Data poor environments

Uncertainty propagation in hydrodynamic modelling

Master Thesis

B.J.T. van der Spek

Graduation committee

prof. dr. ir. M.J.F Stive (TU Delft)

ir. I.L.L. Das (Boskalis)

ir. L. de Wit (Svašek/TU Delft)

dr. R.W.M.R.J. Ranansinghe (TU Delft)

dr. ir. P.H.A.J.M. van Gelder (TU Delft)

April 24, 2013

Delft University of Technology

Data poor environments

Uncertainty propagation in hydrodynamic modelling

Master Thesis
B.J.T. (Bart-Jan) van der Spek
1304224

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

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Hydraulic Engineering at the Delft University of Technology

April 24, 2013

Preface

This thesis constitutes the final stage of my master degree program hydraulic engineering at Delft University of Technology. Initially the subject of my master thesis was different. Boskalis offered me an interesting subject concerning an investigation of the coastal processes of a complex coastal system. The complexity of this system appealed to me and therefore I chose to write a proposal to start on this subject. Since a preliminary study was already performed by Svašek, this company was introduced as second party for my research. As the proposal was approved by the graduation committee a later joined member, Rosh Ranansinghe, noticed the concerning lack of sufficient data. With the available data I sure was not able to deliver a "rock solid" thesis. At that point I began, together with Boskalis, to think about redirecting my subject. The question that remained was how to get insight in the coastal processes while we are dealing with a data poor environment. I devised that stochastic modelling will probably be able to obviate the lack of sufficient data. In the end I think this subject is even more interesting than the initial subject and therefore I would like to thank Rosh for his vigilance. Hopefully the results of my research are useful for the coastal engineering practice and further research will be conducted to improve the applicability of my method.

I would like to use this opportunity to express my gratitude to the people who contributed to my master thesis and study. First of all I would like to thank Lynyrd de Wit and Ingrid Das for their excellent daily supervision. Lynyrd has helped me a lot with practical issues in order to realize the execution of the used models and methods. Not only his expertise and extensive knowledge about numerical modelling, also his reviews of my report has improved my thesis a lot. Ingrid has introduced me to the case study and thought along with me on redirecting the thesis subject. Her sharp comments on my report helped me improve the readability of my thesis. I thank Pieter van Gelder who evaluated my statistics and provide me with various ideas to execute my statistical analysis correctly. Furthermore I want to thank Rosh Ranansinghe for his adequate comments and helping me to get my conclusions sharper. At last I thank professor Marcel Stive who ensured I stayed on track and enhanced the focus of my research.

I also would like to express my gratitude to the people who have facilitated my thesis. Although my presence was little I would like to thank all the people at Hydronamic for the pleasant work environment and moral support, particularly Stefan Aarninkhof who supported me with the determination of my subject. I thank all the people at Svašek for the very pleasant days at the office, useful advise and moral support. Their expertise and incredible knowledge about modelling helped me a lot with the practical part of my thesis. Furthermore I would like to thank my friends, particularly my roommates; Joris for the lovely meals, Thijs for the framework and Steven for his expertise on math. I thank my family and especially my parents for their unconditionally support and trust in my study. And at last I want to thank my girlfriend, Lisa, for her moral support, unlimited patience and trust.

Bart-Jan van der Spek
Delft, April 2013

Abstract

The objective of this study is to propose a method that is able to deal with data poor environments within coastal studies, where detailed insight in coastal processes is desired. Numerical models are useful tools to get insight in the coastal processes of the system being modelled. Inaccurate or poor input data leads to inaccurate or even incorrect model results. By stochastic modelling insight is obtained in the uncertainty propagation through the model and will, in contrast to deterministic modelling, provide output distributions. Output distributions not only provide an estimation of the expected results but they also provide an estimation of the reliability of these results. Knowledge about the effects of input uncertainties on model results improves the interpretation of these results, which is valuable for decision makers. Information about the relative importance of different input variables and their relationships with the output helps the decision maker to focus on the most important model aspects. It gives an indication if and for which variable there is need for more (accurate) data.

A method for stochastic modelling is proposed and assessed for a case study. A coastline suffers from structural erosion and consists of a rocky reef lying oblique in front of the coastline. A preliminary study showed that this reef causes complex (wave driven) rip current patterns which results in poor knowledge about the coastal processes. A model execution with focus on the wave driven currents gives more insight in the governing coastal processes. However accurate information about the reef geometry and sufficient calibration and validation data is missing, which means we are dealing with a data poor environment.

After a detailed evaluation of possible methods for stochastic modelling it became clear that the Monte Carlo method is the most appropriate method for this study. The Monte Carlo method is very robust and is relatively easy to implement. A random sample is generated that corresponds with the determined probability distributions of the used stochastic input variables. This sample defines many scenarios which are all executed with the (deterministic) model which results in a large data set of output values, which in turn describes the output distribution.

To generate a random sample the uncertainties of the used input and model parameters need to be quantified in probability distributions. The quantification is divided in two parts: uncertainties of the input variables and model parameters. The following forcing variables are used as stochastic boundary conditions: wave height, wave direction, wave period, wind speed and wind direction. The uncertainties of these variables stem from the variability and randomness of nature. The uncertainties are quantified by describing the total wave and wind climate in six separate conditions, each with separate probability distribution parameters while taking correlation between the wave and wind variables into account. The uncertainty related to the reef geometry is taken into account by using the unknown reef elevation as stochastic variable. The lack of data leads to a subjective but substantiated determination of the probability distribution parameters. Roughness height is the only model parameter that is used stochastically. This is a commonly used calibra-

tion parameter and therefore a stochastic roughness height partly obviates the missing calibration and validation data. As for the reef elevation appropriate values for the probability distribution parameters for the roughness height are chosen.

The Monte Carlo procedure is executed for one condition. The obtained output distributions are used to determine the total uncertainty propagation. The results show that it is possible to provide an indication of the expected results. The reliability is, as expected, lower for the sediment fluxes than for the hydrodynamics. The obtained data set of output values is used for a comprehensive sensitivity analysis. By examining scatter plots and correlation coefficients insight in the relationships and relative importance between the input and output variables is obtained. Some unexpected relations are found. The reef causes complexity and leads to different variable importance for (spatially) different output variables. The reef causes a set-up behind the reef which in turn causes a (southward) alongshore current and sediment flux behind the reef, including several rip currents. At the southern part of the reef a gap is situated where the most dominant rip current and sediment fluxes are found. It seems that the stochastic reef elevation has more importance for the longshore sediment flux behind the reef than for the cross shore sediment flux through the gap of the reef. More surprisingly, the results show that for the total sediment balance of the system the wind speed and wave period are the most important variables and therefore even more important than the imposed (offshore) wave height.

The found relationships are used to describe a conceptual model of the system and show that the system and governing processes are better understood. A regression analysis is used to quantify the relative importance of the used variables. It gives insight in the direct effect of the variable on the model output. Correlation between input variables makes it more complex but regression analysis is able to determine the direct and indirect effect of correlated variables on the model output.

The proposed method as described by this study gives rise to many applications and opportunities. It has shown an indication of the expected results can be given including an estimation of the reliability of the results. The output distribution provides probabilities of occurrence which can be used for risk analyses. Furthermore the system is better understood and insight in complex coastal processes is obtained. Uncertainty analyses are able to provide the variable importance of the system being modelled. The results of a case study showed some unexpected variable importance of the system processes. Knowledge about the variable importance is valuable for coastal engineers and decision makers as it enables to focus on the most important model aspects and variables. By imposing stochastic model parameters the need for calibration can be determined. Furthermore a conceptual method has shown that it is possible to determine the (financial) value of obtaining more (accurate) data. This can lead to economic efficiency in coastal engineering practice, as it determines the necessity of data collection and calibration.

Given the applicability of the proposed method it is recommended that comprehensive uncertainty analyses like this become a more common approach within coastal studies, regardless of dealing with a data poor environment.

Contents

Preface iii

Abstract iv

Contents viii

1 Introduction 1
- 1.1 Problem description . . . 1
- 1.2 Objective . . . 2
- 1.3 Sub objectives . . . 2
- 1.4 Methodology . . . 2
- 1.5 Outline of report . . . 3

2 Case study 4
- 2.1 Area description . . . 4
  - 2.1.1 Bathymetry . . . 4
  - 2.1.2 Climate . . . 7
- 2.2 Coastal system . . . 7
- 2.3 Implementation . . . 7

3 Uncertainties 8
- 3.1 Classification of uncertainties . . . 8
- 3.2 Quantification of uncertainties . . . 10
  - 3.2.1 Uncertainties of input variables . . . 10
  - 3.2.2 Uncertainties of model parameters . . . 17

4 Methods for stochastic modelling 19
- 4.1 Characteristics of research . . . 19
- 4.2 Criteria . . . 20
- 4.3 Monte Carlo methods . . . 20
  - 4.3.1 Latin Hypercube sampling . . . 21
  - 4.3.2 Importance sampling . . . 22
  - 4.3.3 Generating sample of correlated variables . . . 22
- 4.4 Analytical methods . . . 23
- 4.5 Response surface method . . . 24
- 4.6 Conclusions . . . 25

5 Execution of the proposed method **26**
5.1 Execution process . . . 26
5.2 Numerical model . . . 28
5.2.1 Expected sensitivity . . . 28
5.2.2 Validation . . . 33
5.2.3 Assumptions and limitations . . . 35

6 Results **39**
6.1 Input and focus . . . 39
6.2 Uncertainty propagation . . . 40
6.2.1 Hydrodynamics . . . 40
6.2.2 Sediment fluxes . . . 44
6.2.3 Total sediment balance . . . 50
6.2.4 Statistical uncertainty . . . 52
6.3 Uncertainty importance . . . 55
6.3.1 Scatter plots and (rank) correlation . . . 55
6.3.2 System schematisation . . . 63
6.3.3 Regression . . . 76

7 Applications **83**
7.1 Interpretation . . . 83
7.2 Total wave climate . . . 83
7.3 Reliability . . . 84
7.4 Calibration . . . 84
7.5 Processes . . . 84
7.6 Risk analysis . . . 85
7.7 Value of information . . . 85

8 Conclusions **90**
8.1 Proposed method . . . 90
8.2 Case study . . . 91

9 Recommendations **93**

10 Discussion **95**

References **98**

Appendices **99**

A Statistics **100**
A.1 Estimators . . . 100
A.2 Correlation . . . 101
A.3 Bootstrapping . . . 102
A.4 Maximum likelihood . . . 102
A.5 Kernel density estimation . . . 102
A.6 Regression . . . 103
A.6.1 Coefficient of determination . . . 103
A.6.2 Standard error . . . 104
A.6.3 P-value . . . 105

B Numerical models **106**
B.1 SWAN 106
B.2 FINEL 2D 107
B.2.1 Unstructured grids 107
B.2.2 Hydrodynamics 107
B.2.3 Morphodynamics 108

C Estimates **112**

D Correlations **120**

Chapter 1

Introduction

In different stages of project development, coastal modelling studies are executed to gain more understanding in the coastal processes. Numerical models are a useful tool to get, relatively fast and economically, insight in the important processes of the coastal area that is studied. The study requires more detail in a more advanced stage of the project development. Therefore the amount, reliability and accuracy of the input data usually becomes higher towards a later stage. However, the collected data may be insufficient to elaborate a detailed coastal study and in addition, obtaining data can be very expensive and time consuming. In other words we may have to cope with a *data poor environment* while there is need for reliable insight in processes with as little expenses and time as possible. Poor or inaccurate data leads to uncertainties for the used model input and obviously results in inaccurate and unreliable model output. A model can be schematised with the following equation;

$$\mathbf{Y} = h(\mathbf{X}) \tag{1.1}$$

where $\mathbf{X} = [\mathbf{I}, \boldsymbol{\theta}]$ and consists of $\mathbf{I} = [I_1, I_2, \ldots, I_{nI}]$, a vector of model input variables and $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{n\theta}]$, a model parameter vector. Function h represents the model and $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_{nY}]$ is a vector of model predictions. Assuming the model is accurate, model output $\mathbf{Y}$ represents what the actual output would be if an experiment were performed under the conditions $\mathbf{I}$ and $\boldsymbol{\theta}$. So the approximation of the model depends on one set of input and model parameters; *deterministic approach*. Insight in the uncertainties that propagate through the model leads to insight in the reliability of model prediction $\mathbf{Y}$. Insight in the reliability of the model predictions can be very useful for decision strategies, since it provides a more improved interpretation of the model results. Stochastic output not only provides estimates for the expected results, obtained probability distributions provides opportunities for risk analyses. An uncertainty analysis is able to provide a comprehensive sensitivity analysis which means it is possible to distinguish important variables and parameters. This information enables decision makers to focus on the most important model aspects and variables and predict whether it will be useful to obtain more (accurate) data. Stochastic relationships are able to provide insight in the physical processes and therefore the coastal system is better understood, despite the lack of (inaccurate) data.

1.1 Problem description

Lack of (inaccurate) data leads to unreliable or even incorrect model predictions. Though reliable and accurate insight in the coastal processes is desired. A method that is able to provide stochastic model output not only gives an estimation of the expected results. More important; stochastic

modelling gives insight in the reliability of model results stemmed from the given uncertainties of the input and model parameters.

1.2 Objective

Develop and apply a generic method to cope with data poor environments for hydrodynamic coastal studies.

1.3 Sub objectives

1. Investigate possible probabilistic methods to perform stochastic analysis for hydrodynamic modelling, using literature.
2. Set up a 2DH hydrodynamic model of the study area (see chapter 2). This coastline consist of rocky reef lying oblique to the coastline which induces complex rip current patterns. Insight in the hydrodynamic and morphodynamic processes is desired.
3. Qualify and quantify the uncertainties of the input parameters; forcing and geometry.
   (a) Perform a statistical analysis of the forcing variables; wave and wind climate.
   (b) Quantify uncertainties of unknown geometry parameters.
4. Qualify and quantify uncertainties of unknown model parameters.
5. Couple the proposed method for stochastic modelling with the 2DH hydrodynamic (stationary) model and execute the proposed method with use of the determined stochastic input and model parameters.
6. Evaluate model results of the study area; focus on the uncertainty propagation and uncertainty importance.
7. Evaluate total study and model results with focus on the applicability of the proposed method.

1.4 Methodology

1. By evaluating different existing probabilistic methods from literature the most appropriate method is determined. Characteristics of the different methods for stochastic modelling are compared with the characteristics of this research.
2. A 2DH hydrodynamic and morphodynamic model is set up for the study area. The model consist of two modules, the wave module and flow module. A simple sensitivity analysis, both analytical and numerical, gives insight in the expected relationships and shows whether the model is capable of predicting the processes in a correct way. This crude qualitative validation is extended with a quantitative validation by comparison with laboratory tests performed by Seelig (1983).
3. The quantification of the uncertainties of the input variables is done by means of
   (a) a statistical analysis of the forcing is performed in order to quantify the stochastic relationships of the forcing variables.

   (b) analysis of the available data in order to determine a substantiated subjective quantification of the uncertainties.

4. The subjective quantification of the uncertainties of the model parameters is substantiated by reference projects.

5. The proposed probabilistic method is set up for the hydrodynamic model and executed using the quantified uncertainties. The execution of the method is bounded by calculating sediment fluxes using *stationary* hydrodynamic results; wave driven currents. The study will be focused on stationary processes and therefore morphology, i.e. bottom update, will not be part of this study.

6. Evaluation of the model results gives insight in the total uncertainty propagation and therefore the reliability of the model results and the reliability of the proposed method. Further statistical analysis is able to provide insight in the uncertainty importance. A regression analysis quantifies the relative importance and is able to gives insight in the effects of the used variables. The sediment transport capacities are investigated in order to gain insight in the coastal processes of the system being modelled.

7. The total study with the obtained model results provides insight in the applicability of the proposed method. It gives rise to opportunities and applications for hydrodynamic modelling and decision strategies.

1.5 Outline of report

First a description of the case study is provided in chapter 2, with the implementation into the objective of this study. In chapter 3 the uncertainties of the input and model parameters are classified and quantified. In chapter 4 different methods for stochastic modelling are compared and their applicability for this study is examined. Chapter 5 treats the execution of the proposed method as chosen in chapter 4. And so it treats the coupling of the method for stochastic modelling with the numerical (hydrodynamic) model. Furthermore a sensitivity analysis and crude validation are provided in this chapter. A description of the used numerical model(s) with important assumptions and limitations are also treated in chapter 5. Chapter 6 provides an analysis of the results of the execution of the method. Applications and opportunities stemmed from the study are described in chapter 7 and the most important conclusions are treated in chapter 8. At last some recommendations and a discussion is given in chapter 9 and 10, respectively.

Chapter 2

Case study

The coastline of interest consist of a rocky reef lying oblique in front of the coast, which in many cases can give protection against erosion because of the dissipation of wave energy (Calabrese et al., 2008). However a preliminary study showed that due to the presence of the reef some complex hydrodynamic processes are important for the cause of erosion along this coastline (de Wit, 2012). Breaking of waves on the reef and spatially varying reef elevation leads to complex rip current patterns.

2.1 Area description

The study area reaches from the southern breakwater of the port 12 km south to a small headland, see figure 2.1. The southern breakwater of the port has a length of approximately 5 km. Most parts of the coastline show a very narrow beach or no beach at all. Small houses and buildings are located near the shore. A railway parallel to the shoreline is protected with a small hard defense which mostly contains rock.

2.1.1 Bathymetry

The bathymetry is mainly characterized by the reef. The rocky reef is oblique to the coast starting northwards at the shoreline going more offshore towards the south at the maximum of about 250 meters offshore. Close to a small headland the reef is located further offshore causing a gap between the two reefs. The reef elevation varies spatially. For the preliminary study that has been executed surveying was necessary, however shallow water and strong breaking waves on top of the reef made it difficult to survey (de Wit, 2012). Therefore there is a lack of data and locally the bathymetry and reef elevation are unknown. Figure 2.2 shows the surveyed bathymetry. The plot shows the data interpolated, so in reality the reef elevation may be different.

Some parts of the reef appear above the water surface, which means the reef elevation varies spatially. According to the preliminary study it seems that the higher part of the reef is closer to the shore than the lower part. To get insight in the sensitivity of the reef elevation several reef elevations had been simulated in the preliminary study (de Wit, 2012). From the preliminary study of de Wit (2012) it seemed that the results are not too sensitive for the reef elevation. Soil samples showed that the study area contains sand which is rather coarse with a grain size of approximately $d_{50} \approx 400\mu m$ (figure B.3).

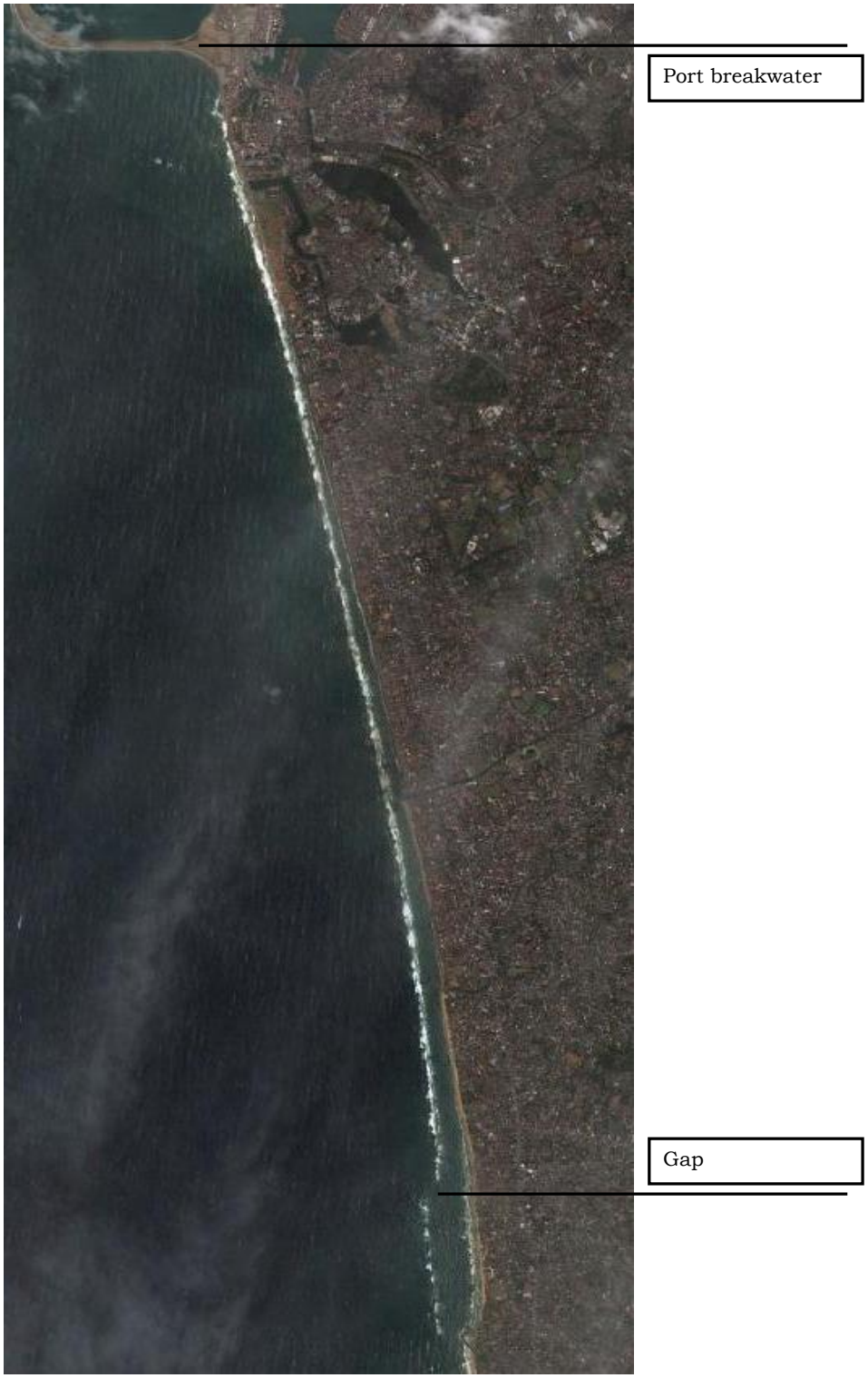

Figure 2.1: Area of interest

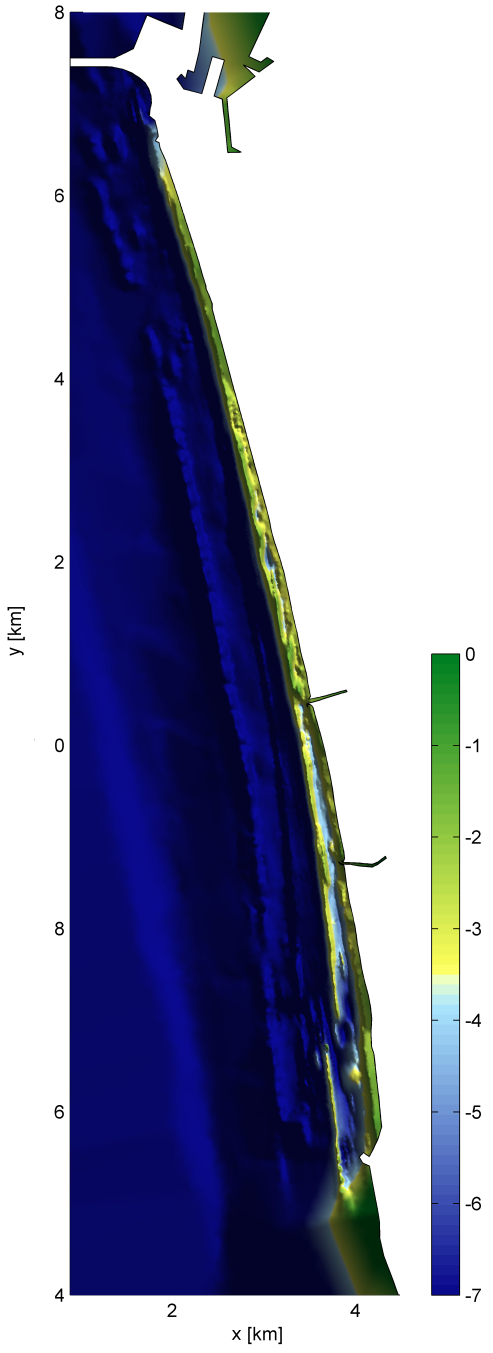

Figure 2.2: Surveyed bathymetry; bottom level with respect to MSL (Note that the plot shows interpolated data, in reality the reef elevation may be different

2.1.2 Climate

The offshore wave climate contains both swell and sea waves (de Wit, 2012). Sea waves are mainly coming from the same direction as the dominated wind, being the west. The dominating swell is coming from south-southwest. There is also a seasonal variation in wave pattern due to wind variability in the monsoon period. Both sea and swell waves can have a significant wave height with a maximum of about 3.5 meter. Along the coast the waves induce a nett alongshore current and sediment transport northwards. The tidal envelope is relatively small with a maximum of approximately 0.80 m.

2.2 Coastal system

The coastline of this case study suffers from structural erosion. This coastline consist of a rocky reef lying in front the sandy coast. In many cases a submerged structure can protect the beach from erosion, however this specific coast is not stable. An executed preliminary study shows that the near shore current patterns along the coast are complex and more processes, beside the nett alongshore current, are causing erosion. Due to breaking of waves on top of the reef and varying reef elevation complex rip currents patterns can be found. The presence of the reef introduces complexity and lack of insight in the processes that causes erosion. In order to prevent problems in the future the possibilities of interventions that will diminish erosion need to be investigated. A further detailed study can help to understand the relevant processes better and help to devise possible mitigation measures. Simulating the processes in the area of interest using a numerical model can give more insight. However there is a lack of data; poor knowledge about the reef elevation and missing validation and calibration data, i.e. flow velocity measurements. Therefore direct calibration and validation of the model is not possible. We are dealing with a *data poor environment*. This means that a detailed study may be not accurate and therefore give unreliable results. A detailed study which can give insight in the hydrodynamic and morphodynamic processes is desired while sufficient and accurate data is not available.

2.3 Implementation

This case study is suitable to test the applicability of the generic method that is studied. By performing an analysis of the uncertainty propagation of this specific case an indication of the expected value, upper and lower bounds of the results can be given. In this case the sediment transport capacities are of most interest in order to obtain insight in the important processes causing erosion. Beside an indication of the most frequent processes an approximation of the reliability of the results can be obtained. The forcing variables and reef characteristics are for this study used as stochastic input variables. Roughness of the reef, a model parameter and a typical calibration parameter is also a stochastic variable for this study (chapter 3).

Chapter 3

Uncertainties

This chapter discusses the main issue of this research. Uncertainties of the input and model parameters are the generator for the lack of reliability of the model output. In the following sections uncertainties that will manifest in this study are mentioned and quantified.

3.1 Classification of uncertainties

The first classification of the uncertainties is based on their source. The main classification is made between uncertainties that are generated by the variability of variables, *inherent* uncertainty, and uncertainties due to the lack of knowledge, *epistemic* uncertainty.

Inherent uncertainty represents the variability and randomness of nature. Even sufficient and long history of data does not provide accurate forecasts for the future. Reduction of this type of uncertainty is not possible. It can be subdivided in time and space.

Epistemic uncertainty stems from lack of knowledge. For example from lack of insight in physical processes or lack of sufficient data. This type of uncertainty can be reduced by an increase of knowledge.

Examples from inherent uncertainties in time are the variability of the forcing variables, like wave height. Unlimited data will not reduce this uncertainty, however uncertainty is generated by insufficient record length.
A common way to express, or *quantify*, uncertainty is the use of (subjective) probability distributions. For the determination of these distributions statistical methods are available. However determination of these parameters generates uncertainty as well. Uncertainty of parameters and distribution type, i.e. uncertainty by the quantification of the uncertainty, is known as *statistical uncertainty* (van Gelder, 2000). Which is an example of epistemic uncertainty. For this research statistical uncertainties are assumed to be very small and therefore neglected.
Classification based on their source can be useful during the quantification of the uncertainties, see section 3.2. However for assessing different methods for stochastic modelling (chapter 4), another classification is desired; based on the manifestation of the uncertainties in the particular model (van der Klis, 2003). The used model for this research is a hydrodynamic (and wave) model (section 5.2) and therefore distinction is made in the relation between uncertainties and model characteristics. According to van der Klis (2003) the following classification at which uncertainties manifest themselves in computer simulation models can be made;

Technical uncertainties are uncertainties in model quantities; input variables and model parameters. Related to both inherent and epistemic uncertainties.

Methodological uncertainties are uncertainties related to the assumptions of the model structure and model equations. This type uncertainties are mainly epistemic uncertainties.

Epistemological uncertainties are uncertainties that are related to the representation of the simulated system by the model. It refers to the completeness and validity of the model. Related to both inherent and epistemic uncertainties.

Model operation uncertainties are uncertainties caused by the numerical and implementation errors.

In this study the focus shall be on *technical uncertainties*. It may be obvious that the other types of uncertainties are relatively difficult to quantify, since they are not uncertainties in (model) *quantities*. Since the focus is only on technical uncertainties, the assumption is made that other types of uncertainties are small compared to technical uncertainties and therefore neglected.
As mentioned technical uncertainties are uncertainties in model quantities. To get more insight of this type of uncertainty it is important to know what model quantities can be defined. According to Morgan and Henrion (1990) distinction can be made between the following different types of model quantities.

Empirical quantities are quantities that are measurable properties from the real world system being modelled.

Defined constants are the quantities that are certain by definition. Examples are mathematical constants like the number π. The uncertainty is always zero for this type of quantity. Physical constants, like the gravitational constant are *not* defined constants but empirical quantities, since they can be measured and are therefore inherently uncertain.

Decision variables are quantities that are controlled by the user of the model, i.e. decision maker. Examples in this case may be variables that are stemmed from designed interventions in the system that are being investigated and modelled.

Value parameters represent aspects that are stemmed from the preferences of the user or decision makers. Example in this is maximum erosion rate; a quantity that is based on value judgment.

Index variables are used to identify a location or cell in the spatial or temporal domain of a model. It makes no sense to be uncertain about this value.

Model domain parameters are quantities that specify the model domain. In this case they are mainly related to the spatial and temporal grid that is used in the model.

Morgan and Henrion (1990) suggests that empirical quantities are from the *only* type that is both uncertain and can have a *true* value; instead of *appropiate* or *good* values. Therefore it is suggested that uncertainties of empirical quantities are the only uncertainties that can be quantified by a stochastic relationship, i.e. probability distributions. In this research the focus shall therefore be only on *technical uncertainties* of *empirical quantities*.

Conclusions

- This research is limited to propagation of *technical uncertainties* of *empirical quantities*
- *Statistical uncertainty* is assumed to be relatively small compared to technical uncertainties and are therefore neglected.

3.2 Quantification of uncertainties

The quantification of uncertainties is divided in two sections regarding the uncertainties of input variables and the uncertainties of model parameters, respectively.

3.2.1 Uncertainties of input variables

The input variables are in the case of hydrodynamic modelling generally *forcing variables*, imposed as boundary conditions, and the *bathymetry* which is in fact (in combination with a surface elevation) the imposed initial condition.

Forcing variables

The forcing variables force the system being modelled and are in principle the only temporal varying variables. The following forcing variables are important for this case study;

- Wave height H_s [m]
- Wave direction φ [degrees]
- Wave period T_p [s]
- Wind speed U [m/s]
- Wind direction ψ [degrees]
- Water level η [m + MSL]

These are all quantities that are subject to variability and randomness of nature and therefore their uncertainties are, as mentioned in section 3.1, inherent uncertainties. However due to the limitation of the data size and unknown reliability epistemic uncertainties are generated as well. Since these uncertainties are not known and not measurable, they are not taken into account. The quantification of uncertainty of these forcing variables are all done with use of the statistical methods provided in appendix A.

Source

The data that are used for the quantification of the forcing variables are taken from the *National Centers for Environmental Prediction* (NCEP) which is part of the *National Weather Service* (NWS) which in turn is part of the *National National Oceanic and Atmospheric Administration* (NOAA). The global wave model WAVEWATCH III® is used for hindcasting wave conditions (Tolman, 2009). NCEP provides hindcast wave and wind data over a period of 12 years with hindcasts every 3 hours. The location of the hindcast is far offshore of the coast and are deep water conditions. Despite the fact that these are model predictions it is assumed that these hindcasts represent the real wave and wind climate correctly; uncertainties generated by this are not taken into account.

Wave climate

In figure 3.1 the offshore (deep water) wave climate near the area of interest is shown. The created plot actually shows a discontinuous bivariate probability density distribution of the wave height and wave direction. Since the probability density is multi modal it is not easy to quantify the uncertainty of the wave height and wave direction as one probability density function. Therefore the

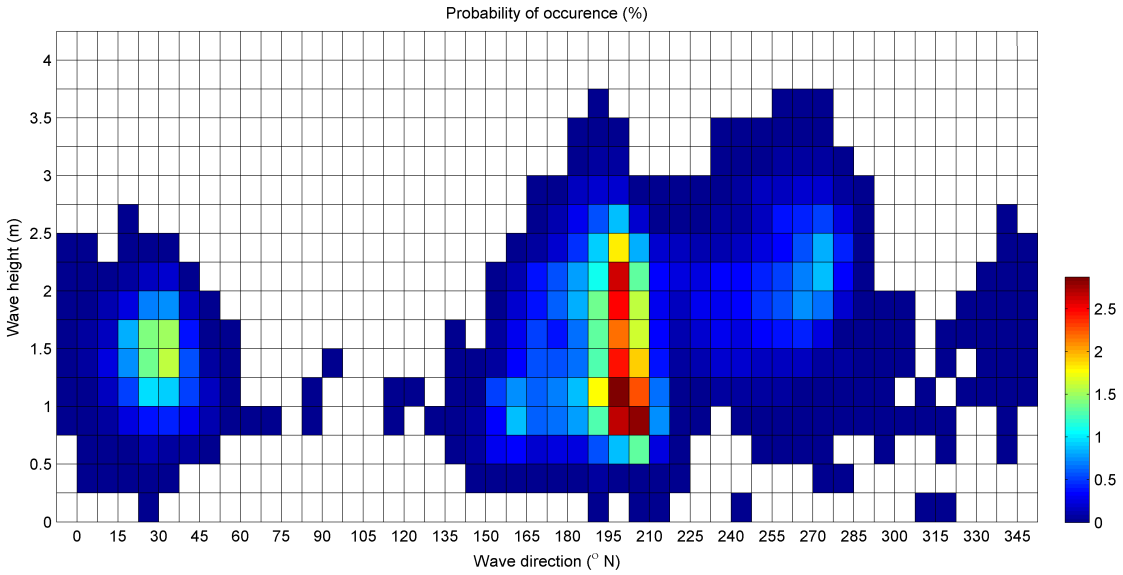

Figure 3.1: Offshore (deep water) wave climate *source: NOAA/NWS/NCEP*

total wave climate is described in six weighted multivariate probability density functions, which are all Gaussian distributed. In figure 3.2 all data points are shown with the corresponding estimated probability density functions. In this way the wave climate is represented in six (weighted) wave conditions with a mean (μ_w) standard deviation (σ_w) and weight(w_w). In order to represent the wave climate in a correct way (non physical) correlation is taken into account; for example shown in figure 3.2 where the most right contour lines are slightly oblique. The probability density functions of the wave conditions are described as follows;

$$\mathbf{D}_w = f(\boldsymbol{\mu}_w, \mathrm{Cov}_w) \quad \text{with} \tag{3.1}$$

$$\boldsymbol{\mu}_w = [\mu_{H_s} \quad \mu_\varphi]_w \qquad w = 1, 2, \ldots, 6 \tag{3.2}$$

$$\mathrm{Cov}_w = \begin{bmatrix} \sigma^2_{H_s} & \mathrm{Cov}(H_s, \varphi) \\ \mathrm{Cov}(\varphi, H_s) & \sigma^2_\varphi \end{bmatrix}_w \tag{3.3}$$

The total multi modal probability density function of the total wave climate then is;

$$D_{total} = \sum_{w=1}^{6} w_w \mathbf{D}_w \tag{3.4}$$

To test the reliability of this method the total probability density function (D_{total}) is integrated over bins (discontinuous) and compared to the (raw) wave data. As can be seen in figure 3.3 the wave climate is represented quite well with the generated probability density function. The probability distribution parameters are estimated using the statistical methods provided in section A.1. In appendix C the estimated probability functions are shown for every wave condition seperately. Clearly not every data set seems to fit perfectly with the estimated distribution, particularly for the extreme percentiles. The estimated functions however fit well for probabilities between 0.05 and 0.95, as shown in the probability plots on the right side in appendix C. The

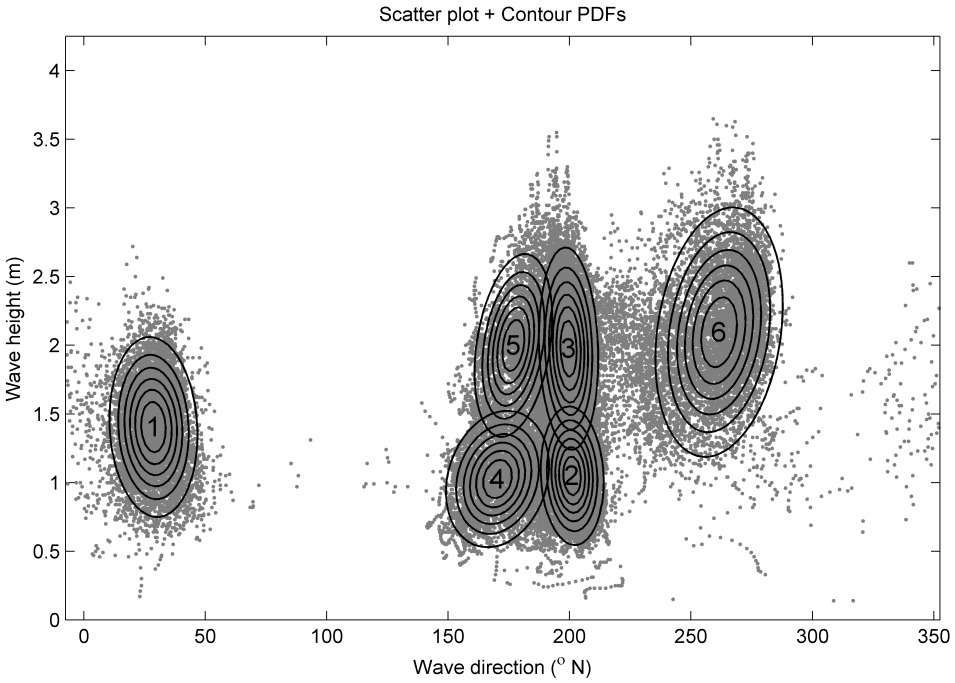

Figure 3.2: Raw data measurements and six weighted probability density functions

focus of the study is to get insight in the uncertainty propagation, so we are not interested in the tails of the distributions.
For some wave conditions, distributions are estimated from non-homogenuous data sets, because some data sets are cut off. In order to represent the total wave climate as correctly as possible, some distributions may overlap to obtain a multimodal distribution. This is done by cutting of data sets in order to allow the tails of the estimated distribution functions to overlap. The (weighted) sum of the estimated overlapping functions show then a multimodal distribution function. Although this method is very crude and straight forward, figure 3.3 proofs that the method is capable of (roughly) representing the climate in a correct way.

Wave period

For the wave period distinction is made between swell and sea waves. The scatter plot in figure 3.5 shows a clear division that can be separated by a line of constant steepness of 0.02. For sea waves there is a clear relationship between wave period and wave height which is described by constant steepness. The steepness of the sea waves seems to fit well with a Gaussian distribution (appendix C, figure C.7. Therefore for sea wave conditions, the uncertainty of the wave period is quantified with use of the distribution for steepness. For swell waves no clear relationship between wave period and wave height is found and there the wave period can be fit with a Gaussian distribution (appendix C, figure C.7). So for swell conditions the uncertainty is quantified direclty from the distribution of the wave periods. Condition 1 and 6 are both sea wave conditions while the other conditions are typical swell waves.

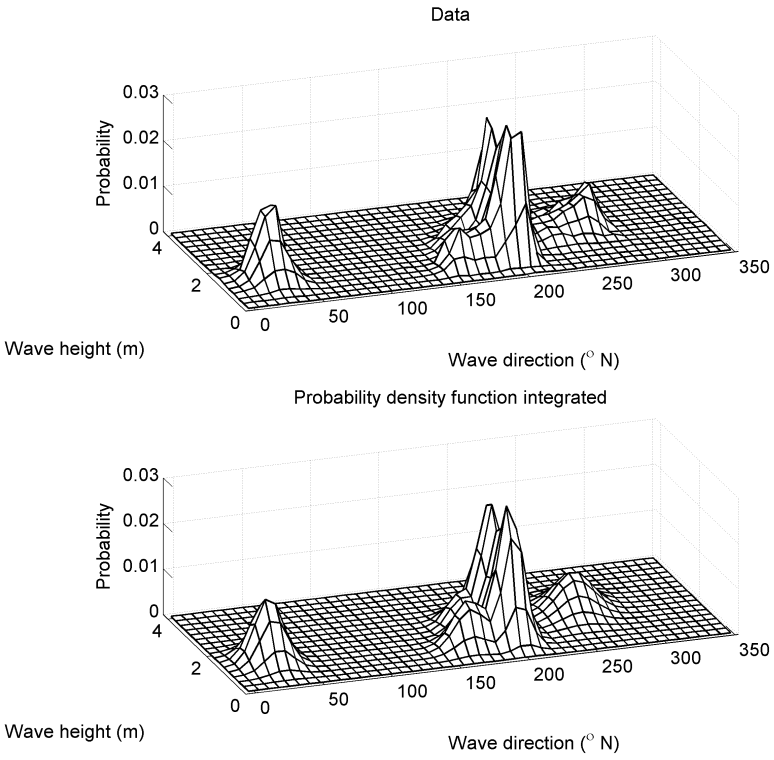

Figure 3.3: Total probability density function integrated over the bins compared to wave climate

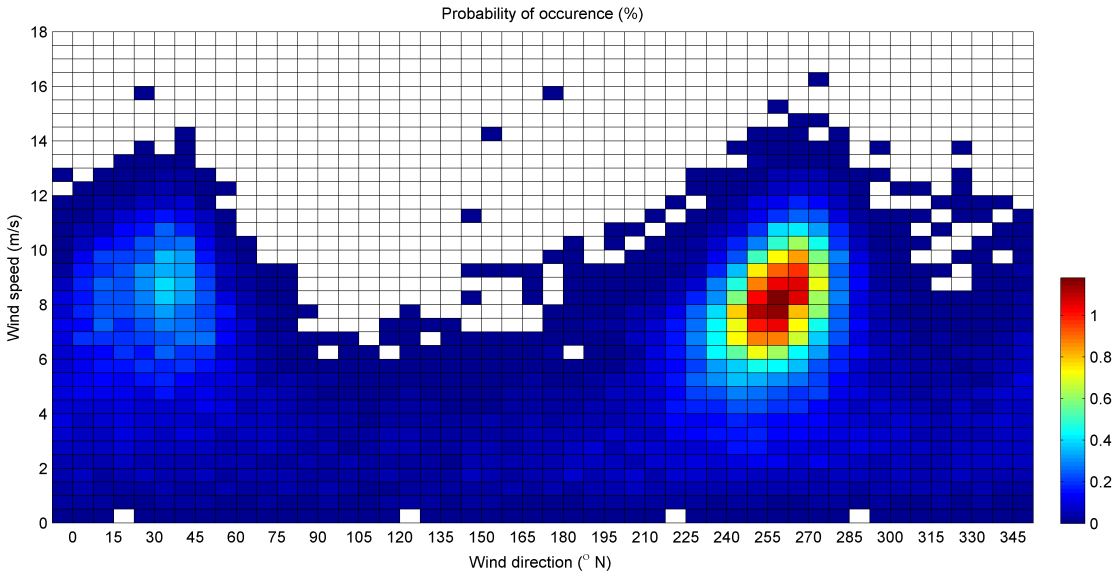

Figure 3.4: Offshore wind climate *source: NOAA/NWS/NCEP*

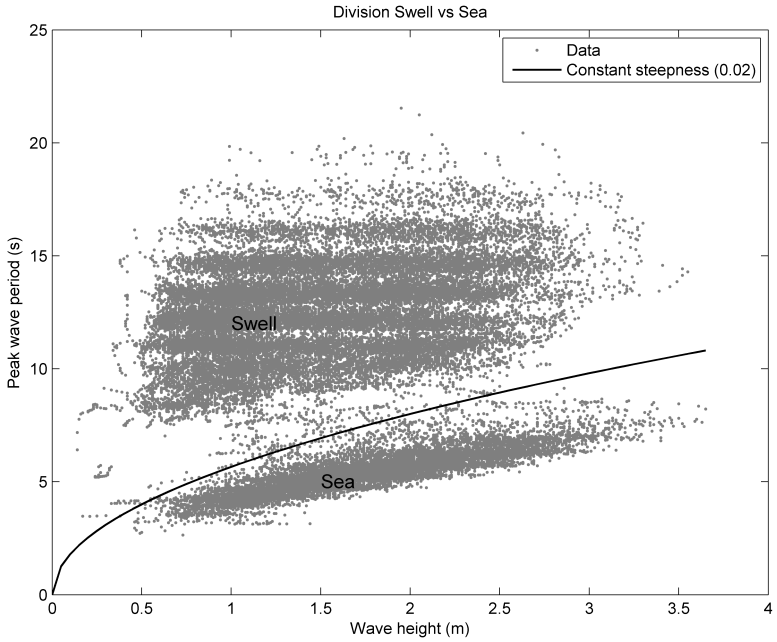

Figure 3.5: Scatterplot data points with line of constant steepness

Wind climate

The wind climate is shown in figure 3.4. Because of the strong relationship between the wave and wind climate the wind climate can not be represented by the variables independently. Therefore equation (3.6) is extended with the wind variables. By looking up the wind conditions at the wave conditions where from the distributions ($\mathbf{D}_w$) for the wave conditions were fitted, distributions for the wind conditions are estimated. The multivariate distributions of the six conditions are now described as follows;

$$\mathbf{D}_w = f(\boldsymbol{\mu}_w, \mathrm{Cov}_w) \quad \text{with} \tag{3.5}$$

$$\boldsymbol{\mu}_w = [\mu_{H_s} \quad \mu_\varphi \quad \mu_U \quad \mu_\psi]_w \qquad w = 1, 2, \ldots, 6 \tag{3.6}$$

$$\mathrm{Cov}_w = \begin{bmatrix} \sigma^2_{H_s} & \mathrm{Cov}(H_s, \varphi) & \mathrm{Cov}(H_s, U) & \mathrm{Cov}(H_s, \psi) \\ \mathrm{Cov}(\varphi, H_s) & \sigma^2_\varphi & \mathrm{Cov}(\varphi, U) & \mathrm{Cov}(\varphi, \psi) \\ \mathrm{Cov}(U, H_s) & \mathrm{Cov}(U, \varphi) & \sigma^2_U & \mathrm{Cov}(U, \psi) \\ \mathrm{Cov}(\psi, H_s) & \mathrm{Cov}(\psi, \varphi) & \mathrm{Cov}(\psi, U) & \sigma^2_\psi \end{bmatrix}_w \tag{3.7}$$

Where the covariance matrix is symmetric because of equation (A.13). As can be seen in appendix C, for two conditions (2 and 4) the wind speed data seems to fit better with a Weibull distribution instead of a Gaussian distribution. For these conditions the method of maximum likelihood is used to estimate the probability distribution parameters (section A.4). Because the generation of a (random) sample for correlated non-normal distributions is relatively difficult (section 4.3.3) these distributions assumed to have no correlation with other variables. Moreover, as shown in appendix D, the correlations for these conditions are rather weak.

The corresponding correlation matrices of the conditions are as follows;

$$\rho_w = \begin{bmatrix} 1 & \rho_{H_s,\varphi} & \rho_{H_s,U} & \emptyset \\ \rho_{\varphi,H_s} & 1 & \emptyset & \rho_{\varphi,\psi} \\ \rho_{U,H_s} & \emptyset & 1 & \rho_{U,\psi} \\ \emptyset & \rho_{\psi,\varphi} & \rho_{\psi,U} & 1 \end{bmatrix}_w \tag{3.8}$$

Where $\emptyset$ means that there is no physical correlation and is therefore neglected and assumed to be zero. Correlation between magnitude and direction of both wind and waves has not been neglected to represent the climate in a better way; magnitude may be higher from a specific direction.
When the correlation between the two magnitudes (wind speed and wave height) or between the two directions (wind direction and wave direction) is small ($\rho < 0.4$) the variables are said to be independent. In appendix D the correlations are shown. The correlations for wave condition 1 and 6 are quite large, which is typical for sea waves.

Bathymetry

Besides the forcing variables, the bathymetry is also an input variable. In this case study the reef is an important feature in the bathymetry, which is mainly responsible for the complex near shore hydrodynamic processes.
The bathymetry used for the model consist of a combination of several data sets. For the deep water bathymetry the global bathymetry and elevation data set SRTM30 PLUS is used with 30 arc second resolution (Becker et al., 2009). Closer to the shore, starting from a depth less than 30 meters, depth contours from admiralty charts are used. At the area of interest, in vicinity of the reef, survey measurements have been used. The bottom has been surveyed in several cross shore transects along the shore line. All the available data is interpolated to the model grid that is used for the model (see section B.2.1). Near the reef top surveying was difficult due to shallow water and strong wave breaking and therefore locally data is missing. Since the reef submergence is an important feature for the hydrodynamic processes but (locally) unknown the reef submergence is considered as stochastic variable. For the rest of the model area it is assumed that the available data is sufficient relative to the desired accuracy and the relative importance low. Therefore uncertainties stemmed from these data set are not taken into account.

Reef submergence

Based on visual observations it is known that the reef elevation varies spatially (de Wit, 2012) and therefore the reef contains higher parts. The quantification of the uncertainty of the reef elevation has been done using the available data in combination with Google Earth images. First the location of the reef top has to be determined. Figures 3.6 and 3.7 show Google Earth images of the area during strong and weak wave conditions. All waves are breaking at the reef top showing an indication of the location of the reef top. Figure 3.8 uses this indication; all survey measurements are plotted taking the reef top indication in the middle (0 m). Obviously onshore of the reef top the amount of measurements are less compared to offshore of the reef top. Northwards the reef top meets the shore line and so therefore the amount data points are diminished onshore of the reef top. The bar plot in figure 3.8 shows clearly a decrease of data points around the indicated reef top. It shows that the assumed reef top location is a good indication for the exact reef top location, assuming that the amount of available data points is inversely proportional to the reef elevation. The assumed width of the reef top is chosen to be 20 metres since there the amount of data points are the least. The rest of the area assumed to have enough data and is directly used (deterministically) for the model. The maximum measured reef elevation at our reef top location

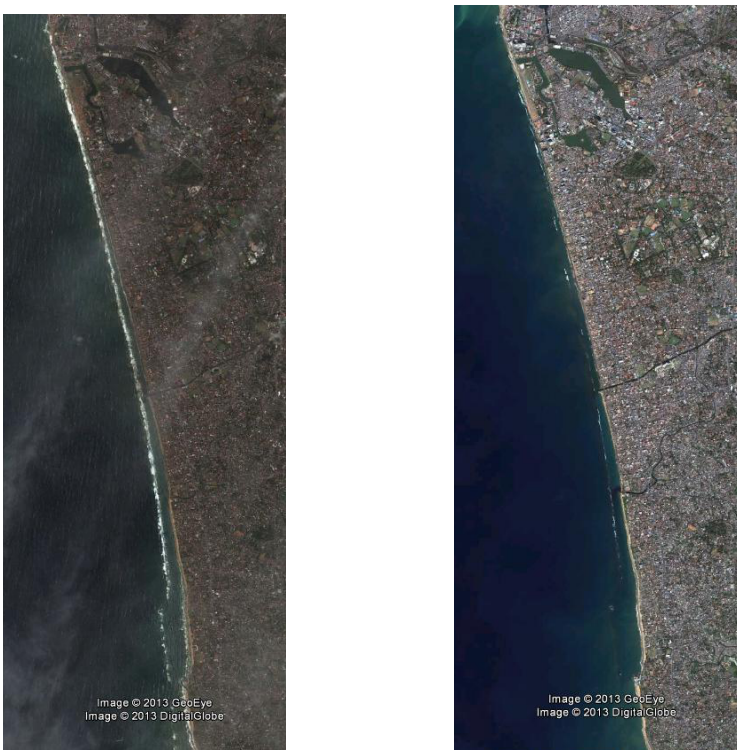

Figure 3.6: Google Earth images during strong (left) and weak (right) wave breaking

was -0.75 m+MSL [1] and the average elevation was -1.50 m+MSL. Since there is not enough data available to use statistical methods the stochastic parameters have to determined subjectively. Assuming that the average reef elevation is higher than the observed maximum the (Gaussian) distribution parameters are chosen to be a mean (μ) of -0.50 m+MSL with a standard deviation (σ) of 0.25 m. In that way the probability that reef elevation will be lower than -0.75 m+MSL will be approximately 15%. The reef elevation varies spatially (de Wit, 2012). Google Earth images (figures 3.6 and 3.7) during weak wave conditions show only wave breaking at several spots along the reef. These locations are used as indication for the (very) shallow parts. These (very) shallow parts have to stochastically 'move' with the total reef elevation; i.e. they are fully correlated ($\rho = 1$). The mean (μ) of the shallow parts is chosen to be at mean sea level (0 m+MSL) with the same standard deviation as for the total reef elevation; $\sigma = 0.25m$. From now on the reef submergence is denoted as positive reef elevation with respect to MSL (h_{reef} [-m + MSL]).

Water level

Despite the fact that the variability of the surface elevation, caused by tide, is also an uncertainty it is not taken into account. Assuming that the wave induced currents are dominating over the tidal currents only the water level with respect the reef top is an important factor for the hydrodynamic processes. Since the reef submergence is already considered as a stochastic variable it is assumed that the uncertainty of the water level is obviated by the variability of the reef top.

[1] MSL = Mean Sea Level

Figure 3.7: Google Earth images during strong (left) and weak (right) wave breaking, zoomed in near gap

3.2.2 Uncertainties of model parameters

Model parameters ($\boldsymbol{\theta}$) are parameters that change the performance of the model. Distinction can be made between physical and numerical parameters. Typical physical parameters are gravitational acceleration, water density or roughness. Numerical parameters are related to the way of obtaining an approximation of the governing equations. The latter parameters are not technical uncertainties and therefore not used in this study. A physical model parameter that does not have a negligible uncertainty is the roughness parameter. Roughness is difficult to measure and is therefore often used as calibration parameter. It is assumed that other parameters are deterministic.

Roughness

The roughness parameter for this study is given as roughness height of Nikkuradse (k_n). Different studies have been executed to determine the roughness height of beachrock and coral reef. By measuring the wave attenuation Vousdoukas et al. (2012) provided an estimation of the bottom roughness for colonized beachrock. Nelson (1996) estimated a roughness height for a coral reef and Lowe et al. (2005) provided for his study an estimation for the roughness of the Barrier reef in Australia including a quantified uncertainty. Below are the estimated roughness heights listed.

- Roughness height for colonized beachrock (equivalent); $k_n \approx 0.13m$ (Vousdoukas et al., 2012)
- Roughness of coral reef; $0.04m < k_n < 0,10m$ (Nelson, 1996)
- Roughness of barrier reef; $k_n \approx 0.16m \pm 0,03m$ (Lowe et al., 2005)

However for this case study the beachrock is not colonized and it is also not consist of coral. Like for the bathymetry no statistical methods can be applied to quantify the uncertainties and therefore the parameters have to be determined subjectively. Using the literature provided the (Gaussian) distribution parameters for the roughness height are chosen to be a mean of $\mu = 0.13$ m with a standard deviation of $\sigma = 0.05$.

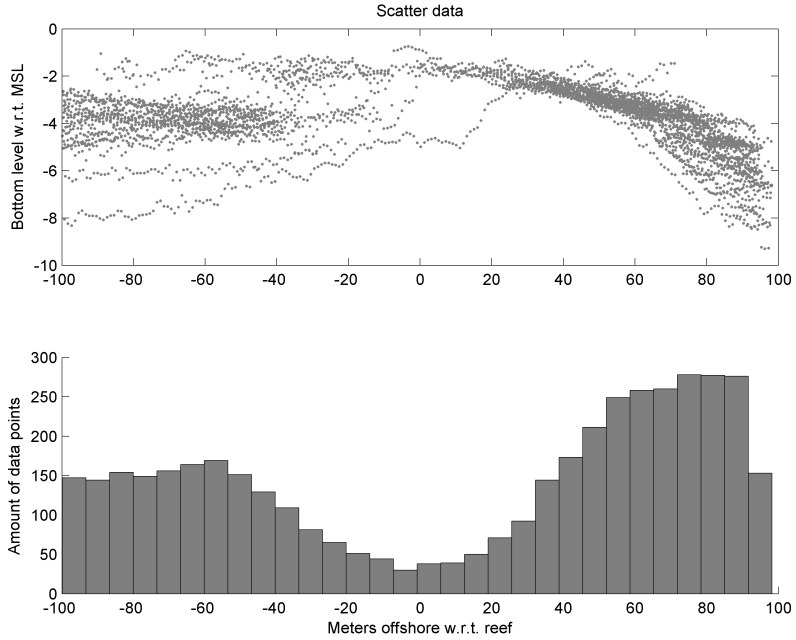

Figure 3.8: Top: Bottom level crosshore with respect to assumed reef top (0 m); positive is offshore. Bottom: Histogram of amount of data points

Chapter 4

Methods for stochastic modelling

In this chapter different methods for stochastic modelling are treated, i.e. methods that can give insight in the uncertainty of the model output. The criteria for the method that meets the research objective are determined. By comparing the criteria, characteristics of the method and the characteristics of this research, the applicability of different methods for stochastic modelling has been evaluated. Since the research only focuses on stationary systems time dependency is not involved. Therefore methods to deal with uncertainty propagating in time, where uncertainties are growing in time, are not mentioned in this research; for example ensemble forecasting (Toth and Kalnay, 1993).

4.1 Characteristics of research

The characteristics of this research that are related to stochastic modelling show the complexity of hydrodynamic models. This is an important factor for the applicability of the proposed methods. The aim of this research is to determine the reliability of the output given a set of input data and model parameters with their uncertainties. By analyzing the *uncertainty propagation* the uncertainties of the model output induced by the uncertainties of the input are calculated (Morgan and Henrion, 1990). Relative contribution and importance of the input uncertainty to the uncertainty of the output is referred as *uncertainty analysis*. Below characteristics of this research are listed, including model characteristics and characteristics of the uncertainties.

- Aim to calculate uncertainty propagation
- Output spatially varying
- Interest in stationary results; only space dependence [1]
- Model characteristics
  - Non-linear
  - Large simulation time
- Characteristics of uncertainties
  - Relative large uncertainty (coefficient of variation untill $CV = 0.5$)

[1] One should realise that the input variables are variable in time but temporal variability is translated into uncertainty. The output is only space dependent

- Uncertainties are quantified by continuous probability distributions functions.
- Statistically dependent variables are present
- Correlated variables are normal distributed
- Uncorrelated variables are normal and non-normal distributed

4.2 Criteria

For a correct assessment of the applicability of the available methods, requirements for the method need to be determined which are partly stemmed from the characteristics of the research. Since the aim of this research is to analyse the uncertainty propagation there is no need for estimating (very) small percentiles. Furthermore; that would require detailed information about the input uncertainties involved that is not available, see section 3.2 (van der Klis, 2003). Therefore insight in the 5^{th} and 95^{th} percentiles would suffice. The aim is to use a relatively easy to implement method with as minimal computational effort as possible.

- Method that gives insight in the 90% confidence intervals
- No need for insight in small probabilities; tails are less important
- Minimize implementation effort.
- Minimize the computational effort; i.e. amount of simulations

4.3 Monte Carlo methods

The uncertainties of input $\mathbf{I}$ and parameters $\boldsymbol{\theta}$ are characterized in a sequence of probability distributions;

$$D_{I,1}, D_{I,2}, \ldots, D_{I,nI} \tag{4.1a}$$
$$D_{\theta,1}, D_{\theta,2}, \ldots, D_{\theta,n\theta} \tag{4.1b}$$

For this method a (random) sample of $\mathbf{I}$ and $\boldsymbol{\theta}$ needs to be generated that is consistent with the probability distributions of equations (4.1a) and (4.1b) (Helton and Davis, 2003);

$$\mathbf{I}_i = [I_{i,1}, I_{i,2}, \ldots, I_{i,nI}] \qquad i = 1, 2, \ldots, m \tag{4.2a}$$
$$\boldsymbol{\theta}_i = [\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,n\theta}] \qquad i = 1, 2, \ldots, m \tag{4.2b}$$

Each set of values ($\mathbf{I}_i$,$\boldsymbol{\theta}_i$) defines a scenario, which is used as input for the model in equation (1.1), computing a corresponding set of output values ($\mathbf{Y}_i$). This process is repeated m times, producing m sets of output values (Morgan and Henrion, 1990). With sufficient sample size m, the output sample can be used for statistical analysis to derive an approximation and quantify corresponding uncertainties. Same methods can be used as been explained in appendix A. In order to generate the set of samples of (4.2a) and (4.2b) different procedures are available, of which the crude Monte Carlo is the simplest. With crude (or random) sampling a sample is generated by drawing values at random from the distribution of each input and parameter variable in $\mathbf{I}$ and $\boldsymbol{\theta}$. When variables are statistically dependent techniques are avalaible to generate a sample, see section 4.3.3. The output sample consists of independent random values from the output distributions, irrespective of nI and $n\theta$. That means that the accuracy of the output distribution parameters is independent of the amount of input and model parameters but dependent on sample size m. Depending on the sampling technique the required sample size m can be different.

Sample size

As mentioned the required sample size m is dependent on the desirable accuracy of the analysis. The desirable accuracy for this analysis is based on the desired insight in the *uncertainty propagation*. Estimation of the upper and lower boundaries, say the 5th and 95th percentiles of the output is therefore sufficient (section 4.2). The required sample size of a crude Monte Carlo analysis can be estimated by a method developed by Morgan and Henrion (1990) and is dependent on the required precision of estimates of the cumulative distribution. The accuracy of for instance fractiles can be determined using the equations provided in appendix A. The confidence interval for fractiles can be estimated using equation (A.9) and (A.10). Those equations can be rewritten to the following equation;

$$m = p(1-p)\left(\frac{c}{\Delta p}\right)^2 \tag{4.3}$$

Where $2\Delta p$ is the required fractile interval with deviation c, which is a function of the required confidence interval α, via equation (A.4). Suppose that the required accuracy is to be 95 % confident that the 90^{th} is in between the 85^{th} and the 95^{th} percentile. Than $p = 0.90$, $\Delta p = 0.05$ and $c = 1.96$, with $\alpha = 0.95$. Using equation (4.3) an required sample size of 144 is estimated.

Characteristics

- Standard statistical analysis methods can be used to estimate distribution parameters of output (Appendix A)
- Accuracy dependent on sample size m
- Sample size independent of amount of input and model parameters
- Method can relatively easy be coded
- For sufficient sample size; estimates are reliable and accuracy can be calculated in a suitable way
- Non-linearities are automatically dealt with
- Computationally inefficient since large amount of simulations are required
- Because of the randomness of the generation of the sample, old samples can be reused, which implies that the sample can easily be extended.
- Relative contribution of uncertainties of input and model parameters to uncertainties of model output can be determined; i.e. uncertainty analysis.

4.3.1 Latin Hypercube sampling

The key value of Monte Carlo analysis is not random sampling but the resulting equidistrubtion of the output sample (Morgan and Henrion, 1990). With that knowledge other sampling techniques have been developed; like *stratified sampling*. With stratified sampling the sample space $\mathbf{W}$ of each input or model parameter is divided in disjoint strata; i.e. subsets of sample space $\mathbf{W}$. Then a random sample is generated separately from each stratum (McKay et al., 2000; Helton and Davis, 2003). An example of stratified sampling is *Latin hypercube sampling* in which the strata are equiprobable distributed. Each input variable and model parameter (X_j,θ_k where $j = 1, 2, \ldots, nX$ and $k = 1, 2, \ldots, n\theta$) is divided in m equiprobable strata with probability $1/m$.

In this way the samples are more equally divided over the sample space compared to random sampling, what means the sample size can be reduced.

Characteristics

- Usually smaller required sample size compared to crude Monte Carlo analysis; less computational effort
- Method can relatively easily been programmed for correlated (normal) distributions.
- Required sample size can not be estimated in advance
- Samples can not be reused, which implies that the sample can not be easily extended like for crude Monte Carlo
- Accuracy can not be easily calculated since then you need to repeat the procedure several times which will undo the sample efficiency (van der Klis, 2003)

4.3.2 Importance sampling

In the described sampling methods above the generated sample is divided over the entire sample space. With *importance sampling* the generated sample points are concentrated around the sample interval of interest; i.e. of most 'importance' (Hammersley and Handscomb, 1964). For example if someone is interested in (very) low probabilities the generated sample is more concentrated around the tails of the probability distribution. This method has been succesfully applied in reliability analyses where sampling is concentrated around failure regions (Beck and da Rosa, 2006).

Characteristics

- Efficient for estimation of (very) small probabilities or reliability analyses
- Smaller sample size compared to crude Monte Carlo

4.3.3 Generating sample of correlated variables

In this study some variables have a probabilistic dependency. To account for correlation in the generation of a (random) sample is not hard for Gaussian distributed variables. However for the generation of a sample of correlated variables with (a mixture of) other marginal distributions the situation gets more difficult. Chang et al. (1994) proposed a method to perform Monte Carlo simulations with correlated variables while preserving their marginal distributions. An other method described by Iman and Conover (1982) works with *rank-order correlations*, known as the Spearman correlation instead of the commonly used Pearson correlation (appendix A.2). Since the correlated variables in this study are all Gaussian distributed a relatively straight forward method can be used (Morgan and Henrion, 1990). Let's assume that a sample is needed of two variables x and y both with a standard normal distribution (Φ) and correlation ρ. An auxiliary variable z is used; (random) samples are generated of x and z of independent normal distributions. These samples are used to compute the corresponding sample for y;

$$y := \rho x + z\sqrt{1-\rho^2} \tag{4.4}$$

Where := denotes a replacement statement. The sample of y is now also a sample of a standard normal distribution but correlated with x, specified with ρ. To transform these samples to Gaussian distributions with parameters (μ, σ) defined in chapter 3, without affecting their correlation, the following expressions are used;

$$x := \mu_x + \sigma_x \times x \tag{4.5}$$
$$y := \mu_y + \sigma_y \times y \tag{4.6}$$

4.4 Analytical methods

Except for very simple linear models it is almost never possible to determine an exact analytical solution for uncertainty propagation (Morgan and Henrion, 1990). Approximations for analytical solutions can be obtained using for instance numerical integration. An other very common used technique for approximation of analytical relationships is based on Taylor series expansion.
The *first order approximation* , also known as method of moments, is often used in reliability analyses as the First Order Reliability Method (Beck and da Rosa, 2006; van der Klis, 2003). First a nominal scenario has to defined, which is in fact a 'base case' scenario. Let's assume that the nominal value of each variable in $\mathbf{X}$ from equation (1.1) is equal to its expected value.

$$X_j^0 = E[X_j] \qquad j = 1, 2, \ldots, nX \tag{4.7}$$

That means that the nominal scenario is equal to the expectation of $\mathbf{X}$. The Taylor series expansion is a way to approximate the deviations of the output from its nominal value[2], $Y - Y^0$ in terms of the deviations of the input and model parameters from their nominal values, $X_j - X_j^0$. The first two terms of the Taylor series expansion is then written as follows;

$$\begin{aligned} Y - Y^0 = &\sum_{j=1}^{nX} (X_j - X_j^0) \left[\frac{\partial Y}{\partial X_j}\right]_{\mathbf{X^0}} + \\ &\frac{1}{2} \sum_{j=1}^{nX} \sum_{k=1}^{nX} (X_j - X_j^0)(X_k - X_k^0) \left[\frac{\partial^2 Y}{\partial X_j \partial X_k}\right]_{\mathbf{X^0}} + O(X^3) \end{aligned} \tag{4.8}$$

Where the subscript $\mathbf{X}^0$ means that the derivatives are evaluated at the nominal scenario and $O(X^3)$ are higher order terms. Now an approximation of the expected deviation of output Y can be derived;

$$\begin{aligned} E[Y - Y^0] \approx &\sum_{j=1}^{nX} E[X_j - X_j^0] \left[\frac{\partial Y}{\partial X_j}\right]_{\mathbf{X^0}} + \\ &\frac{1}{2} \sum_{j=1}^{nX} \sum_{k=1}^{nX} E[(X_j - X_j^0)(X_k - X_k^0)] \left[\frac{\partial^2 Y}{\partial X_j \partial X_k}\right]_{\mathbf{X^0}} \end{aligned} \tag{4.9}$$

For the first order approximation only the first term is taken. And because of the assumption made in equation (4.7) the expected value of Y is approximated by;

$$E[Y] \approx Y^0 = h(\mathbf{X}^0) \tag{4.10}$$

The first order approximation of the variance of output Y is then, using equation (A.11), given

[2] Assuming one output variable in $\mathbf{Y}$

by;

$$\begin{aligned}
\mathrm{Var}[Y] &= E[(Y - Y^0)^2] \approx E\left[\left(\sum_{j=1}^{nX}(X_j - X_j^0)\left[\frac{\partial Y}{\partial X_j}\right]_{\mathbf{X^o}}\right)^2\right] \\
\mathrm{Var}[Y] &\approx \sum_{j=1}^{nX}\sum_{k=1}^{nX} E[(X_j - X_j^0)(X_k - X_k^0)]\left[\frac{\partial Y}{\partial X_j}\right]_{\mathbf{X^o}}\left[\frac{\partial Y}{\partial X_k}\right]_{\mathbf{X^o}} \\
\mathrm{Var}[Y] &\approx \sum_{j=1}^{nX}\sum_{k=1}^{nX} \mathrm{Cov}[X_j, X_k]\left[\frac{\partial Y}{\partial X_j}\right]_{\mathbf{X^o}}\left[\frac{\partial Y}{\partial X_k}\right]_{\mathbf{X^o}}
\end{aligned} \tag{4.11}$$

And because of equations (A.12) and (A.13) the approximation becomes;

$$\begin{aligned}
\mathrm{Var}[Y] \approx &\sum_{j=1}^{nX} \mathrm{Var}[X_j]\left[\frac{\partial Y}{\partial X_j}\right]^2_{\mathbf{X^o}} + \\
&2\sum_{j=1}^{nX-1}\sum_{k=j+1}^{nX} \mathrm{Cov}[X_j, X_k]\left[\frac{\partial Y}{\partial X_j}\right]_{\mathbf{X^o}}\left[\frac{\partial Y}{\partial X_k}\right]_{\mathbf{X^o}}
\end{aligned} \tag{4.12}$$

If the variables are statistically independent the second term in equation (4.12) becomes zero. With a numerical scheme the partial derivatives can be approximated. Although this equation is relatively easy to evaluate, it is a *linear* approximation of the model function at the nominal scenario. That means that this approximation is only reasonable given that the uncertainties of the input variables are small and the model function is smooth (i.e. small second and higher derivatives) in the range of variation of the input variables (Morgan and Henrion, 1990; van der Klis, 2003). Despite the non-linearity this method has been applied successful for several studies (Maurer et al., 1998) . However for the study of van der Klis (2003) with large uncertainties in combination with non-linearity this method seemed not suitable.

Characteristics

- Linear approximation; only reliable if relation is (close to) linear in range of variations of input variables
- Only reliable for relative small uncertainties of input and model parameters. Requirement becomes stronger for stronger non-linearity (van der Klis, 2003).
- Amount of simulations are less compared to crude sampling.
- Suitable for (structural) reliability methods
- Suitable for uncertainty analysis
- Relatively difficult to implement

4.5 Response surface method

A *response surface* is a surface that describes the output($\mathbf{Y}$) as function of the input variables($\mathbf{X}$). This surface can be used for both calculating uncertainty propagation and uncertainty analysis.

With use of a moderate number of model runs a simplified response surface can be fitted, which is an approximate version of the full model. For example Monte Carlo samples can be used to construct a response surface, however less sample scenarios can be sufficient to construct a reliable fitted response surface. The first order approximation is in fact a linearisation of the response surface; it assumes a (hyper)plane at the nominal scenario. With the fitted response surface analytical methods can be used to calculate uncertainty propagation and perform uncertainty analysis. It is a combination of a sampling and an analytical method. According to Morgan and Henrion (1990) there are three key issues in response surface modelling; how to select small sample of scenario's with which to run the large model. How to screen the uncertain inputs and identify which ones need to be modelled explicitly and last, and maybe most important, how best to fit a response surface.

Characteristics

- Relatively much effort to create response surface
- Less simulation required compared to Monte Carlo
- Suitable for uncertainty analysis
- Difficult to examine the accuracy of results

4.6 Conclusions

- *Monte carlo method* is suitable, but has a disadvantage regarding computational effort.
- *Latin hypercube sampling* is suitable and reduces the sample size, but has a disadvantage that sample size can not be determined on forehand and sample can not easily be extended, like for (crude) Monte Carlo.
- *Importance sampling* is more suitable for reliability analyses and has for this study no advantage in relation to other methods.
- *First order approximation* may not be suitable because of the non linearity of the model.
- *Response surface method* is relatively difficult to execute.

Monte Carlo method seems to be the most appropiate method to apply in this study. Latin hypercube sampling and the First order approximation method can reduce a lot of computational effort however their disadvantages are considered stonger than those from (crude) Monte Carlo. Monte Carlo is a very robust method that is relatively easy to implement and can even provide insight in the relative importance of variables, i.e. uncertainty analysis (Morgan and Henrion, 1990).

Chapter 5

Execution of the proposed method

From the previous chapter it became clear that the (crude) Monte Carlo method was the most appropriate probabilistic method for this study. In this chapter the execution of the method for stochastic modelling, i.e. Monte Carlo procedure, is explained in detail here. It gives insight in the coupling between the stochastic method and the numerical model used for this study. Therefore a description of the numerical model, as used for this study, is provided, including a sensitivity analysis and a (crude) validation. Furthermore the most important assumptions and limitations regarding numerical modelling are given here.

5.1 Execution process

Figure 5.1 gives an overview of the execution process. Below an description of the process is given.

1. First a (random) sample is generated that corresponds with the probabilistic distribution parameters and correlations derived in section 3.2. The sample size is chosen to be $m = 160$. The required sample size, calculated in section 4.3, was estimated to be 144. However assuming that some runs fail the sample size is chosen to be larger. After generating the sample there are 160 values available for each input and model parameter, each describing a different scenario.

2. Each scenario is executed with the numerical model (grey area). The numerical model consist of two modules, the wave and hydrodynamic module. The model is used offline, which means that there is no interaction of the hydrodynamics on the wave module[1]. So each scenario is first run with the wave module, SWAN (section B.1) to obtain the stationary wave radiation stresses and local (nearshore) wave conditions.

3. The output of the wave module is used for the hydrodynamic module, FINEL2D (section B.2), to predict the surface elevation, flow field and sediment fluxes. Model parameter(s) of the scenario's are used for both the wave and hydrodynamic model. In this case only the roughness height (k_n) is used as model parameter (section 3.2).

4. All the output of every scenario is stored and when all scenario's are executed the data can be used for statistical analysis.

[1] No output of the hydrodynamic model is used for the wave module, that means there is no iteration procedure. Tests of online calculations showed no significant difference in the results

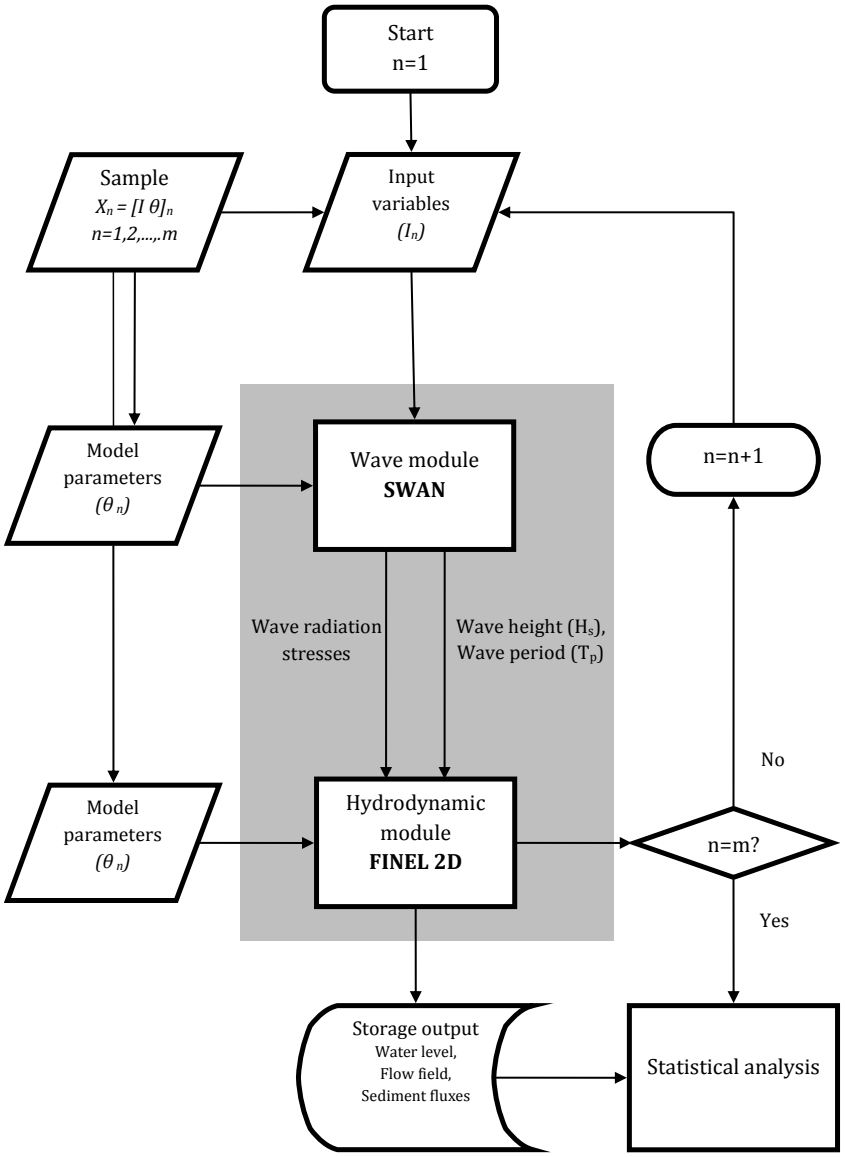

Figure 5.1: Flowchart of model execution

5.2 Numerical model

As mentioned in the previous section a numerical model is used to obtain the predictions for the hydrodynamics and sediment fluxes stemmed from a specified scenario. The total numerical model consist of two models; a wave model to predict the local wave field (SWAN) and a hydrodynamic (and morphodynamic) model to predict the local hydrodynamics and sediment fluxes stemmed from the predicted local wave field (FINEL2D). In appendix B a full description of the numerical models can be found. In this section the results of a sensitivity analysis and a crude validation are presented. Furthermore the most important assumptions and limitations regarding the used numerical model are treated in this section.

5.2.1 Expected sensitivity

The expected relative importance, i.e. (single) sensitivity of the variables gives insight in the expected relationships and their dependency. Analytical and numerical results are compared. The focus is on the most important hydrodynamic process being the (wave generated) set-up which causes complex rip-current systems in vicinity of the reef. By examining the analytical relationship with the numerical model, not only an approximation of the quantitative sensitivity is provided, it shows whether the model is capable of predicting the expected processes in a correct way. Note that in this section only the single sensitivity per variable is examined, i.e. *individual parameter variation* (van der Klis, 2003), in contrast to the global sensitivity as examined with the stochastic method (section 6.3). Quantitative validation is done by means of comparing numerical results with laboratory tests provided by Seelig (1983).

Analytical

With the use of a simple dimension analysis the relation between variables can be predicted and therefore an indication of their sensitivity can be provided. The dimension analysis focuses on the (wave generated) set-up and corresponding velocities in an idealized case.

Set-up

The wave generated set up behind the reef can be approximated by the momentum and mass balances; the balances that many numerical models, like FINEL 2D (chapter B), use for the prediction of hydrodynamics. An idealized one dimensional case is considered with an uniform coast with a simple reef and lagoon configuration. For the dimension analysis first the depth integrated momentum balance equation in the x-direction is used;

$$\frac{\partial hu}{\partial t} + \frac{\partial hu^2}{\partial x} + \frac{\partial huv}{\partial y} + gh\frac{\partial \eta}{\partial x} - \frac{1}{\rho}\tau_b + \frac{1}{\rho}\tau_r = 0 \tag{5.1}$$

Where τ_b and τ_r are respectively the bottom shear stress and the wave induced radiation stresses. The total water depth (h) is the summation of the local bottom level (z) and water level (η) with respect to mean sea level (MSL). Since the focus of this study is on stationary results the first term can be neglected. If the non-linear terms (advective and bottom friction terms) are neglected as well the following equation is left over.

$$gh\frac{\partial \eta}{\partial x} = -\frac{1}{\rho}\tau_r \tag{5.2}$$

The wave induced shear stress τ_r is formulated with the following expression;

$$\tau_r = \frac{\partial S_{xx}}{\partial x} + \frac{\partial S_{xy}}{\partial y} \tag{5.3}$$

An uniform situation in the y-direction is considered so the y-derivative can be neglected. Radiation stress (S_{xx}) is expressed as;

$$S_{xx} = \left(n - \frac{1}{2} + \cos^2\varphi\right) E \tag{5.4}$$

Where n is the ratio between the wave group celerity (c_g) and wave celerity (c); generally $n = {}^1/_2$ for deep water and $n = 1$ for shallow water. The wave direction is denoted as φ and the wave energy, E, is expressed with;

$$E = {}^1/_8 \rho g H_{rms}^2 \tag{5.5}$$

In this situation it can be considered that it is shallow water and therefore $n = 1$. With normal incident waves ($\varphi = 0$) S_{xx} becomes equal to ${}^3/_2 E$. Suppose that all waves will break on top of the reef and therefore all energy is dissipated by wave breaking. In that way the length scale (∂x) for both the radiation stress derivative and the water level gradient can be considered to be constant. If equation (5.2), (5.3), (5.4) and (5.5) are put together it shows that the wave generated set up (η) is proportional to H_{rms}^2.

$$\eta \propto H_{rms}^2 \propto H_s^2 \tag{5.6}$$

This is for normal incident waves. However the wave set up gets smaller if the waves approach the coast with an angle, which is described with equation (5.4). From that equation it can be seen that the wave generated set up is proportional to $\cos^2\varphi$.

$$\eta \propto \cos^2\varphi \tag{5.7}$$

The dimension analysis above is based on conservation of momentum. However an alternative method would be to use the conservation of mass. Loveless et al. (1998) state that the volume of flow backwards and forwards in half wave period of a progressive wave is equal to $HL/\pi T$. However due to a bar like structure, or in this case a reef, there is greater resistance tot the return flow than the overflow. Therefore the net inflow must be balanced by a set up. Loveless et al. (1998) says that if the resistance is dominantly turbulent, which is the case, then the set-up (η) should be proportional with u_o^2, the mean offshore set-up driven velocity. In that way it follows that

$$\eta \propto \left[\frac{H_i L}{hT}\right]^2 \tag{5.8}$$

Where L and T are respectively the wave length and wave period. Equation (5.8) shows also that the set-up is proportional to H^2. Loveless et al. (1998) eventually found out that the dimensionless wave setup for detached submerged breakwaters best describes his results with the following equation;

$$\frac{\eta}{B} = \frac{(H_i L/hT)^2}{8 g D_{n50}} \cdot e^{-20(R_c/h_c)} \tag{5.9}$$

Where D_{n50} is the nominal stone diameter that is used for the breakwater and R_c/h_c the relative submergence. Equation (5.9) shows that the set-up is inversely proportional to the nominal stone diameter, which is in fact a representation of the roughness height. So a larger roughness height means a lower set-up. Which (partly) corresponds with the conservation of momentum (eq. (5.1))

where the set-up is inversely proportional with the bed shear stress. However the bed shear stress is non linear related with the roughness height and is proportional to the velocity squared. Therefore a simple analytical solution is difficult to derive. Nevertheless it is expected that the wave generated set-up is inversely proportional to the roughness height;

$$\eta \propto k_n^{-1} \tag{5.10}$$

Equation (5.9) shows also that the set-up is inversely proportional of the submergence of the reef. Which can be explained by the fact more waves are breaking when the submergence gets smaller and therefore the set-up increases. According to the results of Loveless et al. (1998) the maximum set-up was found when the water level was just below the crest of the breakwater. Following Loveless et al. (1998) the set-up has an exponential relationship with the reef submergence;

$$\eta \propto e^{-h_{reef}} \tag{5.11}$$

Velocities

The set-up is partly diffused by the generation of rip-currents. Rip-currents are caused by the spatial variety of the set-up which causes a (alongshore) water level gradient resulting in a current. The place and magnitude of these currents are difficult to predict. Many effects are important and an exact solution for these current velocities is not available. But if the situation is very simplified an indication of the relationship can be given. If inertia, bottom friction and other external forces are neglected equation (5.1) can be reduced to;

$$\frac{\partial hu^2}{\partial x} + gh\frac{\partial \eta}{\partial x} = 0 \tag{5.12}$$

Now the advective term balances the water level gradient term. Which is in fact a variant of the *Bernoulli equation* (Battjes, 1995). This means that the expected velocities are proportional to the square root of the difference in set-up;

$$\eta \propto u^2 \tag{5.13}$$

Numerical

The expected relations from previous section are checked with the numerical model(s) as used for this study. A base case scenario is chosen and different variables, like wave height and direction, are changed to generate different scenario's to check their (single) sensitivity. The scenario's are executed with the full model that is also used to generate the results for the results presented in chapter 6. The base case scenario is described with the following parameter settings;

- Wave height $H_s = 1.75$ [m]
- Wave direction $\varphi = 197.5$ [degrees north]
- Reef elevation $h_{reef} = 0.75$ [-m +MSL]
- Roughness height $k_n = 0.05$ [m]

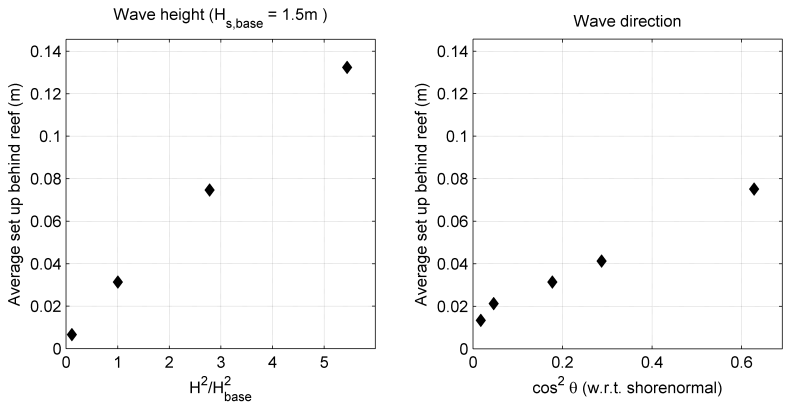

Figure 5.2: Average set-up behind reef versus relative wave height squared and $\cos^2 \varphi$

Set-up

The average wave generated set-up is calculated by taking the average of the water level (w.r.t. MSL) everywhere behind the reef. The set-up from the different scenarios are plotted against the relative wave height squared (H^2/H^2_{base}) and $\cos^2 \varphi$ and are shown in figure 5.2.
Figure 5.2 shows that the relation between these parameters are (close to) linear, which corresponds with equation (5.6) and (5.7). From these plots it is clear that wave height is more sensitive for the mean set-up behind the reef than the wave direction.
The sensitivity for the roughness height (k_n) and the (exponential) submergence (h_{reef}) are shown in the plots in figure 5.3. It looks like the roughness height is inversely proportional to the mean set-up. This is in line with equation (5.10). The right plot shows the results for the relative submergence. Note that the data is plotted against the exponential negative submergence, what means that the relative submergence is also inversely proportional to the set-up. The most right data point corresponds with no submergence; i.e. reef top is on MSL. The calculated mean set-up is (almost) zero with a relative submergence of zero, this is because the numerical model can not deal with overwash and/or overtopping of the waves, what means that there is no interaction of momentum or mass over the reef which results in no set-up. The other data points show an (almost) linear relationship which in line with equation (5.11). From figure 5.2 and figure 5.3 it is clear that the sensitivity for the mean set-up is relatively higher for the wave height compared to the other variables. The roughness height and reef submergence has a relatively low sensitivity to the mean set-up.

Velocities

The occurring velocities in the numerical model are varying a lot spatially, both in direction and magnitude. The most notable rip current that is visible is at the gap of the reef. In figure 5.4 is the absolute velocity field shown in vicinity of the reef, with south the big offshore rip current. The blue polygon shows the area where from the measurements are taken for the sensitivity analysis. The average from the observed (absolute) velocities in the polygon is used as output variable.
Figure 5.5 shows the mean set-up versus the mean velocity for all the sensitivity runs that have been executed. A fitted line shows that the data fits well to a square root function ($u = a + b\sqrt{\eta}$), which is in line with equation (5.13). The sensitivity of the variables are shown in figure 5.6 and figure 5.7. Now it seems that the velocity is proportional to the wave height and the cosine of the wave angle. That corresponds with the relation in figure 5.5. In figure 5.7 it looks like that the

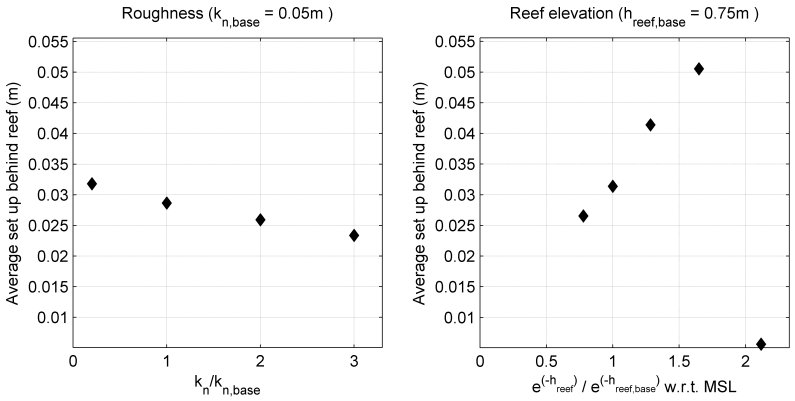

Figure 5.3: Average set-up behind reef versus relative roughness height and relative submergence

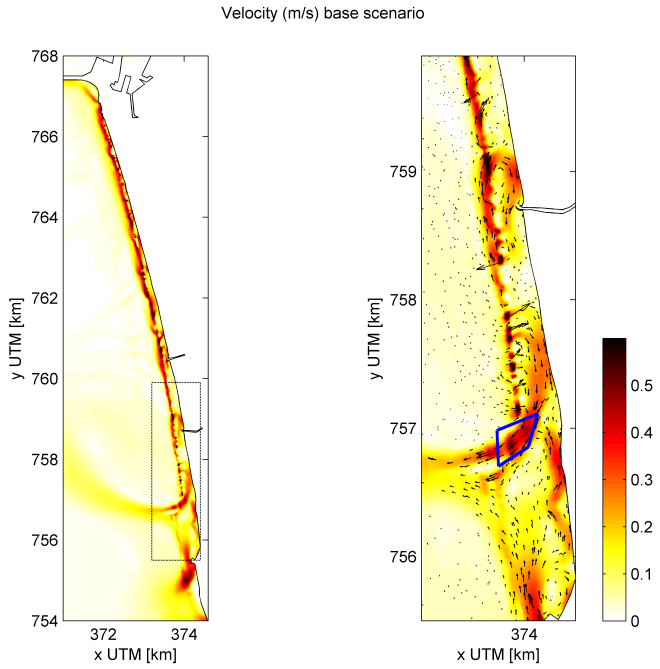

Figure 5.4: Absolute velocity field of base scenario. Blue polygon is area where measurements are taken for sensitivity analysis.

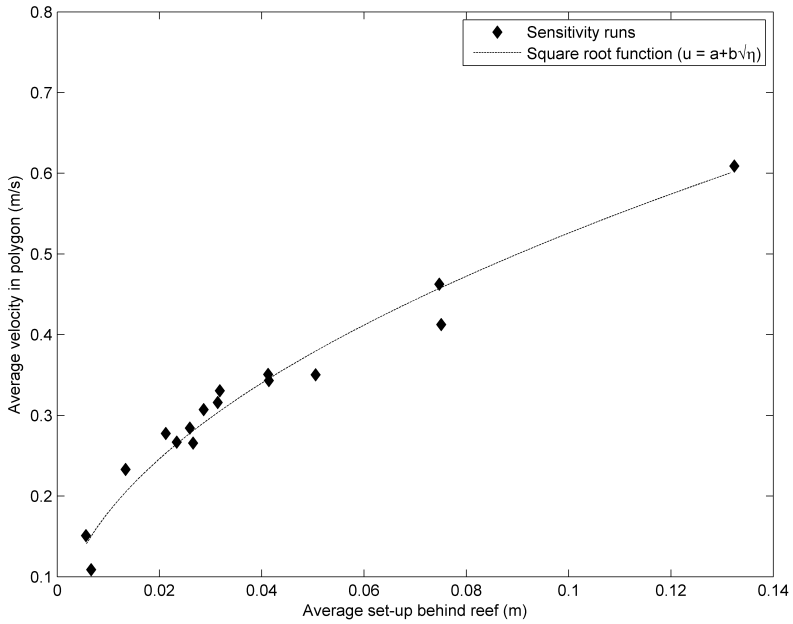

Figure 5.5: The average set-up behind reef versus the average velocity in polygon with a fitted square root function

velocity is also linear proportional to the roughness height. The sensitivity however is relatively large for the velocity compared to the set-up. The relationship for the reef submergence is also shown in figure 5.7. The velocities are inversely proportional to the reef submergence, however a clear relationship has not been found.

5.2.2 Validation

Validation of the numerical model is important in order to be sure the model is able to predict the processes in the model in a correct way. Normally a model is validated and calibrated by use of nearfield (current) data, however since we are dealing with a data poor environment such data is not available. Therefore the validation is done by means of validating the dominant processes of the system. The main characteristic processes are generated by the presence of the reef. The wave generated set-up behind the reef is one of the main causes for the complex rip current system. The sensitivity analysis in section 5.2.1 has shown that the model shows (almost) similar relationships according to the theory and literature. Therefore it is considered that qualitatively the dominant processes are, according to the theory, correctly predicted. Quantitatively the model is tested using a laboratory study. Seelig (1983) performed a laboratory study to investigate the reef-lagoon system and provides an (location specific) empirical relationship for the wave generated set-up on a reef. An idealized reef cross section of the coast of Guam was used for the study of Seelig (1983). In figure 5.8 the reef cross section is visible. The exact reef lagoon configuration of the laboratory study is used for the numerical model. After simulating several scenario's with different wave heights and wave periods the set-up behind the reef is measured and compared to the following empirical relationship (Seelig, 1983);

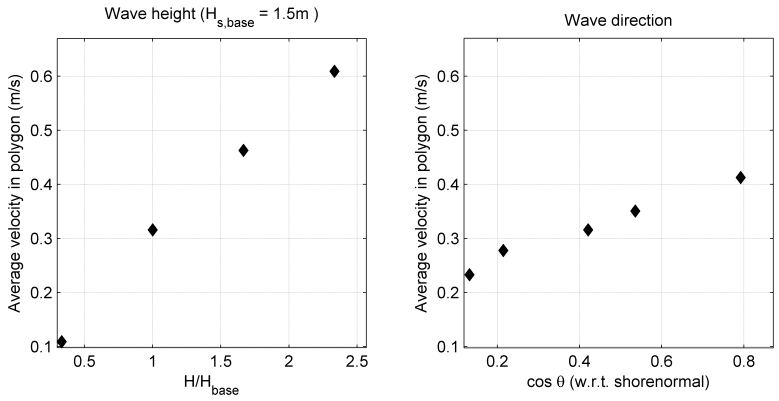

Figure 5.6: Average velocity in polygon versus relative wave height and $\cos\varphi$ w.r.t. shore normal

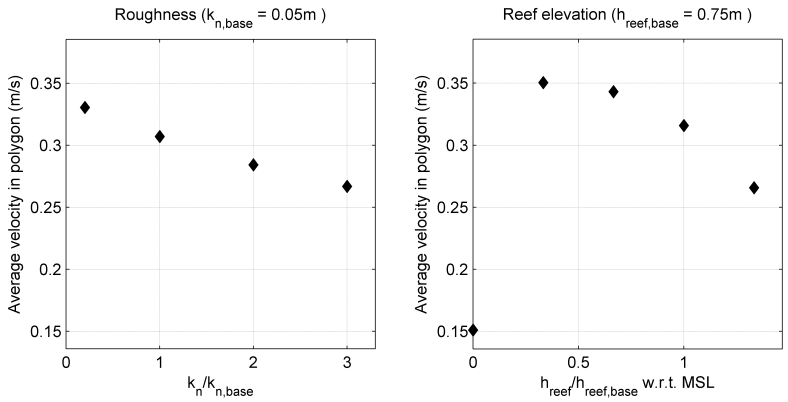

Figure 5.7: Average velocity in polygon versus relative roughness height and relative submergence

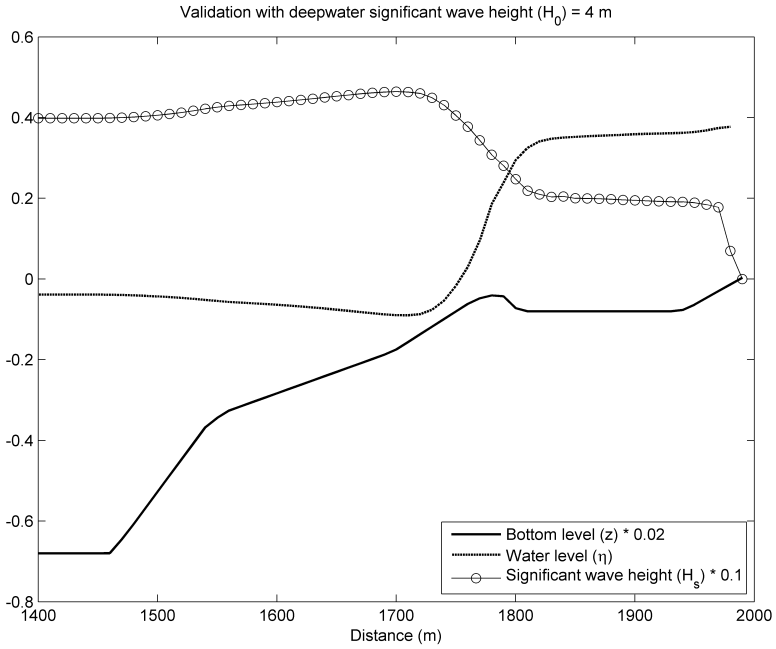

Figure 5.8: Cross section of reef-lagoon system with example calculation of wave height and water level for deep water significant wave height (H_0) of 4 meter.

$$\eta = -1.25 + 0.73 \cdot {}^{10}\log\left(H_0^2 T\right) \tag{5.14}$$

Figure 5.9 shows the results. The model results matches the empirical relation well, although there is a underestimation for some data points. The effect of wave rollers is not taken into account in the model which can lead to less discharge over the reef and results in a smaller set-up. Also the bottom friction coefficient was not given in the article so an estimated value has been used.

5.2.3 Assumptions and limitations

For this study it is assumed that the predictions given by the numerical models represent the reality in a correct way. However assumptions have been made since numerical modelling has limitations. The most important assumptions and limitations regarding the numerical models that are used are listed here. Note that the uncertainties stemmed from these assumptions and limitations are not taken into account in this study. More assumptions, limitations and information about the numerical models can be found in appendix B.

Dry cell criterion The dry cell criterion is a threshold value which determines whether an element is considered wet or dry. Dry cell criteria are required since with (very) small water depths unrealistic results may be obtained (by dividing with water depth for instance). Since we are dealing with reef elevations close to MSL, the dry cell criterion becomes important. The imposed dry cell criterion is equal to 0.05 meter, so if for instance a reef elevation of 0.05 metres below MSL is imposed the elements are considered dry within the model. Dry elements in the model implies that there is no water nor velocity on the elements. This

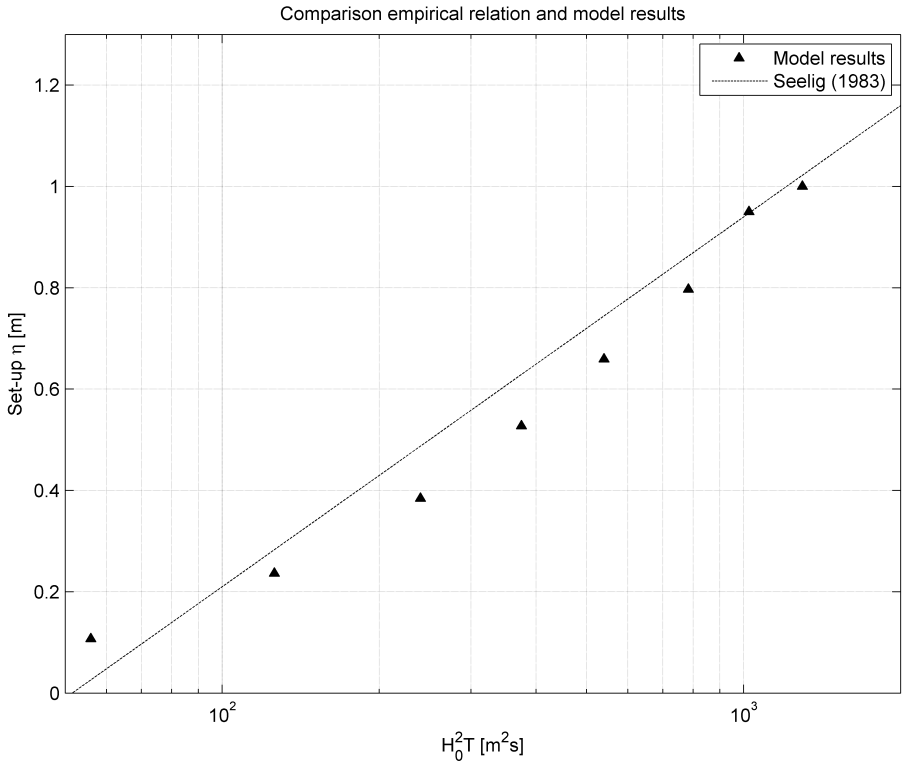

Figure 5.9: Comparison of model results with empirical relation of Seelig (1983)

means when elements of the reef are considered dry no interaction over the reef is possible. In reality interaction over the reef is possible when the reef top is close to MSL, by for instance overwash of waves.

Stationary modelling This study focuses only on stationary systems which means that time dependency is neglected. The wave model (SWAN) provides stationary results. However the hydrodynamic model (FINEL) needs time to get rid of spin-up (numerical oscillations) and to converge to an equilibrium. Assuming that, with constant forcing, an equilibrium exists we can assume that the equilibrium results are the correct stationary results. Figure 5.10 shows the time dependency of the absolute velocity and sediment fluxes for several locations and crosssections in vicinity of the reef. Note that the deviation from the mean is shown. The velocity shows many oscillations which are decaying in time. However even after a model execution of two days oscillations are visible and it seems that for some locations an equilibrium is still not reached. These are caused by eddies formed in front of the reef. However it is assumed that they have not significant effect on the processes behind the reef. Moreover the corresponding deviations are relatively small ($\sim$0.01 m/s). The sediment fluxes behind the reef show no more oscillations after two days and is converged to an equilibrium. Therefore the results used for this study are taken after a model execution of two days and considered as stationary results.

Offline modelling As mentioned; the total numerical model consists of two modules; the wave

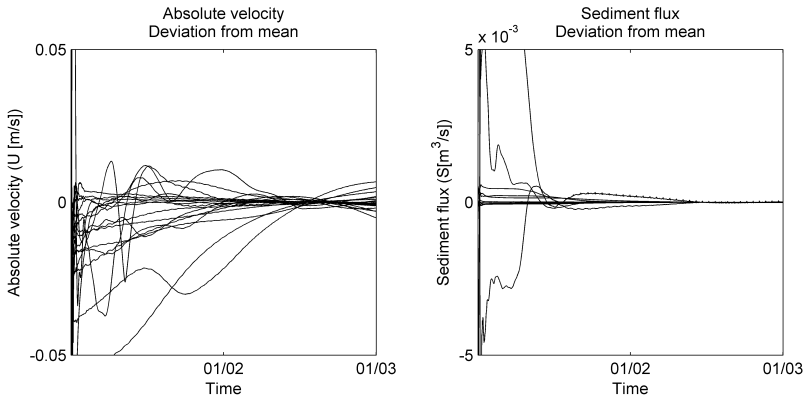

Figure 5.10: Time dependency of the absolute velocity and sediment flux for several locations and crosssections

module and hydrodynamic module. The results from the wave module (SWAN) are used as input for the hydrodynamic module (FINEL). However, the results from the hydrodynamic module are *not* put back as input for the wave module. This is called *Offline modelling*, which implies that there is no interaction of the (wave generated) hydrodynamics on the wave field. However in reality the wave generated set-up and current behind the reef change the conditions for the wave field, which in turn will have influence on the hydrodynamics. *Online modelling* therefore requires iteration to obtain an equilibrium, which means it will become computational inefficient. Online computations showed, however, no significant difference in the results. Therefore, and because of computational efficiency, offline calculations are performed in this study.

Spatial resolution Since numerical modelling uses spatial elements (triangles) the results and imposed conditions are actually discontinuous in space with a specified resolution. The element size varies through the model area, from an element size of about ∼1500 m far offshore to an element size of about ∼20 m in vicinity of the reef. The spatial resolution of the imposed bathymetry depends therefore on the resolution of the numerical grid. Moreover, FINEL2D imposes the depth at the middle of the elements while SWAN imposes the depth at the intersections of the elements. The imposed width of the reef top is chosen to be 20 m (see section 3.2) which is exactly the size of one grid element. The grid is adjusted to be sure that the reef top always will be exactly one element wide. This means that, both for FINEL2D and for SWAN, the reef top is modelled as a squared obstacle. For the higher parts of the reef the complete corresponding element is given a specified depth.

Hard layers Hard layers define area's in the model that are not allowed to erode. However, deposition on hard layers is possible. It is assumed that there is no sediment flux over the reef coming from offshore. In reality only suspended sediment that is available in the top layer of the water column will be transported over the reef. But since we are using a depth averaged model the sand concentration will be constant over the water depth, which implies that, although in reality the suspended (or bed load) concentration will be near the bottom, the model still show transport over the reef. Therefore the area offshore of the reef is defined as non-erodible. Only the area behind the reef and the area south of the gap are defined as erodible. Obviously the reef top itself is also modelled as a non-erodible feature.

Reef schematisation As mentioned in section 3.2 the location of the reef top is determined

with use of Google Earth Images and available survey data. The exact location is not known and therefore this schematisation can have effects on the output prediction. In the model schematisation the northern end of the reef is connected to the shore line based on the knowledge at that time. Current knowledge, however, confirmed a small gap is situated between the coast and the reef at the northern end. This means it is possible that in that area a northward flow, returning seaward, can occur, accompanied by a significant sediment flux. It is expected that this will not have any influence on the later discussed sediment fluxes near the southern gap. However it may have influence on the total sediment balance as more sediment may leave the system through the gap at the northern end. This effect will not be further mentioned in this analysis, but should be taken into account in further research.

Wind In this model only wind growth is included, therefore wind induced currents are not taken into account. The wind conditions are uniformly imposed over the complete model area, therefore land-ocean transformation of the wind is also not taken into account in the model. The wind data is offshore data so the conditions nearshore may be different and therefore may have influence on the results. Although it is expected that the wind growth is dominantly offshore and therefore nearshore conditions will not have significant effect on the waves nearshore. Furthermore it is assumed that wind induced currents are inferior to wave induced currents.

Sand constants In figure B.3 the sieve curves are shown of soil samples taken from different locations in vicinity of the reef. Clearly the grain size distributions differ spatially. Though it is assumed that the grain size distribution is constant over the complete model area. Note that this is only used for the erodible area's. The derivation of the sand constants can be found in section B.2.3.

Numerical errors Numerical modelling will always involve errors. For the simulation of the hydrodynamics a depth averaged two dimensional model is used (FINEL2D). This means that typical 3D effects, like return flow, are not taken into account. Furthermore, like mentioned in chapter B the numerical method guarantees strict mass and momentum conservation but suffers from numerical diffusion in stream wise direction. These effects are assumed to have negligible influence on the accuracy of the results.

Chapter 6

Results

In this chapter the results of the execution of the proposed probabilistic method, as described in the previous chapter are described. The statistical methods provided in appendix A are used to analyze the obtained data set. First the uncertainties of the several output variables are treated. Furthermore an estimation of uncertainty importance of the different input variables are given.

6.1 Input and focus

Since this study is focused on the applicability of the proposed probabilistic method and the computational effort has to be minimized, only one wave condition has been executed. The wave condition that is used to execute the method is wave condition 3 (see section 3.2 and appendix C, page 115). Wave condition 3 is a swell wave condition, although there is a (imposed) correlation with windspeed (figure D.3). Table 6.1 shows the probability distribution parameters that are used. The imposed correlation matrix is shown in equation (6.1), derived from the correlations shown in appendix D. This wave condition has weight $w = 0.26$, which means that this wave condition has an expected yearly occurence of 26%.

Variable	μ	σ
Wave height (H_s [m])	1.97	0.37
Wave direction (φ [degrees])	199.57	6.13
Wind speed (U [m/s])	7.56	1.40
Wind direction (ψ [degrees])	257.39	16.00
Peak wave period (T_p [s])	12.52	2.12
Roughness height (k_n [m])	0.13	0.05
Reef elevation (h_{reef} [- m+MSL])	0.50	0.25

Table 6.1: Probabilistic distribution parameters for model execution; Wave condition 3. All parameters are Gaussian distributed.

$$\rho_{condition3} = \begin{array}{c} \\ H_s \\ \varphi \\ U \\ \psi \end{array} \begin{array}{c} \begin{array}{cccc} H_s & \varphi & U & \psi \end{array} \\ \begin{bmatrix} 1 & 0 & 0.55 & 0 \\ 0 & 1 & 0 & 0 \\ 0.55 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \tag{6.1}$$

Since the model output consists of many (spatially varying) variables this study only focuses on the following important output variables;

Average set-up behind reef(η) The mean surface elevation behind the reef is calculated in the same way as for the sensitivity analysis (section 5.2.1).

Average velocity in polygon (u) The mean velocity is taken from the area shown in figure 5.4, the same as for the sensitivity analysis (section 5.2.1).

Sediment transport fluxes (S) The sediment transport fluxes are calculated using cross sections. Along the reef several longshore and cross shore sections are imposed; see figure 6.5. The sediment fluxes across these sections are measured and used for statistical analysis.

Total sediment balance (S_{tot}) Since the sediment fluxes are known along the reef the total sediment balance behind the reef can be determined.

6.2 Uncertainty propagation

In this section the results of *uncertainty propagation* are presented. The total influence of the uncertainties of the input and model parameters **X** on model output **Y** is shown in empirical (cumulative) probability distributions and different *estimators* (see appendix A).

6.2.1 Hydrodynamics

In this section the uncertainties of the hydrodynamic output variables are treated. Figure 6.1 and 6.2 shows the mean and relative standard deviation [1] of the hydrodynamics in vicinity of the reef. The relative standard deviation is relative constant over the area. On top of the reef the deviation is much higher because of the stochastic reef submergence. The reef top and the higher parts are able to become dry and combined with a small mean (μ) the relative standard deviation becomes large. The southern part has a mean water level (very) close to zero (MSL) which will lead to a (very) high relative standard deviation. The mean velocity (μ) in figure 6.2 shows expected results; a strong current on top and in front of the reef (northwards longshore current) and large rip-current at the gap of the reef. That current pattern is induced by the strong offshore current through the gap which is turned northwards by the longshore current more offshore. The (stochastic) spatial variability of that rip current causes a high relative standard deviation around the center of the large rip-current. Like for the surface elevation, the relative standard deviation is high at the higher parts of the reef. Also in areas with a low mean absolute velocity the relative standard deviation is high.

Empirical distributions

Figure 6.3 shows the empirical cumulative distributions of the average set-up behind the reef and the average velocity in the polygon as described in section 6.1. Included are the *bootstrap confidence intervals* (section A.3) and the bootstrap confidence interval width, which provide an indication of the reliability of the statistical results; i.e. statistical uncertainty. The empirical distributions are an estimation of the real distribution and since the sample size is limited statistical uncertainty is involved. This is measured by the bootstrap method, which implies resampling (with replacement) of the obtained data set where from the confidence intervals are determined, see also section section 6.2.4 and A.3. The confidence intervals become larger towards the tails

[1] Relative standard deviation $= \left|\frac{\sigma}{\mu}\right|$, which is a variant of the *coefficient of variation* (CUR, 1997)

of the distribution which is because of the diminished amount of data points at the tails. The estimated distribution parameters are shown in table 6.2. From the relative standard deviations it is clear that the uncertainties are not out of proportion. A difference in the median and mean may indicate a asymmetrical distribution. For the set-up a slight difference in the median and mean is found, however this is not significant. Figure 6.4 shows the *Kernel density estimation* (section A.5) with the corresponding output data points. Although the Kernel density function is positive for a surface elevation below zero, we know that the average set-up behind the reef will never be negative. As the data points confirm that last statement, the smoothing of the Kernel density estimation may give a distorted view. This does not affect the values of the estimators, since the Kernel density estimation is only used for graphical purposes. Following this we may conclude that the real distribution of the average set-up will be asymmetrical. The distribution of the velocity does look like it's symmetrical as may be hold for the real distribution. Beside for the extreme percentiles a good indication of the uncertainty of the hydrodynamics stemmed from the uncertainty of the input variables is given; i.e. the uncertainty propagation.

	Average set-up (η [m])	Average velocity (u [m/s])
Mean (μ)	0.066	0.432
Standard deviation (σ)	0.026	0.096
Median (50^{th} percentile)	0.065	0.434
Relative standard deviation $\left(\left\|\frac{\sigma}{\mu}\right\|\right)$	0.40	0.22

Table 6.2: Estimators of average set-up behind reef and average velocity in polygon

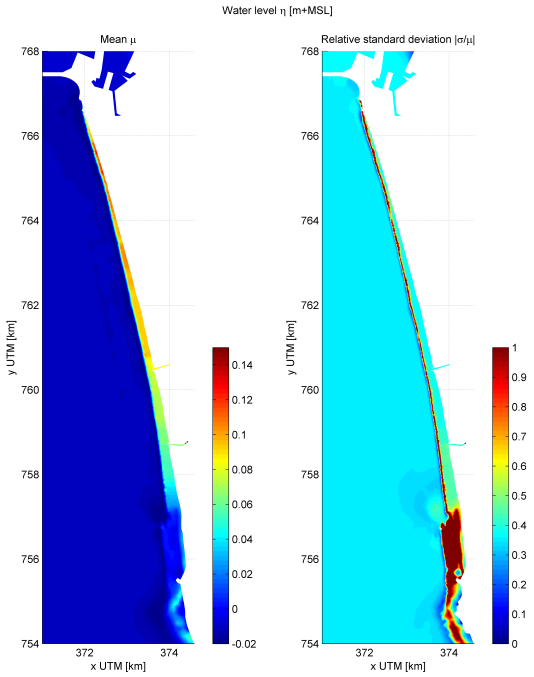

Figure 6.1: Plot of mean and relative standard deviation of the surface elevation [m] in vicinity of the reef

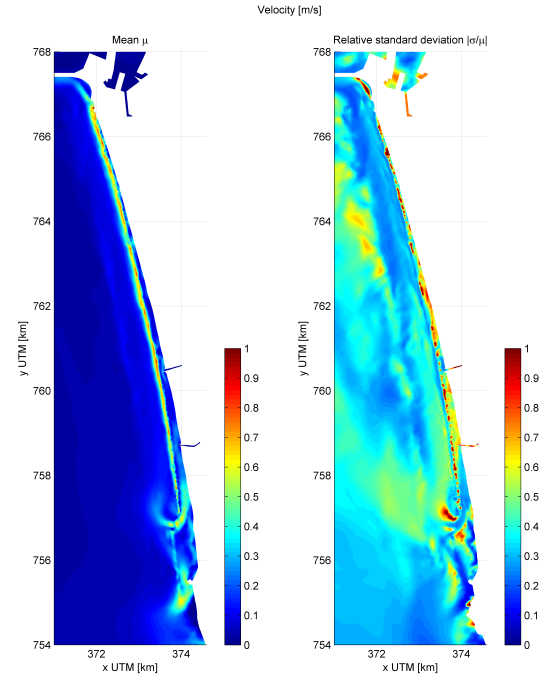

Figure 6.2: Plot of mean and relative standard deviation of the absolute velocity [m/s] in vicinity of the reef

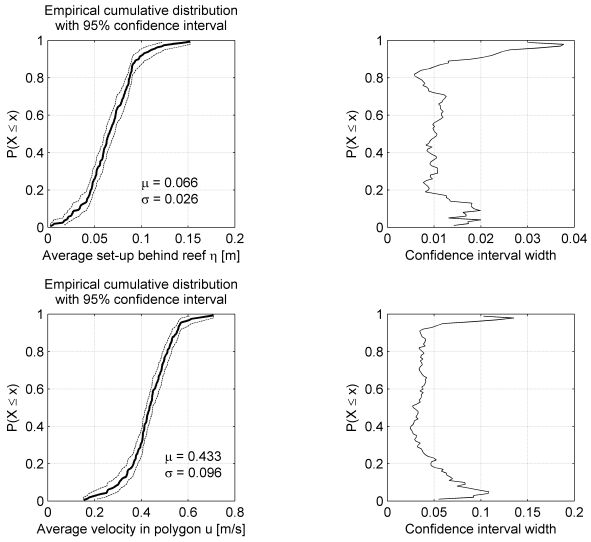

Figure 6.3: Empirical cumulative distribution plots for the average set-up behind reef and average velocity in polygon as described in section 6.1

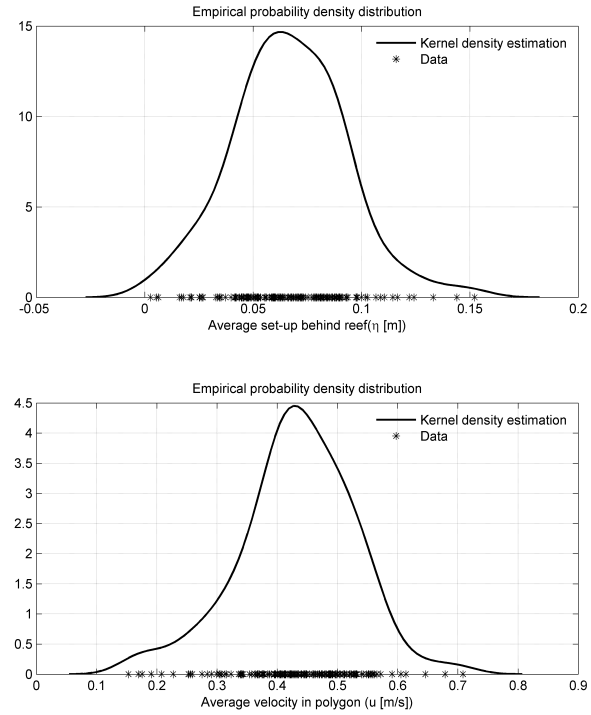

Figure 6.4: Empirical probability density distribution plots for the average set-up behind reef and average velocity in polygon as described in section 6.1

6.2.2 Sediment fluxes

In this section the uncertainties of the sediment fluxes are treated. Figure 6.7 shows the mean, median and the 5^{th} and 95^{th} percentiles for the cross shore and longshore sediment fluxes. As shown the uncertainty at the gap of the reef is significant compared to other locations. Although there is significant discharge at the northern part of the reef (figure 6.6) the threshold current velocity in equation (B.5) is preventing significant sediment fluxes. Despite the fact that the longshore discharge at the gap is always directed southwards, still northwards longshore sediment transport is occurring at that location. This phenomenon will be discussed later in section 6.3.1. Since the most notable sediment fluxes are at the gap of the reef we will focus on that location from now on.

Empirical distributions

The empirical distributions are determined from the cross- and longshore sections near the gap (red sections in figure 6.5). Figure 6.8 shows the empirical cumulative distributions (with the bootstrap confidence intervals) for the sediment fluxes near the gap. Notably is the small tail for the northwards longshore transport (negative values). Also here the confidence intervals are getting larger towards the tails of the distribution. The estimators of the distribution are shown in table 6.3. It is clear that the uncertainty is relatively large compared to the uncertainties from the hydrodynamics. The standard deviation of the longshore flux is even higher than its mean, resulting in a relative standard deviation greater than 1. This can also be seen in the kernel density estimations shown in figure 6.9. Both distributions have a long thick positive tail, mainly caused by extreme outliers. As for the hydrodynamics, the Kernel density estimation may give a distorted view for the crossshore sediment flux; no negative data points are found, though due to the smoothing the Kernel density estimation show a positive probability density for negative values. This does not affect the values of the estimators, since the Kernel density estimation is only used for graphical purposes. The density distribution of the longshore flux has at both sides a long thick tail because the output showed also some northward transport, while most of the transport is southward. From the difference between the mean and median and the kernel density estimations in figure 6.9 we may conclude that both distributions are asymmetrical.

	Longshore flux (S $[m^3/\text{day}]$)	Cross-shore flux (S $[m^3/\text{day}]$)
Mean (μ)	1135	2967
Standard deviation (σ)	1457	2212
Median (50^{th} percentile)	872	2555
Relative standard deviation $\left(\left\|\frac{\sigma}{\mu}\right\|\right)$	1.28	0.74

Table 6.3: Estimators of sediment fluxes near the gap

Conditional probability

From the empirical distributions of the longshore sediment flux it is clear that the (negative) outliers are (strongly) causing the very high degree of uncertainty. By looking at the distribution of the data points in figure 6.9 it looks like the (negative) northwards data points are not really part of the rest of the data set. Therefore we may conclude that the data set is not *homogeneous*. For good statistical analysis only homogeneous data sets may be used. A method to neglect the (negative) outliers is the use of *conditional probability*. That is the probability given a certain condition, denoted as; $P(x|A)$ where P is the probability of x, given condition A. From investigating the

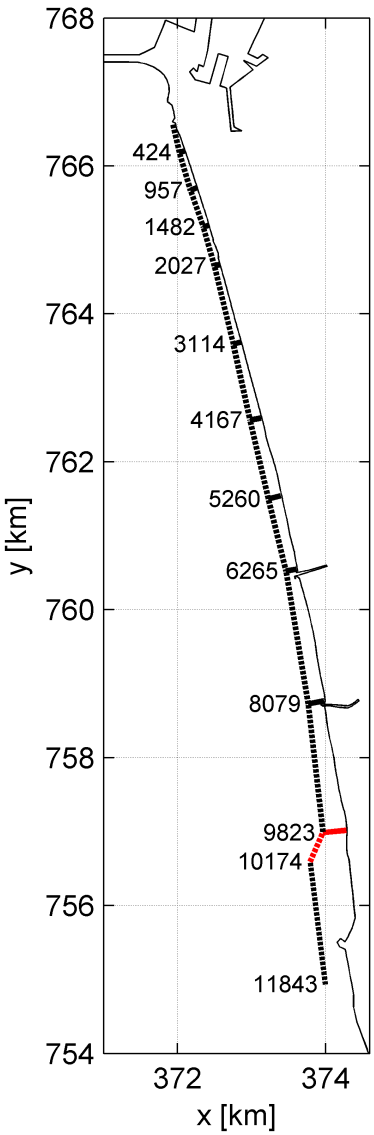

Figure 6.5: Location of cross sections for determination of sediment fluxes with in red the cross sections near the gap used for the determination of the empirical distributions

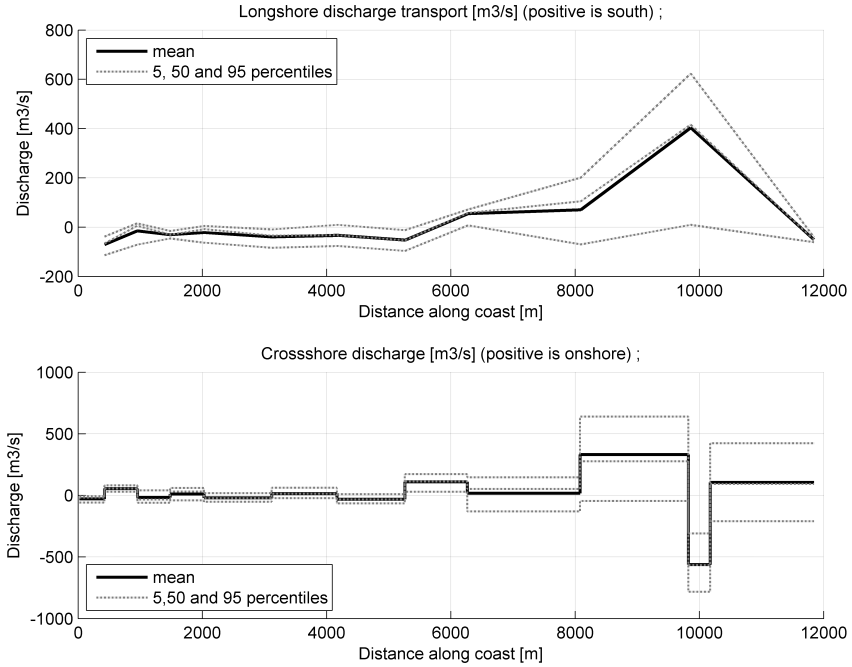

Figure 6.6: Mean, median and confidence intervals of the cross shore and long shore discharges; the gap is located at the distance of 10 km

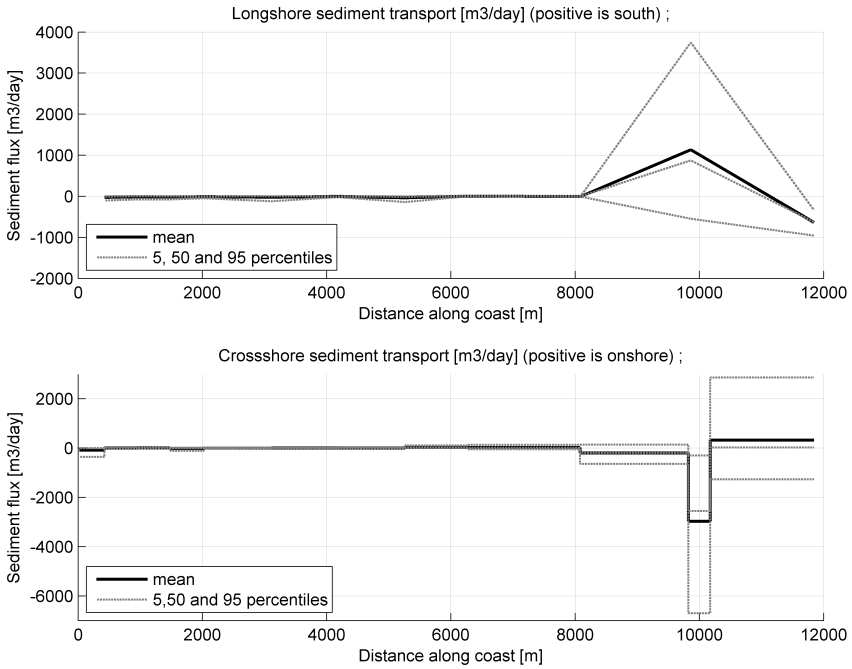

Figure 6.7: Mean, median and confidence intervals of the cross shore and long shore sediment fluxes; the gap is located at the distance of 10 km

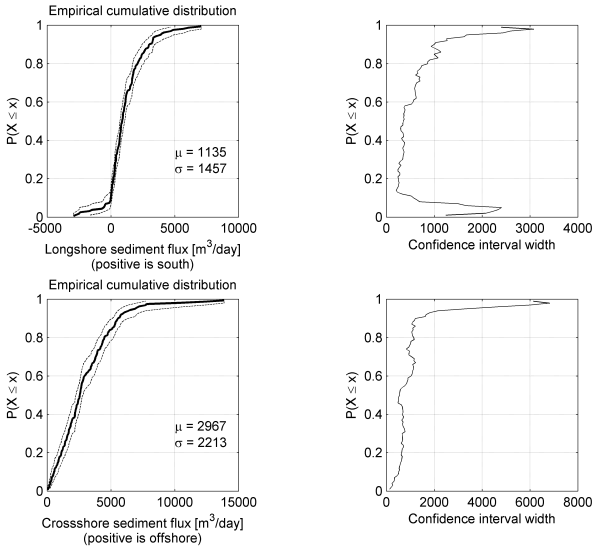

Figure 6.8: Empirical cumulative distribution including bootstrap confidence intervals for the cross- and longshore sediment flux at the gap

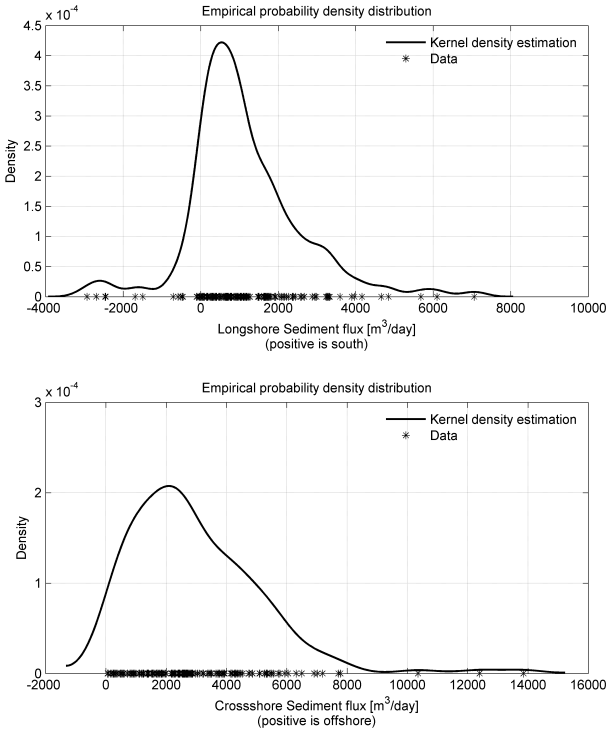

Figure 6.9: Empirical probability density distribution including bootstrap confidence intervals for the cross- and longshore sediment flux at the gap

scatter plots it became clear that the reef submergence was mostly responsible for the northwards longshore fluxes (see section 6.3.1). For negative reef submergence, so actually emergence of the reef, the longshore flux is always northwards (negative). Since we know that the reef is never totally emerged we now can determine a certain condition; $A = h_{reef} > 0$. The same plots are made as in the previous section but now given that the reef submergence is always positive; i.e. the reef is never fully emerged. Figure 6.10 and 6.11 show respectively the empirical cumulative distribution and kernel density estimation of the sediment fluxes given condition A. The left tail of the longshore sediment flux distribution is gone and the distributions of the longshore and crossshore sediment flux now look the same. This gives, almost certainly, a better representation of the real distribution of these variables. The distribution parameters are also determined following this condition and are shown in table 6.4. The relative uncertainty is clearly diminished for the longshore flux but still significant. The distribution for the cross-shore flux is not changed by this condition. More information about the involved processes see section 6.3.1. Also for these variables it holds that beside for the extreme percentiles a good indication of the uncertainty of the hydrodynamics stemmed from the uncertainty of the input variables is given; i.e. the uncertainty propagation.

	Longshore flux (S $[m^3/\text{day}]$)	Cross-shore flux (S $[m^3/\text{day}]$)
Mean (μ)	1323	2990
Standard deviation (σ)	1295	2227
Median (50^{th} percentile)	936	2547
Relative standard deviation $\left(\left\|\frac{\sigma}{\mu}\right\|\right)$	0.97	0.74

Table 6.4: Estimators of sediment fluxes near the gap given that reef submergence is positive

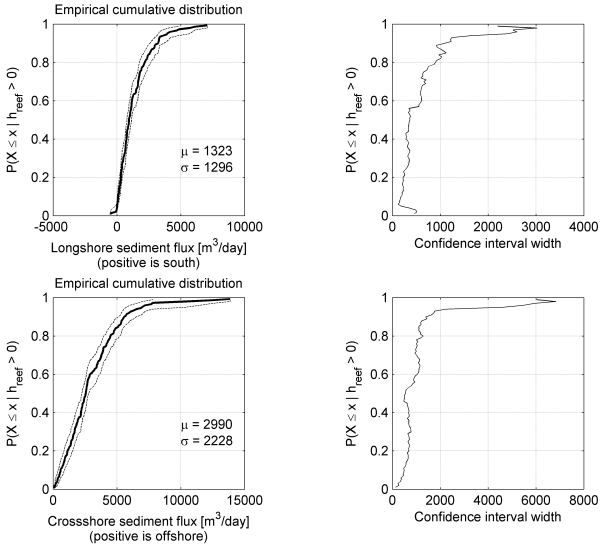

Figure 6.10: Empirical cumulative distribution including bootstrap confidence intervals for the cross- and longshore sediment flux at the gap given that the reef submergence is positive

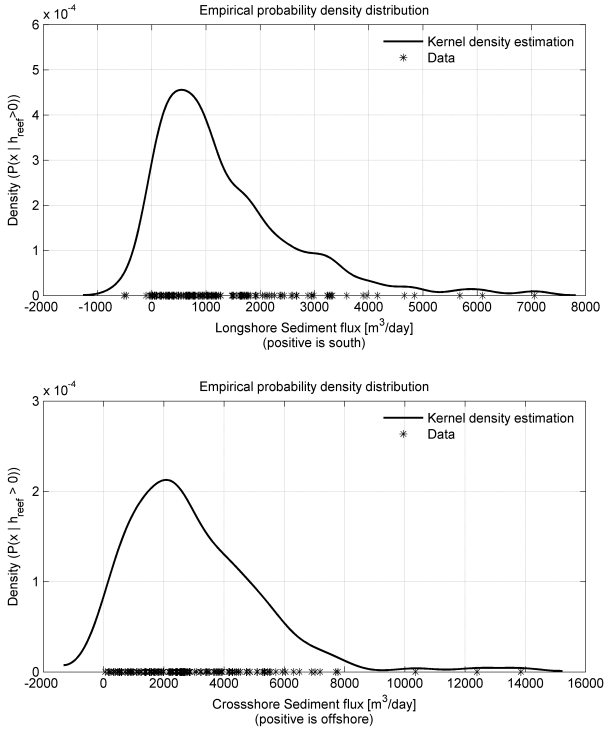

Figure 6.11: Empirical probability density distribution including bootstrap confidence intervals for the cross- and longshore sediment flux at the gap given that the reef submergence is positive

6.2.3 Total sediment balance

Since the sediment fluxes along the reef are known, the total sediment balance for the whole area can be determined. The fluxes of all the alongshore sections (dashed lines in figure 6.5) and the most southern cross section are added up to get an estimation of the total sediment balance [2]. A negative sediment balance therefore implies that sediment is leaving the system. Again the same plots are made for this output variable.

Empirical distributions

The empirical cumulative distribution and the Kernel density estimation are shown in figure 6.12. The statistical uncertainty for this variable is for some regions larger than for the previous discussed output variables. The cumulative distribution looks less smooth compared to the other examined output variables. This may be due to the fact that the total sediment balance is more dependent on the spatial variability of the results and therefore dependent on more factors compared to the other discussed variables. This causes more irregular variability in the output. And combined with random sampling therefore a more irregular distribution. The Kernel density estimation shows a multi modal density distribution. This might be caused by the same effect and coincidence because of random sampling. Therefore the expected (exact) distribution would probably be a smooth single modal distribution.

	Total sediment balance $[m^3/\text{day}]$
Mean (μ)	-2202
Standard deviation (σ)	1888
Median (50^{th} percentile)	-2037
Relative standard deviation $\left(\left\lvert\frac{\sigma}{\mu}\right\rvert\right)$	0.85

Table 6.5: Estimators of the total balance

Conditional probability

Like for the sediment fluxes conditional probability distributions, given that the reef submergence is positive, can be made. However significant difference in the results are not found. The Kernel density estimation shows sharper peaks, while the distribution parameters are staying almost constant. Probably due to the fact less data points have been used for the conditional distributions the peaks get sharper but not change the distribution parameters significantly. Therefore we may conclude that the real distribution will not change (significantly) due to this condition.

[2] Note that this includes the area south of the gap

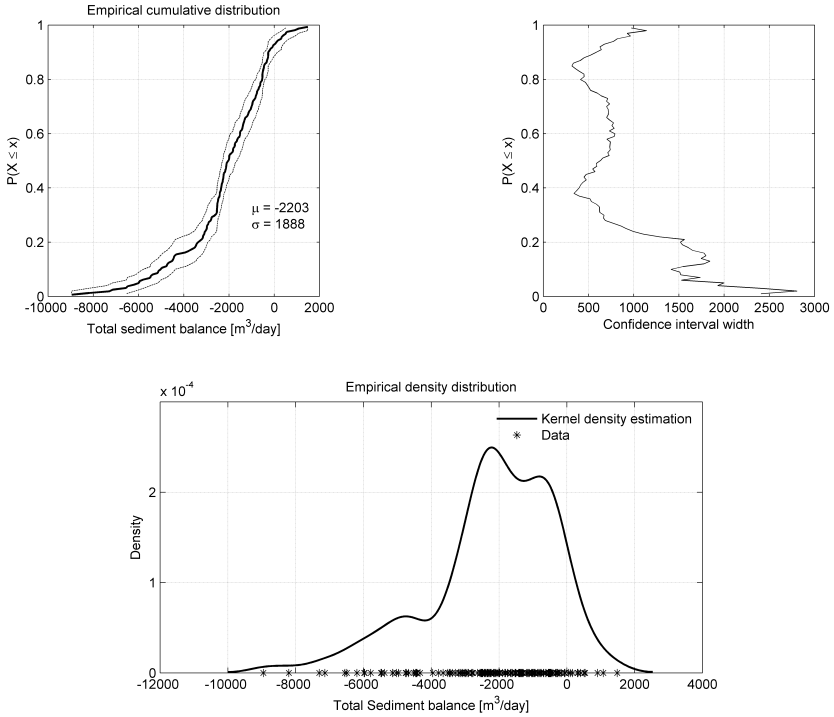

Figure 6.12: Empirical cumulative distribution including bootstrap confidence intervals and Kernel density estimation for the total sediment balance

6.2.4 Statistical uncertainty

Like mentioned earlier, the width for the bootstrap confidence interval shown in the cumulative distribution plots is a measure for statistical uncertainty. Statistical uncertainty gives insight in the accuracy and reliability of the statistical results. As discussed in section 4.3 the accuracy is dependent of the sample size used for the Monte Carlo analysis.

Estimation of the mean

The accuracy of the estimation of the mean of the output distributions can be estimated using the expressions provided in appendix A. There we found that due to the *central limit theorem* the distribution of the mean, with large sample size m, is normally distributed with a mean equal to its *sample mean*($\bar{x}$) and a standard deviation equal to equation (A.3);

$$S_m = \frac{\hat{\sigma}}{\sqrt{m}} \tag{6.2}$$

Where $\hat{\sigma}$ is the estimated standard deviation of the output distribution. For the estimation of the 95% confidence interval for the mean, deviation c in equation (A.4) is equal to 1.96 (≈ 2). Using equation (A.5) we find that the total width of the confidence interval (CI) for the mean is therefore estimated by;

$$CI_\mu \approx \frac{4\hat{\sigma}}{\sqrt{m}} \tag{6.3}$$

So with an increasing sample size the confidence interval width is decaying with $1/\sqrt{m}$. In figure 6.13 the confidence interval width is shown as a function of sample size m for the estimation of the mean of the total sediment balance. In this figure the estimated confidence interval width according to equation (6.3) is plotted. The estimated standard deviation for the output distribution ($\hat{\sigma}$) differs as the sample size is increasing, therefore the plot does not show a smooth curve. Figure 6.13 shows that equation (6.3) show very good correspondence with the estimated confidence intervals by the bootstrap method. And so the bootstrap method seems to be an appropriate method to estimate confidence intervals. This relationship can also be used to estimate the required sample extension if more accurate results are desired.

Estimation of percentiles

The accuracy of the estimation of percentiles can also be estimated using the expressions provided in appendix A. There we found that the estimate of a *percentile* of the output distribution seems to be *binomial distributed*, which in turn can be approximated by a Gaussian distribution with mean $\mu_p = p$ and standard deviation $\sigma_p = \sqrt{\frac{p(1-p)}{m}}$. Where p is the percentile. Using again equation (A.4) the 95% confidence interval is estimated with the following percentiles of the output distribution;

$$\left(p - 2\sqrt{\frac{p(1-p)}{m}}, p + 2\sqrt{\frac{p(1-p)}{m}} \right) \tag{6.4}$$

For the estimation of the 10^{th} percentile p is set equal to 0.10. For the determination of the confidence interval the percentiles in equation (6.4) are obtained from the empirical (cumulative) distribution. Figure 6.14 shows the confidence interval width estimated using the bootstrap method and from the percentiles in equation (6.4). The bootstrap confidence interval shows similar results compared to the confidence interval described by the binomial distribution. Therefore

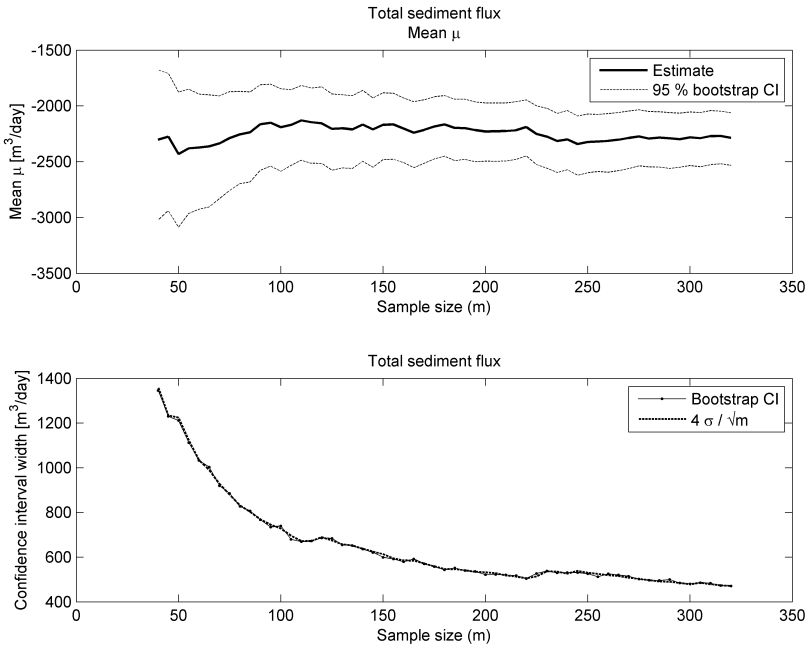

Figure 6.13: Top: Estimation of mean of total sediment balance as function of sample size m including estimation of the 95% bootstrap confidence interval. Bottom: Estimation of the confidence interval width by the bootstrap method and by equation (6.3) using the estimated standard deviation ($\hat{\sigma}$)

it seems that the bootstrap method is an appropriate method to estimate the confidence interval for the percentiles. Also this relationship can be used to estimate the required sample extension if more accurate results are desired.

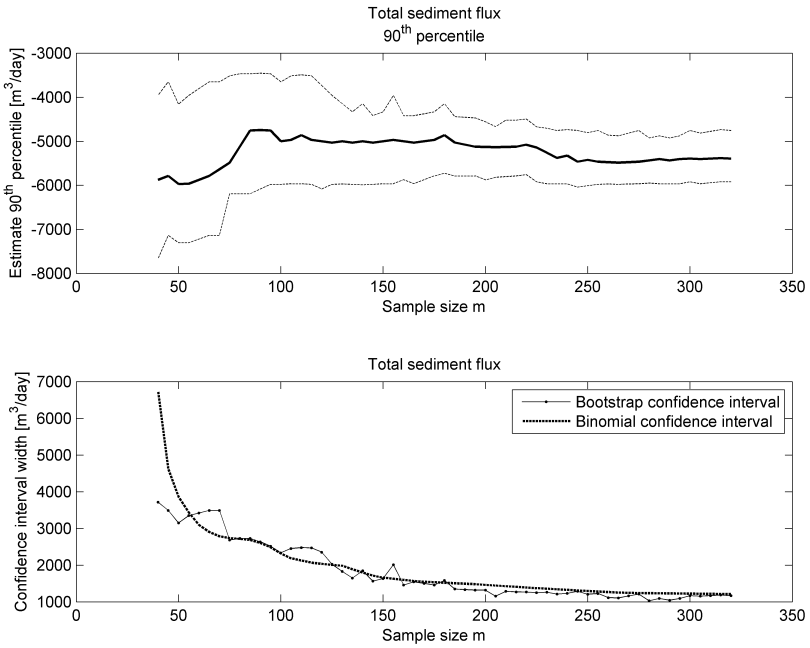

Figure 6.14: Top: Estimation of the 10^{th} percentile of the total sediment balance as function of sample size m including estimation of the 95% bootstrap confidence interval. Bottom: Estimation of the confidence interval width by the bootstrap method and by the binomial distribution, see equation (6.4).

6.3 Uncertainty importance

The uncertainty importance is a measure for the influence of the uncertainty of a specific input variable on the uncertainty of the output variable. Note that this is not the same as the sensitivity of the variable, which is - in contrast to the uncertainty importance - independent of the degree of uncertainty of the input variable. Hamby (1994) has made distinction between *important* parameters whose uncertainty contributes substantially to the uncertainty in the model results and *sensitive* parameters which have a significant influence on the model results. Important parameters are always sensitive but sensitive parameters may not be important parameters since they may be known exact and therefore have no contribution to the output uncertainty. Nevertheless sensitivity of the input variables is a good measure for their uncertainty importance. Many sensitivity analysis techniques are available (Hamby, 1994; Helton and Davis, 2003; Helton et al., 1985; Morgan and Henrion, 1990). Some of those methods are treated in this section.

6.3.1 Scatter plots and (rank) correlation

To understand the relationship between input and output it is often useful to examine scatter plots. These can provide a lot of insight, show nonlinear effects, thresholds and so on (Morgan and Henrion, 1990). For every output variable determined in section 6.1 scatter plots are shown which will give insight in the relationships. Later on these scatter plots are used to get insight in the hydrodynamic and morphological processes and to certify the observed relationships.

Hydrodynamics

In figures 6.15 and 6.16 the scatter plots are shown of the hydrodynamic output variables $\mathbf{Y}$, determined in section 6.1, versus the stochastic input variables $\mathbf{X}$. The plots give quick insight in the relationships. As can be seen the wave height and the wind speed are (far) more correlated with the output variables compared to the other variables. The wind speed is correlated with the wave height, which can explain the relative large correlation for the wind speed. Neglecting outliers, the wave height seems to have a non-linear (quadratic) relationship with the set-up, while for the velocity the relationship looks more linear. This corresponds with the expected sensitivity derived in section 5.2.1.
Estimates for the *rank order correlation coefficients* are given in tables 6.6 and 6.7. These coefficients are obtained given condition A; so given that reef is never totally emerged. Rank order correlations, also known as the *Spearman's* correlation, are a measure for the strength of monotonic relationship, whether they are linear or not (Morgan and Henrion, 1990). Normal (Pearson) correlation coefficients are a measure for the strength of linear relationship, however since we are dealing with non-linearities rank order correlations are used instead. More information about correlation can be found in section A.2. The bootstrap confidence intervals are measure for the uncertainty of the determination of the rank correlation coefficient. It is clear from the obtained values that the most correlated variables are the wave height, wave direction and the wind speed. The rank correlation coefficient between the velocity and the reef elevation is slightly positive, however we know from the sensitivity analysis that the reef elevation must have a negative correlation. The scatter plots of the reef elevation shows that the data points with a reef elevation close to zero show for both the set-up and the velocity (very) low values and that the maximum values are occurring with a reef elevation close to 0.50 metres. This phenomenon is discussed later in section 6.3.2. Moreover, the calculated correlation coefficients will never be exactly zero, therefore (very) low correlations are not reliable and may be caused by coincidence. This can also be confirmed by examining the bootstrap confidence intervals.

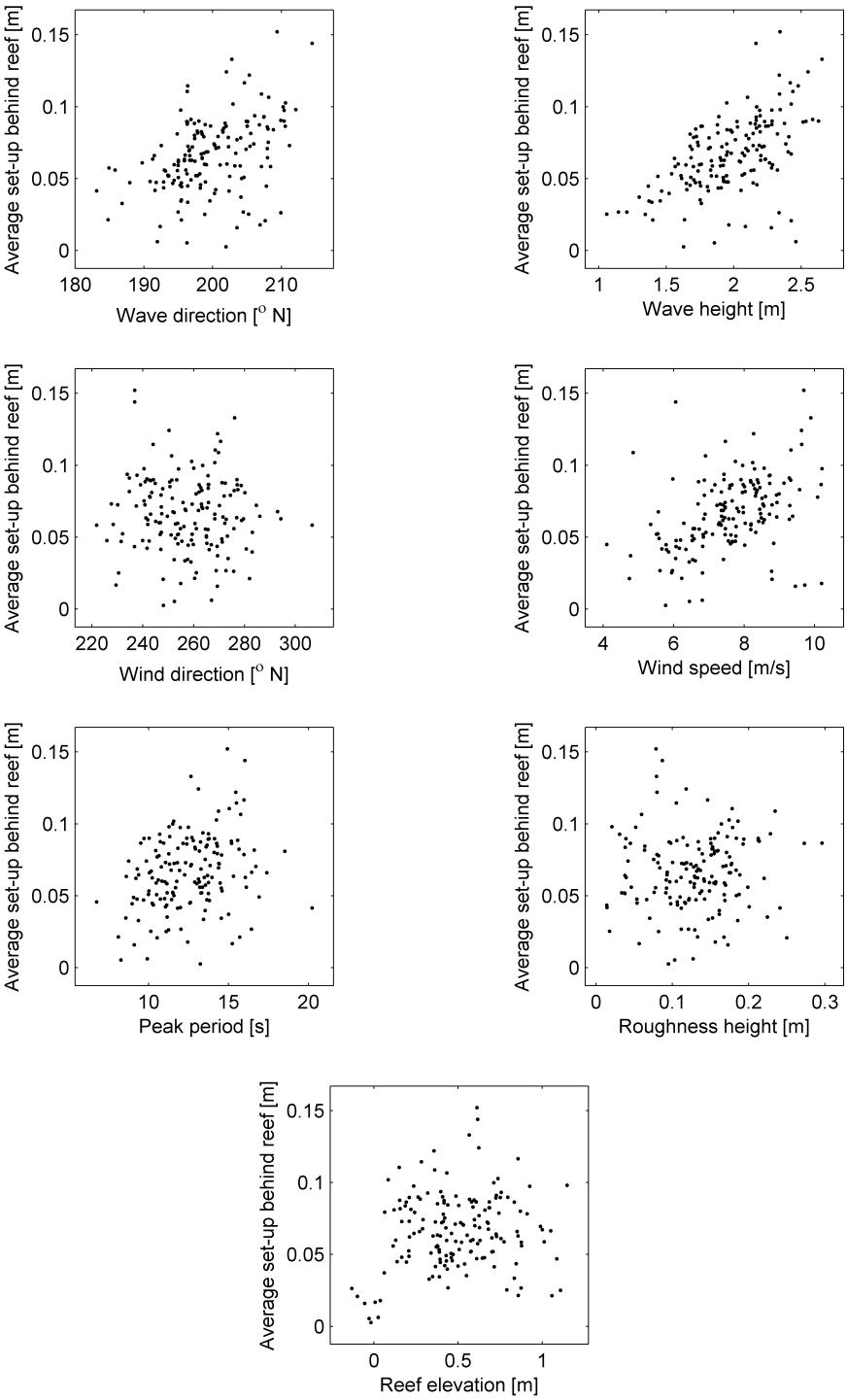

Figure 6.15: Scatter plots for average set-up behind the reef versus the stochastic input variables

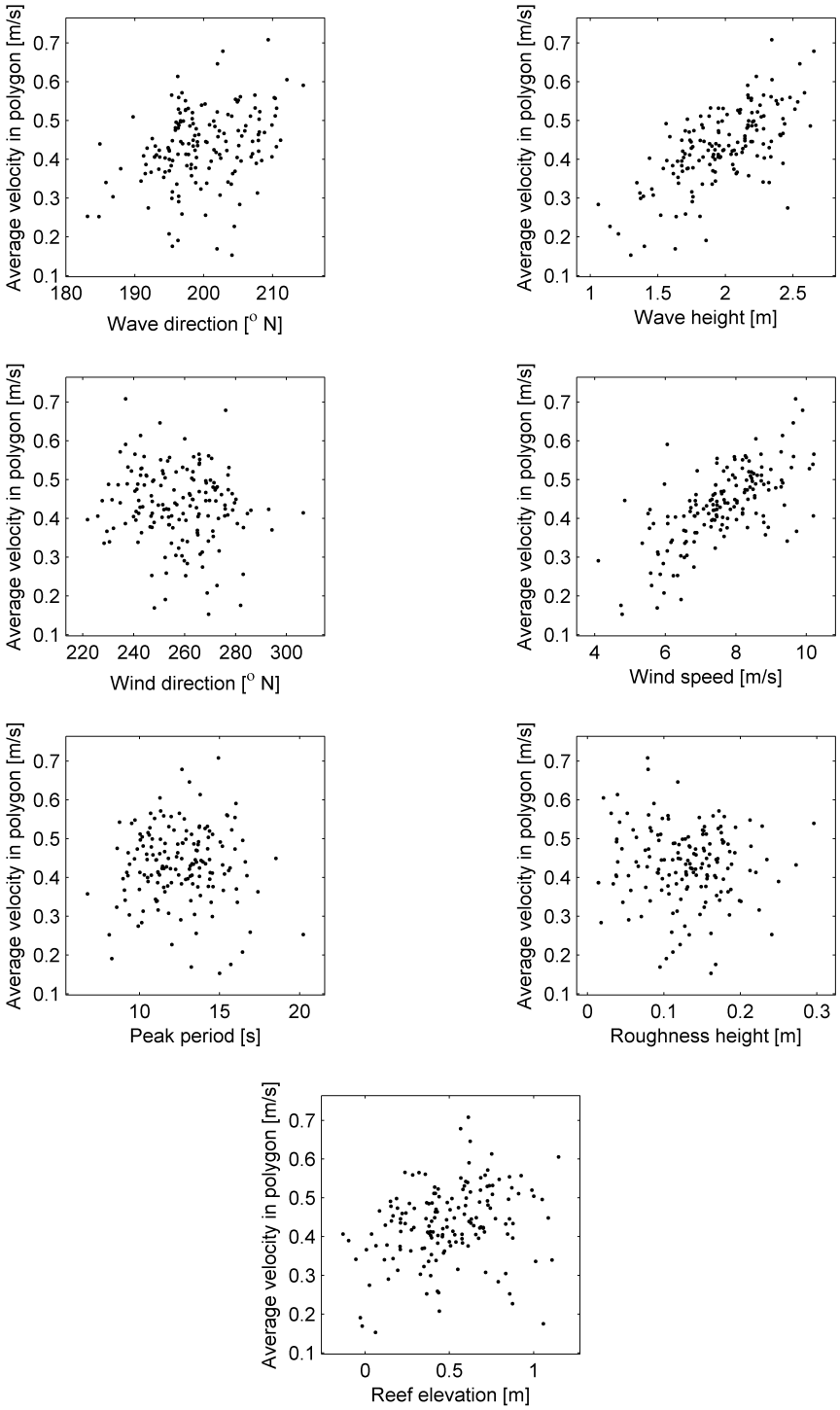

Figure 6.16: Scatter plots for average velocity in polygon versus the stochastic input variables

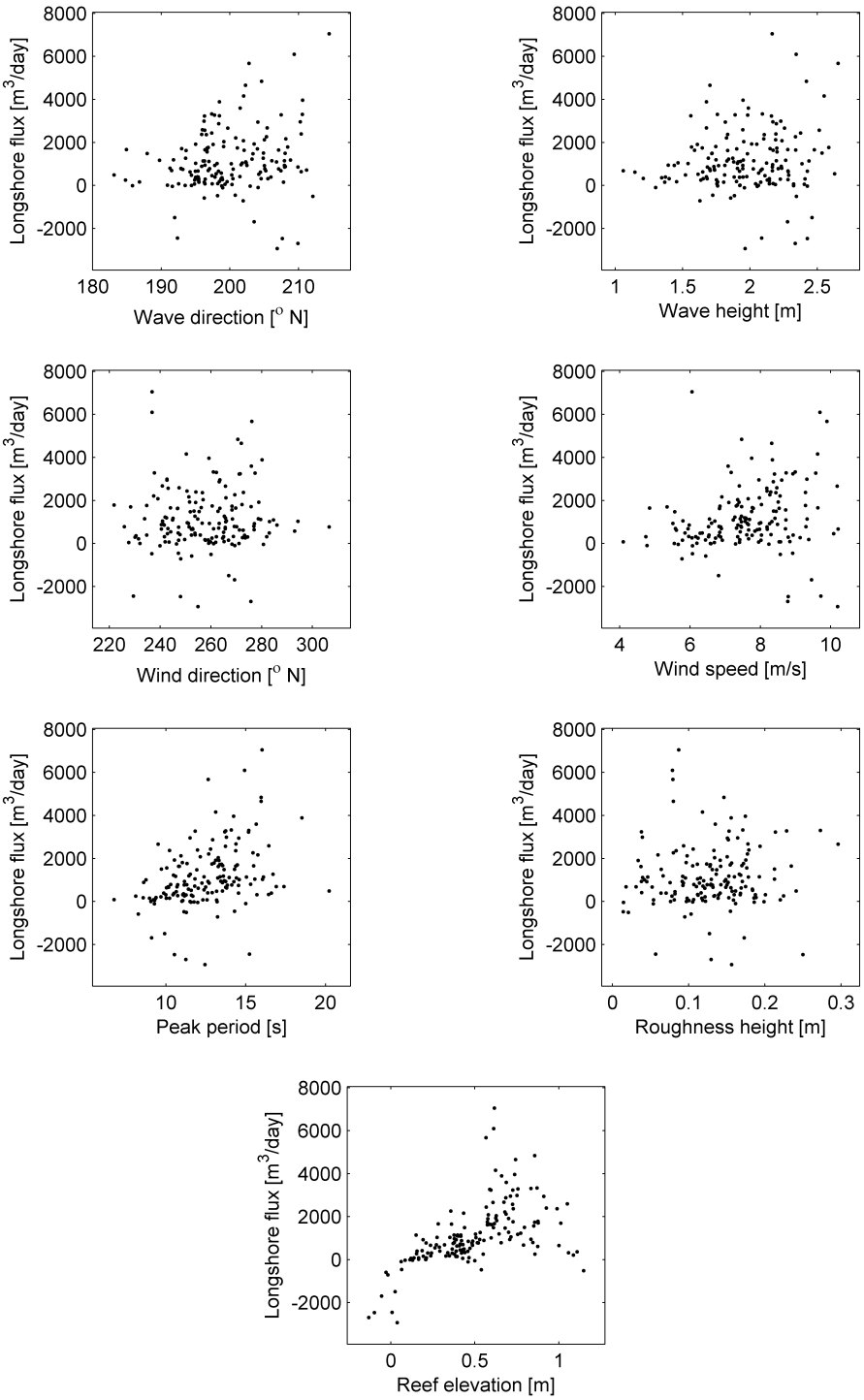

Figure 6.17: Scatter plots for longshore sediment flux near the gap versus the stochastic input variables; positive is southwards

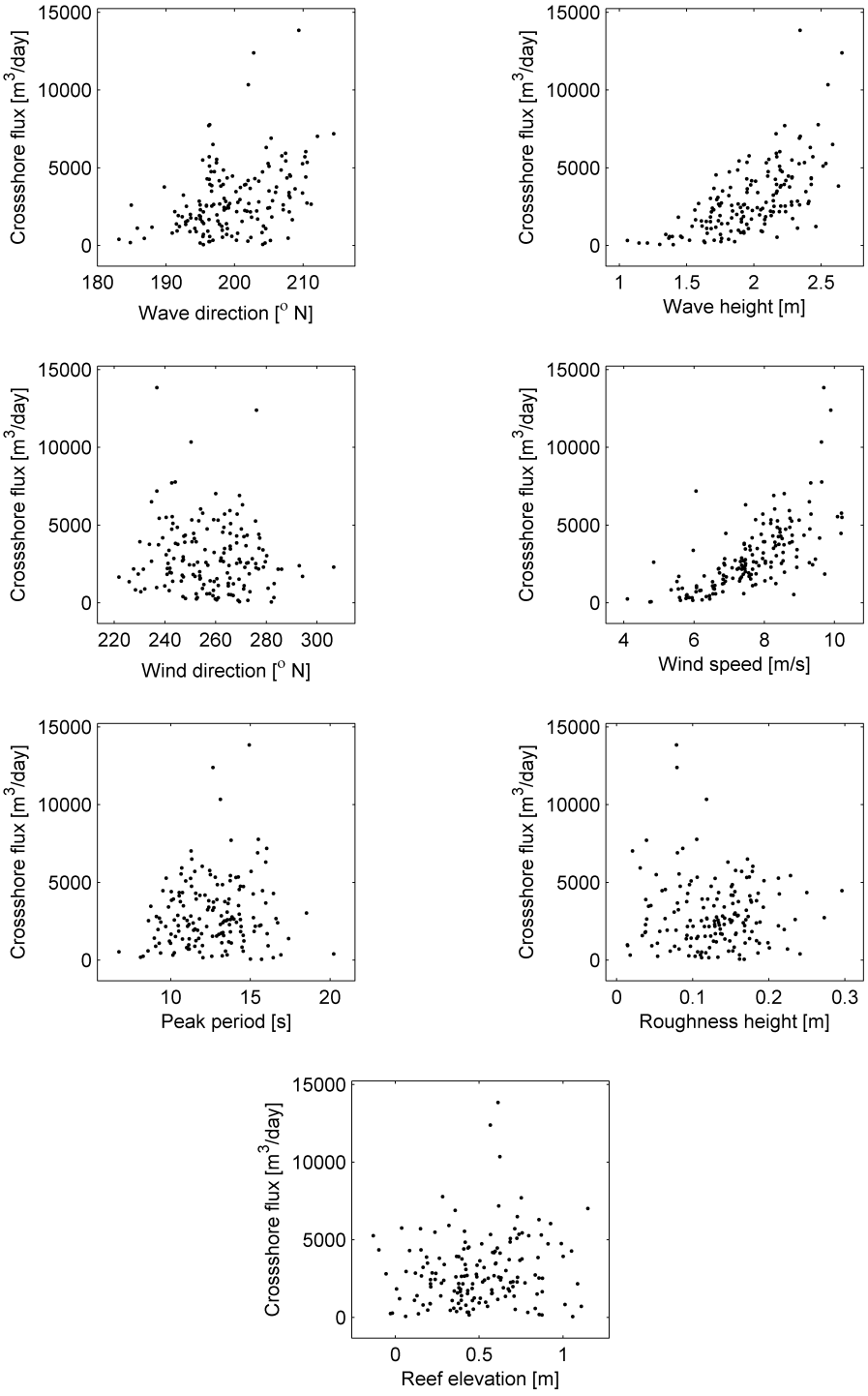

Figure 6.18: Scatter plots for crossshore sediment flux near the gap versus the stochastic input variables; positive is offshore

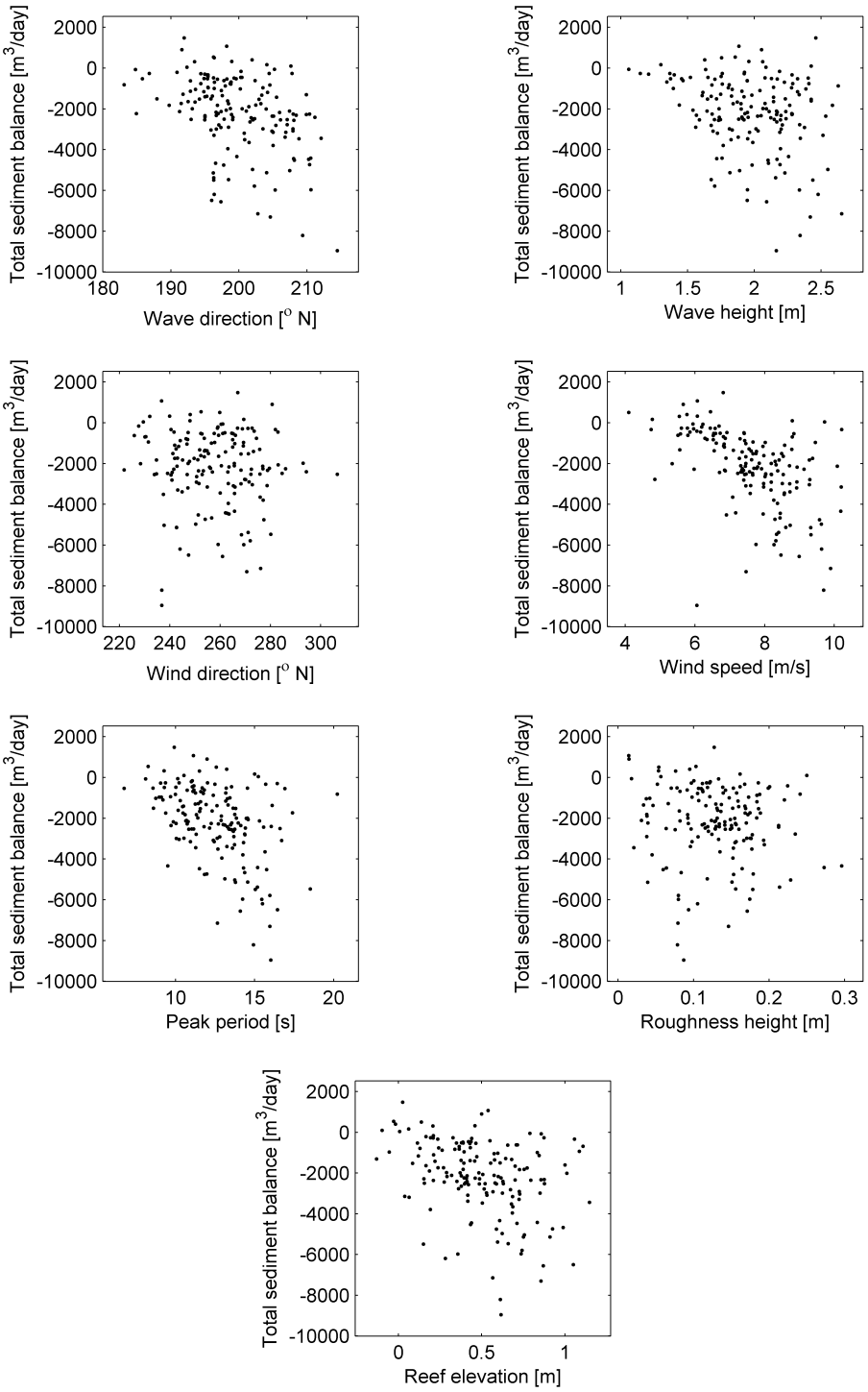

Figure 6.19: Scatter plots for the total sediment balance versus the stochastic input variables

Rank correlation; Average set-up behind reef			
Variable	Best estimate	Bootstrap confidence intervals	
Wave direction	0.515	0.379	0.632
Wave height	0.657	0.548	0.748
Wind direction	-0.084	-0.243	0.079
Wind speed	0.545	0.391	0.661
Peak period	0.262	0.101	0.410
Roughness height	0.095	-0.084	0.264
Reef elevation	-0.037	-0.203	0.127

Table 6.6: Rank correlation coefficients with bootstrap confidence intervals for the average set-up behind reef, given that reef submergence is positive

Rank correlation; Average velocity in polygon			
Variable	Best estimate	Bootstrap confidence intervals	
Wave direction	0.367	0.219	0.501
Wave height	0.679	0.566	0.763
Wind direction	-0.121	-0.273	0.031
Wind speed	0.715	0.600	0.797
Peak period	0.014	-0.153	0.186
Roughness height	-0.054	-0.225	0.118
Reef elevation	0.202	0.030	0.365

Table 6.7: Rank correlation coefficients with bootstrap confidence intervals for the average velocity in polygon, given that reef submergence is positive

Sediment fluxes

In figures 6.17 and 6.18 the scatter plots are shown of respectively the longshore and crossshore sediment flux near the gap versus the stochastic input variables. The scatter plots for the longshore sediment flux differs a lot with the scatter plots for the crossshore sediment flux. The relationships for both output variables are different indicating that also the driving processes differ from each other. In tables 6.8 and 6.9 the estimated rank correlation coefficients are shown for the sediment fluxes. Clearly the correlation of the wave- and wind variables for the longshore sediment flux are (very) small compared to those for the crossshore sediment flux. However the interpretation of this rank correlation coefficient is important. For example, the wave height has a negligible correlation for the longshore flux, but as can be seen in the scatter plot there is definitely a dependency; large (absolute) fluxes are only found with large wave heights. Though the calculated rank correlation coefficient is close to zero. However we may conclude that the forcing variables are less dependent for the longshore sediment flux than for the crossshore sediment flux. The peak period and reef elevation are far more correlated for the longshore sediment. Moreover, the reef elevation has a large positive correlation for the longshore flux, while we would expect a negative correlation (section 5.2.1). The scatterplot shows a unexpected non monotonic relationship. However it is clear that a total reef emergence, always shows a northward sediment flux. In section 6.3.2 this relationship is discussed in detail.

Total sediment balance

Also for the total sediment balance scatter plots are made and are shown in figure 6.19. At first sight it looks like the scatter plots are a combination of the scatter plots obtained for the longshore

Rank correlation; Longshore sediment flux near the gap		
Variable	Best estimate	Bootstrap confidence intervals
Wave direction	0.303	0.155 0.438
Wave height	0.140	-0.023 0.295
Wind direction	0.049	-0.111 0.213
Wind speed	0.363	0.201 0.509
Peak period	0.453	0.309 0.574
Roughness height	0.121	-0.051 0.288
Reef elevation	0.570	0.419 0.682

Table 6.8: Rank correlation coefficients with bootstrap confidence intervals for the longshore sediment flux near the gap, given that reef submergence is positive

Rank correlation; Crossshore sediment flux near the gap		
Variable	Best estimate	Bootstrap confidence intervals
Wave direction	0.379	0.232 0.506
Wave height	0.667	0.560 0.752
Wind direction	-0.090	-0.243 0.064
Wind speed	0.737	0.619 0.818
Peak period	0.122	-0.044 0.288
Roughness height	0.011	-0.162 0.181
Reef elevation	0.147	-0.024 0.307

Table 6.9: Rank correlation coefficients with bootstrap confidence intervals for the crossshore sediment flux near the gap, given that reef submergence is positive

and crossshore sediment fluxes. Which may be obvious since the total sediment balance is highly dependent on the sediment fluxes near the gap (figure 6.7). The rank correlation coefficients are given in table 6.10. Remarkable is the relative large correlation of the wave period, like we already saw for the fluxes. The scatter plots and rank correlation coefficients show that the wind direction and roughness height are both variables that do not have a clear relationship nor a large influence on the output.

Rank correlation; Total sediment balance		
Variable	Best estimate	Bootstrap confidence intervals
Wave direction	-0.446	-0.566 -0.309
Wave height	-0.288	-0.431 -0.127
Wind direction	-0.102	-0.263 0.058
Wind speed	-0.612	-0.724 -0.463
Peak period	-0.453	-0.581 -0.307
Roughness height	-0.135	-0.298 0.041
Reef elevation	-0.310	-0.457 -0.147

Table 6.10: Rank correlation coefficients with bootstrap confidence intervals for the total sediment balance, given that reef submergence is positive

6.3.2 System schematisation

From the scatter plots and obtained rank correlation coefficients the following relationships and phenomena are observed;

- For all output variables the roughness height and wind direction are not (significant) correlated nor show a clear relationship.
- For the hydrodynamics, the forcing variables; wave height, wind speed and wave direction are the most correlated variables and show an expected relationship (section 5.2.1).
- For all output variables the reef elevation showed a positive correlation, while a negative correlation is expected. Particularly for the longshore sediment flux, the reef elevation showed an unexpected relationship.
- The forcing variables, except for the wave period, are more correlated for the crossshore flux than for the longshore flux. While the reef elevation and wave period are more correlated for the longshore flux than for the crosshore flux.
- The scatter plots for total sediment balance looks like a combination of the scatter plots of the longshore and crossshore sediment fluxes.

To get more insight in the observed relationships, the hydrodynamic and morphodynamic processes are examined. As mentioned, the scatter plots for the reef elevation, particularly for the longshore sediment fluxes, shows an unexpected relationship. Therefore we are going to have a closer look into that scatter plot. Figure 6.20 shows again the scatter plot for the longshore sediment flux versus the reef elevation. Now the different markers denote a specific wave direction sector and the thick black line shows the *running average*[3]. The different markers show that (approximately) the same relationship holds irrespectively of the wave direction. The same plots using subsets of other input variables show similar results.
After an analysis of the data set and corresponding running average it is clear that the relationship shows some thresholds, which are denoted as dashed lines in figure 6.20. These thresholds correspond exactly with respectively the submergence of the total reef and the submergence of the higher parts[4]. To understand the processes the system is divided in different stages following the thresholds found in the scatter plots;

Stage 1	$h_{reef} < 0$	Total reef emerged
Stage 2	$0 < h_{reef} < 0.5$	Total reef submerged, higher parts emerged
Stage 3	$h_{reef} < 0.5$	Total reef submerged

For each stage the processes are described in detail with use of the scatter plots in figures 6.23 and 6.21. Note that an additional velocity measure is used in figure 6.21-b. The average velocity in polygon (fig. 6.21-c) is the mean velocity taken from the polygon defined in figure 5.4; which is a measure for the (crossshore) velocity in the gap. The average velocity behind the reef (6.21-b) is the average velocity taken from the area behind the reef (near the gap); which is a measure for the longshore velocity behind the reef (near the gap). The velocity squared (U^2) is used to obtain

[3]The running average, also known as a *moving average*, is a filter used to analyze a set of data points by creating a series of averages of different subsets of the full data set. In this case the average of a subset of 15 data points is used.

[4]The submergence of the total reef and the higher parts are respectively 0 and 0.5 metres below MSL, however taking the dry cell criterion of the model into account (0.05 m) the model thresholds become slightly higher

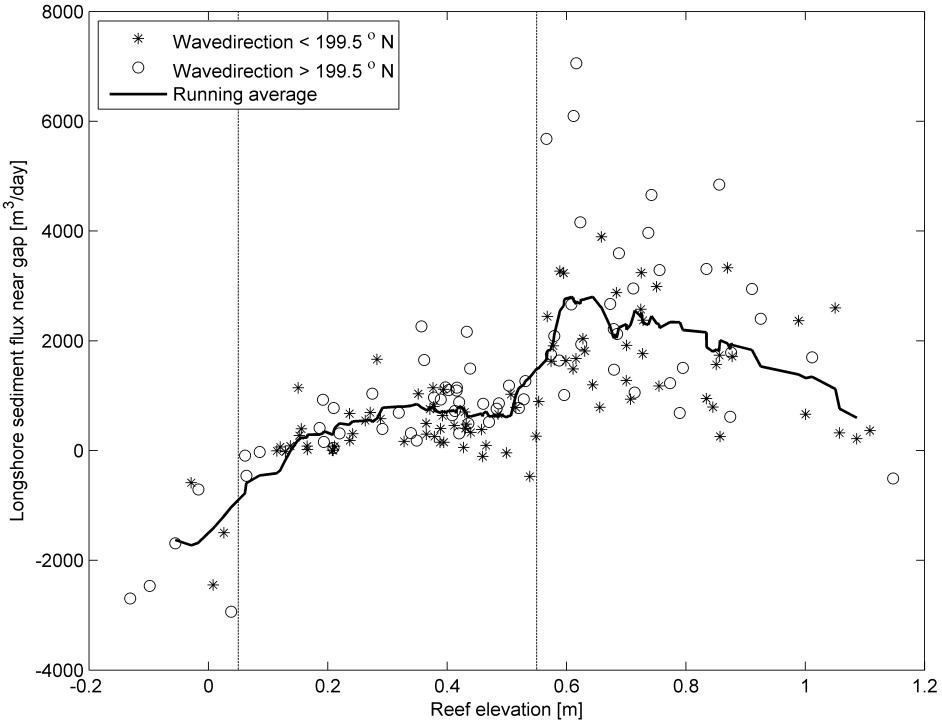

Figure 6.20: Scatter plot for the longshore sediment flux near the gap versus the reef elevation; including separation of two wave direction sectors and the running average

the influence on the sediment suspension and therefore on the sediment flux. [5]. The term between the squared brackets, including its power, in the sediment transport formula (equation (B.5)) of Soulsby-Van Rijn, determines the equilibrium concentration (c_{eq}). The concentration is therefore dependent on both the wave orbital velocity term and the current velocity term. The terms that describe their leverage to the equilibrium concentration are therefore as follows;

$$\text{Current velocity term} = \bar{U}^2 \tag{6.5}$$

$$\text{Wave orbital velocity term} = \frac{0.018}{C_D} U_{rms}^2 \tag{6.6}$$

However from figure 6.22 it is clear that the current velocity term (eq. 6.5) always dominates the wave orbital velocity term (eq. 6.6) for the suspended sediment concentration behind the reef. The current velocity term becomes even more dominant for the total sediment flux because of the entrainment of suspended sediment by this same current velocity.

[5]Since the bed load term is significant smaller than the suspended sediment flux , here we focus only on the suspended sediment concentration. Though the results presented in this report includes bed load transport

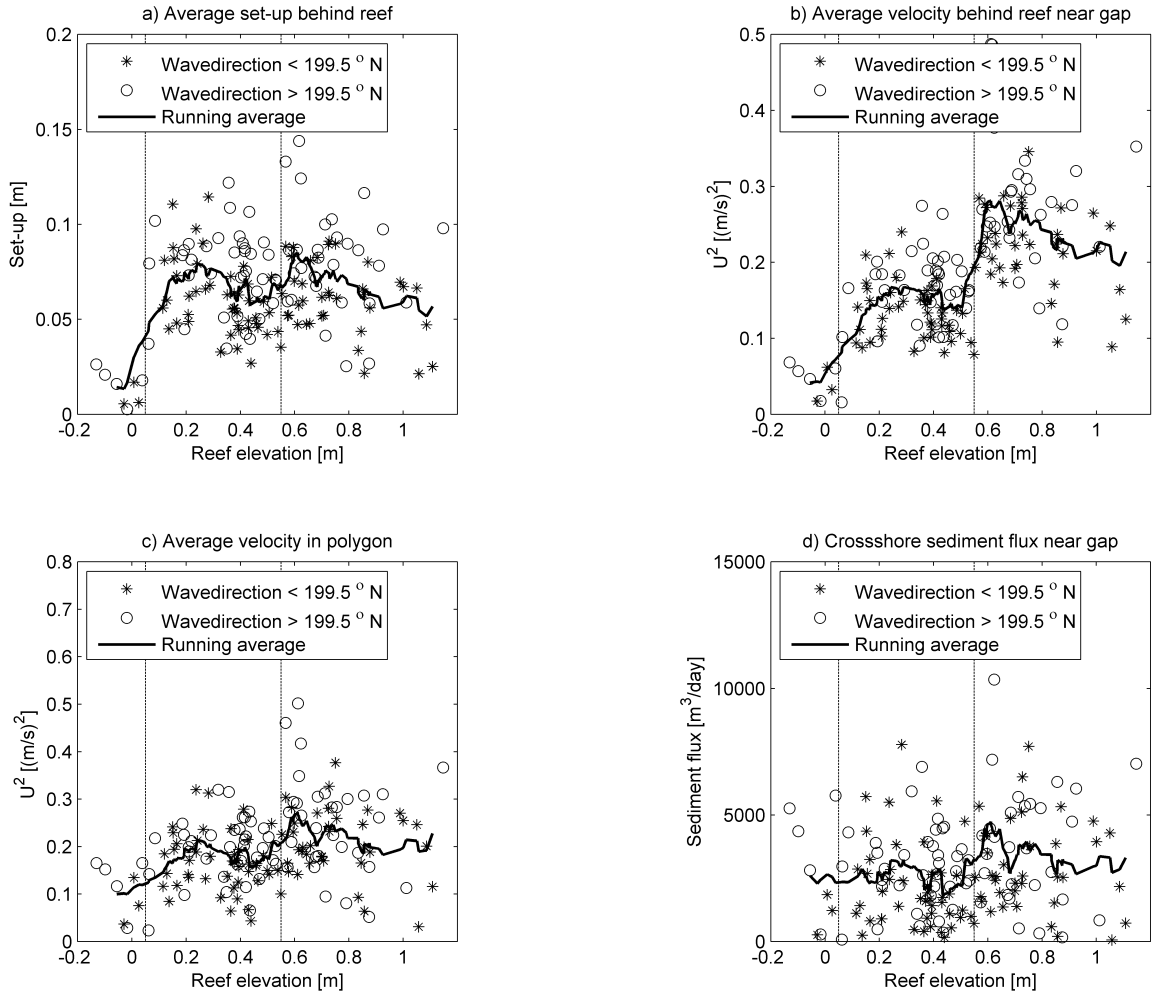

Figure 6.21: Scatter plots for the reef elevation versus the average set-up, average velocity behind the reef, average velocity in polygon and the crosshore sediment flux; including separation of two wave direction sectors and the running average

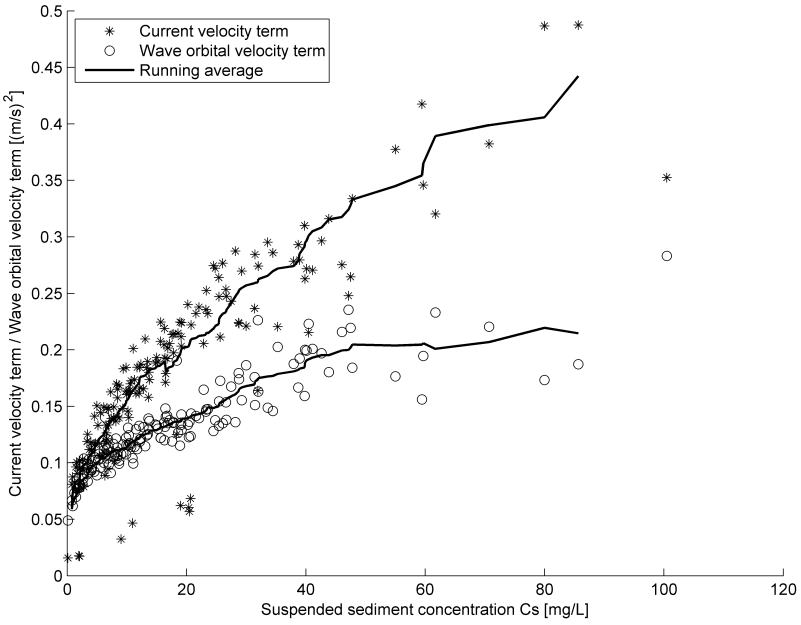

Figure 6.22: Current velocity term and wave orbital velocity term versus the suspended sediment concentration behind the reef; including their running average

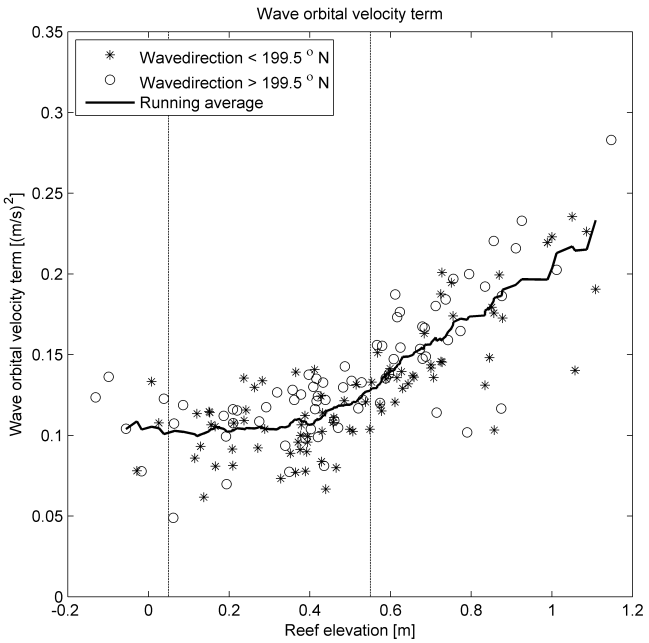

Figure 6.23: Scatter plot for the reef elevation versus the wave orbital velocity term (eq. 6.6); including separation of two wave direction sectors and the running average

Stage 1; Total emergence

In figure 6.24 a schematisation of the system in stage 1 is shown, so a situation where the total reef is emerged. Figure 6.25 shows the flow field and suspended sediment concentration of an example modelled scenario with a system in stage 1. The following phenomena are observed;

- Due to the total emergence of the reef there is no interaction possible over the reef.
- This means that no (significant) wave induced set-up is generated behind the reef. This is shown in figure 6.21-a.
- Since there is no set-up no (significant) velocity is visible. See also the plot in figure 6.21-b.
- This means that there is no current velocity and negligible wave orbital velocity, because there is no wave energy behind the reef. [6] behind the reef no sediment is suspended behind the reef.
- The oblique incident waves (green arrows) induce a northward longshore current along the entire coast (red arrows).
- The northward longshore current travels behind the reef (red arrows). However because of continuity the flow has to bend back southwards.
- Wave breaking at the shore line south of the gap and a significant northward longshore current brings sediment in suspension. As clearly can be seen in figure 6.25.
- Because of the suspended sediment in the south and no suspension behind reef there is a nett northwards longshore sediment flux behind the reef. See also the plot in figure 6.20.
- The northward flow is bended offshore through the gap (red arrows) which is shown in the plot in figure 6.21-c.
- The offshore flow through the gap results in a offshore directed crossshore sediment flux through the gap, which is shown in figure 6.21-d.

From this we may conclude that if the reef is totally emerged it leads to a nett northwards longshore sediment flux behind the reef and a offshore directed crossshore sediment flux.

[6] The wave orbital velocity term is positive with an emerged reef (figure 6.23 which can be explained by the fact the waves are oblique incident to the coast so wave energy may travel northward behind the reef

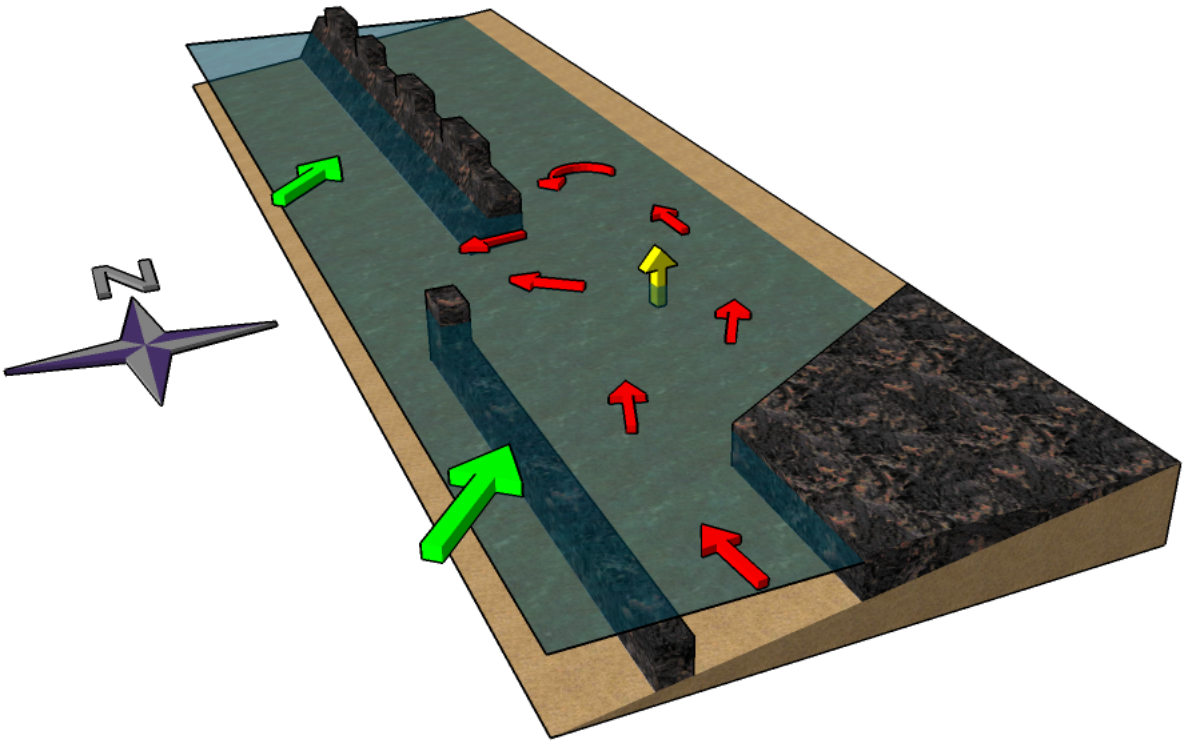

Figure 6.24: Schematisation of stage 1; green arrows = wave energy, red arrows = current velocity field, yellow arrows = sediment suspension

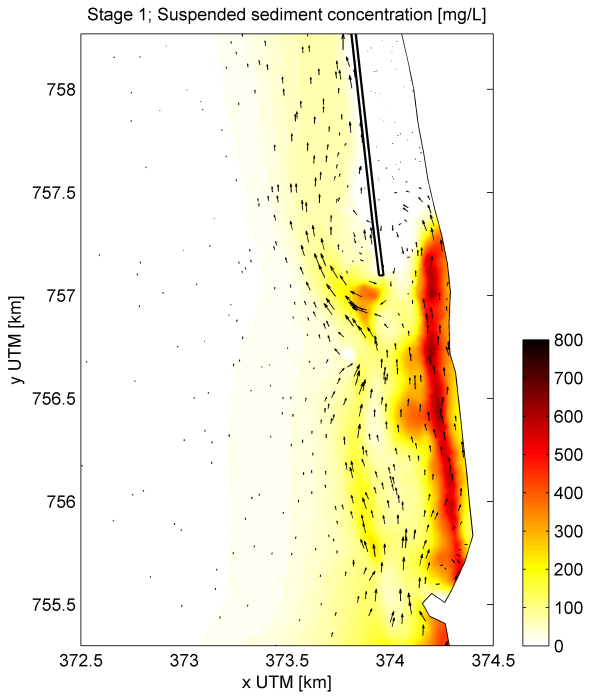

Figure 6.25: Flow field and suspended sediment concentration of an example scenario in stage 1

Stage 2; Partly submergence

In figure 6.26 a schematisation of the system in stage 2 is shown, so a situation where the reef is submerged but the higher parts are emerged. Figure 6.27 shows the flow field and suspended sediment concentration of an example modelled scenario with a system in stage 2. The following phenomena are observed;

- As the reef is just below MSL;
  - Waves are breaking on top of the reef and as the water column on top of the reef is increasing the energy and discharge over the reef increases (red arrows).
  - Therefore there is an increase in the average set-up behind the reef, which is shown in the plot in figure 6.21-a.
  - This causes an increase in the velocity behind the reef (red arrows) as can be seen in the plot in figure 6.21-b.
  - Now the velocity is increased sediment is brought into suspension (yellow arrow)
  - This results in a southward longshore sediment flux, see also the plot in figure 6.20.
  - The southward directed flow is mixed with northward flow and is pushed offshore through the gap. Therefore the velocity in the gap is slightly increased as can be seen in the plot in figure 6.21-c.
  - The longshore sediment flux behind the reef is small compared to the sediment flux coming from the south, because more sediment is brought in suspension at the south (yellow arrows). Therefore no significant difference is found in the crosshore (offshore) sediment flux in figure 6.21-d
- As the reef gets lower;
  - The water column above the reef is increasing even more and therefore less wave energy is dissipated on top of the reef and part of the wave energy is traveling over the reef and is dissipated at the shore line. A slight increase of the wave orbital velocity term in figure 6.23 can be found.
  - This leads to a decrease of the set-up which can be found in the plot in figure 6.21-a.
  - The velocity behind the reef is decreasing and so is the velocity in the gap (polygon), see also the plot in figure 6.21-b,c.
  - Although the wave orbital velocity behind the reef increases the decreasing current velocity dominates the decrease in sediment suspension (figure 6.22).
  - However it leads to just a slight decrease in the longshore sediment flux, see figure 6.20.

Although it expected that the maximum set-up will occur with a reef submergence just below MSL (Loveless et al., 1998), the model limits the discharge over the reef and therefore the maximum occurs with a higher submergence. As the reef gets lower the set-up decreases and therefore the velocity and sediment flux decreases. So as the reef is just below MSL the increase of mass transport over the reef is dominating and as the reef is getting lower the decrease of wave energy is dominant.

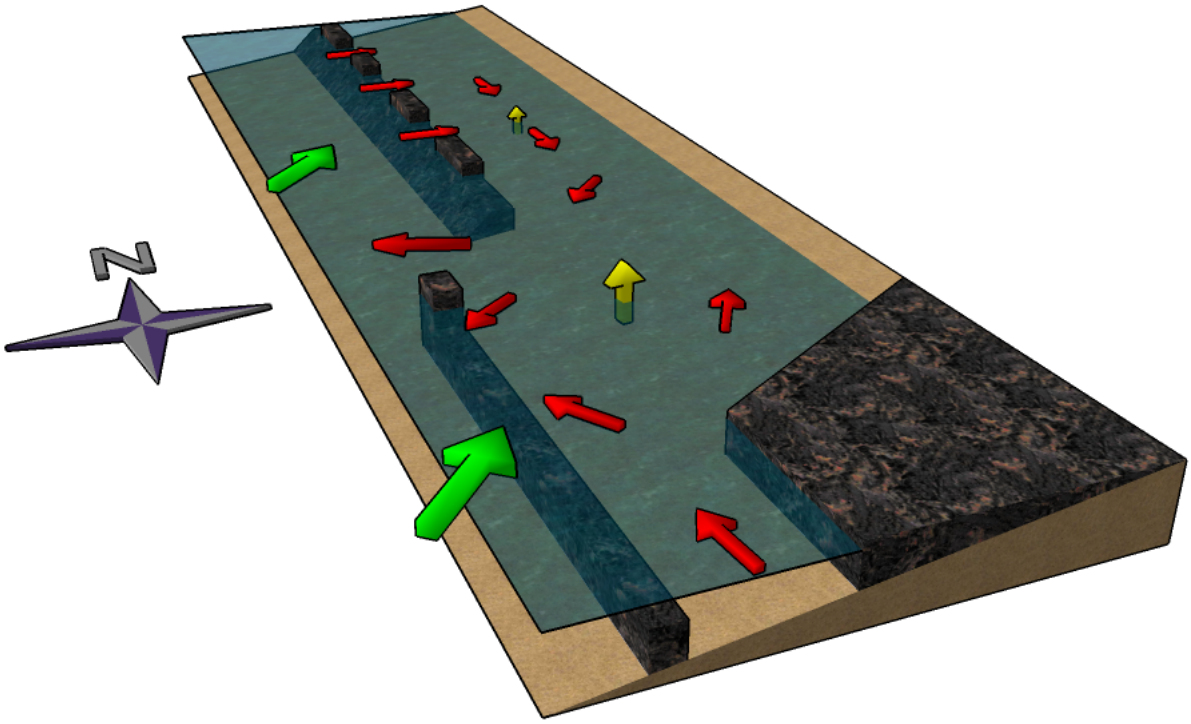

Figure 6.26: Schematisation of stage 2; green arrows = wave energy, red arrows = current velocity field, yellow arrows = sediment suspension

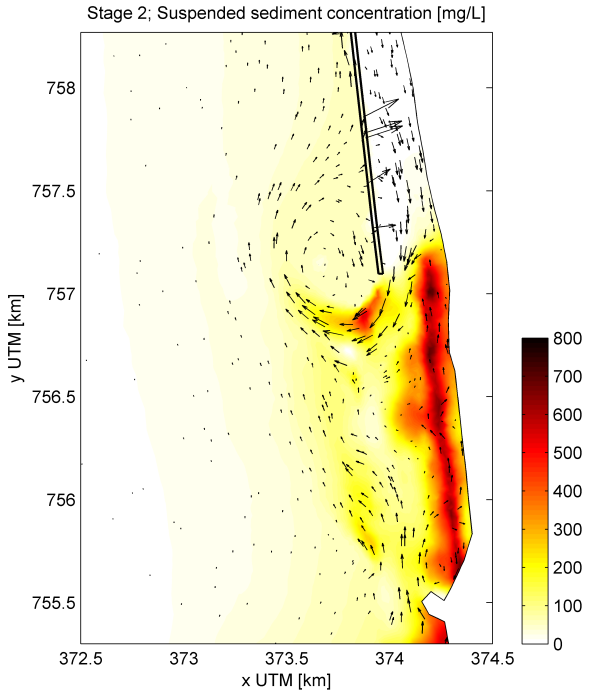

Figure 6.27: Flow field and suspended sediment concentration of an example scenario in stage 2

Stage 3; Total submergence

In figure 6.28 a schematisation of the system in stage 3 is shown, a situation where the reef is totally submerged. Figure 6.29 shows the flow field and suspended sediment concentration of an example modelled scenario with a system in stage 3. The following phenomena are observed;

- As the higher parts of the reef are just below MSL;
  - Compared to stage 2 the cross sectional area is suddenly increased a lot, because now everywhere along the reef interaction is possible.
  - Both the discharge and the dissipated wave energy increases over the reef.
  - This leads to an increase of the set-up, as shown in the plot in figure 6.21-a.
  - Because the set-up increases and the mass transport over the reef (red arrows) that has to be drained from the system is increased a lot, the velocity increases even more than the set-up. This is shown in the plot in figure 6.21-b.
  - The southward current is mixed with the northward current from the south and pushed through the gap which results in a notable peak for the velocity in the gap (polygon) in figure 6.21-c.
  - With an increase of the velocity and the wave orbital velocity behind the reef more sediment is suspended by the reef, which is visible in figure 6.29
  - This results in a sharp increase in the longshore sediment flux as can be seen in figure 6.20.
  - The increase of the longshore sediment transport and the increase in of the velocity in the gap results in a notable peak for the crossshore sediment flux in figure 6.21-d.
- As the reef gets lower;
  - The water column above the reef is increasing and therefore less wave energy is dissipated on top of the reef and part of the wave energy is traveling over the reef and is dissipated at the shore line. Which can be seen by the increase of the wave orbital velocity term in figure 6.23.
  - The set-up decreases as shown in the plot in figure 6.21-a. This results in a decreasing velocity behind the reef and in the gap as can be seen in figure 6.21-b,c.
  - Although the wave orbital velocity term is increasing (figure 6.23) the current velocity term is dominating (figure 6.22) and leads to a decrease of the suspended sediment.
  - This leads to decrease of the longshore sediment flux shown in figure 6.20.
  - The velocity behind the reef can get even that low that the northward longhore current dominates which can result in a nett northward longshore transport behind the reef.

From these observations we may conclude that the submergence of the higher parts of the reef and the total reef results in a sharp increase of the longshore sediment transport, which is even notable in the crossshore sediment flux. The increase in mass transport over the reef and the increase of the wave energy dissipation results in the increase of velocity and sediment transport. However as the reef gets lower the decrease of the wave energy dissipation dominates the increase of the water column on top of the reef and leads to lower velocities and therefore for a decrease of sediment flux. By investigation of the processes divided in the three stages it became clear that the reef has, beside the expected relationship (section 5.2.1 and Loveless et al. (1998)), a discrete effect on the output variables. It has more influence on the longshore sediment flux than on the crossshore sediment flux. Moreover the forcing variables are therefore less important for the longshore transport than for the crossshore transport, which was already discussed in section 6.3.1.

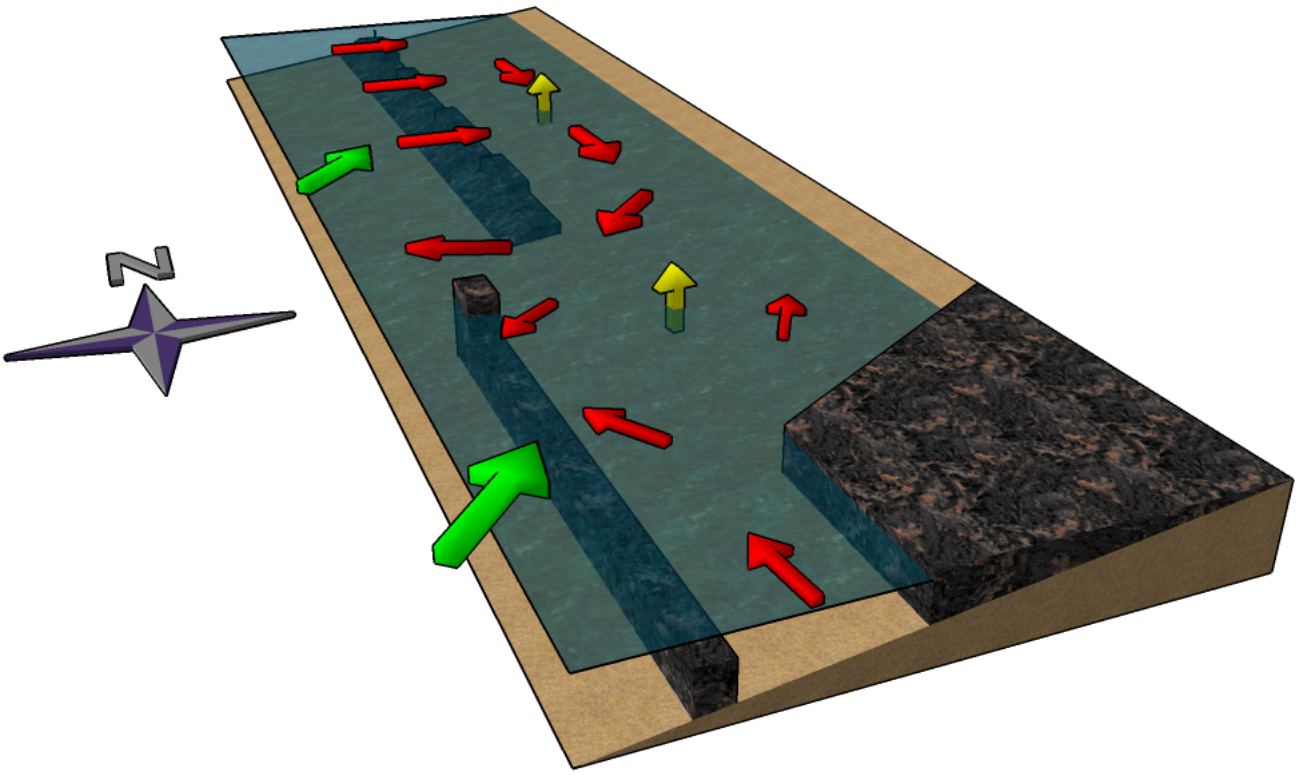

Figure 6.28: Schematisation of stage 3; green arrows = wave energy, red arrows = current velocity field, yellow arrows = sediment suspension

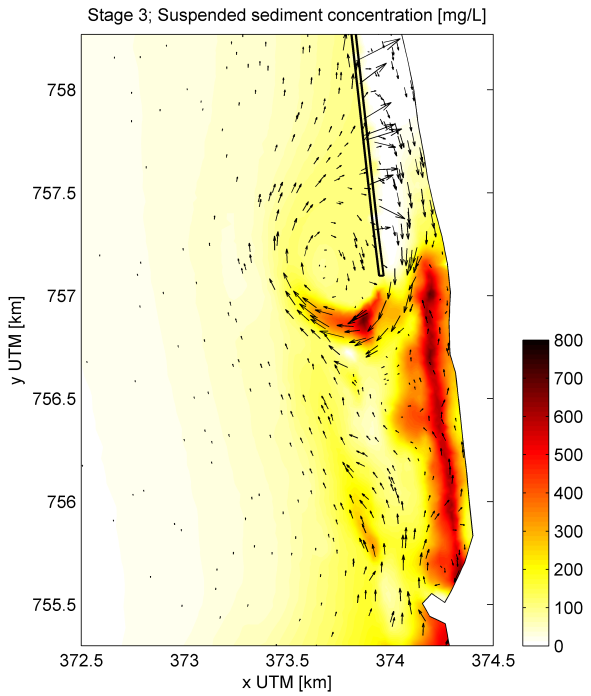

Figure 6.29: Flow field and suspended sediment concentration of an example scenario in stage 3

Wave period

From the scatter plots and calculated rank correlation coefficients we observed a notable dependency of the wave period for the longshore sediment flux and the total sediment balance (tables 6.8 and 6.10). In figure 6.30 the modelled wave height is shown of two scenarios with same imposed boundary conditions, but with a different wave period. Both scenarios look very different and the wave period seems to have significant influence on the spatial variability of the wave height. This can be explained by the fact that when waves approach shallow water, longer waves feel the bottom earlier than shorter waves. The ratio between the wave group celerity (c_g) and wave celerity (c), denoted as n[7], is decaying earlier for longer waves. Ratio n is a function of kh and wave number k is smaller for longer waves than for shorter waves (Bosboom and Stive, 2011). Therefore the waves with a wave period (T_p) of 12 seconds undergoes transformation at the continental shelf while for the other scenario with a wave period of 8 seconds no difference is observed (figure 6.30). This results in a spatial variability of the wave height in vicinity of the reef for longer waves while for shorter waves the wave height along the reef is more or less constant. Moreover, the wave height is locally higher for the situation with a longer wave period. The shoaling factor K_{sh} is also purely a function of kh (Bosboom and Stive, 2011). Therefore the shoaling factor is larger for longer waves resulting in higher waves in front of the reef. This can have significant influence on the hydrodynamics, and therefore the sediment fluxes, in vicinity of the reef.

Wind speed

From the results we observed a remarkable high dependency of the wind speed. For some output variables, like for the total sediment balance, the wind speed is even more dominant than the wave height. This seems quite odd, but it should be noticed that the dependency of the wave height is that from the *imposed* wave height at the offshore boundary and not from the local nearshore wave height. The local wave climate is not only dependent on the imposed offshore wave height but also on the imposed wind conditions. As mentioned, the imposed wave condition is a swell wave condition. This means that the offshore wave energy density spectrum will show a sharp narrow peak. However the local wind conditions is able to develop sea waves and therefore it is possible that the nearshore wave energy density spectrum will show an additional broader peak. Figure 6.31 shows two wave energy density spectra in vicinity of the reef for different scenarios with different imposed wind speeds. The scenario with a higher imposed wind speed shows an additional broad (in both direction and frequency) peak. The total energy (E) is dependent on the volume under the total wave energy density spectrum; the zero order moment (m_0) (Holthuijsen, 2007). So the local total wave energy is also dependent on the local wind climate and can even be more dominant than the imposed wave height. The large directional spreading of the additional sea waves also explains the low dependency of the wind direction for all the output variables.

[7]Ratio n is also a measure for the difference between shallow, intermediate and deep water. It is used in section 5.2.1

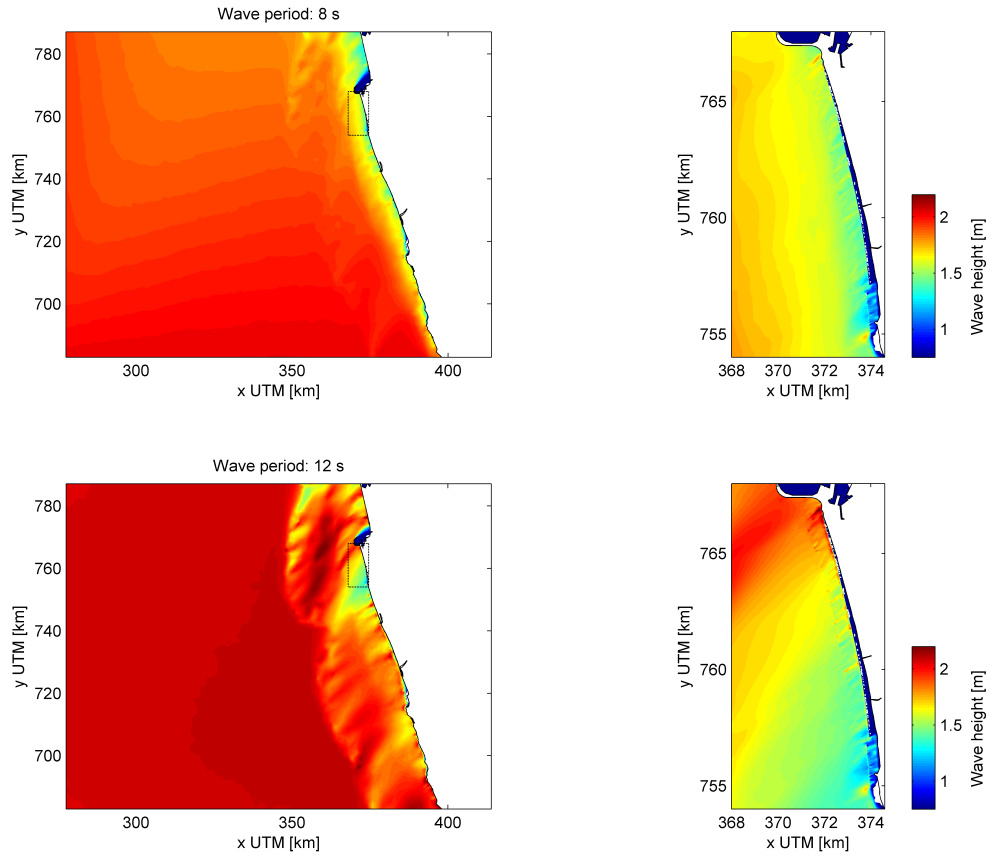

Figure 6.30: Wave height of two scenarios with different wave periods

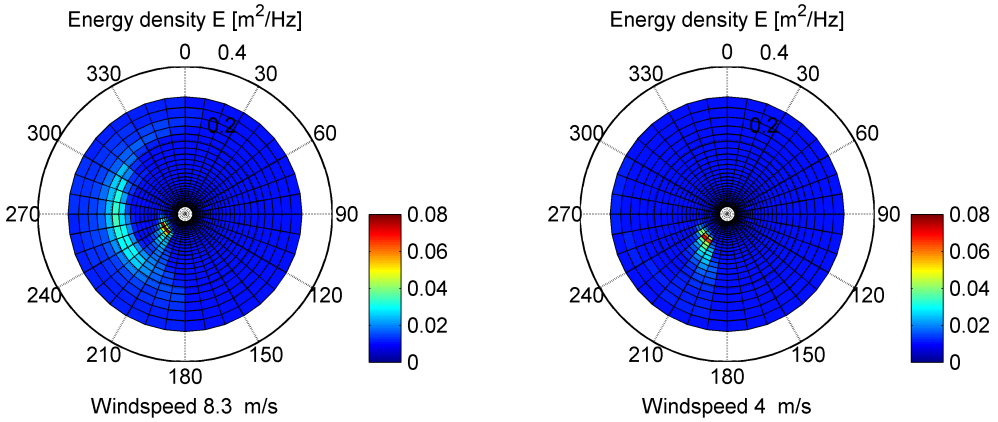

Figure 6.31: Wave energy density spectra in vicinity of the reef for two scenarios with different wind speeds

6.3.3 Regression

A linear multiple regression model is used to replace a complex model with a simplified *response surface* (Hamby, 1994), see section 4.5. It provides an estimate of the linear sensitivity of the input variables on the output variables, which are given by the *regression coefficients* b_j. For the explanation of the regression model and the derivation of the regression coefficients see appendix A.6. In many cases it useful to perform *stepwise regression* (Hamby, 1994; Helton and Davis, 2003; Morgan and Henrion, 1990). This means that variables are added to the initial regression model in a sequence depending on their *p-value*. The *p-value* is a measure for the significance of a variable to the total regression model. More information about the *p-value* can be found in section A.6.3. The initial model starts with no variables and for every variable single regression is performed. The most significant variables, those with the least *p-value*, are added to the model first. If a variable has a *p-value* larger than significance level $\alpha = 0.05$ the variable is not added to the model, since they are not improving the total model.
In order to get rid of the different scales and the units between the different variables the data is first normalized. The *standard scores*, or often called the *z-scores*, are computed for the data sets of both the output and the input variables. The regression coefficients estimated from these normalized data sets are called the *standardized regression coefficients* (SRC) (Hamby, 1994; Morgan and Henrion, 1990).
Like for the correlation coefficients it may be useful to use *rank transformations* in order to linearize the underlying relationships and facilitate regression for non-linear models. The regression coefficient stemmed from rank transformed data sets are called *standardized rank regression coefficients* (SRRC) (Helton and Davis, 2003). Since rank transformed data sets are not dependent on units or scale it is not required to normalize the data.
Since we know that the reef is never totally emerged the same condition A as determined in section 6.2.2 (page 44) is used for this regression model. So the (standardized) regression coefficients are determined given that the reef elevation is always positive, i.e. no total emergence of the reef.
The results of the stepwise regression are given in tables 6.11, 6.12, 6.13, 6.14 and 6.15. The variables are listed in order of sequence in which they are added to the model. Included are the *standard errors* (SE) of the regression coefficients. The standard error is a measure for the uncertainty of the estimation for the regression coefficient. For the definition and derivation of the standard error of regression coefficients see section A.6.2. Note that this is not a confidence interval. Also included are the (cumulative) *coefficients of determination*, denoted as R^2. The coefficient of determination is measure for the global fit of the model. Formally, the value of R^2 is the fraction of the total sample variation that is explained by the regression model. For the definition and derivation of the coefficient of determination see section A.6.1. As variables are added to the regression model R^2 is increasing.
Beside the obtained regression coefficients, the sequence in which the variables are selected is a useful measure for their uncertainty importance, as is the increment in the R^2 value they produce. In figure 6.32 the regression coefficients are visualised to get an better overview of the results.
The results give quick insight in which variables are of most importance for this study. The regression coefficients should be interpreted as the fraction of influence of one standard deviation of the input variable (explanatory variable) on one standard deviation of the output variable (variable to be explained). For some variables the rank transformed data gives better values R^2 which indicates a more accurate regression model. The overall R^2 values show that the regression models are quite accurate, particularly for the hydrodynamics. For the longshore transport the errors are a relatively larger which results in a slightly lower value for R^2. This corresponds with the observed phenomena described in section 6.3.1, where we found that the relationships for the longshore sediment flux are rather complex. Therefore it should be noted that regression provides a measure for the global linear sensitivity of the variables. Particularly for the reef elevation the

y = Average set-up behind given that the reef is submerged								
	Raw data				Rank transformed data			
Step	Variable	SRC	SE	R^2	Variable	SRRC	SE	R^2
1	Wave height	0.61	0.0272	0.46	Wave height	0.59	0.0288	0.43
2	Wave direction	0.52	0.0249	0.75	Wave direction	0.51	0.0267	0.72
3	Peak period	0.33	0.0248	0.85	Peak period	0.32	0.0266	0.82
4	Wind speed	0.27	0.0272	0.90	Wind speed	0.31	0.0288	0.89
5	Reef elevation	-0.08	0.0249	**0.91**	Reef elevation	-0.08	0.0268	**0.90**

Table 6.11: Stepwise regression for the average set-up behind the reef, including standard errors and (cumulative) R^2 values with in bold the value of the final regression model.

y = Average velocity in polygon given that the reef is submerged								
	Raw data				Rank transformed data			
Step	Variable	SRC	SE	R^2	Variable	SRRC	SE	R^2
1	Wind speed	0.47	0.0273	0.52	Wind speed	0.48	0.0322	0.51
2	Wave height	0.56	0.0278	0.72	Wave height	0.54	0.0324	0.70
3	Wave direction	0.35	0.0249	0.86	Wave direction	0.33	0.0300	0.82
4	Roughness height	-0.18	0.0250	0.89	Reef elevation	0.17	0.0301	0.85
5	Reef elevation	0.13	0.0251	0.90	Roughness height	-0.13	0.0298	0.87
6	Wind direction	-0.06	0.0250	0.91	Peak period	0.07	0.0298	**0.87**
7	Peak period	0.06	0.0248	**0.91**				

Table 6.12: Stepwise regression for the average velocity in polygon, including standard errors and (cumulative) R^2 values with in bold the value of the final regression model.

coefficients therefore may give a distorted view. However although the relationship is complex, we observed that the reef elevation is of significant importance for the longshore sediment flux (figure 6.20). This can also be observed in the results of the regression analysis, where the reef elevation is even denoted as the most sensitive variable (table 6.13). The wind direction and roughness height are both for almost every output variable very insensitive, which corresponds with the observations in section 6.3.1. For some output variables these variables are not even included in the final regression model. For the total sediment balance it seems that the forcing variables, particularly the wind speed, are of most importance. However the wave height seems to be less sensitive than the reef elevation. This was already observed by the examination of the rank-order correlations in table 6.10. From these results we may conclude that the forcing variables, beside the wind direction, are almost always the most important variables. The reef elevation can have significance influence on the sediment fluxes which is also notable for the total sediment balance. The roughness height seems to have no significant influence on the output. Also, as mentioned earlier, the wind speed and the wave period are (unexpected) the most dominant variable for the total sediment balance. See page 74 for an explanation of this remarkable variable importance.
Another method is the use of *robust regression*, which is less sensitive for outliers. However several tests showed no significant difference in the results and is therefore not mentioned further in this analysis.

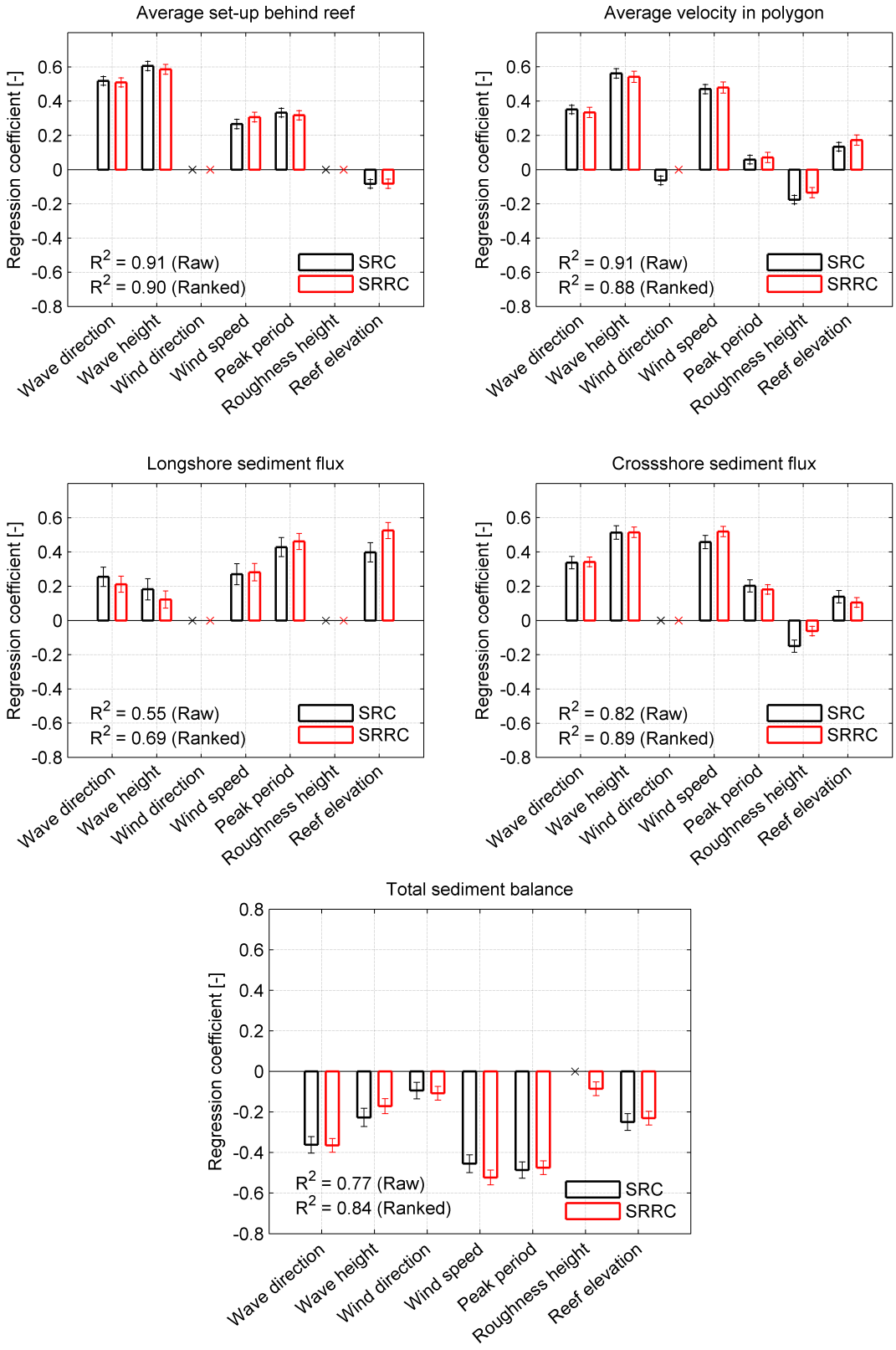

Figure 6.32: Regression coefficients of raw and rank transformed data including their standard error and corresponding (cumulative) R^2 value of the final regression model. The 'x' means that the variable is not included in the final regression model

y = Longshore sediment flux near the gap given that the reef is submerged								
	Raw data				Rank transformed data			
Step	Variable	SRC	SE	R^2	Variable	SRRC	SE	R^2
1	Reef elevation	0.40	0.0565	0.17	Reef elevation	0.53	0.0471	0.32
2	Peak period	0.43	0.0561	0.33	Peak period	0.46	0.0467	0.52
3	Wind speed	0.27	0.0616	0.47	Wind speed	0.28	0.0506	0.64
4	Wave direction	0.25	0.0564	0.53	Wave direction	0.21	0.0470	0.68
5	Wave height	0.18	0.0617	**0.55**	Wave height	0.12	0.0506	**0.69**

Table 6.13: Stepwise regression for the longshore sediment flux near the gap, including standard errors and (cumulative) R^2 values with in bold the value of the final regression model.

y = Crossshore sediment flux near the gap given that the reef is submerged								
	Raw data				Rank transformed data			
Step	Variable	SRC	SE	R^2	Variable	SRRC	SE	R^2
1	Wind speed	0.46	0.0392	0.46	Wind speed	0.52	0.0300	0.54
2	Wave height	0.51	0.0396	0.62	Wave height	0.51	0.0301	0.72
3	Wave direction	0.34	0.0359	0.74	Wave direction	0.34	0.0279	0.85
4	Peak period	0.20	0.0357	0.78	Peak period	0.18	0.0277	0.88
5	Roughness height	-0.15	0.0358	0.80	Reef elevation	0.10	0.0280	0.88
6	Reef elevation	0.14	0.0360	**0.82**	Roughness height	-0.06	0.0278	**0.89**

Table 6.14: Stepwise regression for the crossshore sediment flux near the gap, including standard errors and (cumulative) R^2 values with in bold the value of the final regression model.

y = Total sediment balance given that the reef is submerged								
	Raw data				Rank transformed data			
Step	Variable	SRC	SE	R^2	Variable	SRRC	SE	R^2
1	Wind speed	-0.46	0.0443	0.30	Wind speed	-0.52	0.0365	0.35
2	Peak period	-0.49	0.0403	0.54	Peak period	-0.48	0.0337	0.61
3	Wave direction	-0.36	0.0405	0.68	Wave direction	-0.37	0.0339	0.75
4	Reef elevation	-0.25	0.0407	0.73	Reef elevation	-0.23	0.0340	0.80
5	Wave height	-0.23	0.0448	0.76	Wave height	-0.17	0.0369	0.82
6	Wind direction	-0.09	0.0406	**0.77**	Wind direction	-0.11	0.0338	0.83
7					Roughness height	-0.09	0.0338	**0.84**

Table 6.15: Stepwise regression for the total sediment balance, including standard errors and (cumulative) R^2 values with in bold the value of the final regression model.

Multicollinearity

When two or more predictor variables are (highly) correlated within a multi regression model we may deal with *multicollinearity*. *Collinearity* is a linear relationship between two explanatory variables, which in this study exists (imposed) between the wave height and wind speed. Multicollinearity can be a problem as it may give inaccurate estimates of the regression coefficients stemmed from the multi regression model. In appendix A the standard deviation of regression coefficient b_j (SE) is denoted in equation (A.36). However the variance of regression coefficient b_j may also be decomposed as (Heij et al., 2004);

$$s_{b,j}^2 = \frac{s^2}{(m-1)s_{x_j}^2} \cdot \frac{1}{1-R_{a,j}^2} \tag{6.7}$$

Where s^2 is the estimated variance of the total regression model, $s_{x_j}^2$ the estimated variance of explanatory variable x_j and m the sample size. The last term consist of $R_{a,j}^2$ which is the coefficient of determination of an *auxiliary regression*. Auxiliary regression in this case implies regression of explanatory variable x_j on the remaining explanatory variables. So output variable y in this auxiliary regression is left out. From equation (6.7) it is clear that the accuracy is high for a large sample size, large variation in the relevant regressor, small error variance and small collinearity with the other regressors. The last term in equation (6.7) is denoted as the *variance inflation factor* (VIF); the factor by which the variance increases because of collinearity of the j^{th} regressor with the other regressors (Heij et al., 2004).

$$\text{VIF} = \frac{1}{1-R_{a,j}^2} \tag{6.8}$$

According to rules of thumb a variation inflation factor larger than 4 (or some say 10) indicates severe multi collinearity and may cause problems for the estimation of the regression coefficients (O'brien, 2007). Table 6.16 shows the coefficients of determination and the variation inflation factors for all the variables. It shows that some collinearity exists, for wave height and wind speed, however the values are rather low which indicates that multicollinearity will not be a serious problem for this study.

Variable	R_a^2	VIF
Wave direction	0.024	1.025
Wave height	0.177	1.216
Wind direction	0.019	1.019
Wind speed	0.158	1.188
Peak period	0.012	1.012
Roughness height	0.019	1.019
Reef elevation	0.031	1.032

Table 6.16: Coefficients of determination of auxiliary regression and variation inflation factors

Direct, indirect and total effects

The results from the stepwise regression analysis provided in previous section gives insight in the *direct effects* of each variable on the output variable. That is the effect under the assumption that all other variables remain fixed. However since a correlation is imposed between wave height and wind speed, higher wind speed can *cause* a higher (imposed) wave height. Therefore those variables may have two effects; a *direct effect* and a *indirect effect* via the other variable. Those

two effects together is then the *total effect* on the output variable. In such case it may be hard to keep one variable fixed and therefore it may be more natural to look at the total effect rather than the direct effect (Heij et al., 2004) [8]. The total effect can be determined by regression of a *restricted model*, which implies a regression model where the other dependent variable is left out of the model. So for the determination of the total effect of the wave height (x_1) the variable wind speed (x_2) is left out the model and vice versa. The obtained regression coefficient b_R is the total effect of the explanatory variable (x_1) and consist, as mentioned, of two components;

$$b_R = b_1 + Pb_2 \tag{6.9}$$

$$\text{'total'} = \text{'direct'} + \text{'indirect'} \tag{6.10}$$

Here b_1 is the regression coefficient of the stemmed from the *unrestricted* model, so the results provided in previous section and therefore the direct effect of variable x_1. The second term is denoted as the indirect effect and is dependent on regression coefficient b_2 of dependent variable x_2 stemmed from the *unrestricted* model and the effect of x_1 on x_2 (P). Consider first input variable wave height as x_1 and wind speed as x_2 and vice versa. In table 6.17 the direct, indirect and total effects are shown of the wave height and wind speed on the total sediment balance. The indirect effects in this table are the effects that run via the other explanatory variable x_2. Clearly the indirect effect of the wave height is significant and larger than the indirect effect of the wind speed. This can be explained by the graphical presentation of the effects for the rank transformed data in figure 6.33. The effect of x_1 on x_2 (P) is, as expected, for both situations the same. But since the direct effect of the wind speed is larger than the direct effect of the wave height, the *indirect* effect is larger for the wave height than for the wind speed. From these observations we may conclude that despite the fact these variables are correlated the wind speed dominates the wave height for the total sediment balance.

	Raw data			Rank transformed data		
Variable (x_1)	Direct (b_1)	Indirect (Pb_2)	Total (b_R)	Direct (b_1)	Indirect (Pb_2)	Total (b_R)
Wave height	-0.227	-0.190	-0.417	-0.171	-0.210	-0.381
Wind speed	-0.455	-0.092	-0.548	-0.523	-0.065	-0.588

Table 6.17: Direct, indirect and total effect of the wave height and wind speed for the total sediment balance

[8] Note that this not imply that the estimate for the direct effect is wrong

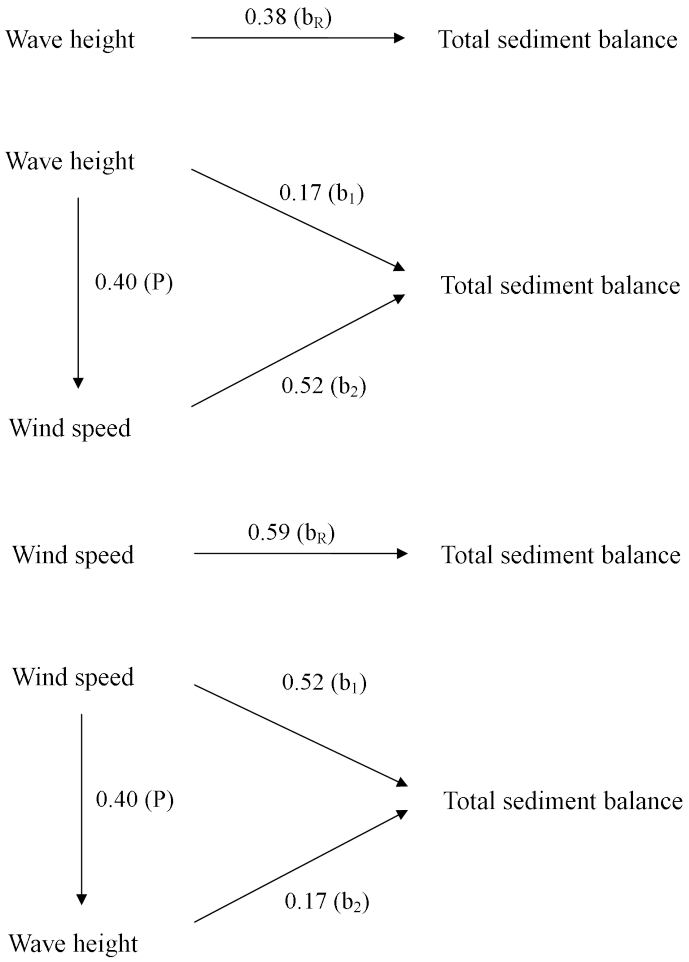

Figure 6.33: Direct, indirect and total effects of wave height and wind speed on the total sediment balance for the rank transformed data. The indirect effect of x_1 is the effect of x_1 on x_2 (P) times the direct effect of x_2 (b_2). The direct effect and the indirect effect together is the total effect (b_R).

Chapter 7

Applications

The results presented in previous chapter provide a lot of opportunities and gives rise to some useful applications. This chapter provides ways to interpret the results and gives insight in the possibilities and applications stemmed from the results of the performed method.

7.1 Interpretation

The results of provided in previous chapter are divided in two segments; *uncertainty propagation* and *uncertainty importance*. The uncertainty propagation is a measure for the deviation, or uncertainty, of the model output stemmed from the specified and determined uncertainty of the model input and parameters compared with a deterministic approach. The distribution of the model output provides an indication of the expected results given the specified and determined uncertainties. To translate this to reality we have to define what the quantified uncertainty of the input an model parameters actually stands for. We can say that the quantified uncertainties are the uncertainties given the *expected* scenario. The expected scenario is equal to the expected values of the input variables ($\mathrm{E}[x]$) which is in this case are equal to their mean (μ). In this case the expected scenario is equal to the second column in table 6.1. In this way the uncertainty propagation therefore provides the expected distribution of results that will occur given the expected scenario and specified uncertainties, stemmed from both variability and limited knowledge. In fact it gives an indication of the expected distribution of the (stationary) results if the coast is subject to wave condition 3, neglecting uncertainties stemmed from other than those described in this study. This provides us with an indication of the results given the limited knowledge about some parameters; a data poor environment. The calculated (sample) mean, 5^{th} and 95^{th} percentiles give an approximation of the expected value, lower and upper bound, respectively.

7.2 Total wave climate

The method is executed for only wave condition. The results should therefore be interpreted as described in the section above. The translation to reality lacks in a way that in reality the coast is subject to more wave conditions. In section 3.2 we considered the total wave and wind climate as the sum of six weighted conditions;

$$D_{total} = \sum_{w=1}^{6} w_w \mathbf{D}_w \tag{7.1}$$

This function still holds for the obtained distributions of the output variables. If the method is executed for all wave conditions for every wave condition distributions of output variables are available. Taking the weighted sum of these distributions, with use of for instance a Monte Carlo procedure, the distribution stemmed from the total wave climate is obtained. This also means that it is allowed to extend the results to yearly values since the yearly wave climate is now taken into account.

7.3 Reliability

The distributions of the model output variables not only give an indication of the expected results, it also provides insight in the reliability of the results. A higher uncertainty corresponds with less reliable results. The relative standard deviation is a good measure for the relative magnitude of the uncertainty irrespectively of their mean. In this study reliability can be seen from two different perspectives; as the reliability of the *results* or as the reliability of the *analysis*. The first one is related to the observed distributions and uncertainties of the model output, while the other is related to statistical uncertainty, which is stemmed from the limited amount of data samples. By taking both into account insight in the total reliability is given. For instance from the results observed it is clear that the reliability for the hydrodynamics, both related to statistical uncertainty and output uncertainty, are larger compared to the sediment fluxes. We may wonder how much this decrease in reliability is related to model reliability.

7.4 Calibration

One of the drawbacks dealing with data poor environments is the lack of performing a correct calibration or validation of the numerical model. A crude validation is carried out (section 5.2.2) in order to test whether the model is capable of predicting the processes in a correct way, however an exact location specific validation is not given. By imposing model parameters as uncertain variables the lack of calibration is obviated. Then the question arises if the reliability is significant decreased by this lack of correct validation and calibration. The only uncertain model parameter that is used for this study is the roughness height. The roughness height is often difficult to measure in real life and is therefore often used as calibration parameter. From the results we observed that the roughness height has a low uncertainty importance and has therefore no significant influence on the output of the model. Assuming that the roughness height is the only parameter that is related to model reliability we may conclude that calibration is inferior to the uncertainty of the other variables. However this is very hard conclusion since more model parameters are available. It is therefore not concluded that the validation is also obviated. Though crude validation showed that the model is capable of predicting the processes in a correct way. So in fact there is only a lack of quantitative validation, which can partly be obtained with use of correct calibration. This would therefore mean that also validation is obviated by the probabilistic method. However due to strong schematisation exact validation, for this study, is not obtained. It should be noted that the proposed method is capable of *obviating* calibration and is able to provide an approximation of the need of calibration.

7.5 Processes

By the execution of the proposed method, described in this study, an useful data set is obtained. As shown in the previous chapter not only an indication of the uncertainty propagation is given also insight in the relationships between the parameters are obtained. Examination of scatterplots

can reveal underlying relations, thresholds or non-linear effects. The observed relations can be checked by comparing them with physical relationships. In that way insight is given in which processes are dominant. In this study the scatterplots gave insight in the important complex hydrodynamic processes and the role of the reef into the total system. The data set in fact consists of all possible scenario's that may occur for the coast that is subject to the imposed wave condition. With examination of separate scenario's and overall relationships the system is better understood. Moreover the correlation and regression coefficients give insight in the most important parameters and their dominance related to the processes that occur.

7.6 Risk analysis

An important advantage compared to traditional deterministic modelling is the availability of probabilities of (unwanted) events. *Risk analyses* combine the probability of occurrence with the consequences of an event. Interests are for instance the risk involved when the erosion exceeds any specified limit. Empirical (cumulative) distributions provide an estimation of the probability of exceedance of any specified value. Usually the interests are the other way around; the value that corresponds with a desired probability of exceedance or undershooting. For example if one is interested in the sediment balance that is undershooted with a probability of α, the following value is looked for;

$$S(\alpha) = \min_{x}(P(x \leq S_{tot}) \geq \alpha) \tag{7.2}$$

Since estimations of the (bootstrap) confidence intervals are available, intervals for the determined value can be estimated as well. Note that this study has used a relatively small sample size and therefore values corresponding with extreme probabilities (α) are very unreliable. Moreover the estimations of the probability distributions of the input variables (appendix C) are less accurate for the tails of the distribution.

7.7 Value of information

As we already observed from the results the reef elevation can have significant influence on the sediment fluxes, particularly for the longshore sediment flux. Also on the total sediment balance the reef elevation has influence. It is possible to reduce the uncertainty of the reef elevation by for instance a additional detailed survey. Therefore it may be interesting to get an idea of the total uncertainty reduction by knowing the reef elevation exactly. By obtaining sufficient data we get more *information*. *Perfect information* (PI) would than be that the reef elevation would be known exactly, which implies that the uncertainty of the reef elevation is reduced to zero. Obviously this uncertainty reduction will also lead to an uncertainty reduction of the output distribution. Uncertainty reduction of the results is valuable, because the reliability is increased. Therefore obtaining extra information, or perfect information, has a certain value. In this section a conceptual method is given that is able to estimate the *expected value of perfect information* (EVPI), which is based on methods given by Morgan and Henrion (1990) and Howard (1966). The EVPI is described by the difference between the *expected costs including uncertainty* (ECIU) and the *expected costs with perfect information* (ECPI);

$$EVPI = ECIU - ECPI \tag{7.3}$$

Expected costs including uncertainty

The expected costs including uncertainty (ECIU) is obtained using the results of the execution given the uncertainty of the reef elevation; i.e. *including uncertainty* (IU). In this case we are interested in the total sediment loss behind the reef and therefore the empirical distribution of the total sediment balance is used. Consider that the total sediment balance corresponds with the required nourishment to keep the coast stable. Therefore the costs are proportional to the *negative* sediment balance. We are looking for a value that corresponds with a specified probability of exceedance (α). For example if the required total nourishment is set equal to the negative total sediment balance (S_{tot}) with an undershooting probability of 10% [1]. The ECIU is then denoted as ($\alpha = 0.10$);

$$ECIU = -\min_{x}(P(x \leq S_{tot}|IU) \geq \alpha) \tag{7.4}$$

Which can be obtained with use of the empirical cumulative distribution shown in figure 6.12. The costs are then given in required cubical meters of sand.

Expected costs with perfect information

The costs given the state of perfect information (PI) is more difficult to determine. The exact deterministic reef elevation is not known and is Gaussian distributed with parameters described in section 3.2. This means that PI is also not known and described by the same distribution. The costs given the state of perfect information ($C|PI$) is estimated using the following equation;

$$C|PI = -\min_{x}(P(x \leq S_{tot}|PI) \geq \alpha) \tag{7.5}$$

Since PI is stochastic the corresponding costs $C|PI$ must be stochastic as well, which is described by an unknown distribution. The *expected costs with perfect information* (ECPI) can be obtained by determining the expected value of the distribution;

$$ECPI = E[C|PI] \tag{7.6}$$

This means that only if the distribution of the costs given perfect information ($C|PI$) is known the expected value of perfect information (EVPI) can be obtained. This distribution is, however, not known before hand. An execution of, for instance, a Monte Carlo method is able to provide an estimation of the distribution of $C|PI$. By the generation of z samples of PI, z scenarios can be executed. So executing the proposed method with the same parameters, but given the state of perfect information (deterministic reef elevation). This will provide z different output distributions and z different estimations of the costs ($C|PI$), where from the distribution can be estimated. Note that this means that a Monte Carlo method is used within a Monte Carlo method.
There is a lack of information regarding the underlying distribution of the costs with perfect information. Below an example sample execution is given that can be used for the Monte Carlo analysis for the determination of the underlying distribution of $C|PI$.

Example scenario

The results of one sample scenario execution are given in figure 7.1, 7.2 and 7.3. In this case PI is set equal to the mean of the reef elevation so $h_{reef} = -0.50$ m+MSL. It gives insight in the uncertainty reduction stemmed from knowing that the reef elevation is exactly -0.50 m+MSL . The results

[1] For this value the statistical uncertainty is still acceptable, for more extreme percentiles more samples are required for an accurate estimate, see section 6.2.4

show that the uncertainty reduction is much larger for the longshore sediment flux than for the crossshore sediment flux. This was expected since it was concluded that the reef elevation was more important for the longshore sediment flux (section 6.3). Hardly any uncertainty reduction is observed for the crossshore sediment flux, beside the fact that the right tail is slightly smaller and thinner. For the total sediment balance the uncertainty reduction is also significant although, as expected, not as large as for the longshore sediment flux. The costs with perfect information ($C|PI$) is in this case equal to 4216 m^3/day, see figure 7.3. With an ECIU of 5024 m^3/day the value of perfect information for this scenario is equal to 808 m^3/day. Repeating this calculation z times for z different deterministic values for the reef elevation eventually leads to an estimation of the underlying distribution and therefore can lead to an estimation of the *expected* value of perfect information.

Translation to full climate

The $C|PI$ is given in cubical metres of sand, however the financial costs can be estimated as well since indications of costs per cubical meter are available. Assumed that nourishing of sand will approximately cost €$10/m^3$ of sand, the $C|PI$ for the example scenario becomes €42.160,- and the expected cost including uncertainty (ECIU) becomes €50.240,-. These costs are given *per day* for the system that is subject to wave condition 3. Since we know that the coast is subject to more wave conditions, the other wave conditions need to be taken into account to get the total value of information for a longer period, for instance per year. By taking the weighted sum of the EVPI of all wave conditions (equation 7.1), it is allowed to extend the EVPI to yearly values.

This concept shows it is possible to determine the *expected value of information*. This value can be used to get idea whether it would be useful to obtain more (accurate) data. For instance if the costs of an additional survey will be less or equal than the computed EVPI it may be useful to perform the survey since the risk reduction is greater than the expected costs. Note that this concept is based on the information about the total reef elevation, so spatial variability nor resolution is mentioned in this analysis. Correlation and regression coefficients provides approximations of the relative importance of the parameters and so give indication for which parameters it it is useful to reduce their uncertainty. For some parameters the uncertainty is stemmed from inherent uncertainty and therefore it is impossible to reduce their uncertainty. However for parameters which are uncertain because of the lack of (accurate) data, it is possible to reduce their uncertainty.

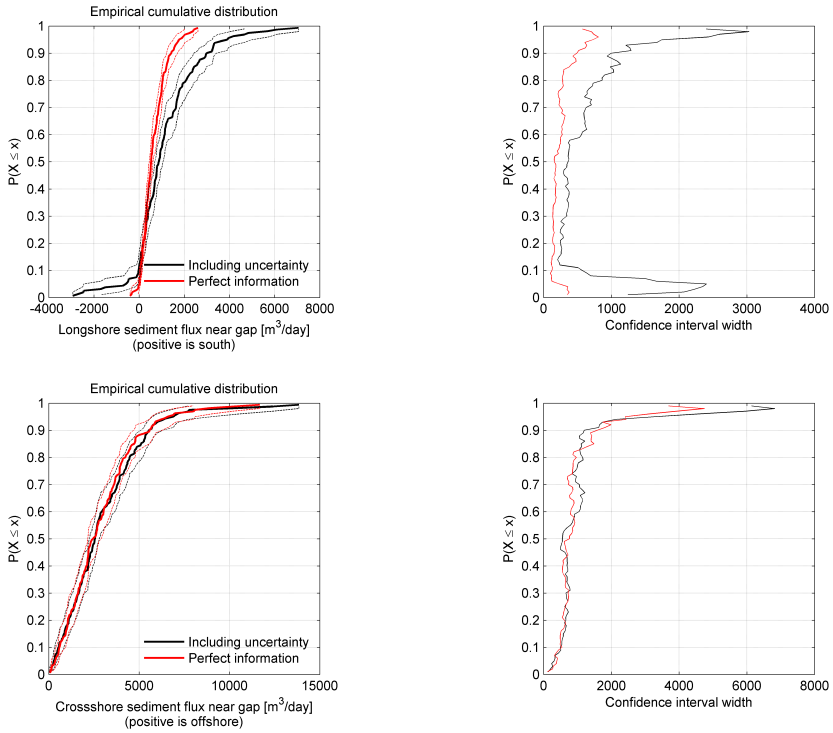

Figure 7.1: Empirical cumulative distributions for the sediment fluxes near the gap for the executions; including uncertainty and given the state of perfect information

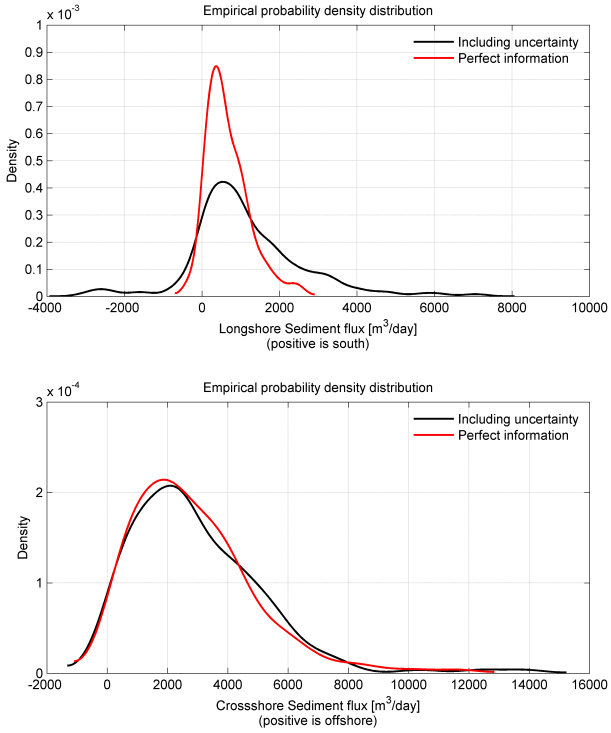

Figure 7.2: Empirical probability density distributions for the sediment fluxes near the gap for the executions; including uncertainty and given the state of perfect information

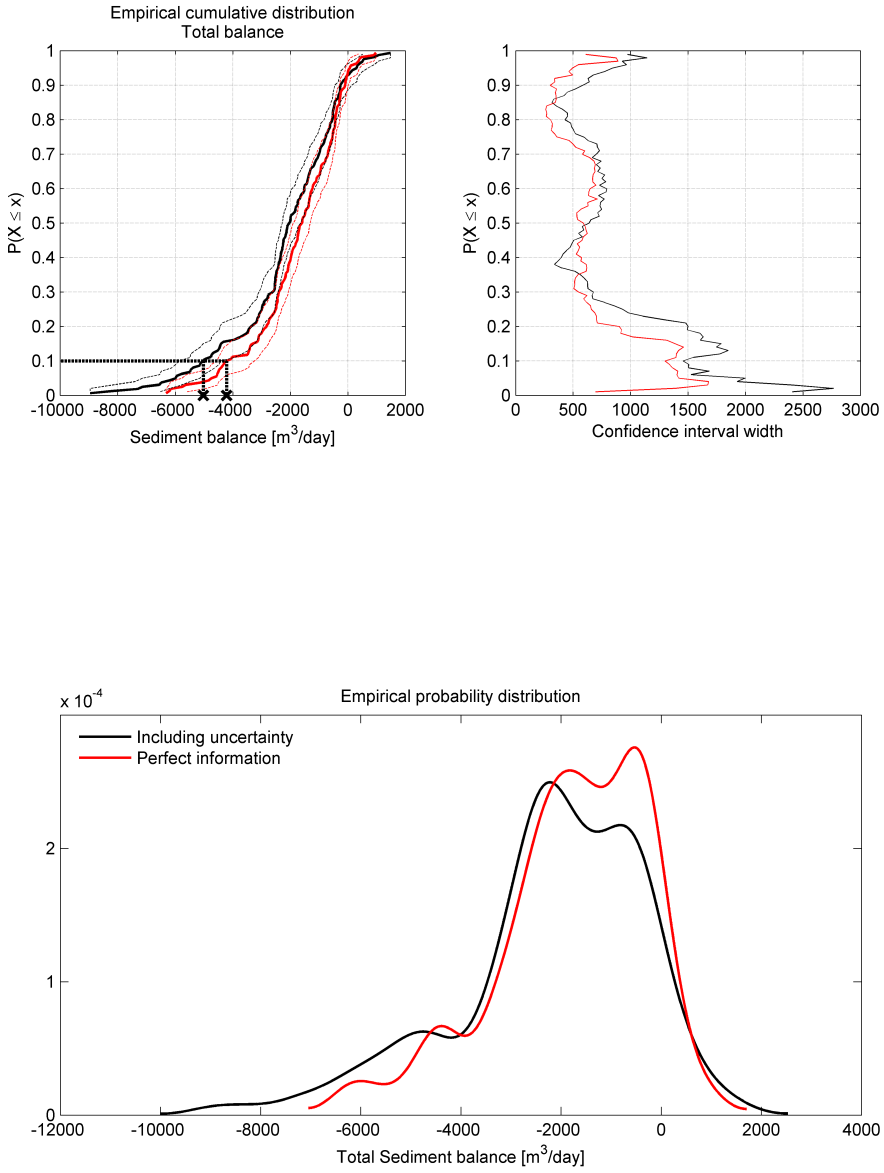

Figure 7.3: Empirical cumulative and probability density distributions for the sediment fluxes near the gap for the executions; including uncertainty and given the state of perfect information

Chapter 8

Conclusions

This report describes a specific approach for stochastic modelling, including a quantification of uncertainties for input variables, realization of the method and an analysis of the results. Therefore a method is proposed for dealing with uncertainties in hydrodynamic and morphodynamic modelling, i.e. dealing with modelling of a data poor environment. The following general conclusions are drawn;

- Uncertainty analyses provide an improved interpretation of model results compared to deterministic modelling, as output distributions provides a range of possible results including their expected probability of occurrence.
- Uncertainty analyses enable coastal engineers and decision makers to focus on the most important model aspects and variables, particularly when dealing with data poor environments. This can lead to economic efficiency in coastal engineering practice, as it determines the necessity of data collection and calibration.

These general conclusions are substantiated by the following conclusions of this research. They are separated by conclusions regarding the proposed method and conclusions regarding the case study.

8.1 Proposed method

Regarding the proposed method the following conclusions are drawn;

1. It is possible to get insight the hydrodynamics and morphodynamics in a data poor environment, because;
2. It is possible to get insight in the uncertainties and reliability of hydrodynamic and morphological model output stemmed from specified and determined uncertainties of the input and model parameters, and therefore it is possible to get an indication of the processes that most likely will occur in the system being modelled.
3. Monte Carlo method is a suitable and robust method to perform stochastic modelling.
4. It is possible to quantify the uncertainties of forcing variables by describing the full climate as a function of separate probability distribution functions.

5. A sample size of (only) 160 data points per probability distribution provides reliable insight in the total uncertainty propagation and uncertainty importance for non extreme probabilities (probabilities between 0.05 and 0.95).
6. An increase of the sample size would provide a more reliable insight and better estimates for more extreme probabilities.
7. It is possible to get insight in the accuracy of the statistical results by use of the bootstrap method.
8. (Rank) correlation coefficients provides an approximation of the linear dependency of input and model parameters on the output
9. Stepwise regression is able to quantify the relative variable importance into the total system being modelled.
10. The proposed method gives rise to many opportunities and applications;
    (a) The method for the quantification of uncertainties of the forcing variables provides a translation to get insight in the uncertainties stemmed from a full climate, in contrast to one wave condition as been done for this case study. (section 7.2)
    (b) The method provides insight in the reliability of the results, as well related to the uncertainties from the output variables as related to statistical uncertainty; stemmed from limited sample size (section 7.3).
    (c) The method is able to obviate calibration and is able to get an approximation of the need for calibration (section 7.4)
    (d) The method provides a useful data set from which relationships and relative importance between variables can be obtained which in turn provides insight in the important (hydrodynamic and morphological) processes. Therefore the system being modelled is better understood.
    (e) The method provides the applicability to perform risk analyses
    (f) A conceptual method proves the possibility to determine the expected value of information.
11. The proposed method as applied for this study also holds for other case studies. The following aspects will not affect the applicability of the proposed method and (statistical) analysis of the results;
    (a) The amount of input and output parameters.
    (b) The magnitude of the uncertainty of input and model parameters.
    (c) The use of other types of probability distributions
    (d) The use of other hydrodynamic/morphological (numerical) models.

8.2 Case study

The proposed method is executed for a case study (chapter 2). From the analysis the following conclusions are drawn;

1. Insight is obtained in the hydrodynamic and morphological processes, despite the fact we are dealing with a data poor environment

2. Indication of the most probable results is obtained, despite the fact we are dealing with a data poor environment
3. Insight is obtained in the reliability of the results, stemmed from the quantified uncertainties of the input and model parameters
4. Regarding the observed uncertainties we may conclude the following
   (a) The hydrodynamics show reliable results, with respect to statistical uncertainty. The relative uncertainty is acceptable.
   (b) The sediment fluxes show less reliable results, with respect to statistical uncertainty, compared to the hydrodynamics. Also the relative uncertainty is larger.
   (c) The total sediment balance show the least reliable results, with respect to statistical uncertainty. And also the relative uncertainty is large.
5. Regarding the observed processes we may conclude, for parameters inside the used sample space, the following;
   (a) The reef elevation can have significant influence on the sediment flux, particularly for the longshore sediment flux near the gap
      i. A lower reef elevation not necessarily means a decrease in sediment transport nor a higher reef elevation does. For information about the relationship for the reef elevation see section 6.3.1.
   (b) The forcing variables, beside the wind direction, have significant influence on both the hydrodynamics and sediment fluxes.
      i. The wind speed and the wave period are the most important variables for the total sediment balance of the system.
      ii. Higher waves and higher wind speeds both leads to higher sediment fluxes near the gap.
      iii. Larger wave periods lead to higher (longshore) sediment fluxes
      iv. Smaller wave angle with respect to shore normal leads to a higher set-up behind the reef and therefore to higher sediment fluxes.
      v. Wind direction has negligible influence on the results.
   (c) The roughness height seems to have negligible influence on the results.

Chapter 9

Recommendations

Given the applicability of the proposed method it is recommended that, irrespectively of dealing with a data poor environment, comprehensive uncertainty analyses like this become a more common approach within coastal studies. This not necessarily implies a full Monte Carlo computation needs to be executed. *Awareness* of the influence of uncertainty propagation and relative importance can already contribute to a better interpretation of the model results. The proposed method as described in this research gives rise to further research topics. To increase and improve the applicability of this method the following recommendations are given;

1. Include time dependency and investigate the uncertainty propagation through time. It enables to include morphology, i.e. bottom update, into the model and it would give insight in the uncertainties of longer periods in time.

2. Perform a more advanced and therefore more accurate quantification of the uncertainties of the input and model parameters. The method is used for this study is rather crude, but is able to represent the full wave climate in a correct way.

3. Include more variables and model parameters in order to investigate whether better insight in the overall uncertainty propagation is obtained.

   (a) Include more model parameters to determine and quantify the need for calibration.

4. Extend the sample size to get more accurate estimates.

5. Investigate the possibilities to deal with extreme values within an uncertainty analysis applied to hydrodynamic modelling. Extreme values of the used input variables provide insight in the processes during storm conditions.

6. Apply the proposed method on more case studies in order to investigate the applicability on other coastal systems and processes. Applying the method on a case study where sufficient data is available allows for good comparison between deterministic and stochastic modelling.

7. Further research on the determination of the expected value of information. A conceptual method is proposed, however a detailed research may provide a method to quantify the value of obtaining more (accurate) data accurately.

8. Investigate the applicability of other methods for stochastic modelling

(a) Investigate the applicability of other sampling techniques to reduce the required sample size and therefore reduce the computational effort. Investigate the applicability of analytical methods for stochastic modelling in order to reduce the (computational) effort.

Chapter 10

Discussion

The method as described in this report shows a way to deal with uncertainties in hydrodynamic modelling. The study shows that uncertainty analyses give rise to many opportunities and applications, and therefore show that an uncertainty analysis can be very valuable in coastal engineering practice. Depending on the research objective of coastal studies the focus is different. A more detailed study would not necessarily mean that more accurate results are desired. More understanding about the coastal system, both globally and detailed, can also meet the research objective. Therefore it is important to get the explicit focus of the study clear. In coastal engineering practice model results are often used to substantiate decisions. Interpretation of the model results is therefore a very important factor in decision strategy. More accurate results are able to substantiate the decisions better, however wrong interpretation of the results can lead to wrong decisions. Uncertainty analysis provides a range of possible results and can provide more understanding of the system being modelled, which can lead to a better interpretation of the results. We therefore may wonder whether more accurate results are really desired or that an uncertainty analysis also provides the desired substantiation. If still more accuracy is desired; an uncertainty analysis will let engineers and decision makers focus on the most important model aspect and variables, which enables to discuss the necessity of collecting more data for the most important parameters. One of the characteristics of uncertainty analyses for hydrodynaimc modelling is the relative large required computional effort. The last decade computer power is increased very fast. As the computer power is increasing even further it will amplify the possibility to let uncertainty analyses become a more common approach within coastal studies.

References

Battjes, J. (1995). *Vloeistofmechanica*, TU Delft Civil engineering.

Beck, A. T. and da Rosa, E. (2006). Structural reliability analysis using deterministic finite element programs, *Latin American Journal of Solids and Structures* **3**: 197–222.

Becker, J., Sandwell, D., Smith, W., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., S., I., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharach, A., Trimmer, R., Von Rosenberg, J., Wallace, G. and Weatherall, P. (2009). Global bathymetry and elevation data at 30 arc seconds resolution: Srtm30 plus, *Marine Geodesy* **32:4**: 355–371.

Bosboom, J. and Stive, M. J. F. (2011). *Coastal Dynamics 1*, VSSD Delft.

Calabrese, M., Vicinanza, D. and Buccino, M. (2008). 2d wave setup behind submerged breakwaters, *Ocean Engineering* **35**: 1015–1028.

Chang, C.-H., Tung, Y.-K. and Yang, J.-C. (1994). Monte carlo simulation for correlated variables with marginal distributions, *Journal of Hydraulic Engineering* **120**: 313–331.

CUR (1997). Kansen in de civiele techniek. deel 1: Probalistisch ontwerpen in theorie., *Technical Report 190*, CUR, Ministerie van Verkeer en Waterstaat, Directoraat Generaal Rijkswaterstaat.

de Wit, L. (2012). Internal memo boskalis, *Technical report*, Svašek Hydraulics.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical uncertainty, *Statistical Science* **1**: 54–77.

Galappatti, R. (1983). A depth-integrated model for suspended transport., *Technical report*, Delft university of Technology.

Glaister, P. (1993). Flux difference splitting for open-channel flows, *International Journal for Numerical methods in Fluids* **16**: 629–254.

Hamby, D. (1994). A review of techniques for parameter sensitivity analysis of environmental models, *Environmental Monitoring and Assesment* **32:2**: 135–154.

Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods*, Methuen & Co Ltd, London.

Heij, C., de Boer, P., Franses, P., Kloek, T. and van Dijk, H. (2004). *Econometric Methods with Applications in Business and Economics*, Oxford University press.

Helton, J. C. and Davis, F. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability engineering and system safety* **81**: 23–69.

Helton, J., Iman, R. and Brown, B. (1985). Sensitivity analysis of the asymptotic behavior of a model for the environmental movement of radionuclides, *Ecological Modelling* **28**: 243–278.

Holthuijsen, L. (2007). *Waves in oceanic and coastal waters*, Cambridge University Press.

Howard, R. A. (1966). Information value theory, *IEEE Transactions on systems science and cybernetics* **sec-2, no. 1**: 22–26.

Hughes, T. (1987). *The Finite element method, Linear static and dynamic Finite element analysis*, Prentice-Hall, Inc.

Iman, R. L. and Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables, *Communications in Statistics - Simulation and Computation* **11:3**: 311–334.

ITI (2012). Test report, internal report boskalis, *Technical Report CP - 1204004 - A*, Materials Laboratory, Industrial Technology Institute.

Komen, G., Hasselman, S. and Hasselman, K. (1984). On the existence of a fully developed wind-sea spectrum, *Journal physical Oceanography* **14**: 1271–1285.

Loveless, J., Debski, D. and MacLeod, A. (1998). Sea level set-up behind detached breakwaters, *Proceedings of the International Conference on Coastal Engineering, ASCE* .

Lowe, R., Falter, J., Bandet, M. and Pawlak, G. (2005). Spectral wave dissipation over a barrier reef, *Journal of geophysical research* **110**: C04001.

Maurer, M., Kelane, Y. and Bechteler, W. (1998). The effects of inaccurate input on the modelling of deposition of suspended sediment, *Modelling Soil Erosion, Sediment Transport and Closely Related Hydrological Processes* **249**: 363–374.

McKay, M. D., Beckman, R. J. and Conover, W. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* **42:1**: 55–61.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative risk and Policy Analysis*, Cambridge University Press.

Nelson, R. (1996). Hydraulic roughness of coral reef platform, *Applied Ocean Research* **18**: 265–274.

O'brien, R. (2007). A caution regarding rules of thumb for variance inflation factors, *Qualtiy & Quantity* **41**: 673–690.

Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33 - 3**: 1065–1076.

Seelig, W. (1983). Laboratory study of reef-lagoon system hydraulics, *Journal of Waterway, Port, Coastal and Ocean Engineering* **109**: 380–391.

Soulsby, R. (1997). *Dynamics of marine sands*, Thomas Telford Publications.

the SWAN team (2006a). Swan technical documentation, *Technical report*, Delft University of Technology.

the SWAN team (2006b). *SWAN user manual*, swan cycle iii version 40.51 edn, Delft University of Technology.

Tolman, H. (2009). User manual and system documentation of wavewatch iii version 3.14, *Technical report*, U.S. Department of Commerce National Oceanic and Atmospheric Administration National Weather Service National Centers for Environmental Prediction.

Toth, Z. and Kalnay, E. (1993). Ensemble forecasting at nmc: The generation of pertubations, *Bulletin of the American M* **74**: 2317–2330.

van der Klis, H. (2003). *Uncertainty analysis applied to Numerical Models of River Bed Morphology*, PhD thesis, Technical University Delft.

van Gelder, P. H. A. J. M. (2000). *Statistical methods for the Risk-Based Design of Civil Structures*, PhD thesis, Technical University Delft.

Vousdoukas, M., Velegrakis, A., Paul, M. and Dimitriadis, C. (2012). Field observations and modeling of wave attenuation over colonized beachrocks, *Continental Shelf Research* **48**: 100–109.

Vreugdenhil, C. (1994). *Numerical methods for shallow-water flow*, Kluwer Academic Publishers, Dordrecht.

Appendices

Appendix A

Statistics

In this appendix standard statistical techniques are explained that are used to get an *estimation* of a parameter of the distribution(ϕ), like moments or fractiles. An *estimator* is a function of an observed sample that provides an estimate. Considering the classical or frequenstic point of view each parameter has a single, true, if unknown value. That is why a parameter should not be treated as a random variable (Morgan and Henrion, 1990).

A.1 Estimators

Consider a sample of observations of a random variable x with a sample size of m. The estimator for the mean and variance are;

$$\bar{x} \equiv \hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{A.1}$$

$$S^2 \equiv \hat{\sigma}^2 = \frac{1}{m-1}\sum_{i=1}^{m} (x_i - \bar{x})^2 \tag{A.2}$$

Note that the 'hat' denotes an estimation. Because of the central limit theorem the uncertainty of the sample mean is represented by a normal distribution. In that way the standard deviation of the sample mean can be estimated by;

$$S_m = \frac{S}{\sqrt{m}} \tag{A.3}$$

This can be used to estimate a confidence interval α for the mean of x. Consider Φ as a random variable of a standard normal distribution with a mean and variance of 0 and 1. Then let c be the deviation such that the range $(-c, c)$ encloses α probability;

$$P(-c < \Phi < c) = \alpha \tag{A.4}$$

In that way the confidence interval for the mean (μ) of x is described as;

$$\left(\bar{x} - c\frac{S}{\sqrt{m}}, \bar{x} + c\frac{S}{\sqrt{m}}\right) \tag{A.5}$$

For the estimation of fractiles it is convenient to have the sample in ascending order. Without loss of generality the sample is relabeled;

$$x_1 \leq x_2 \leq \ldots x_m \tag{A.6}$$

Now suppose to that the p^{th} fractile X_p needs to be estimated. A simple estimate is the sample value x_j, where $j = mp$. This requires that j is rounded to the nearest integer. This means that x_m is the best estimate for the maximum $X_{1.0}$. However the true maximum is almost certain to be larger than the observed maximum. Therefore a slightly better approach is to interpret the x_i observed value as the $i/(m+1)$ fractile;

$$\hat{X}_{i/(m+1)} = x_i \tag{A.7}$$

Now it is also possible to estimate a confindence interval for the fractile. Let c again be the deviation as in equation (A.4). The interval with confidence α for the pth fractile is approximately;

$$(mp - c\sigma, mp + c\sigma) \tag{A.8}$$

With mean $\mu = mp$ and variance $\sigma^2 = mp(1-p)$; a binomial distribution. Therefore the confidence interval for the p^{th} fractile can be estimated with (x_i, x_k) where;

$$i = \left\lfloor mp - c\sqrt{mp(1-p)} \right\rfloor \tag{A.9}$$

$$k = \left\lceil mp + c\sqrt{mp(1-p)} \right\rceil \tag{A.10}$$

Where the notation $\lfloor\rfloor$ and $\lceil\rceil$ stands for rounding respectively up and down to nearest integer.

A.2 Correlation

The (Pearson) correlation coefficient is a measure for the linear dependence between two variables. The *covariance* (Cov) between two variables is important for the determination of the correlation. The analytical expression for covariance is as follows (CUR, 1997);

$$\mathrm{Cov}[X_j, X_k] = E[(X_j - \mu_{X_j})(X_k - \mu_{X_k})] \tag{A.11}$$

Where also the following rules can be applied;

$$\mathrm{Cov}[X_j, X_j] = \mathrm{Var}[X_j] = \sigma^2_{X_j} \tag{A.12}$$

$$\mathrm{Cov}[X_j, X_k] = \mathrm{Cov}[X_k, X_j] \tag{A.13}$$

The correlation coefficient is then described as;

$$\rho_{X_i, X_j} = \frac{\mathrm{Cov}[X_j, X_k]}{\sigma_{X_j}\sigma_{X_k}} \tag{A.14}$$

The *sample correlation coefficient* can be calculated using the expressions denoted in section A.1 and is calculated with the following expression (Morgan and Henrion, 1990);

$$\rho_{x,y} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 \times \sum_{i=1}^{m}(y_i - \bar{y})^2}} \tag{A.15}$$

The (Pearson) correlation gives insight in the strength of linear relationship between two variables. However for non linear relationships it may not give a correct measure for statistical dependence. A method to reduce the effect of non-linear data is the use of *rank transformation* (Hamby, 1994).

Rank order correlation or the Spearman correlation coefficient is a measure for the strength of (non linear) monotonic relations (Morgan and Henrion, 1990). The observations of the two variables, (x, y) are given a specified rank (rank transformation) and from the ranked data sets the Pearson correlation coefficient is calculated using equation (A.15).

A.3 Bootstrapping

The bootstrap method is a method to get an estimation of *statistical uncertainty*; i.e. uncertainty of an estimated parameter (ϕ) of a distribution (Efron and Tibshirani, 1986). The method is based on resampling (with replacement) of a data set of observations, $\mathbf{x} = (x_1, x_2, \ldots, x_m)$. Consider B resampled data sets of $\mathbf{x}$;

$$x^{*^1}, x^{*^2}, \ldots, x^{*^B} \tag{A.16}$$

Now a parameter estimation can be made of every bootstrap sample which results in B parameter estimates;

$$(\phi^{*^1}, \phi^{*^2}, \ldots, \phi^{*^B}) = \phi^* \tag{A.17}$$

With these set of parameter estimates statistical techniques, described in previous section, can be used to derive an empirical distribution or confidence intervals of the parameter estimate.

A.4 Maximum likelihood

In many cases in the quantification of the uncertainties (section 3.2) the used distributions are Gaussian. For the estimation of the parameters of a Gaussian distribution ($\phi = (\mu, \sigma)$) the statistical methods described in section A.1 can be used, which is also referred as *method of moments*. However for some variables the data fits better with a Weibull distribution. For these cases the method of moments can not be used and the method of *maximum likelihood* is used instead. This method uses the likelihood function (CUR, 1997);

$$\mathcal{L}(\phi|\mathbf{x}) = f_x(x_1|\phi) f_x(x_2|\phi) \ldots f_x(x_m|\phi) \tag{A.18}$$

$$= \prod_{i=1}^{m} f_x(x_i|\phi) \tag{A.19}$$

Now the estimate of parameter ϕ is chosen with the maximum value for the likelihood function. To find this maximum value built in functions of MATLAB are used to eventually get the estimated parameter.

A.5 Kernel density estimation

The kernel density estimation is a non-parametric way to estimate the probability density fucntion of a random variable. The definition for the estimated probability density function $\hat{f}(x)$ is as follows (Parzen, 1962);

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) \tag{A.20}$$

where K is the *Kernel*; a symmetric but not necessarily positive function that integrates to one. The variable h is a smoothing parameter, mostly called the bandwidth. The bandwidth has to be chosen as small as the data allows for accurate results. For this study built-in functions of MATLAB had been used to determine the kernel density estimation.

A.6 Regression

A linear multiple regression model is used to replace a complex model with a simplified *response surface* (Hamby, 1994). The generalized form of the linear regression model denotes;

$$\hat{y}_i = b_1 + \sum_{j=2}^{k} b_j x_{i,j} + \epsilon_i \quad i = 1, 2, \ldots, m \tag{A.21}$$

Where y is the output variable or *variable to be explained* and x_j are the *regressors* or *explanatory variables*. The regression coefficients b_j are a measure for the linear sensitivity of variable x_j on output y. The last term ϵ are the *residuals* or the *error term*. Note that the index i is over m (sample) observations and index j is over k input variables. Equation (A.21) can be written in matrix form for m sample observations (Heij et al., 2004). Let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{2,1} & \cdots & x_{k,1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{2,n} & \cdots & x_{k,m} \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix} \quad e = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix} \tag{A.22}$$

In this notation equation (A.21) can be rewritten as;

$$y = bX + e \tag{A.23}$$

Where y is a $m \times 1$ vector of output observations, X is a $m \times k$ matrix of sample observations per variable, b is a $k \times 1$ vector of estimated parameters and e is a $m \times 1$ vector of residuals.
A method to obtain an estimation of regression coefficients b is by using the *least square estimator*, which is the sum of squares of the residuals. The vector of residuals can easily be computed using equation (A.23);

$$e = y - bX \tag{A.24}$$

The least square estimator is therefore denoted as;

$$S(b) = \sum e_i^2 = e'e = (y - bX)'(y - bX) \tag{A.25}$$

The estimates of b are found by the minimization of the least square estimator $S(b)$. For a full derivation see Heij et al. (2004).

A.6.1 Coefficient of determination

The coefficient of determination, commonly denoted as R^2 , is usually known as a measure for the 'goodness' of fit of the regression model. It is actually formally defined as the *relative explained sum of squares*. First consider a linear regression model with one variable;

$$y_i = a + bx_i + \epsilon_i \tag{A.26}$$

Where a is the intercept (b_1). Equation (A.26) can be rewritten in the following form; (Heij et al., 2004);

$$y_i - \bar{y} = b(x_i - \bar{x}) + \epsilon_i \tag{A.27}$$

So the difference from the mean $(y_i - \bar{y})$ can be decomposed as a sum of two components, a component corresponding tot the difference from the mean of the explanatory variable and a component described by the residual ϵ_i. The sum of squares of the first term $(y_i - \bar{y})$ consists also of two components;

$$\sum (y_i - \bar{y})^2 = b^2 \sum (x_i - \bar{x})^2 + \sum e_i^2 \tag{A.28}$$

$$SST = SSE + SSR \tag{A.29}$$

Where SST is called the *total sum of squares*, SSE the *explained sum of squares* and SSR the *sum of squared residuals*, which is the same as the least square estimator. The relative explained sum of squares, the coefficient of determination, is then denoted as;

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \tag{A.30}$$

From this we may conclude that R^2 is *the fraction of the total sample variation explained by the regression model*. If equation (A.30) and (A.29) are combined it shows that the R^2 is actually the squared correlation coefficient between y and its explained part $\hat{y} = Xb$ (Heij et al., 2004). In matrix form the coefficient of determination is denoted as follows;

$$R^2 = \frac{b'X'NXb}{y'Ny} = 1 - \frac{e'e}{y'Ny} \tag{A.31}$$

with

$$N = I - \frac{1}{m}L \tag{A.32}$$

where I is the identity matrix and L is a $m \times m$ matrix of ones. For the full derivation of the coefficient of determination see Heij et al. (2004)

A.6.2 Standard error

To derive a standard error for the regression coefficient b, it is assumed that the disturbances ϵ_i are normally distributed. The unbiased estimator for the variance of the model is then given by (Heij et al., 2004);

$$s^2 = \frac{1}{m-2} \sum \epsilon_i^2 \tag{A.33}$$

Considering the regression model of equation (A.26), the standard deviation of the regression coefficient is then given by;

$$s_b = \frac{s}{\sqrt{\sum (x_i - \bar{x})}} \tag{A.34}$$

The square root of equation (A.33) (s) is called the *standard error of the regression* and s_b the *standard error of the regression coefficient* b. In matrix form the unbiased estimator for the variance is therefore given by;

$$s^2 = \frac{e'e}{m-k} \tag{A.35}$$

The standard error for the j^{th} regression coefficient b_j is then given by;

$$s_{b,j} = s\sqrt{a_{j,j}} \tag{A.36}$$

Where $a_{j,j}$ is the j^{th} diagonal element of $(X'X)^{-1}$. For the full derivation and further assumptions on the derivations of the standard errors see Heij et al. (2004).

A.6.3 P-value

In this study the *p-value* is used for stepwise regression. It is a measure for a variable whether it is significant enough to include in the regression model. First a null hypothesis is determined, which is $H_0 : b_j = 0$. Which means that the regression coefficient is zero and therefore the variable has no influence on the output. The null hypothesis is tested against the hypothesis $H_1 : b_j \neq 0$. The *p-value* is then a probability of obtaining a test statistic, assuming that the null hypothesis is true. Actually it is more common to say that when the *p-value* is lower than a determined significance level α the null hypothesis H_0 is *rejected*. Which means that it is very unlikely that the observed result will be under the null hypothesis. Note that the *p-value* is not the probability that the null hypothesis is true. The significance level in this study is $\alpha = 0.05$. That means that if variables have a *p-value* higher than 0.05 the variable is not included into the model, because then it is likely that they would have a regression coefficient of zero and will not improve the full regression model. More information about test statistics and *p-values* can be found in Heij et al. (2004).

Appendix B

Numerical models

In this appendix technical background of the numerical models is provided. The total model consists of two modules; the wave module and the hydrodynamic (and morphological) model.

B.1 SWAN

For the computation of the local wave field the SWAN wave model is used. It is a third generation wave model that is widely used by scientist and engineers for research and consultancy practice (Holthuijsen, 2007). The SWAN model is based on implicit propagation schemes and is therefore always numerically stable. It is based on the action balance equation since SWAN accounts for wave-current interactions. For small scale computations the formulation of the action balance equation in Cartesian coordinates reads;

$$\begin{aligned}\frac{\partial N(\sigma,\theta;x,y,t)}{\partial t}+\frac{\partial c_{g,x}N(\sigma,\theta;x,y,t)}{\partial x}+\frac{\partial c_{g,y}N(\sigma,\theta;x,y,t)}{\partial y}\\ +\frac{\partial c_\theta N(\sigma,\theta;x,y,t)}{\partial \theta}+\frac{\partial c_\theta N(\sigma,\theta;x,y,t)}{\partial y}=\frac{S(\sigma,\theta;x,y,t)}{\theta}\end{aligned} \tag{B.1}$$

Where the action density $N(\sigma,\theta)=E(\sigma,\theta)/\sigma$, σ is the relative radian frequency and θ direction. Energy is propagating with group velocity c_g and the last term S is the source term. For more information about this balance see Holthuijsen (2007). This equation can be transformed to spherical coordinates for large scale computations (Holthuijsen, 2007). In stationary situations and in the absence of ambient currents [1], as in this study, the equation can be reduced to the following energy balance equation;

$$\frac{\partial c_{g,x}E(\omega,\theta;x,y,t)}{\partial x}+\frac{\partial c_{g,y}E(\omega,\theta;x,y,t)}{\partial y}+\frac{\partial c_\theta E(\omega,\theta;x,y,t)}{\partial \theta}=S(\omega,\theta;x,y,t) \tag{B.2}$$

Where $E(\omega,\theta)$ is the energy density spectrum, dependent on the absolute radian frequency (ω) and direction (θ). For this study a JONSWAP-spectrum is imposed at the boundary as initial wave condition with an imposed directional spreading depending on the type of wave condition; swell or wind wave condition. A directional distribution of $\cos^2\theta$ is used with $\theta=10^\circ$ for swell conditions and $\theta=30^\circ$ for wind wave conditions. Exponential wind growth is taken from Komen et al. (1984) with dissipation caused by whitecapping taken into account. Depth induced breaking is used with a constant breaker parameter $\gamma=0.73$. For dissipation by bottom friction the JONSWAP model

[1] The model is used off-line so no wave-current interactions are taken into account

is used with constant $C = 0.038$ for swell conditions and $C = 0.067$ for wind wave conditions. Also the effect of both quadruplet and triad wave-wave interactions are taken into account. This and more information about the SWAN wave model can be found in (the SWAN team, 2006a,b) The model is also capable of using unstructured grids (see section B.2.1), however the variables are determined at the triangles intersections instead of the elements. The results of the SWAN computation are interpolated to the elements so they can be used as input for the hydrodynamic module, FINEL 2D.

B.2 FINEL 2D

For the simulation of the hydrodynamic and morphological processes, numerical model FINEL 2D is used. FINEL stands for FInite ELement method where the model is based on. The model is capable of simulating both 2DH and 2DV situations, however for this study only the 2DH (depth-averaged) variant is used.

B.2.1 Unstructured grids

The model uses unstructured grids where the model area is divided in triangles, the so-called elements. Since the triangles may vary in size and shape along the area it allows for very flexible schematisations. In this case the area in vicinity of the reef has a higher grid resolution compared to the rest of the model area.

B.2.2 Hydrodynamics

The hydrodynamics are simulated by solving the (two dimensional) shallow water equations. The governing equations are the continuity and momentum equation. For an overview on the shallow water flows see Vreugdenhil (1994). For the solution the discontinuous Galerkin method is used (Hughes, 1987), in which the three flow variables (waterlevel and two velocity components) are taken constant in each element, see figure B.1.

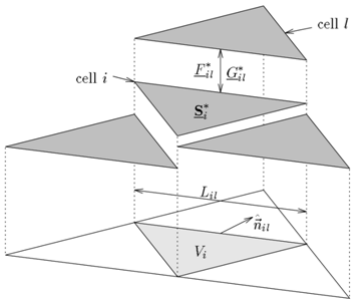

Figure B.1: Discontinuous Galerkin method; constant flow variables in each element

As the momentum equations only contain first order derivatives in space, they can be written as;

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F} = \mathbf{H} \tag{B.3}$$

where

$$\mathbf{U} = \begin{pmatrix} h \\ uD \\ vD \end{pmatrix} \quad \mathbf{F} = \begin{pmatrix} uD & vD \\ u^2 + {}^1/_2 gh^2 & uvD \\ uvd & v^2 D + {}^1/_2 gh^2 \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} 0 \\ \frac{1}{\rho}\tau_{tot,x} - f_c vD - gDi_{b,x} \\ \frac{1}{\rho}\tau_{tot,y} + f_c vD - gDi_{b,y} \end{pmatrix} \tag{B.4}$$

Where h, u and v are respectively the water level and depth integrated velocity in x- and y-direction. The total water depth is denoted as $D = h + z_b$, while i_b is the bottom level gradient in x- or y-direction. The first term ($\mathbf{U}$) is the acceleration term, second ($\mathbf{F}$) the advective term and third ($\mathbf{H}$) source term. The source term contains the influence of bottom slope, Coriolis (f_c) and other external stresses (τ_{tot}), like bottom, wind or wave induced stresses. The flow variables are determined at the elements and not at the element sides, therefore the fluxes are not known beforehand. Such a problem involves a Riemann problem which is solved with a method of Roe (Glaister, 1993). This method guarantees strict mass and momentum conservation, but suffers from some numerical diffusion in stream-wise direction. An explicit time integration scheme is used. As this method restricts the time step, the time step is controlled automatically for optimum performance.

B.2.3 Morphodynamics

For the derivation of the sediment fluxes the formula of Soulsby - Van Rijn is used (Soulsby, 1997). The formula applies to total (bedload and suspended load) sediment transport in combined waves and currents on horizontal and sloping beds. The method is intended for conditions in which the bed is rippled. The formula is described as follows;

$$q_t = (A_{sb} + A_{ss})\,\bar{U}\left[\left(\bar{U}^2 + \frac{0.018}{C_D} U_{rms}^2\right)^{1/2} - \bar{U}_{cr}\right]^{2.4} (1 - 1.6\tan\beta) \tag{B.5}$$

where

$$A_{sb} = \frac{0.005h\left(d_{50}/h\right)^{1.2}}{\left[(s-1)gd_{50}\right]^{1.2}}$$
$$A_{bb} = \frac{0.012d_{50}D_*^{-0.6}}{\left[(s-1)gd_{50}\right]^{1.2}}$$
$$\bar{U} = \text{depth-averaged currrent velocity}$$
$$U_{rms} = \text{root-mean-square wave orbital velocity}$$
$$C_D = \text{drag coefficient due to current alone}$$
$$= \left[\frac{0.40}{\ln(h/z_0)-1}\right]^2$$
$$\bar{U}_{cr} = \text{treshold current velocity}$$
$$\beta = \text{slope of bed in streamwise direction, positive if flow runs uphill}$$
$$h = \text{water depth}$$
$$d_{50} = \text{median grain diameter}$$
$$z_0 = \text{bed roughness length} = 0.006 \text{ m}$$
$$s = \text{relative density of sediment}$$
$$g = \text{acceleration due to gravity}$$
$$\upsilon = \text{kinematic viscosity of water}$$
$$D_* = \left[\frac{g(s-1)}{\upsilon^2}\right]^{1/3} d_{50}$$

The suspended sediment transport is usually 5-8 times larger than the bed load transport. The root-mean-square wave orbital velocity (U_{rms}) is calculated using the local significant wave height (H_s), peak period (T_p) and water depth (D) and is determined using figure B.2 which requires the zero-crossing period $T_z = 0.781T_p$ and $T_n = (D/g)^{0.5}$. In this case the JONSWAP curve is used since SWAN is using the same spectrum shape.

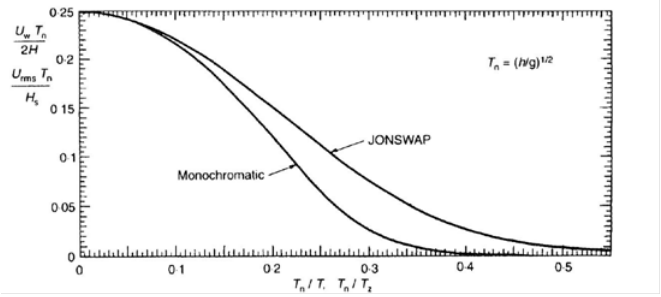

Figure B.2: Determination of U_{rms}, from Soulsby (1997)

The threshold current velocity ($\bar{U}_{cr}$) is calculated using the following expressions (Soulsby, 1997);

$$\bar{U}_{cr} = 0.19(d_{50})^{0.1} \log\left(\frac{4h}{d_{90}}\right) \qquad \text{for} \quad 0.1 \leq d_{50} \leq 0.5mm \tag{B.6a}$$

$$\bar{U}_{cr} = 8.50(d_{50})^{0.6} \log\left(\frac{4h}{d_{90}}\right) \qquad \text{for} \quad 0.5 \leq d_{50} \leq 2mm \tag{B.6b}$$

For the determination of threshold current velocity both the median grain size d_{50} and the d_{90} grain size are required. These variables are determined using different samples taken from the area of interest. Sieve curves can be made of the sample data and are shown in figure B.3. By taking the average of all the median and 90^{th} percentile grain sizes the d_{50} and d_{90} are determined;

$$d_{50} = 360 \quad \mu m$$
$$d_{90} = 760 \quad \mu m$$

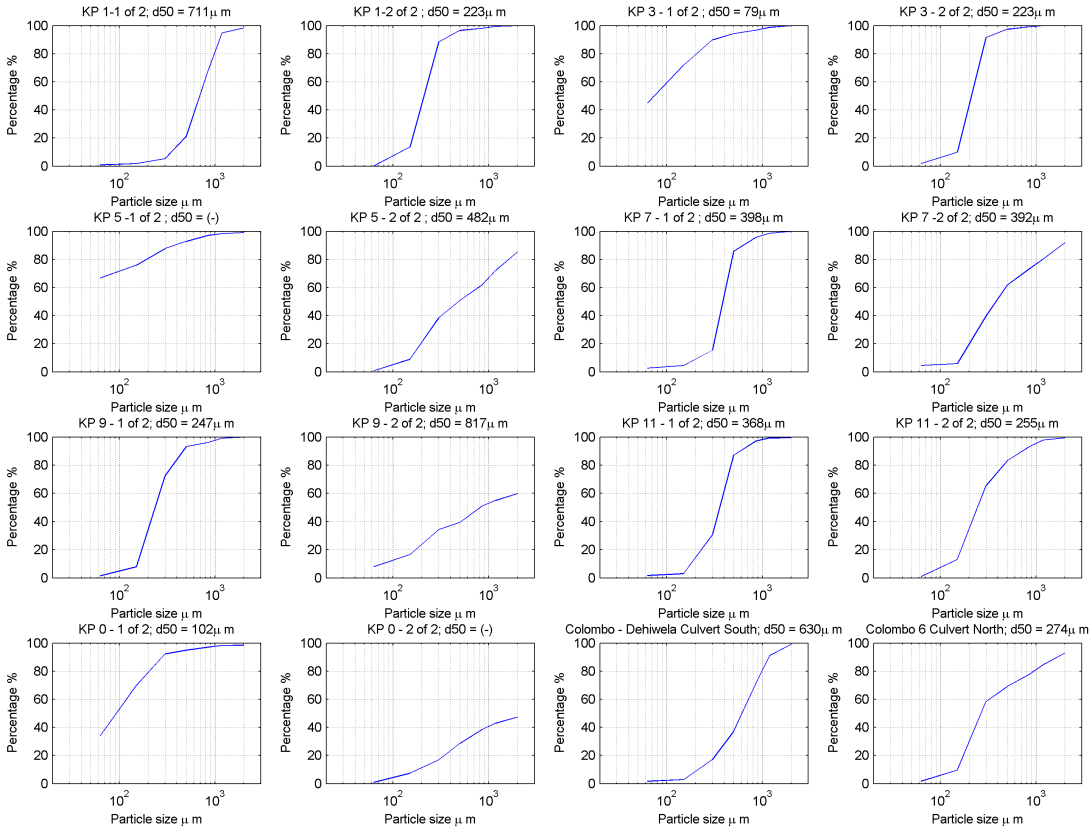

Figure B.3: Sieve curves of sample data; data taken from (ITI, 2012)

Suspended transport usaully lags the hydrodynamics forcing. In FINEL2D this is implemented according to the method of Galappatti (1983), where the lag is included with adaptation time T_a. The sediment concentration c is then calculated with the following advection equation;

$$\frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} + v\frac{\partial c}{\partial y} = \frac{c_{eq} - c}{T_a} \tag{B.7}$$

in which

$$T_a = \frac{h}{w_s} \tag{B.8}$$

Where w_s is the sediment settling velocity which is calculated using the nominal grain diameter and water temperature. The equilibirium sediment concentration c_{eq} is derived directly from the sediment transport equation.

Appendix C

Estimates

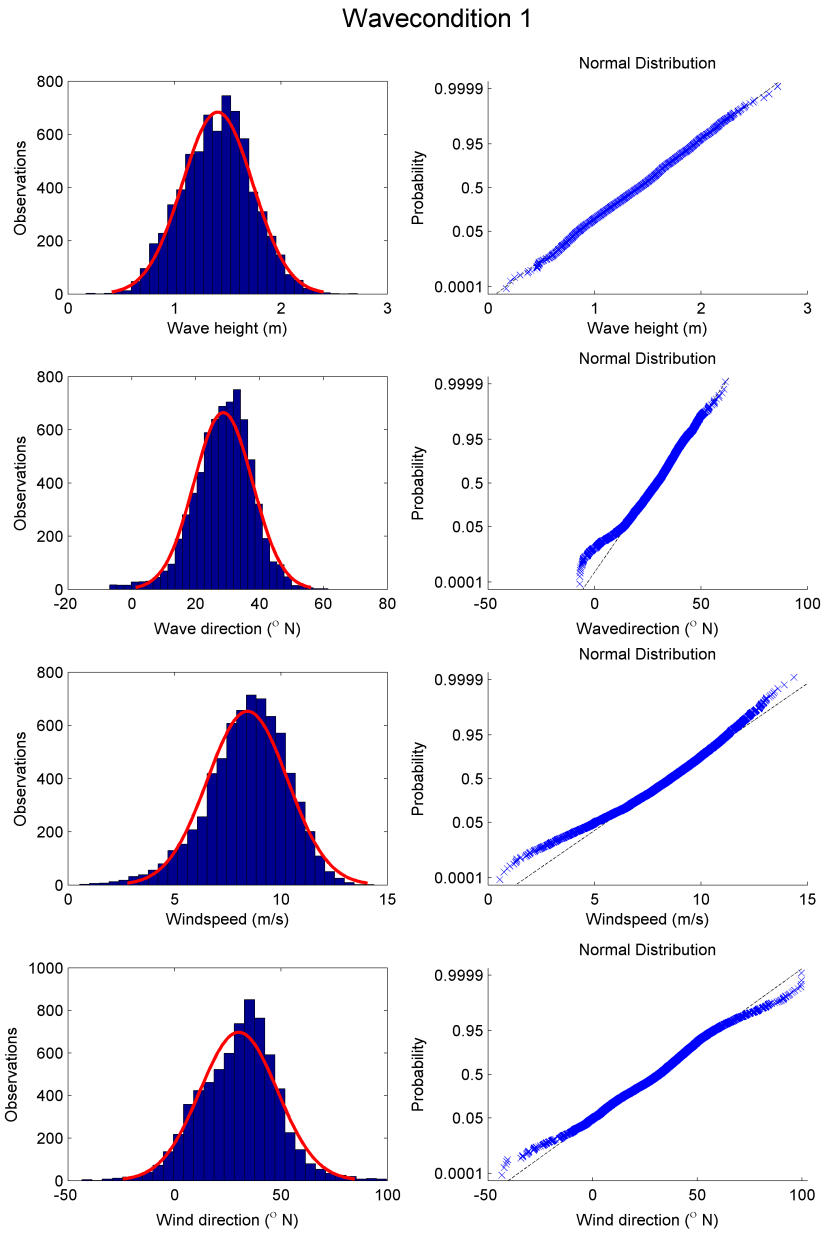

Figure C.1: Distribution estimations - Wavecondition 1

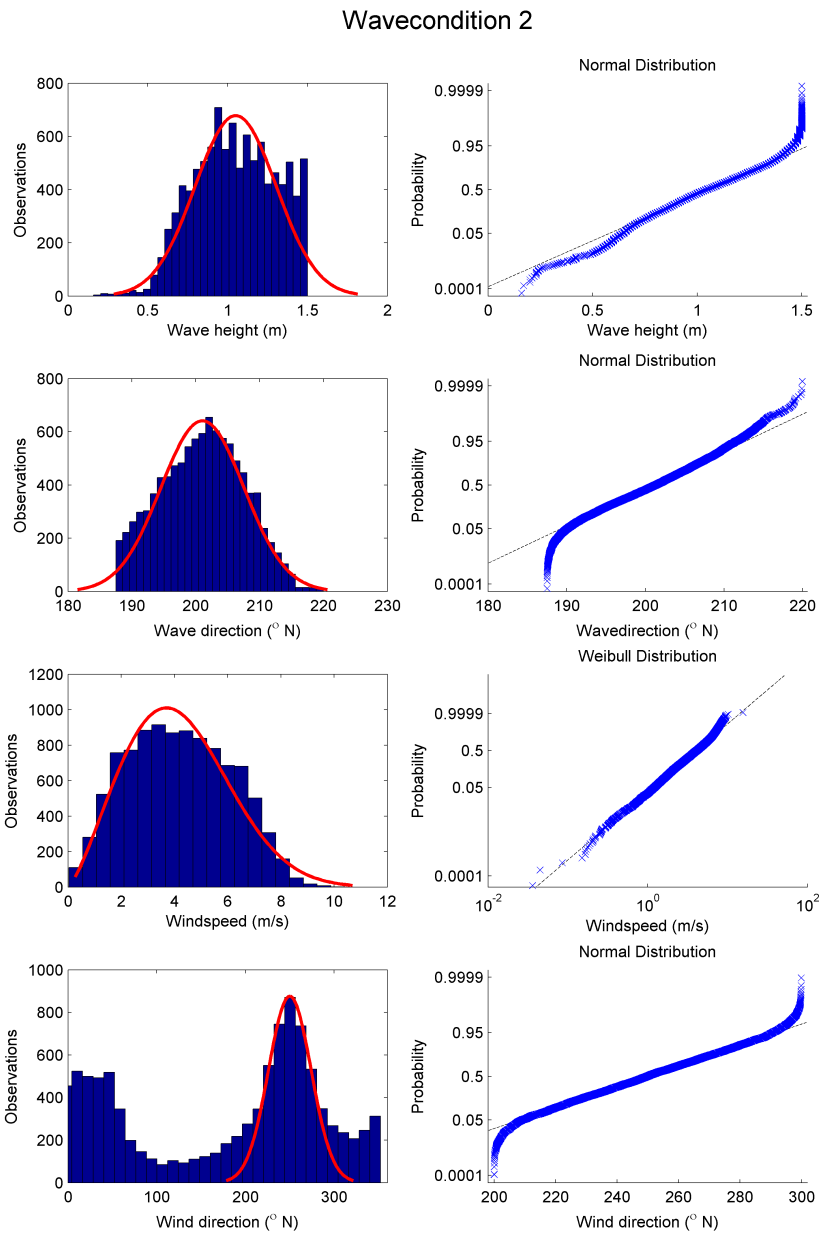

Figure C.2: Distribution estimations - Wavecondition 2

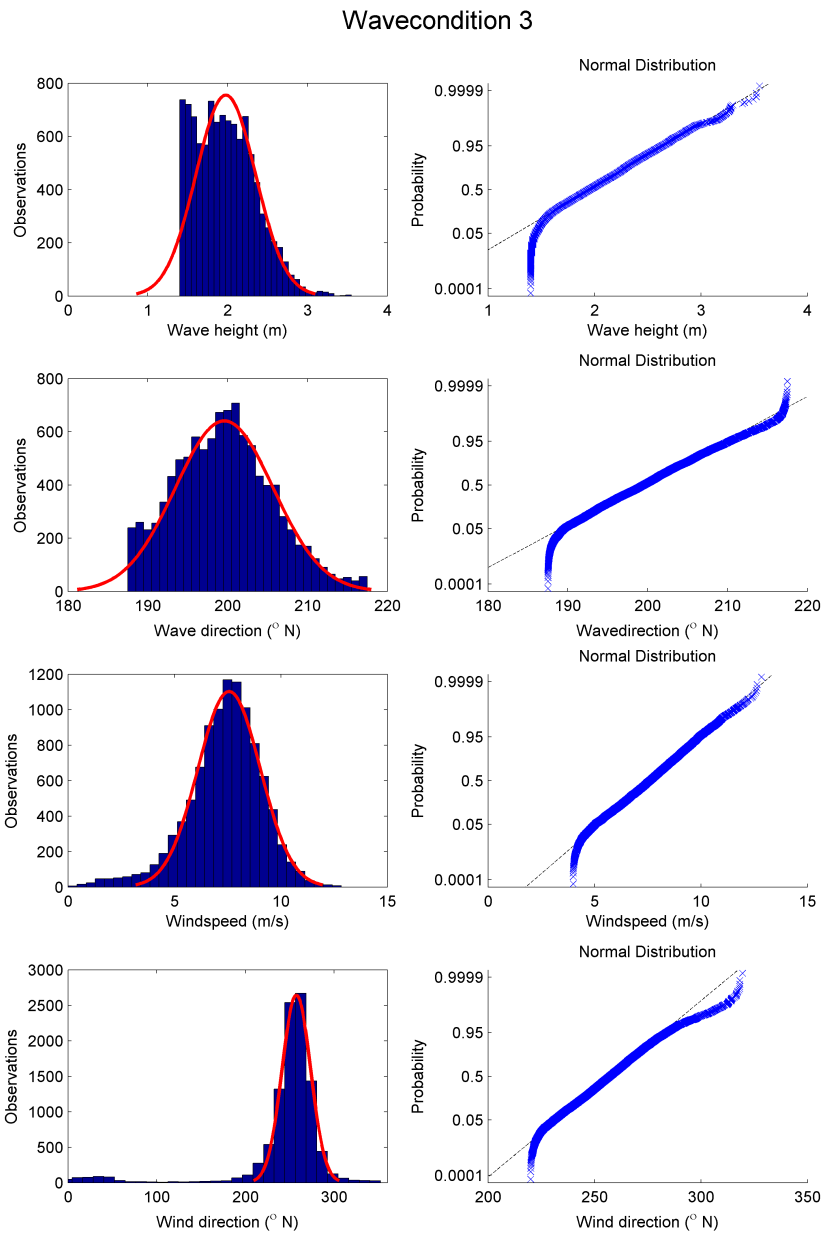

Figure C.3: Distribution estimations - Wavecondition 3

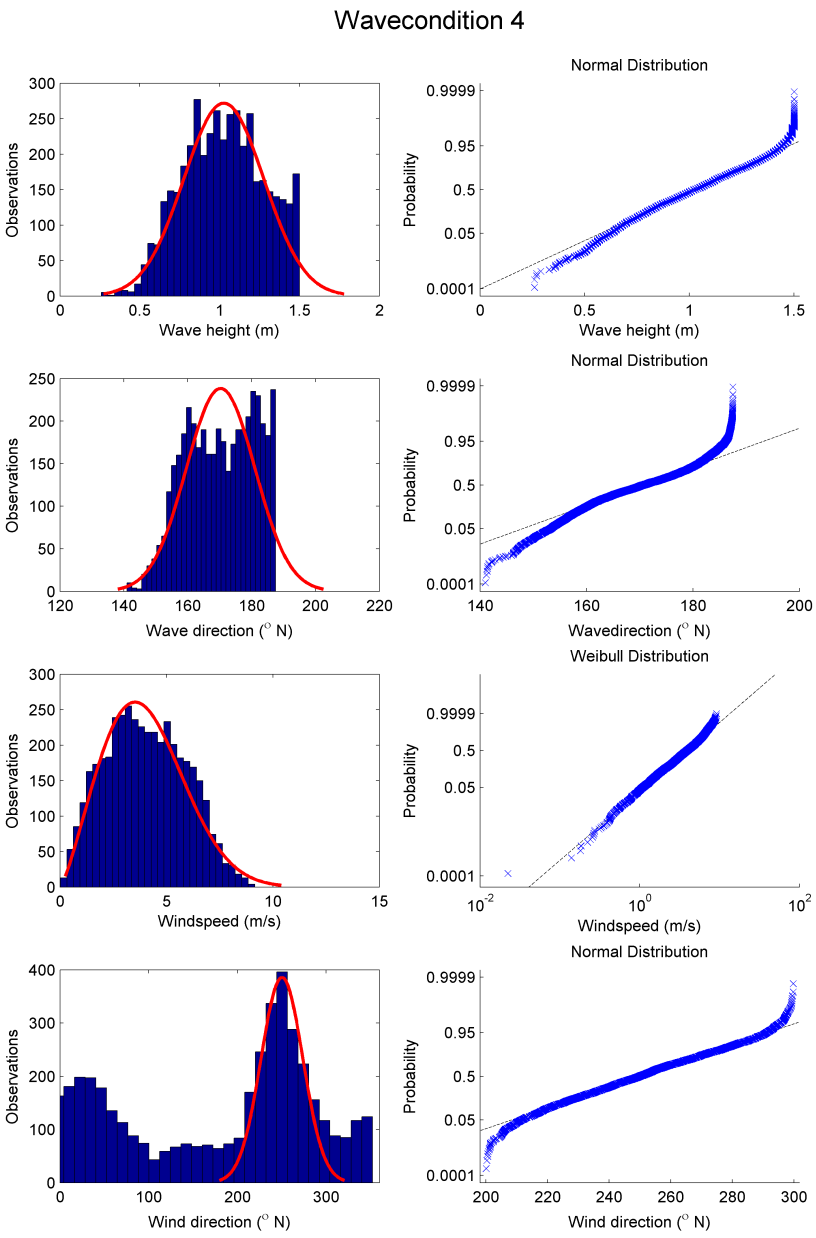

Figure C.4: Distribution estimations - Wavecondition 4

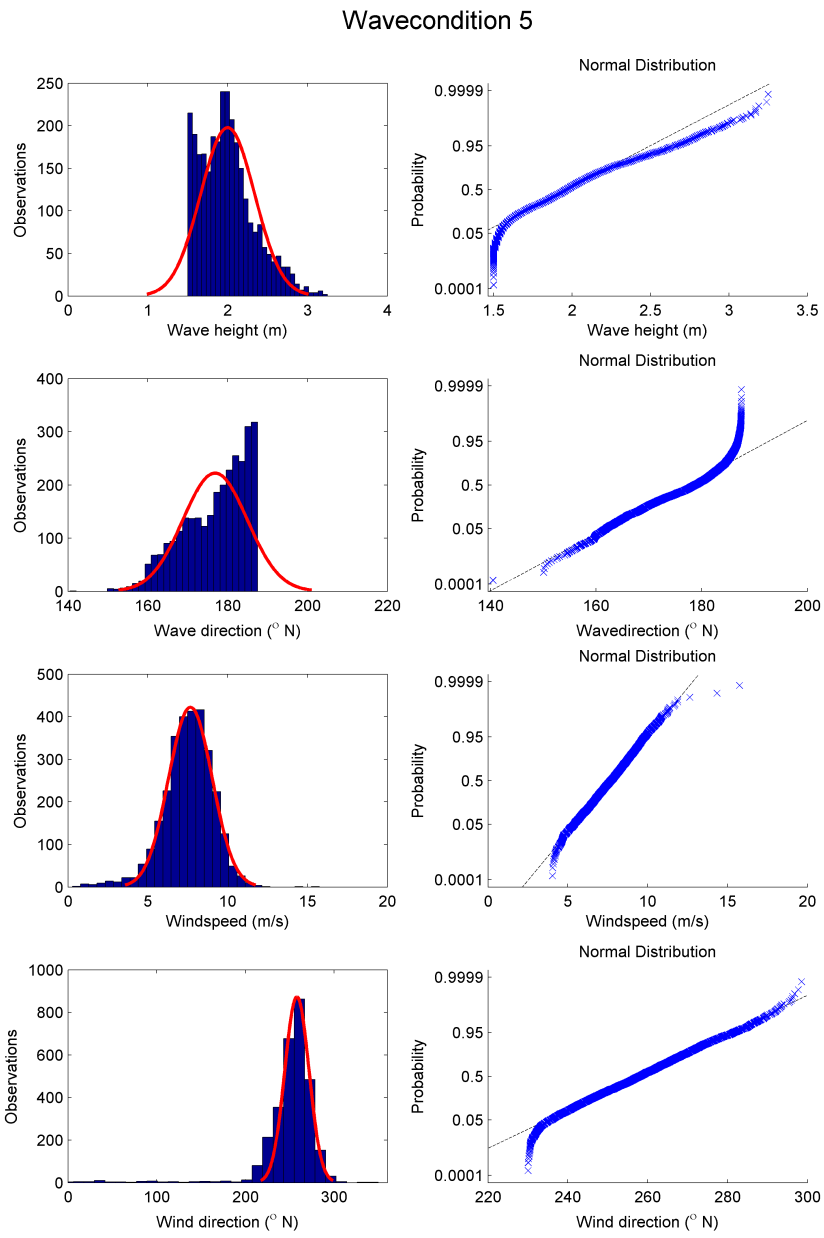

Figure C.5: Distribution estimations - Wavecondition 5

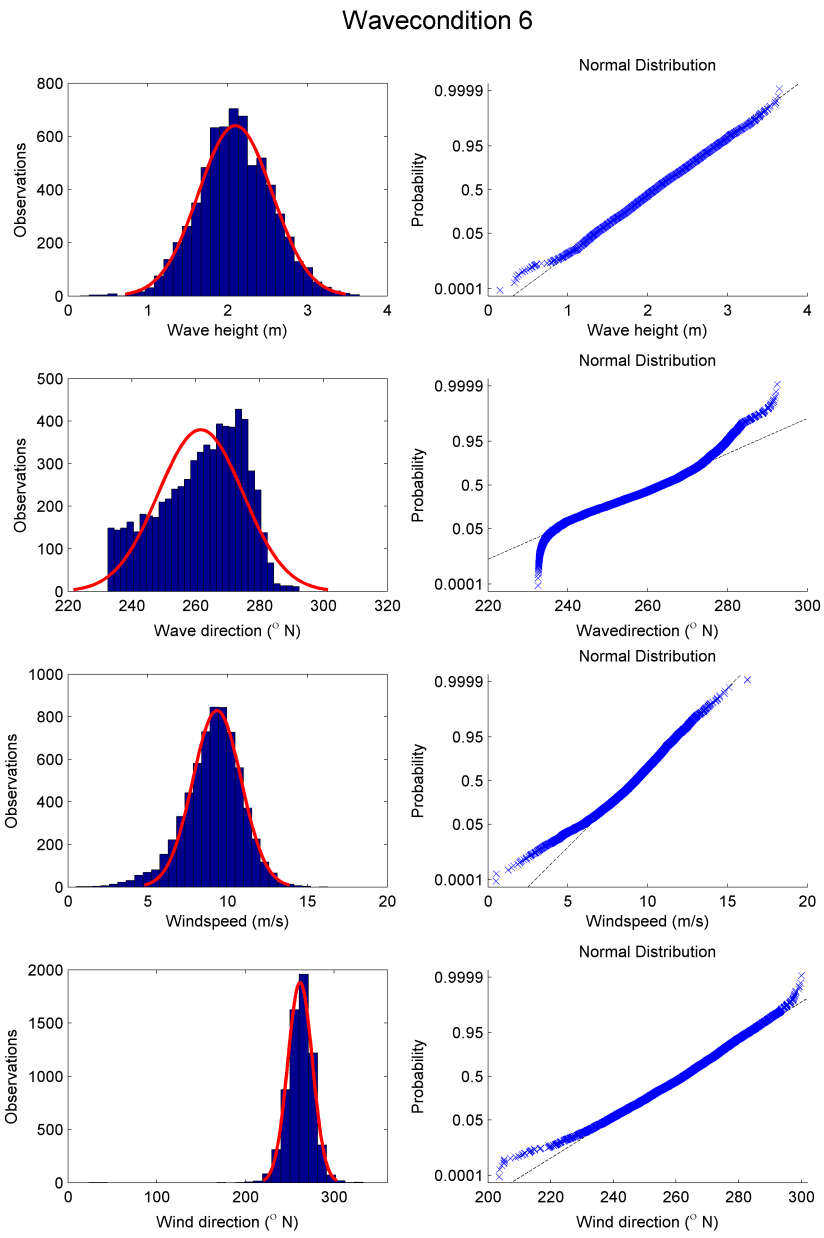

Figure C.6: Distribution estimations - Wavecondition 6

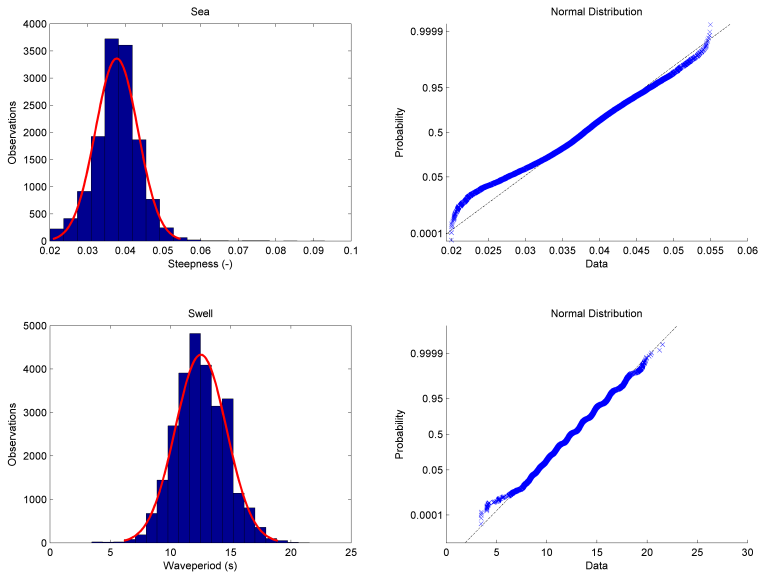

Figure C.7: Distribution estimations - Waveperiods

Appendix D

Correlations

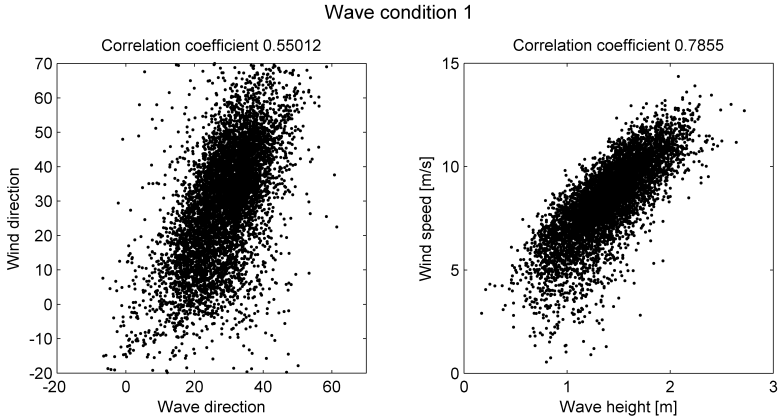

Figure D.1: Correlation coefficients - Wavecondition 1

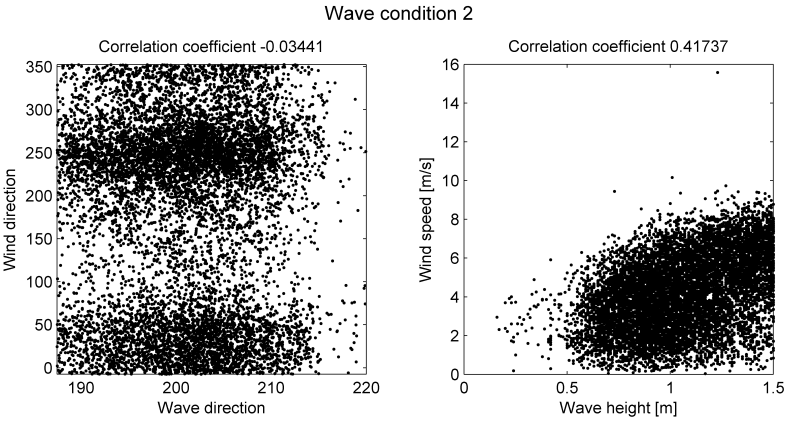

Figure D.2: Correlation coefficients - Wavecondition 2

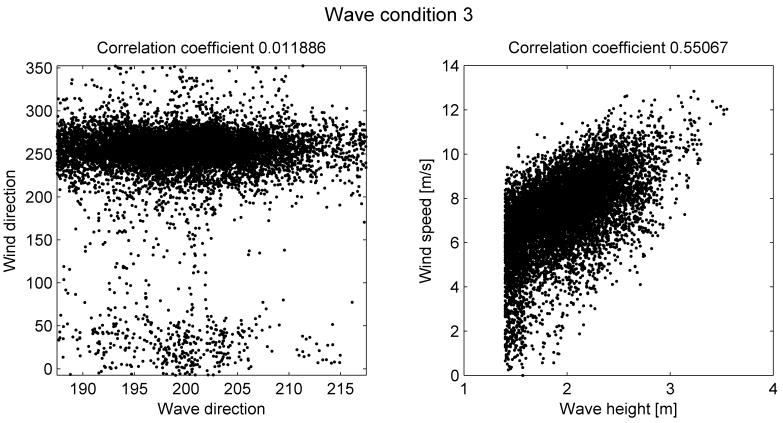

Figure D.3: Correlation coefficients - Wavecondition 3

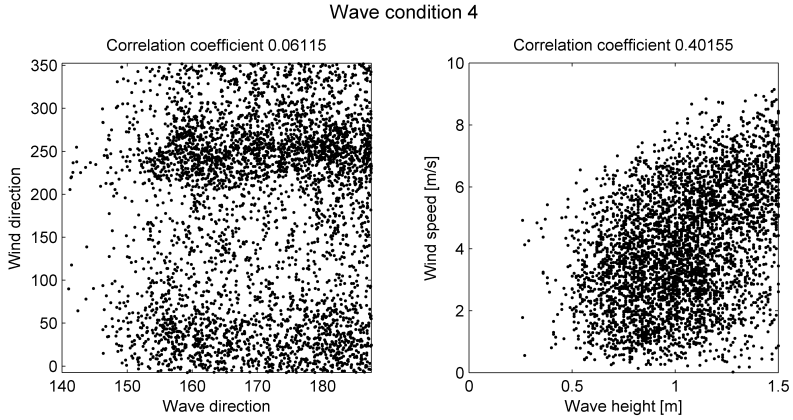

Figure D.4: Correlation coefficients - Wavecondition 4

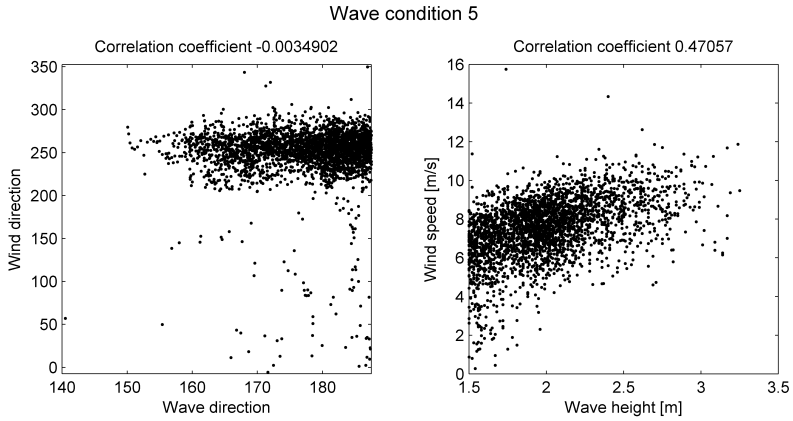

Figure D.5: Correlation coefficients - Wavecondition 5

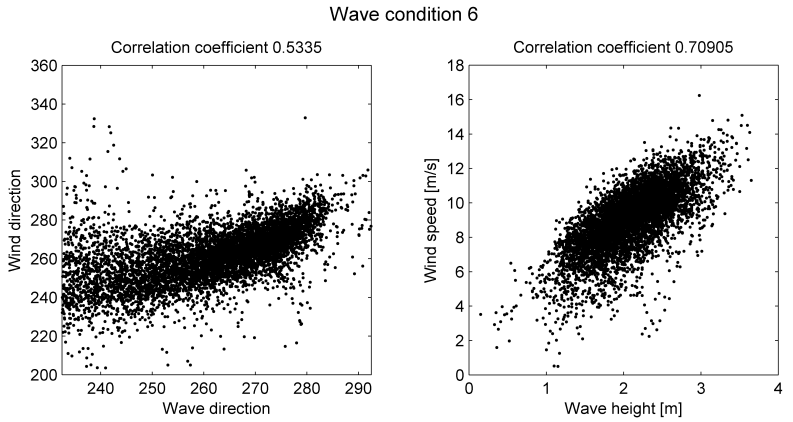

Figure D.6: Correlation coefficients - Wavecondition 6