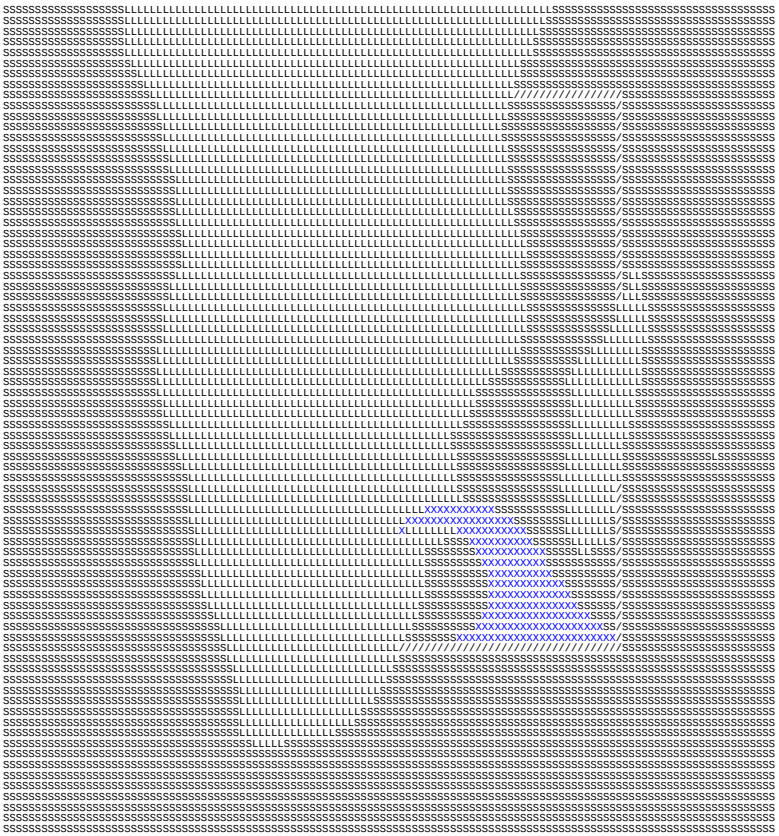
100-year waves in the Mozambique Channel

Master thesis in Physical Oceanography



Rasmus Kanstad Kulseng

UNIVERSITY OF BERGEN
Geophysical Institute
· June 2010 ·

Acknowledgements

First of all, I want to convey my sincere gratitude to my supervisors Knut Barthel and Øyvind Breivik for being forthcoming and flexible in the process of providing me a thesis, in which both personal and academical motivation has driven me. A special thanks goes to Øyvind for his highly impressive and educational manner of guiding me through my speed bumps.
Secondly, Ole Johan Aarnes, deserves a big thanks for sharing his support and expertise within statistical modeling of extreme values.
My friends and colleagues at the University of Marine Sciences in Quelimane. Your local knowledge and input has been vital throughout this work. A big thanks goes to Tor Gammelsrød and Dr. Antonio Mubango Hoguane, who have, on behalf of the NOMA project, given me the opportunity to spend a memorable month at this unique university.

But how can I forget...
... my mom and dad for their never ending care and generosity, but most importantly for their influence on my passion to the Sea.
Two unquestionable thanks go to Mia and Madel for just being!
Lastly, and with regret, I wish to thank my fellow students for their friendships and laughs. It is almost sad it is over..

Rasmus K. Kulseng

Bergen · June 2010 ·

Abstract

Six-hourly significant wave height (SWH) data covering the period September 1957 to August 2002 is obtained from the ERA-40 reanalysis. This data, generated by ECMWF's own configured global WAM model, stems from the first reanalysis in which a wave model is coupled to an atmospheric model. By accounting for the wave-atmosphere interaction this is the most comprehensive ocean state dataset available (Uppala *et al.*, 2005). Based on this data, return values of SWH are estimated at five locations within the Mozambique Channel. In this work an extreme value is defined as the highest value of SWH which, on average, is only exceeded once during a period of 100 years, namely the 100-year return value. These extreme values are found by fitting both a Generalized Extreme Value (GEV) and Generalized Pareto (GP) distribution to SWH data, and thereafter, based on the behavior of the model, extrapolated to yield the 100-year return value. A thorough introduction to both extreme value models will be given. Furthermore, an introduction to goodness-of-fit tests which assess the validity of the model fits is given, followed by a routine which estimates the confidence interval (CI) for the return values. However, as the ERA-40 data does not account for bottom-wave interaction and, additionally, underestimates high wave values (Caires & Sterl, 2003a), the SWH data is refined by use of an altimeter-validated high resolution WAM model, covering the Mozambique Channel during 2001. Functional relationships are established between the five ERA-40 locations and the corresponding location of the self-run WAM model. By applying these on the ERA-40 data, the superior properties from the high resolution model, including bottom interaction, is extrapolated back to 1957. Along with return values, the temporal variability of SWHs within the Mozambique Channel is addressed. The results show return values up to 7.0 m with a 95% CI ranging from 6.6 to 7.5 m in the south, a value which decreases to 5.0 m in the northern opening of the Mozambique Channel. Based on monthly means, the extreme values are likely to occur during the winter months of June and July.

CONTENTS

Chapter 1

INTRODUCTION

Awareness regarding the most extreme conditions occurring at sea is of life saving importance for the high concentration of human settlements along the coast and marine activities, like fishing and shipping. When engineers design ships, offshore oil rigs and other coastal structures, information regarding extreme events are crucial for the calculation of the structure's strength. For the 2 700 km coast line of Mozambique and its adjacent ocean this is no exception. The newly achieved political stability and the improving infrastructure both for roads and gas transmission, result in a rapid growth in natural gas and oil exploration and production offshore as well as onshore (Zacarias, 2009).
40% of the Mozambiquan population lives in the coastal districts (Chemane *et al.*, 1997), of these about half is sustained by the natural resources which include both industrial and artisanal fisheries and marine fauna (Hoguane, 2007). De Young (2006, and references therein) estimates the fleet to consist of 186 registrated industrial vessels and 15 269 artisanal vessels. These operate predominantly in the *Delagoa Bight* in the south and the *Sofala Bank* in the centre, which are the two main shelves. With a continental shelf extending 145 km offshore, the *Sofala Bank* is the biggest and is where most of the industrial shallow-water shrimp fishery is done (Saetre & Silva, 1979; De Young, 2006), contributing to the country's economy (De Young, 2006; Hoguane, 2007).
In the southern region of Mozambique, referred to as the region of lagoons, the direct impact of the waves give rise to an unbalanced displacement and distribution of sand. This, accompanied with strong tidal currents, cause a constantly changing coastal regime. Furthermore, international shipping routes to and from the southern African harbors navigate through the Mozambique Channel (Schreier *et al.*, 2007), potentially exposed to hazardous extreme wave events.

As a means of precaution and awareness motivated by the above mentioned events, the most extreme wave conditions experienced, for example during an offshore installation's N years of lifetime, are estimated. This N-year return value is the highest wave value, in this case significant wave height (SWH), which is expected to be exceeded on average once every N year (Coles, 2001), and will be denoted as H_s^N. In this study, in accordance to previous work (Cooper & Forristall, 1997; Soares & Scotto, 2001; Caires & Sterl, 2003a; Alves & Young, 2003; Caires & Sterl, 2005; Naess & Gaidai, 2009), the 100-year return value will be addressed.

1.1 Area of investigation

The map in Figure 1.1 shows the locations at which the wave study will be focused. At each location a statistical analysis of 45 years of wave height data will be used to yield the N-year return value. Flag one represent an area of the coast which due to its facing angle is heavily exposed to the incoming propagating swells. The narrow continental shelf in this region allows swell to propagate close to the coast without significant dissipation, permitting the swell to arrive with the bulk energy intact. In this area erosion has become an increasing threat to the coastal communities (A. Hoguane Mubango[1], pers. comm.). Flag three is located on the *Sofala Bank* at depth of 12 m and is motivated by the high fishing activity. The *Rovuma Delta Basin* with its expanding gas and oil industry is represented by flag five. Flags two and four are located south and north, respectively, of the volcanic *Europe Island* and several volcanic sea mounts. These positions are chosen to give a general overview of the Mozambique Channel. The five geographical coordinates and their respective depths are presented in Table 1.1. Following is a description of the hydrography and meteorology of the Mozambique Channel region.

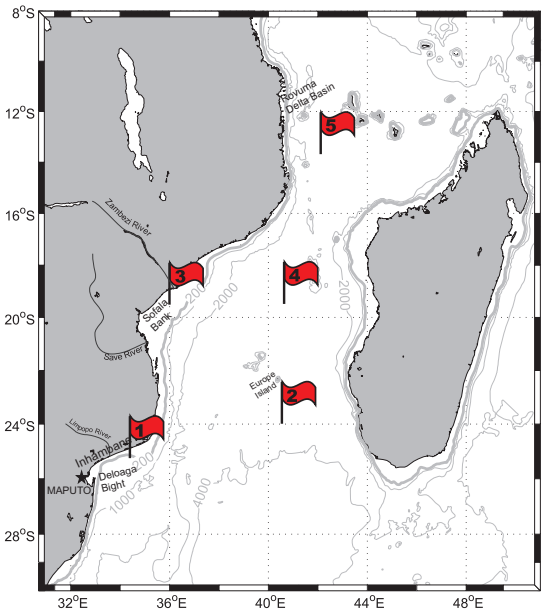


Figure 1.1. *Bathymetry of the Mozambique Channel and location at which the return value is to be estimated. See Table 1.1 for depths, coordinates and names.*

Table 1.1. *Positions and depths for the selected locations as illustrated in Figure 1.1.*

	Location name	**Depth [m]**	**Coordinates**
Flag 1	Surf location	192	34.5°E 25.5°S
Flag 2	South of island	3375	40.5°E 24.0°S
Flag 3	Sofala Bank	12	36.0°E 19.5°S
Flag 4	North of island	2552	40.5°E 19.5°S
Flag 5	Oil location	3017	42.0°E 13.5°S

[1]Director School of Marine Sciences, Quelimane, Mozambique.

1.1.1 Bathymetry

The coastline of Mozambique is in general smooth and covered by sandy beaches, but with the occurrences of few large headlands. The continental slopes and shelves are typically steep and narrow (Shaumann, 1998). Close to the Zambezi, Save and Limpopo Rivers we find exceptions, where large fan shaped continental shelves have been created due to the rivers' deposition of heavy sediments. The biggest of these is the *Sofala Bank*. Separating Mozambique and Madagascar we find the Mozambique Channel, which at the narrowest point in the North measures 400 km across. The depth ranges from 2000 to 3000 m with features of small volcanic islands and sea mounts. In the southern opening we find the Mozambique Basin with depths over 5000 m, encircled by the Mozambique and Madagascar Ridge on each side. To the west of the Mozambique Ridge we find the Natal Valley, a depression with a maximum depth around 4900 m. Here, the bottom is covered with sediments transported by the coastal currents. Further south the Natal Valley merges in to the deeper Transkei Basin in the south, extending beyond the latitude of the southernmost point of South Africa. These features are seen in Figure 1.2 along with the general surface currents, which in the following are addressed.

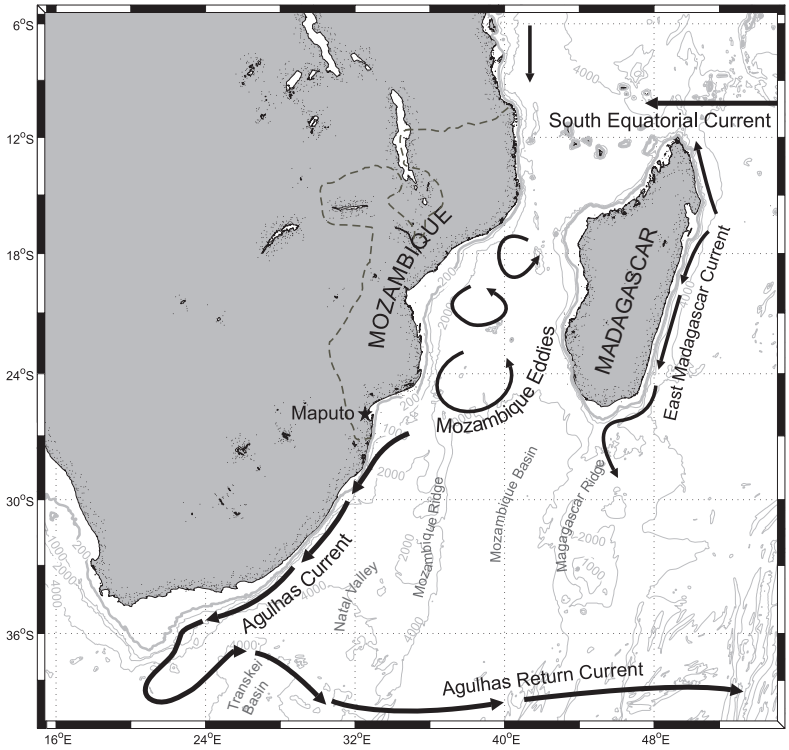


Figure 1.2. *Bathymetry and dominating surface currents of the southwest Indian Ocean (after Lutjeharms (2004)). Depth contours showing 200 (bold), 1000, 2000 and 4000 m according to GEBCO (2003).*

1.1.2 Currents

The general current description of the southwest Indian Ocean was first described by Michaelis (1923) and later by Sverdrup *et al.* (1942). Both descriptions closely correspond to the current system illustrated in Figure 1.2, where the South Equatorial Current feeds both the East Madagascar Current and the Mozambique Current (MC). Since the first descriptions, it is mostly the knowledge regarding

the MC which has changed. Several cruises collecting data (e.g. R/V Fridtjof Nansen 1977, 1978) and work based on the data, e.g. by Saetre & daSilva (1984) and Saetre (1985), established that the southerly moving MC in fact was the western section of channel-sized anticyclonic eddies within the Mozambique Channel.
Recent work by Harlander *et al.* (2009), based on long-term observations, supports the eddy phenomenon, now named the Mozambique Eddies, and further shows that these discontinuous southerly migrating eddies only appear four to five times a year. The Mozambique Eddies along with the MC supply the strong and narrow Agulhas Current with water from the South Indian Ocean and Red Sea. In a global oceanic climate perspective the Agulhas Current plays an important role through its inter basin exchange between the South Indian and the Southern Atlantic Oceans (Biastoch *et al.*, 1999, and references therein). In addition to the influence of the main current systems, the shallow coastal waters of the channel is expected to be influenced by both dominating winds and tides (Lutjeharms, 2004).

1.1.3 Wind pattern

Regional

The wind system in the Mozambique Channel is divided at 20°S, the latitude of the southernmost position of the north-south migrating inter tropical convergence zone (ITCZ) (Saetre, 1985). In the region north of 20°S, northerly and northeasterly winds from the monsoonal wind system of the Indian Ocean dominate during the austral summer (October to February). In this period the winds are in general weaker (Lutjeharms, 2004). Along the coast of Mozambique this division is set at 15°S, below which the monsoon has no influence. Between 20° and 25°S the southerly and southeasterly winds prevail (Saetre, 1985), whereas in the most southern part of the Mozambique Channel easterly winds dominate throughout the year.

Large scale

The most important wind system generating ocean waves is the consistently strong westerly winds that prevails over the Southern Ocean (Young, 1999). In fact Young (1999) obtained results confirming the Southern Ocean as the roughest ocean on earth. As seen from storm track studies done by Hoskins & Hodges (2005), the storm trajectories which originate over the southern part of the South America pass just south of Africa. Although Mozambique has a predominately easterly facing coastline, the waves produced by these storms are expected to hit the coast, particularly in the austral winter months (Young, 1999).

1.1.4 Tropical cyclones

During the months of summer, tropical cyclones (TC) of severe intensity are frequently observed in the southwest Indian Ocean. In fact, the TCs generated here account for 14% of the world total (Jury, 1993). Generally, the TCs migrate poleward (Lutjeharms, 2004), and pass east of Madagascar more frequently than through the Mozambique Channel (Williams *et al.*, 1984). The sea and swell generated by the intense winds of TCs cause potential risk to ship and coastal settlements. However, calculating the wave properties near TCs is often challenging

due to the strong horizontal wind gradient and the complexity of assessing the fetch area (Chang-Seng & Jury, 2010). Nevertheless, the most severe damage is caused by storm surges which may occur when TCs make landfall. In a case study by Chang-Seng & Jury (2010), the worst TC since 1951 is addressed in detail as it makes landfall in Mozambique the 29th of February 2000. With maximum wind speed recorded at 62 m/s and a swell height of 3 m, reports claimed 700 people dead, 800 000 affected and *us*$ 0.5 billion worth of destruction.
The extent to which TCs contribute to extreme wave conditions is not much studied in the southwest Indian Ocean. Yet, waves of 10 m are reported (Chang-Seng & Jury, 2010) hitting the coast of Madagascar, indicating a potential contribution to extreme wave events.

1.1.5 Tides

In the area around the *Sofala Bank*, the tidal range is more than 6 m, one of the highest in Africa (Coughanowr *et al.*, 1995). This tidal amplitude decreases both to the north and south as a product of a double standing wave system driven from both ends (Schwiderski, 1980). At the channel openings the amplitude is around 2 m (Lutjeharms, 2004, and references therein). As a consequence of these vast tidal ranges salt marshes and mangrove swamps are found in the low lying coastal regions and estuaries (Lutjeharms, 2004). This strong tidal current is an important contributor to the water motions at the wide and shallow parts of the *Sofala Bank*, where the sand banks are continuously moved (Lutjeharms, 2004).

1.2 Objective

The aim of this work is to produce reliable estimates of 100-year return values by means of two different extreme value distribution approaches. Opposed to similar work, which has been carried out on a global scale, e.g. Alves & Young (2003); Caires & Sterl (2005), the estimates obtained here are believed to be of a more accurate nature as the global wave data is enhanced to locally fit the locations of interest. This is done by means of an altimeter-validated nested model which is implemented over the region of interest, making it possible to find a functional relationship by which the global wave data is refined. Furthermore, data from both a coarse and nested model will be applied to address the wave climate in the Southern Ocean and the Mozambique Channel, respectively, both with respect to spatial and temporal variation.
Chapter 2 will give an introduction to the fundamental wave theory along with numerical modeling of ocean waves. In chapter 3 the details regarding the model configuration and inputs will be addressed in addition to the altimeter data which is used in the process of validating the model. The chapter proceeds by presenting the 45 years of historical wave data on which the return value estimates will be based on, followed by the theoretical aspect regarding the two extreme value approaches. The results are presented in chapter 4, with successive discussions that gradually evolve towards the final results. In the end, an appendix is found in which statistical parameters are defined, along with computational routines developed in *MatLab*, which may be utilized in further work addressing extreme value problems.

Chapter 2

THEORY

Ocean surface waves are propagating oscillations on the ocean-atmosphere interface. Figure 2.1 illustrates the common attributions used to describe a simplified sinusoidal wave:

- The wave height, H, is the vertical distance from the wave crest to the previous wave through. In the simple case of the sinusoidal wave $H = 2a$, where a is the maximum fluctuation around the zero level. The wave height is normally measured in meters.
- The wavelength, λ, is the horizontal crest-to-crest distance in meters.
- The period, T, is the time interval in seconds for two successive crests to pass a fixed point.

Other basic definitions used to describe surface waves are

- Steepness, H/λ, which is the ratio of the wave height to the wavelength. In the open ocean the steepness is typically ranging from 0.01 to 0.06 (Wright *et al.*, 1999).
- Frequency, f, is the amount of crests passing a fixed point during one second. The unit is Hertz (number per seconds), and is equivalent to $1/T$.
- Wave number, k, is a measure of the number of wave crests per unit distance and is given by
$$k = 2\pi/\lambda.$$

The characteristics of a wave reflect the forces by which the wave has been created. The period and wavelength are commonly used to make such a classification. Waves with period and length ranging from around 1 to 30 seconds and a few centimeters to several hundred meters, respectively, are termed ordinary gravity waves (Wright *et al.*, 1999). The predominant generating force for these waves is the wind. The perturbation of the sea surface is caused by the wind induced stress, and continues oscillating due to the restoring force exerted by the Earth's gravitation which strives to regain the equilibrium state. These wind induced gravity waves are the type dealt with in this thesis. In the following sections the basic theory will be addressed followed by an introduction to wave modeling.

2.1 Basic relationship

Consider a wave propagating in the x direction as illustrated in Figure 2.1. A fixed position on the wave interface, here represented with the value F, is

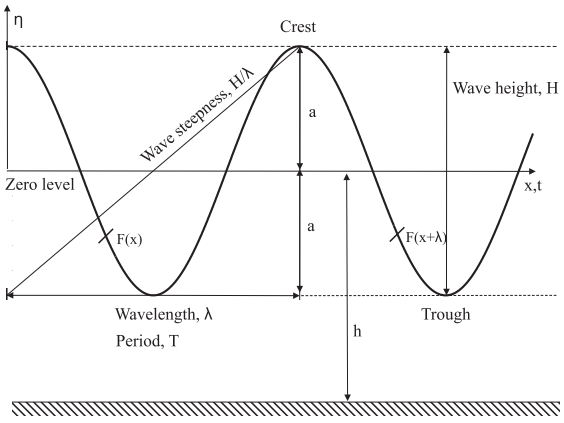


Figure 2.1. *Basic parameters used to describe wave attributes of a simplified sinusoidal wave.*

periodically symmetrical with the horizontal displacement of the wavelength: $F(x,t) = F(x+\lambda,t)$. The same periodicity is seen under a temporal displacement during the wave period: $F(x,t) = F(x,t+T)$. As the constant value F is moving with the speed c; $x = ct$, the function F must be of the form $F(x,t) = G(x-ct)$. Since $G(x-ct) = G[(x+\lambda)-c(t+T)] = G[x-ct+\lambda-cT]$ then $\lambda - cT = 0$ and $c = \lambda/T$, where c is termed the phase speed. This relationship holds for all truly periodic progressive waves. The vertical displacement of the free surface, η, around the zero level changes over time and space (t and x, respectively) and is expressed by: $\eta = A\cos(kx - \omega t)$, where $\omega = 2\pi/T$ is the radian frequency.
For further physical description of the waves, some assumptions must be made in order to apply the known dynamics of the ocean on the rather complex behavior of the surface waves (WMO 1998):

1. The water is incompressible: $D\rho_w/Dt = 0$, i.e. the density is constant and the continuity equation can be derived describing the mass flow in and out of a volume.

2. The particle motion is irrotational: $\nabla \times \vec{v} = 0$. This assumption allows us describe the flow by the gradient of a velocity potential ($\vec{v} = \nabla\Phi$), ϕ.

3. The water is an inviscid fluid: $\mu_w = 0$, where μ_w is the dynamic molecular viscosity of water, i.e. the water has no resistance to shear stress and therefore friction can be neglected.

With these assumptions, the linearized and frictionless equation of motion where the Coriolis term is neglected due to the small scale of waves, can be solved. By applying the boundary value conditions on the general wave solution, the important relation between phase speed and wave length is obtained:

$$c^2 = \frac{g}{k}\tanh kh, \tag{2.1}$$

where g is the gravitational acceleration. This equation holds for all depths, h, but, as we will see, can be simplified as we generalize into short (deep water) and long (shallow water) waves.

2.2 Influence of water depth

2.2.1 Deep water

In deep water ($h > \frac{\lambda}{2}$) the water particles move in almost closed circles as illustrated in Figure 2.2. At the surface the orbital diameter is the same as the wave height. The particle is moving in the same direction as the propagating wave at the crest of the wave, whereas in the trough the particle is moving in the opposite direction. During one wave period the water particle is close to its previous position, but is now slightly relocated horizontally with a net component of forward motion. This small forward displacement is called Stokes drift (e.g. λ=100m, H=3m and c=12.5 m/s gives a surface Stokes drift of 0.1 m/s (Pond & Pickard, 1983)) and is due to the decrease of particle speed with depth and the forward motion of the wave. The orbital diameter of deep water waves decreases exponentially with increasing depth, see Figure 2.2 (a), and at a depth equal to half the wavelength, the particle diameter is reduced to 4% of the surface wave height (Pond & Pickard, 1983). As $\lambda \ll h$ in deep water, $\frac{2\pi}{k} \ll h$, implying that

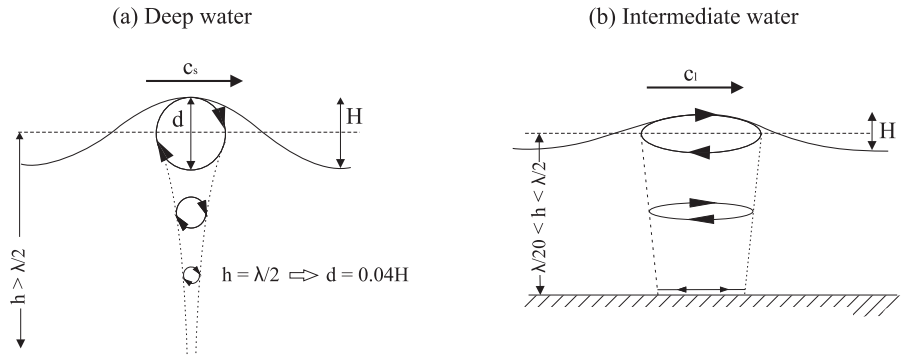


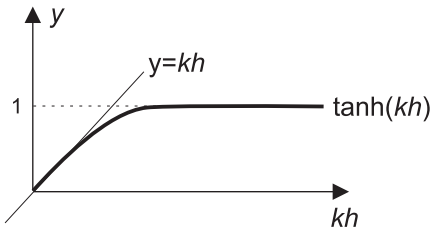
Figure 2.2. *The water particles for deep water waves (a) move in circles, whereas for shallow water waves (b) the path is elliptical.*

$kh \gg 2\pi$. This allows us to simplify Eq. (2.1) to

$$c^2 = \frac{g}{k} = \frac{g\lambda}{2\pi} \tag{2.2}$$

as the hyperbolic tangent ($\tanh kh \Rightarrow 1$) reaches unity for $kh \gg 1$ (see Figure 2.3 for illustration).

Figure 2.3. *The hyperbolic tangent reaches unity for large values of kh, while for small values of kh the line $y = kh$ tangents trough origo.*



Eq. (2.2) shows that the phase speed of deep water waves (short waves) is increasing with increasing wavelength. Hence, long waves will propagate faster, which explains why these waves are the first to reach the coast after propagating over long distances of ocean from the area of generation.

2.2.2 Shallow water

Nearly all the characteristics of waves propagating into shallower water change. Only the period remains unchanged. From Table 2.1 it can be seen that the wave speed decreases with decreasing dept, which owing to the relation $c = \lambda/T$ implies that the wavelength must decrease. In transitional water, $\frac{\lambda}{20} < h < \frac{\lambda}{2}$, the particles start to 'feel' the bottom; the orbits gradually become flatter (Figure 2.2(b)) and adopt elliptical shapes. The semi major axis decreases linearly with increasing depth, while at the bottom a to and fro movement is observed.
In shallow water, which is defined as $h < \frac{\lambda}{20}$, the particle motion is still elliptical. The minor axis decreases linearly with depth opposed to the major axis which is constant. In this case the dispersion relation, Eq. (2.1), can now be simplified as we have $\lambda \gg h \Rightarrow \frac{2\pi}{k} \gg h \Rightarrow kh \ll 2\pi$ for shallow water (long waves). As illustrated in Figure 2.3 the hyperbolic tangent ($\tanh kh \Rightarrow kh$) approaches kh for $kh \ll 1$ reducing Eq. (2.1) to

$$c^2 = gh, \tag{2.3}$$

indicating that the phase speed is determined by depth only.
For transitional water depth ($\frac{\lambda}{25} < h < \frac{\lambda}{2}$) the Eq. (2.1) in its complete form should be used.

2.3 Open ocean waves

The idealized sinusoidal wave described previously does not suffice when describing open ocean waves. An observed state of wind waves will most likely be a combination of locally generated waves and waves propagating beyond the direct influence of the wind (swell). Figure 2.4(a) illustrates how two simple wave profiles (upper panel) with slightly different wavelengths merge to be a more realistic representation (lower panel) of the observed waves. This principle, where waves differing in height, length and direction are added, is called superposition. With a complex pattern of the sea surface (many wave components), new parameters are needed in order to adequately describe the waves. Some of these parameters are listed below and are illustrated in Figure 2.4(b).

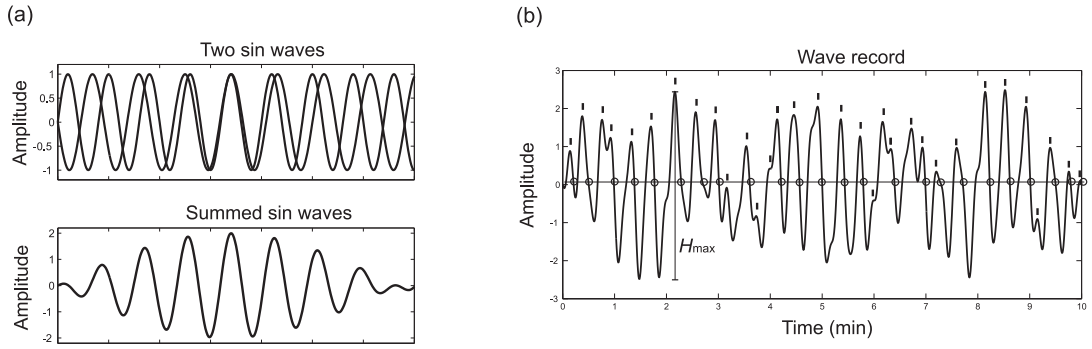


Figure 2.4. *The upper panel in (a) illustrates two waves with slightly different wavelength, while the bottom panel illustrates the sum of these two waves. In (b) an example of a realistic sea state is seen marking the positions of the down crossings (circles), peaks (I) and the highest wave (H_{max}).*

- $\overline{H}$: Average wave height, $\sum_1^n \frac{H_i}{n}$.
- $\overline{T}_z$: The average zero crossing wave period. $\overline{T}_z$ is obtained by dividing the length of the wave record by the number of down or up crossings.
- H_{max}: The maximum wave height recorded during the measurements.
- H_s: The significant wave height. The average height of the $\frac{1}{3}$ highest waves from the wave record.

Using Fourier analysis the sea state can be decomposed, giving the individual wave's amplitude, direction, frequency and phase. The surface elevation varying with time, $\eta(t)$, can then be expressed:

$$\eta(t) = \eta_0 + \sum_{j=1}^{n} a_j \sin\left(j\omega_0 t + \phi_j\right), \tag{2.4}$$

where

- η_o is the zero level as illustrated in Figure 2.1,
- j is the number of wave component,
- a_j is the amplitude of the jth component,
- ω_0 is the angular frequency corresponding to the longest wave in the record,
- ϕ_j is the phase angle of the jth component,
- n is the total amount of wave components.

In practice, this method is used when modeling waves and deriving new wave parameters, a subject covered in section 2.6.4

2.4 Generation of waves

There are three factors limiting the growth of the wind waves:

- Strength of the wind.
- Duration of the wind.
- The area (fetch) over which the wind is blowing.

A diagram illustrating the relationship between these factors and wave height and length is shown in Figure 2.5.

In order to describe the physical process under which the waves are generated let us consider the onset of a wind over a calm ocean. The initial stage of wave generation is the creation of small pressure fluctuations of the sea surface caused by the turbulent nature of the air flowing over the surface (Pond & Pickard, 1983). Through resonance, namely *Phillips' resonance*, these wavelets grow linearly. When the wavelets reach a size which is sufficient to affect the air flow over the surface, the second stage of wave growth, which is the most prominent, commences. This stage is characterized by shear flow instability, where the wave growth increases as the wind is pushing on the troughs and sucking on the crest, and is explained by the *Miles theory*. The rate of this growth depends on the existing sea state and is therefore exponential (Bouws *et al.*, 1998). The magnitude of the final state of the waves now depends on the three previously mentioned factors. For further details see Komen *et al.* (1996). Exemplifying, we see from Figure 2.5 that a 20m/s wind blowing over a 600km area for 24 hours generate

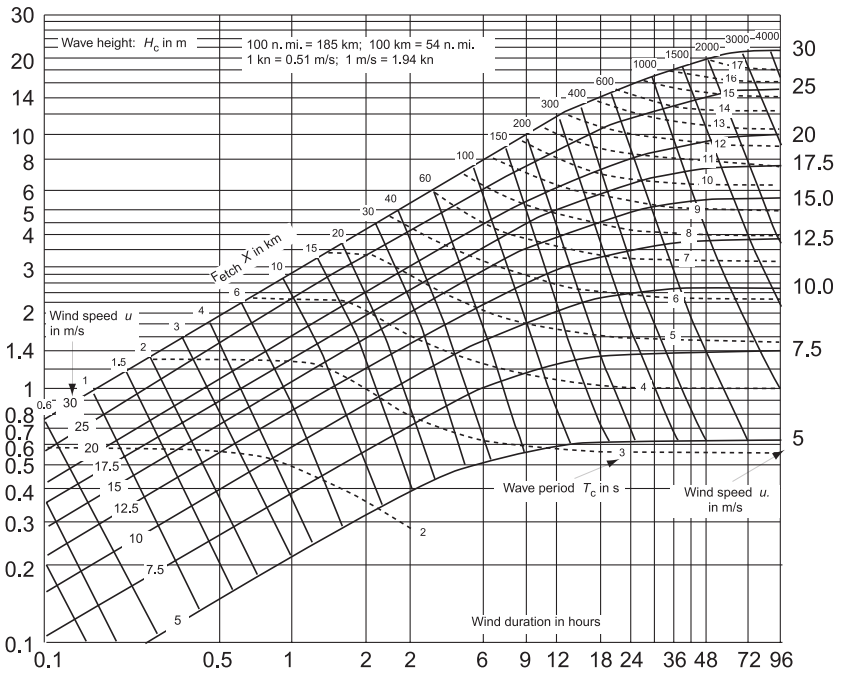


Figure 2.5. *A manual wave forecasting diagram by Groen & Dorrstein (1976), translated by Bouws* et al. *(1998), illustrating the dependence of the tree growth limiting factors and the properties of the waves. Note how the wave heights reach equilibrium and therefore are bounded at the upper end.*

waves with significant wave height of 8 m and a period of 12 seconds.

The waves propagating away from the generating area or the waves which persist after the wind fades are called swell. These waves of long wavelengths (typically 300 to 600 m (Wright *et al.*, 1999)) will diminish in height as they gradually move away from the area of generation. The reason for this is the 'fan' shaped formation through which they propagate, where the energy is spread over a constantly widening front. In fact, the angular spreading is considered to be the main energy loss as far as swell is concerned (Bouws *et al.*, 1998). Referring back to the dispersion relation, Eq. (2.1), we have mentioned that the phase speed increases with increasing wavelength, and for this reason, away from the generation area, waves with close wavelengths will tend to group together in so called wave groups. It is within these groups that the wave energy is carried. Per definition, the group velocity is defined as

$$c_g = \frac{\partial \omega}{\partial k} \tag{2.5}$$

where $\omega = ck$. Again generalizing into deep and shallow water, by use of Eq. (2.2) and (2.3), respectively, the group velocity becomes

$$c_g^{deep} = \frac{c}{2} \tag{2.6}$$

$$c_g^{shallow} = c, \tag{2.7}$$

where c is the phase velocity.

A peculiar feature regarding the energy in waves is the equal partition between kinetic and potential energy (Bouws *et al.*, 1998). For calculating the total amount of energy in the water column per wavelength and crest length the formula

$$E = \frac{1}{8}\rho_w g H^2 = \frac{1}{2}\rho_w g a^2, \tag{2.8}$$

can be used. This (keeping in mind that the energy travels with the group velocity) implies that when a wave approaches the coast, as both group velocity and wavelength decreases, the wave height must increase. This is true for a system where the energy is conserved, i.e. no friction.

2.4.1 Wave spectra

As Eq. (2.8) shows, the energy is proportional to the square of the amplitude, a. After decomposing the sea state by Fourier analysis into individual components E can now be expressed as

$$E \approx \frac{1}{2}\sum_{i=1} a_i^2, \tag{2.9}$$

where $\rho_w g$ is omitted. A distribution where the square of the amplitude is plotted against its frequency is called a *wave-variance spectrum*, $S(f)$, and can be interpreted as the distribution of wave energy over frequency, also referred to as a *wave-energy spectrum*, $E(f)$, when the $\rho_w g$ term is included. Theoretically the wave spectrum is continuous, but for computational reasons it is computed at discrete frequencies. For this reason, the variance or energy value at a given frequency, say 0.20 Hz, is in fact the mean value typically ranging from 0.195 to 0.205 Hz. Therefore, the value, which is divided by the width of the frequency interval is more correctly called an *energy-density spectrum*, with units of $\frac{m^2}{Hz}$. Different spectral shapes are seen for evolving sea state stages. A swell is represented in the spectrum with a narrow peak over its small range of frequencies (periods). Multiple peaks may occur, owing to separately generated swell, or different peaks may be merged together as a wide hump. As models predominantly use energy-density spectra to describe wave fields, deriving the wave height from the spectra is essential and is treated in section 2.6.4.
In the next section, the nature of waves in shallow water will be addressed more closely.

2.5 Waves approaching the coast

Referring to Eq. (2.1) and (2.3), the phase speed of waves in shallow (transitional) water is dominated (affected) by the water depth. In general, moving into shallower water, the phase speed decreases and as the wave period remains constant the wavelengths will decrease, a phenomenon known as shoaling. Waves which are not entering perpendicularly to the coast will be affected by the phenomenon called refraction; a turning of the wave direction. The part of the wave which first encounters shallow water will slow down, causing the wave crest to turn towards shallower water as illustrated in Figure 2.6. In case of a local current, refraction may occur as a response to current-phase speed interaction, irrespective of the direction of approach. Proceeding towards the coast, the final situation will be wave crests breaking parallel to the beach regardless of its deep water direction. In addition to the turning of wave direction, the wave heights increase while the wavelengths decreases toward the coast. Depending on the topography (submarine ridges or canyons) a concentration or rarefaction of wave energy

will occur. An example of refraction is the focusing of wave energy on headlands, whereas the energy in bays is diminished due to the divergence of wave rays. Reflection is another phenomenon, occurring when waves hit the coast in areas of great depth close to land.

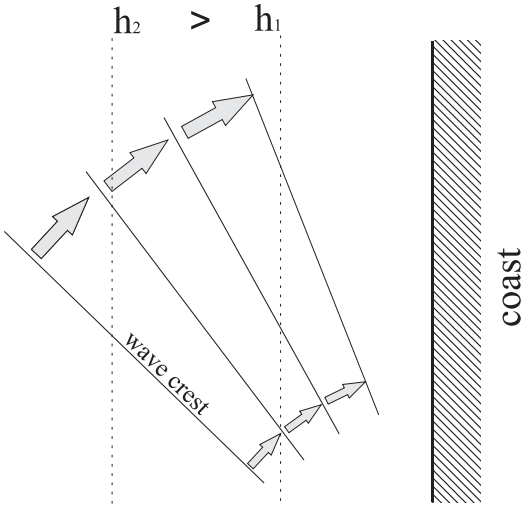


Figure 2.6. *Wave crests (solid lines) approaching the coast change direction as parts of the crests are being slowed down by the depth influence (dashed lines).*

Shoaling, as mentioned, is the bottom influence on waves when they are not subjected to a change of direction. The waves enter the coast perpendicularly to the bottom contours, and results in an increase of wave height, a decrease of wavelength and phase speed and an unchanging period. Table 2.1 exemplify the shoaling effect on a 1.0 m high wave with an 8 seconds period and a 100 m wavelength. When the water depth is the same order as the wave height, the waves start to break, or more specifically when $h = 1.28H$ (Bouws *et al.*, 1998). The upper part of the wave spills over the forward face of the wave as the speed in the lower part of the wave is smaller relative to the upper. Depending of the slope of the bottom, the waves will break differently. In heavily sloping surf zones the water from the crest can plunge forward in free fall.

Table 2.1. *Example of the shoaling effect on wave parameters based on an example by Pond & Pickard (1983).*

Depth, h [m]	>50	10	5	2
Phase speed, c [$\frac{m}{s}$]	12.5	8.9	6.6	4.3
Wavelength, λ [m]	100	71	53	35
Wave height, h [m]	1.0	1.2	1.4	1.7

2.6 Wave modeling

Numerical ocean wave models use surface wind as energy input to calculate the wave conditions. Over the considered time scale the surface winds are the only energy source, and its accuracy is therefore of high importance for the model's output accuracy. With an offset in the wind the model will as a consequence, over time, become inaccurate. The model time step and spatial scale is carefully adjusted to the natural scale of the gravity waves, making sure the distance moved by a wave during one time step is less than a grid length (numerical stability)(Bouws *et al.*, 1998). Over these scales the wave conditions are averaged and assumed temporally stationary and spatially homogeneous. The physical processes controlling the waves are modified into statistical expressions by use of the wave spectra. The time and space evolution of the waves are expressed by the spectral energy balance/transport equation which reads (Bouws *et al.*, 1998)

$$\frac{\partial E}{\partial t} + \nabla \cdot \left(c_g E\right) = S = S_{in} + S_{nl} + S_{ds}. \tag{2.10}$$

The left hand side accounts for the evolution and advection of waves, respectively, where c_g denotes the group velocity in deep water. $E\left(f,\theta,x,t\right)$ is the wave spectrum in two dimensions depending on frequency, f, and direction of propagation, θ. Combined, the left hand side represents the total temporal derivative of E. The S is termed the source function and accounts for energy input by the wind (S_{in}), non linear transfer of energy by wave to wave interaction (S_{nl}) and dissipation of energy (S_{ds}). On the form expressed here Eq. (2.10) is only valid in deep water and does not account for refraction nor currents (Bouws *et al.*, 1998). Following is a short insight in each of the source terms.

2.6.1 Wind input, S_{in}

Any wave model depends heavily upon the quality of the wind input, both regarding resolution and accuracy. As mentioned in section 2.4 there is an interaction between the waves and the atmospheric boundary layer. The most common way of expressing the rate at which the wind energy is transferred into the waves, accounting for the feedback mechanism, is given by:

$$S_{in} = A\left(f,\theta\right) + B\left(f,\theta\right) E\left(f,\theta\right), \tag{2.11}$$

where the first term on the right hand side, $A\left(f,\theta\right)$, represents the linear growth of wavelet through *Phillips' resonance*. The second term is the feedback term expressing the stage at which the predominant wave growth takes place.

2.6.2 Non linear interaction, S_{nl}

This term describes the weakly non linear interaction between waves of different frequencies through resonance, but also the propagation of surface waves. This is purely a redistribution of energy within the spectrum and does not affect the overall energy in the wave field, however the spectral shape might be altered. Despite not being classified as energy input, the non linear interaction is due to the redistribution of energy the main growth factor for waves below the spectral peak. This is illustrated in Figure 2.7. What happen is, as seen in the upper panel, that in the region close to the spectral peak the wind input is greater than the

dissipation. This abundance of energy is transferred by the non linear interaction to either side of the peak. The energy shifted towards higher frequencies is lost through dissipation, whereas the energy transferred towards the lower end of the spectra induces growth of new wave components; a phenomenon called *downshift*.
It must also be mentioned that it is the handling of S_{nl} that distinguishes second from third generation wave models. The third generation is, although by greater computational cost and special integration techniques, capable of explicitly calculating this term, whereas the second generation model was forced to parameterize.

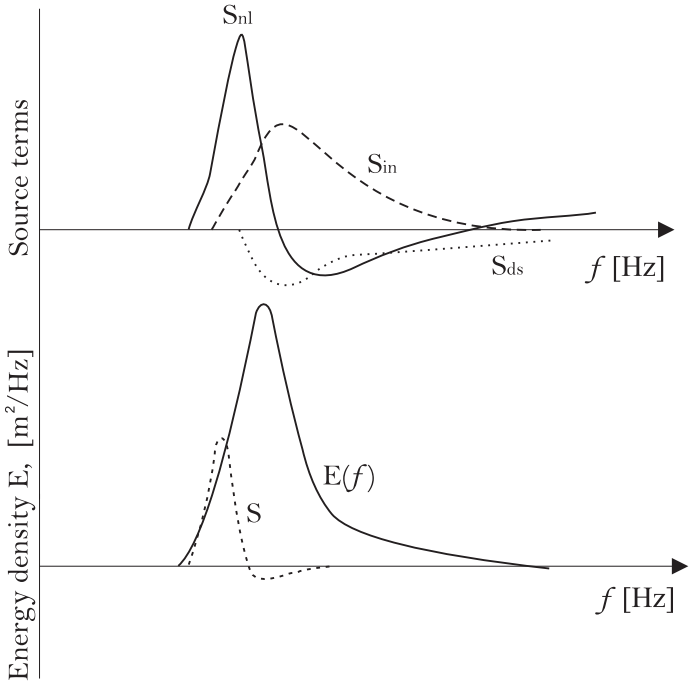


Figure 2.7. *The upper panel illustrate the three source terms S_{in}, S_{nl} and S_{ds} and their rate of growth influence with respect to frequencies. In the bottom panel the sum of the source terms, S, is seen within the frequency spectra, $E(f)$. Based on a plot by Bouws* et al. *(1998).*

2.6.3 Dissipation, S_{ds}

The dissipation term is the sum of the energy lost by the three processes white capping, wave-bottom interaction and surf breaking. White capping is the dominating energy loss in deep water, however the effects on swell is negligible (Bouws *et al.*, 1998). White capping happens when waves become too steep in the growth phase, and is an important growth limiting factor which transfers energy to underlying currents. Energy loss by wave-bottom interaction involves different mechanisms like bottom friction, movement of the bottom material and water percolation into the sea bed. In water with depth of same order as the wave height, surf-breaking becomes the dominating energy loss (Bouws *et al.*, 1998).

2.6.4 Wave height derived from the spectrum

The form of the spectrum is described by use of moments of the distribution. To calculate the spectrum's *nth* order moment, m_n, the following definition is

used:

$$m_n = \int_0^\infty f^n E(f)\, df, \tag{2.12}$$

where $E(f)$ is the discrete variance density value at frequency, f, hence $E(f)\, df$ denote the variance $a_i^2/2$ within the ith interval of f and $f + df$. Integrated, taking into account for that $E(f)$ is discrete, Eq. (2.12) becomes:

$$m_n = \sum_{i=0}^{N} f_i^n \frac{a_i^2}{2}. \tag{2.13}$$

According to the definition of m_n in Eq. (2.12), the zero-order moment, m_0, which represents the area under the spectral curve becomes:

$$m_0 = \sum_{i=0}^{N} \frac{a_i^2}{2} = \frac{a^2}{2}. \tag{2.14}$$

As we soon will see m_0 can be used to deduce wave height parameters. Now, considering a single sinusoidal wave with the same energy as the actual sea state, its corresponding wave height, H_{rms}, is obtained by rearranging Eq. (2.8) from section 2.4:

$$H_{rms} = \sqrt{\frac{8E}{\rho_w g}}. \tag{2.15}$$

This *root-mean-square wave height* has empirically been found to relate to the significant wave height by a multiplication of $\sqrt{2}$. The wave parameter H_{m0} which we now obtain is the model deduced parameter corresponding to H_s and can in turn be expressed in terms of m_o by:

$$H_{m0} = \sqrt{2}\sqrt{\frac{8E}{\rho_w g}} = 4\sqrt{m_0}, \tag{2.16}$$

which theoretically is valid only for swell. However, the bias between H_s and H_{m0} is small and is empirically found on average to be:

$$H_{m0} = 1.05 H_s. \tag{2.17}$$

As mentioned in section 2.3 the significant wave height is an average over the $1/3$ highest waves in a typically six hour wave record. The maximum wave height, H_{max}, occurring in the same wave record is related to H_s and H_{m0} by:

$$H_{max} \cong 2.0 H_s \cong 1.9 H_{m0}. \tag{2.18}$$

Chapter 3

METHOD

3.1 WAM cycle 4

A coarse and nested model was implemented and run from 00:00 01.01.2001 to 24:00 31.12.2001, hereafter referred to as the 2001 model run, covering the areas shown in Figure 3.1. The model used was the WAve Model (WAM) cycle 4, as modified by the Norwegian Meteorological Institute. This third generation wave model, developed by the Wave Modeling Group, is characterized by the lack of assumptions concerning the spectral shape, the explicit solving of the transport equation (see section 2.6) and improvements on the finite-depth version of the model (The WAMDI group; Hasselmann *et al.*, 1988). The latest improvement of the third generation WAM model is the cycle 4 version (Hasselmann *et al.*, 1992), where new wind physics is included and current refraction (not enabled in this model run) is supplemented (Komen *et al.*, 1996).

The high degree of freedom in WAM allows the user to freely choose a regional or global domain with arbitrary resolution in time, space, direction and frequency. The settings used in the 2001 model run are displayed in Table 3.1. As for the frequency resolution, both the nested and coarse model consists of 25 predefined frequency bands which are with high resolution appointed to the low frequencies. This resolution distribution increases logarithmically from 0.042 Hz to 0.41 Hz with an increment-to-frequency ratio equal to 0.1. The choices regarding the grid spacing and time steps were based on computational efforts along with the demand of satisfying the CFL criteria, i.e. avoiding numerical instability.

Table 3.1. *Resolution settings for the coarse and nested WAM 2001 model run.*

	Spatial resolution	**time step**	**# of frequencies**	**# of directions**	**# of sea points**
Coarse	$0.5^\circ \approx 55.5$ km	15 min	25	24 (15°)	11 251
Nested	$0.1^\circ \approx 11.1$ km	4 min	25	24 (15°)	936

The coarse model was run first; providing the nested model with boundary input. Hence, waves generated outside the nested domain were accounted for. Thereafter, the nested model produced ocean state parameters, as listed in Table 3.2, with high temporal and spatial resolution. The inputs on which both the coarse and nested models were provided with are addressed in the following section.

Table 3.2. *Output parameters given by the WAM cycle 4 model.*

Output parameters
Wind stress fields
Mean wave direction
Mean wave frequency
Significant wave height
Swell wave height and direction
2D wave spectrum at chosen grid points

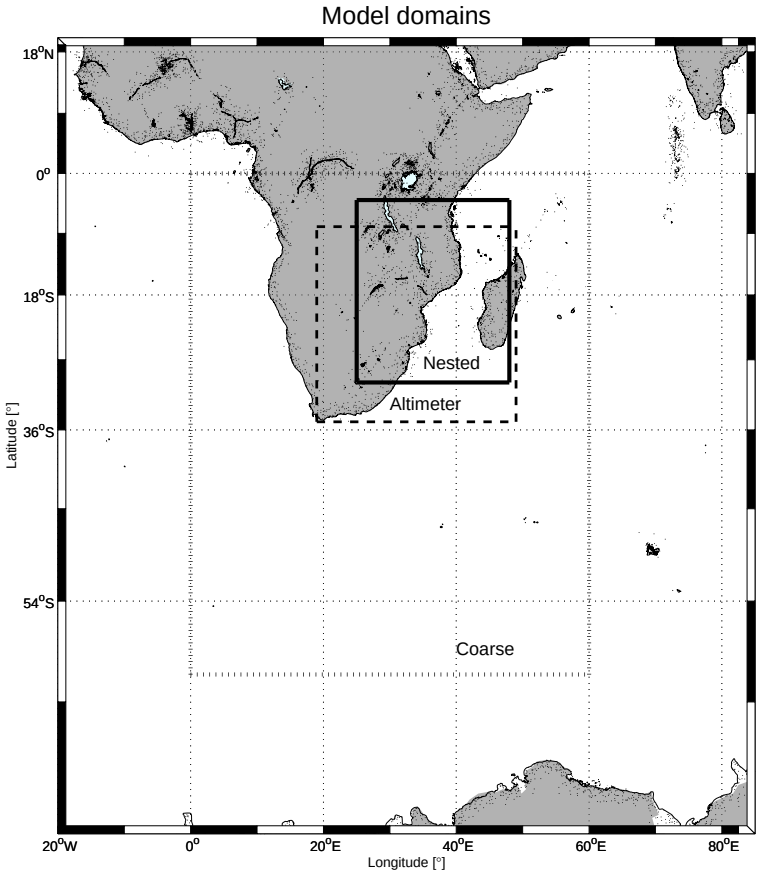


Figure 3.1. *Illustration of the model domains. The coarse and nested models are represented with dotted and solid lines, respectively. The area in which the altimeter data was retrieved is within the dashed lines.*

3.1.1 Input

Wind

With the extensive earth-system model called the Integrated Forecast System (IFS) the European Centre for Medium-Range Weather Forecasts (ECMWF) produces a range of global atmospheric and oceanic parameters (ECMWF, 2006). Covering both the coarse and fine meshed model domains, synoptic (six hours interval) winds with a horizontal resolution of 39 km ($\approx 0.35^\circ$ of latitude and longitude) was extracted. As required by the WAM model, the wind input file

was extrapolated and interpolated to exactly match the grid of the coarse and nested model, respectively. Given the location of the region of interest this is the best resolution available. However, the grid spacing is not sufficiently small to catch the full small scale synopsis of the strong pressure gradients of tropical cyclones (Hart & Evans, 2001). As a result underestimation of the intensity of the tropical cyclones occur (Uppala *et al.*, 2004). For a full description of the assimilation method and changes undergone by IFS during 2001 see ECMWF (2006).

Bathymetry

A global data set of topography, with a resolution of 2 minutes of longitude and latitude, was taken from the Earth Topography Two Minutes Grid (ETOPO2) (NGDC & NOAA, revised 2009). This data set includes the ocean bathymetry and was in 2001 assembled from various global and regional data sets by NGDC & NOAA (revised 2009). The superior resolution of the topography data set has no restrains of the models accuracy. In fact, a smoothing of the bathymetry data had to be done in order to customize the bathymetry to each model domain and their respective spatial resolution.

3.2 Altimeter

Without available in situ measurements from buoys, altimeter data is the only accessible means to validate the model output. Presently, the altimeter SHWs accuracy is suitable to validate wave models, despite the fact that time series are not readily constructed. However, during satellites' life time trends and differences in the SWH data are seen caused by electronic drift, calibration and processing changes. Therefore, linear corrections proposed by Queffeulou (2004), listed in Table 3.3, are applied on the original altimeter SWHs to compensate. Additional care should be taken when using altimeter data on locations close to land as degraded quality of the data occurs when the satellite moves from land to sea, compromising the data for the first few measurements (Krogstad & Barstow, 1999). With an altimeter spatial resolution of typically 7 km this implies that data from the roughly 20 first kilometers from land should be avoided. Through the *Center for Satellite Exploitation and Research* (CERSAT), a department of the *French Institute of Research for the Exploitation of the Sea* (Ifremer), SWH data covering the region within the dashed lines in Figure 3.1 was obtained. The measurements were done by the *ERS-2*, *TOPEX* and *GEOSAT Follow-On* satellites, owned by the *European Space Agency* (ESA), *U.S Space Agency* (NASA) and *French Space Agency* (CNES) and the *U.S Navy*, respectively. This combined satellite altimeter data set consisted of 455 571 recordings from 2001. For more details regarding the principle of altimeter measurements see Krogstad & Barstow (1999).

Table 3.3. *Corrections to altimeter SWH measurements proposed by Queffeulou (2004), utilized by Ifremer (Queffeulou & Croize, 2009), on the form* $SWH_{cor} = \boldsymbol{a} \times SHW + \boldsymbol{b}$.

Satellite	a	b
ERS-2	1.0642	0.0006
TOPEX	1.0237	-0.0476
GEOSAT FO	1.0625	0.0754

This altimeter SWH data was collocated with SWHs from the 2001 nested WAM run; chronologically, starting in January 00:00, the altimeter data which deviated in time and space with less than one hour and 30 minutes and 0.07° of longitude and latitude ($\approx$ 7.8 km) was paired with the model SWH. In this case 0.07° corresponds to half the diagonal within the nested grid box. With this vast collocated data set, linear regression and correlation coefficients (as defined in Eq. (A.1)) can reveal the functional relationship and linear dependence of the model with respect to the assumed true altimeter data. After applying the functional relationship on the nested model output, the H_s data can now be considered validated (hereafter addressed as H_s^{val}). In order to increase the data set from which the return values are to be calculated H_s^{val} will be assessed against a 45 year reanalysis of global SWHs, addressed in the following section. The concept is to extract points of interest from this reanalysis and compared with the same validated point from the nested WAM model, a linear functional relationship will be applied and extrapolated over the remaining 44 years of data as illustrated in Figure 3.2.

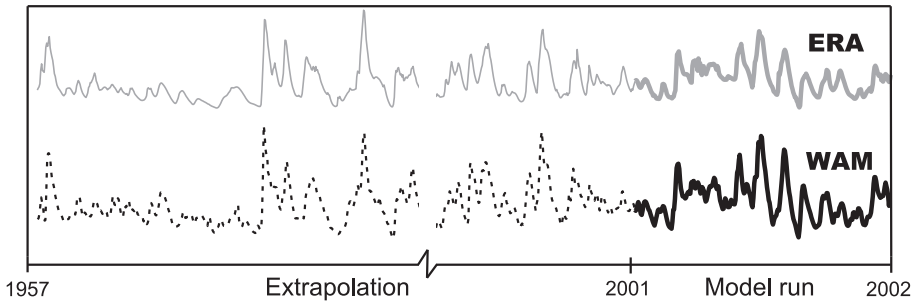


Figure 3.2. *Illustration of the extrapolation principle. A functional relationship between the validated WAM data and the ERA-40 data is found for 2001 (bold lines). Thereafter the relationship is applied on the ERA-40 data to extrapolate the WAM data (dashed line) backwards in time. This is done individually for the five points of interest.*

3.3 The *ERA-40* reanalysis

ECMWF has, in collaboration with many institutions, accumulated global meteorological observations from September 1957 to August 2002. This vast amount of meteorological observations includes everything from traditional radiosonding and synoptic surface observations from ship and land to an increasing number of ocean buoys, aircraft and satellite observations. The globally assembled data was used as assimilation in the 45 year ECMWF Re-Analysis (*ERA-40*) using the IFS. The reanalysis provides global meteorological wind, temperature and humidity fields, stratospheric ozone and deep water SWH, making it the longest and most comprehensive wave data set available (Uppala *et al.*, 2005). The *ERA-40* was, in fact, the first reanalysis carried out in which an ocean wave model had been coupled to an atmospheric model (Uppala *et al.*, 2004). This allows for ocean atmosphere interaction to be accounted for, and should be considered important as the ocean waves to a large extent control the surface wind flow, and in turn might make the model more accurate (Janssen *et al.*, 1989). The wave model used in the IFS system was ECMWF's own configured WAM

cycle 4 model, where both the current and finite depth options were disabled (Caires *et al.*, 2004). It uses spherical coordinates and provides a global 1.5 ° times 1.5 ° latitude and longitude resolution with synoptic time scales. With this ocean wave data available on a global basis, the studies previously limited only to the northern hemisphere due to higher data coverage, climatology and extreme wave values can now be studied everywhere in equal accuracy. But, as mentioned in section 3.1.1, the spatial resolution makes the model underestimate the intensity of tropical cyclones. Nevertheless, a detection rate close to 100% is reported by Uppala *et al.* (2004, 2005) in the southern hemisphere. This value accounts for the period after 1980 only, and decreases to around 75% in the begining of the reanalysis. A validation assessment by Caires & Sterl (2003a) revealed that *ERA-40* underestimate high H_s values by more than 20%, which in this work may result in low estimates of calculated return values, if not accounted for.

3.4 Extreme value theory

In extreme value theory statistical techniques and models describe unusual rather than usual events. Widely used today are two approaches, namely the Generalized Extreme Value (GEV) and the Generalized Pareto (GP) distribution. Mutual within these two techniques is the extrapolation of a statistical distribution representing measurements from a recorded period to an unobserved period in which the extreme values may occur (Coles, 2001). Following is a brief introduction to both techniques.

3.4.1 Generalized Extreme Value distribution

The origin of the extreme value theory, where the maximum samples were expressed as asymptotic functions can be traced back to Fisher & Tippett (1928). Three different limit distributions were distinguished by Gnedenko (1943) which later was unified by Jenkinson (1955). This unification of the distributions allows for a continuous range of possible shapes to be modeled within one distribution function expressed as the Generalized Extreme Value (GEV) distribution (Coles, 2001). Independence and identical distribution (iid) is a prerequisite for data to be modeled by the GEV distribution. Wave height data recorded continuously with short intervals over several years will not fulfill these requirements as the measurements are exposed to strong seasonal variabilities. As a means to remove this dependence, the data is often divided into bulks of measurements recorded over periods of time sufficiently long to isolate the seasons. Nevertheless, the method is still applicable when data with weak dependence is used (Soares & Scotto, 2001). The values, z_n, modeled by the GEV distribution is the maxima events during these bulks. For this reason this method is often referred to as the *Bulk method*, or in the case where annual maxima are used as the *Annual Maxima* (AM) method, with a distribution given by

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi} \right\}, \tag{3.1}$$

defined for values of z: $1 + \xi\left(\frac{z-\mu}{\sigma}\right) > 0$,
where:

ξ : shape parameter $-\infty < \xi < \infty$,
μ : location parameter satisfying $-\infty < \mu < \infty$,
σ : scale parameter $\sigma > 0$.

The value of ξ regulates the shape of the distribution which is fitted to the data:

· For $\xi = 0$ a GEV *type I* or *Gumbel* distribution is obtained.
· For $\xi > 0$ a GEV *type II* or *Fréchet* distribution is obtained.
· For $\xi < 0$ a GEV *type III* or *Weibull* distribution is obtained.

Of the many techniques proposed for parameter estimation, the method of *Maximum Likelihood* (ML) is applied in this work. This method is implemented in both the statistical *R*-based (R, 2009) package *extRemes* (Gilleland *et al.*, 2009) and in *Matlab*'s *Statistics Toolbox*. Despite that numerical methods are needed (Dong & Takayama, 2002), this is the most frequently used method and is recommended among others by Coles (2001). The principle of the ML method is to decide which family of distributions assigns the highest probability to the data (Coles, 2001). The ML estimators, σ, ξ and μ, are defined as the parameter values which maximize the log-likelihood function, ℓ, given as

$$\begin{aligned}\ell(\mu,\sigma,\xi) = &- m\log\sigma - (1+1/\xi)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] \\ &- \sum_{i=1}^{m}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-1/\xi},\end{aligned} \tag{3.2}$$

where $i = 1, ..., m$ denotes the data index.
In the case of $\xi = 0$, the Gumbel limit of the GEV distribution, the log-likelihood becomes:

$$\ell(\mu,\sigma) = -m\log\sigma - \sum_{i=1}^{m}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\left(\exp\left\{-(\frac{z_i-\mu}{\sigma}\right\}\right). \tag{3.3}$$

The maximized likelihood estimate for the united GEV families is obtained by derivation of Eq. (3.2) or (3.3) with respect to the parameter vector (μ, σ, ξ) and equating to zero, i.e. maximizing with respect to each of the parameters.

With a model fitted to the data, return levels, z_N, defined as the level expected to be exceeded on average once every N years, can now be calculated. Combining the probability for non-exceedance, $P(z < z_N) = 1 - 1/N$, and Eq. (3.1), the system $G(z_N) = 1 - 1/N$ is solved for z_N by inversion, yielding (Palutikof *et al.*, 1999):

$$z_N = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-1/N)\}^{-\xi}\right] & \text{if } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-1/N)\} & \text{if } \xi = 0. \end{cases} \tag{3.4}$$

When using multiple observations during a year, i.e. the bulk method, this must be accounted for by replacing the $1/N$ term with $1/\left(N \times n_y\right)$, where n_y is the number of measurements during a year. By this, we see that the return level depends on the number of observations used in the distribution.

An important property regarding the three different distribution functions is illustrated when plotting Eq. (3.4) against $y_p = -log(1-1/N)$ on a logarith-

mic scale. This is illustrated by a *return level plot* in Figure 3.3. In this example the scale and location parameters are held constant for values of 1.0 and 2.0, respectively, while different shape values invoke the characteristic shapes of each GEV family. The Gumbel distribution is here seen as a straight line, in contrast to the Fréchet and Weibull distributions which have concave and convex shapes, respectively. It appears from this that the Weibull distribution is the only family which is bounded in the upper end. For the other two families an unphysical behavior is seen in the upper end where they are unbounded, indicating return levels of infinite size. This characteristic behavior prevails irrespective of the values of the scale and location parameters, which control the slope and intercept on the y-axis, respectively. Although exaggerated, Figure 3.3 indicates that, despite being bounded in the upper end, return values obtained by the Weibull family are always smaller than what the other families would estimate Harris (2004, 2005).

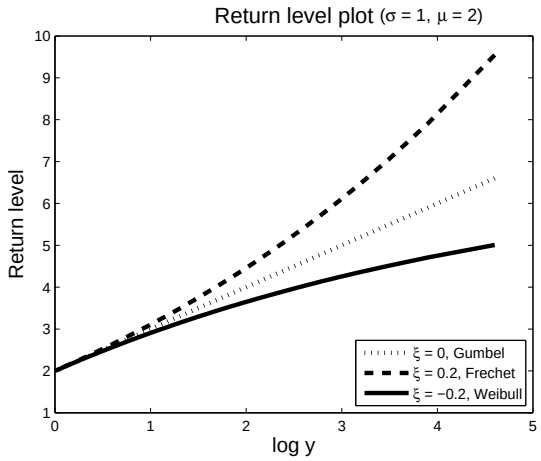


Figure 3.3. *The concave, linear and convex shapes of the Fréchet, Gumbel and Weibull families, respectively, are well illustrated when plotted on a logarithmic scale. In this example the scale and location parameters of a GEV distribution are set to 1.0 and 2.0, respectively.*

A drawback for the GEV model is the critical data exploitation, where only the single highest observation during a year is used while the rest of the data is discarded (Pickands, 1975; Soares & Scotto, 2001; Caires & Sterl, 2005). This might cause questionable return levels estimates as the highest values below the AM might exceed an AM recorded any other year. High uncertainty is an additional consequence, predominately seen for return levels corresponding to long return periods, as the extrapolation is based on an inadequate data set. Motivated by this weakness a new asymptotic distribution for extreme value analysis, called the Generalized Pareto (GP) distribution, was developed. This distribution combines the GEV approach, and is addressed in the following section.

Finally, it should be mentioned that AM data is used in this work, in which a 'year' is defined from October 1st to September 30th. By dividing the year at the beginning of the austral summer the full austral winter is accounted for in the cycle of a 'year', hence avoiding that valuable data is left out by the occurrence of two extreme seasons within the same bulk.

3.4.2 Generalized Pareto distribution

When data of higher resolution than AM is available, e.g. monthly maxima or six-hourly in this case, modeling using block maxima is a waste in the approach to finding an accurate extreme value (Coles, 2001). The method to be described now avoids the blocking procedure and characterizes the observation as an extreme if it exceeds a high threshold, hence making the most out of the available data. Within the community of extreme value statistics the Peaks over Threshold (POT) method seems highly favored compared to the AM method (Soares & Scotto, 2001; Dong & Takayama, 2002; Alves & Young, 2003; Caires & Sterl, 2005; Caires *et al.*, 2006b; Neelamani *et al.*, 2007; Naess & Gaidai, 2009).

Let $X_i = X_1, X_2, ..., X_n$ be a time series of recorded SWH measurements. The observations which exceed a fixed threshold, u, are termed exceedances, $z_1, ..., z_k = X_i - u$, where $X_i > u$, and occur according to a Poisson process[1]. This classification of extremes is illustrated in Figure 3.4. Groups of consecutive exceedances

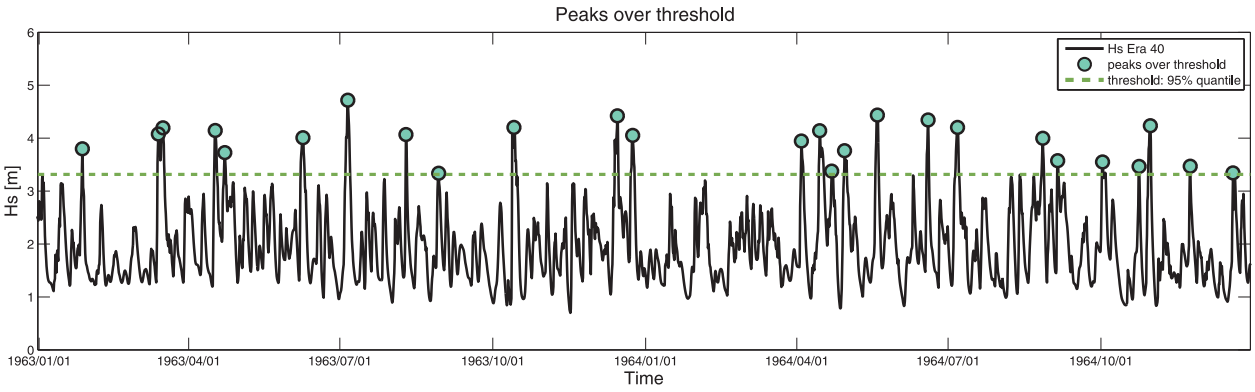


Figure 3.4. *Example showing the peaks over threshold extraction for two random years of the ERA-40 data. The exceedance values range from the individual peaks marked by a circle to the dashed threshold line, which in this case represent the* 95% *quantile.*

are called clusters and are generated by the same storm. This dependence must be removed to fulfill the requirement of independence, and is done by choosing only the peak exceedance within a cluster. In the case where the storm intensity decreases for thereafter to intensify, multiple clusters may be generated by the same storm, an example of which is seen in March 1963 in Figure 3.4. Again, also this dependence is removed, as proposed by Alves & Young (2003), Caires & Sterl (2005) and Caires *et al.* (2006b), by separating the clusters with a minimum of 48 hours. The distribution of the exceedances can now be approximated by the GP distribution (Coles, 2001):

$$G(z) = \begin{cases} 1 - \left(1 + \frac{\xi z}{\tilde{\sigma}}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{z}{\tilde{\sigma}}\right) & \text{if } \xi = 0, \end{cases} \tag{3.5}$$

defined for $z > 0$ and $\left(1 + \frac{\xi z}{\tilde{\sigma}}\right) > 0$, where u is the threshold value and

$$\tilde{\sigma} = \sigma + \xi (u - \mu). \tag{3.6}$$

[1]Any process with continuous events occurring independently. A positive parameter, n_y, exists which represents the average rate of occurrence during a year (Hogg & Tanis, 2006).

Again, this allows for a continuous range of distributions combining three families with different tail behaviors. The parameters ξ and μ are the same as defined for the GEV distribution in Eq. (3.1). Coles (2001) states that if the AM from a set of observation can be fitted to a GEV distribution, observations above a threshold can be fitted to a GP distribution. Similarly as for the GEV distribution, ξ is dominant in determining the behavior of the GP distribution. The same conditions regarding the values of ξ apply and assigns the same distinct shapes as illustrated in Figure 3.3. This implies that for the GP distribution only the Weibull family is fully appropriate to model a naturally bounded phenomenon as waves. Once a threshold is selected, a topic addressed below, the model parameters can be estimated. Again, this is done by the ML method, in the same manner as described above, but now the log-likelihood function, ℓ, in the case of $\xi \neq 0$, is given by

$$\ell\left(\tilde{\sigma}, \xi\right) = -k \log \tilde{\sigma} - \left(1 + 1/\xi\right) \sum_{i=1}^{k} \log\left(1 + \frac{\xi z_i}{\tilde{\sigma}}\right), \tag{3.7}$$

where $i = 1, ..., k$ is the index of the exceedances. For $\xi = 0$, ℓ is given as

$$\ell\left(\tilde{\sigma}\right) = -k \log \tilde{\sigma} - 1/\tilde{\sigma} \sum_{i=1}^{k} z_i. \tag{3.8}$$

With a model assigned to the data, the return level, z_N, is derived solving $G\left(z_N\right) = 1 - 1/\left(N n_y \lambda_u\right)$ for z_N. Here $G\left(z_N\right)$ stems from Eq. (3.5), and the probability of non-exceedance is $P\left(z < z_n\right) = 1 - 1/\left(N n_y \lambda_u\right)$. This yields (Coles, 2001)

$$z_N = \begin{cases} u + \frac{\tilde{\sigma}}{\xi}\left[\left(N \lambda_u n_y\right)^{\xi} - 1\right] & \text{if } \xi \neq 0 \\ u + \tilde{\sigma} \log\left(N \lambda_u n_y\right) & \text{if } \xi = 0, \end{cases} \tag{3.9}$$

where λ_u represents probability of an observation exceeding the threshold, u, and $n_y = k/45$ is the average number of exceedances per year.

Threshold selection

The challenge when fitting a GP distribution to a set of data, is the selection of an appropriate threshold. A threshold too high will render little data on which to base the estimation of the parameters on, hence resulting in large variance of the parameter estimates. On the contrary, a threshold selected to low will give biased parameter estimates but with small variance. Furthermore, if the threshold is lowered too much, events belonging to the central part of the distribution are included. This may violate the independent distribution condition of the GP distributions as the observations are clustered too close, but also include observations not classified as extremes. The idea is to find a threshold high enough for the underlying theoretical foundation to be valid, but low enough so that there are sufficient data with which to make an accurate fit. Despite the threshold selection is vital to the outcome of the return values (Dong & Takayama, 2002), no standardized procedure exists to select an appropriate threshold. Alves & Young (2003) recommend tuning the threshold until the number of extracted maxima corresponds to the average number of storms per year. Caires & Sterl (2005) achieved good results fixing the threshold at the 93% quantile of the full time series when using the *ERA-40* data. And from Breivik *et al.* (2009, and references

therein) the threshold is recommended to be chosen such that the number of exceedances are approximately 10 per year, or $2-3$ times the mean value of the significant wave height. However, as the threshold selection is expected to vary for each location, the threshold choice for each position studied in this work is individually assessed by using two visual techniques proposed by Coles (2001).

The first method is based on the mean of the GP distribution, and is illustrated in Figure 3.5. The concept is to find the lowest threshold for which the plot is nearly linear, and ideally choose the lowest threshold at which the GP distribution is valid (Caires & Sterl, 2005). In this plot the sum of the excesses over the threshold u is divided by the amount of data points which exceed the threshold u and is plotted against u. The curve describes the expected overshoot of a threshold once an exceedance occurs. If the curve has a negative gradient it indicates that the observations follow a GP distribution with a negative ξ, and vice versa for positively sloped curves (Gencay *et al.*, 2002). In the example illustrated in Figure 3.5 the mean exceedance line is seemingly linear with a negative gradient for thresholds between 2.4 and 3.1 m, an indication that the data should be modeled with the Weibull distribution.

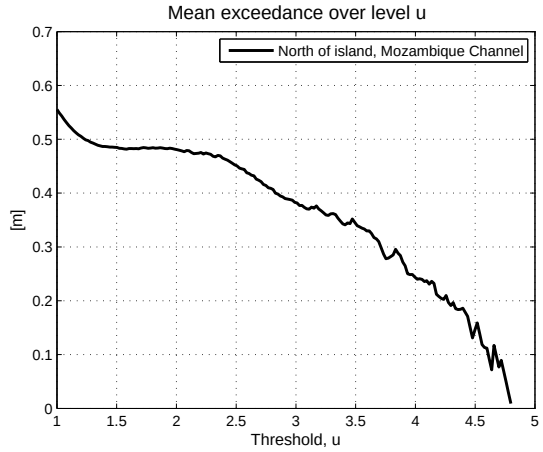


Figure 3.5. *Estimated expected exceedance over the threshold, u, as a function of u for flag 4, North of island. Linearity with a negative slope is seen for values of u ranging from* 2.4 *to* 3.1 *m, implying the Weibull distribution as a likely model to fit the data.*

In the second method the GP distribution is fitted to a range of thresholds, looking for stability in the parameter estimations. The procedure is based on that if a GP distribution serves as a reasonable fit to the exceedances of a threshold u_0, then the exceedance of a higher threshold, $u > u_0$, should also fit a GP distribution. For both these two distributions ξ is identical, whereas the scale parameter estimated when using $u > u_0$ is σ_u, and is by use of Eq. (3.6) given as:

$$\sigma_u = \sigma_{u_o} + \xi\,(u - u_0)\,. \tag{3.10}$$

The fact that σ_u changes with u unless $\xi = 0$ is accounted for by reparameterizing Eq. (3.10) to $\sigma^* = \sigma_u - \xi u$, which is constant with respect to u. Accordingly, both estimates of σ^* and ξ should be constant for $u > u_0$. Figure 3.6 illustrates an example of the estimates for the modified scale and shape parameter for a range of 20 different thresholds. Accounting for the sampling variability the parameters seem stable around 2.5 m, appointing this as an appropriate threshold selection.

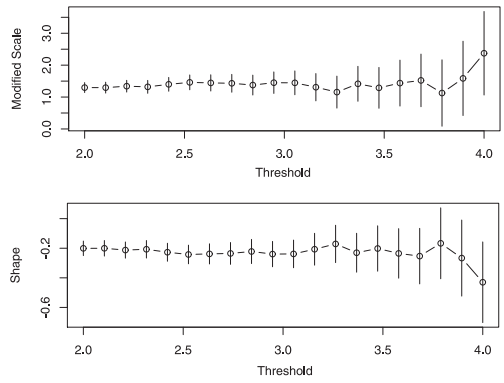


Figure 3.6. *Parameter estimates against 20 thresholds ranging from 2.0 to 4.0 m for significant wave height observation. In this example a threshold selection of 2.5 m seems appropriate.*

3.4.3 Model checking

Although it is impossible to quantify the reliability of an extrapolation based on a GEV or GP model, different graphical goodness-to-fit checks are used in which the validity of the model is assessed with reference to the observed data. *Probability plots* compare the empirical and fitted cumulative distribution functions, where any observed departures from the unit diagonal indicate shortcomings of the model. However, as both the empirical and fitted distribution functions are bound to approach 1 for observations of large values, little information is provided in this important low density region of the model. This weakness is avoided when the same information is expressed at a different scale in the *quantile-quantile (Q-Q) plot*. Here observations from different quantiles are assessed against the corresponding modeled quantile. Again, departures from linearity provide evidence of a failing model. Another graphical technique is the *return level plot* where Eq. (3.4) or (3.9) is plotted against $y_p = -log\,(1-P)$ on a logarithmic scale. P represents here $1/N$ or $1/Nn_y\lambda_u$ for the GEV and GP distribution, respectively. The advantages of this technique are the simple interpretation and the compressing of the tail by the choice of a logarithmic scale which highlights the effect of extrapolation for long return levels. Additionally, the linearity of the plot in the case of $\xi = 0$ makes the distinction of the different model families easily distinguished. To further increase the informativeness of the plot, confidence intervals (CI) can be added. This allows the return level plot to be used as a goodness-to-fit test where observations outside the CI band suggest an inadequate model. The last graphical check is different from the others in the way that it is based on the density function rather than the model-based and empirical distribution function. Here a modeled probability density function is compared with a scaled histogram of the observed data. However, these density plots are less informative regarding model checking as the shape of histograms vary markedly with the choice of bin intervals. Therefore, there is no unique parameter assigned to different shapes of density functions of the respective families.

3.4.4 Confidence intervals

When extrapolating return values the uncertainty increases for values corresponding to long return periods. In this work a CI of 95% will be applied on the

data. However, one should be aware of that this CI is based on the assumption that the model is correct, and therefore should be regarded as the lower bounds of uncertainty. If the uncertainty regarding the model's validity were taken into account, the CI would have been considerably higher (Coles, 2001). Two different approaches, namely the *Delta* and the *Profile likelihood* method are used to estimate the CI of return values in this work. For a thorough explanation, see Coles (2001).

The *Delta method* is applied when using both *MatLab* and the *R*-based *extRemes* package. It assumes that the ML parameter estimates are distributed with a standard normal distribution, and will therefore symmetrically flank the best estimate. The approach involves finding the variance, $Var\left(z_N\right)$, of the return value, z_N, for thereafter to obtain the approximate 95% CI by $z_N \pm 1.96 \times SE$, where SE is the standard error ($\sqrt{Var\left(z_N\right)} = SE\left(z_N\right)$). The variance of the return value is defined as (Coles, 2001)

$$Var\left(z_N\right) \approx \nabla z_N^T V \nabla z_N, \tag{3.11}$$

where V is the variance-covariance matrix of the ML estimated parameters. The transposed matrix, ∇z_N^T, is given by

$$\nabla z_N^T = \left[\frac{\partial z_N}{\partial \mu}, \frac{\partial z_N}{\partial \sigma}, \frac{\partial z_N}{\partial \xi}\right], \tag{3.12}$$

when using a GEV distribution. However, the same formula is applicable for the GP distribution by substituting μ with λ_u in the first column. Differentiating, by use of Eq. (3.4) and (3.9), Eq. (3.12) is rewritten as (Coles, 2001)

$$\nabla z_N^T, \text{ GEV} = \left[1, -\xi^{-}1\left(1 - y_p^{-\xi}\right), \sigma\xi^{-2}\left(1 - y_p^{-\xi}\right) - \sigma\xi^{-1} y_p^{-\xi} log y_p\right], \tag{3.13}$$

and

$$\begin{aligned} \nabla z_N^T, \text{ GP} = \Big[&\sigma m^{\xi} \lambda_u^{\xi-1}, \xi^{-1}\left\{(m\lambda_u)^{\xi} - 1\right\}, \\ &-\sigma\xi^{-2}\left\{(m\lambda_u)^{\xi} - 1\right\} + \sigma\xi^{-1}(m\lambda_u)^{\xi} log\left(m\lambda_u\right)\Big], \end{aligned} \tag{3.14}$$

for the GEV and GP distribution, respectively, evaluated at their respective parameter estimates. In Eq. (3.14), m corresponds to the return period multiplied by the number of data points per year, i.e. $m = Nn_y$. Either equation can now be substituted into Eq. (3.11) to yield the return value variance, and in turn provide the CI. This procedure was implemented in *MatLab* routines which are attached in the appendix for illustration.

The *Profile likelihood* method generally generates better approximations of the CI (Coles, 2001). The technique involves a direct reparameterization of z_N by use of Eq. (3.4) or (3.9). The profile log-likelihood is obtained by maximizing the respective model's log-likelihood function, ℓ, with respect to the parameters in which z_N is now represented. The parameterizations reads (Coles, 2001)

$$\mu = \begin{cases} z_N + \frac{\sigma}{\xi}\left[1 - \{-\log\left(1 - 1/N\right)\}^{-\xi}\right] & \text{for } \xi \neq 0 \\ z_N + \sigma \log\left[-\log\left(1 - 1/N\right)\right] & \text{for } \xi = 0 \end{cases} \tag{3.15}$$

in the case of a GEV distribution and

$$\tilde{\sigma} = \begin{cases} \frac{(z_N - u)\xi}{\left(Nn_y\lambda_u\right)^{\xi} - 1} & \text{for } \xi \neq 0 \\ \frac{z_N - u}{log\left(Nn_y\lambda_u\right)} & \text{for } \xi = 0 \end{cases} \tag{3.16}$$

for the GP distribution.
The CI is now found in the profile plot where a horizontal line representing a critical log-likelihood value intercepts with the return level profile. This critical value is defined as the border above which the model parameters are in close agreement with the observed data. An example of a profile log-likelihood is seen in Figure 3.7, where data from the *Oil location* is modeled with a GP distribution. Note that the CI will in this case not be symmetrical, which is owing to the assumed right skewed chi-square distribution of the ML parameters (Meeker & Escobar, 1995). However, this assumption is appropriate regarding waves, as the density of information decreases for extreme observations, hence reflected by the more precise CI.

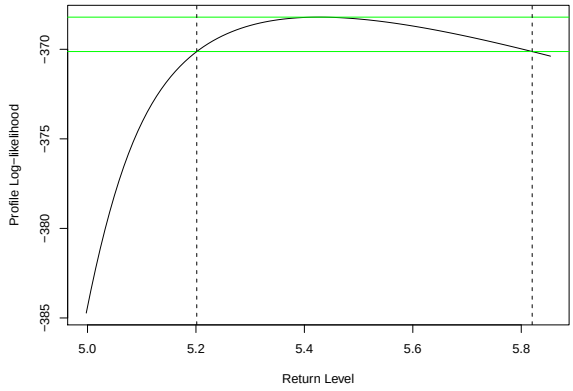


Figure 3.7. *Profile likelihood for 100-year return level from the Oil location. The* 95% *CI of the* H_s^{100} *is found between the intercepts of the log-likelihood and the critical negative log-likelihood value of* -370*, flanked by dotted vertical lines. The best estimate of the* H_s^{100} *is located at the maxima of the log-likelihood.*

Chapter 4

RESULTS

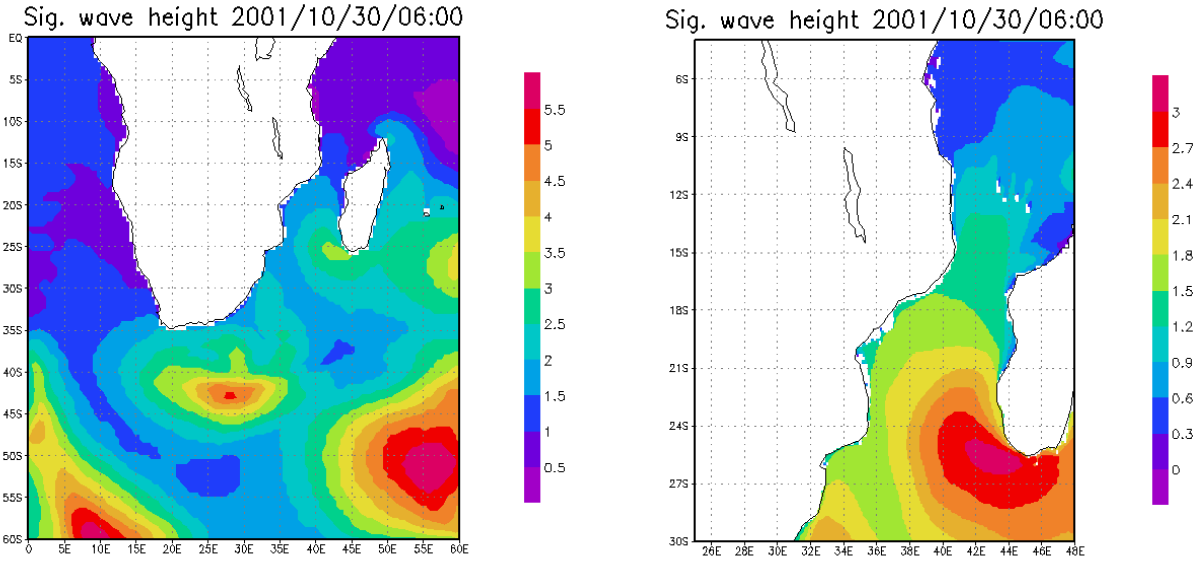
4.1 The model run

First off in this chapter the coarse and nested 2001 WAM model runs will be addressed. The enhancement of spatial resolution in the nested model will be illustrated, followed by single point model comparisons where significant wave height data is extracted from one exposed and one shallow water position.

4.1.1 Coarse versus nested run

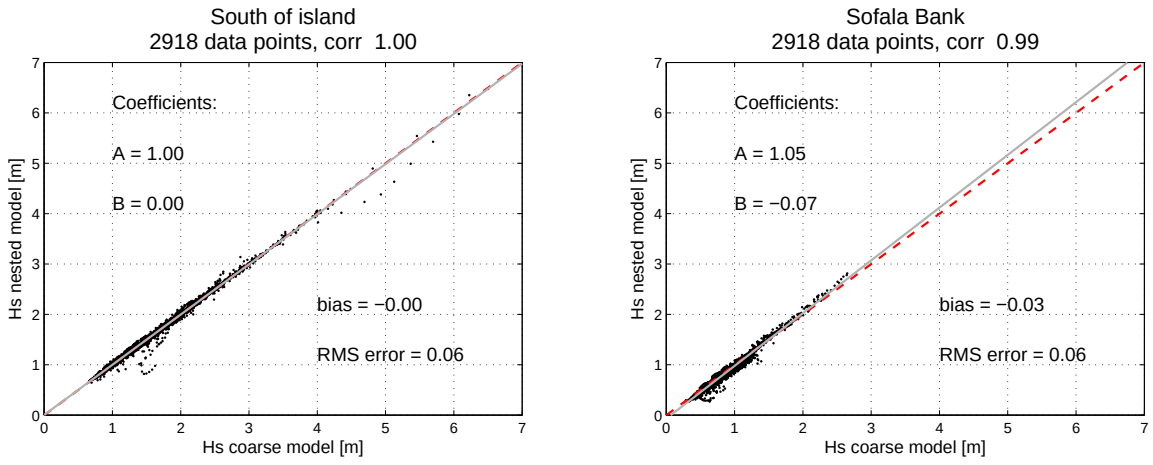
Figure 4.1, which illustrates the significant wave height conditions for both model domains at 06:00 the 30th of October 2001, shows matching large scale features in the Mozambique Channel. However, as the resolution of the nested model resolves higher geographical details, like coastal features and islands, some enhancements can be seen in Figure 4.1(b). Illustrated here is a shading effect from the islands located in the northern opening of the channel, where the northward propagating waves which does not make land fall on the island are seen as fingers extending to the north. For this reason the nested model is expected to yield output more realistic in areas close to land, islands or in regions with shallow water depths.

In order to test the coarse and nested model correlation, scatter plots were made with SWH data extracted from the location *South of island*, Flag 2 in Figure 1.1, an exposed position in deep water, and from the shallow *Sofala Bank* location, Flag 3. The linear regression and correlation coefficients along with the root-mean-square error, *RMS*, as defined in Eq. (A.2), and bias, as defined in Eq. (A.3) in the Appendix, are seen in Figure 4.2(a) and (b). At the *South of island* location a correlation coefficient of 1.0 implies that an increase or decrease in SWH occurs simultaneously in both the nested and the coarse model, and the functional relationship reveal a one-to-one relationship. Little deviation is seen between corresponding data points, illustrated by the insignificant *RMS* error of 0.06 m and the zero bias value. Regardless of the persuasive correlation values, overestimation performed by the coarse model during some single events is seen. As for the shallow *Sofala Bank* the 2918 data points are again perfectly correlated and densely distributed, with a positive correlation of 0.99 and a *RMS* error of 0.06 m. However, at the *Sofala Bank* the linear regression coefficients reveal a slight overestimation of SWHs smaller than 1.5 m by the coarse model and an underestimation of the exceeding SWHs, given by the ratio and intercept values 1.05 and -0.07, respectively.



(a) Coarse model (b) Nested model

Figure 4.1. *Significant wave height map illustrating the difference in spatial resolution of the coarse and nested models. Both color bar units are given in meters, however, with different scales.*



(a) (b)

Figure 4.2. *Point comparison between coarse and nested model. The coarse to nested ratio (A) and vertical displacement value (B) indicate an excellent coincidence for the exposed and open water position South of island, whereas a small offset is experienced over the shallow Sofala Bank. The dashed red line represents ideal values of ratio $A = 1$ and zero offset, $B = 0$, while the gray solid line has slope and intersection values as given by A and B on the plots.*

In the next section an immediate application of the model data is seen when the seasonal pattern is found by calculating monthly means of SWH and primary wave direction for both the coarse and nested model domain.

4.1.2 Monthly means

The seasonal variations are illustrated in Figure 4.3 and 4.4, where the monthly means of the primary wave direction and significant wave height for both the coarse and nested models are presented. The small scale pattern from the nested model plots are in accordance with the main features of the corresponding plots based on the coarse model. Throughout the year waves generated by the consistent westerly winds in the Southern Ocean propagate towards either side of the South African coast. However, the extent of which they penetrate into the Mozambique Channel vary, with a minimum wave intrusion observed in January, where only waves of modest height are observed further north than the *Sofala Bank*. At this time of the year the easterly winds dominate south of the channel, producing waves with an average significant wave height of 1.6 to 1.8 m which travel towards the region of Maputo, the capital. North of the channel waves generated by the northeasterly monsoon travel southwards and meet with the northward moving waves in the middle of the channel, as seen in Figure 4.4(a). For the rest of the year the waves north of the channel travel northwestward towards the border between Mozambique and Tanzania.

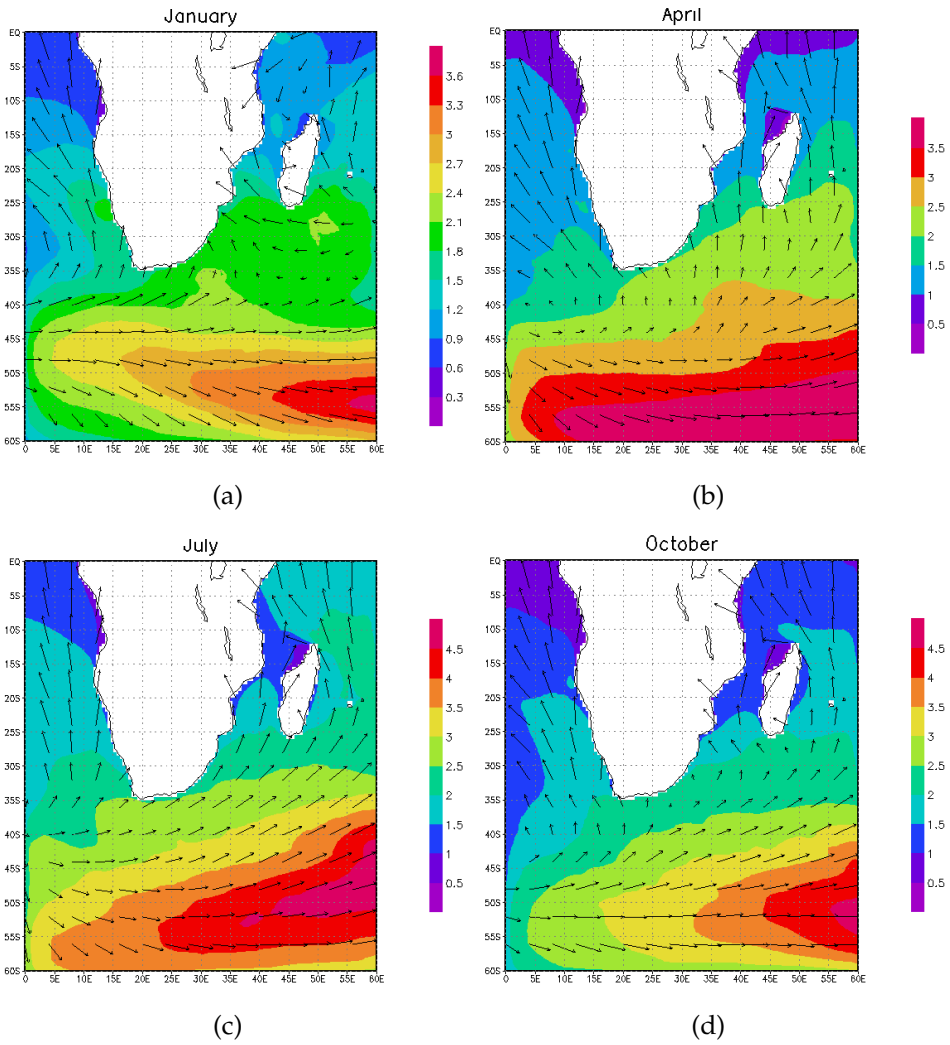


Figure 4.3. *Monthly means of the primary wave direction and the significant wave height for January, April, July and October from the coarse model run of 2001. The primary wave direction is averaged and displayed for every 12th grid point. The units of the color bars is meter and are subjected to a change of scale between some panels.*

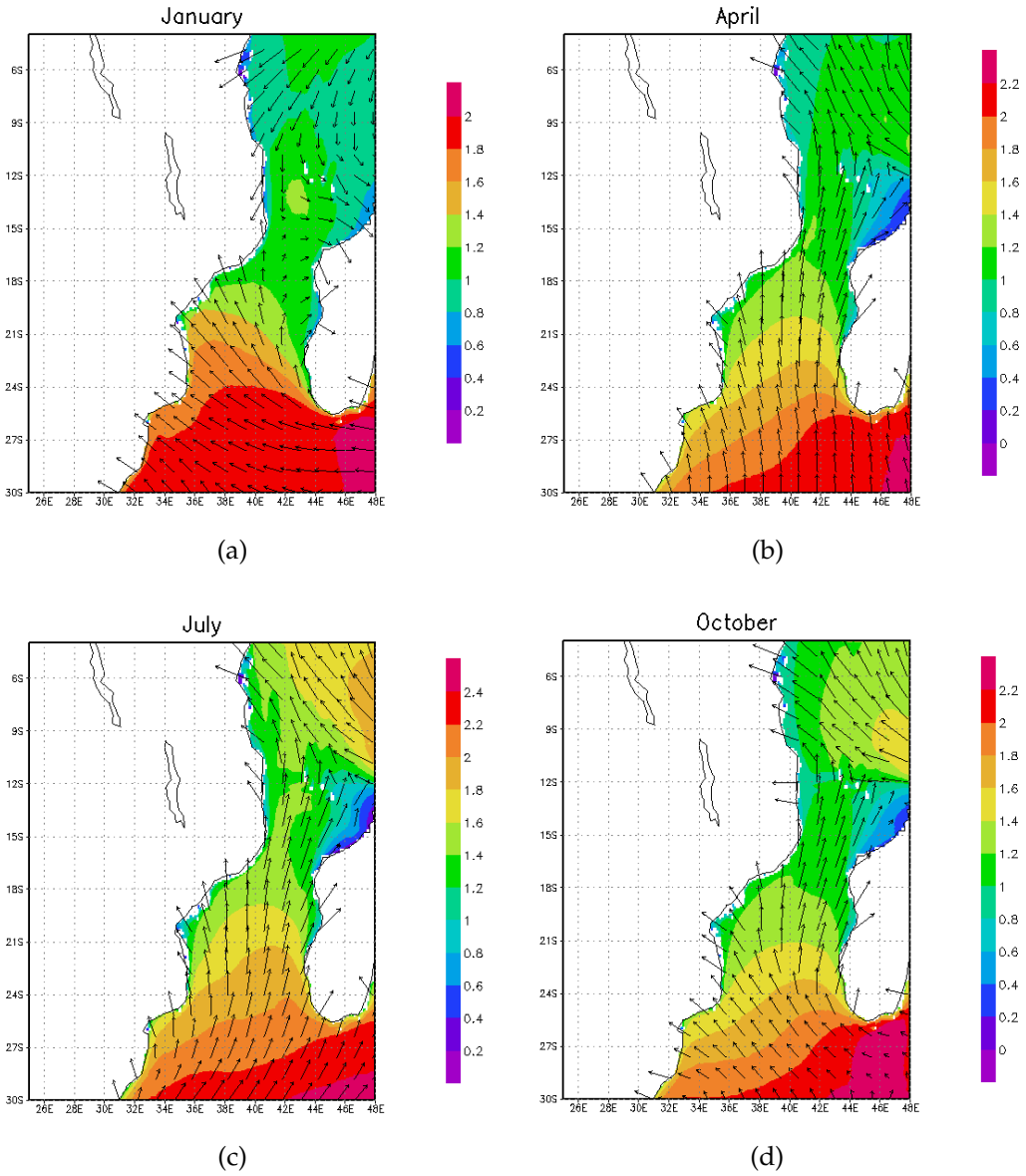


Figure 4.4. *Directional and significant wave height monthly mean for January, April, July and October from the nested model run of 2001. The primary wave direction is averaged and displayed for every 12th grid point. The scales of the color bars vary and have units given in meters.*

4.1.3 Discussion

Despite the plots presented in Figure 4.3 and 4.4 are based on 2001 data alone, the maps correlate well both in wave size and spatial distribution when compared with monthly means plotted in the ERA-40 based *Global Wave Climatology Atlas* by Caires *et al.* (2006a). This is assuring regarding the accuracy of the model run, as the *Global Wave Climatology Atlas* is based on corrected data, whereas the model data from this work is not yet validated.
As for the coarse and nested model run an appropriate correlation is observed, both regarding spacial pattern and wave size. However, the small deviations presented by the statistical parameters in Figure 4.2(b) might indicate that within the shallow regions of highly varying bathymetry, the nested model will most likely predict values more accurately than the coarse model. Then again, as the same wind input is applied for both model runs, only small departures are expected.

4.2 Validation

In this section validation of the nested model will be assessed by use of altimeter data. Initially, the total collocated WAM-altimeter data set will be used, and a functional relationship for the full area over which the nested domain and the altimeter regions overlaps will be estimated. Further, as to assess the local spatial variability of the model's accuracy, linear dependencies will be estimated using collocated SWH extracted from a 2° by 2° latitude/longitude domain with each of the five locations of interest as a center.
Once a linear relationship which validates the nested WAM model is established and applied, functional relationships between the five locations from the corrected WAM model and five *ERA-40* data locations will be estimated for the purpose of extrapolating the WAM model backward in time. This will be addressed in a successive section, where scatter diagrams for each location will be presented. As a last step in the validation process the correspondence between the *ERA-40* data and the satellite data will be assessed using a new data set in which untampered *ERA-40* is collocated with altimeter SWH.

4.2.1 WAM versus altimeter

Taking the absolute value of the difference from all collocated data pairs, and averaging the points which fall within a 0.5° longitude/latitude box gives the *mean absolute deviation* plot seen in Figure 4.5. In the north and towards the southern opening of the Mozambique Channel altimeter and modeled SWHs correlate well, where small deviations ranging up to 0.3 and 0.4 m are seen. Further south a sporadic increase of deviations occur, reaching values of 0.7 m. Yet, it is along the coast the highest values of discrepancy are observed.

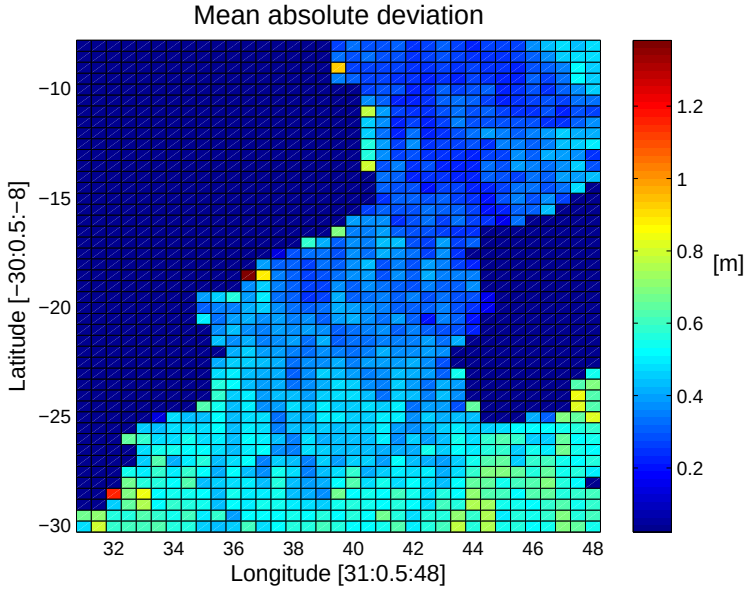


Figure 4.5. *Mean absolute deviation from the collocated WAM and altimeter data with a spatial resolution of* 0.5° *latitude/longitude.*

Figure 4.5 relates to the geographical credibility of the nested model yet it does not convey any information regarding over which wave dimensions the model's

credibility is highest. This is illustrated in Figure 4.6, where the collocated SWHs are plotted with color codes indicating the probability of occurrences. The first striking feature is the model's underestimation which increases in magnitude for higher waves. As for the smaller values, < 1.0 m, the belt of the most frequently occurring wave heights is nearly underlying the red dashed line which indicates the one-to-one ratio. It is also evident from the probability density plot that the SWHs from 2001 rarely exceed 3 m, seen by the dark blue color indicating a probability of occurrence close to zero. However, as it is this upper tail of waves which are most prone to underestimation by the model, and also the waves of interest in this work, it is obvious that in order to ensure accurate return values, the underestimation observed must be accounted for.

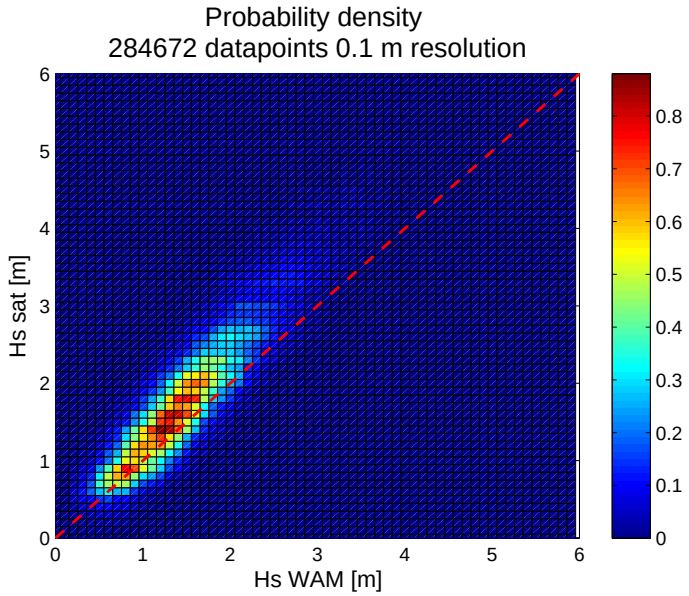


Figure 4.6. *Density probability plot of the collocated nested model and altimeter data. The wave height resolution is 0.1 m and the probability of any occurrence is indicated by the color bar. The red dashed line indicates the perfect ratio between model and altimeter.*

The functional relationship which relates the modeled SWHs to the measured altimeter SWHs is given by the regression coefficients A and B in the scatter plot seen in Figure 4.7. This data is based on the full collocated SWH data set, and yields a very strong linear association between the variables given by the positive correlation of 0.89. The ratio, $A = 1.19$, and vertical displacement, $B = 0.03$, implies a general underestimation of 19% by the nested WAM model, whereas the small RMS error of 0.54 m indicates a relatively dense distribution pattern. With this robust linear dependence, based on 284 672 data points, applied on the nested model, the underestimation will be accounted for. SWH values close to the altimeter values from any random grid location may now be extracted from the nested model for comparison with the *ERA-40* time series. This is done in the following section, but first, the spatial variation of the model validation is assessed.

The scatter plots, based on the locally extracted data, in Figure 4.8 all reveal, though varying in value, tendencies of model underestimation, the same as observed in Figure 4.7. However, for the *North of island*, *South of island* and the *Surf location* the correlation and regression coefficients along with the RMS error and the bias are of equal size as the values obtained when applying the full collocated

data set. For the *Oil location* the underestimation amounts solemnly from the vertical displacement, B, an offset of 0.18 m. Furthermore, this is the location where both the lowest correlation coefficient, reading 0.81, along with the smallest *RMS* error and bias are seen. As for the *Sofala Bank*, the ratio and the intercept values of 1.10 and 0.26, respectively, are contributing factors for the model underestimation. General for these five scatter plots is the sparse number of applied data, ranging from 2969 to 5644 data points, makeing the RMS error and the bias sensitive to single events such as the departures seen in the upper end of (b) and (d).

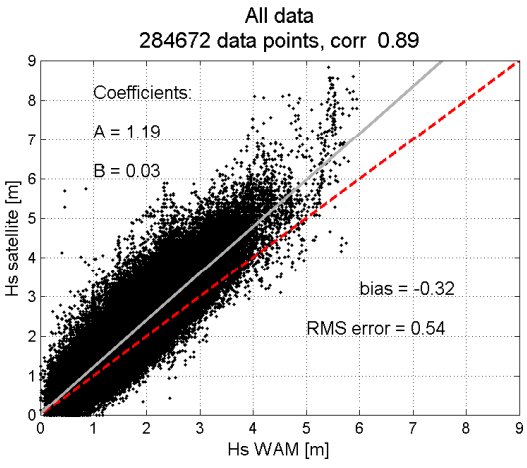


Figure 4.7. *Scatter plot with statistical coefficients. The linear dependence of the model and altimeter is described by the straight (gray) regression line $y = Ax + B$, where A is the slope and B is the intercept on the y axis. The number of data points of which the regression is based on is given in the title along with the correlation coefficient.*

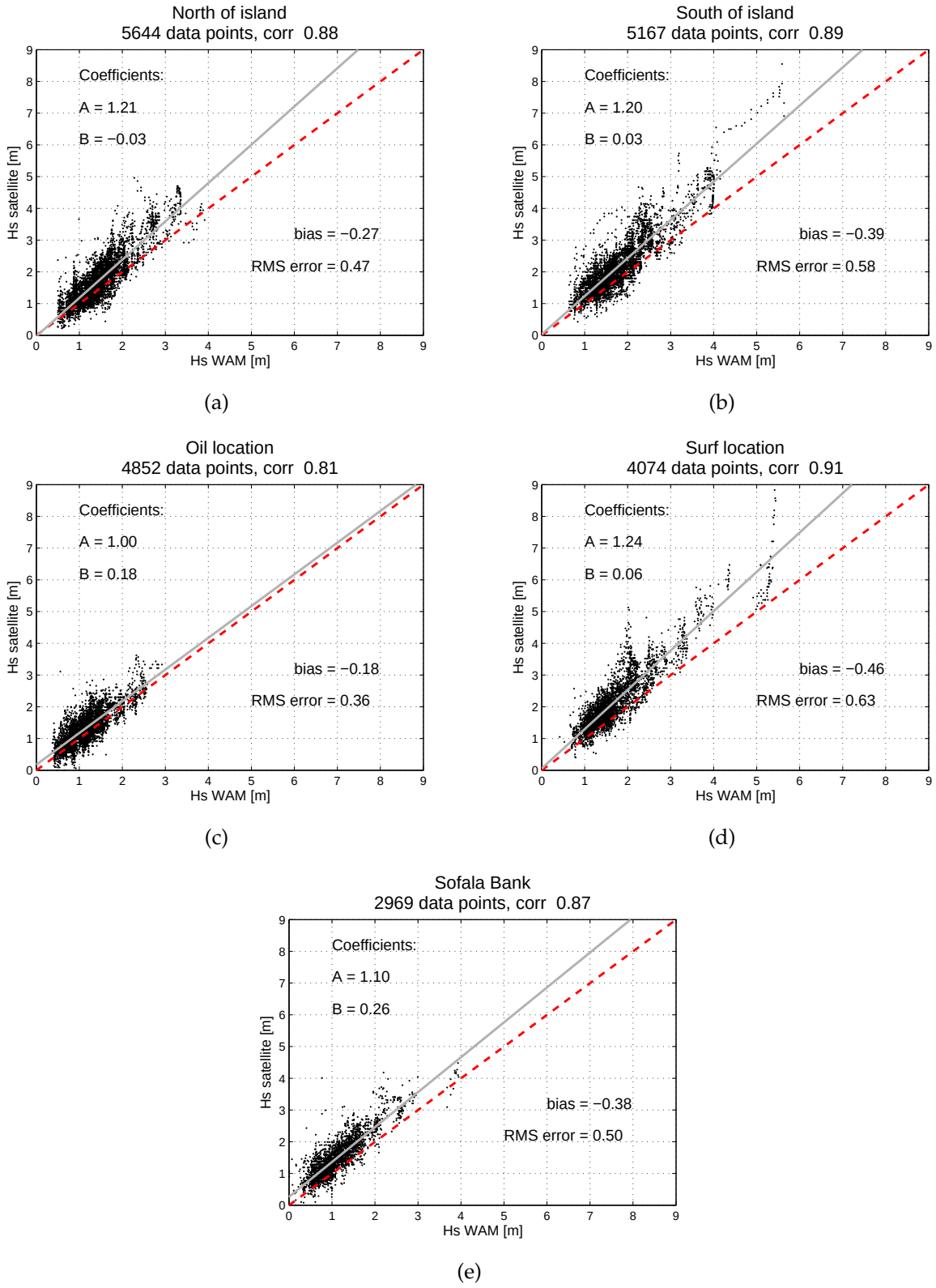


Figure 4.8. *Scatter plots with statistical coefficients. For each location collocated data has been extracted from a 2° latitude/longitude box with its respective position in the center. The linear dependence between the model and altimeter is described by the straight (gray) regression line $y = Ax + B$, where A is the slope and B is the intercept on the y axis. The number of data points on which the regression is based is given in the title along with the correlation coefficient.*

Discussion

The functional relationship obtained when using all collocated data points, yielding $Hs_{SAT} = Hs_{WAM} \times 1.19 + 0.03$, is regarded robust due to the vast amount of data points used in the linear regression. Despite coastal altimeter values are included in the 284 672 collocated data points, the effect on the functional relationship inflicted by the small amount of erroneous data is less important. The functional relationship in Figure 4.7 shows that the modeled SWHs are generally 19% too low compared with the altimeter data. This coincides with results from a global model experiment carried out by Romeiser (1993) where the WAM modeled SWHs validated by *GEOSAT* altimeter data indicated an underestimation of about 20% in the Southern Hemisphere winter. Geographically closer, Bauer *et al.* (1992) found the modeled SWH to be $30 - 40\%$ lower than the altimeter data over the Indian Ocean. However, these studies were carried out using global WAM models with spatial resolutions of $1.5^\circ \times 1.5^\circ$ and $3^\circ \times 3^\circ$ latitude and longitude, respectively. In the latter studie the high underestimation is most likely a resolution effect where the coarse resolution in time and space miss out on the high wind speeds. Also the small *RMS* error of 0.50 m obtained by Janssen *et al.* (1997), again from a global WAM model, matches well with the 0.54 m obtained here. However, regional differences are reported regarding both functional relationships and statistical constants. This is illustrated in Figure 4.8, where the deep water locations *South of island*, *North of island* and *Surf* tend to relate better to the all-inclusive functional relationship.

When the linear regression is carried out separately for each of the five locations, the coefficients are based on 2969 to 5644 data points. Though this is not as robust as the coefficients seen in Figure 4.7, it may in turn be regarded as more exact for its respective location. The sparse number of data points are likely sufficient for single events like the storm observed in the upper end of Figure 4.8(b) to slightly influence the result, therefore using this regression should only be done if the robust regression is considered inaccurate for a specific location. In other words, determining which of the regression coefficients to apply on the model data is a predicament between sturdiness and accuracy. Therefore, as the regression coefficients and correlation obtained from Figure 4.7 are closely related to the values in Figure 4.8(a),(b) and (d), the functional relationship of $Hs_{SAT} = Hs_{WAM} \times 1.19 + 0.03$, obtained from Figure 4.7, will be applied on the locations *South of island*, *North of island* and the *Surf*. With this, the validation is both robust and at the same time the accuracy is preserved.

For the *Oil location* the model and altimeter provide data with good correlation, however the regression coefficients are markedly different from those in Figure 4.7, yielding an one-to-one ratio and a constant underestimation of 0.18 m. This implies a smaller underestimation for higher SWH values compared to the all-inclusive 19% underestimation. However, if the $Hs_{SAT} = Hs_{WAM} \times 1.19 + 0.03$ dependence is applied at this location a higher estimate of the return value is obtained. Table 4.1 exemplifies this effect, where both the all-inclusive and the individually obtained functional dependencies are applied on the 20 highest SWHs occurring at each location in the *ERA-40* time series. This table illustrates that the *Oil location* is the only location which will suffer from an underestimation, and the *Surf location* the opposite, if the locally obtained linear dependence were to be applied. At the remaining locations a neglectable effect is seen whichever functional dependence is applied. A possible reason for this distinctive functional

Table 4.1. *Illustrating the effect of the applied correction versus the individual linear dependence presented in Figure 4.8.*

	Average top 20 *ERA-40* measurements, [m]		
Location	Full: $1.19 + 0.03$	Local, as Figure 4.8	Difference
North of island	4.99	**5.02**	-0.03
South of island	6.14	**6.19**	-0.05
Oil location	**4.55**	3.96	0.59
Surf location	6.38	**6.69**	-0.31
Sofala Bank	**3.40**	3.38	0.02

relationship at the *Oil location* is the absence of the high swells produced by the persisting westerly winds in the south. This may further imply that it indeed is over the higher end of the SWHs where the discrepancies start to occur between model and altimeter. Another, but less likely reason, is based on that the *Oil location* is located in the northern outskirts of the nested and coarse model domain. Therefore it might occur that the boundary input provided by the coarse model may depart from the true wave conditions as the model area in which the south-west moving waves are generated is too small to reproduce all generated waves. This is particularly the case in the summer when the weak northeasterly winds prevail.
Furthermore, the *Oil location* is in the region where tropical cyclones occasionally influence wave conditions. Again, this occurs during summer, and is not possible to fully include in the model calculations until winds with resolution adequate to resolve the full nature of TCs are available. However, as implied by the very low RMS error in this location, this effect is not very big. Nevertheless, as a means of making the validation more robust, and simultaneously avoiding underestimation of the return values, the functional relationship obtained by the all-inclusive scatter plot in Figure 4.7 is the one which will be applied when validating the model data.

At the shallow *Sofala Bank* the locally obtained functional relationship yields $Hs_{SAT} = Hs_{WAM} \times 1.10 + 0.26$. This is different from the other functional relationships as both the ratio and the intercept constants contribute to the underestimation, an effect seen influencing even the small SWHs. At first this dependence was regarded doubtful due to the *Sofala Bank's* vicinity to land and the altimeter's weakness over this area. However, as Figure 4.9 illustrates, the distance separating the location from land is sufficient to provide accurate altimeter data, a statement which holds for all locations. Another concern regards the coarse spatial altimeter resolution of typically 7 km. This footprint, over which the measurements are averaged, does not allow for the potentially large transformation in wave conditions which takes place from deep to shallow water, c.f. Table 2.1, to be fully resolved (Krogstad & Barstow, 1999). With this weakness in mind, but also referring to the small difference the individual functional dependence would play on the highest SWHs, see Table 4.1, the decision of applying the dependence from the full collocated data set was made.

To summarize, the functional relationship which will be applied on the selected nested model locations reads $Hs_{SAT} = Hs_{WAM} \times 1.19 + 0.03$, and is the proposed functional relationship for all data extracted from this model.

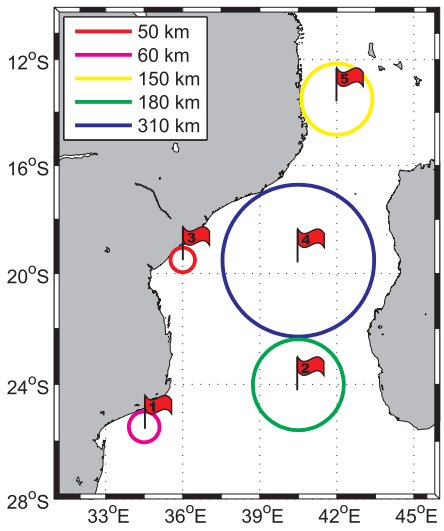


Figure 4.9. *Map showing the minimum radial distance from each position to either land or islands. The Flags represent locations as presented in Table 1.1*

4.2.2 Validated WAM versus *ERA-40*

To extrapolate the five validated model locations backward in time functional relationships with the *ERA-40* locations from 2001 must be established. After synchronizing the validated model data to fit time intervals and recording times of the 2001 *ERA-40* data the scatter plots in Figure 4.10 were made. The *RMS* error seen at all locations relates to values of high accuracy, while the correlation coefficients, ranging from 0.85 to 0.91, indicate a very strong degree of linear dependence between the model and *ERA-40*. However, a general underestimation by the *ERA-40*, though varying in magnitude, is observed at all deep water locations. For the *North of island*, *Oil* and *Surf* locations the vertical displacements read $-0.04, -0.04$ and -0.01 m, respectively, insignificantly small values which may be neglected compared to their high values of slopes which yield $11\%, 23\%$ and 9%. For the *South of island* location the displacement constant is of a higher value, and the model to *ERA-40* ratio reads $Hs_{WAM\ validated} = Hs_{ERA-40} \times 1.18 - 0.15$, indicating a slight overestimation for *ERA-40* SWHs just below one meter. At the *Sofala Bank* the functional relationship reads $Hs_{WAM\ validated} = Hs_{ERA-40} \times 0.80 + 0.01$, indicating an overestimation for *ERA-40* SWHs which increases towards higher wave heights. With the functional relationships listed in Table 4.2 the validated model can now be extrapolated back to September 1957.

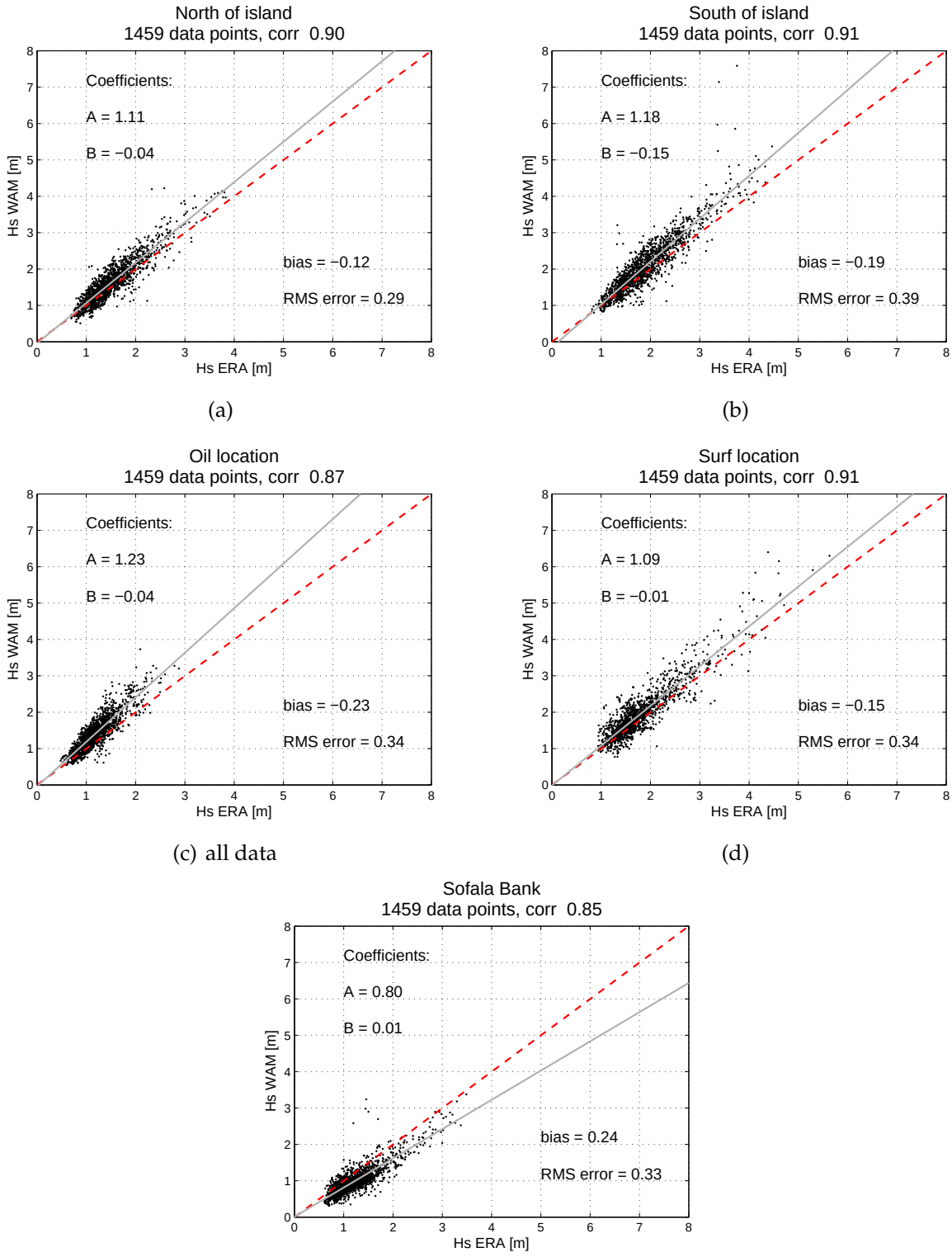


Figure 4.10. *Scatter plots with estimated functional relationships coefficients for the 2001 validated nested model and ERA-40 SWHs. A and B correspond to the slope and vertical displacement at the y axis of the gray least-square line, respectively. The number of data points on which the regression is based is given in the title along with the correlation coefficient.*

Table 4.2. *Functional relationships used to extrapolate the validated model data back to 1957.*

Location	Functional relationship
North of island	$Hs_{WAM\ validated} = Hs_{ERA-40} \times 1.11 - 0.04$
South of island	$Hs_{WAM\ validated} = Hs_{ERA-40} \times 1.18 - 0.15$
Oil location	$Hs_{WAM\ validated} = Hs_{ERA-40} \times 1.23 - 0.04$
Surf location	$Hs_{WAM\ validated} = Hs_{ERA-40} \times 1.09 - 0.01$
Sofala Bank	$Hs_{WAM\ validated} = Hs_{ERA-40} \times 0.80 + 0.01$

Discussion

Unfortunately no previous work with which to compare this locally validated WAM model exists. However, *ERA-40* validation assessments carried out by Caires & Sterl (2003a) and Caires & Sterl (2003b) found 20% and 17% underestimation on a global basis, respectively. This concur well with the averaged underestimation of the four deep water locations which adds up to 15%. For the *Sofala Bank* this does not apply as the *ERA-40* model does not account for shallow water effects, implying the SWH data is only valid in deep water regions. The functional relationship seen at the *Sofala Bank* is therefore not comparable with the latter mentioned works. The *Sofala Bank* overestimation is likely an effect caused by the coarse resolution where the bathymetry is not adequately resolved, hence missing the frictional effect from the bottom.
Large spatial resolution might also be the cause for the general underestimation experienced by the *ERA-40* data, where the modeled high wind speeds are likely underestimated, causing a negative feedback for the waves. Caires & Sterl (2003a,b) further reported an overall effect of overestimation which was found over SWH smaller than 1.5 m. This is not seen in the validated *WAM* and *ERA-40* scatter plots. Although not vital in this work, this effect might be revealed when the collocated altimeter and *ERA-40* data is assessed in the following section.

4.2.3 ERA-40 versus altimeter

The final validation assessment carried out consider the collocated altimeter and unaltered *ERA-40* data extracted from the altimeter domain illustrated in Figure 3.1. In this collocation each altimeter measurement is paired with the closest *ERA-40* grid point, with a maximum deviation of 3 hours and approximately 1° latitude/longitude in time and space. Again, the maximum distance, in this case 1°, corresponds to half the diagonal of the respective model grid. The scatter plot based on this data set is seen in Figure 4.11 along with statistical properties. In the case where successive altimeter measurements occur within the region of 1°, multiple altimeter records are assigned to the same *ERA-40* point, a circumstance appearing as vertical concentrations in the scatter plot.

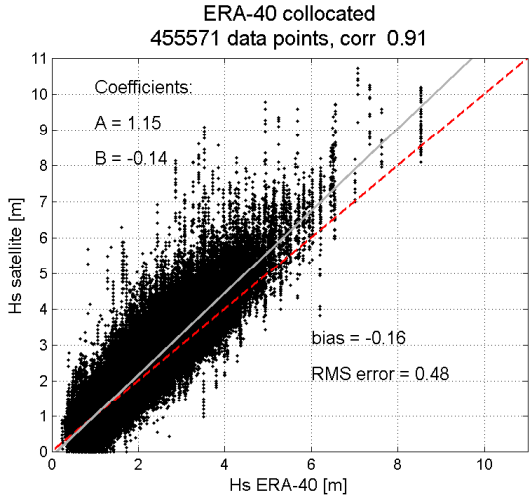
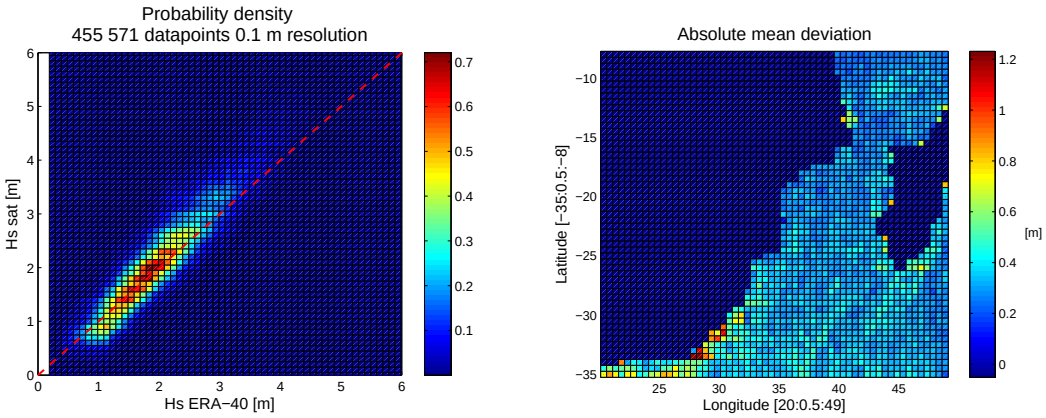


Figure 4.11. *Scatter plot with statistical coefficients. The linear dependence of ERA-40 and altimeter is described by the straight (gray) regression line $y = Ax + B$, where A is the slope and B is the intercept on the y axis. The number of data points of which the regression is based on is given in the title along with the correlation coefficient.*

A functional relationship of $Hs_{altimeter} = Hs_{ERA-40} \times 1.15 - 0.14$ is illustrated by the gray least-square line. The negative intercept value of -0.14 m causes the gray least-square line to cross the red dashed one-to-one ratio line, indicating a region of overestimation below the point of intersection of one meter. As for the correlation coefficient, *RMS* error and bias, the values correspond to highly linear dependent variables with a distinct degree of accuracy.
In Figure 4.12 the probability density distribution within the altimeter and *ERA-40* SWH range is seen in panel (a), whereas panel (b) illustrates the mean absolute deviation between altimeter and *ERA-40* for the collocated region. The probability density plot displays a high probability belt with the peak occurrence of 2.0 m, only slightly shifted away from the one-to-one ratio line. For the upper tail the underestimation of the *ERA-40* model is stronger, nevertheless, also here the overestimation for waves of modest height is evident. The mean absolute deviation map depicts the same pattern as seen for the full collocated nested WAM and altimeter plot, namely that the highest deviations are found along the coast. In this plot, the coast of Madagascar, the southernmost region of the African coast, and the coastal area inshore from the *Oil location* are regions over which large deviations prevail.



(a) (b)

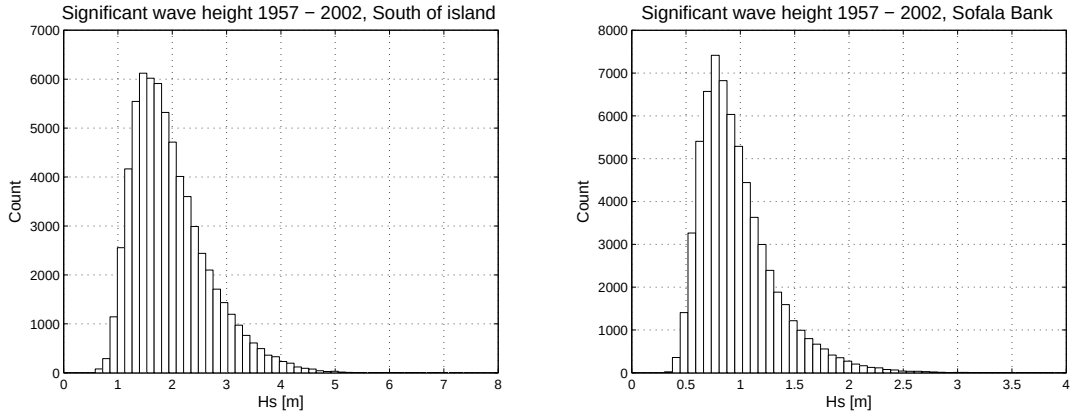
Figure 4.12. ***(a)*** *Density probability plot of ERA-40 and altimeter data. The wave height resolution is 0.1 m and the probability of any occurrence is indicated by the color bar. The red dashed line indicates the one-to-one ratio between the ERA-40 model and altimeter SWHs.* ***(b)*** *Mean absolute deviation from ERA-40 and altimeter data with a spatial resolution of* 0.5° *latitude and longitude.*

Discussion

The functional relationship from the two different approaches of *ERA-40* validation, namely the averaged deep water individually validated WAM locations and the direct altimeter approach, relates well, reading $Hs_{WAM\ validated} = Hs_{ERA-40} \times 1.15 - 0.06$ and $Hs_{sat} = Hs_{ERA-40} \times 1.15 - 0.14$, respectively. However, a slightly larger vertical displacement is seen in the latter, a property which makes it in better agreement with the results by Caires & Sterl (2003a), where overestimation for SWH less than 1.5 m is reported compared the corresponding approximately 1.0 m found here. This overestimation is not pronounced in the functional relationship between the validated WAM and altimeter data, except at the *South of island* location. Still, the underestimation of 15% found in both assessments, corresponds well with the 17% found by Caires & Sterl (2003b).
Regarding the correlation coefficients, the values of 0.91 and 0.90 obtained from the collocated *ERA-40* versus altimeter and the averaged deep water validated WAM locations versus *ERA-40*, respectively, are superior to 0.79 found over the Indian Ocean during January 1993 to December 2001 by Uppala *et al.* (2005). A comparison which also was based on altimeter data.
Reinforcing the credibility of the collocated *ERA-40* and altimeter validation assessments is the excellent RMS error correspondence with Caires *et al.* (2005). In this work Caires *et al.* (2005) found a RMS error of 0.46 m by validating *ERA-40* data with *TOPEX* altimeter data over the Indian Ocean in 2001, a value only slightly lower than the 0.48 m which was found here. As for the RMS error value found when averaging the four values obtained from the individually assessed validated WAM versus *ERA-40* scatter plots, an even higher accuracy of 0.34 m is seen. The discrepancy of the *ERA-40* data is most likely a resolution effect, where the low and high wind speeds are over and underestimated, respectively, which accordingly effects the waves in the same manner.

4.3 Wave climate

With five locations, each possessing 65 744 validated data points, equivalent to a time series of 45 years, the wave climate can now be assessed. Figure 4.13 displays histograms from two extreme locations; the *South of island* location and the *Sofala Bank*, where the highest and lowest SWHs are observed, respectively. Right skewed distribution patterns are seen for both locations, whereas a factor of two differ the SWH between the locations. At the *Sofala Bank* the mode is located at 0.75 m compared to 1.5 m *South of the island*. Extreme SWHs of 7.4 m and 3.9 m are modeled at the *South of island* and the *Sofala Bank*, respectively, during the historical wave record.



(a) (b)

Figure 4.13. *Histograms illustrating the ERA-40 SWH distribution for the South of island location,* ***(a)****, and Sofala Bank,* ***(b)****. The occurrences which fall within the 50 bins on the x-axis are counted on the y-axis.*

Despite 45 years of data is expressed in Figure 4.13, little information is conveyed regarding the extremes occurring within this period. For this reason the single highest SWH measurements from each year are plotted in Figure 4.14. Here both the *South of island* and *Sofala Bank* are represented along with a least square line indicating the linear trend. The curves show a highly fluctuating nature. For either position a trend of increase is observed throughout the 45 years, approximating to 0.4 and 0.2 m for the *South of island* and *Sofala Bank*, respectively. However, by assuming stationarity of the data, statistical distribution models can extrapolate the data into a period in which no data is obtained yet. This is addressed in the following sections, where first the AM method is applied, followed by the POT method, to estimate the return values of the SWH which is likely to be exceeded only once during a period of 100 years.

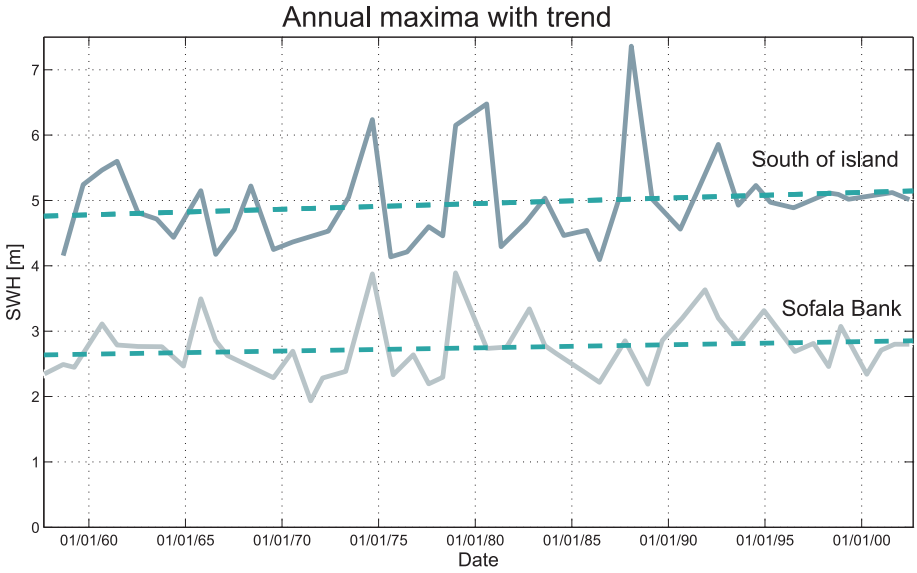


Figure 4.14. *Annual maxima plot for the South of island location and the Sofala Bank with linear trends indicated by the dashed lines.*

4.3.1 Annual Maxima method

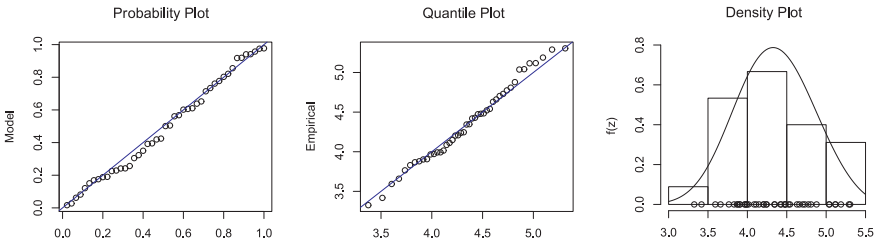
For each location the GEV distribution is fitted to AM data. The parameters were estimated by ML and are listed in Table 4.3 along with CIs calculated by both the Delta method and the profile likelihood method. From this table it appears that for all locations, except at *South of Island*, negative estimated values of ξ are obtained. This implies that a Weibull distribution is fitted to the AM data. However, accounting for the CIs of ξ, the model association is not convincing as the CI spans both negative and positive values of ξ, indicating an attraction to the unphysical Fréchet distribution. For *North of island* this is not the case, where the data is appointed to only one distribution family, namely the Weibull, even after the CI is accounted for. It should also be noticed that the CI found by the profile likelihood method is more confined than the CI based on the Delta method.

Table 4.3. *Parameter estimates found by the ML method. The 95% CIs of ξ are calculated by the Delta method, CI_{dm}, and Profile likelihood method, CI_{pl}.*

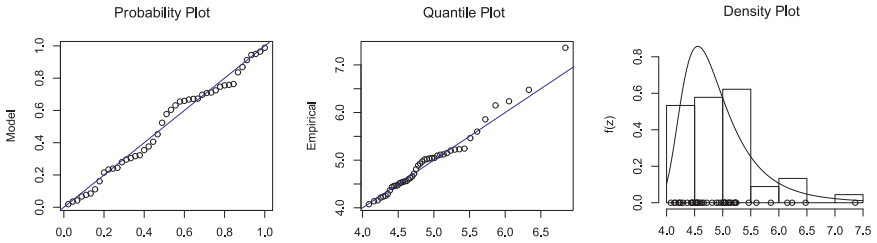
	Parameter estimates		
Location	ξ, CI_{dm} and CI_{pl}	σ	μ
North of island	-0.28 (-0.54, -0.02) (-0.46, -0.10)	0.49	4.17
South of island	0.15 (-0.12, 0.42) (-0.02, 0.32)	0.43	4.62
Oil location	-0.09 (-0.40, 0.23) (-0.27, 0.09)	0.45	3.51
Surf location	-0.15 (-0.34, 0.04) (-0.31, 0.08)	0.67	4.95
Sofala Bank	-0.03 (-0.23, 0.17) (-0.15, 0.09)	0.34	2.56

The graphical goodness-of-fit tests seen in Figure 4.15 further impair the credibility of the model fits. Here deviation from linearity is seen in both the probability and Q-Q plots, albeit to a lesser extent at the *North of island* and *Surf location*. The highest deflections from the diagonal are seen at the higher end of the AM values, best expressed in the Q-Q plots, but still, deflections are also clearly evident throughout the probability plots. In the density plots strong discrepancies between the probability density function of the model and the histogram based

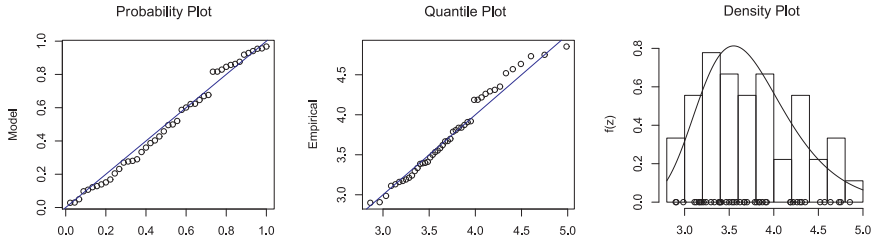
on the AM are seen, but are hard to interpret due to the histogram's shape dependence of bin selection. Despite the goodness-of-fit tests imply a quality of fit which is not perfect, all values of AM are within the CI band of the return levels seen in Figure 4.16. In Figure 4.16 (b) and (c) the AM data attraction to two different GEV families is well illustrated by the concave and convex shape of the upper and lower CI, respectively. Whereas, at panel (e) an attraction to all three GEV families is observed as the best estimate line which occurs linear, indicating a Gumbel fit, is flanked by both the Fréchet and Weibull families. At the *North of island* both the best estimate and the CI curves have convex shapes.
For all panels the increase in uncertainty involved when extrapolating to high leveled return values is highlighted, particularly so for models with CIs assigned to different model families.



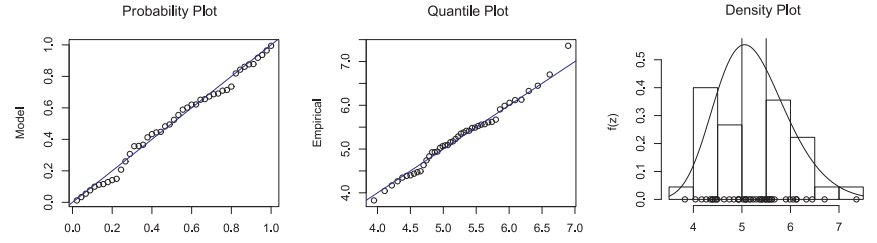
(a) North of island



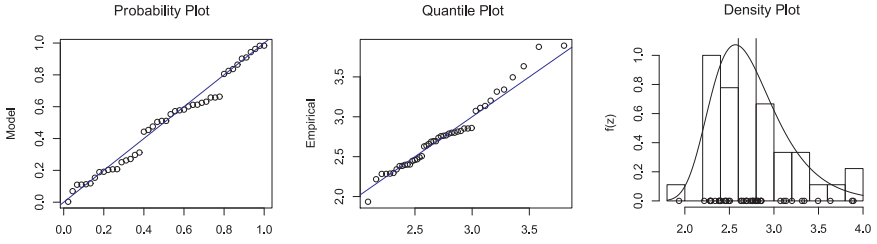
(b) South of island



(c) Oil location



(d) Surf location



(e) Sofala Bank

Figure 4.15. *Goodness-of-fit tests for the GEV modeled AM data. Deviation from linearity of the circles in the probability and quantile-quantile plots indicates a mismatch of distribution model (solid line) and SWH data (circles). The units in the center panel are in meters. The panels to the right represent the modeled probability density function (solid line) against a scaled histogram based on the SWH observations (circles) in meters. Plotted by the extRemes package (Gilleland* et al.*, 2009).*

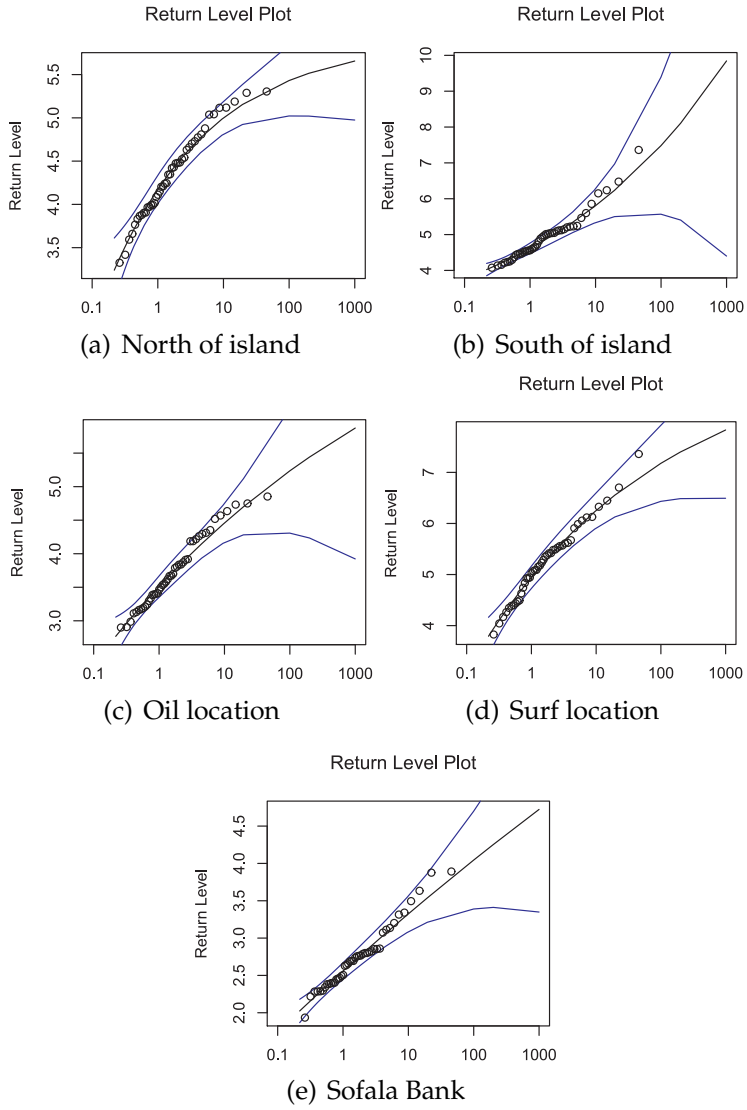


Figure 4.16. *Return level plots with* 95% *CI based on the GEV modeled AM data. The x-axis is given in years on a logarithmic scale, whereas the y-axis denotes SWH in meters. Note that the scale on the y-axis varies for each location. The CIs are found by the Delta method, and are plotted by the extRemes package (Gilleland* et al.*, 2009).*

The return level estimates for a period of 100 years range from 4.0 m at the *Sofala Bank* to 7.5 m at *South of island*, without accounting for the CIs. However, for a H_s^{100} a CI should be given, but as illustrated in Figure 4.16 and Table 4.4, the CIs are far to extensive to yield a return level of satisfying accuracy. An exception is seen at the *North of island* location where the CI spans 0.8 m only. As these broad CIs are likely owing to the sparse number of data on which these extrapolations are based, improvements are expected in the following section when the extrapolation is based on far more data extracted by the POT method and thereafter modeled by the GP distribution.

Table 4.4. *100-year return value estimates found by the GEV approach accompanied by the* 95% *CI found by the Delta method.*

Location	**GEV H_s^{100}, CI_{dm} [m]**
North of island	5.4 (5.0, 5.8)
South of island	7.5 (5.6, 9.4)
Oil location	5.2 (4.3, 6.2)
Surf location	7.2 (6.4, 7.9)
Sofala Bank	4.0 (3.4, 4.7)

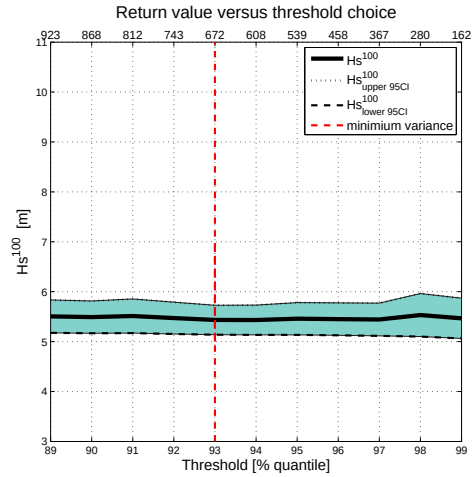
4.3.2 Peaks Over Threshold method

As previously mentioned, the threshold selection is vital in finding an accurate return value. In Figure 4.17, we see that this is particularly so when assessing the CI: The CI broadens as the threshold value increases, a direct consequence of the decreasing number of data points on which to base the CI calculation. Despite this broadening of the CI is justified by theory, anomalies at the *North of island* and the *Oil location* are observed. Here the CIs have a minimum variance for a centrally selected threshold, and the CI does not follow the generality of broadening with increasing thresholds.
The H_s^{100} estimate, on the other hand, remains surprisingly steady over the 89 to 99 percentile range, despite a minor increase for higher thresholds. Nevertheless, this plot is useful in the aspect of seeing how the threshold values influence on the number of peaks which exceed the threshold, and how this affects the CI.
It is here illustrated that the *Surf location*, followed by the *South of island* location, are prone to the highest 100-year return values, and, not surprisingly, the *Sofala bank* to the lowest values.

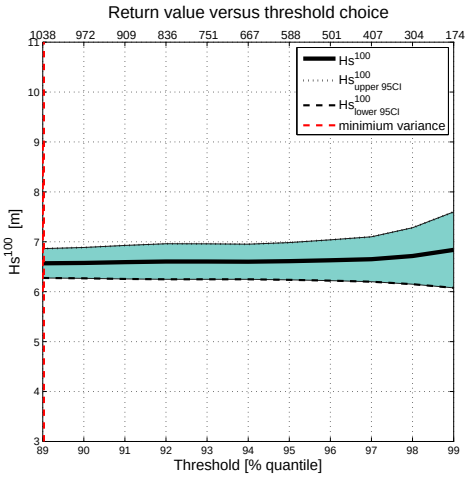
Before addressing the return values with reliable CIs, the threshold must first be selected. This is accomplished by the two previously mentioned exploratory techniques: Figure 4.18 illustrate these two methods applied on the five time series in which every storm is separated by 48 hours. The lowest threshold which represents a region in the parameter stability plots where both the modified scale and shape parameters are stable, and simultaneously belonging to an appurtenant linear segment of the mean exceedance plot is selected. Despite this is a subjective approach in finding a threshold, it is still the foundation on which new automated methods by e.g. Thompson *et al.* (2009) are based on. The selected threshold values for its respective location along with the corresponding quantiles are listed in Table 4.5.

Table 4.5. *Threshold selection according to Figure 4.18 and the coinciding quantile of the full time series.*

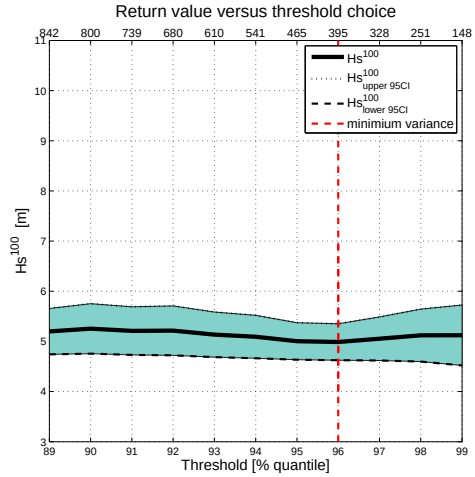
Location	**Thresholds [m]**	**Quantile [%]**
North of island	2.55	93.7
South of island	3.20	93.8
Oil location	2.40	95.9
Surf location	2.80	93.7
Sofala Bank	1.50	91.8



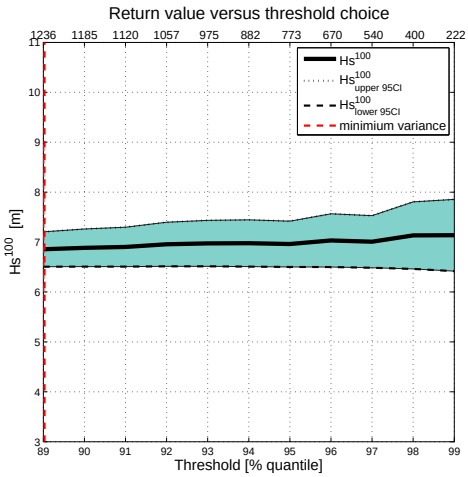
(a) North of island



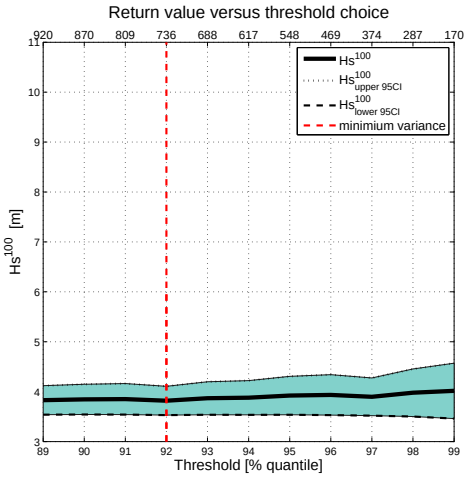
(b) South of island



(c) Oil location

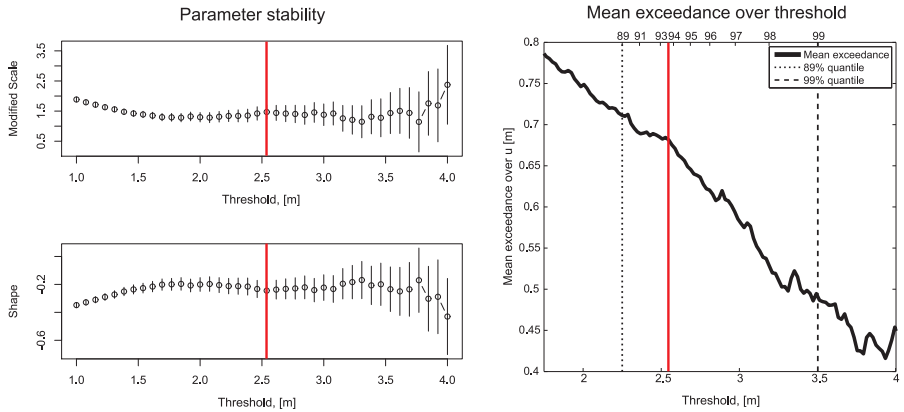


(d) Surf location

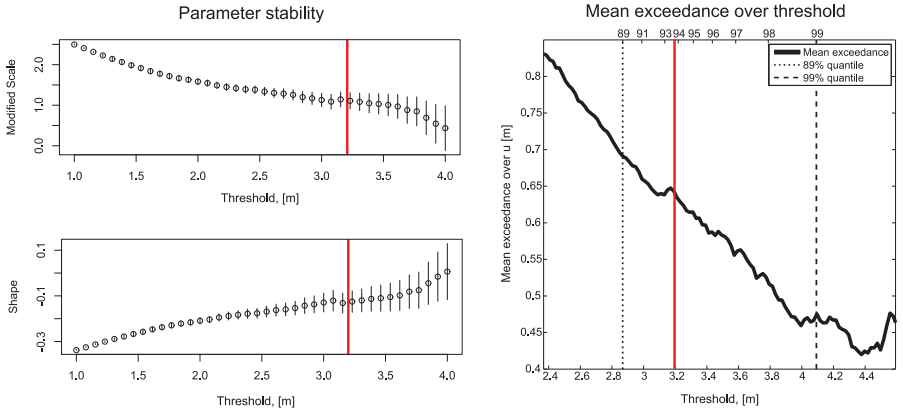


(e) Sofala Bank

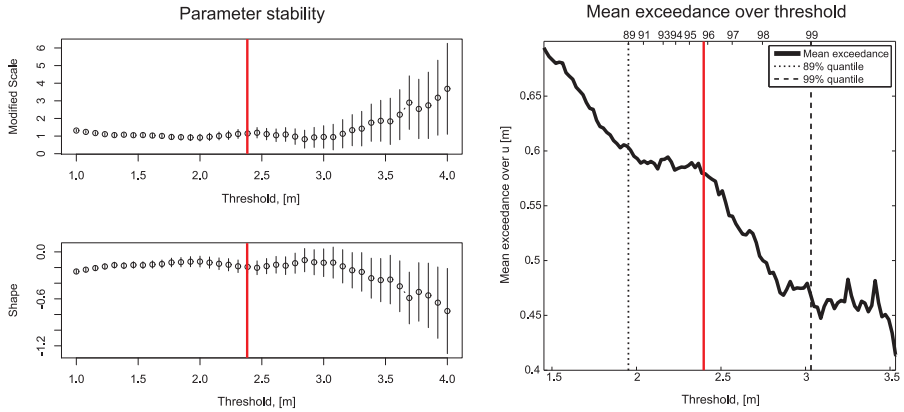
Figure 4.17. *100-year return level for all five locations plotted against a range of thresholds. The solid line represents the estimates of the H_s^{100} and is flanked by the upper (dotted) and lower (dashed) boundaries of the 95% confidence interval obtained by the Delta method. The vertical dashed line localizes the threshold at which the smallest variance of the return value occurs. The unit of the lower x-axis represents the threshold quantile of the full ERA-40 time series of the respective location. At the upper x-axis the number of data points corresponding to the thresholds quantiles are seen, whereas the y-axis is the 100-year return level of SWH measured in meters.*



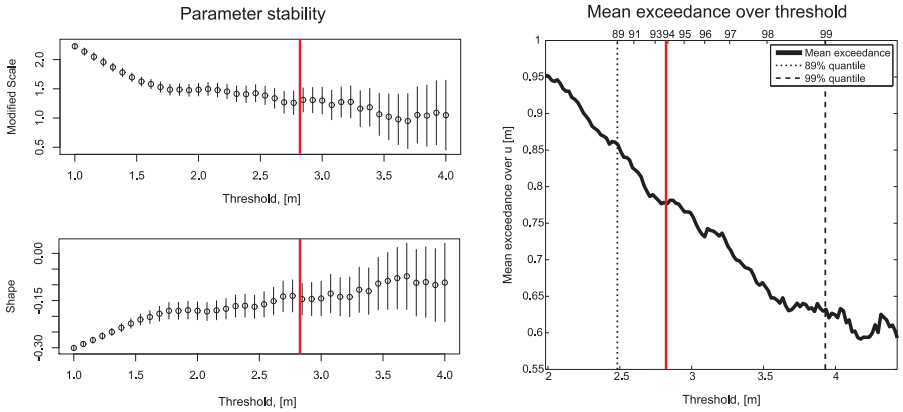
(a) North of island



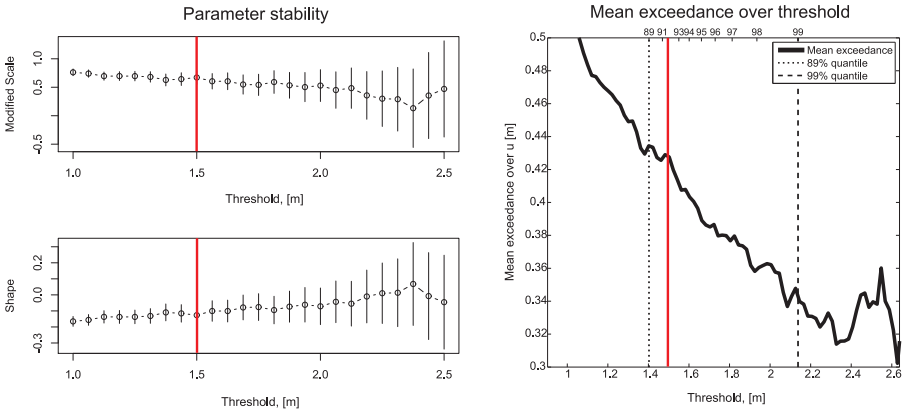
(b) South of island



(c) Oil location



(d) Surf location



(e) Sofala Bank

Figure 4.18. *Graphical techniques applied to fix the threshold for each location as described in section 3.4.2. The solid vertical red lines represent the threshold selections, and are fixed after a visual examination taking both the parameter stability and mean exceedance plot into account. At the upper x-axis of the mean exceedance plot the quantile [%] of the full time series which the threshold corresponds to are seen. Additionally, the dotted and dashed vertical lines correspond to the 89 and 99% quantiles, respectively, and delimit the region in which a threshold is likely located. The parameter stability plots were generated using the extRemes package by Gilleland* et al. *(2009).*

With a threshold appointed to each location the model parameters are estimated by the *ML* method using the exceedances. These parameters are seen in Table 4.6. In contrast to the shape parameters obtained by the GEV distribution, both the *Delta* and the *Profile Likelihood* methods yield shape parameters which span negative values only. Therefore, the exceedances will be modeled with the *type III* family, and both the best return value estimate and the CI will have convex shapes in the return level plots, implying an appropriately chosen model which have an upper bound. Lastly in Table 4.6 the average number of data points which annually exceeds the threshold are tabulated. From this we see that the POT approach provides us with 9.2 to 20.5 times more data than what the AM approach did.

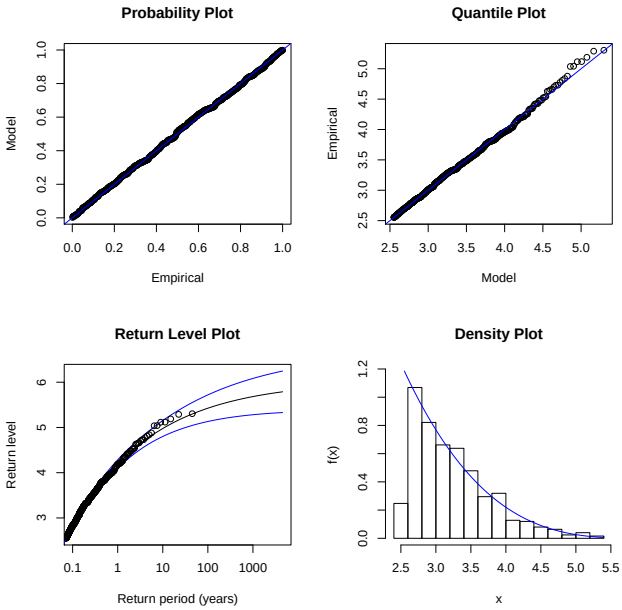
Table 4.6. *Parameter estimates found by the ML method. The 95% CIs of ξ are calculated by the Delta method, CI_{dm}, and Profile Likelihood method, CI_{pl}. n_y represents the average number of data points exceeding the threshold per year.*

	Parameter estimates		n_y
Location	ξ, CI_{dm} and CI_{pl}	$\tilde{\sigma}$	
North of island	-0.24 (-0.31, -0.18) (-0.30, -0.17)	0.84	13.9
South of island	-0.13 (-0.18, -0.09) (-0.17, -0.08)	0.72	15.0
Oil location	-0.19 (-0.28, -0.11) (-0.27, -0.10)	0.69	9.2
Surf location	-0.14 (-0.19, -0.09) (-0.18, -0.08)	0.88	20.5
Sofala Bank	-0.13 (-0.18, -0.07) (-0.18, -0.07)	0.48	16.5

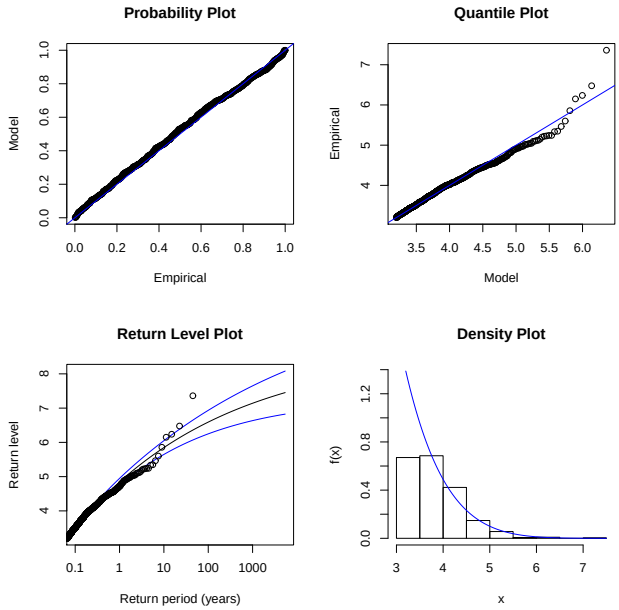
The goodness-of-fit of the models are assessed in Figure 4.19, along with the return value plots. Neither the probability plot nor the Q-Q plots give any reason to doubt the quality of the model fit as the observations occur linear. However, some deviations from linearity are observed at the upper end of the Q-Q plots, albeit minor at the *North of island* and the *Oil location*. Reasonable fits are also seen in the density plots, particularly if neglecting the lowest exceedance bin. The bounded nature of the curves in the return level plots confirms the association to the *type III* family. In addition, the CIs are narrower than what was obtained from the GEV approach, stemming directly from the increased number of data. However, this narrowing of the CI results in occasional outliers which are seen beyond the bounds of the CI in the return level plots. In comparison with the estimated 100-year return values obtained by the GEV approach, the GP estimates are slightly lower. The major distinction between the models occurs for the CIs; revealed in Table 4.7, the CIs from the GP approach are considerably more confined.

Table 4.7. *100-year return levels estimates from both the GEV and GP approach with 95% confidence interval by the Delta method.*

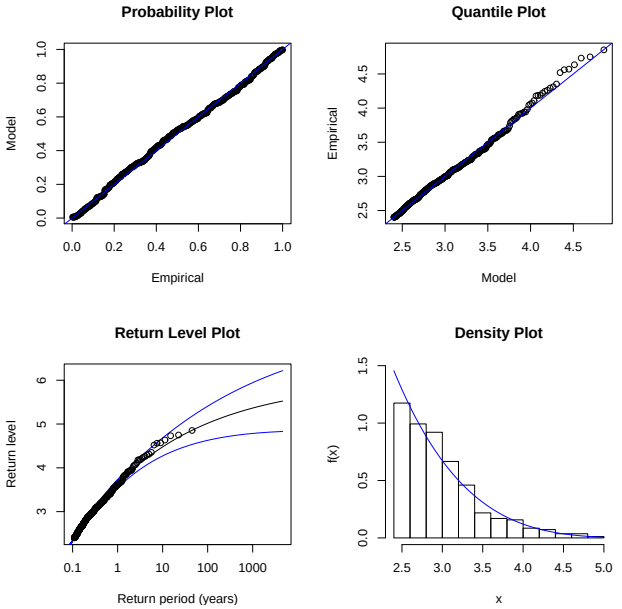
	100-year SWH return value, H_s^{100} **[m]**	
Location	**GEV**, CI	**GP**, CI
North of island	5.4 (5.0, 5.8)	5.4 (5.1, 5.7)
South of island	7.5 (5.6, 9.4)	6.6 (6.3, 6.9)
Oil location	5.2 (4.3, 6.2)	5.0 (4.6, 5.4)
Surf location	7.2 (6.4, 7.9)	7.0 (6.5, 7.5)
Sofala Bank	4.0 (3.4, 4.7)	3.8 (3.5, 4.1)



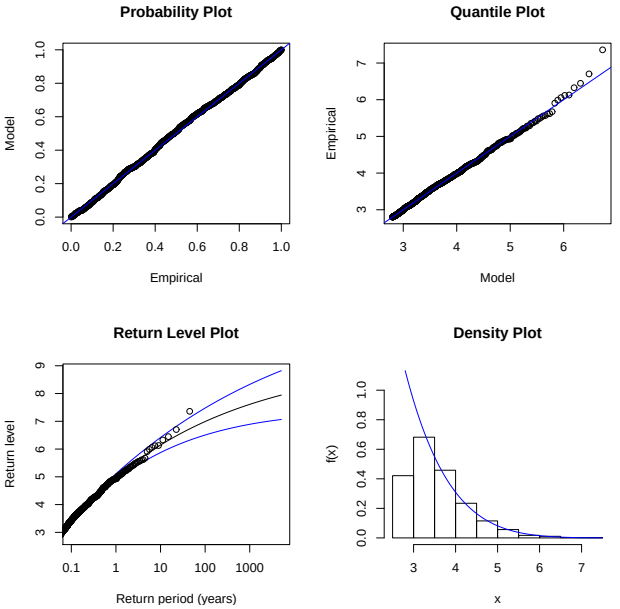
(a) North of island



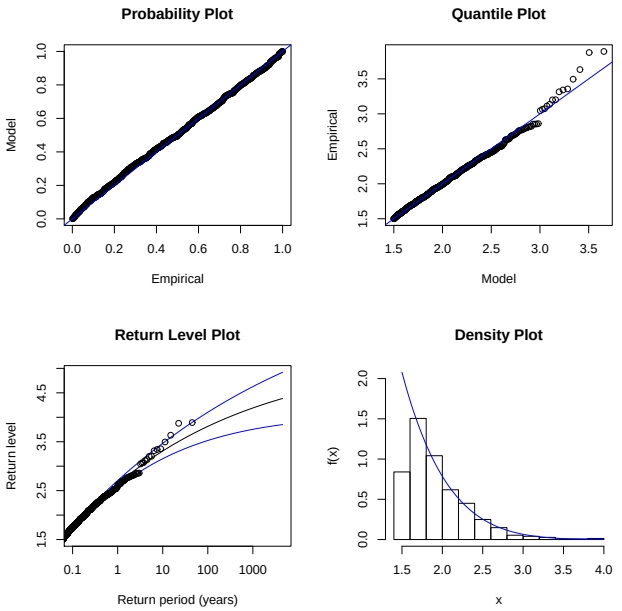
(b) South of island



(c) Oil location



(d) Surf location



(e) Sofala Bank

Figure 4.19. *Goodness-of-fit tests for the GP modeled POT data. Deviation from linearity of the circles in the probability and Q-Q plot indicate a mismatch of distribution model (solid line) and SWH data (circles). The panels on the bottom right represent the modeled probability density function (solid line) against a scaled histogram based on the SWH observations in meters. The return level plots with 95% CI found by the Delta method is seen at the bottom left panel. The x-axis is given in years on a logarithmic scale, whereas the y-axis denotes SWH in meters. All plots made by the extRemes package (Gilleland* et al.*, 2009).*

4.3.3 Discussion

According to the goodness-of-fit tests, the most suitable model fits were achieved by the GP distribution. This is supported by the alternating sign of the shape parameter's CI obtained by the GEV approach. The attraction to multiple domains is an indication of a non-persuasive model fit and that the GEV approach is not entirely appropriate for this purpose. For the GP approach the consistency of the shape parameters argue the opposite; an appropriately chosen threshold yielding exceedances which are suitably modeled by the upper bounded Weibull distribution. However, as it appears from Table 4.8, which also includes the Profile likelihood based CIs, the H_s^{100} estimates found by parameters from the GEV distribution are in reasonable agreement with those found by the GP distribution, albeit slightly overestimated. This overestimation, which is particularly evident at the *South of island* location, supports the statement from Harris (2004, 2005), which claims that return values obtained from the Fréchet distribution always exceeds those from a Weibull distribution. The concept on which this is based is illustrated in Figure 3.3. As for the overestimation at the other locations, this statement does not hold as the shape parameters indicate an attraction to the same family. However, as the shape parameters generally are lower in value for the GEV approach, the same principle might justify the overestimation.

The most significant deviation is observed in the CIs of the different approaches. Figure 4.20 illustrates this well; where the CIs of the GEV estimates (disregarding which CI method applied) ranges well beyond those of the GP approach. As previously mentioned, this feature is caused by the sparse availability of data in the GEV approach. Figure 4.20 further displays the symmetrical CIs associated with the Delta method in contrast to the skewness related to the CIs achieved by the Profile likelihood method. This skewness stems from an assumption used in the approach when maximizing the log-likelihood function, and represents the increasingly weaker information provided at the high levels of the waves, hence representing the true nature of waves to a higher extent. As a direct consequence, also illustrated in Figure 4.20, the Profile likelihood based CI surpasses the variance of the Delta method based CI at the upper end, whereas at the lower end the Delta method yields an estimate of larger variance. For the GP approach, which well justified can be claimed to be of a higher accuracy, the CIs from the Delta method and Profile likelihood method do not deviate drastically. In fact, the deviance is never more than 0.1 m at the upper and lower estimated interval.

Table 4.8. *100-year return levels with 95% confidence interval obtained by GEV and GP. CI_{dm} and CI_{pl} corresponds to confidence intervals obtained by the Delta method and the Profile likelihood method, respectively. The last column represents the number of observations which are greater than or equal to the lower CI, colored blue, found by the GP approach using the Profile Likelihood method.*

	100-year SWH return value, H_s^{100} **[m]**		**# of SWH**
Location	**GEV**, CI_{dm} and CI_{pl}	**GP**, CI_{dm} and CI_{pl}	$\geq$**GP,** $\mathbf{CI}_{pl}^{lower}$
North of island	5.4 (5.0, 5.8) (5.2, 6.0)	5.4 (5.1, 5.7) (**5.2**, 5.8)	3
South of island	7.5 (5.6, 9.4) (6.4, 10.4)	6.6 (6.3, 6.9) (**6.3**, 7.0)	2
Oil location	5.2 (4.3, 6.2) (4.7, 6.6)	5.0 (4.6, 5.4) (**4.7**, 5.5)	3
Surf location	7.2 (6.4, 7.9) (6.7, 8.3)	7.0 (6.5, 7.5) (**6.6**, 7.5)	2
Sofala Bank	4.0 (3.4, 4.7) (3.6, 5.0)	3.8 (3.5, 4.1) (**3.6**, 4.1)	3

Even if the CI[1] of the GP approach gives the H_s^{100} estimates with accuracy ranging from $\pm$ 0.25 to 0.45 m, a further narrowing at the lower CI is implied by the last column in Table 4.8. Here the number of observations which are greater than or equal to the lowest CI level for each location are tabulated. As we see, 2 to 3 occurrences exceed the lowest CI, an indication that the H_s^{100} range is not as low in value as the CI imply.
A corresponding inference is not applicable at the upper end. However, an inference on the maximum limit of the upper CIs can be made, based on the fitted models. This value is regarded as the upper CI of the return value with 'infinite return period' and is the most severe condition which according to the models can ever occur. The values are found by inserting $N = \infty$ in Eq. (3.9) when $\xi < 0$, which yields $z^{\infty} = u - \frac{\tilde{\sigma}}{\tilde{\xi}}$. Applying this relation on the upper CI of the parameters given in Table 4.6 and the thresholds given in Table 4.5, yields 7.5, 12.2, 9.3, 13.8 and 8.4 m, in the same order as tabulated. Nevertheless, one should be aware that all values, both the H_s^{100} and its CIs, are strictly applicable only within the period in which the data are sampled, while values obtained by extrapolation are merely guesses based on the behavior of the data Coles (2001).

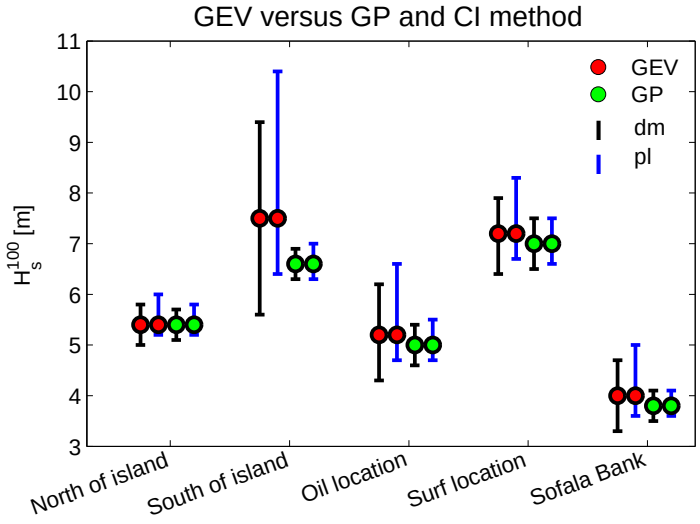


Figure 4.20. *Graphical illustration of Table 4.8. Red and green circles indicate GEV and GP, respectively. The black and blue lines represent the method by which the CI is calculated, namely the Delta method and the Profile likelihood method, respectively.*

Regarding the credibility of the return value estimates presented, there are several delicate issues which in the following are addressed:

- As illustrated in Figure 4.14, the annual maxima data applied in the GEV approach has a positive linear trend. However small such a trend might be, it challenges the applicability of the traditional GEV and GP models, as these are based on stationary processes (Coles, 2001). As a means to account for seasonal effects, annual maxima values were selected, hence avoiding seasonality altogether. The possible long term climate change and atmospherical cycles with periodicity greater than a year are unfortunately not

[1] As the Profile likelihood method yielded the narrowest CIs and additionally accounts for the skewed distribution of information which prevails for natural phenomenons as waves, the CI referred to as from now will be the CI obtained by this method.

accounted for. When adopting the models to non-stationary conditions, it is reasonable to assume that even though the SWH changes linearly over time, the distribution remains unchanged. This might be accounted for by modifying the location parameter to be time dependent. The GEV distribution becomes $Z_t \sim GEV(\mu(t), \sigma, \xi)$, where $\mu(t) = \beta_0 + \beta_1 t$, in which β_0 and β_1 are parameters representing the linear trend's intercept with the vertical axis and the annual rate of change, respectively. The GP approach is modified in a similar manner, additionally with the threshold being time dependent, $u(t)$. Without going further into the methods of non-stationary modeling, it suffices to say, that the current climate alone should not be used to predict future return values without accounting for possible trends. Without these adjustments, the return values might be regarded as lower estimates, that is, if the existence of a positive trend has been established.

- The estimates given in this work are based on SWHs from the global *ERA-40* data set. These SWHs represent the average conditions during the IFS model's respective spatial and temporal resolution. For this reason the return value estimates presented here are averages, which are likely exceeded at shorter spatial and temporal scales. Most likely, the highest individual wave will occur during the period of the highest SWH, however there is a small probability that it might occur during a time with lower values of SWH. For methods that derive return value estimates based on individual wave height see the work by Stanton (1984, and references therein). Yet, the relation given in Eq. (2.18), indicates a factor two increase based on the SWH return values.

- As the intensity of tropical cyclones (TC) is not fully resolved in the large spatial and temporal scale of the *ERA-40* data, an underestimation of SWH might occur. Little is known regarding the extent to which TCs affect the return values and the severity of the TC generated waves in the Mozambique Channel. However, as the ERA-40 reanalysis does detect TCs (see section 3.3), there is little doubt that TC generated waves are incorporated in the data on which the return values have been based. This is supported by Figure 4.21, where the 413 data points which exceeds the threshold of 3.2 m are plotted in monthly bins in a histogram for the *Oil location* during the full ERA-40 period. Here a concentration of exceedances is seen during the peak TC season, January and February, and illustrates the possible contribution of TC generated waves to the return value estimates. This clustering of exceedances during the summer occur exclusively at the *Oil location*. The other four locations are less exposed to TC generated waves, except when in rare occasions the TC pass directly into the Mozambique Channel. The histograms for the remaining locations describe a distribution which have a peak during June and July and minima during the summer.
  To further complicate this matter it must be mentioned that the basic extreme value models applied in this work are not capable of modeling both synoptic scale and TC generated waves. This is because the models assume that the extreme value observations must come from the same probability distribution, an assumption which does not hold if TC generated waves would have been included as they are generated by a different physical process (Bouws *et al.*, 1998). Nevertheless, in turn this might be yet another reason which suggests that my results are too low.

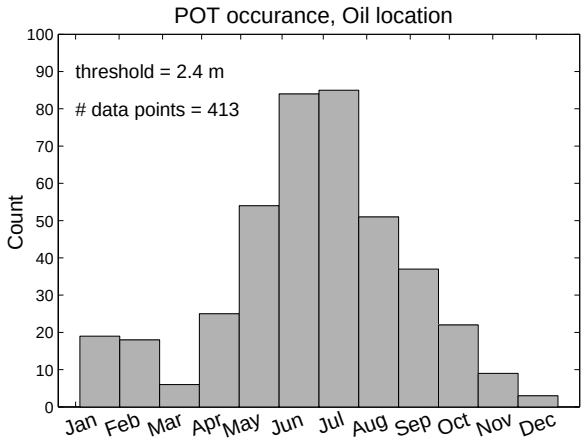


Figure 4.21. *Histogram illustrating the monthly occurrence of the 413 exceedances during the ERA-40 data at the Oil location. The cluster of exceedances occurring during January and February illustrates the possible inclusion of TC generated waves in the GP model.*

- The fact that the *ERA-40* reanalysis was run for deep water only makes the return value estimate obtained at the shallow *Sofala Bank* questionable. Even as the linear relationship from the WAM model was applied on the data, an uncertainty regarding this shallow location's validity remains.
- The last caveat, regarding the return values, addresses the domains of the coarse and nested WAM model. Without considering the computational cost, the domain should have been extended, particularly in the westward direction. The coarse model resolves wind seas to the fullest, as these build up over short distances compared to swell. However, there is a possibility that swell generated from a storm outside the model domain is allowed to propagate undetected into the region of interest. The effect of these missing swells is most likely insignificant, however, a comprehensive study of storm track history (e.g (Young, 1999; Hoskins & Hodges, 2005)) should have been carried out prior to the domain designation.

The results obtained in this work are not readily validated nor assessed by means of previous work. The spatial resolution of the works by Alves & Young (2003) and Caires & Sterl (2005), in which global 100-year return values have been considered, makes a comparison rudimentary.

Chapter 5

SUMMARY AND CONCLUSION

Based on 45 years of SWH data, 100-year return values for five locations within the Mozambique Channel have been estimated. In the approach of refining and validating the ERA-40 SWHs, data from a high resolution WAM model was applied. Accordingly, the results obtained in this work feature a high resolution and accuracy which will be prominent for coastal regions especially.
The overall results indicate that the southern region of the channel is prone to higher return values than what the northern region is. At the *Surf location* the 100-year return value is estimated to 7.0 m with a 95% CI of 6.6 to 7.5 m. Second highest is the open ocean location, *South of island*, located approximately 600 km further east and 170 km further north, with a value reading 6.6 m. 500 km north from here the return value has decreased to 5.4 m. At the location which represents the region in which the oil activities take place, 670 km further north from the latter location, a value of 5.0 m is found. At the shallow *Sofala Bank*, located centrally in the channel, the lowest value of 3.8 m is found.

By means of goodness-of-fit tests along with tables, it has been shown that the 100-year return level estimates obtained by the GP approach are preferable to those obtained by the GEV approach. This statement is heavily based upon the AM data's attraction to several families, but also on the superior CIs yielded by the GP approach. Further, it is evident that the threshold selection, which initially was believed to markedly influence the return values, in fact is not of utmost importance. Therefore, the results in this work support the statement by Caires & Sterl (2005), in which a threshold selection appointed at the 93% quantile of the full time series seems reasonable. This should also be sufficient when assessing point wise data from a larger area.
Once sufficient data is available with which to estimate the model parameters, the method chosen to assess the CI is a matter of preference, as both yield a closely related interval band. However, if the *Profile likelihood* method is adopted, a computational routine should be made in order to implement this approach in *MatLab*.

5.1 Further work

Along with the need of implementing the Profile likelihood method into *MatLab* further subjects require additional assessment and development. Following is a brief listing of such fields.

- The validity of ERA-40 data in coastal regions requires more attention. Even after the data has been refined by use of a high resolution model, uncertainty remains as the WAM model in the ERA-40 reanalysis did not account for bottom effects.
- Little is known about the extent to which TC generated waves are contributing to extreme events in the Mozambique Channel. However, it is likely that the occasional TCs which make their way into the channel generate waves with heights exceeding the threshold selection in the GP approach. For this reason the credibility of the traditional extreme value models have been challenged. Techniques, such as the one described by Spillane & Dexter (1976), should therefore be adopted in order to obtain results more appropriate for tropical regions.
- The vast amount of data available from the ERA-40 reanalysis opens for a wide range of topics for further work. Among these is a seasonal or monthly breakdown of the return values. Also a study regarding the mean wave period which corresponds to the return values is of interest. Another challenge is to account for characteristics in the data which change systematically over time, such as long term climate change. This is addressed in detail by Coles (2001).

Appendix A

A.1 Statistical formulas

Correlation coefficient

$$corr(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right]}} \tag{A.1}$$

Root mean square error

$$RMS\ error = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \tag{A.2}$$

Bias, mean error

$$bias = \frac{\sum_{i=1}^{N}(X_i - Y_i)}{N} \tag{A.3}$$

A.2 Matlab scripts

Generalized Extreme Value distribution, Delta method

```
function zpci=gev_ci(X,m)
%
% zpci=gev_ci(X,m)
%
% gev_ci returns the best estimate of the m-year retrun value, flanked
% by 95% confidence intervals found by the Delta method.
%
% X: input data. Annual Maxima (AM) data.
% m: return period
%
% see 'An introduction to statistical modeling of extreme values' by
% Sutart Coles, 2001, chapter 3 for supplementary information.

%% Annual Maxima data___optional_____________________________________
step=365.25*4; %average number of days
teller=1;
for i=121:step:((43*365.25*4)+121)      % starting October 1st
        AM(teller)=max(X(i:i+step));     %65700
        teller=teller+1;
end
last_year=max(X(44*365.25*4+122:end)); %data from 31-aug-2002 to
%                                      30-sept-2002 are missing to
%                                      complete the full year
AM=[AM last_year];
AM=AM';
%%_____________________________________________________________________

[parms,parmci]=gevfit(AM);%Obtaining the paramete estimates (ML method)
[nlogL,acov_1] = gevlike(parms,AM); %obtaining the information matrix

acov=[acov_1(3,3) acov_1(3,2) acov_1(3,1);...
      acov_1(2,3) acov_1(2,2) acov_1(2,1);... % Adjusting according to
      acov_1(1,3) acov_1(1,2) acov_1(1,1)];   % Coles 2001 page 59

SE=sqrt(diag(acov)); % Standard error of the parameters

p=1/m; % P value: probability for non-exceedance
yp=-log(1-p); %definition Coles page 56

gradzp=[1;... %inverse of gradzp, se Coles 2001 page 56
     -parms(1)^-1*(1-yp^-parms(1));...
     parms(2)*parms(1)^-2*(1-yp^-parms(1))-parms(2)*parms(1)^-1*yp^-...
     parms(1)*log(yp)];

zp=gevinv(1-p,parms(1),parms(2),parms(3));%Best estimate return value

var=gradzp'*acov*gradzp; %variance of zp
zpci=[zp-1.96*sqrt(var) zp zp+1.96*sqrt(var)];
% ci of zp, NOTE:+/-1.96*sqrt(var) corresponds to approximatly the
% 95% ci
symetri=[zp-zpci(1) zpci(2)-zp];%check for symmetri
```

Generalized Pareto distribution, Delta method

```
function xmci=gp_ci(X,threshold,numyr,m)
%
% gp_ci(X,threshold,numyr,m)
%
% gp_ci returns the best estimate of the m-year retrun value, flanked
% by the 95% confidence intervals found by the Delta method.
%
% X: input data. NOTE: should be i.i.d
% threshold: value of the threshold
% numyr: number of years which the data spans
% m: return period
%
% see 'An introduction to statistical modeling of extreme values' by
% Sutart Coles, 2001, chapter 4 for supplementary information.



varcovar=zeros(3,3);%Variance-covariance (VC) matrix shape
p=find(X>threshold); %locating the exceedances
data=X(p)-threshold; %Applying data which exceeds the threshol
zeta=length(data)/length(X); %Probability for a random observation to
% exceed the threshold

[parms,parmci]=gpfit(data); %ML parameter estimates
[nlogL,acov_1] = gplike(parms,data); %obtaining the information matrix
% Adjusting the order of acov so as to match the order in Coles 2001
% page 82.

acov=[acov_1(2,2) acov_1(1,2);...
      acov_1(2,1) acov_1(1,1)];
SE=sqrt(diag(acov)); % Standard error of the parameters

varcovar(2:3,2:3)=acov;%inserting parameter variance in the VC matrix
varcovar(1,1)=zeta*(1-zeta)/length(X); %inserting the var of zeta in
% VC matrix
M=m*length(X)/numyr; % return value times 'data per year'
mzeta=(M*zeta)^parms(1); %inverse of gradxm, se Coles 2001 page 82
gradxm=[parms(2)*M^parms(1)*zeta^(parms(1)-1);...
  parms(1)^-1*(mzeta-1);...
 -parms(2)*parms(1)^-2*(mzeta-1)+parms(2)/parms(1)*mzeta*log(M*zeta)];

P=1-1/(length(data)*m/numyr); % P value: probability for non-exceedance

xm=gpinv(P,parms(1),parms(2),threshold);%Best estimate of return value
var=gradxm'*varcovar*gradxm; %variance of xm
xmci=[xm-(1.96*sqrt(var)) xm xm+(1.96*sqrt(var))]; % ci of xm,
                           % NOTE:+/-1.96*sqrt(var)
                           % corresponds to approximatly the 95% ci
symetri=[xm-xmci(1) xmci(2)-xm]; %check for symmetri
```

BIBLIOGRAPHY

Alves, G. M., & Young, I. R. 2003. On estimating extreme wave heights using combined Geosat, Topex/Poseidon and ERS-1 altimeter data. *Applied Ocean Research*, **25**(4), 167–186.

Bauer, E., Hasselmann, S., Hasselmann, K., & Graber, H. C. 1992. Validation and Assimilation of Seasat Altimeter Wave Heights Using the WAM Wave Model. *Journal of Geophysical Research*, **97**(C8), 12671–12682.

Biastoch, A., Reason, C. J. C., Lutjeharms, J. R. E., & Boebel, O. 1999. The importance of flow in the Mozambique Channel to seasonality in the greater Agulhas Current system. *Geophysical Research Letters*, **26**(21).

Bouws, E., Draper, L., Shearman, E. D. R., Laing, A. K., Feit, D., Mass, W., Eide, L. I., Francis, P., Carter, D. J. T., & Battjes, J. A. 1998. *Guide to Wave analysis and forecasting. WMO-No. 702*. World Meteorological Organization.

Breivik, Ø., Gusdal, Y., Furevik, B. R., Aarnes, O. J., & Reistad, M. 2009. Nearshore wave forecasting and hindcasting by dynamical and statistical downscaling. *Journal of Marine Systems*, **78**, 235–243.

Caires, S., & Sterl, A. 2003a. On the estimation of return values of significant wave height data from the reanalysis of the European Centre for Medium-Range Weather Forecasts. *Safety and Reliability: Proceedings of the ESREL 2003 Conference, Maastricht, the Netherlands, 15-18 June 2003*, 353.

Caires, S., & Sterl, A. 2003b. Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research-Oceans*, **108**(C3).

Caires, S., & Sterl, A. 2005. 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *Journal of Climate*, **18**(7), 1032–1048.

Caires, S., Sterl, A., Bidlot, J. R., Graham, N., & Swail, V. 2004. Intercomparison of different wind-wave reanalyses. *Journal of Climate*, **17**(10), 1893–1913.

Caires, S., Sterl, A., & Gommenginger, C. P. 2005. Global ocean mean wave period data: Validation and description. *Journal of Geophysical Research-Oceans*, **110**(C2).

Caires, S., Sterl, A., Komen, G., & Swail, V. 2006a. *Global Wave Climatology Atlas, derived from 45-years of ECMWF reanalysis data*. Website. accessed 11.05.2010: www.knmi.nl/waveatlas/.

Caires, S., Swail, V. R., & Wang, X. L. L. 2006b. Projection and analysis of extreme wave climate. *Journal of Climate*, **19**(21), 5581–5605. Caires, Sofia Swail, Val R. Wang, Xiaolan L.

Chang-Seng, D. S., & Jury, M. R. 2010. Tropical cyclones in the SW Indian Ocean. Part 2: structure and impacts at the event scale. *Meteorology and Atmospheric Physics*, 1–16.

Chemane, D., Motta, H., & Achimo, M. 1997. Vulnerability of coastal resources to climate changes in Mozambique: a call for integrated coastal zone management. *Ocean and Coastal Management*, **37**(1), 63–83.

Coles, S. 2001. *An introduction to statistical modeling of extreme values*. Springer Verlag.

Cooper, C. K., & Forristall, G. Z. 1997. The use of satellite altimeter data to estimate the extreme wave climate. *Journal of Atmospheric and Oceanic Technology*, **14**(2), 254–266.

Coughanowr, C. A., Ngoile, M. N., & Linden, O. 1995. Coastal zone management in Eastern Africa including the island states: A review of issues and initiatives. *Ambio, Research and Capacity Building for Sustainable Coastal Management*, **24**(7/8), 448–457.

De Young, C. 2006. *Review of the state of world marine capture fisheries management: Indian Ocean*. Food and Agriculture Organization of the United Nations. Fisheries Technical Paper.

Dong, S., & Takayama, T. 2002. Improved Least Square Method for Selecting Design Wave Height. *Proceedings of The Twelfth (2002) International Offshore and Polar Engineering Conference*.

ECMWF. 2006. *The ECMWF operational analysis and forecasting system*. Website. accessed 20.01.2010: http://tinyurl.com/yergc4k.

Fisher, R. A., & Tippett, L. H. C. 1928. Limiting forms of the frequency distribution of the smallest and the largest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**.

GEBCO. 2003. *Centenary edition of the GEBCO digital atlas. Published on CD-ROM on behalf of the Intergovernmental Oceanographic Commission (IOC) and the International Hydrographic Organization (IHO) as part of the General Bathymetric Chart of the Oceans (GEBCO)*. British Oceanographic Data Centre, Liverpool.

Gencay, R., Selcuk, F., & Ulugulyagc, A. 2002. EVIM: A software package for extreme value analysis in Matlab. *Studies in Nonlinear Dynamics and Econometrics*, **5**(3), 213–239.

Gilleland, E., Katz, R., & Greg, Y. 2009. *extRemes: Extreme value toolkit.* R package version 1.60.

Gnedenko, B. 1943. Sur la distribution limite du terme maximum dune serie aleatoire. *Annals of Mathematics*, **44**(3), 423–453.

Groen, P., & Dorrstein, R. 1976. Opstellen op Oceanografisch en Maritiem Meteorologisch Gebied. *Koninklijk Nederlands Meteorologisch Instituut (KNMI)*, **11**.

Harlander, U., Ridderinkhof, H., Schouten, M. W., & de Ruijter, W. P. M. 2009. Long-term observations of transport, eddies, and Rossby waves in the Mozambique Channel. *Journal of Geophysical Research*, **114**.

Harris, R. I. 2004. Extreme value analysis of epoch maxima–convergence, and choice of asymptote. *Journal of Wind Engineering and Industrial Aerodynamics*, **92**(11), 897–918.

Harris, R. I. 2005. Generalised Pareto methods for wind extremes. Useful tool or mathematical mirage? *Journal of Wind Engineering and Industrial Aerodynamics*, **93**(5), 341–360.

Hart, R. E., & Evans, J. L. 2001. A Climatology of the Extratropical Transition of Atlantic Tropical Cyclones. *Journal of Climate*, **14**(4), 546–564.

Hasselmann, S., Gunter, s., & Janssen, P. 1992. The WAM Model cycle 4 (revised version). *Technical Report No. 4*.

Hogg, R. V., & Tanis, E. A. 2006. *Probability and statistical inference*. Upper Saddle River, N.J.: Pearson/Prentice Hall.

Hoguane, M. A. 2007. Diagnosis of Mozambique Costal Zone. *Journal of Integrated Coastal Zone Management*, **7**(1), 68–82.

Hoskins, B. J., & Hodges, K. I. 2005. A new perspective on Southern Hemisphere storm tracks. *Journal of Climate*, **18**(20), 4108–4129.

Janssen, P., Lionello, P., & Zambresky, L. 1989. On the interaction of wind and waves. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **329**(1604), 289–301.

Janssen, P., Hansen, B., & Bidlot, J.R. 1997. Verification of the ECMWF wave forecasting system against buoy and altimeter data. *Weather and forecasting*, **12**(4).

Jenkinson, A. F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, **81**, 158–171.

Jury, M. R. 1993. A preliminary study of climatological associations and characteristics of tropical cyclones in the SW Indian Ocean. *Meteorology and Atmospheric Physics*, **51**(1), 101–115.

Komen, G. J., Cavaleri, L., Donelan, M., Hasselmann, S., & Janssen, P. 1996. *Dynamics and modelling of ocean waves*. Cambridge Univ Pr.

Krogstad, H. E., & Barstow, S. F. 1999. Satellite wave measurements for coastal engineering applications. *Coastal Engineering*, **37**(3-4), 283–307.

Lutjeharms, J. R. E. 2004. The Coastal Oceans of South Eastern Africa. *The Sea*, **14**.

Meeker, W. Q., & Escobar, L. A. 1995. Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, **49**(1).

Michaelis, G. 1923. *Die Wasserbewegung an der Oberflache des Indischen Ozeans im Januar und Juli*. Mittler.

Naess, A., & Gaidai, O. 2009. Estimation of extreme values from sampled time series. *Structural Safety*, **31**(4), 325–334.

Neelamani, S., AlSalem, K., & Rakha, K. 2007. Extreme waves for Kuwaiti territorial waters. *Ocean Engineering*, **34**(10), 1496–1504. Neelamani, S. Al-Salem, K. Rakha, K.

NGDC, & NOAA. revised 2009. *National Geophysical Data Center and National Oceanic and Atmospheric Administration, U.S. Dept. of Commerce. ETOPO2v2 Global Gridded 2-minute Database*. Website. accessed 20.01.2010: http://tinyurl.com/yc8aqy4.

Palutikof, J. P., Brabson, B. B., Lister, D. H., & Adcock, S. T. 1999. A review of methods to calculate extreme wind speeds. *Meteorological Applications*, **6**(2), 119–132. 10.1017/S1350482799001103.

Pickands, J. L. 1975. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, **3**(1), 119–131.

Pond, S., & Pickard, G. L. 1983. *Introductory Dynamical Oceanography, 329 pp*. Elsevier, New York.

Queffeulou, P. 2004. Long-term validation of wave height measurements from altimeters. *Marine Geodesy*, **27**(3), 495–510.

Queffeulou, P., & Croize, F. D. 2009. *IFREMER Online Global Altimeter SWH data set documentation*. http://tinyurl.com/yaevjoo.

R, Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Romeiser, R. 1993. Global validation of the wave model WAM over a one-year period using GEOSAT wave height data. *Journal of Geophysical Research*, **98**(C3), 4713–4726.

Saetre, R. 1985. Surface currents in the Mozambique Channel. *Deep-Sea Research Part a-Oceanographic Research Papers*, **32**(12), 1457–1467.

Saetre, R., & daSilva, A. J. 1984. The circulation of the Mozambique Channel. *Deep-Sea Research Part a-Oceanographic Research Papers*, **31**(5), 485–508.

Saetre, R., & Silva, R. P. 1979. *The marine fish resources of Mozambique*. Institute of Marine Research.

Schreier, M., Mannstein, H., Eyring, V., & Bovensmann, H. 2007. Global ship track distribution and radiative forcing from 1 year of AATSR data. *Geophysical Research Letters*, **34**.

Schwiderski, E. W. 1980. On charting global ocean tides. *Reviews of Geophysics and Space Physics*, **18**(1), 243–268.

Shaumann, E. H. 1998. The Coastal Ocean off Southeast Africa, Including Madagascar. *The Sea*, **11**.

Soares, C. G., & Scotto, M. 2001. Modelling uncertainty in long-term predictions of significant wave height. *Ocean Engineering*, **28**(3), 329–342.

Spillane, K. T., & Dexter, P. E. 1976. Design waves and wind in the Australian tropics. *Australian Meteorological Magazine*, **24**, 37–58.

Stanton, B. R. 1984. Return wave heights off South Uist estimated from seven years of data. *Institute of Oceanographic Sciences*, **164**.

Sverdrup, H. U., Johnson, M. W., & Fleming, R. H. 1942. *The oceans: their physics, chemistry, and general biology*. Englewood Cliffs, N.J.: Prentice-Hall.

The WAMDI group; Hasselmann, S., Hasselmann, K., Bauer, E., Janssen, P., Komen, G. J., Bertotti, L., Lionello, P., Guillaume, A., Cardone, V. C., Greenwood, J. A., & Reistad, M. 1988. The WAM model, A third generation ocean wave prediction model. *Journal of Physical Oceanography*, **18**(12), 1775–1810.

Thompson, P., Cai, Y., Reeve, D., & Stander, J. 2009. Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, **56**(10), 1013–1021.

Uppala, S. M., Kaallberg, P. W., Hernandez, A., Saarinen, S., Fiorino, M., Xu, L., Onogi, K., Sokka, N., Andrae, U., & Da Costa Bechtold, V. 2004. ERA-40: ECMWF 45-year reanalysis of the global atmosphere and surface conditions 1957 to 2002. *ECMWF Newsletter*, **101**(Summer/Autumn), 2–21.

Uppala, S. M., Kaallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, Paem, Jenne, R., McNally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., & Woollen, J. 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**(612).

Williams, F. R., Renard, R. J., Jung, G. H., Tomkins, R. D., & Picard, R. R. 1984. *Forecasters Handbook for the Southern African Continent and Atlantic/Indian Ocean Transit*.

Wright, J., Colling, A., & Park, D. 1999. *Waves, tides and shallow-water processes*. Oxford: Butterworth-Heinemann, in association with the Open University. 2nd ed.

Young, I.R. 1999. Seasonal variability of the global ocean wind and wave climate. *International Journal of Climatology*, **19**(9), 931–950. 10.1002/(SICI)1097-0088(199907)19:9<931::AID-JOC412>3.0.CO;2-O.

Zacarias, C. 2009. *Mozambique 4th License Round*. Online Presentation, accessed 4.02.2010: http://tinyurl.com/ygan9we. Institute of National Petroleum (INP), Mozambique.