6. The three plots show a significant correlation between each variable, but also show a large scatter. It is important that these characteristics are also exhibited by the synthetic dataset.

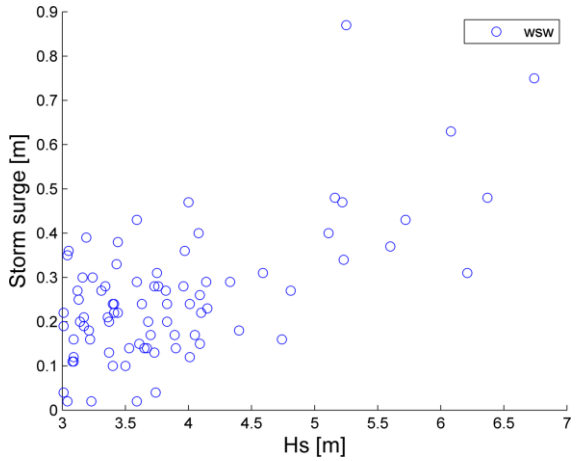

Figure 4-5 Surge dataset: storm surge and significant wave heights for individual storm events.

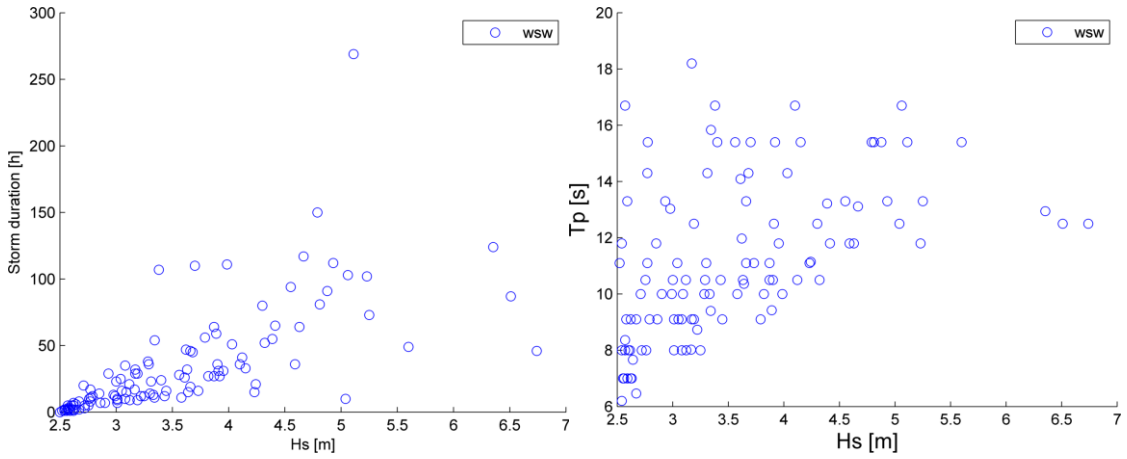

Figure 4-6 Duration dataset: storm duration and peak period for individual storms plotted against the significant wave height.

4.3.2 MARGINAL DISTRIBUTION FITTING

The first step in obtaining a synthetic dataset that mimics the properties of the measured dataset is to look at the individual variables separately. The marginal distribution is the probability distribution that describes a single variable. In this case there are two sets of random variables: the two datasets described in the previous section: *surge & duration*. The individual variables in the sets are the significant wave height, the peak period, the storm duration and the surge level.

For each variable several distributions have been fit and ranked according to the Akaike information criterion (AIC).[9] The top ranked distributions are shown in Table 4-1, for an overview of all fitted distribution types and their ranking is referred to appendix C.1. The best fit for the surge is a generalized extreme value distribution. This, however, allows the variable to reach values smaller than zero, indicating that a negative surge is possible which does not make sense. The second ranked distribution for the surge is a Weibull distribution, which does not allow the variable to reach values below zero, and is therefore the preferred distribution.

Table 4-1 Best marginal distribution fits according to the Akaike information criterion.

Variable	Best fit distribution	Used distribution
Hs (duration)	Generalized Pareto	Generalized Pareto
Tp (duration)	Lognormal	Lognormal
Duration	Exponential	Exponential
Hs (surge)	Generalized Pareto	Generalized Pareto
Surge	Generalized extreme value	**Weibull**

To visualize the fit of the distributions the empirical cumulative distribution function (ECDF) and the parametric cumulative distribution function (CDF) can be compared. The parametric cumulative distribution function uses the formulation of a certain distribution type (e.g. Generalized Pareto) and shows the probability of a certain value of the random variable. The ECDF estimates the underlying CDF of the points in the dataset. It makes steps of 1/n in which n is the number of data points. In Figure 4-7 the ECDF and CDF for the significant wave height of the *duration* dataset is shown. For the marginal distribution fits of each of the variables and their respective parameters is referred to appendix C.1. These plots give an extra, visual, check on top of the AIC, concluded is that all distributions are well estimated by the chosen parametric distributions and that no anomalies are seen.

[9] There are more criteria available to rank the fit of different probability distributions on a dataset. The choice to use the AIC is based on its use in Corbella & Stretch (2013) that follows a similar approach.

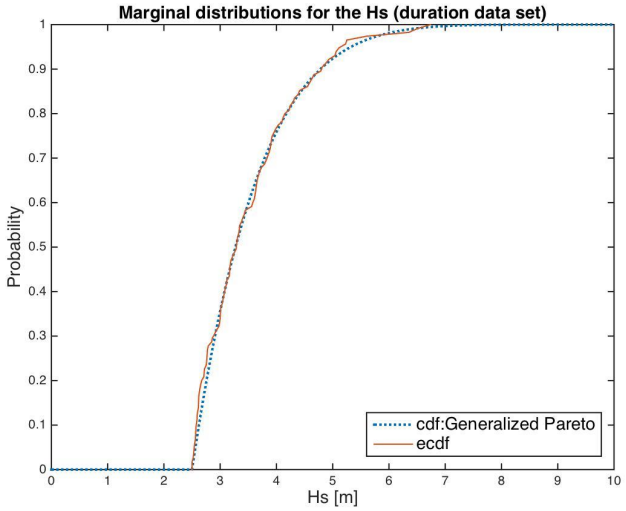

Figure 4-7 ECDF and CDF for the significant wave height of the duration data set.

4.3.3 COPULA FITTING

This chapter will explain how the different copulas are fit to the variable pairs. Three variable pairs have been identified for which a copula needs to be fit:

1. Significant wave height and the storm duration (Hs – Duration)
2. Significant wave height and the storm surge level (Hs – Surge)
3. Significant wave height and the peak period (Hs – Tp)

For the 1st and the 2rd pair four copulas are tried: Gaussian, Clayton, Frank and Gumbel. An additional copula has been fit for the 3rd pair, the skew-t copula, as the dependency displays certain skewness.

The first step in fitting the copulas is to transform the marginal distributions of the variables to uniform distributions by using their respective CDFs, as explained in section 2.1.5**.** The transformation is supposed to return uniformly distributed variables, which can be visually checked with a histogram. Figure 4-8 shows a scatterplot and histogram of the transformed variables duration D and significant wave height H_s. The histograms show that the transformed variables are not fully uniformly distributed, but uniform enough. This is explained by the fact that a limited amount of data points are available (in the order of 100 points). If more data points are available a better fit for the parametric distribution could be reached and thereby a more uniform distribution for the transformed variables. The scatterplot indicates the dependence structure of the random variables, isolated from their marginal behavior. This is the behavior that the copula will try to mimic. For plots of the other variables is referred to appendix C.2.

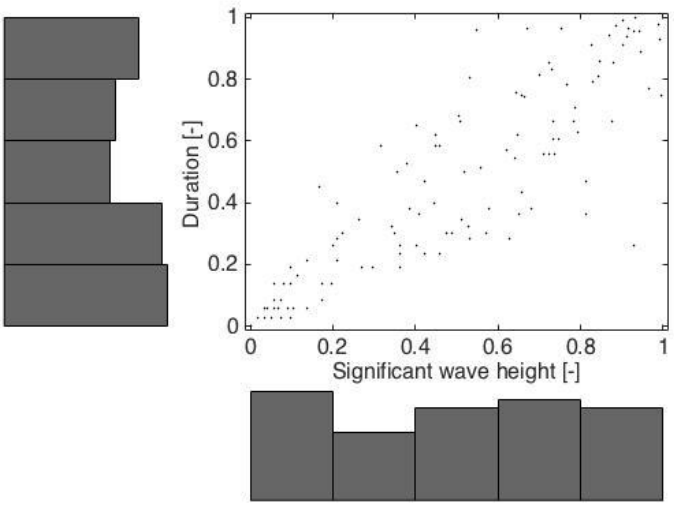

Figure 4-8 Scatterplot of the uniformly distributed variables of Hs and Duration.

The second step in fitting the copulas is determining their parameters. The parameters are based on the rank correlation, Spearman's rho, between the variables is determined by:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{6}$$

For which $d_i = x_i - y_i$ and x x and y are the ranked variables (wave height and duration) and n the size of the sample.

The association between Spearman's rho and the copula parameters has been determined using the Matlab statistics toolbox for each variable pair. Using the same Matlab toolbox samples have been created for each copula. In Figure 4-9 the same scatter plot as in Figure 4-8 is shown for the variable pair D and H_s, now including 10,000 samples of the four copulas. For the copula fits of the other variable pairs is referred to 8C.2. Each copula shows a slightly different behavior in trying to mimic the original dataset. For example, consider the Clayton and Gumbel copulas, seen in the right panels in Figure 4-9. These copulas show opposing tail dependencies; the Clayton copula has a concentration of points in the bottom left corner that fans out towards the top right where the Gumbel copula shows the opposite.

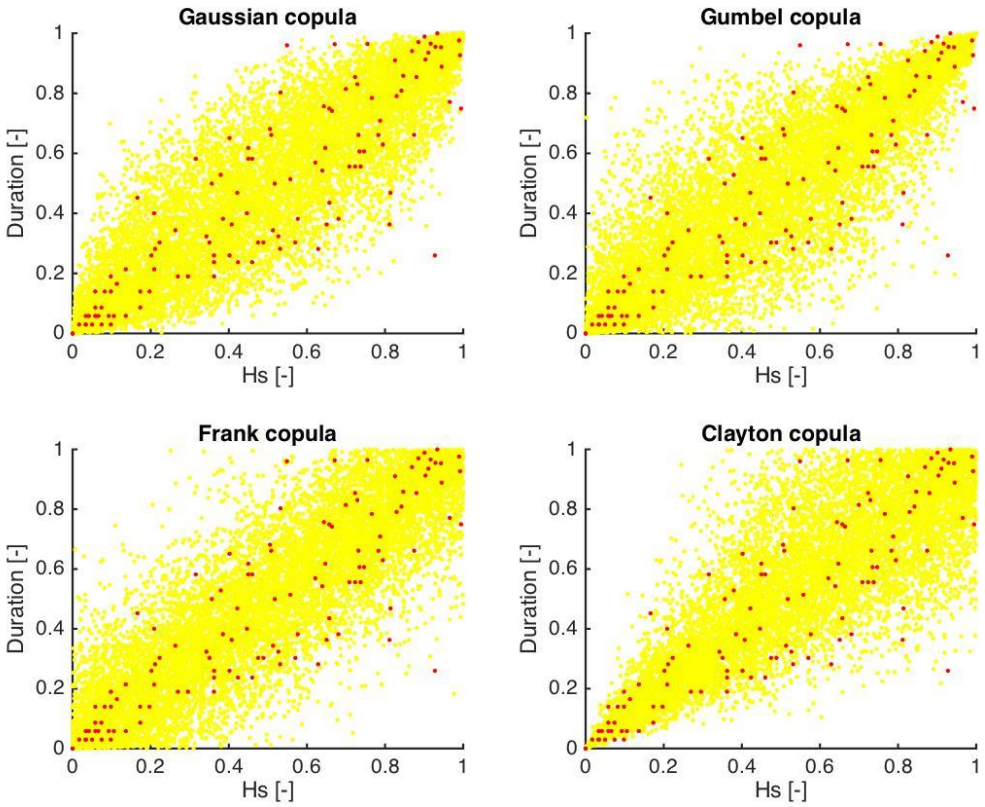

Figure 4-9 Different copulas fit to the Hs and Duration dataset. The red dots indicate the original dataset and the yellow dots the data points generated with the copula.

For the variable pair significant wave height H_s and peak period T_p an extra copula is fit: the skew-t copula, for which the fit is shown in Figure 4-10. The variable pair $H_s - T_p$ shows skewness in the scatterplot leaving the bottom right corner empty. This occurs due to the physical process of wave breaking; when the wavelength becomes too short with respect to the wave height a wave breaks. As the wave length and period are related this is reflected in the dataset. The copula describing the dataset therefore needs to mimic this behavior.

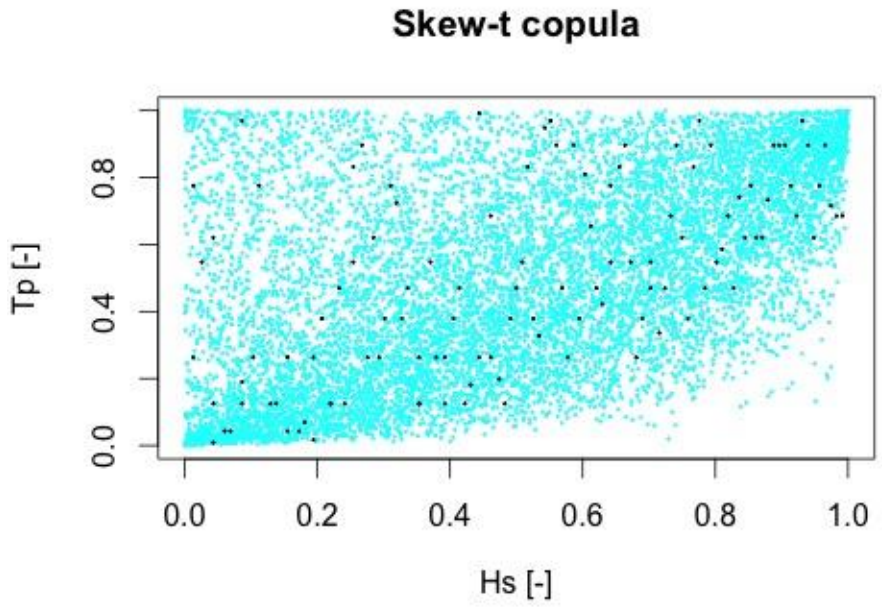

Figure 4-10 Skew-t copula fit for the wave height and peak period.

Other than looking at the scatter plot of Figure 4-10, the goodness of fit is judged with two methods: one numerically and one visually as done in Jäger & Morales (2014).

The numerical method is based on the Cramer von Mises (CvM) statistical test and the results of are summarized in Table 4-2. The test is relative, meaning that the scores can only be compared within the variable pair. A lower test score indicates a better performance of the copula fit.

Table 4-2 Cramor von Mises test results for each copula and variable pair.

Copula	Hs - Duration	Hs – Surge	Hs – Tp
Clayton	1.1137	1.6256	1.6256
Frank	1.4905	1.2669	1.2669
Gaussian	**1.0852**	1.1930	1.1930
Gumbel	1.3927	**0.9756**	**0.9756**
Skew-t	-	-	1.0083

The visual test compares the copula fits to the actual data in the quadrants of the scatterplots. For this the intervals $[0,0.25]\times[0.75,1]$, $[0.75,1]^2$, $[0.75,1]^2$ and $[0.75,1]\times[0,0.25]$ are considered (top right, top left, bottom right and bottom left, respectively). At the intervals is looked at the conditional probability for, in this case, $P(T_p \mid H_s)$. So for the top right quadrant the markers show the probability that $T_p > q$ given that $H_s > q$, where q rangers from 0.75 to 1. The result is seen in Figure 4-11. The plots can be interpreted as following: if the conditional CDF of the copula is above the original data points ($\bigtriangledown$) it is overestimating and if it is underneath it is underestimating (in the respective quadrant).

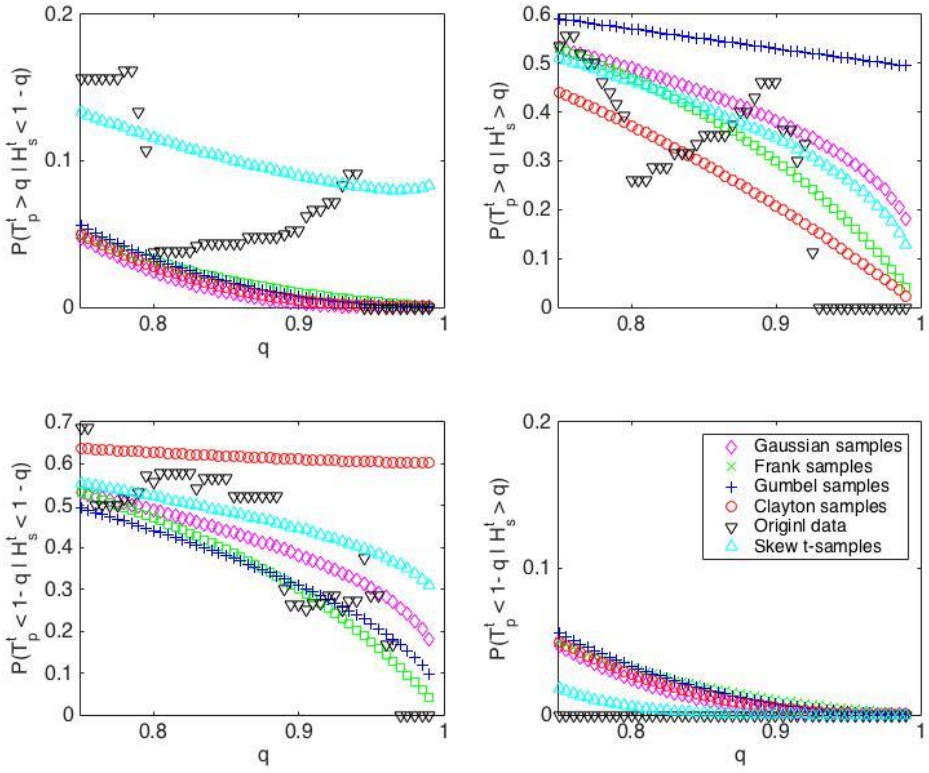

Figure 4-11 Comparison of conditional probabilities for $T_p \mid H_s$ on the intervals $[0, 0.25]\times[0.75, 1]$, $[0.75, 1]^2$, $[0.75, 1]^2$ and $[0.75, 1]\times[0, 0.25]$ (top left, top right, bottom left and bottom right respectfully).

Based on the CvM and the visual test the Gaussian copula is chosen for the $H_s - D$ variable pair and the Gumbel copula for the $H_s - S$ variable pair (Table 4-2, Figure C-10 and Figure C-11 in appendix C.2.). For the $H_s - T_p$ variable pair the Gumbel copula shows a slightly better performance than the other copulas according to the CvM test.[10] However, when looking at the Gumbel fit (+) in the quadrants in Figure 4-11, it is not performing so well when comparing it to the skew-t copula. Gumbel underestimates in the top left panel, and highly overestimates in the top and bottom right panels. The skew-t copula may overall not have a better fit than the Gumbel copula, but performs better in the extremes and better models the physical process of wave breaking. For these reasons the skew-t copula is chosen to model the $H_s - T_p$ variable pair. The parameters for the copulas are:

1. $H_s - D$: Gumbel, $\rho = 0.8814$
2. $H_s - S$: Gaussian, $\alpha = 1.4621$
3. $H_s - T_p$: Skew-t, $\rho = 0.71$, $v = 5$, $\gamma_1 = 0.9$, $\gamma_2 = 0$

[10] Prezado leitor, Muito obrigado por ler isso! Ganhou uma cerveja! Assino, o autor.

4.3.4 SAMPLING FROM THE COPULAS

One million points are generated with each copula to generate a large synthetic dataset. These points are uniformly distributed on the interval $[0,1]$ and need to be transformed to usable values. This is done with the marginal distribution functions of the variables; by applying the inverse of the CDF to the variable (see section 2.1.5).

The result is shown in Figure 4-12, the blue dots are the original dataset and the orange dots the one million synthetic events. Since the BN needs to be trained evenly over the whole range of possible storms, a uniformly distributed set of storms is chosen on the interval $[3m, 8.8m]$ significant wave height from the one million synthetic events to be run in XBeach: the dark orange dots in Figure 4-12.

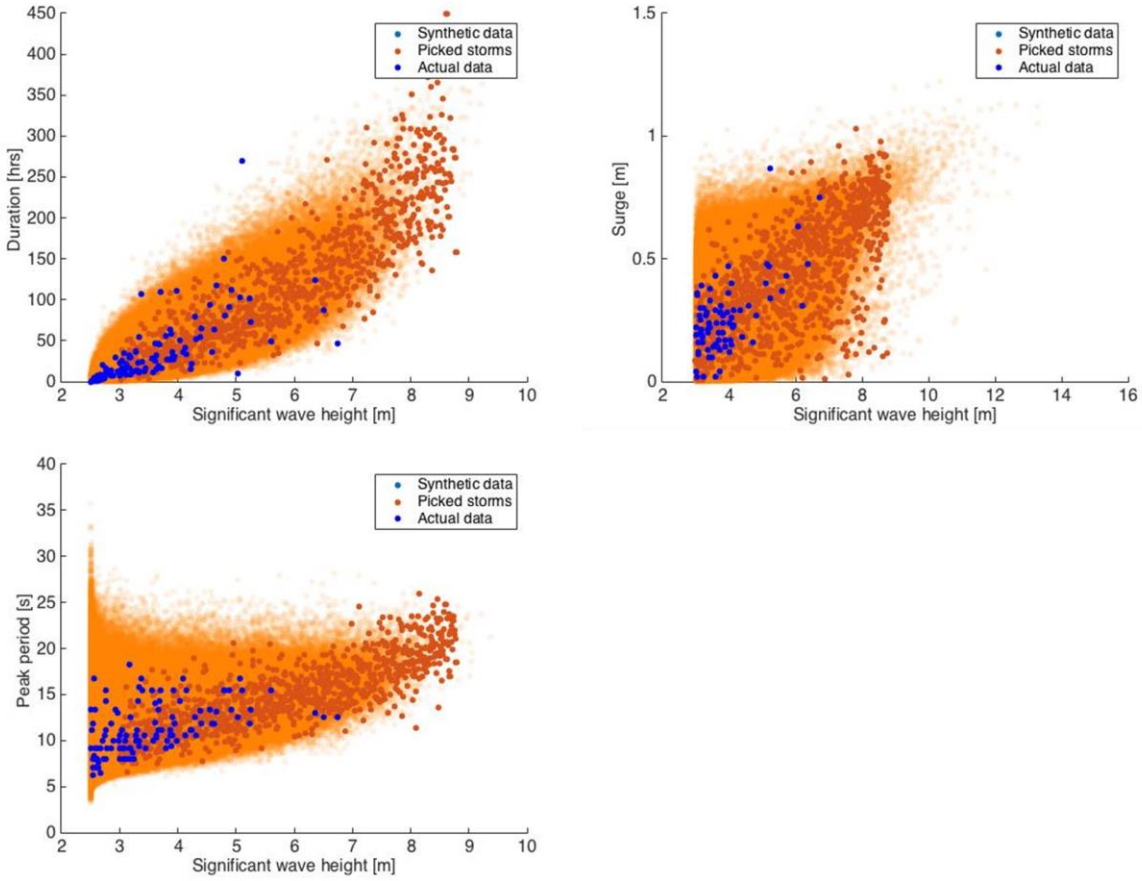

Figure 4-12 Synthetic dataset for the variable pairs: H_s D , $H_s S$ and $H_s T_p$.

4.4 XBEACH

The local storm conditions and morphological changes are determined with the process based model XBeach (Deltares, 2015). The setup of the model, the calibration efforts and the efforts done to speed up the model are discussed in detail in appendix B. In this section an overview will be given of the most important aspects of the XBeach modeling.

4.4.1 TOPOGRAPHY AND BATHYMETRY

Topographic and bathymetric information has been provided by the University of The Algarve. Information from the summer of 2009 is available for a stretch of 8 kilometers on the Ancão peninsula. It consists of bathymetric cross sections of the wet areas and a LIDAR of the dry area. The LIDAR gives an accuracy of 2 by 2 meters.

An overview of the available topography and bathymetry is given in Figure 4-13. The area of interest has a total width of 1 kilometer and is centered on the entrance road and parking lot. The red box indicates the research area. Another important location that lies within the red box is a former video monitoring station that has been used to collect data during a number of storms. This dataset will be used for calibrating the model.

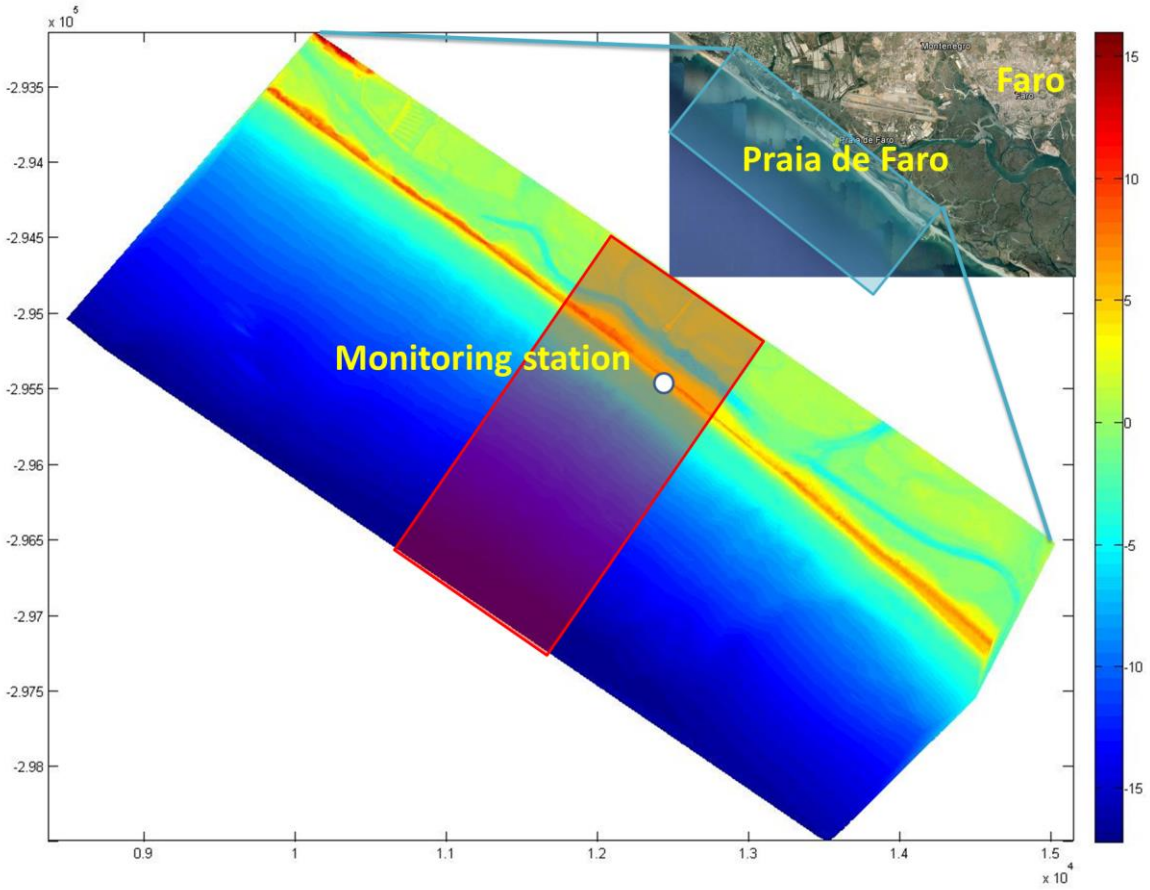

Figure 4-13 Available LIDAR and bathymetric data and location of research area.

4.4.2 GRID

The XBeach model setup is a balance between several sometimes contradicting requirements. The most important requirements that have to be dealt with are listed here.

- The depth at the offshore boundary has to be sufficiently large for the assumption of intermediate depth for long wave propagation to hold.
- The width of the domain has to be determined so as to minimize the effect of the shadow zone.
- The grid resolution at the shoreline has to be fine enough to model the beach face and dune erosion and overwash.
- The total number of grid cells should be kept at a minimum to reduce the runtime.
- In case of varying grid cell sizes the transition between two cells has to be smooth enough to prevent numerical instabilities.

First the grid is set up for which the grid cell sizes have to be small enough to give detailed information about dune and beach face erosion as well as overwash events. However, the total number of grid cells needs to be kept at a minimum to reduce the total runtime of a model.

To do this the grid cells are varied in the alongshore and cross shore directions to have the highest level of detail only in the area of interest. The model is centered on the area of interest and therefore the grid cells are finest in the center of the grid. In the cross shore direction the grid cells vary from 0.65 at the beach face to 30 meters offshore. Alongshore they are varied from 5 meters in the center, around the parking lot, to 20 meters at the edges. The total number of grid cells is 154 x 229 = 35266.

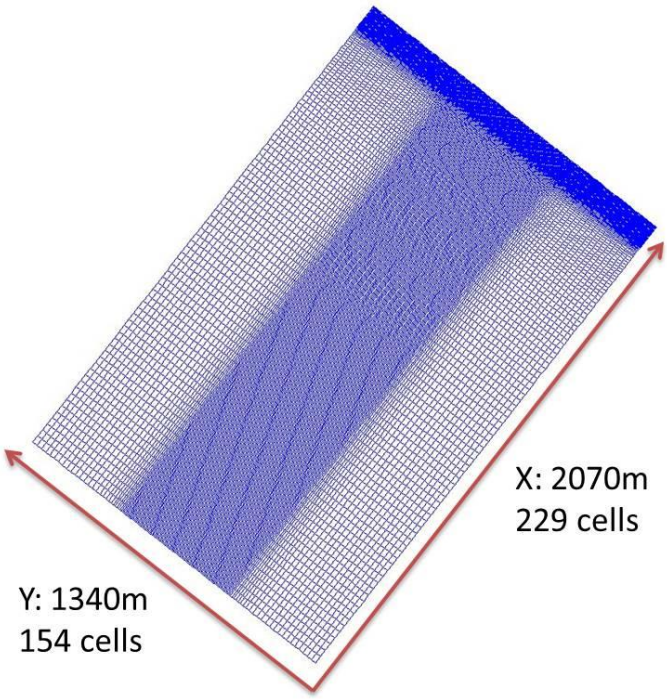

Figure 4-14 XBeach grid setup:. (max dx = 30m, min dx = 6.5m, max dy = 20m, min dy = 5m. The y direction is alongshore and the x direction is cross-shore pointing towards the shoreline.

Each grid point is filled with topographic and bathymetric information. The grid has been extended seawards to a depth of 20 meters, to fulfill the deep water assumption. For this it reaches 2 km in the cross shore direction. Alongshore the model spans 1.3 km and is centered on the parking lot and includes the video monitoring station: the most vulnerable part of Praia de Faro. The row of dunes and a part of the land behind it is included to monitor the onshore effects of the storms.

The waves considered for this model enter the domain at an angle creating the shadow zone indicated in Figure 4-15Figure 4-14. This zone contains less wave energy because waves can only enter the domain at the offshore boundary. The results at the shoreline are therefore distorted and have to be neglected from the output.

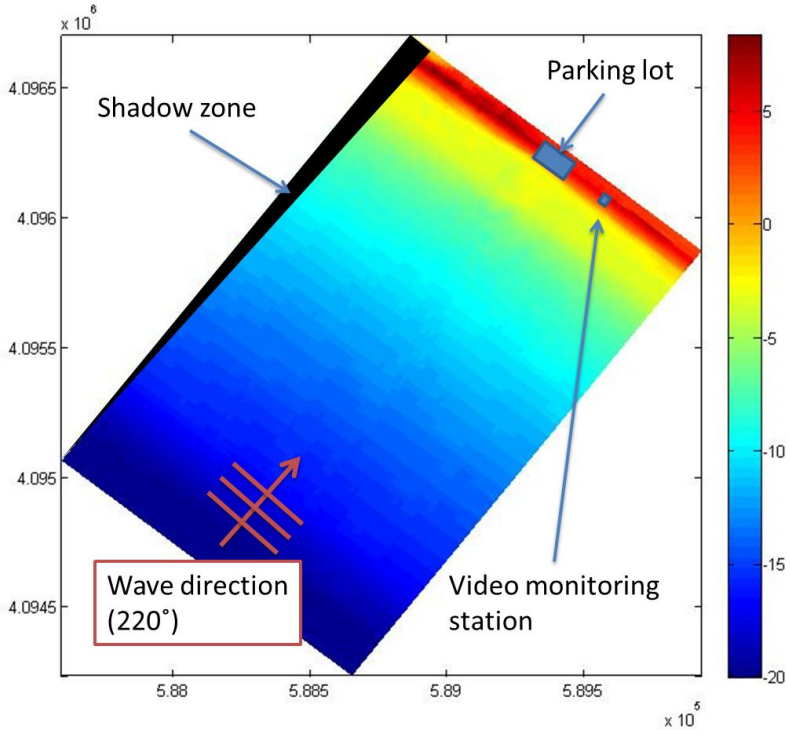

Figure 4-15 XBeach model setup showing the most important locations, the wave direction and shadow zone.

4.4.3 HYDRODYNAMICS

There are different options to supply XBeach with wave boundary conditions, ranging from stationary conditions to time varying two-dimensional wave spectra. The wave conditions supplied are deep-water waves and have to be supplied at the offshore boundary. Since there is no available spectral analysis for the sea state in the south of Portugal a JONSWAP spectrum is assumed.

A wave spectrum is a statistical description of the sea state as a density spectrum of the frequencies of the waves (Figure 4-16). The JONSWAP spectrum is a type of spectrum that assumes that the sea state is not yet fully in equilibrium with the wind. This assumption means that the fetch and duration that the wind blows is limited, which is often true. The JONSWAP spectrum has one peak; this means that a mixed sea state (more than one peak) cannot be properly represented and is thus a limitation. A mixed sea state consists of sea waves and swell, both of which may have significantly different

periods and directions. In this setup the boundary conditions have been supplied as hourly one-dimensional (JONSWAP) wave spectra. (Holthuijsen, 2007)

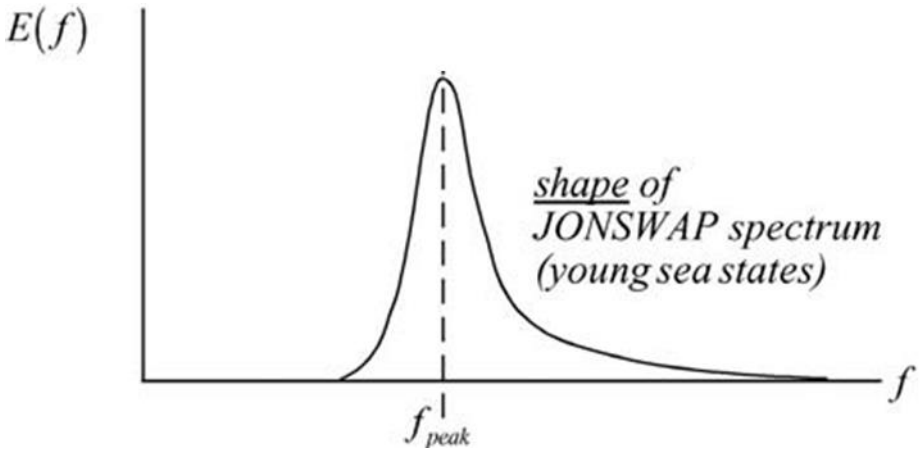

Figure 4-16 Shape of the JONSWAP spectrum. The energy is given in the vertical axis and the frequency of the wave on the horizontal axis. (Holthuijsen, 2007)

Water level boundary conditions in XBeach, similar to the waves, can also be defined stationary or varying in time. Water boundary conditions can be used to impose a tide and surge on the model. For the modeling of storms at Praia de Faro it is important to include these, as there is a significant variation in the water level, especially due to the tide and less due to storm surge. The tidal level and the storm surge are added together to give one water level for each time step as described in the beginning of this chapter.

4.4.4 CALIBRATION EFFORTS

Calibration of the XBeach model is necessary to ensure the results obtained are reliable and of high enough accuracy. The required outputs of the model are the coastline erosion and the occurrence and magnitude of overwash events. These physical processes occur around the beach and dune face and the calibration of the model is therefore focused on this area. This section does not treat the steps taken in the calibration but rather focusses on the issues faced during the calibration. The calibration itself is elaborated in appendix B.4.

The calibration is performed using a dataset that covers a number of high energy events in the period December 2009 until January 2010. During this period wave and tidal data have been collected by offshore measuring stations and topographic data have been collected for a small area at the beach using a video monitoring station.

The area over which topographic information has been collected is very small compared to the computational grid described previously (±100 m wide). Furthermore the area only covers the dry part of the beach. This limits the calibration to a small area within the computational grid and also to the upper part of the cross-shore profile. This makes calibration difficult for two reasons: (1) the model cannot be run on the measured initial profiles as they are only available for a very small region and (2) there is no information on the 'wet' cross-shore profile so that the performance of the

model for this area cannot be checked. The dry profiles do, however, give good insight in the height of the run-up due to storms.

The hydrodynamics conditions are well described but lack any information about the shape of the wave spectrum. An assumption therefore has to be made about the shape of the spectrum which is an issue as the response of the beach is sensitive to the shape. For example: a very narrow spectrum enhances the groupiness of the waves and thereby the forcing of the infragravity waves. A larger forcing of infragravity waves will result in higher runup values. A spectrum with two peaks means that there is a mixed sea state containing swell and wind waves. Swell waves have higher periods and therefore carry more energy resulting in a higher impact. These things are unknown and therefore an issue for the calibration.

Due to these reasons it was not possible to fully calibrate the XBeach model. However, the calibration settings should be in the correct order of magnitude.

4.4.5 RUNTIME REDUCTION

The XBeach model has to be run for all storms of the synthetic dataset. Currently this means that the model has to be run 300 times (100 storms were picked from the synthetic data set that have to be run for the three tidal amplitudes), making it desirable to reduce the runtime. Two options are used to create a reduction in the runtime: (1) use the internal morphological acceleration factor and (2) reduce the number of grid cells of the model setup.

The morphological acceleration factor in XBeach can be triggered with the keyword *morfac*. This option decouples the hydrodynamical and morphological time in the model. Setting *morfac = 10* results in 10 times more erosion and sedimentation for a given time step. To obtain the same results as without the *morfac* option the model run time is shorten by a factor ten. (Roelvink et al., 2010)

The grid size of the original model has been reduced to reduce the run time of XBeach. The grid size has been approximately divided by 2, 4 and 8 in the along shore direction (Table 4-3 & Figure 4-17). The cross shore direction has not been reduced as a reduction in the cell sizes here will result in losing the necessary detail to model the beach and dune erosion, or lose the shape of the profiles completely.

Table 4-3 Number of grid cells of the courser grids.

Model	X cells	Y cells
Original	229	154
/2	229	77
/4	229	39
/8	229	20

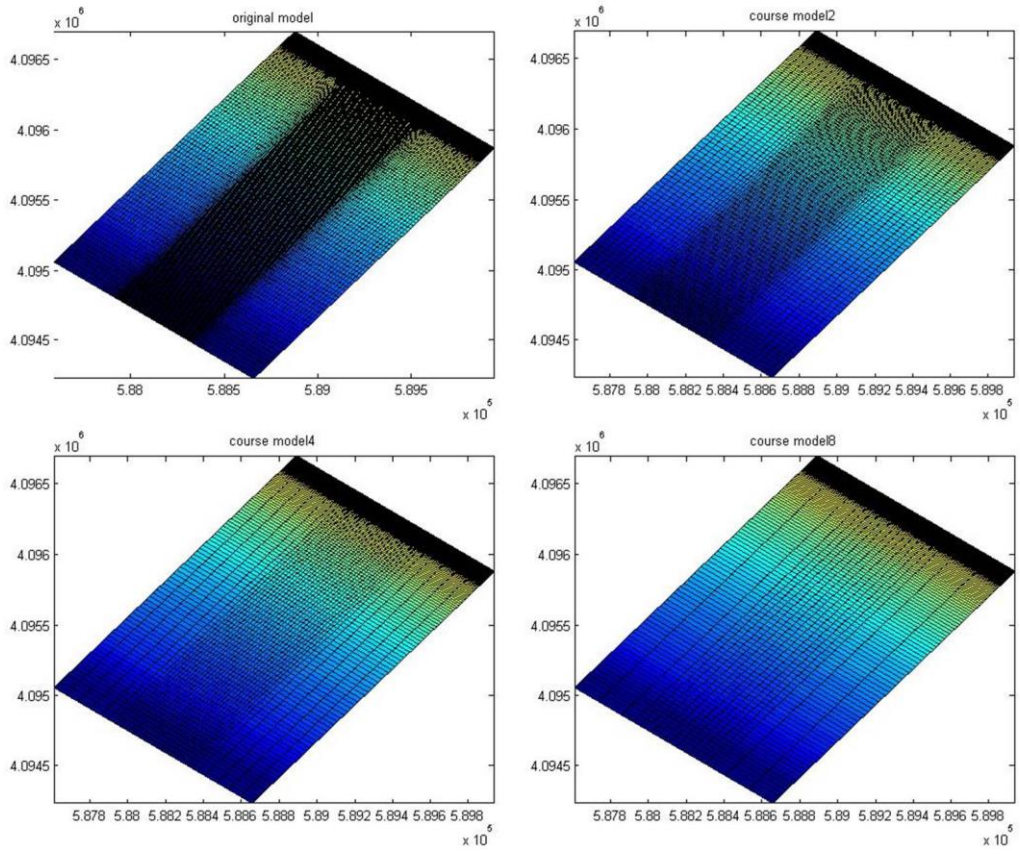

Figure 4-17 Courser grids. Top left: original. Top right: original/2. Bottom left: original/4. Bottom right: original/8.

The coarser models are tested with three storms events (B, D and E as described in appendix B.48B.4). The grid with 1/8 times the grid cells of the original immediately has been discarded as the large differences between the cross shore and alongshore cell sizes resulted in numerical instabilities in XBeach causing runs to crash.

The results for the final sedimentation and erosion of the original model have been compared to the final results of the courser models. The comparison has been done using a Brier skill score between 0 and 1, for which 1 is the best score (Table 4-4) (van Rijn et al., 2003).

$$BSS = 1 - \left[\frac{\left\langle \left(\left| z_{b,c} - z_{b,o} \right| \right)^2 \right\rangle}{\left\langle \left(z_{b,i} - z_{b,o} \right)^2 \right\rangle} \right] \tag{7}$$

In which $z_{b,c}$ =computed bed level of the courser grid, $z_{b,o}$ = is the computed bed level of the original grid, $z_{b,i}$ is the initial bed level and $\langle ... \rangle$ indicates averaging over the considered grid points. Only the morphological active zone has been used for the comparison to prevent bias in the skill score due to the fact that large parts of the cross sections (at large depth or behind the dunes) do not change at all. Since the brier skill score results are excellent the most course grid is chosen as the final grid.

Table 4-4 Brier skill scores for the courser grids, for events B, D and E.

Model	Event	Brier skill score (0 to 1, 1 is best)
2	B	0.9715
2	D	0.9811
2	E	0.9696
4	B	0.9599
4	D	0.9726
4	E	0.9712

4.5 BAYESIAN NETWORK DEVELOPMENT

The BN is a computational tool to describe a system in a probabilistic way. The system consists of variables and dependencies. The relation between variables in a system comes from prior knowledge about the system. In this case the system has to describe the effect of a storm on the coastline of Praia de Faro. The general configuration of the system is therefore relatively straightforward and shown in Figure 4-18.

The intensity of coastal hazards at a certain stretch of coastline is determined by the intensity of the storm and the local characteristics of the coastline. A storm is described by offshore boundary conditions, as treated in the previous section. The local hazards have been identified as overwash and coastal erosion. Where overwash causes damage due to salt water flooding houses and streets and coastal erosion may undermine the foundation of the local structures. This study focusses on the link from offshore storm conditions to onshore hazard intensities and does not incorporate the link to damages. This link has been research before and can be found in van Verseveld et al. (2015).

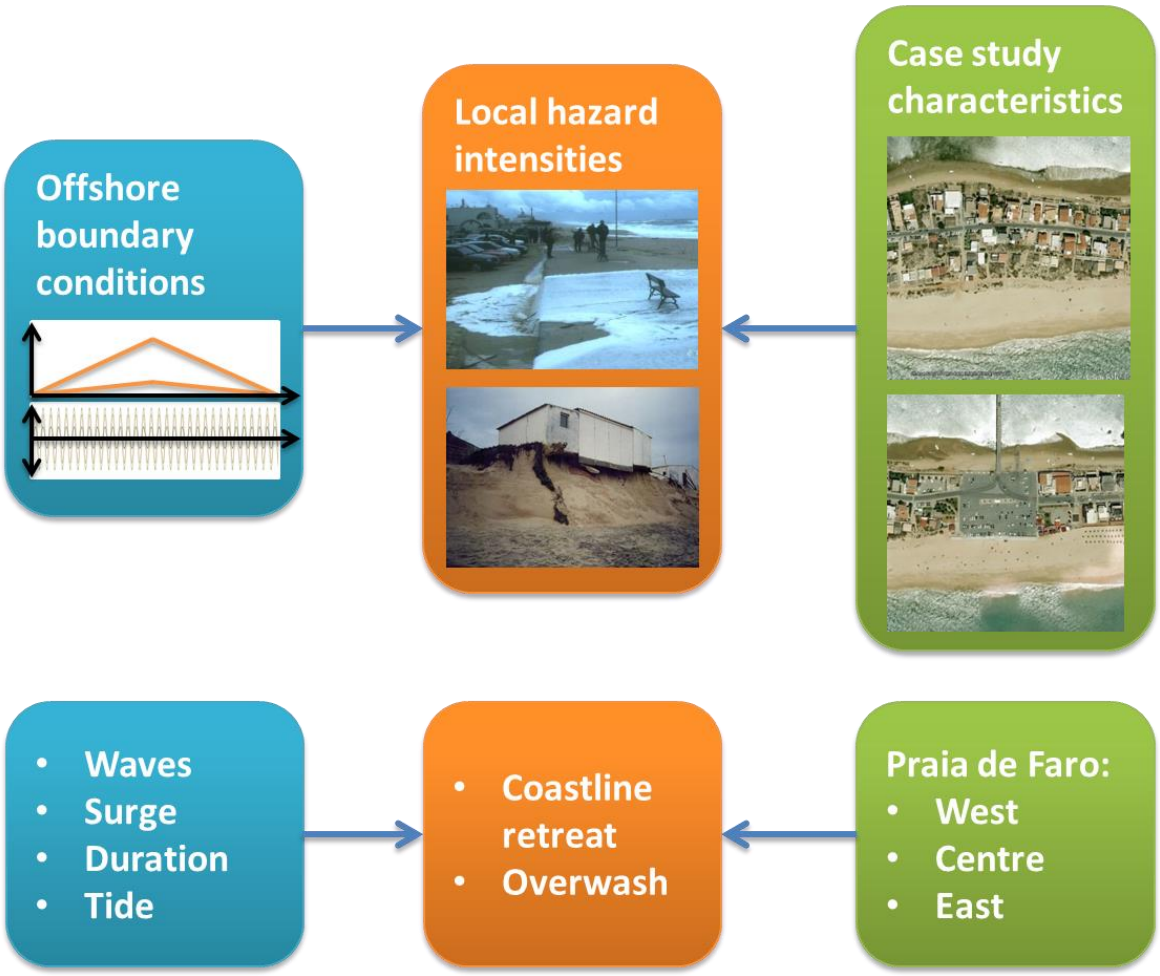

Figure 4-18 Relation between offshore boundary conditions, local hazard intensities and the case study site characteristics.

Setting up a BN roughly consists of two steps: (1) setting up the structure by identifying the variables that need to be included in the network and the dependencies between the variables and (2) training the network with cases (a dataset).

4.5.1 STRUCTURE OF THE BAYESIAN NETWORK

As explained in previous sections, the structure of the BN consists of nodes and arcs. The nodes contain the individual variables and the arcs describe the dependencies. Each node is subdivided into bins that indicate the state of the variable and may be discrete or continuous and should cover the full range of the dataset.

The hazard intensities are grouped into two variables: the maximum offshore wave height and water level that occur during the peak of the storm. The coastal hazards are divided into coastline retreat (changed to coastal buffer in the next section) and overwash, where the overwash is defined both as a water depth and flow velocity. The overwash depth is defined as the water depth between the current bathymetry and water level. The coastline retreat and overwash depth are measured in meters and the overwash velocity in meters per second. The locations are grouped into one node, where the different areas will be included as discrete bins. The maximum wave height and maximum

water level during a storm have been identified as the storm boundary conditions. The water level includes the tidal level and the storm surge and the wave height includes the other storm parameters (duration and peak period) through the copulas. One may argue that another arc is needed from the wave height to the maximum water level since the surge is included in both variables and they are therefore no longer independent. This is true, however, there is no real gain from adding this node since both variables are conditioned-on when using the network. Furthermore, the boundary conditions are defined offshore, meaning that the water depth does not influence the maximum wave height.

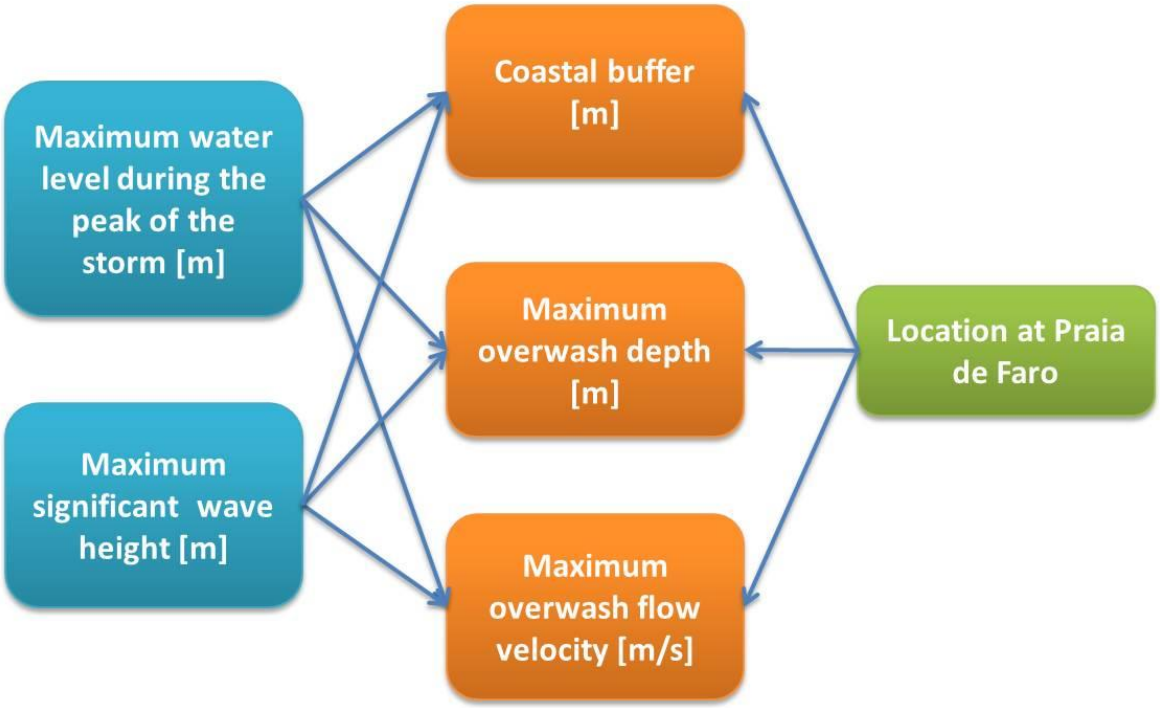

Figure 4-19 BN structure, showing the variables and dependencies in nodes and arcs.

4.5.2 CREATING CASES FOR THE BAYESIAN NETWORK

The in- and output of the storm run in XBeach are used to generate the cases that are needed to train the BN. Information is needed to satisfy all the individual nodes:

Input:

- Locations at Praia de Faro.
- Maximum water level
- Maximum significant wave height

Output:

- Coastline retreat
- Maximum overwash depth
- Maximum overwash velocity

The stretch of coastline that is modeled in XBeach varies in the lateral direction. There height of the dunes and land varies as well as the distance at which houses and infrastructure are built from the sea. Because of the inhomogeneity the coastline should be separated into areas with similar coastal response. The western part of Praia de Faro consists of a row of dunes in front of the houses, that is a higher than in the rest of the peninsula. The parking lot in the central part of Praia de Faro is the lowest lying area where there is no row of dunes left between the sea and the infrastructure. In the east houses are partially build on top and behind a row of dunes that is slightly lower than in the eastern part.

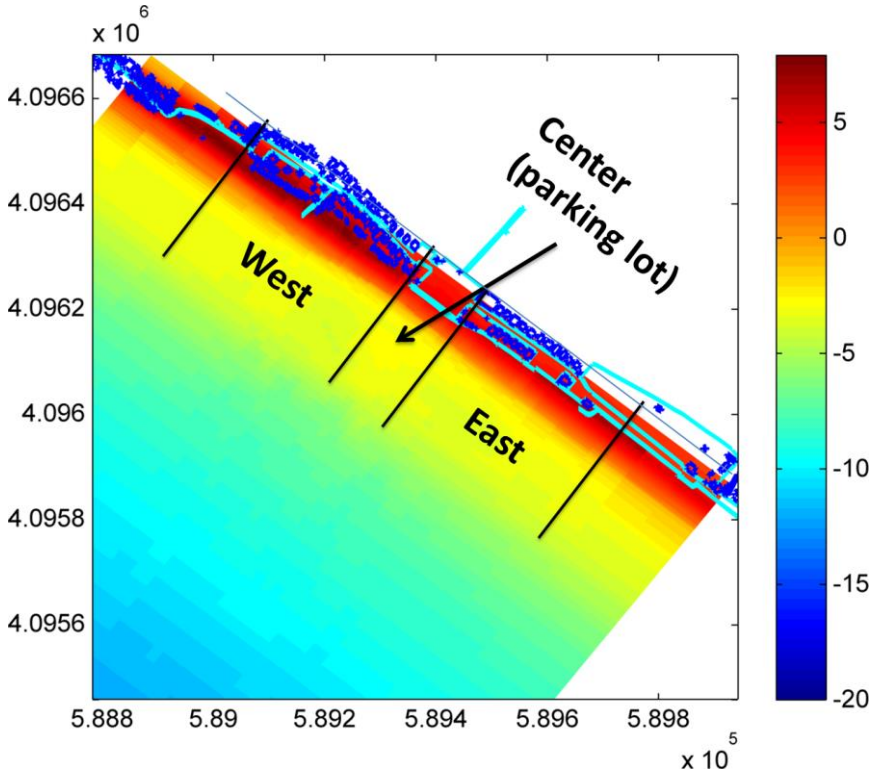

Figure 4-20 Locations specified for the Bayesian Network.

The maximum water level is obtained from the input file that specifies the water levels at the offshore boundary. These water levels are a combination of the tidal signal and the surge. A water level needs to be defined that is normative for the storm. The assumption is made that joint occurrence of the highest waves and water levels during the peak of the storm are normative for the effect of the storm on the coastline. The peak of the storm is then defined as a percentage of the total duration of the storm in which the highest conditions occur. The maximum water level and wave height are then extracted that occurs during this time.

Since the peak of the storm is arbitrarily defined as a percentage of time that the storm intensity is highest, several percentages are evaluated, namely 10%, 15%, 20%, 25% and 30%. Resulting in five different sets of training data for the BN that will be evaluated for the best performance. Figure 4-21 illustrates that for a different definition of the duration of the peak of the storm a different maximum water level may be obtained. The effect of the choice for a certain duration of the peak of the storm will most likely only have an effect on storms with a shorter duration. If the tide is semi-diurnal and the duration of the peak of the storm is less than 12 hours, than it is possible that the

highest tidal level does not coincide with the highest wave and surge levels. The shorter this period, the larger the chance becomes that this happens. Oppositely so for larger storms in which at least one full tidal cycle occurs during the peak of the storm.

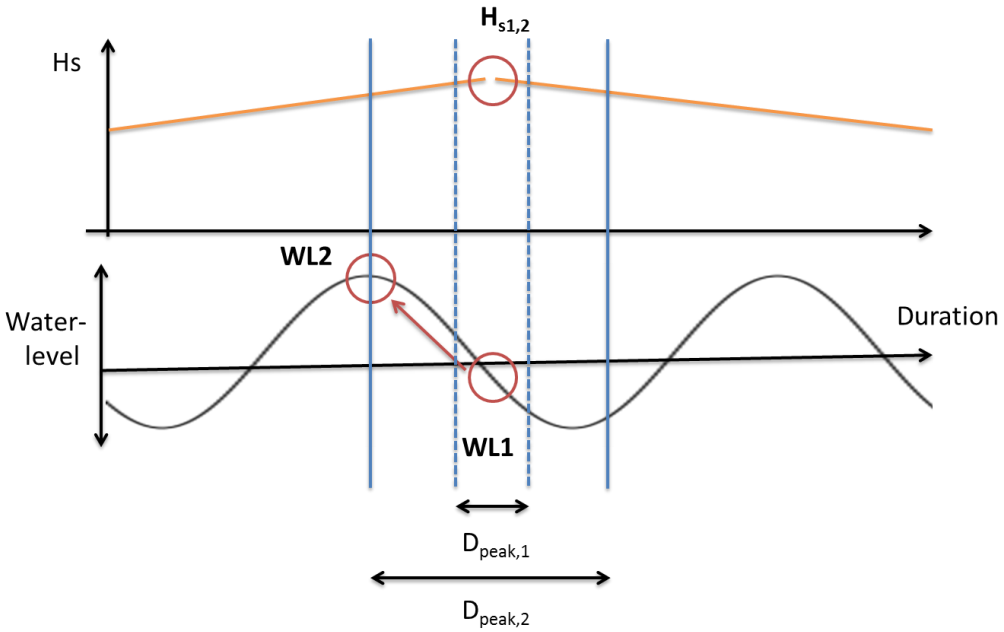

Figure 4-21 Definition of the peak of the storm, zoomed in on Figure 4-4.

Different houses and infrastructure are located at varying distances from the sea. A good indication of the hazard due to coastline retreat for a certain area is the number of meters of coastline left untouched until the first structure. Rather than indicating the coastline erosion, a better indication is thus specified as a *buffer*. In Figure 4-22 two plots are shown of different parts of Praia de Faro. The plots show the locations of buildings and infrastructure, the XBeach grid and a plot of the total sedimentation and erosion that occurred due to an arbitrary storm. The pink dots are the grid points that are closest to either a house or infrastructure. The buffer is then deduced per cross section by calculating the distance between the pink dots and the first grid cell that indicates erosion. The smallest distance is then chosen to be normative for an entire area.

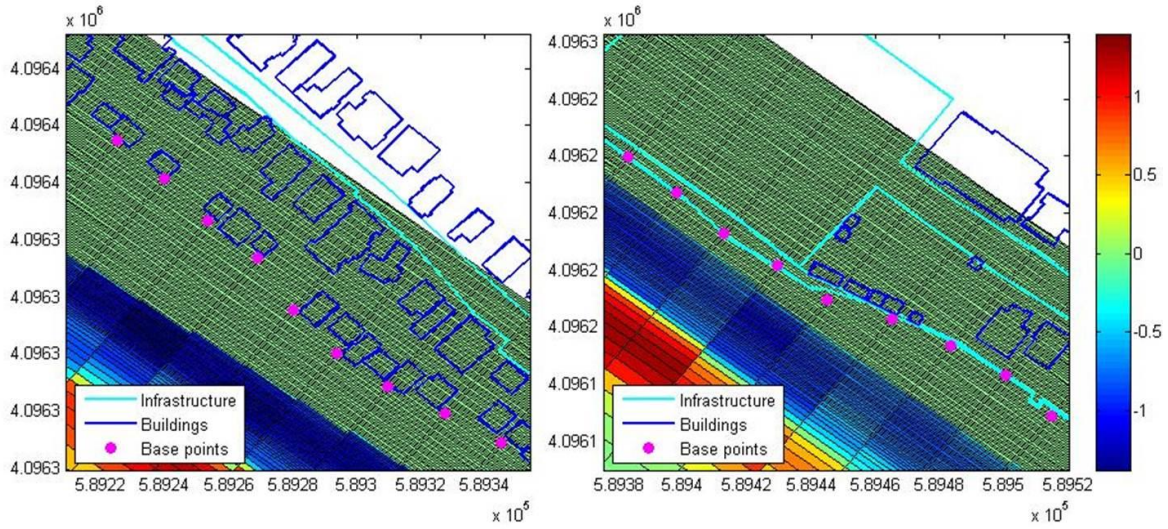

Figure 4-22 Locations of points to which the coastline retreat is measured based on the location of the structues.

Overwash is an event that occurs during a very short period of time. To record it output has to be generated for water levels and flow velocities at very high frequency. Recording this output at high frequencies is problematic as it generates very large data files, quickly filling up the storage space of a computer or hard drive. To solve this problem output is only generated for 19 points, rather than for the full grid, at intervals of 2 seconds. The 19 points are strategically located on the highest point of the dune of the cross-section it is on. If any water level and flow velocity is recorded at these points it can only be due to overwash or inundation. As we know that the water level will not exceed the top of the dunes during the storms that are modeled, it must be due to an overwash event. The output locations are indicated in Figure 4-23.

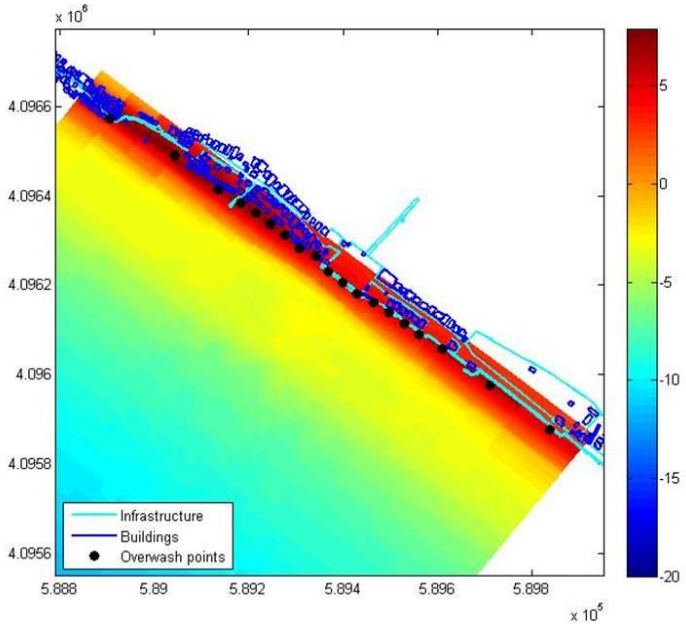

Figure 4-23 Overwash output points in XBeach.

4.5.3 DISCRETIZATION OF THE BAYESIAN NETWORK

There are different ways to go about discretizing the random variables in the BN. All methods, however, have to balance a number of competing interests. A heuristic method is explored.

1. The bin sizes have to be narrow enough to provide forecast utility to the end user.
2. The bin sizes have to be wide enough to contain enough data to satisfy the joint probability tables.
3. Since the BN comes at the end of a model train of which the water level and wave height is used for as the input of the BN, the bins sizes of these input variables have to be wider than the expected error in their prediction.
4. The edges should preferably be rounded to either the nearest integer, half or tenth.

Continuous random variables can be described with a probability density function (PDF). The area underneath the probability density function is described with its integral: the CDF. When a random variable is discretized into bins, its probability density function is also discrete and becomes a probability mass function (PMF). For the probability mass function to give a good description of the behavior of the variable the bin sizes should not be too large such that large changes in its shape are lost in the size of the bin. To get a feeling for the right bin sizes one can look at the empirical cumulative density function (ECDF) of the discrete random variable to find near linear sections. Large changes in the shape of the ECDF should also appear in the shape of the PMF.

Another aspect is that the dependence between variables is not constant over the range of values that a variable can take; as soon as a variable is conditioned on the distributions for other variables change and so do their ECDFs. The discretization should in theory be done such that the interdependence is maximized. This is not checked by looking at the individual ECDF's and as such the followed method is an approximation of a good discretization.

The ECDFs and the PMFs of the input and the output variables of the XBeach simulations are shown in Figure 4-24. The PMFs are shown with 10 bins of equal size for each variable. From this figure can be seen that the variables show large differences in their distributions and therefore should be treated individually. The resulting discretization is shown in Figure 4-25, and elaborated below.

The water level has very little data up to the 1 meter point (top left panel, Figure 4-24). The bins before this point contain very little data and will satisfy only few joint distribution tables in the BN. It makes sense to group the low water levels up to a certain threshold and have smaller bins for the larger water levels. Starting around 0.5 meter the shape of the ECDF is starting to change so this value is chosen as the threshold. Furthermore, bin sizes larger than 0.5 meter do not make a lot of sense since the expected error in the prediction of the water level from the overlying model train is expected to be in this range.

The wave height is uniformly distributed between 3 meters and 8.8 meters. The bin width is determined at 1 meter, with bins spanning from 3 to 9 meter, to absorb any small errors in the prediction of the wave height (from the preceding model train) and for ease of use for the end user.

The distribution of the coastal buffer shows that a large part of the dataset is focused at the edge of the spectrum (middle left panel, Figure 4-24). In this initial discretization these are captured in a bin

with a width of 10 meters. The ECDF, however, indicates that most of the information may be captured in a much smaller bin size, to increase the accuracy for this part of the network. This bin spans from -65 to -60 meter, indicating that all dry parts of the XBeach model are actually affected. The improved discretization is seen in the middle left panel of Figure 4-25.

Similar to the coastal buffer variable, the overwash variables also have a lot of information focused at the edge of the spectrum (middle right and bottom left panels, Figure 4-24). The reason for this is that for a large part of the modeled storms no overwash occurs and the values in the most left bins are zero. To capture these zero values and distinguish no overwash from overwash a bin is created with edges 0 to 0.1 (see Figure 4-25 for the result).

The overwash water level variable also contains a very little amount of data for a depth larger than 3 meters. To satisfy the joint probability tables and since any overwash larger than this depth is not very interesting to the end user anyhow, they are grouped together into one bin.

The resulting discretization is shown in Figure 4-25 and summarized in Table 4-5. Observed is that the new discretization follows the changes in the ECDF much better than before, meaning that less information is lost in the discretization.

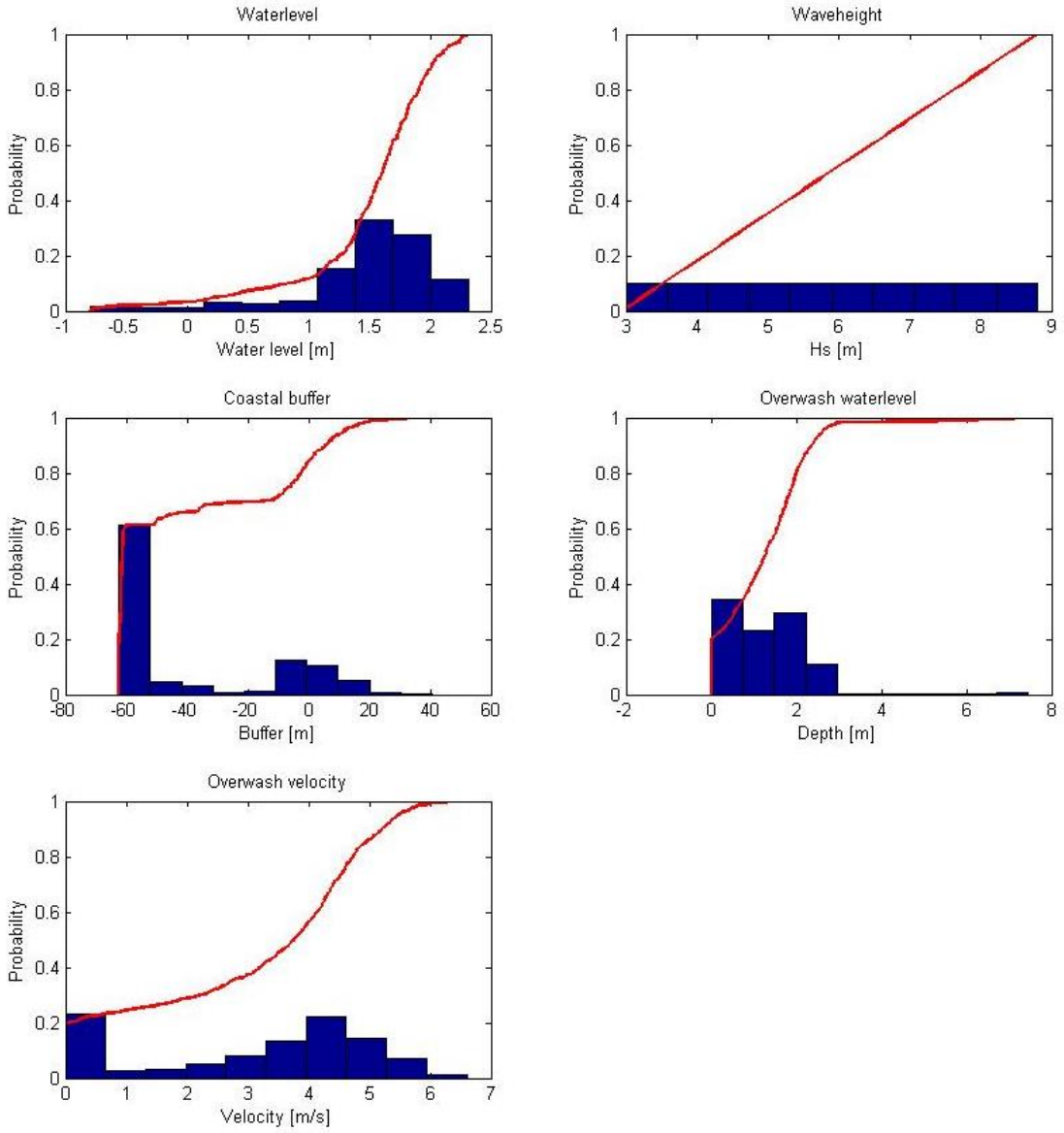

Figure 4-24 Empirical cumulative density functions and probability mass functions for the initial state of the random variables in the Bayesian Network.

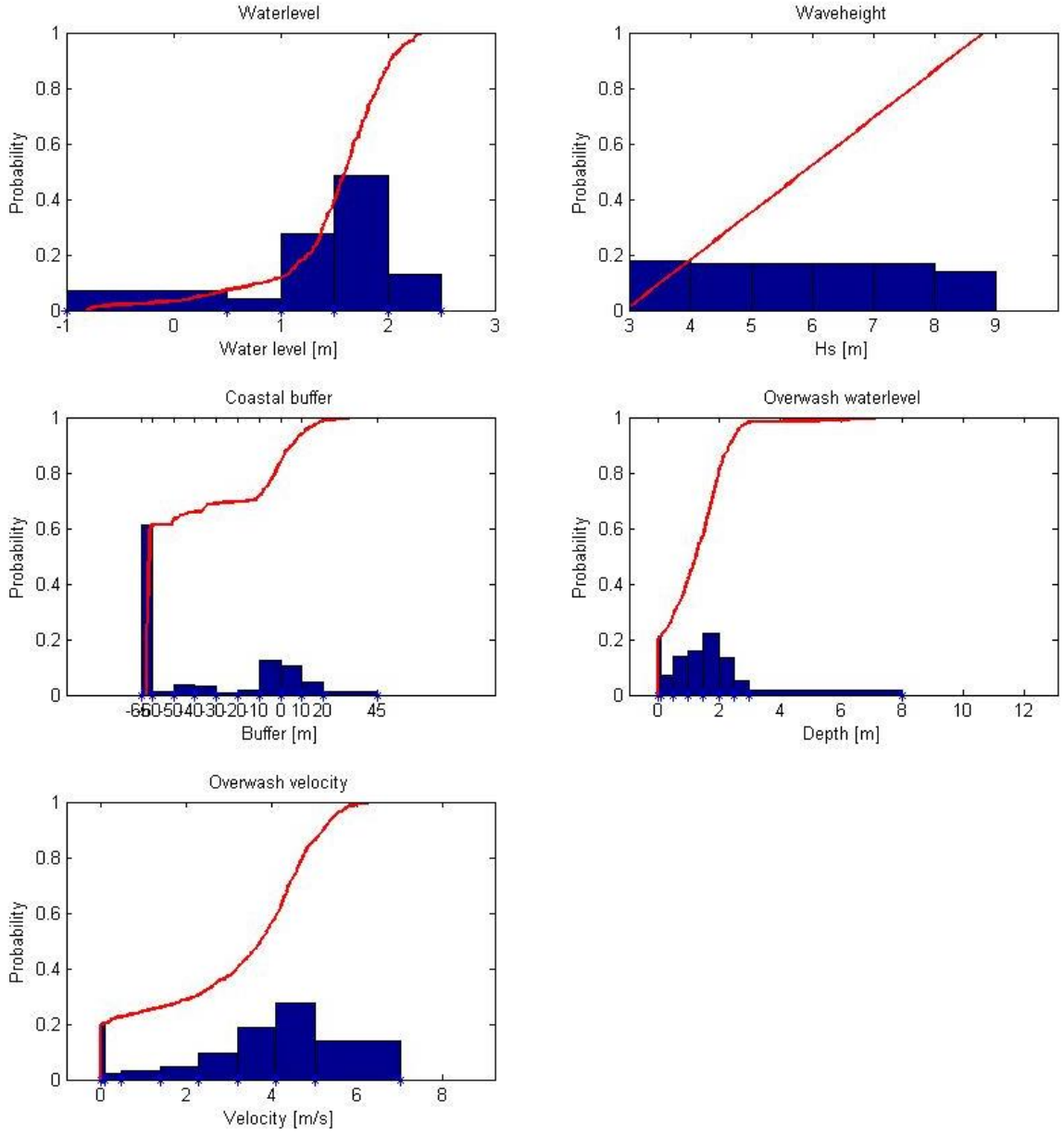

Figure 4-25 Empirical cumulative density functions and probability mass functions for the initial state of the random variables in the Bayesian Network for optimized bin sizes.

Table 4-5 Edges of the bins of the discretization of the random variables.

Variable	Bin edges									
Water level	-1	0.5	1	1.5	2	2.5				
Wave height	3	4	5	6	7	8	9			
Coastal buffer	-65	-60	-50	-40	-30	-20	-10	0	10	>20
Overwash water level	0	0.1	0.5	1	1.5	2	2.5	>3		
Overwash velocity	0	0.1	0.5	1.4	2.3	3.2	4.1	>5		

4.5.4 TRAINING THE BAYESIAN NETWORK

The BN is implemented using the software package Netica (Norsys, 2013).

Prior to training the network with data the marginal distributions are seeded with uniform distributions. This means that the poorly trained parts of the network will return a uniform distribution as a result, rather than giving a zero probability. For example if a node consists of two bins and is not supplied with any information the BN will give each bin an equal probability: fifty-fifty. The effect of this is that a poorly trained network will have a relatively large rest probability in the less likely states and in reverse a well-trained network will have small “rest probabilities” in the less likely states.

The nodes and arcs are setup conform the previous sections and trained with the dataset that is extracted from the results of the XBeach model runs. A total of 900 cases are supplied to the network, 300 for each location and 100 storms per tidal signal (amplitudes 1, 1.25 and 1.5). There are, however, 5 different dataset to train the networks with, depending on the definition of the peak of the storm, to be evaluated in the next section. The network trained with the peak of the storm defined as 10% of the duration is shown in Figure 4-26, for the other networks is referred to appendix D8D.

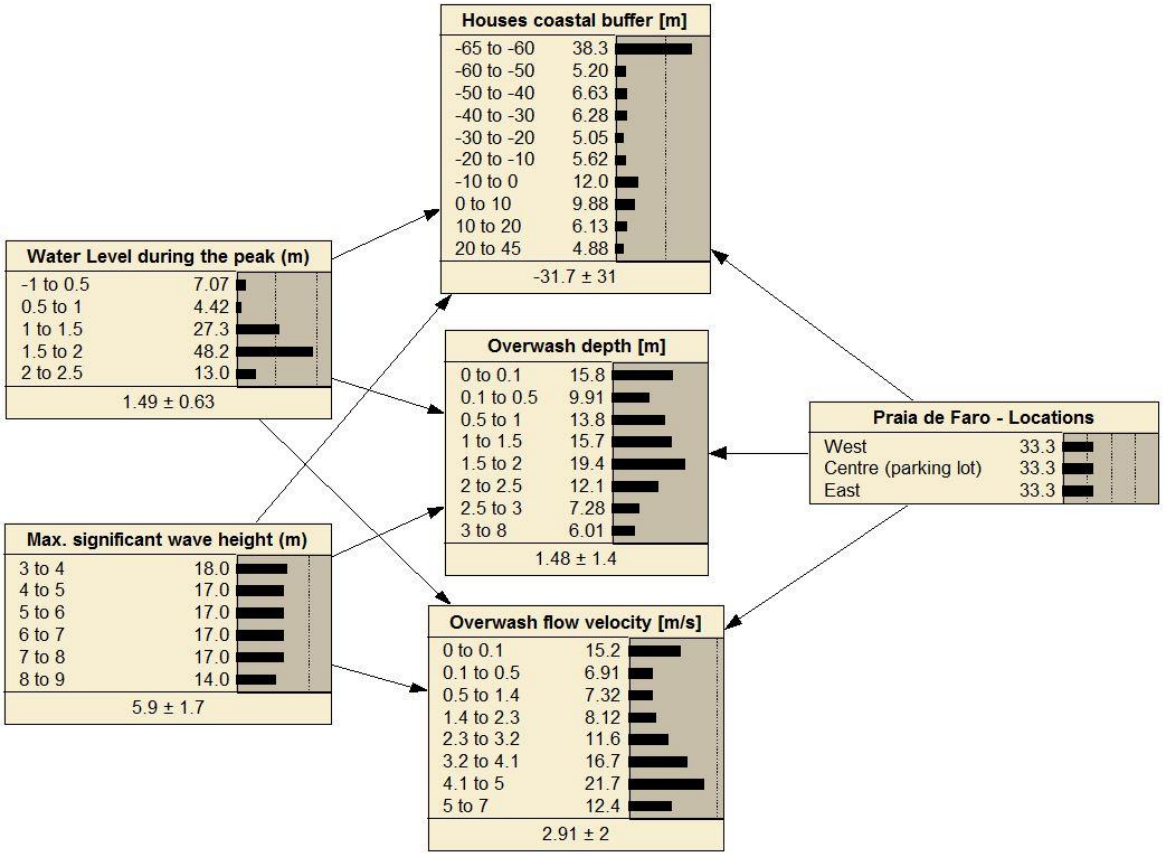

Figure 4-26 Bayesian Network for Praia de Faro, trained with all cases for the peak of the storm duration of 10%.

5 RESULTS

In this chapter is first explained how a BN can be tested and how its performance can be measured. The results are then presented for different aspects: the overall performance, the performance of different networks, the performance over the range of predictions and the value of a prediction.

5.1 METHOD OF TESTING THE BAYESIAN NETWORK

The purpose of the BN is to provide a decision maker with a tool to take justified action. The main question is then whether or not the outcome of the BN can be trusted, or more specifically: what is the quality of a prediction of the BN? The Netica software package provides several tools to analyze the BN. The basis of the analysis is to test the trained network with cases to see how well the predictions of the BN agree with actual cases. To test the different networks they are trained with 90% of the data and tested with the remaining 10%, chosen randomly. Since the nodes of interest are the hazard indicators (coastal erosion, overwash depth and velocity) the network has to be tested on the prediction of these nodes. The test works as follows:

The coastal hazard nodes are treated as unobserved nodes (coastal buffer, overwash depth and velocity), this means that Netica does not know their values during its inference. This is intuitive since prior to a storm these effects are unknown, but the boundary conditions of the storm are known and so are the different areas. A case file is entered into Netica, who will use the three observed variables (water level, wave height and location) to make a prediction for the three unobserved variables by conditioning on the known variables. It will do this for all case files that are entered for the test and produce several test results in the following four flavors (Norsys, 2015):[11]

1. Confusion matrix

The confusion matrix compares the predicted states of each unobserved node in the BN to the actual state of the node that is being tested. The most likely state in the BN is chosen as the prediction value and entered in the confusion matrix. If the BN is performing well then the numbers along the diagonal will be large compared to the others.

2. Testing real value

Several values are calculated that give insight in the performance of the prediction of a certain node.

a. *Mean absolute error (MAE)* and the *root mean square error (RMSE)*, defined as: $\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i|$, and $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2}$. In which f_i is the mode of the prediction and y_i is the actual value and n are the number of cases for which the BN is being evaluated. The MAE gives the average of the absolute errors where the RMSE gives the average of the root of the squares of the errors.

[11] Since the case file entered does contain the actual values for the unobserved nodes, tests can be done to see if the BN is able to predict these values well.

b. $\frac{Standard\ deviation}{Mean\ absolute\ error}$: It is assumed that the error is normally distributed. Dividing the standard deviation by the mean absolute error gives a percentage. This percentage makes it possible to compare the performance of different nodes.

c. *Confidence intervals*: Another good indication of the performance of a node is by looking at the confidence intervals. A confidence interval indicates a certain percentage of the sample to be within that interval. They are often expressed in standard deviations, where one, two and three standard deviations indicate the confidence intervals of approximately 68%, 95% and 99.7%. These are the standard deviations of the mean absolute error. The confidence interval then counts for the mode of the prediction, giving the range of values the state of a variable may take with a certain level of confidence.

3. Scoring rules

Several scoring rules are used to describe the performance of the BN. The reason for using several methods is that when they are in agreement it gives more solid proof of the performance.

a. *Logarithmic loss* $= \frac{1}{n}\sum_{i=1}^{n}\left(-\log\left(P\right)\right)$, $P = \begin{cases} P_{ci}, & \text{if the prediction is correct} \\ 1-P_{ci}, & \text{if the prediction is incorrect} \end{cases}$

in which P_{ci} is the probability predicted for the correct state of the i-th case and n is the number of cases being tested. The logarithmic loss is calculated by giving a score to the most likely prediction. For example: if the prediction of the most likely state is 75% and it is the correct state it would receive a score of $-\log(0.75) = 0.29$ but if the prediction proved incorrect it would receive the score $-\log(1-0.75) = 1.39$. The average of all cases is taken, producing a value between zero and infinity. The goal is to minimize this score.

b. *Quadratic loss (Brier skill)* $= \frac{1}{n}\sum_{i=1}^{n}\left(1-2P_{ci}+\sum_{j=1}^{k}P_j^2\right)$, in which P_{ci} is the probability predicted for the correct state, P_j is the probability predicted for state j and k is the number of states. The quadratic loss score also gives a score to a prediction and is (in this form) rated between zero and two, with zero being the best score. [12]

c. *Spherical payoff* $= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{P_{ci}}{\sqrt{\sum_{j=1}^{k}P_j^2}}\right)$, in which P_{ci} is the probability predicted for the correct state, P_j is the probability predicted for state j and k is the number of states. The spherical payoff is scored between zero and one with one being the best score.

4. Calibration table

The calibration table links the belief of a certain state to the percentage that the actual outcome was in that specific state. It therefore indicates whether or not the belief of a certain state is appropriate,

[12] Not that this differs from the more well-known Brier skill score that is scored between 0 and 1.

or well calibrated. The table gives the belief of a variable in one column and the percentage of times the actual value was in that state in the adjacent column. Since this is very ambiguous an example is given:

Consider a part of Table 5-6, for the coastal buffer node, shown below in Table 5-1. The belief columns indicate a range of beliefs (or predictions) in percentages for a certain state (in this case -65 to -60 m). These predictions are based on the 90% of the cases on which the network is trained. The actual columns give the percentage of times the actual outcome was in this state of the tested 10% of the cases, while the belief of the adjacent column occurred. So for all tested cases that caused a belief between 10% and 40% for the state -65 to -60 m, 54.2% of the time the actual value of the tested cases fell within this bin. This does not mean that 54.2% of all tested cases fall within this bin, this only counts for cases that have caused a belief of 10% to 40% for this state.

Table 5-1 Part of a calibration table for the coastal buffer node, for illustrative purposes. The belief column indicates a range of predictions of a state and the actual columns indicate the percentage of times the actual number occurred when the prediction was in the range of the belief column.

Coastal buffer	Belief (%)	Actual (%)	Belief (%)	Actual (%)	Belief (%)	Actual (%)	Belief (%)	Actual (%)
-65 to -60	0-10:	0	10-40:	54.2	40-60:	82.6	60-80:	100

5.2 PREDICTIVE SKILL OF THE BAYESIAN NETWORK

To get an overview of the performance of the BN it will be evaluated from different angles. The overall predictive skill will be looked at as well as the predictions over the range of the variables, since some parts of the network may perform better than others. The five different definitions for the peak of the storm will also be evaluated simultaneously. All these together should give an idea of the value of a prediction made by the BN.

5.2.1 OVERALL PERFORMANCE

The overall predictive skill is determined by the results of the confusion matrix and the skill score ratings. Since the confusion matrix for a tested node is rather large and there are three nodes per network and five tested nodes they are not given. Instead a set of values are given that summarize the outcomes of the confusion matrices.

The scoring rules used are the Log Loss, the Quadratic Loss and the Spherical Payoff, as described in section 5.1. The results of all scoring rules for each tested node and all networks are shown in Table 5-2.[13]

Table 5-2 Scoring rule results for all tested nodes and BNs. The numbers in the top row indicate the definition of the peak of the storm duration (10%, 15%, etc.).

Coastal buffer	10%	15%	20%	25%	30%
Log loss	1.073	1.046	1.046	0.9987	0.9987
Quad loss (brier skill)	0.4631	0.453	0.4529	0.4382	0.4382

[13] As a reminder: Log loss is scored from 0 to infinity where 0 is best, quadratic loss is scored from 0 to 2 where 0 is best and spherical payoff is scored from 0 to 1 where 1 is best.

Spherical payoff	0.7425	0.7492	0.7478	0.7649	0.7649
Overwash depth					
Log loss	1.347	1.351	1.362	1.36	1.36
Quad loss (brier skill)	0.656	0.6647	0.6738	0.6735	0.6735
Spherical payoff	0.5776	0.5761	0.5708	0.5678	0.5678
Overwash velocity					
Log loss	1.327	1.348	1.346	1.345	1.345
Quad loss (brier skill)	0.6154	0.6293	0.6335	0.6332	0.6332
Spherical payoff	0.6093	0.6005	0.6014	0.5987	0.5987

The overall results are a spherical payoff score above 0.5, a quadratic loss lower than 1 and a Log loss between 1 and 1.35. These results indicate that the performance is definitely not flawless and predictions have to be looked at with care. The magnitude of the errors of the prediction can be observed by looking at the real test values and the results from the confusion matrix (Table 5-3).

First observed are the absolute errors. The mode of the coastal buffer has a mean absolute error and root mean square error of around 20 meters, indicating that the mean error of a prediction is plus or minus 20 meters off. With the bin size being 10 meters this basically comes down to the prediction being off by 2 bins on average. For the overwash depth this value is about 0.4 to 0.5 m, which is the same size of one bin (0.5 m). This is the same for the overwash velocity where the MAE and the RMSE are in the order of 0.8 to 1 m and the bin size is 0.9 m/s.

Table 5-3 Real value test and confusion matrix results for the tested node and for each BN. The numbers in the top row indicate the definition of the peak of the storm duration (10%, 15%, etc.).

Coastal buffer	10%	15%	20%	25%	30%
MAE (m)	18.96	18.76	18.87	18.74	18.74
Standard deviation (m)	27.86	27.54	27.62	27.52	27.52
Std. dev. / MAE	1.47	1.47	1.46	1.47	1.47
RMSE (m)	20.76	20.62	20.85	20.95	20.95
Error rate	30%	26.67%	30%	25.56%	25.56%
Overwash depth					
MAE (m)	0.4312	0.4144	0.4131	0.3953	0.3953
Standard deviation (m)	1.206	1.172	1.174	1.151	1.151
Std. dev. / MAE	2.80	2.83	2.84	2.91	2.91
RMSE (m)	0.5356	0.5132	0.5198	0.503	0.503
Error rate	62.22%	57.78%	55.56%	55.56%	55.56%
Overwash velocity					
MAE (m/s)	0.8153	0.8203	0.8284	0.8903	0.8903
Standard deviation (m)	1.664	1.642	1.651	1.795	1.795
Std. dev. / MAE	2.04	2.00	1.99	2.02	2.02
RMSE (m/s)	1.001	1.006	1.02	1.007	1.007
Error rate	48.89%	48.89%	48.89%	48.89%	48.89%

Assuming that the errors in the predictions are normally distributed, another indication of the error is its standard deviation. The standard deviation is a measure of the spread that can be expected of

the error in a prediction. This spread is often described with a confidence interval which is related to the standard deviation by the 68% – 95% – 99.7% confidence rule for respectively one – two – three standard deviations. Basically, 68% of the outcomes will lie within one standard deviation of the mean and likewise. The confidence intervals obtained are given in Table 5-4. The confidence intervals for all nodes are large when considering the use of the BN.

Table 5-4 Confidence intervals for the mode of the prediction of each node.

Node \ confidence interval	68%	95%	99.7%
Coastal buffer (m)	27.5	55	82.5
Overwash depth (m)	1.2	2.4	3.6
Overwash velocity (m/s)	1.7	3.4	5.1

5.2.2 PERFORMANCE OF DIFFERENT NETWORKS

When considering the performance of the different tested networks most changes are incremental; several trends are however observed. The first is that the scoring rule results (Table 5-3) of the coastal buffer node changes for the better as the duration of the peak of the storm increases from 10% to 30%, while they change for the worse for the overwash depth and velocity. This can be explained by the following the combination between the water level and the significant wave height.

By increasing the length of the peak of the storm the weight of the water level as compared to the wave height is increased in determining the coastal hazards. This is because when a length of the peak of the storm is chosen, it also chooses a combination of wave height and water level that defines the coastal hazards. The highest wave height however is always included in this definition but the highest occurring water level may not be. By increasing the duration of the peak of the storm a higher or at least as high water level is obtained. The new combination now defines the coastal hazard, in which the water level is now relatively more important. The predictions for the coastal erosion become better as the highest water level is included, indicating that this is a more important factor than the maximum wave height. Reversely so the quality of the prediction for the overwash decreases as the peak of the storm duration is increased. This is because for the overwash the combination of the two variables is more important; the water level and wave height that occur at the same time govern the overwash, a time lag distorts the prediction.

Another observation is the trend in the error rate in combination with the relative standard deviation (Std. dev. / MAE, Table 5-3). For the coastal buffer the error rate is decreasing and the relative standard deviation stays the same, indicating an increase in the performance, which is in line with the better scoring rule results. The overwash depth also shows a decrease in the error rate but at the same time has an increase in the relative standard deviation; the state with the highest belief is more often correct than before but the spread in the error is larger. There is no observed change for the overwash velocity, but there is an increase in the relative standard deviation, which is also in line with the trend in the results of the scoring rules.

A possible explanation for the relatively better performance of the coastal buffer node is that most of the results of the XBeach modeling indicate a coastal buffer in the -60 to -65 bin. The results of the overwash have a much larger spread and are therefore predicted with less belief; it is harder to

predict the correct bin within the node. This makes it relatively easy to perform better for the coastal buffer node.

5.2.3 PERFOMANCE OVER THE RANGE OF PREDICTIONS

A more in depth performance of the individual nodes is obtained by looking at their respective calibration tables. The calibration table, as explained in the beginning of this chapter, gives the belief of a certain state (based on the 90% trained data) and the percentage of times the actual outcome was in that state (based on the 10% tested data). Insight is gained over which parts of the node are better calibrated than others, as well as insight into over- or underestimation. Two tables are considered, Table 5-5 and Table 5-6, respectively: the overall calibration table that gives the total calibration of a node and the calibration tables that give the calibration for each bin within a node. For the latter the probability ranges change for the different states. This is done by Netica to ensure a certain minimum degree of accuracy; the bin sizes adapt to have a minimum number of cases falling in them.

Table 5-5 Overall calibration of the three coastal hazard nodes. The top row gives the belief as a percentage and the values in the table indicate the actual occurrence as a percentage. Green: well calibrated. Yellow: mediocre calibrated. Red: Poor calibrated.

Node \ Belief (%)	0-5:	5-10:	10-15:	15-20:	20-30:	30-40:	40-60:	60-100:
Coastal buffer	0.19	3.13	7.5	35	28.6	59.3	82.6	100
Overwash depth	0.35	2.27	15.5	35.1	31.1	45.5	33.3	61.1
Overwash velocity	0.415	2.96	15.9	18.2	27.5	36.4	63.3	77.8

The overall calibration table (Table 5-5) has the ranges of predicted values in the top row and the actual occurred percentages for those predictions filling the table. The color scales indicate the calibration, green if the actual percentage falls within the predicted ranges, yellow if it falls just outside the predicted bounds and red if the actual occurrence differs greatly from the predictions. Several observations are made from this table. The coastal buffer is the least well calibrated and has the tendency to underestimate cases (the actual occurrence percentage is higher than the belief). The overwash depth is badly calibrated in the mid-range of the belief and well calibrated if the belief reaches the extremes. Lastly the overwash velocity is overall best calibrated, increasingly so as the belief increases.

Table 5-6 Calibration tables for the nodes coastal buffer, overwash depth and overwash velocity. The belief of a bin is compared to the actual percentage of the occurrence. Green: well calibrated. Yellow: mediocre calibrated. Red: Poor calibrated.

Coastal buffer	Belief (%)	Actual (%)	Belief (%)	Actual (%)	Belief (%)	Actual (%)	Belief (%)	Actual (%)
-65 to -60	0-10:	0	10-40:	54.2	40-60:	82.6	60-80:	100
-60 to -50	0-5:	1.67	5-10:	0				
-50 to -40	0-5:	0	5-10:	6.67	10-100:	20		
-40 to -30	0-5:	0	5-10:	0	10-100:	0		
-30 to -20	0-5:	0	5-15:	0				
-20 to -10	0-5:	0	5-10:	5.88	10-100:	0		
-10 to 0	0-5:	0	5-15:	0	15-30:	40	30-100:	28.6
0 to 10	0-5:	0	5-30:	25	30-100:	66.7		
10 to 20	0-5:	0	5-30:	9.52	30-100:	50		

20 to 45	0-5:	0	5-100:	13.3				
Overwash depth								
0 to 0.1	0-5:	0	5-15:	0	15-50:	41.7	50-100:	71.4
0.1 to 0.5	0-5:	0	5-10:	11.1	10-20:	25	20-100:	35.7
0.5 to 1	0-5:	0	5-10:	0	10-30:	34.8	30-50:	26.3
1 to 1.5	0-10:	2.78	10-20:	16	20-40:	37.9		
1.5 to 2	0-5:	3.85	5-15:	4	15-40:	44.4	40-100:	50
2 to 2.5	0-5:	0	5-10:	0	10-50:	30.8		
2.5 to 3	0-5:	0	5-10:	0	10-100:	45.5		
3 to 8	0-5:	0	5-10:	0	10-100:	0		
Overwash velocity								
0 to 0.1	0-5:	0	5-15:	0	15-50:	41.7	50-100:	71.4
0.1 to 0.5	0-5:	0	5-10:	4.17	10-100:	28.6		
0.5 to 1.4	0-5:	0	5-10:	5	10-30:	5.26		
1.4 to 2.3	0-5:	2.94	5-10:	0	10-15:	25	15-100:	33.3
2.3 to 3.2	0-5:	0	5-10:	0	10-30:	20	30-100:	0
3.2 to 4.1	0-10:	7.41	10-20:	22.5	20-40:	47.6	40-100:	100
4.1 to 5	0-10:	0	10-20:	4.55	20-50:	46.7	50-100:	85.7
5 to 7	0-5:	0	5-10:	0	10-40:	26.3	40-100:	75

A number of observations can be made from the calibration tables of the individual nodes (Table 5-6). The coastal buffer node holds most of its information in the states -65 to -60 and in the range from 0 to 45. The bad calibration for the states in between seems understandable since there is hardly any data to calibrate this part. The underestimation (actual occurrence is higher than the predictions) seen in the overall calibration table for this node can be attributed to a single bin: -65 to -60. For the 'dry' part of the node (for storms that do not erode the coastline to the first buildings) from 0 to 45 it is actually well calibrated. The overwash depth performs well as soon as the belief for a state is higher than 10-20%, for any lower beliefs the calibration is generally poor. The overwash depth and velocity perform well in predicting that a storm has no overwash, but are not very well calibrated when predicting the magnitude of an overwash event.

5.3 VALUE OF A PREDICTION

One of the main questions that arise from the prediction of the BN is: how should a prediction be interpreted and valued? There is no single answer to this question. This section evaluates a possibility of how this can be done. Two predictions are examined and their results analyzed. For this the BN trained with all data will be considered with the definition of the peak of the storm of 25% of the duration.

Two predictions for the coastal buffer node are shown in Figure 5-1, for two different conditioning sets. The prediction on the left shows a bimodal distribution (two peaks) where the one on the left has one very pronounced peak.

The mean absolute error (MAE) and its standard deviation, calculated in the previous section, give the overall error for the node. These values say that on mean average the error of the prediction of the **most likely** state is 18.7 m with a standard deviation of 27.5 m, for all the tested cases. This is a rather large error and an even larger standard deviation, giving very large confidence intervals. The

magnitude of the MAE can be explained when considering the bimodal prediction on the left in Figure 5-1. If the actual value is not in the most likely state (-65 to -60 m) then it is probably in in the -10 to 0 state, giving an error of about 60 meters in the prediction. The most likely state in the right panel has a belief 19% higher than that of the left panel and shows a distribution with only one mode. It seems unreasonable to use the same MAE for both predictions, in fact the errors in the predictions change for every combination of input variables. For these cases the MAE is likely to underestimate the error in the left panel and overestimate the error in the right panel.

Physically the two modes for the coastal buffer node indicate that either the dune is breached and all of the hinterland is affected, or that the barrier is not breached and that only part of the dune is eroded.

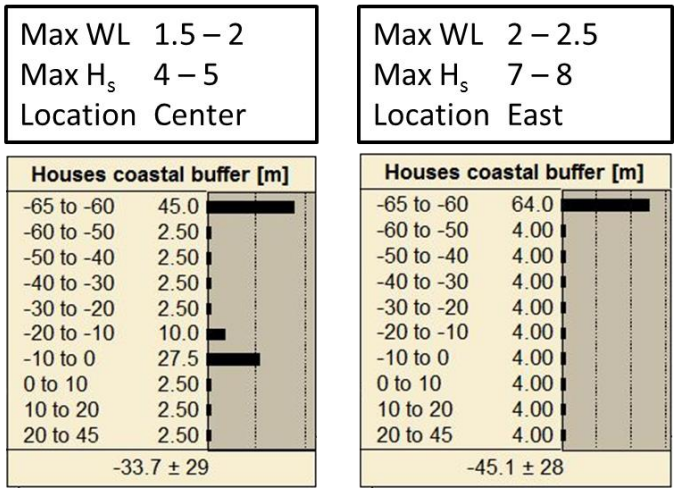

Figure 5-1 Prediction for the coastal buffer node for two different sets of conditioning on the water level, wave height and location.

Another way to look at the predictions of the BN is by considering the distributions, in this case the probability mass functions. The probability mass function itself can also be used to determine exceedance probabilities. The exceedance probability is the chance that a value is beyond a certain threshold value. The probability is by calculating the area underneath the curve of the PDF, beyond the threshold, as indicated in Figure 5-2. For the probability mass functions that the BN gives these values can also be calculated. As an example the 10% exceedance value for the coastal buffer node is calculated by linear interpolation for the left prediction in Figure 5-2:

$$x = Bin\ edge + \frac{Bin\ size}{Bin\ belief}(1 - confidence) = -65 + \frac{-60 - -65}{0.45}(1 - 0.9) = -63.9\,m \qquad (8)$$

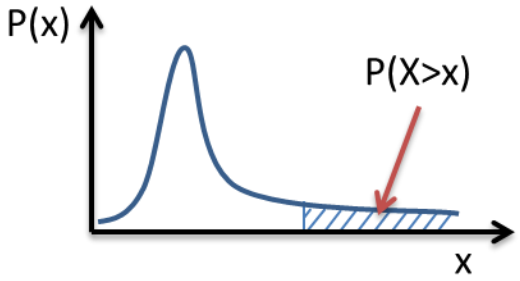

Figure 5-2 Schematic probability density distribution with exceedance probability.

In Figure 5-3 the predictions for the overwash nodes are shown. The MAE's for these nodes are, 0.4 m for the overwash depth and 0.9 m/s for the overwash velocity, with standard deviations of 1.15 m and 1.8 m/s (Table 5-3). Since the distributions obtained for these predictions generally have a single mode, this number is much more representative of the error that can be expected for any given prediction. Also for these nodes, however, exceedance probabilities can be determined for each individual prediction.

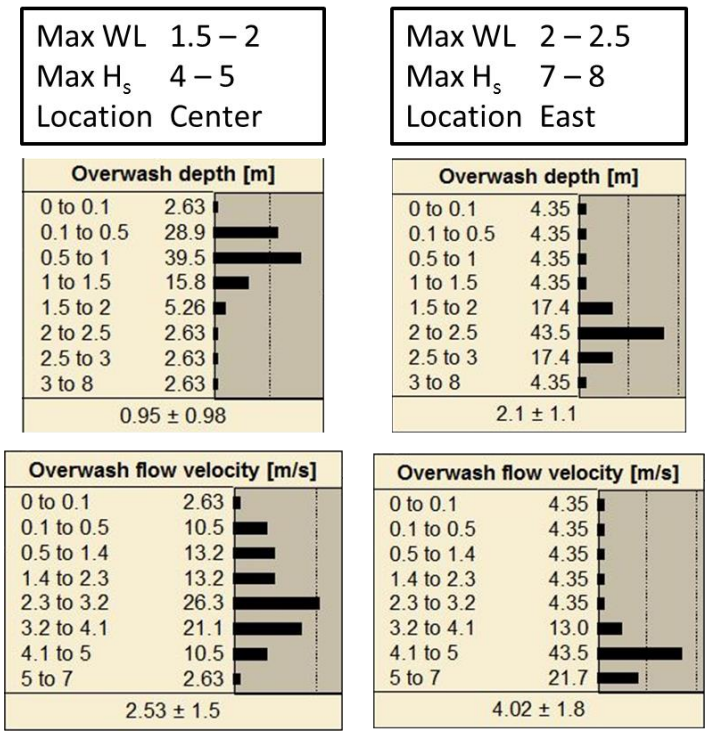

Figure 5-3 Prediction for the overwash nodes for two different sets of conditioning on the water level, wave height and location.

6 CONCLUSIONS

This study investigated the development of a surrogate for the coastal morphodynamic part of an Early Warning System (EWS) for marine coastal hazards using a Bayesian Network (BN). For this Praia de Faro, located in the south of Portugal, has been selected as a case study site for the application of the concept. The aim of the BN was to predict onshore hazard intensities given offshore storm boundary conditions.

To develop the modeling concept several steps have been taken. The case study site has been thoroughly analyzed with a focus on the coastal response to high energy, low frequency events. Based on this analysis a site has been chosen for the implementation of the EWS. The implementation on the chosen site consisted of five steps: (1) creating a synthetic storm dataset, (2) setting up an XBeach model for the selected site, (3) setting up a BN, (4) training the BN and (5) evaluating the BN.

The main research question was: *How can Bayesian Networks be used as part of an Early Warning System for spatially varying coastal hazards at sandy coastlines?* This question has been delineated by setting a main objective, *to quantify spatially varying coastal hazard intensities resulting from different forcing scenarios (e.g. storms) using a Bayesian Network*, and five sub objectives. The conclusions drawn during the development of the EWS will be given with respect to the set of five sub objectives.

1. **Development of a Bayesian Network that can act as a surrogate for a 2D XBeach model.**

A 2D XBeach model is capable of translating offshore storm boundary conditions to near shore hydrodynamics and morphological change; it gives insight in the response of the coastline to a storm. Since it works in two dimensions the spatial variability of coastal hazards can be extracted. The BN therefore has to translate offshore boundary conditions to onshore hazard indicators, including spatial variability. This had to be achieved with the knowledge that higher complexity results in a higher demand for training data and has led to a BN with as little complexity as possible. The complexity has been reduced by using the physical correlations between the significant wave height, peak period, storm duration and storm surge levels. The BN only uses the offshore maximum significant wave height and water level that occur during a storm to predict the coastline erosion and overwash intensities. The spatial variability has been included by identifying three areas based on their similar elevation and structures. The resulting BN, with low complexity, makes it easier to train the network as less data is necessary, coming at the cost of losing a lot of information produced with the XBeach model. For this reason only storms with a similar course can be included; a storm with two peaks would currently not fit the model.

2. **Development of an XBeach model that generates suitable data for the Bayesian Network.**

The XBeach model for Praia de Faro has been created using bathymetric data from the summer of 2007. The grid of the model has been setup following the basic rules as laid out in the XBeach manual. The waves in XBeach have been modeled using a wave spectrum. Since no spectral analysis of the waves is available for the South of Portugal a JONSWAP spectrum has been assumed. This has two consequences: (1) Mixed sea states, consisting of both swell and wind waves, cannot be properly represented. (2) It is unsure if this type of spectrum describes a storm well for this region

since it assumes that the sea state is not yet fully in equilibrium with the wind. For the short storm durations this assumption may be valid, however, as the storm duration increases this may become invalid, especially considering the large possible fetch over the Atlantic Ocean. For the results of the modeling this means that a spectrum containing both swell and waves will cause overwash much sooner than a spectrum containing only wind waves.

The XBeach model could not be calibrated well due to the limited dataset that was available for it. Measured profiles were only available for a very small section of the grid and only above mean sea level. For this reason the model could not be run on the initial profiles nor calibrated for the wet section of the profiles. Furthermore the response of the coastline to the storms that have been used for the calibration was limited; no real overwash or major dune erosion is observed. Many of the storms run for the BN are much larger causing large overwashes and dune erosion in the XBeach model. The effects of these storms seem exaggerated and are passes on to the BN which predicts very large coastal erosion and overwash for the larger storms.

The combination of the assumption for the spectrum and the lack of field data to calibrate XBeach for extreme conditions means that the results are less trustworthy for these events.

3. **Investigate an efficient method for generating sufficient data for the Bayesian Network.**

One of the main issues of this research is that the data needed to train the BN is produced with XBeach. A BN needs a lot of data to be trained well and XBeach has high computational times. On top of this overwash had to be extracted from the XBeach simulations, posing a difficulty since overwash events occur in a very short period of time. To record overwash, water levels and flow velocities have to be recorded at very high frequencies, creating large data files that quickly fill up storage space. To reduce the runtime of XBeach two options have been used: (1) the internal morphological acceleration option and (2) a reduction in the number of cross shore profiles in the grid. It was found that the detailed grid could be reduced to a fourth of the total number of grid cells. A further reduction caused numerical instabilities within the XBeach model causing simulations to crash. The results of the coarser grid were very similar to the results obtained with the original grid; a comparison of the two with the Brier skill score (scored from 0 to 1, where 1 is best) gave a score of 0.97. An added advantage of the reduced grid size is that less output is generated and so reducing the demand for storage space. Overwash has been measured at only 19 locations, set on top of the dune crest, at which output has been extracted every 2 seconds. This sufficiently reduced the generated data while still obtaining the necessary output to observe overwash events.

4. **Apply the Bayesian Network to the case study site of Praia de Faro on the Portuguese coastline.**

A part of applying the modeling concept to Praia de Faro is covered by setting up an XBeach model for the area. However, the XBeach model needs to be run with a set of storms to train the BN. The local storm dataset that was available did not contain enough storms and did not cover the range of storms necessary to train the BN. This is a universal problem as storm datasets are often limited to a few decades and the EWS has to be trained for storm with a return period of at least once in 100 years. Furthermore these datasets will never contain the necessary amount of storms needed to train a BN. This problem has been solved by creating a synthetic dataset, based on the existing local storm dataset. The variables of the local dataset have been used to create copulas that describe

their inter-relations. With these copulas new data points can be sampled that bear the same characteristics as the parent dataset.

The advantage of using this method over using, for example linear regression lines, is that the natural variability is incorporated in the synthetic dataset and thus in the BN. A limitation is that a large variability will make it harder to train a BN as the predictions will also show this variability; ultimately leading to a higher demand for training data (so more storms need to be run). In other words including the variability means the model is a closer representation of reality, however making it hard to model.

5. **Evaluate the performance of the Bayesian Network with respect to its use as an EWS.**

The evaluation of the performance of the BN has been focused on the method of testing and determining the performance of a BN in a general sense. A comparison between the predictions of the BN and expected outcomes for storms at Praia de Faro, based on literature, has been excluded. The reason for this is that the XBeach model results are not realistic enough to make a good comparison and justify any conclusion based on it.

Five different BNs have been created based on the extraction of the maximum occurring water level and wave height from the peak of the storm from XBeach to be used as input for the BN. The peak of the storm has been varied between 10% and 30% of the total duration of the storm. The BNs have been evaluated by training them with 90% of the data and testing them with the remaining 10%. A number of test results have been considered that give insight into the performance of the BNs on different levels. For this each hazard node has been considered individually.

Based on the skill score tests (log loss, quadratic loss and spherical payoff), the overall performance of the different networks is very similar and can be described as decent but definitely not flawless. However, without a benchmark these numbers are hard to interpret and the absolute errors give more insight. The errors in the predictions of the three hazard nodes, coastal buffer, overwash depth and overwash velocity, have a MAE (mean absolute error) of respectively 19 m, 0.4 m and 0.9 m/s, with standard deviations of, respectively, 27.5 m, 1.2m and 1.7 m/s. The 90% confidence intervals are therefore 55 m, 2.4 m and 3.4 m/s. With respect to predicting coastal erosion and overwash these errors and confidence intervals are very large, and could be considered too large for its use as an EWS. However, the results of the calibration table have shown that the overwash nodes are not very accurate in predicting the magnitude of an overwash but are good at predicting whether or not overwash will occur. These errors are caused by two effects: (1) for the coastal buffer the predictions often consists of a distribution with two modes, causing very large mean errors and accompanying standard deviations and (2) the large spread that is seen in the synthetic storm dataset is translated to the BN, giving a large spread in the predictions. In conclusion, the BN either needs more training data to make better predictions or more complexity (more nodes and smaller bins) needs to be added, also implying a need for more training data. Furthermore, when the value of an individual prediction is considered it is better to use the given probability mass function than the calculated MAE. The MAE can over or under predict the error for an individual prediction and should only be used as a general performance indicator. A note also has to be made on the way the BN is tested; by training it with 90% and testing it with the remaining 10%. If the remaining 10% is an ‘unlucky’ sample in the sense that it does not reflect the other 90% well, it will negatively influence the performance results.

Lastly, the test results for the different BNs indicate that as the definition of the duration of the peak of the storm is increased that the prediction for coastal erosion is improved and of the overwash intensity is worsened. By increasing the duration of the peak of the storm, the possible time lag between the combination of the highest water level and wave height is increased. This indicates that for the coastal erosion node the occurrence of the highest water level is more important than its joint occurrence with the highest wave height for the total amount of erosion. However, for the overwash intensity the time lag between the two distorts the prediction as this process is governed by the joint occurrence of the wave height and water level and so the predictions worsen as the definition of the peak of the storm increases.

7 RECOMMENDATIONS

There have been many degrees of freedom within the development of the EWS. These degrees of freedom have meant that a number of decision and assumptions have had to be made to develop and implement the modeling concept. As this is a pilot study there are many aspects that can be improved and optimized of which 6 major improvements are recommended here.

1. The large spread that is observed in the predictions is the direct result of the large spread that is seen in the synthetic dataset. The spread in the synthetic dataset cannot be reduced as this would give a false representation of reality. Two recommendations are therefore including more detail in the BN, by adding an extra boundary condition node or decrease the size of the bins, and simulating more storms. Suggested is that the best node to add for this case study site is the peak period. The reasons being that the surge is already included in the water level node and the duration shows a much higher correlation to the significant wave height than the peak period for the local storm dataset (Figure C-6).
2. In the setting up of the BN the structure has been relatively straight forward. However, the discretization of the individual nodes is not. Currently a heuristic method is used that considers the prior distributions. It would be better to discretize the network such that the interdependencies between variables are maximized. For this algorithms can be used such as the CAIM (class-attribute interdependence maximization) algorithm (Kurgan & Cios, 2004).
3. Currently the BN uses the water level, wave height and location to make a prediction for the onshore hazard intensities. Another option would be to consider the highest expected runup per location based on a runup formula. The expected runup for each individual storm can be calculated using the time series of the wave height and water level. The advantage is that always the simultaneously occurring water level and wave height are considered. Disadvantages are that the BN becomes less transparent for the end user, since the water level and wave height nodes would be grouped into a runup node, and that prior to a prediction a calculation over the full time series of a storm needs to be performed.
4. A major step toward attaining an operational EWS is making sure that the results of the XBeach model can be trusted. For this the local model needs to be well calibrated and the modeling of a storm needs to be correct. In the modeling of storms two important assumptions have been made that need to be verified. These are the usage of a JONSWAP spectrum for the wave input and the triangular shape of the course of a storm.
5. The modeling concept in its current form has been developed based on the conditions at Praia de Faro. For validation purposes the same concept should be applied to another case study site with varying conditions and hazards to gain more insight in the general applicability.
6. Ultimately the prediction of hazards in the BN should be coupled so that expected damages and possibly loss of life are also predicted. This could be incorporated with risk reduction measures such as evacuation, removal of houses or the placement of sand bags. To accomplish this, hazards need to be predicted at a higher level of detail. The course grid of XBeach does not allow extraction of data at this level of detail; however this could be obtained by, for example, interpolation of the results. For this the data extraction from XBeach needs to be further researched.

8 REFERENCES

Almeida, L.P., Ferreira, Ó., Vousdoukas, M.I. & Dodet, G. (2011a). Historical variation and trends in storminess along the Portuguese South Coast. *Natural Hazards and Earth System Science*. 11. p.pp. 2407–2417.

Almeida, L.P., Vousdoukas, M.V., Ferreira, Ó., Roderigues, B.A. & Matias, A. (2011b). Thresholds for morphological changes on an exposed sandy beach as a function of wave height. *Earth Surface Processes and Landforms*. [Online]. 36 (4). p.pp. 523–532. Available from: http://doi.wiley.com/10.1002/esp.2072. [Accessed: 22 December 2014].

Almeida, L.P., Vousdoukas, M.V., Ferreira, Ó., Rodrigues, B. a. & Matias, A. (2012). Thresholds for storm impacts on an exposed sandy coastal area in southern Portugal. *Geomorphology*. [Online]. 143-144. p.pp. 3–12. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0169555X11003412. [Accessed: 22 December 2014].

Carrasco, A.R., Ferreira, Ó., Matias, A. & Freire, P. (2012). Flood hazard assessment and management of fetch-limited coastal environments. *Ocean & Coastal Management*. [Online]. 65. p.pp. 15–25. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0964569112000865. [Accessed: 23 March 2015].

Corbella, S. & Stretch, D.D. (2013). Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering*. [Online]. 76. p.pp. 68–78. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378383913000240. [Accessed: 21 July 2015].

Costa, M., Silva, R. & Vitorino, J. (2001). CONTRIBUIÇÃO PARA O ESTUDO DO CLIMA DE AGITAÇÃO MARÍTIMA NA COSTA PORTUGUESA. In: *2as Jornadas Portuguesas de Engenharia Costeira e Portuária - Associação Internacional de Navegação*. 2001, p. 20.

Deltares (2015). *XBeach*. [Online]. Available from: http://oss.deltares.nl/web/xbeach/.

Demarta, S. & McNeil, A.J. (2005). The t Copula and Related Copulas. *International Statistical Review*. [Online]. 73 (1). p.pp. 111–129. Available from: http://doi.wiley.com/10.1111/j.1751-5823.2005.tb00254.x.

Van Dongeren, A., Ciavola, P., Viavattene, C., De Kleermaeker, S., Martinez, G., Ferreira, O., Costa, C. & McCall, R. (2014). RISC-KIT: Resilience-Increasing Strategies for Coasts - toolKIT. *Journal of Coastal Research*. [Online]. p.pp. 366–371. Available from: <Go to ISI>://WOS:000338176100063.

EU (2007). Directive 2007/60/EC of the European Parliament of the council of 23 October 2007 on the assessment and management of flood risks. *Official Journal of the European Union*. L (288). p.pp. 27–34.

Ferreira, Ó. (2006). The role of storm groups in the erosion of sandy coasts. *Earth Surface Processes and Landforms*. [Online]. 31 (8). p.pp. 1058–1060. Available from: http://doi.wiley.com/10.1002/esp.1378.

Genest, C. & Favre, A. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of hydrologic engineering*. [Online]. (July/August). p.pp. 347–368. Available from: http://ascelibrary.org/doi/abs/10.1061/(ASCE)1084-0699(2007)12:4(347). [Accessed: 21 July 2015].

GTL (2014). *Gestão da Zona Costeira O Desafio da Mudança*.

Guedes Soares, C. (1984). Representation of double-peaked sea wave spectra. *Ocean Engineering*. [Online]. 11 (2). p.pp. 185–207. Available from: http://linkinghub.elsevier.com/retrieve/pii/0029801884900192.

Gutierrez, B.T., Plant, N.G. & Thieler, E.R. (2011). A Bayesian network to predict coastal vulnerability to sea level rise. *Journal of Geophysical Research*. [Online]. 116 (F2). p.p. F02009. Available from: http://doi.wiley.com/10.1029/2010JF001891. [Accessed: 10 September 2015].

Hapke, C. & Plant, N. (2010). Predicting coastal cliff erosion using a Bayesian probabilistic model. *Marine Geology*. [Online]. 278 (1-4). p.pp. 140–149. Available from: http://dx.doi.org/10.1016/j.margeo.2010.10.001.

Den Heijer, C., Knipping, D.T.J.A., Plant, N.G., Van Thiel de Vries, J.S.M., Baart, F. & Van Gelder, P.H.A.J.M. (2012). Impact assessment of extreme storm events using a Bayesian network. In: *Proceedings of the Coastal Engineering Conference*. [Online]. 2012. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84884938068&partnerID=tZOtx3y1.

Holthuijsen, L. (2007). *Waves in oceanic and coastal waters*. [Online]. New York: Cambridge University Press. Available from: http://books.google.com/books?hl=en&lr=&id=7tFUL2blHdoC&oi=fnd&pg=PA1&dq=Waves+in+Oceanic+and+Coastal+Waters&ots=g3nDYhnhpC&sig=HbMpepYc1ngJX_bw__h0qU6G9Yc. [Accessed: 24 March 2015].

Jäger, W. (2013). *Using Dynamic Nonparametric Bayesian Belief Nets (BBNs) to Model Human Influences on Safety*. [Online]. Delft University of Technology. Available from: http://repository.tudelft.nl/assets/uuid:ad111d2e-6049-4aef-a3ce-6c4b3c715ae2/final_Thesis_EPA_Wiebke_Jager.pdf. [Accessed: 15 December 2014].

Jäger, W.S. & Morales, O.N. (2014). *Sampling Joint Time Series of Significant Wave Heights and Periods in the North Sea*.

Kahn, C.E.J., Roberts, L.M., Shaffer, K.A. & Haddawy, P. (1997). Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine*. 27 (1). p.pp. 19–29.

Kurgan, L.A. & Cios, K.J. (2004). CAIM Discretization Algorithm. *IEEE Transactions on Knowledge and Data Engineering*. 16 (2). p.pp. 145–153.

Masselink, G. & Hughes, M.G. (2003). *Introduction to Coastal Processes & Geomorphology*. 1st Ed. London: Hodder Education.

McCall, R.T., Van Thiel de Vries, J.S.M., Plant, N.G., Van Dongeren, A.R., Roelvink, J.A., Thompson, D.M. & Reniers, A.J.H.M. (2010). Two-dimensional time dependent hurricane overwash and erosion modeling at Santa Rosa Island. *Coastal Engineering*. [Online]. 57 (7). p.pp. 668–683.

Available from: http://linkinghub.elsevier.com/retrieve/pii/S037838391000027X. [Accessed: 15 December 2014].

Norsys (2013). *Netica*. [Online]. Available from: www.norsys.com.

Norsys (2015). *WebHelp*. [Online]. 2015. Available from: https://www.norsys.com/WebHelp/NETICA.htm. [Accessed: 18 August 2015].

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Fransisco, CA: Morgan Kaufmann Publishers Inc.

Pilkey, O.H.J., Neal, W.J., Monteiro, J.H. & Dias, J.M.A. (1989). Algarve Barrier Islands: A Noncoastal-Plain System in Portugal. *Journal of Coastal Research*. 5 (2). p.pp. 239–261.

Pires, H.O. (1998). *Project INDIA. Preliminary report on wave climate at Faro*. [Online]. Lisboa, Portugal. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Project+INDIA.+Preliminary+report+on+wave+climate+at+Faro#0. [Accessed: 18 March 2015].

Pires Silva, A.A. & Sarmento, A.J.N.A. (1991). Double-peaked wave spectra modelling: a case study in the Portuguese coast. In: *The First International Offshore and Polar Engineering Conference (ISOPE 91)*. 1991, Edinburgh, pp. 76–82.

Plomaritis, T. a., Benavente, J., Laiz, I. & Del Río, L. (2015). Variability in storm climate along the Gulf of Cadiz: the role of large scale atmospheric forcing and implications to coastal hazards. *Climate Dynamics*. [Online]. Available from: http://link.springer.com/10.1007/s00382-015-2486-4. [Accessed: 18 March 2015].

Van Rijn, L.., Walstra, D.J.., Grasmeijer, B., Sutherland, J., Pan, S. & Sierra, J.. (2003). The predictability of cross-shore bed evolution of sandy beaches at the time scale of storms and seasons using process-based Profile models. *Coastal Engineering*. [Online]. 47 (3). p.pp. 295–327. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378383902001205.

Rodrigues, B.A., Matias, A. & Ferreira, Ó. (2012). Overwash hazard assessment. *Gelologica Acta*. 10 (4). p.pp. 427–437.

Roelvink, D., Reiniers, A., Van Dongeren, A., Van Thiel de Vries, J., Lescinski, J. & McCall, R. (2010). *XBeach Model Description and Manual*.

Roelvink, D., Reniers, A., van Dongeren, A., van Thiel de Vries, J., McCall, R. & Lescinski, J. (2009). Modelling storm impacts on beaches, dunes and barrier islands. *Coastal Engineering*. [Online]. 56 (11-12). p.pp. 1133–1152. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378383909001252. [Accessed: 14 July 2014].

Sallenger, A.H.J. (2000). Storm impact scale for barrier islands. *Journal of Coastal Research*. [Online]. 16 (3). p.pp. 890–895. Available from: http://www.jstor.org/stable/4300099. [Accessed: 17 March 2015].

Schmidt, T. (2007). Coping with Copulas. In: J. Rank (ed.). *Copulas: From theory to application in Finance*. London: Risk Books, pp. 3–34.

Sun, S., Zhang, C. & Yu, G. (2006). A Baysian Network Approach to Traffic Flow Forecasting. *IEEE Transactions on Intelligent Tra*. 7 (1). p.pp. 124–132.

UNISDR (2002). Guidelines for Reducing Flood Losses. *United Nations*. [Online]. Available from: http://www.unisdr.org/files/558_7639.pd.

Van Verseveld, H.C.W. (2014). *Impact Modelling of Hurricane Sandy on the Rockaways*. Delft University of Technology.

Van Verseveld, H.C.W., van Dongeren, a. R., Plant, N.G., Jäger, W.S. & den Heijer, C. (2015). Modelling multi-hazard hurricane damages on an urbanized coast with a Bayesian Network approach. *Coastal Engineering*. [Online]. 103. p.pp. 1–14. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378383915000927. [Accessed: 7 June 2015].

Vila-Concejo, A., Matias, A., Ferreira, Ó., Duarte, C. & Dias, J.M.A. (2002). Recent evolution of the natural inlets of a barrier island system in Southern Portugal. *Journal of Coastal Research*. [Online]. (SI 36). p.pp. 741–752. Available from: http://www.academia.edu/download/30945525/Recent_Evolution_of_the_Natural_Inlets_of_a_Barrier_Island_System_in_Southern_Portugal_Vila-Concejo_Matias_Ferreira_Duarte_Dias_pp._741-.pdf. [Accessed: 17 March 2015].

Vousdoukas, M.I., Almeida, L.P.M. & Ferreira, Ó. (2012a). Beach erosion and recovery during consecutive storms at a steep-sloping, meso-tidal beach. *Earth Surface Processes and Landforms*. [Online]. 37 (6). p.pp. 583–593. Available from: http://doi.wiley.com/10.1002/esp.2264. [Accessed: 22 December 2014].

Vousdoukas, M.I., Ferreira, Ó., Almeida, L.P. & Pacheco, A. (2012b). Toward reliable storm-hazard forecasts: XBeach calibration and its potential application in an operational early-warning system. *Ocean Dynamics*. [Online]. 62 (7). p.pp. 1001–1015. Available from: http://link.springer.com/10.1007/s10236-012-0544-6. [Accessed: 17 December 2014].

Vousdoukas, M.I., Wziatek, D. & Almeida, L.P. (2011). Coastal vulnerability assessment based on video wave run-up observations at a mesotidal, steep-sloped beach. *Ocean Dynamics*. [Online]. 62 (1). p.pp. 123–137. Available from: http://link.springer.com/10.1007/s10236-011-0480-x. [Accessed: 22 December 2014].

Wright, L.D. & Short, A.D. (1984). Morphodynamic variability of surf zones and beaches: a synthesis. *Marine geology*. [Online]. 56. p.pp. 93–118. Available from: http://www.sciencedirect.com/science/article/pii/0025322784900082. [Accessed: 17 March 2015].

APPENDICES

A COASTAL VULNERABILITY

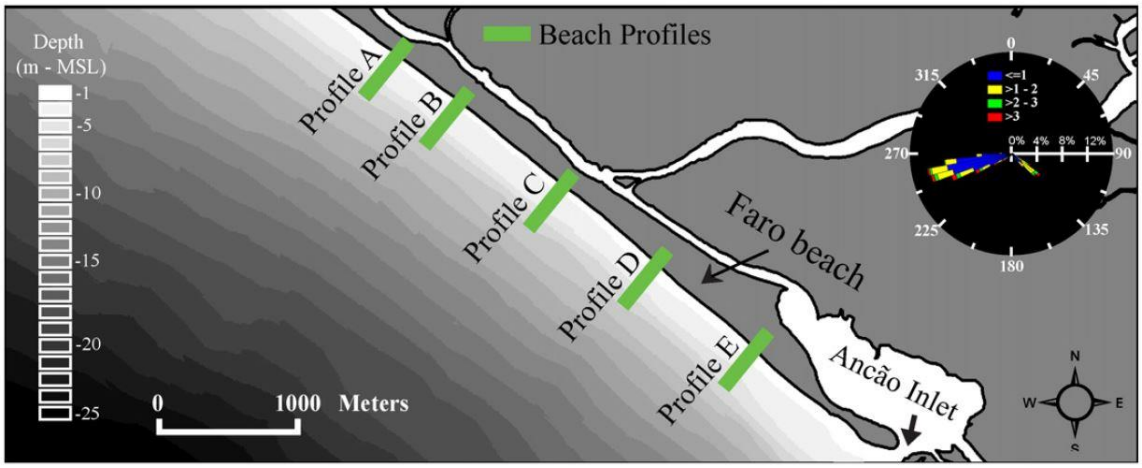

Figure A-1 Locations of the cross sections for which runup calculations have been made to determine storm impact thresholds. (Almeida et al., 2012)

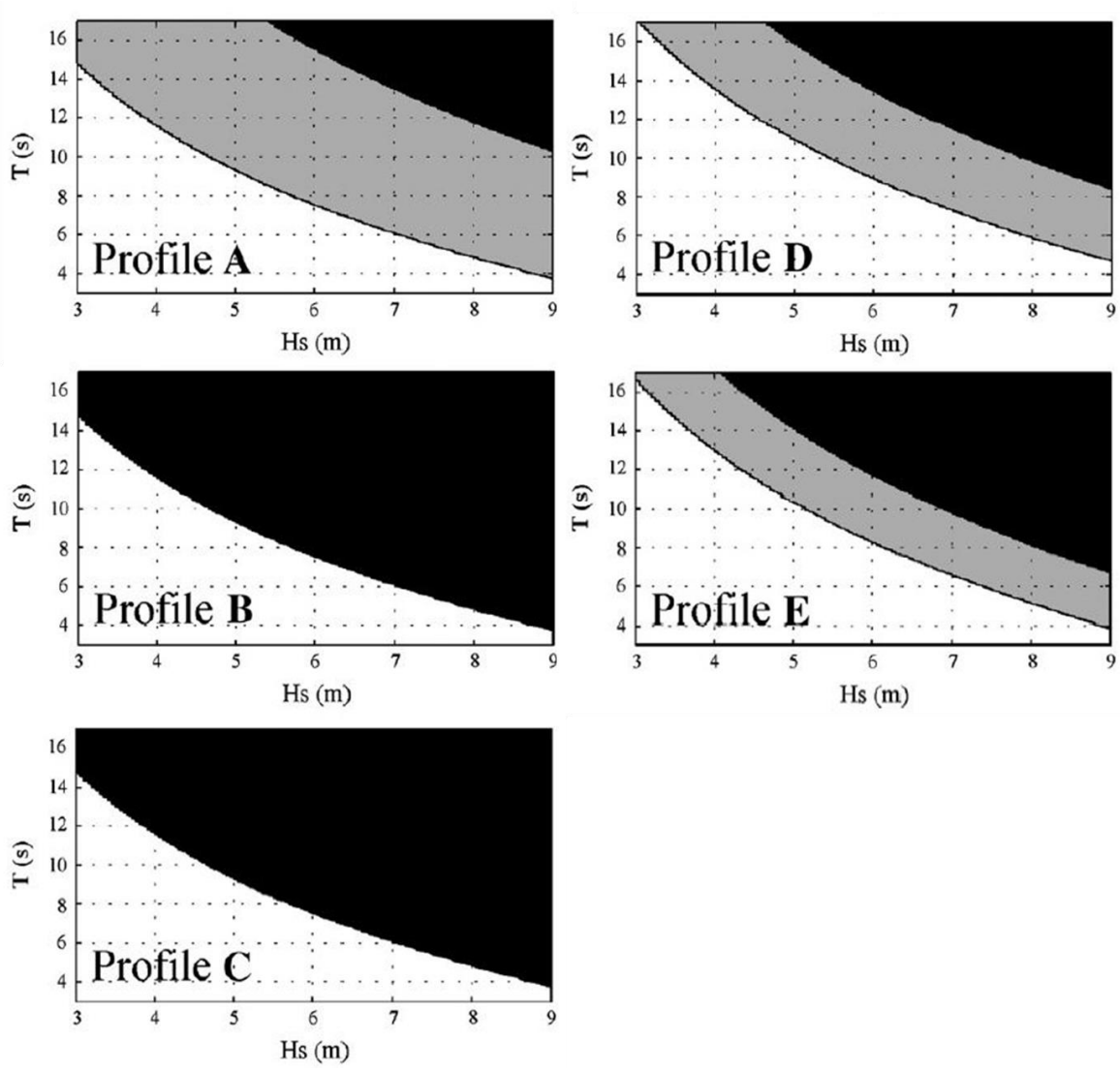

Figure A-2 Thresholds for storm impacts with varying period and wave height (surge levels are coupled to wave heights). Black represents overwash regime, grey collision regime and white swash regime. (Almeida et al., 2012)

B XBEACH SETUP

B.1 Topography and Bathymetry

Topographic and bathymetric information has been provided by the University of The Algarve. Information from the summer of 2007 is available for a stretch of 8 kilometers on the Ancão peninsula. It consists of bathymetric cross sections of the wet areas and a LIDAR of the dry area. The LIDAR gives an accuracy of 2 by 2 meters.

An overview of the available topography and bathymetry is given in Figure B-1. The area of interest has a total with of 1 kilometer and is centered on the entrance road and parking lot, as indicated by the red box. Another important location is a former video monitoring station of which the data will be used for calibrating the model which also falls within this model domain.

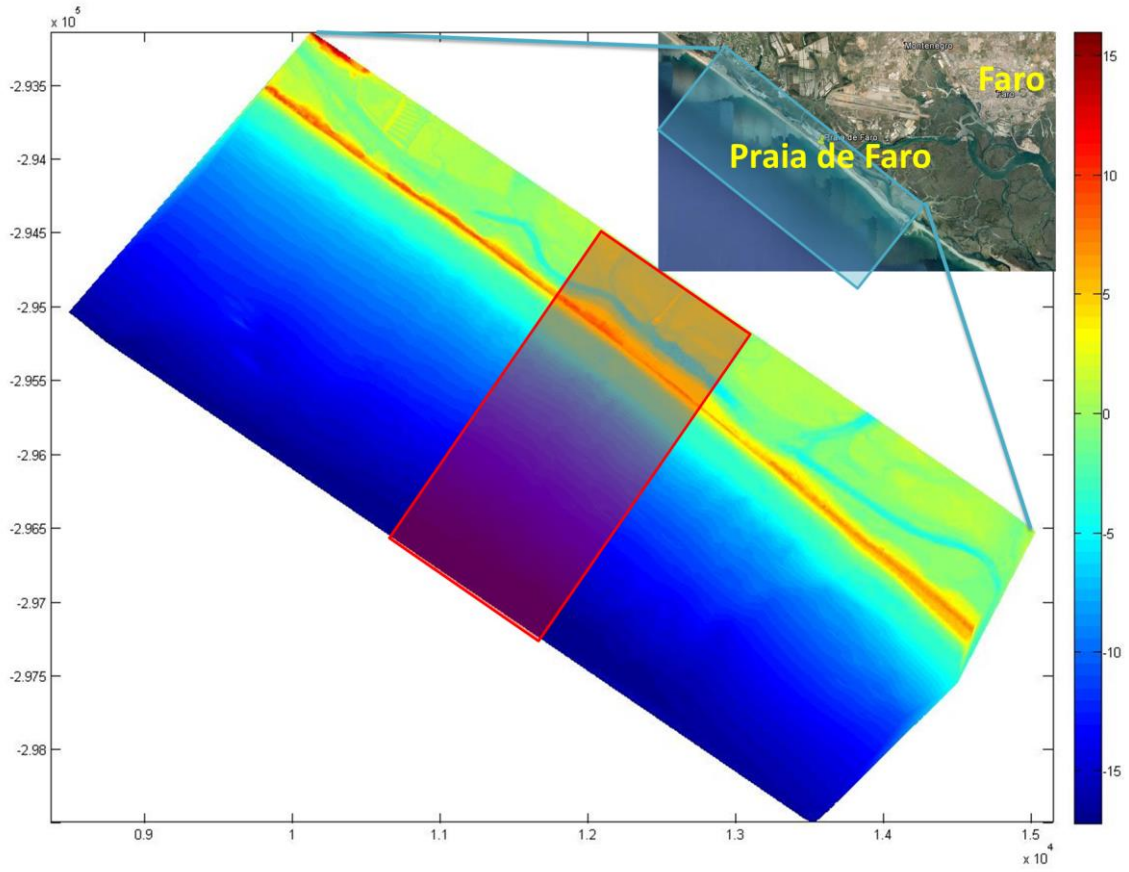

Figure B-1 Available lidar data and location of research area.

B.2 Grid Setup

A grid is defined based on several, sometimes contradicting, requirements that will be explained in this paragraph. The most important factors are listed below.

- The depth at the offshore boundary has to be sufficiently large for the assumption of deep water for wave propagation to hold.

The waves supplied as input for XBeach are deep-water waves, meaning that the water depth has negligible influence on their propagation. If the depth at the offshore boundary is not large enough for this assumption to hold then the waves should have deformed before they entered the domain. It is therefore important that the depth is large enough so that the waves may deform within the XBeach model. Generally for XBeach models this depth is chosen at 20 meters.

- The width of the domain has to be determined according to the effect of the shadow zone.

Waves enter the grid only at the offshore boundary, and not at the lateral boundaries. When waves enter the domain at an angle a triangular shadow zone will be formed with much less wave energy, as illustrated in Figure B-2: the light grey area is the shadow zone for waves entering the domain at an angle of 220 degrees (nautical). It is important that this zone does not affect the area of interest, as it will distort the results.

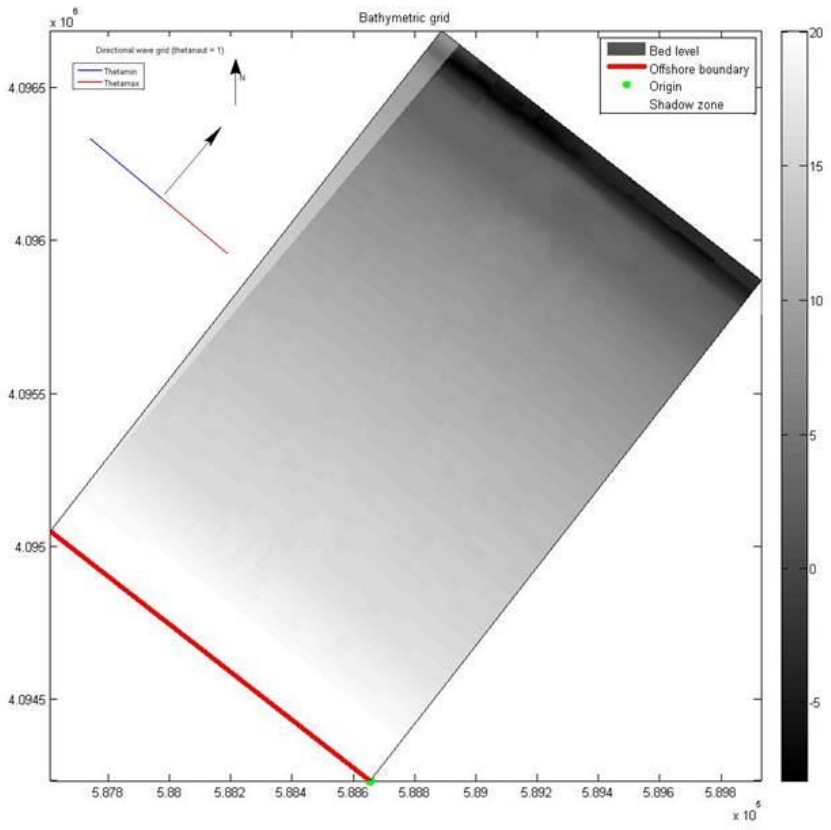

Figure B-2 Shadow zone in XBeach due to oblique wave attack.

- The grid resolution at the shoreline has to be fine enough to model the beach face and dune erosion and overwash.

The purpose of this model is to give results concerning coastline retreat and overwash events. The shape of the beach face and dune should therefore be well incorporated in the model. Changes to this area are desired output and also affect the run-up and overwash. The grid size around this area is therefore chosen at approximately 0.65 meters.

- The total amount of grid cells should be kept at a minimum to reduce the runtime.

The runtime of XBeach largely depends on the amount of grid cells in the grid. More grid cells simply mean more calculations per time step and should therefore be kept at a minimum.

- In case of varying grid cell sizes the transition between two cells has to be smooth enough to prevent numerical instabilities.

Varying grid cell sizes are favorable as it allows the grid to be detailed in areas of interest and less detailed in other areas. However, large differences between neighboring grid cells give rise to numerical instabilities in XBeach, causing runs to fail or output to reach unrealistic values. This translates to a small cross shore grid cell size at the beach face, in this case around 0.65 meters, and a large grid cell size at the offshore boundary of about 30 meters. In the long shore the grid size can also be varied to focus on the area of interest purchased. In the center, at the location of the parking lot and video monitoring station the grid is 5 meters wide, varying to a maximum of 20 meters at the edges (Figure B-3).

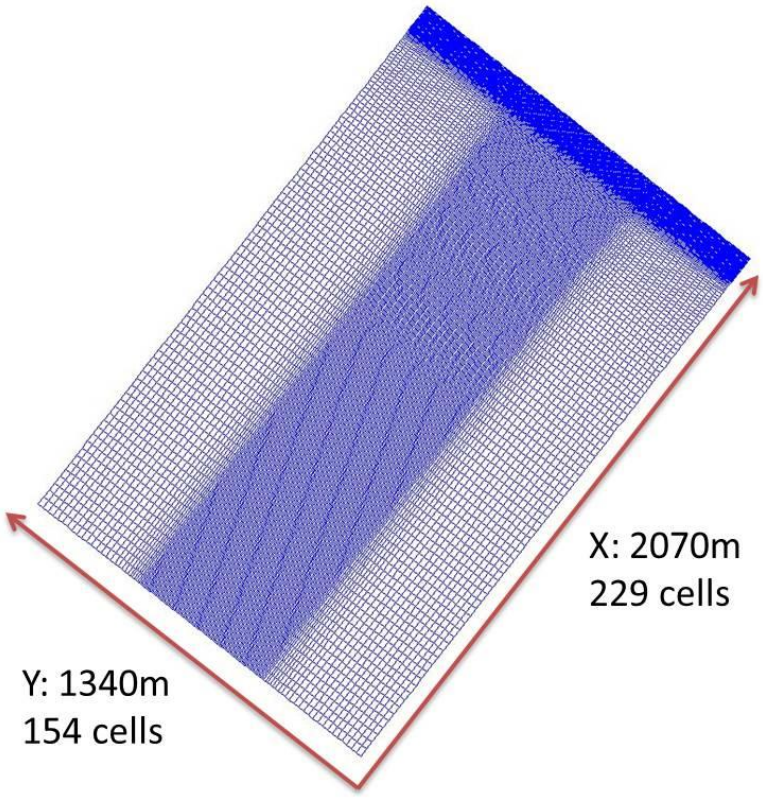

Figure B-3 XBeach grid setup:. (max dx = 30m, min dx = 6.5m, max dy = 20m, min dy = 5m

B.2.1 Model Domain

The coordinate system of XBeach is defined as shown in Figure B-4.The y-axis has to be oriented approximately parallel to the coastline and the x-axis perpendicular to the coastline. Waves enter the model at the offshore boundary, labeled as front. The coordinates can be either supplied in world coordinates or in a local coordinate system for which an origin (xori, yori) and rotation (alfa) have to be given.

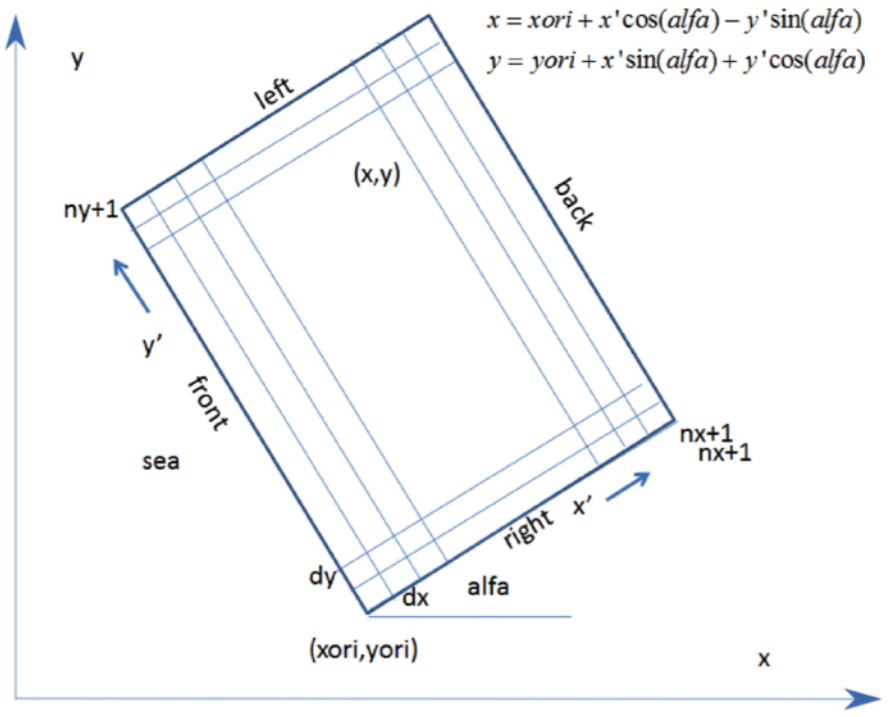

Figure B-4 Coordinate system of XBeach

For the setup of the grid the world coordinate system is chosen, which has to be metric. The raw topographic data have been supplied in the spherical coordinate system WGS83 and have been converted to the metric UTM29 coordinate system. The main wave direction considered is the southwest. The orientation of the coastline is from northwest to southeast so that the shadow zone is very small. This final grid is shown in Figure B-7**,** indicating the main wave direction, shadow zone and most important locations.

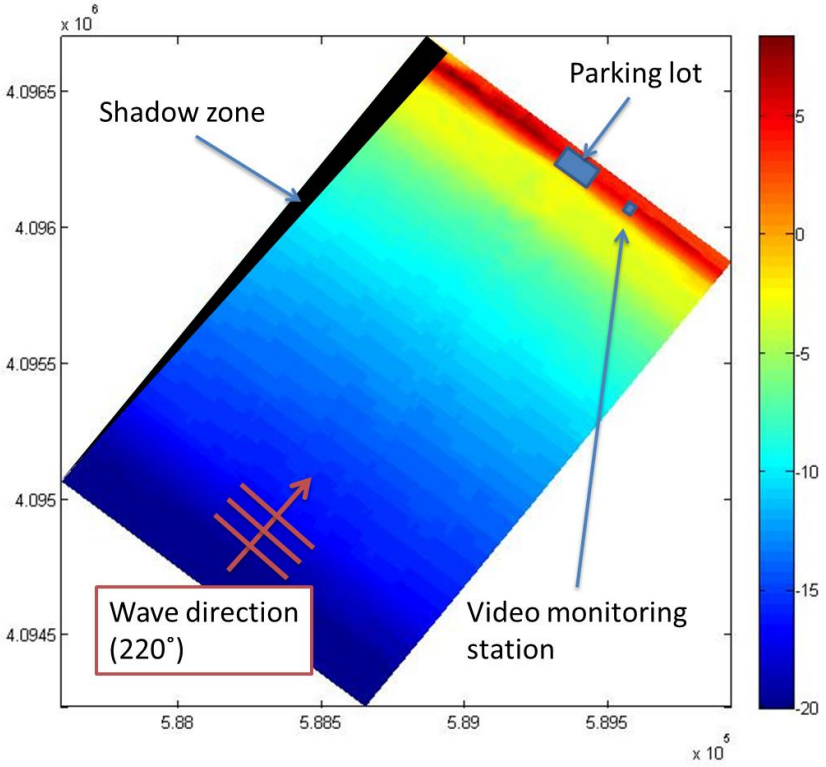

Figure B-5 XBeach model setup showing the most important locations, the wave direction and shadow zone.

The supplied topographic data include a large berm, which is generally present during summer. This berm has large influence on the total amount of runup, limiting coastline erosion and overwash. Since the berm is not present during the winter season, in which most storms occur, it is removed from the profile.

The berm is removed by substituting it with a straight line from the point at which the profile is 1,5 meter above the mean sea level until 10 grid cells forward of the top of the dune. The final difference in volume at the beach is approximated to be 260 m^3. The result of the removal of the berm is shown in Figure B-6 and Figure B-7**.** It can be seen that most of the sand is redistributed along the topography, rather than being completely removed, which is important for the total sediment budgets.

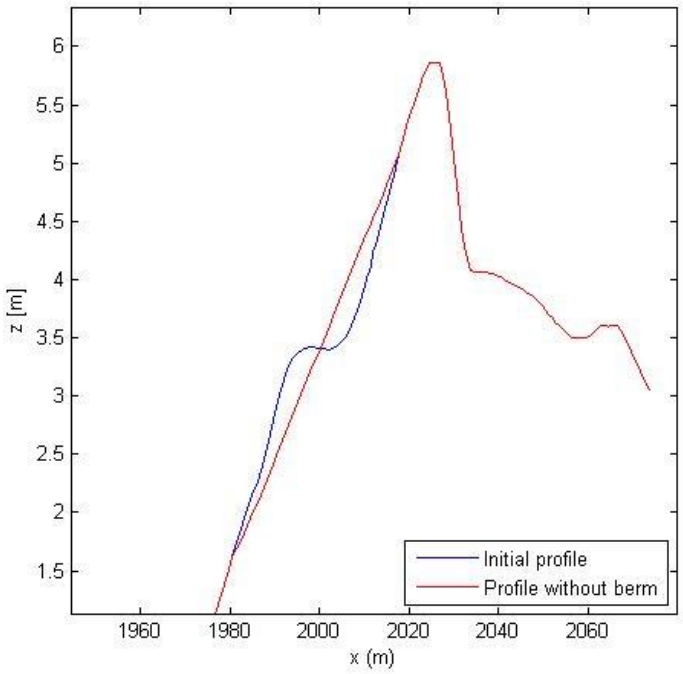

Figure B-6 Cross sections of the topography with and without a berm at the beach face.

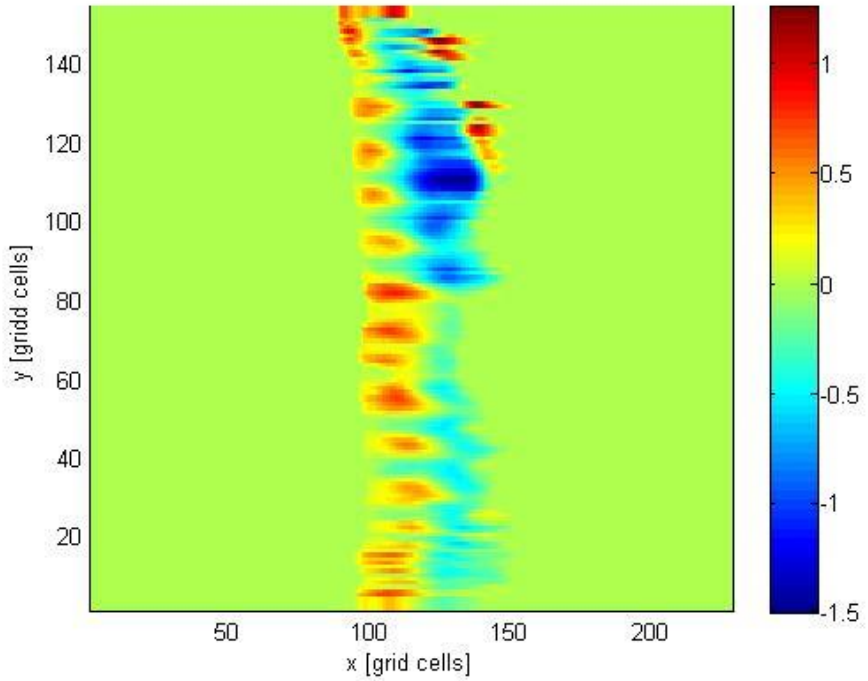

Figure B-7 Difference plot between the bathymetry with and without the berm at the beach. Axis are grid cells; the colors indicate change in depth.

B.3 Waves

There are different options to supply XBeach with wave boundary conditions, ranging from stationary conditions to time varying two-dimensional wave spectra. The wave conditions supplied are deep-water waves and have to be supplied at the offshore boundary.

To specify a JONSWAP spectrum for as a boundary condition for XBeach a file has to be created containing the parameters for the wave spectra for each individual hour. The parameters that have to be given are:

- H_s, the significant wave height (m)
- T_p, the peak period (s)
- Mainang, the main direction of the waves (°)
- s, the directional spreading coefficient: $s = \frac{2}{\sigma^2} - 1$ (-) , in which σ is the directional spreading.
- Gammajsp, the peak enhancement factor of the JONSWAP expression (-)
- Duration, in this case an hour, specified in seconds.

Since there is no available spectral analysis of the sea state in the south of Portugal the default value of the peak enhancement factor has been used for the shape of the JONSWAP spectrum (3.3). The directional spreading coefficient has been set to 20, resulting in a directional spreading of 0.31 radians (or 17.7 degrees).

B.4 Calibration

To perform well the XBeach model has to be calibrated to include the effects of the local conditions. Preceding this research, efforts have been made at the University of The Algarve to calibrate a 2D XBeach model for Praia de Faro. These settings are unpublished and will be described in the next section. For further calibration a dataset is available from a research concerning beach morphological changes during consecutive storms at Praia de Faro. Although the usability of the dataset is not very high for calibration purposes it has been used as such. The dataset and calibration efforts are treated in the following sections.

B.4.1 Preceding Calibration Efforts

Preceding this research there have been calibration efforts for an XBeach model at Praia de Faro. These efforts have led to a number of settings that deviate from the XBeach defaults and are summarized in Table B-1. Furthermore there are new and improved default settings for XBeach, referred to as the WTI settings, which are also yet unpublished. These settings are summarized in Table B-2.

Table B-1 XBeach settings of preceding calibration efforts

Parameter	Value	Units	Description
D50	0.005	m	D50 grain size per grain type
D90	0.002	m	D90 grain size per grain type
bedfriction	Manning	-	Bed friction formulation
bedfriccoef	0.02	$s/m^{1/3}$	Bed friction coefficient
delta	0.1	-	Fraction of wave height to add to water depth

Table B-2 XBeach WTI settings

Parameter	Value	Units	Description
cf	0.001	-	Friction coefficient flow
gammax	2.364	-	Maximum ratio wave height to water depth
beta	0.138	-	Breaker slope coefficient in roller model
wetslp	0.260	-	Critical avalanching slope under water (dz/dx and dz/dy)
alpha	1.262	-	Wave dissipation coefficient in Roelvink formulation
facSK	0.375	-	Calibration factor time averaged flows due to wave skewness
facAs	0.123	-	Calibration factor time averaged flows due to wave asymmetry
gamma	0.541	-	Breaker parameter in Baldock or Roelvink formulation

B.4.2 Dataset

The dataset available consists of two parts; beach profiles obtained from time stack images and hourly offshore storm conditions. The data have been collected during a series of storms over the period December 2009 – January 2010, in which several storm occurred. The wave data originates from a wave buoy at 93 meters depth from the Portuguese Hydrographic Institute (Figure B-8). The topographic data are collected from a video station located on top of a roof of a building facing the beach at Praia de Faro (Figure B-9). Images have been acquired every hour for 10 minutes, during daylight. Tidal data have also been collected using a pressure transducer that has been deployed at 14 meters depth offshore of the study area. (Vousdoukas et al., 2012a)

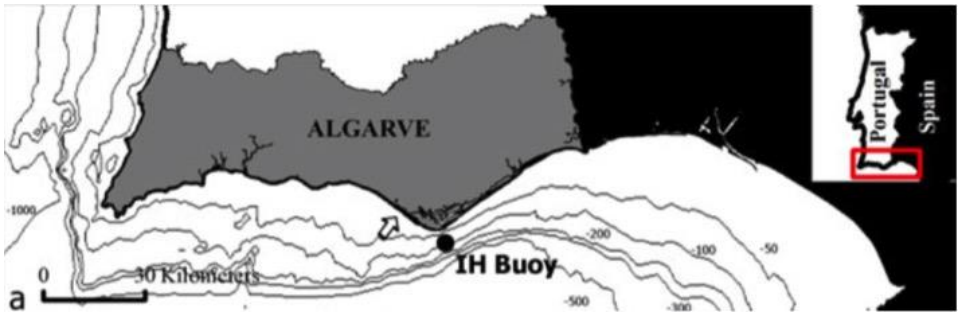

Figure B-8 Map of the southern region of Portugal, showing the location of the wave buoy and the location of the study site.
(Vousdoukas et al., 2012a)

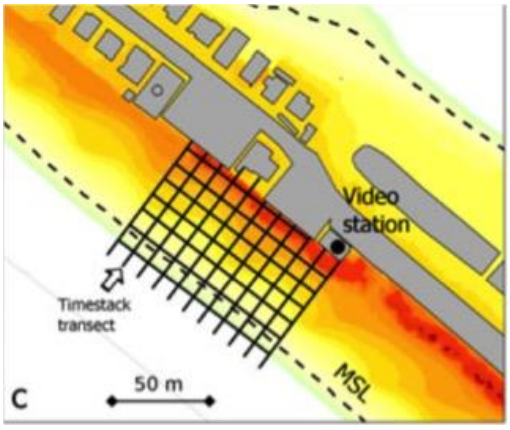

Figure B-9 Map of the study site showing the location of the video monitoring station and the topographic survey grid. (Vousdoukas et al., 2012a)

The hydrodynamics conditions during the study period are displayed in Figure B-11. In the period six separate events are identified labeled A through F, with a significant wave height of 2 meters or higher. The changes in the beach profile above MSL have been measured for these 6 events and are summarized in Table B-3.

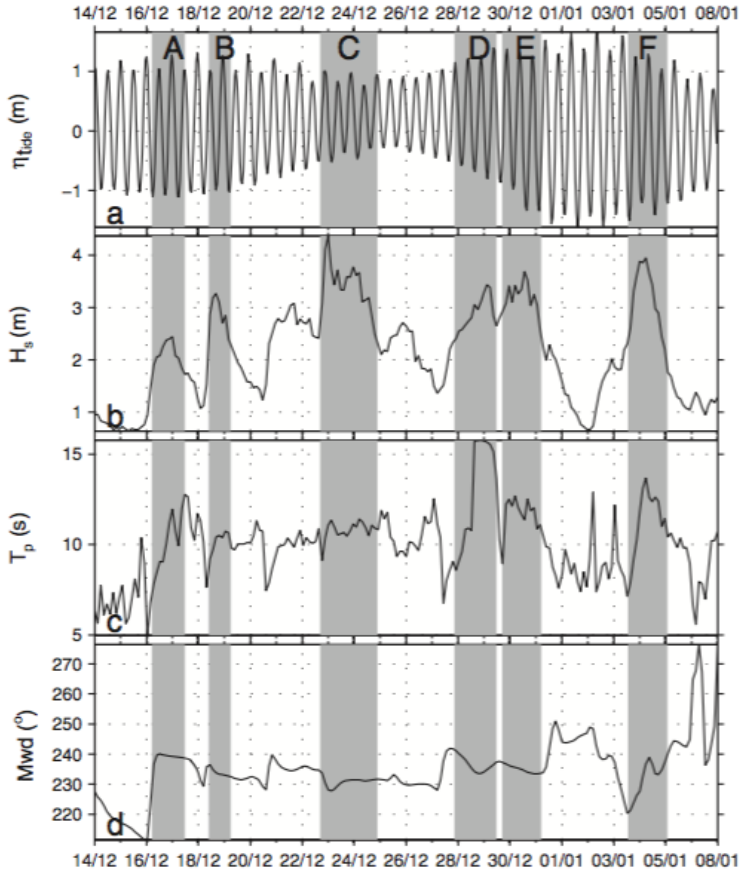

Figure B-10 Hydrodynamic conditions collected by the Hydrographic Institute during the period December 2009 - January 2010. (Vousdoukas et al., 2012a)

Table B-3 Erosion volumes per event. dV is the difference between the post and pre storm sediment volumes above MSL.

Event	dV (m3/m)
A	-14
B	-9
C	-2.5
D	-13
E	-13
F	-1

B.4.3 Calibration Efforts

The calibration efforts have been focused on two aspects, the erosion volumes and the shape of the beach profile. The calibration on the erosion volumes has been done for three selected events: B, D and E. The calibration on the shape of the beach profile (including runup) has been focused at event B. This choice is made since this event has the shortest duration and therefore more XBeach runs can be performed in the least amount of time.

Calibration for the erosion volumes can be done by tweaking the calibration factors of the wave asymmetry and skewness in XBeach (keywords: facAs and facSK). They are accounted for in the advection-diffusion equation in XBeach by influencing a velocity component $u_a = (f_{Sk} S_k - f_{As} A_s) u_{rms}$. For a detailed description is referred to the XBeach manual. In essence the higher the value for u_a the higher the onshore sediment transport. Nine iterations have been performed on storms B, D and E to establish the best set of asymmetry and skewness calibration factors. The storms have been run using the time series from Figure B-10. The waves have been modeled using a JONSWAP spectrum with gamma = 3.3 and a directional spreading coefficient of 10. The values used for facAs and facSk for the iterations are shown in Table B-4. For three cross sections of the XBeach model that are located in the same area in which the change in the beach volumes have been measured the change in volume above MSL have been determined. The results are shown below in figures Figure B-11, Figure B-12 and Figure B-13. From these results the calibration values of iteration 2 are thought to give the best fit and these will be used for further calibration.

Table B-4 Calibration values used for the facAs adn facSk for the iterations.

Iteration	facas	facsk
1	0.4	0.2
2	0.4	0.3
3	0.4	0.4
4	0.5	0.2
5	0.5	0.3
6	0.5	0.4
7	0.6	0.2
8	0.6	0.3
9	0.6	0.4

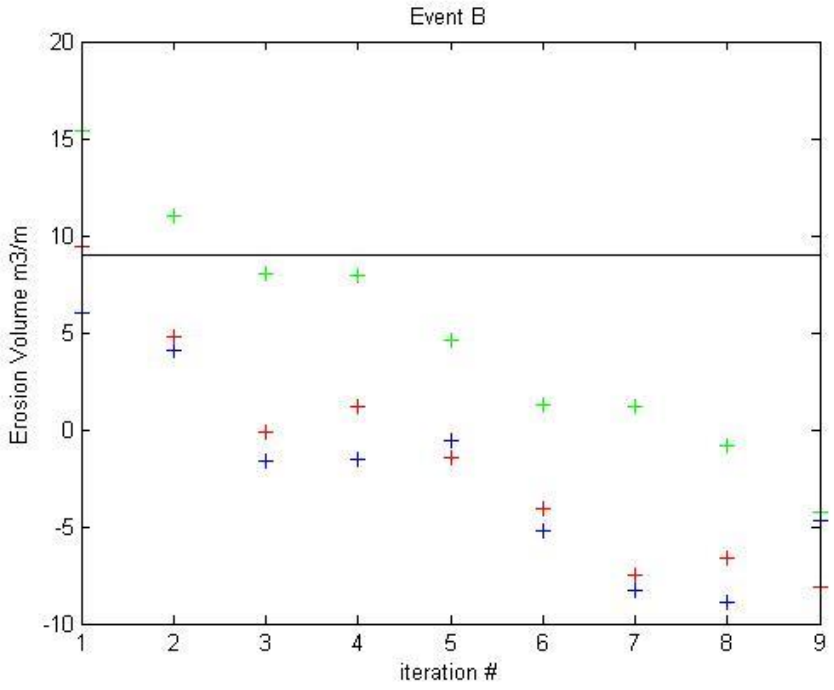

Figure B-11 Erosion volumes for three cross sections at the video monitoring station for event B. The horizontal line indicates the measured erosion volume.

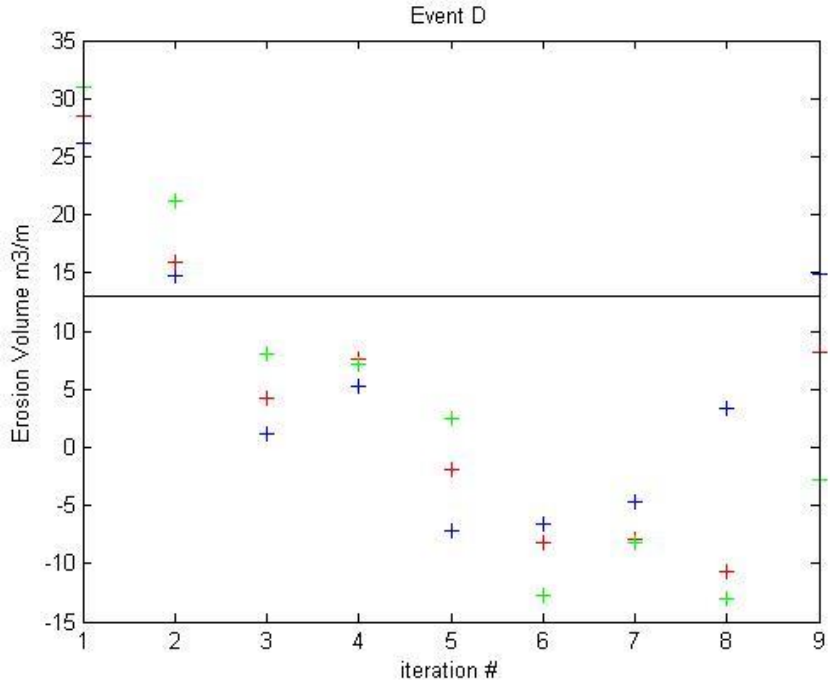

Figure B-12 Erosion volumes for three cross sections at the video monitoring station for event D. The horizontal line indicates the measured erosion volume.

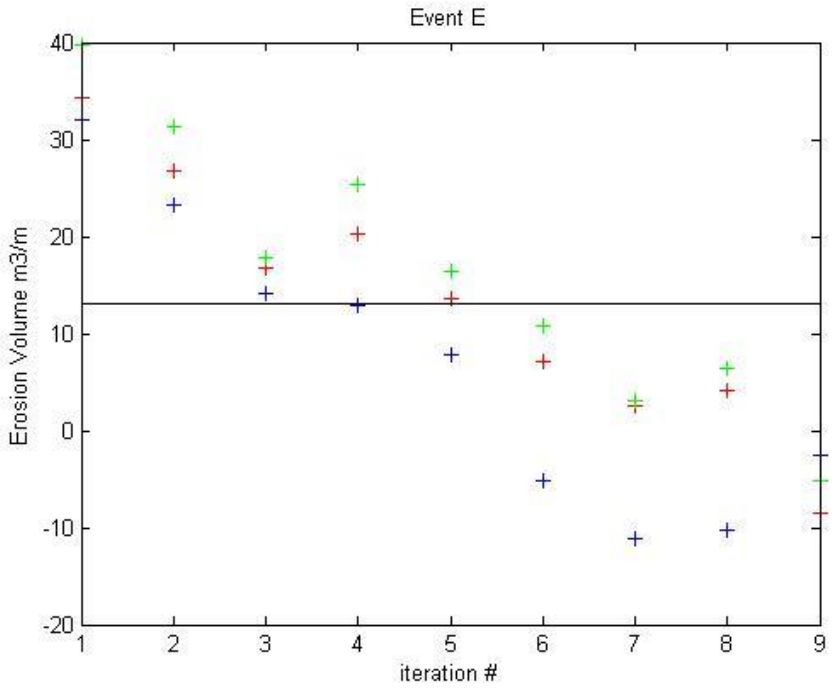

Figure B-13 Erosion volumes for three cross sections at the video monitoring station for event E. The horizontal line indicates the measured erosion volume.

The measured profiles from the area show that there is no scarping during the storm. XBeach however, does induce a scarp under the given conditions, see Figure B-14. As soon as the scarp is formed it starts to limit the maximum runup and thereby limits the beach face erosion to the location of the scarp. Without a scarp waves could cause higher runup and the beach face would be eroded up to a higher level. To limit the formation of the scarp and to increase the runup levels several things can be adjust in XBeach.

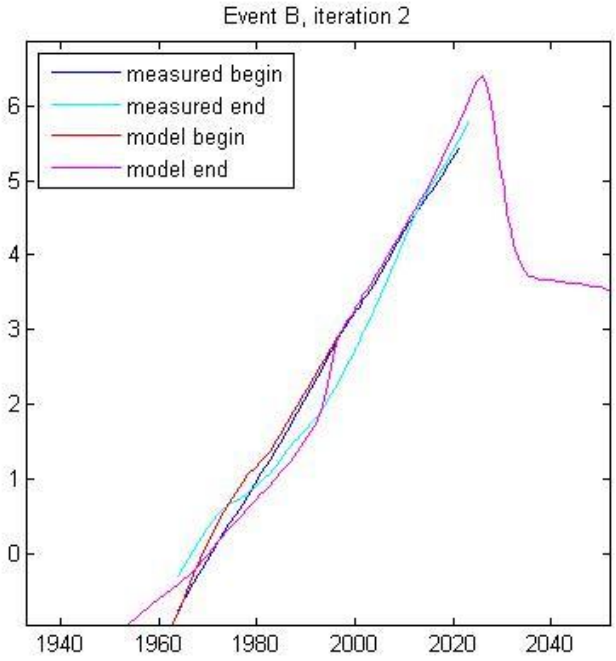

Figure B-14 Measured and modeled profiles for iteration 2, event B. A scarp is formed by XBeach that is not seen in the measured profiles.

The scarp formation is the effect of the avalanching algorithm used in XBeach. The avalanching algorithm causes slumping of sand from the dune to the foreshore during storm erosion. Since wet sand is more prone to slumping than dry sand a critical slope is defined for both with the keywords *wetslp* and *dryslp*. The criteria that determines if sand should be treated as dry or wet is controlled with the keyword *hswitch*. It defines a minimum water level needed for sand to be treated as wet. The default values are: *wetslp* = 0.3, *dryslp* = 1 and *hswitch* = 0.1 m. To decrease the scarp formation the default values have been lowered. The effect is visual but does not prevent the scarp formation from occurring.

Another factor that can influence the total runup is the wave groupiness. If the groupiness of the waves is higher the forcing due to infragravity waves is also larger. When specifying the wave input one factor that has to be given is the directional spreading coefficient, *s*. Higher directional spreading can smoothen the groupiness and therefore lead to lower runup heights. The directional spreading coefficient is inversely proportional to the directional spreading and therefore a higher value causes more runup. The default value is 10, giving a directional spreading of 24.4 degrees.

Initially the surge levels during the storm had not been added in the modeling since these were very small. These have been added later to see the effect on the runup, and as will be seen later the runup is only a little higher.

Since there is no research on the shape of the wave spectrum for this part of the Portuguese coastline the effect of using a JONSWAP spectrum with a gamma of 3.3 is observed by modeling the Xynthia storm of 2010. From a larger SWAN model for the area full 2D spectra are extracted at the

location of the offshore boundary of the XBeach model. The model is forced with these wave spectra as well as with JONSWAP wave spectra.

In the table below an overview is given of all combinations that have been run in XBeach. The resulting profiles are also given in the following figures.

Table B-5 Calibration values for iterations 10 to 15 and the Xynthia model runs.

Iteration	*avalanching*	*wetslp*	*dryslp*	*hswitch*	*surge*	*s*
10	no	0.1	1	0.1	no	10
11	yes	0.1	0.5	0.1	no	10
12	yes	0.1	0.5	0.05	no	10
13	yes	0.1	0.5	0.05	yes	10
14	yes	0.1	0.5	0.05	yes	20
15	yes	0.1	0.5	0.05	yes	30
xynthia-2D swan	yes	0.1	0.5	0.05	yes	10
xynthia-JONSWAP	yes	0.1	0.5	0.05	yes	10

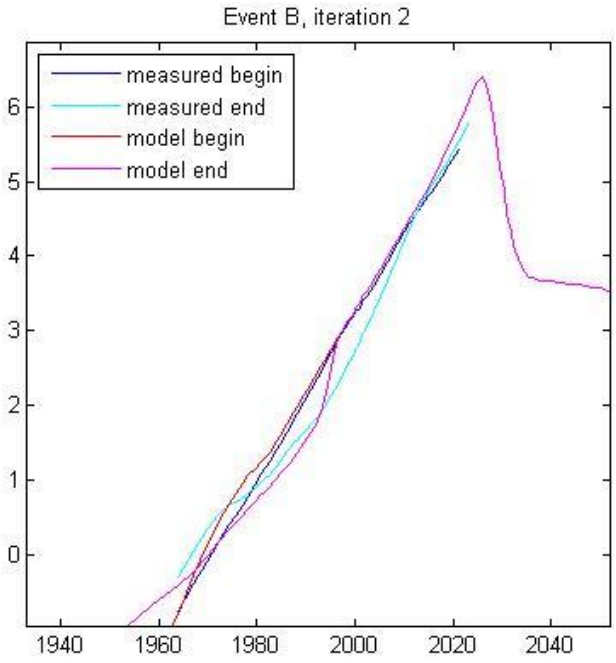

Figure B-15 Measured and modeled profiles for iteration 2, event B.

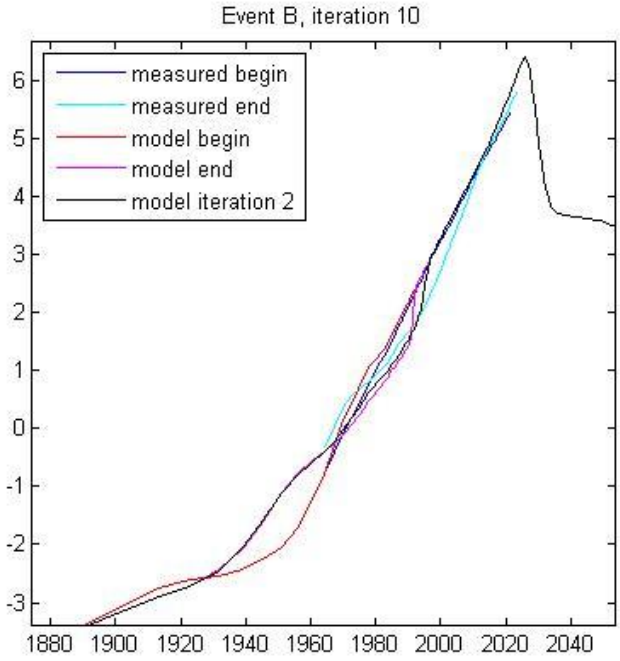

Figure B-16 Measured and modeled profiles for iteration 10, event B.

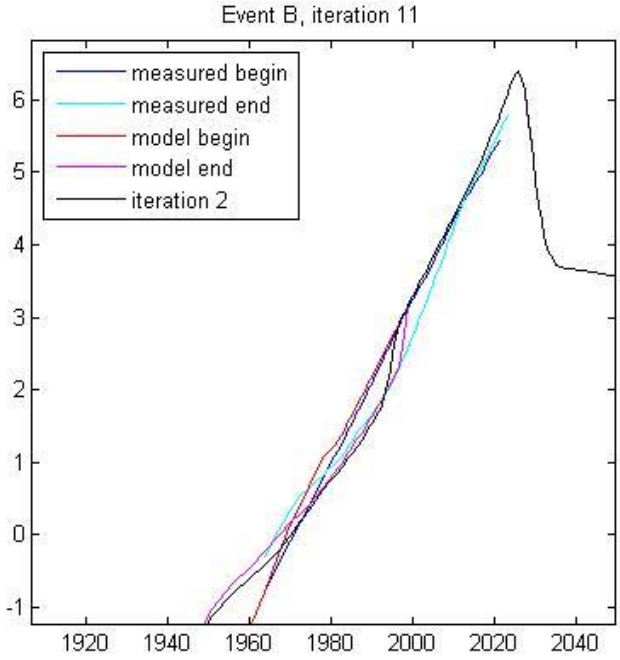

Figure B-17 Measured and modeled profiles for iteration 11, event B.

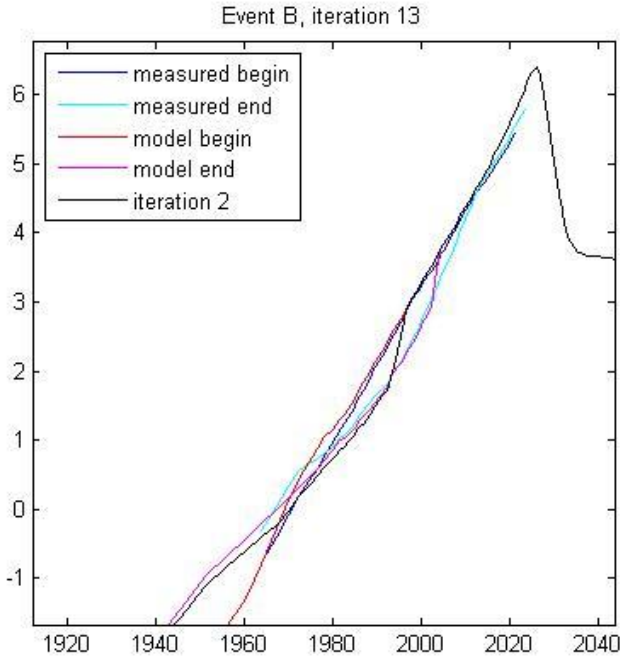

Figure B-18 Measured and modeled profiles for iteration 13, event B.

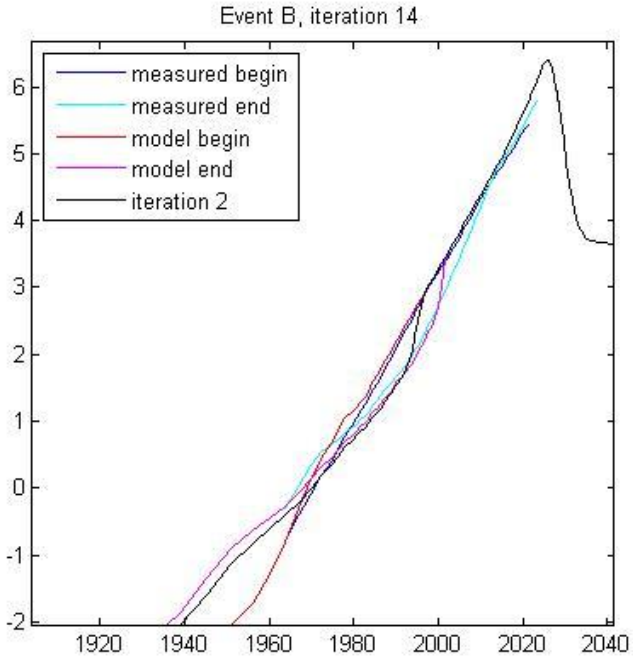

Figure B-19 Measured and modeled profiles for iteration 14, event B.

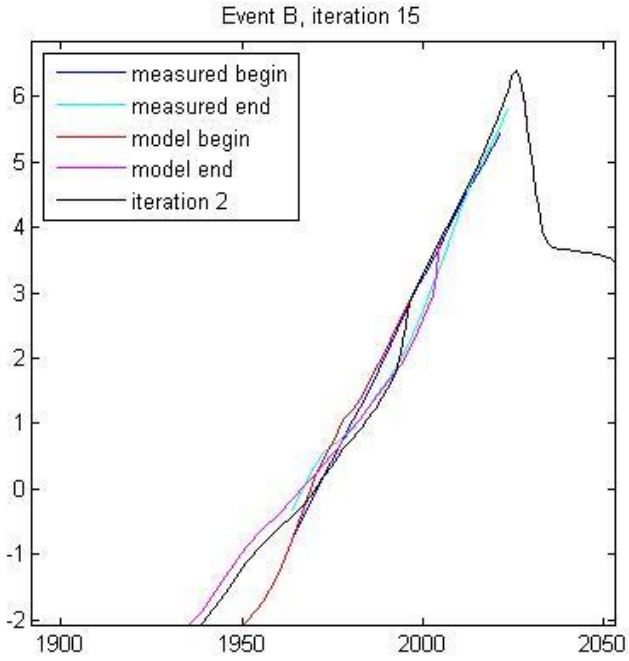

Figure B-20 Measured and modeled profiles for iteration 15, event B.

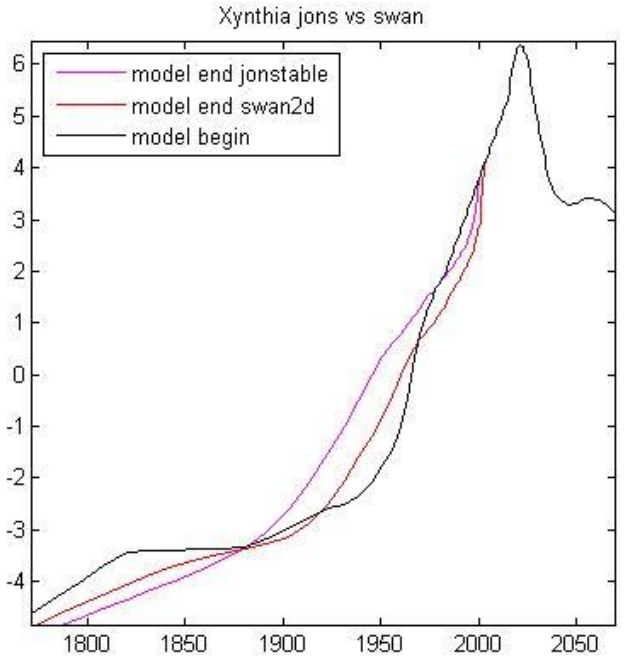

Figure B-21Measured and modeled profiles for Xynthia.

B.4.4 Final Calibration Settings

The final calibration settings have been determined based on the previous section and based on the expert judgment of the developers of XBeach. The settings are summarized in Table B-6.

Table B-6 Final calibration settings of XBeach.

Parameter	Value	Units	Description
D50	0.005	m	D50 grain size per grain type
D90	0.002	m	D90 grain size per grain type
bedfriction	Manning	-	Bed friction formulation
bedfriccoef	0.02	$s/m^{1/3}$	Bed friction coefficient
delta	0.1	-	Fraction of wave height to add to water depth
cf	0.001	-	Friction coefficient flow
gammax	2.364	-	Maximum ratio wave height to water depth
beta	0.138	-	Breaker slope coefficient in roller model
alpha	1.262	-	Wave dissipation coefficient in Roelvink formulation
facSK	0.3	-	Calibration factor time averaged flows due to wave skewness
facAs	0.4	-	Calibration factor time averaged flows due to wave asymmetry
gamma	0.541	-	Breaker parameter in Baldock or Roelvink formulation
dryslp	1	-	Critical avalanching slope above water.
wetslp	0.1	-	Critical avalanching slope under water.
hswitch	0.05	m	Water level at which is switched from wetslp to dryslp.
s	20	-	Directional spreading coefficient

C SYNTHETIC DATASET

C.1 Marginal Distribution Fitting

The ranking of the marginal distributions that have been fit to the individual variables are listed in Table C-1, the values in the table are determined using the Akaike information criterion. In Table C-2 the parameters of the chosen marginal distributions are given. Furthermore the plots of the parametric and empirical CDFs are shown in the figures below.

Table C-1 Ranking of the fitted marginal distributions according to the Akaike information criterion.

Distribution	Hs (duration)	Tp	Duration	Hs (surge)	Surge
Exponential	521,28	787,04	**1037,95**	397,05	-60,83
Extreme value	365,18	588,50	1280,71	259,03	-28,80
Gamma	288,39	557,33	-	197,82	-94,76
Generalized extreme value	265,59	558,54	1055,90	162,76	**-96,94**
Generalized Pareto	**237,92**	561,06	1038,15	**147,31**	-81,63
Logistic	302,54	569,62	1146,91	204,01	-91,14
Loglogistic	284,83	563,99	-	187,93	-90,83
Lognormal'	281,29	**556,84**	-	191,18	-78,24
nakagami	297,13	559,09	-	205,38	-95,12
Normal	308,03	563,39	1176,28	213,72	-80,41
Rayleigh	383,45	648,17	-	291,85	-95,87
Rician	307,16	563,00	-	213,28	-93,87
Tlocationscale	303,81	565,39	1125,77	197,28	-92,07
Weibull	318,24	566,73	-	227,26	-95,87

Table C-2 Parameter values of the used marginal distributions

Variable	Used distribution		Parameters	
Hs (duration)	Generalized Pareto	k = -0.1514	sigma = 1.1732	theta = 2.5
Tp (duration)	Lognormal	mu = 2.3829	sigma = 0.2481	
Duration	Exponential	mu = 33.2522		
Hs (surge)	Generalized Pareto	k = -0.0166	sigma = 0.8674	theta = 3.5
Surge	Weibull	A = 0.2841	B = 1.7900	

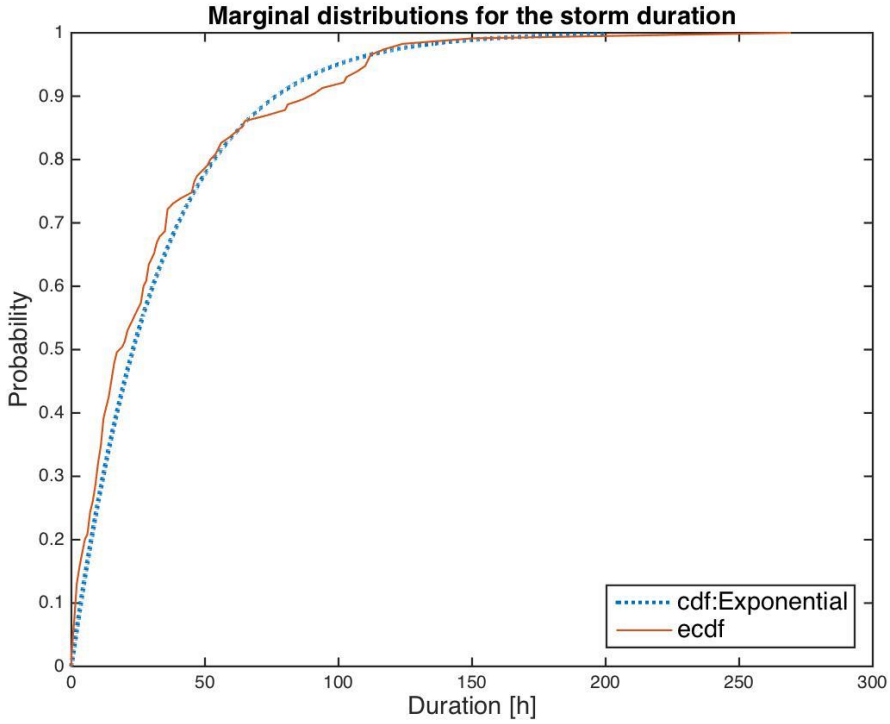

Figure C-1ECDF and CDF for the storm duration.

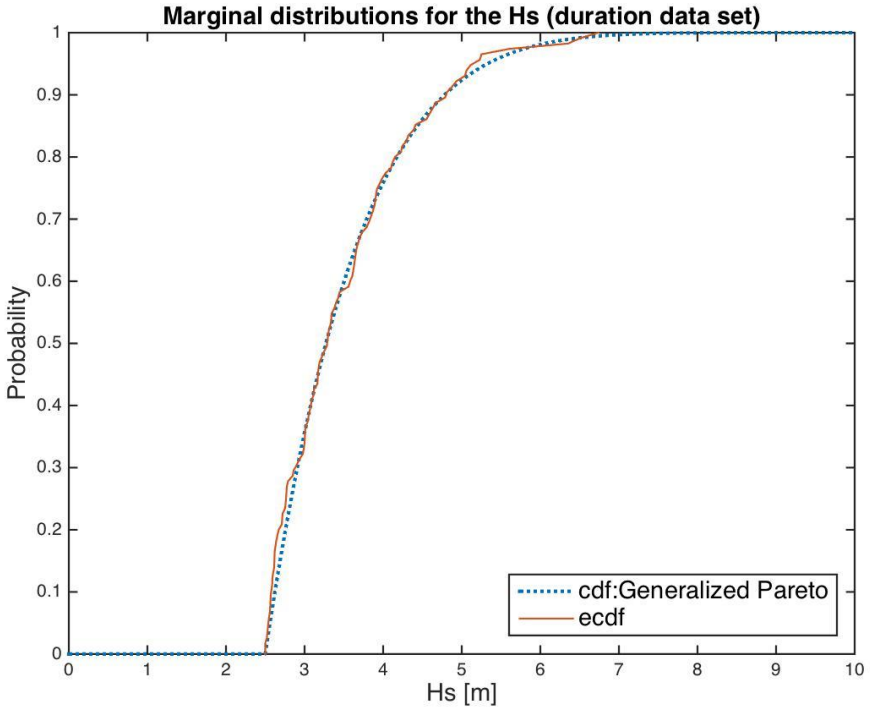

Figure C-2 ECDF and CDF for the significant wave height of the duration data set.

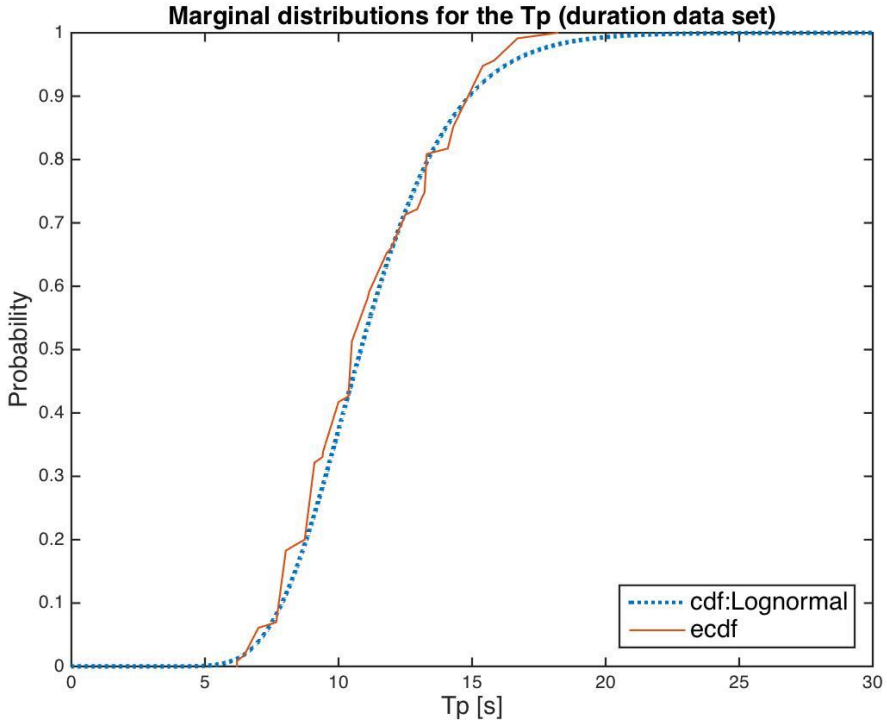

Figure C-3 ECDF and CDF for the peak period of the storms.

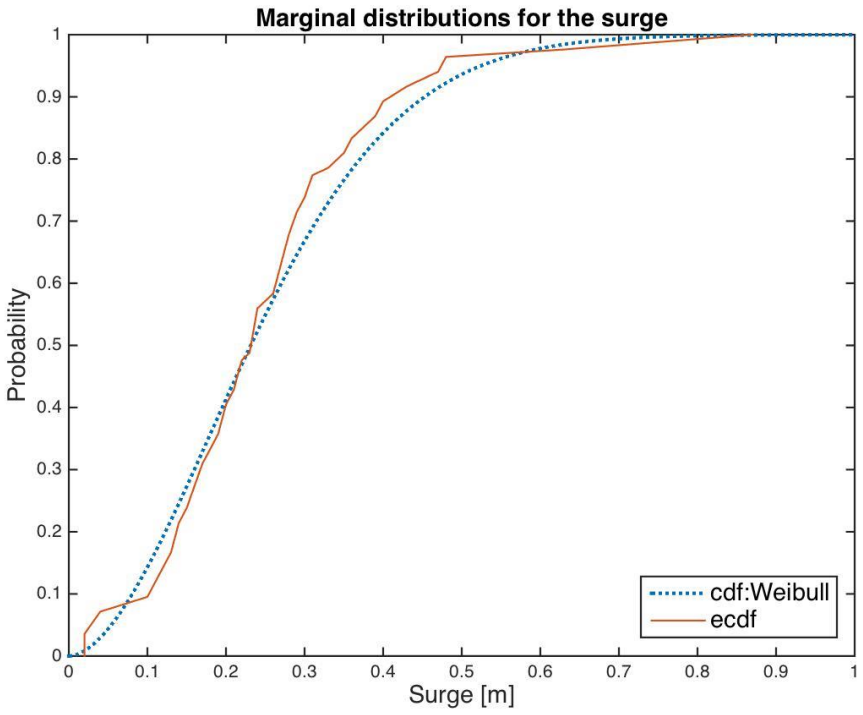

Figure C-4 ECDF and CDF for the storm duration.

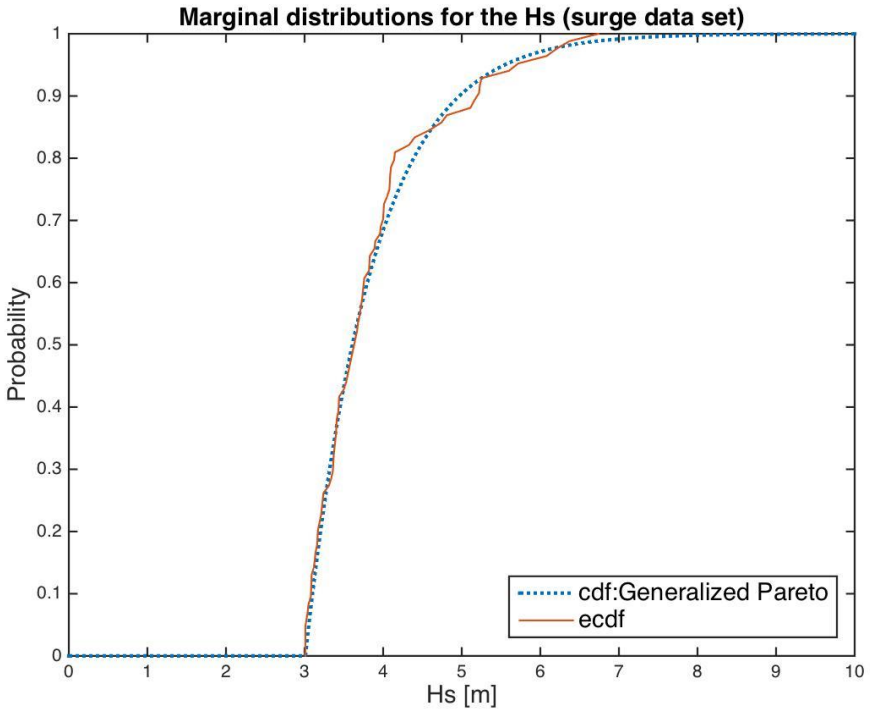

Figure C-5 ECDF and CDF for the storm surge levels.

C.2 Copula fitting Figures

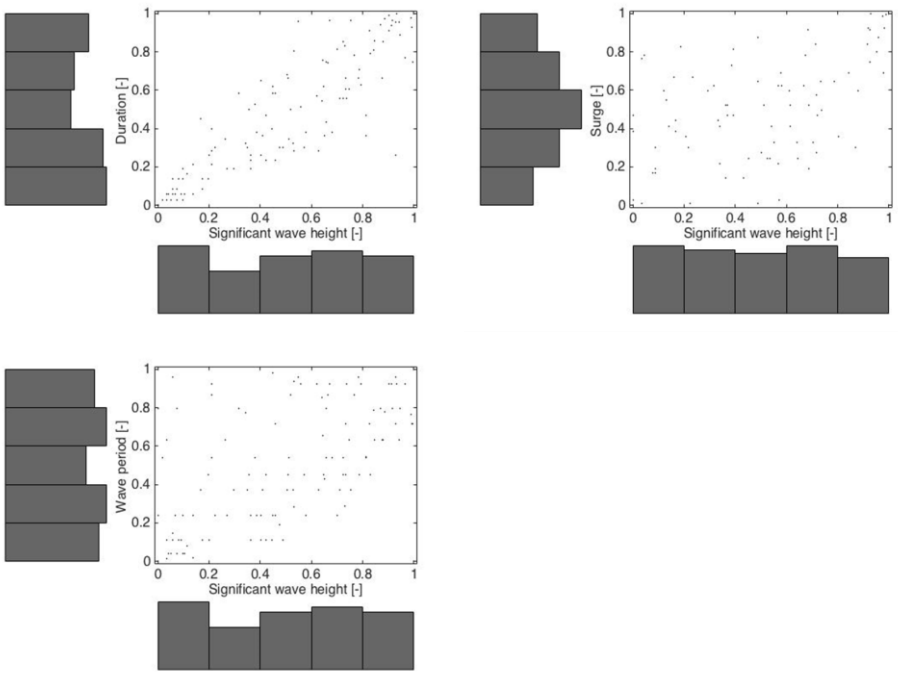

Figure C-6 Scatter plots of the uniformly distributed variable pairs.

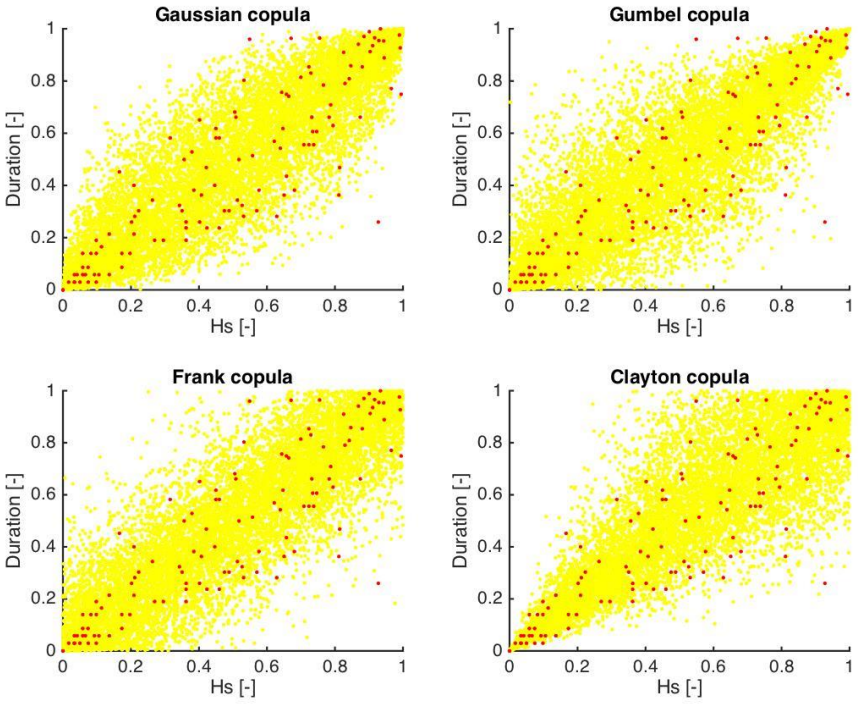

Figure C-7 Different copulas fit to the Hs and Duration dataset. The red dots indicate the original dataset and the yellow dots the data points generated with the copula.

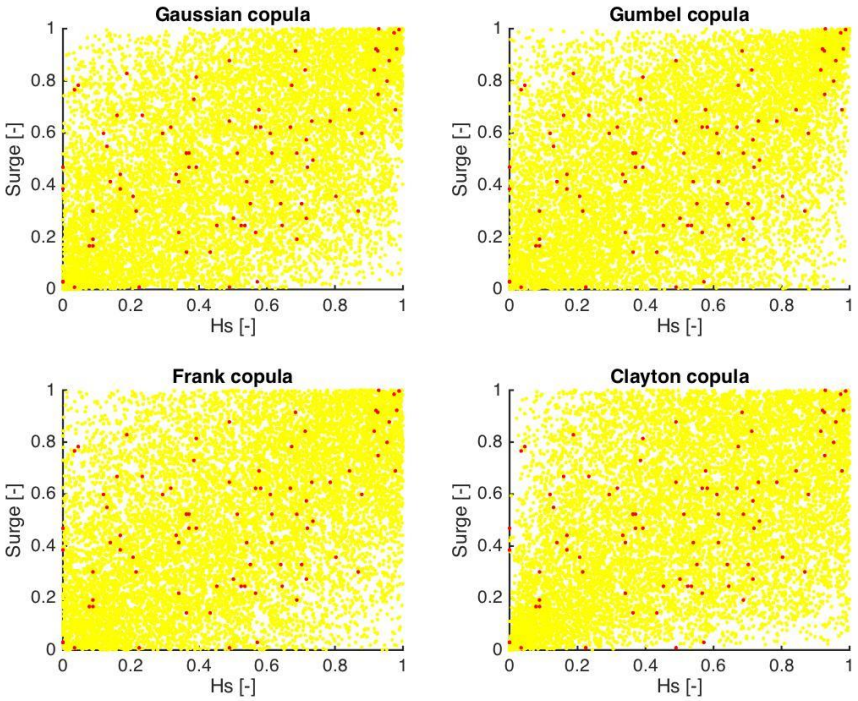

Figure C-8 Different copulas fit to the Hs and Surge dataset. The red dots indicate the original dataset and the yellow dots the data points generated with the copula.

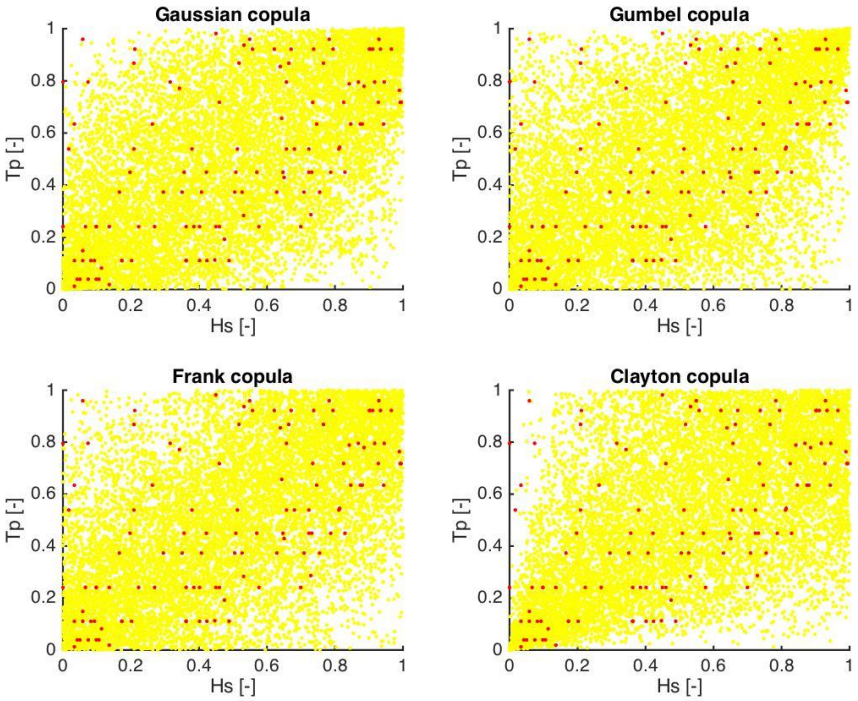

Figure C-9 Different copulas fit to the Hs and peak period dataset. The red dots indicate the original dataset and the yellow dots the data points generated with the copula.

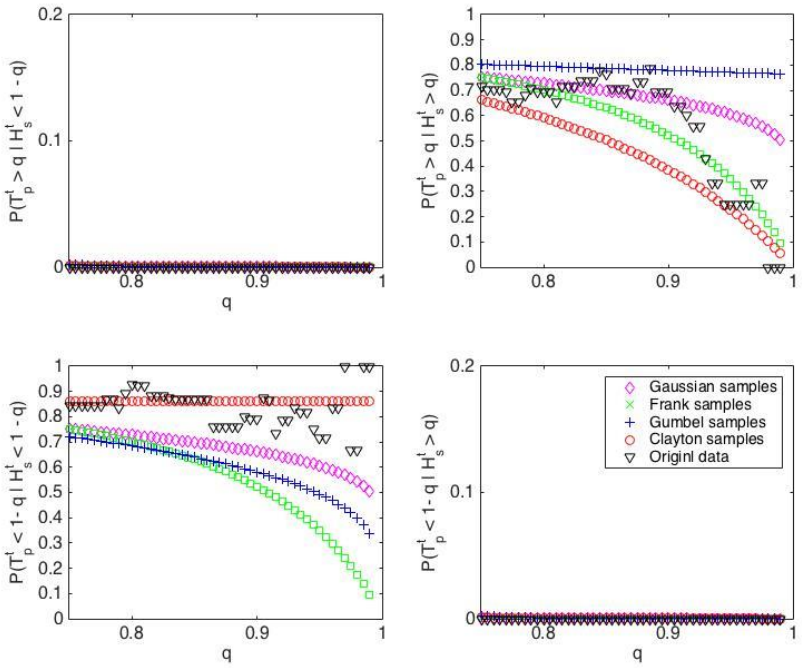

Figure C-10 Comparison of conditional probabilities for $D \mid H_s$ on the intervals $[0, 0.25]\times[0.75,1]$, $[0.75,1]^2$, $[0.75,1]^2$ and $[0.75,1]\times[0, 0.25]$ (top left, top right, bottom left and bottom right respectively).

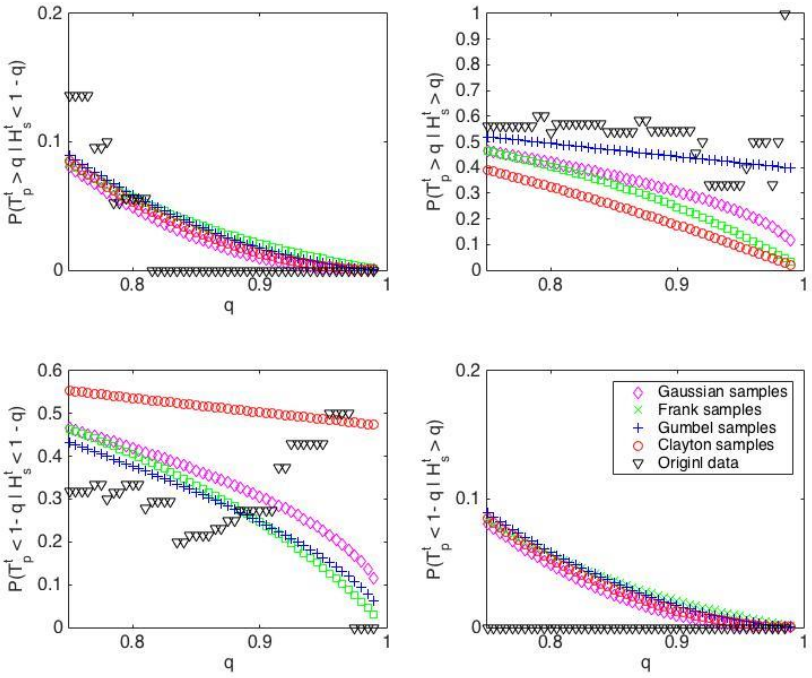

Figure C-11 Comparison of conditional probabilities for $S \mid H_s$ on the intervals $[0, 0.25]\times[0.75,1]$, $[0.75,1]^2$, $[0.75,1]^2$ and $[0.75,1]\times[0, 0.25]$ (top left, top right, bottom left and bottom right respectively).

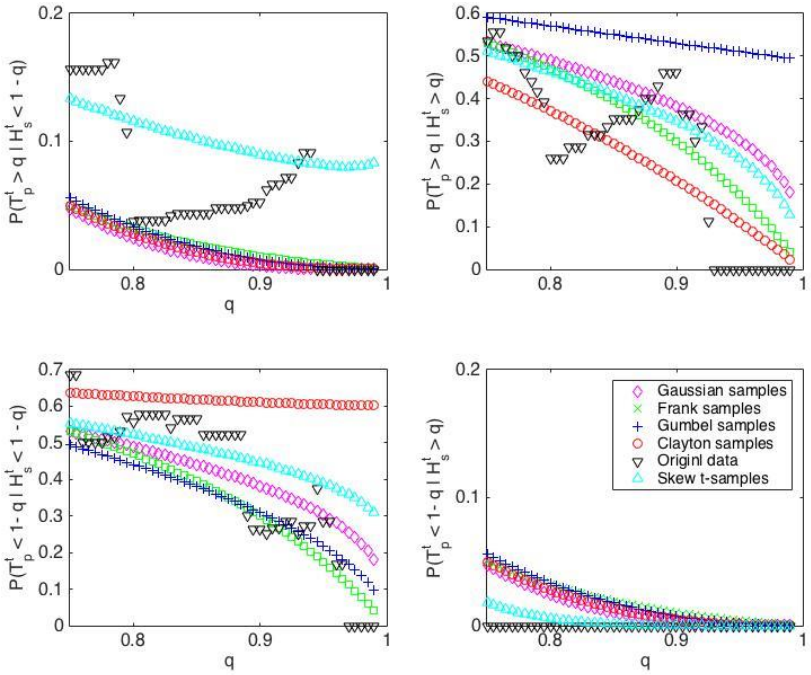

Figure C-12 Comparison of conditional probabilities for $T_p \mid H_s$ on the intervals $[0, 0.25]\times[0.75,1]$, $[0.75,1]^2$, $[0.75,1]^2$ and $[0.75,1]\times[0, 0.25]$ (top left, top right, bottom left and bottom right respectively).

D BAYESIAN NETWORKS

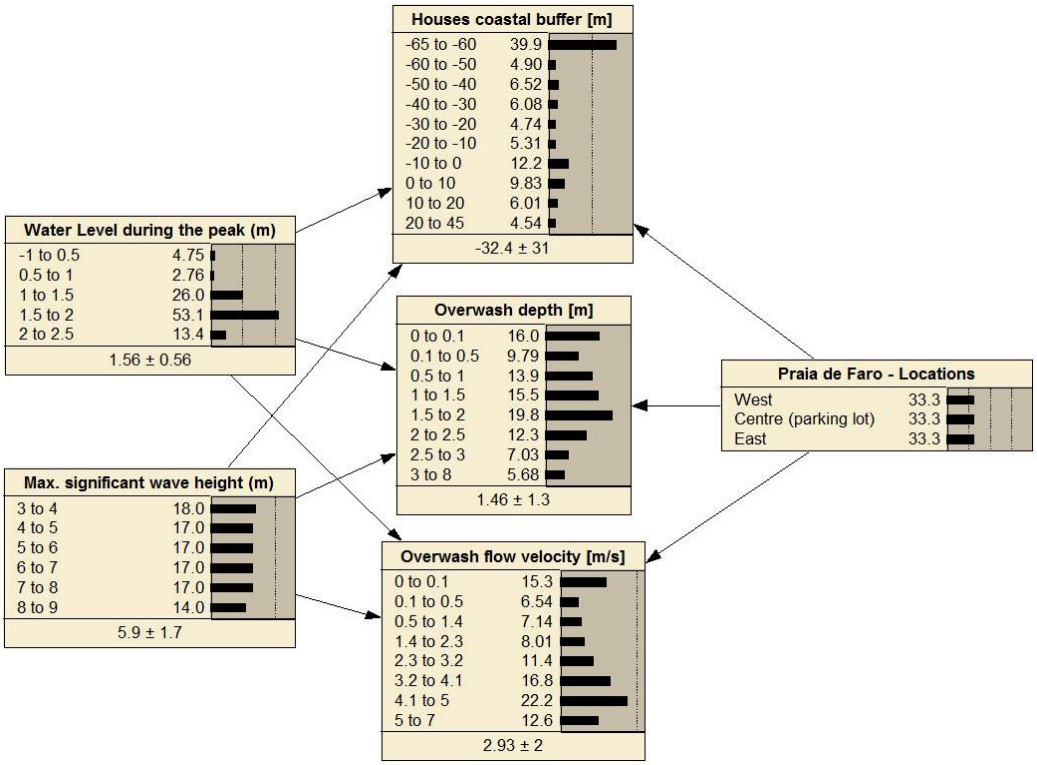

Figure D-1 Bayesian Network for Praia de Faro, trained with all cases for the peak of the storm duration of 15%.

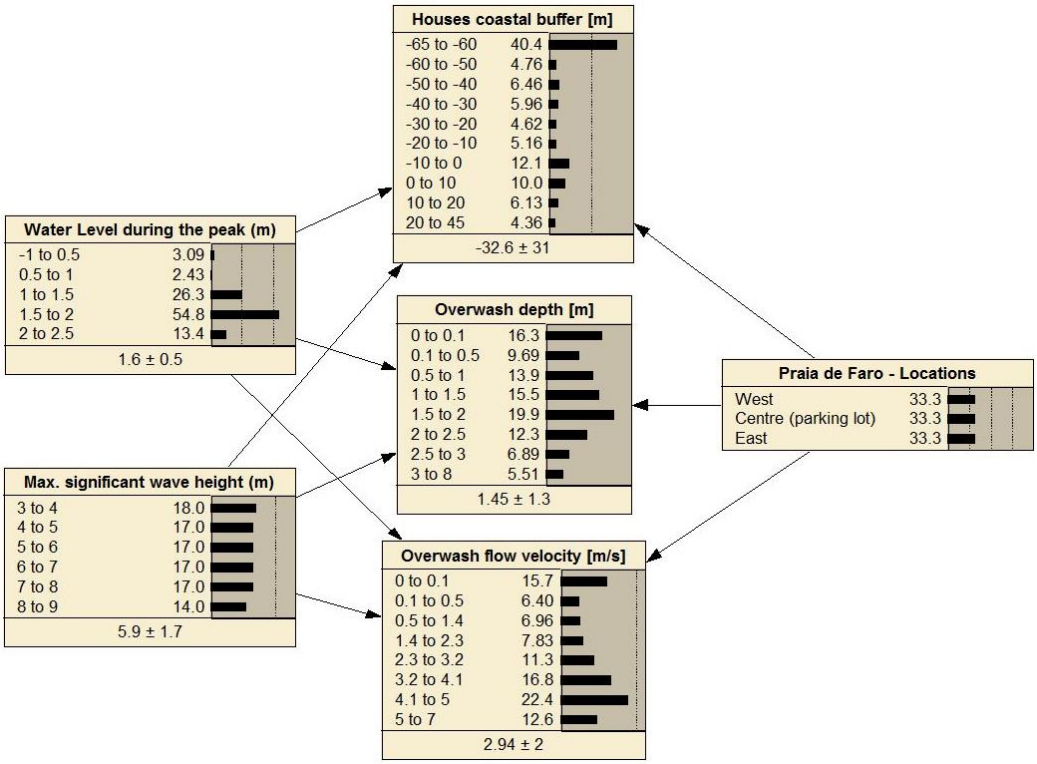

Figure D-2 Bayesian Network for Praia de Faro, trained with all cases for the peak of the storm duration of 20%.

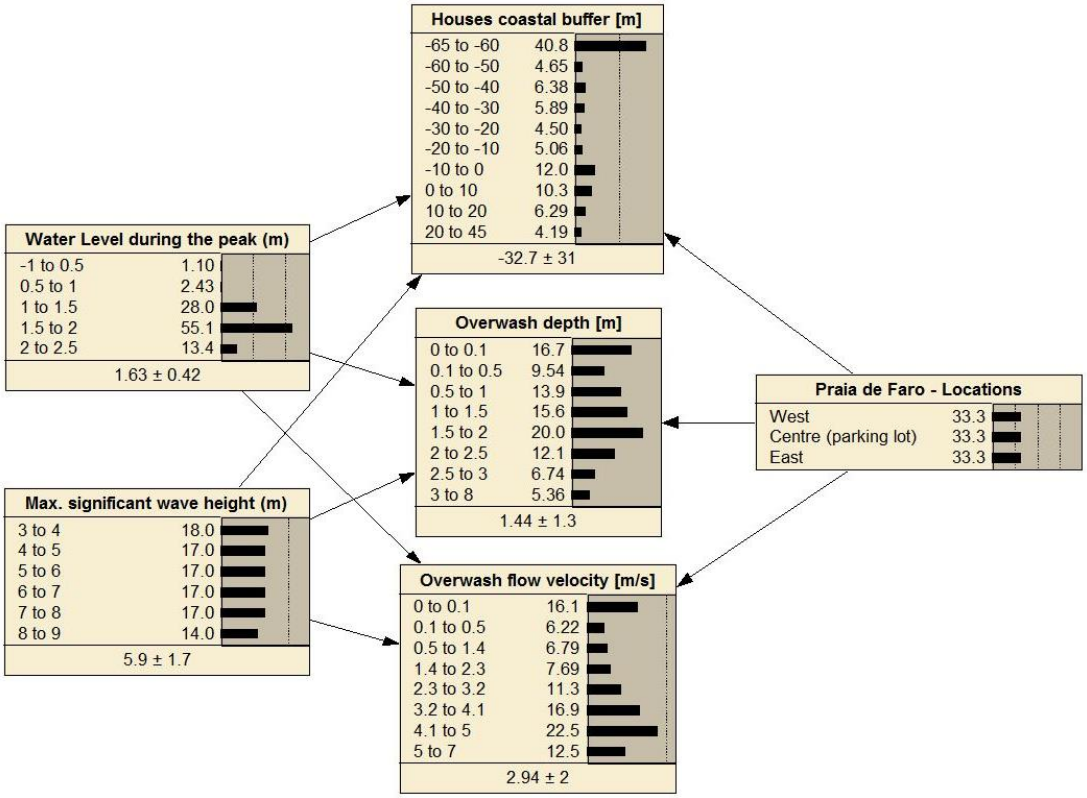

Figure D-3 Bayesian Network for Praia de Faro, trained with all cases for the peak of the storm duration of 25%.

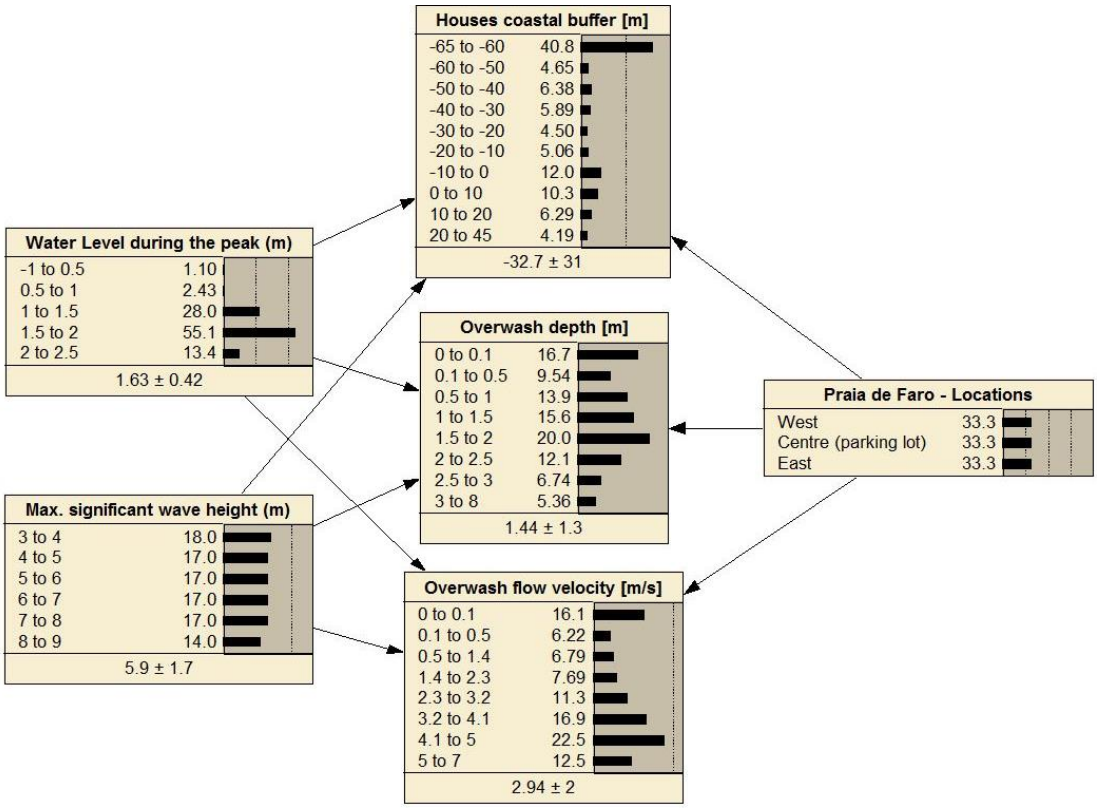

Figure D-4 Bayesian Network for Praia de Faro, trained with all cases for the peak of the storm duration of 30%.

E LIST OF FIGURES

F LIST OF TABLES

G LIST OF ABBREVIATIONS

2D	Two dimensional
3D	Three dimensional
BN	Bayesian Network
CDF	Cumulative density function
CRAF	Coastal Risk Assessment Framework
CvM	Cramer von Mises
DSS	Decision Support System
DST	Decision Support Tool
ECDF	Empirical cumulative density function
EU	European Union
EWS	Early Warning System
HFA	Hyogo Framework for Action
MAE	Mean Absolute Error
MSL	Mean sea level
NAO	North Atlantic Oscillation
PDF	Probability density function
PMF	Probability mass function
RISC-KIT	Resilience Increasing Strategies for Coasts tool KIT
RMSE	Root mean square error
SE	South East
SW	South West
UN	United Nations
UTM	Universal Transverse Mercator