On the Probability and Averaged Shape of Extreme Water Waves

Tianning Tang

Somerville College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2020

On the Probability and Averaged Shape of Extreme Water Waves

A thesis submitted for the degree of
Doctor of Philosophy at the University of Oxford

By Tianning Tang, Somerville College, Michaelmas 2020

Extreme waves are reported more frequently than the probability predicted from standard statistical models in the open ocean. This departure from the classical statistical distributions is primarily due to nonlinear physics. Wave nonlinearity also modifies the most probable shape of extreme water waves as these waves propagate relative to that expected in a linear model. Better insights into both aspects of extreme water waves are crucial for the safety design of carriers and marine structures.

The horizontal asymmetry and contraction of the averaged shape of extreme events can be observed in naturally occurring water waves. Under experimental conditions, horizontal asymmetry is a somewhat transient behaviour whereas contraction seems to be permanent. We also observe limited finite water effect on these nonlinear modifications, particularly for extremely steep ones.

The probability of extreme events can be measured by the kurtosis of the free surface as a convenient proxy. The analytical derivations for how this evolves in wave flumes are found to be accurate for low steepness. For high steepness cases, the analytical models miss the transient maxima and over-predict the kurtosis at the steady state. This is primarily because the kurtosis is dependent on the spectral bandwidth in such situations. The analytical models underestimate the steady state kurtosis at finite water depth, particularly for the critical water depth.

Data driven methods are applied to predict the space-time probability distribution of extreme events. Two practical examples with second order wave fields indicate the substantial potential of data driven methods in predicting the probability of extreme events. These benefits are particularly clear as the complexity of the problem increases.

In this study, we closely examine the averaged shape and the probability of extreme water waves. Nonlinear modifications on both aspects lead to significant departure from the linear theory. These nonlinear modifications challenge the safety of the engineering design.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Thomas Adcock. Throughout my research, he has provided thoughtful advice and practical suggestions. He has also offered substantial encouragement with considerable patience during the hard times. He has also provided a significant number of opportunities for me to explore life in academia. These experiences become the starting point of my early career. I have benefited tremendously from experiences in the group, and I appreciate his guidance.

I want to thank Prof. Paul Taylor from the University of Australia, and Prof. Ton van den Bremer for their thoughtful suggestions during my research, particularly for the advice on my first two journal publications. I am grateful to Dr Peter Tromans for providing the Lake George data and useful engineering insights. I would also like to thank Prof. Alistair Borthwick for providing suggestions on experiments and journal publications during my stay in China.

I would also like to thank Prof. Ye Li for his support on the experiments and the thoughtful advise on my academic career.. I want to thank the laboratory team at Shanghai Jiaotong University, especially Wentao Xu, Zhiyong Yan, Huimei Xie.

I want to thank my colleagues Dylan Barratt, Yan Li, Wenhua Zhao, Xiaochen Li, Mark McAllister, Binzhen Zhou, for their support on my research. I would also like to thank my friends Xiaodong Shi, Qian Ma, Yin Liang, Tuo Wang, Yiqing Liu, Ting Zhong, Zhaoxian Lin, Kunwei Wu, Jingyi Yang, Bohan Chen for their support during my stay at Oxford.

My special thanks go to my fiancée Zeyu Li for her love and constant support during my studies. My greatest thanks go to my mother Min Bao, my father Wenwei Tang and my cousin Ruizhe Zhao for their encouragement and support throughout my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The ocean environment

Many mariners sporadically report a quasi-mythical event, during which a giant wave with a significantly larger amplitude than the background sea-state occurs. Reports of these extreme ocean surface waves, which have been referred to as 'wall of water', can be traced back to the beginning of the 20th century. On 29 August 1916 at about 4:40 p.m., United States Navy cruiser USS Memphis was damaged in Santo Domingo harbour in the Dominican Republic due to large waves of up to 21 meters in height. This accident caused 40 deaths and in addition 204 people on board got injured. Meanwhile, United States Navy gunboat USS Castine got nearly wrecked. The extreme waves associated with the two accidents are believed to have been generated by passing hurricanes (Smith 2006). There are much more documented reports on huge ocean waves from sailors' testimonies and evidence of damages on ships in the last ten decades. A relevant example is a six-year-old barge carrier named MS München that sank at mid-Atlantic after sending an emergency message at 3 a.m. on 12 December 1978. Later on people found twisted attachment pins of an unreleased lifeboat that were then stored at 20 meters above the water line of the said carrier.

This became a piece of evidence that the carrier had experienced violent force caused by extremely large waves. Moreover, there are more than 22 super-carriers that are believed to be damaged or lost due to gigantic ocean waves between 1969 and 1994 (Kharif & Pelinovsky 2003).

The aforementioned examples suggest that extremely large waves are especially dangerous to human activities in the ocean, a reason why they were called the 'monster' in the open ocean. For example, the master of Queen Elizabeth 2 described these large wave as "came out of the darkness" and "looked like the White Cliffs of Dover" after witnessing a 29 metres rogue wave during Hurricane Luis in the North Atlantic. The ocean large waves reported usually have the characteristics as follows:

(i) a crest with significantly larger elevation when compared to the crest in front,

(ii) a peak crest with a fairly steep front, which looks like a 'wall of water',

(iii) a deep trough like a hole appears either in the front or at the back of the huge crest.

Although there have been many reports of waves having such characteristics in the open ocean for decades, these observations have limited scientific use beyond providing evidence that our understanding of the ocean is imperfect (Adcock & Taylor 2014). The limited scientific use of anecdotal reporting is partly due to the poor visibility at these extreme events during a storm. The objectivity of the observer can also affect the description of these waves. The first scientific evidence of the existence of rogue waves is commonly held to be obtained at Draupner platform (Haver 2004) in 1995. Since this date, an increasing number of scientific efforts have been made in the field of extreme waves (see review by Dysthe et al. (2008)) and these extremely large waves are referred to as rogue, abnormal, or freak waves by scientists.

After the observation of the famous Draupner wave, more rogue wave events have been captured by scientific measurement devices, which has further confirmed the

existence of rogue waves. In 2004, during Hurricane Ivan, pressure sensors from Naval Research Laboratory, which was installed on the ocean floor at the Gulf of Mexico, detected a freak wave of 27.7 meters. During Hurricane Dorian, an offshore buoy recorded several extreme waves with the largest waves reaching 30 meters on 8 September 2019, in the Cabot Strait off Channel-Port aux Basques, Newfoundland.

1.2 Offshore engineering design challenges

The presence of abnormal waves raises concern on the safety of the offshore structures at the design stage. For example, there will be serious implications if a rogue wave is large enough to hit the underside of a deck of an oil platform. To avoid possible damage, the designer of the fixed oil platform needs to leave enough air gap between the mean sea level and the deck, which requires an accurate estimate of the size and return period of these rare waves. Additionally, due to the global energy crisis, renewable energies such as offshore wind energy play an increasingly important role in energy generation. The design of offshore wind turbines also involves consideration of wave load on the tower structure and the avoidance of direct impact on wind turbine blades from rogue waves. In engineering practice, the short term crest distributions are usually estimated based on sea state parameters obtained through field measurements or wave hindcast data. Hence, an accurate wave statistical model is essential to predicting the probability of extreme events based on the extreme sea states during storms. The performance of these models will directly affect the cost and the safety of offshore structures.

Wave statistical models based on single-point measurements are widely used for the design of offshore structures. Nevertheless, as those models are found to underestimate the size of realistic waves, it would thus lead to an underestimated prediction of wave loads on structures, posing risks on offshore structures at service. Especially,

it is found the waves measured within an area can lead to much larger maximum crests than those at a single point (Forristall 2006). This addresses the need of taking into account the size of an area for obtaining more reasonable wave models in practices in offshore engineering.

Apart from the direct damages that can be caused by wave impact loads in an extreme wave event, the designer also needs to consider extreme and fatigue loads caused by wave-structure interaction. For example, for the design of offshore platforms, the horizontal wave force on offshore structures is of great importance as it is closely related to the base shear and the overturning moment. If the Keulegan–Carpenter number is large, the horizontal force on the support structures is dominated by the drag force, and its magnitude depends on the square of the wave particle velocity. Another example of this is offshore wind turbines with a dynamic response such as the 'ringing' effect, which requires better perception on the kinematics of these extreme events (Grue 2011, Zang et al. 2010). The predictions of the shape of extreme events are also critical to shipping and floating production storage and offloading vessel when calculating the response of the ship and the wave slamming effects under extreme conditions. These types of calculation require further understanding of the kinematics of these extreme events. The shape of these large events becomes essential for these types of calculations, as wave kinematics are closely associated with the wave shape.

Model tests play an important role in the design of offshore and coastal structures. It provides a feasible way to obtain short term wave variability in natural occurring wave fields. During model tests, irregular or random waves are generated by wave paddles based on a specified wave spectrum with random phase to each Fourier component. However, particularly for unidirectional waves, nonlinear interactions lead to both correlations and energy transfer between different components, which increase the rogue wave density (i.e. the number of rogue waves during given amount of time)

and also modify the average shape of extreme waves from that predicted from the linear theory (Janssen 2003). It is important for offshore and coastal structures designers to understand this departure from linear physics. For instance, if the designer generates a target sea state at the wavemaker, the loads measured on the object will be over-estimated if the nonlinear effects are not taken into account. Additionally, these nonlinear effects also lead to the change in the averaged shape of extreme events, which also has an impact on extreme force estimations.

1.3 Aspects of rogue waves

Further insight into the formation mechanisms of rogue waves in the open ocean has the potential to lead to improvements in design guidelines for more reliable offshore structures and carriers. In this study, we will focus on three main aspects of rouge waves, which correspond to the offshore engineering design challenges addressed in section 1.2:

(i) the averaged shape of rogue wave events,

(ii) spatial evolution of the rogue wave density of steep unidirectional random waves,

(iii) wave statistics model for maximum crest distribution over a single point and an area.

With the detailed research in these three main aspects of the rogue wave events, we aim to provide better understanding of the physics underlying the formation of rogue waves and also offer useful guidelines that would potentially improve offshore engineering designs.

Chapter	Methods	Probability/Shape	Wave type
3. Field Measurements...	Field data	Shape	• Omnidirectional • Deep water
4. Comparison of MNLS models...	Numerical simulation	Probability & Shape	• Unidirectional • Deep water
5. Long distance evolution...	Numerical simulation, Experiments	Probability & Shape	• Unidirectional • Deep water
6. Finite water depth effect...	Numerical simulation	Probability & Shape	• Unidirectional • Finite water depth
7. Data driven analysis...	Numerical simulation, Data driven	Probability	• Directional spread • Finite water depth

Table 1.1: Overview of different methodologies applied in this study.

1.4 Methodology

In order to examine the three aspects mentioned in the previous section, we have applied varied methods to provide a more comprehensive understanding of the physics behind extreme events in the open ocean. The detailed methods for each chapter are summarised in Table 1.1.

1.4.1 Field data analysis

Field data provide direct information about surface waves in the real ocean. Analysing field data is of great importance in this study as it shows the behaviour of waves in nature. The relevant physics in typical open ocean environment might include the wind-wave interactions, nonlinear wave evolution, wave-current interaction, wave

breaking, topography, diffraction and radiation etc. All the characteristics of surface waves observed from field measurements are the consequence of multiple physical processes.

However, the limitation of field measurements is apparent when scientists are attempting to understand the physics behind rogue wave events. The combined effects from different physical processes on wave behaviour raise challenges in identifying the impact of the individual element. Due to high cost of installing long term *in situ* measurement devices in the open ocean, the sampling frequency from field data is often insufficient for detailed spectral analysis. Rigorous quality control processes are also required to remove wave recordings with nonphysical spikes or measurement errors. It is not possible to control the sea state in the open ocean, which leads to insufficient data for relative steep sea state during storms and rogue wave events. Additionally, the radical change of sea state during storms also leads to difficulties for statistical analysis. Currently, most of the high-quality measurement devices are installed on offshore platforms, which might include reflected wave from the supporting structures. In this study, we analysed two datasets from Lake George, Australia and the North sea. We focused our study on the average shape of the largest events at these locations.

Alternative instrumentations like radars, laser scanners, and stereo cameras have been utilised for field measurements, which can provide more accurate description of the wave field. However, due to the high initial costs and difficulties in calibrations, the applications of these instrumentations in field measurements are limited (see some prior work Bergamasco et al. (2017), Alvise et al. (2017)).

1.4.2 Wave experiments

Experiments are essential in this study for obtaining a better understanding of the nonlinear physics in the evolution of surface waves. For instance, the high-quality

measured data from capacitance or resistance wave probes in experiments can provide a detailed description of the wave elevation profile. Experiments also enable scientists to generate a repeatable steep sea state or even an isolated rogue wave event, which allows for examination of targeted nonlinear processes. However, experiments also suffer from clear limitations, most of which are due to physical constraints. For example, for relatively small wave tanks, the friction from the sidewall and the bottom of the wave tank and the viscosity will have an impact on the evolution of the wave trains.

Additionally, the physical constraints of the flume size also limit experimental investigation on wave evolution over very long distances or a large area with directional spread waves without reflections. It is challenging to carry out wave statistical analysis in the wave flume partially because of wave reflections and also the expense of performing experiments. There are also additional error sources in the wave generation part for the experimental campaign. Second order error waves and evanescent waves can affect the wave profile near the wave paddle. Finally, it is challenging to obtain the spatial profile of waves along the wave flume, which leads to difficulties to obtain a complete picture of wave evolution along the flume.

In this study, we performed experiments on the spatial evolution of wave kurtosis at Shanghai Jiaotong University to further understand the impact of modulational instability on the probability of extreme events.

1.4.3 Numerical simulation

Numerical simulations are always excellent complements to experimental results. Recreating experimental results in a numerical tank can be challenging, but numerical simulation results can provide a better description of spatial wave profile. Numerical simulation can also help researchers to understand the physical process better by ruling out some of the other effects. For example, the success on recreating experimental

results with a numerical code with inviscid assumption indicates the viscous effects during the experiments is negligible.

Furthermore, numerical wave tanks do not have any physical constraints. With sufficient computational resources, wave evolution over very long distances or over a large area with directional spread waves can be achieved. Since computational power is usually far less expensive than experiment costs, a large number of realisations can be achieved with more stable statistical results.

However, one must be careful with numerical results partially because of the underlying assumption of these numerical methods. Improper initial or boundary conditions can also lead to very different results. Hence, detailed verification and validation of numerical results are essential.

In this study, we used fully nonlinear potential flow solver, Modified Nonlinear Schrödinger equation, and enveloped based second order simulation tool to investigate the spatial evolution of wave kurtosis in deep water and at finite water depth and the maximum crest distribution over an area.

1.4.4 Data driven method

Advances in the understanding of extreme events and wave statistics are associated with data from field measurements, experiments and numerical simulations. Unsurprisingly, data driven methods gain in popularity in fluid mechanics community recently because of their excellent capability of information extraction and generalisation from datasets. Additionally, the advances in high-performance computing and new machine learning algorithms also accelerate the wide spread of data driven methods for both research and engineering purposes (Brunton et al. 2020, Sapsis 2020*b*).

For direct modelling, data driven models require substantial amount of computation power for complex problems, and their performance is also directly related to

the quality of training dataset. However, when they are fully trained, these models can make predictions very fast and require almost no computation power. Hence, these models have huge potential in quick-response onsite wave predictions. Recent applications of data driven method with reduced order modelling (see Wan et al. (2018), Farazmand & Sapsis (2017)) and dynamic mode decomposition (see Williams et al. (2014), Takeishi et al. (2017)) also show great potential for modelling complex problems, which may also provide some insights into wave physics.

Furthermore, recent advances in active sequential sampling (Mohamad & Sapsis 2018, Sapsis 2020*a*, Ragone & Bouchet 2019, Gramstad et al. 2020) and probabilistic decomposition (Mohamad & Sapsis 2016, Farazmand & Sapsis 2017, Mohamad et al. 2016) also show the great penitential of data driven methods in reducing the computational cost for extreme statistics.

In this study, data driven methods are applied for predicting maximum crest height distributions based on second order simulations. The preliminary results show the huge potential of data driven methods in predicting wave statistics under complex situations.

1.5 Overview

This thesis follows the integrated thesis format. Chapters 3 to 7 are taken directly from published or submitted papers. The symbols are uniquely defined within each chapter and there are some repetitive definitions between different chapters.

Chapter 2 first reviews the basic concepts commonly used for describing the sea states and rogue wave events. This chapter further details the mathematical equations used to model the surface water wave problems and introduces the critical nonlinear physics that is likely associated with the formation of rogue wave events in the open ocean. We also introduce the fully nonlinear potential flow solver used in this study

at the end of chapter 2.

Chapter 3 explores the averaged shape of the extreme events in naturally occurring water waves based on two field datasets from Lake George, Australia and the North Sea. This project confirms two crucial features of the averaged shape of extreme events predicted by the nonlinear theories in the open ocean: largest wave tends to move to the front of the wave group, and the wave group tends to contract in the mean wave direction relative to the linear theory. This chapter has been published as Tang, Tromans & Adcock (2019). Analysis of an additional data set was presented at OMAE Tang, Yelland & Adcock (2019). This additional study was consistent with that presented here but contained no new conclusions and issues were identified with the data used. These additional results have therefore not been included in this thesis.

Chapter 4 examines the performance of two numerical simulation tools used in this study. OceanWave3D solves fully nonlinear potential flow equations, which provide accurate predictions on surface waves profiles. We also consider two versions of the modified nonlinear Schrödinger equation, which are compared against the results of OceanWave3D in both wave group profiles and random wave statistics. This chapter was presented at the OMAE conference Tang, Li, Bingham & Adcock (2020).

In chapter 5, we study the evolution of unidirectional deep water waves with uncorrelated Fourier components as inputs. We examine the evolution of kurtosis as a convenient measure of the probability of a rogue wave with experiments. We also extend the experimental results in parameter space with two numerical schemes verified in chapter 4. Both the experimental and numerical results are compared against the existing analytical models. We also examine how nonlinear physics modifies the average shape of extreme events by comparing the simulated averaged shape of extreme events against the shape predicted by the linear theory. This chapter has been published as Tang, Xu, Barratt, Bingham, Li, Taylor, van den Bremer & Adcock

(2020). This paper has numerous authors; however, the experiments as well as both parts of the numerical analysis were carried out by the author.

We further extend our results in chapter 5 to finite water depth in chapter 6. The impact from finite water depth on the spatial evolution of kurtosis and the average shape of extreme events is examined with fully nonlinear potential flow simulations. We also compare the simulation results against the theoretical predictions to provide proper guidance to model tests at finite water depth. A paper based on this chapter has been submitted to a journal jointly authored with the supervisor.

In chapter 7, we study the maximum crest distribution over an area. We propose data driven methods as an engineering solution for predicting maximum crest distribution over an area. We examine the performance of existing theoretical models and data driven methods by comparing the predictions against second order simulation results. A paper based on this chapter has been submitted to a journal jointly authored with the supervisor.

In chapter 8, we summarise our main findings of this work and briefly discuss the limitation of this study. We also propose possible research topics to be explored in the future.

Chapter 2

Literature Review

In the chapter below, we first present some widely used basic concepts, which are fundamental to the understanding of water waves and are the basis for 'classic' design of offshore structures. For simplicity this chapter focuses on 'deep water'. In this study, deep water is defined as the condition where the relative water depth $k_0 d$ is greater than 3 (where k_0 is the dominant wave-number and d is the water depth). We further introduce the basic mathematical equations used to model the wave problem and also the key nonlinear physics associated with rogue wave events in the open ocean. Finally, the fully nonlinear potential solver is introduced at the end of this chapter.

2.1 Basic concepts

2.1.1 Wave spectrum

The simplest assumption in wave analysis is that the waves are actually formed by the superposition of different sinusoidal waves at different frequencies and directions. A wave spectrum can directly show the energy distribution of a wave record at different frequencies and directions, which provides the fundamental description of the time

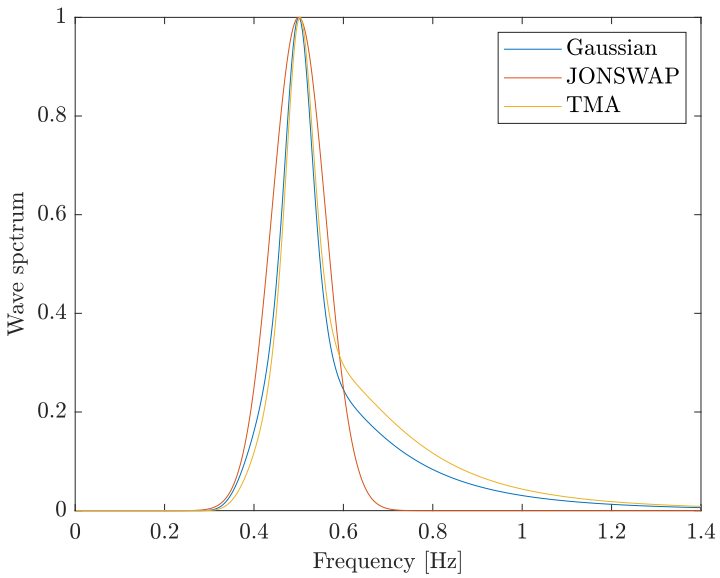

Figure 2.1: Example of different wave spectrum with peak period of 2 seconds, and unity wave spectral amplitude.

series in the Fourier domain.

There are many standard formulae for the shapes of unidirectional wave spectra for different conditions. An example of different wave spectra are shown in Figure 2.1.

Gaussian spectrum

One of the most commonly used wave spectrum is the Gaussian spectrum. The Gaussian input spectrum is defined as:

$$S_{\text{Gaussian}}(f) = \left(\frac{H_s}{4}\right)^2 \frac{1}{\Delta f\sqrt{2\pi}} \exp\left(-(f - f_0)^2/(2\Delta f^2)\right), \tag{2.1}$$

where f_0 is the peak frequency, H_s is the significant wave height given in Equation 2.9, and Δf controls the bandwidth of the spectra. Gaussian spectrum is symmetric about the peak frequency, and with a relatively short spectral tail. Hence, the bandwidth parameters for Gaussian spectrum are more stable as these parameters are sensitive to the high frequency spectral tails (Prasada Rao 1988, Serio et al. 2005). Gaussian

spectrum is also widely used in theoretical derivations due to its simplicity (Fedele et al. 2010, Janssen & Janssen 2019).

JONSWAP spectrum

The JONSWAP spectrum is associated with fetch limited conditions in deep water. This was first reported by Hasselmann et al. (1973) in the Joint North Sea Wave Observation Project, where the waves in open ocean were found to be hardly ever in an equilibrium state. Instead, the waves tend to evolve through non-linear, wave-wave interactions for a long time and distance. Hence, a new spectrum formula with additional artificial factors was suggested as:

$$S_{\mathrm{JONSWAP}}(f) = G(f)\alpha_{PM} g^2 (2\pi)^{-4} f^{-5} \exp\left(5/4\left(\frac{f_0}{f}\right)^4\right), \tag{2.2}$$

where

$$G(f) = \gamma^{\exp\left(\frac{(f-f_0)^2}{2\sigma_{\mathrm{J}}^2 f_0^2}\right)}, \tag{2.3}$$

where α_{PM} is the coefficient from Pierson-Moskowitz spectra, the value of which depends on the wind conditions, γ is the peak enhancement factor, and σ_{J} is defined as:

$$\sigma_{\mathrm{J}} = \begin{cases} 0.07 & f \leq f_0 \\ 0.09 & f > f_0. \end{cases} \tag{2.4}$$

The extra parameters provide extra flexibility. Hence, JONSWAP spectrum can be customised for generating random waves based on specified requirements in both numerical simulations and flume experiments. Additionally, the peak enhancement factor γ provides an estimate of how narrow-banded the spectrum is (Serio et al. 2005).

TMA spectrum

In contrast to deep water, the spectrum of waves in finite water depth does not maintain its shape along the fetch. The high frequency tail becomes less steep and evolves from f^{-5} in deep water towards f^{-3} when wave travel along the fetch in finite water depth (Knowles 1982). Hence, Bouws et al. (1985) proposed a low frequency cut off transfer function, which can be applied to JONSWAP spectrum, to describe the wave behaviour in finite water depth, which is given as:

$$\phi_{\mathrm{TMA}}(f,d) = \frac{1}{2n_r}\tanh^2(k_0 d), \tag{2.5}$$

where n_r is the ratio of group velocity over phase velocity, k_0 is the dominant wave number and d is the water depth. The TMA spectrum is given as:

$$S_{\mathrm{TMA}}(f) = \phi_{\mathrm{TMA}}(f,d) S_{\mathrm{JONSWAP}}(f) \tag{2.6}$$

The shape of this spectral formula is verified against observations in Young & Verhagen (1996) with excellent agreement.

Directional spectrum

As opposed to the unidirectional spectrum, the directional spectrum $S_{dir}(f,\theta_s)$ contains the information of energy distribution not only across different frequencies but also in different directions. To indicate the energy distribution among different directions, a normalized directional distribution function $D(\theta_s)$ is added to the unidirectional spectrum $Sp(f)$:

$$S_{dir}(f,\theta_s) = D(\theta_s) S_p(f). \tag{2.7}$$

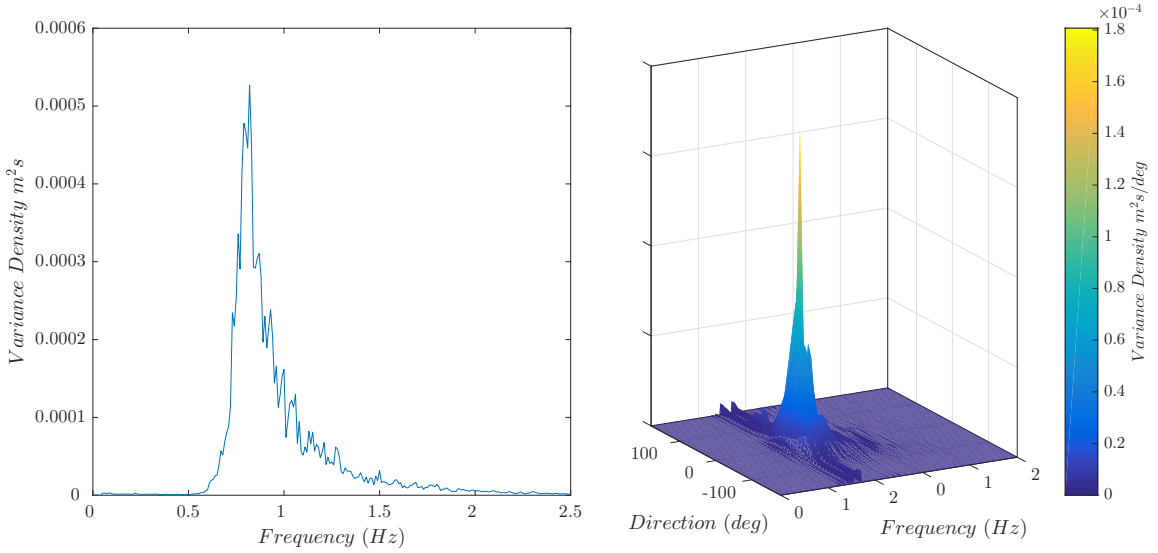

Figure 2.2: Comparison between omnidirectional spectrum (left) and directional spectrum (right) from the Lake George dataset.

The directional spectrum gives important additional information about the sea state.

An example of omnidirectional and directional sprectrum is shown in Figure 2.2.

2.1.2 Statistical parameters

Spectral moments

Many useful definitions exist in terms of the moments of the wave spectrum $S_p(f)$.

The n^{th} spectral moment is defined as:

$$m_n = \int_0^\infty f^n S_p(f)\ \mathrm{d}f. \tag{2.8}$$

Significant wave height

In modern practice, the significant wave height H_s is usually defined by the zeroth moment of the wave spectrum m_0 as:

$$H_s = 4\sqrt{m_0} = 4 \cdot std, \tag{2.9}$$

where *std* is the standard deviation of wave time series about the mean. For a sea state dominated by linear dynamics, the significant wave height is approximately the average height of the third largest waves (see Tucker & Pitt (2001) for a history of this important parameter).

Spectral bandwidth parameters

For standard wave spectra, there is always at least one parameter designated to control the spectral bandwidth. For an arbitrary shaped spectrum, spectral width parameters are broadly used to quantify the bandwidth. Although all the bandwidth parameters can describe the width of the spectrum, their sensitivity to the high frequency tail is different.

Cartwright & Longuet-Higgins (1956) suggested the parameter ε as a measure of bandwidth:

$$\varepsilon = \left(1 - \frac{m_2^2}{m_0 m_4}\right)^{1/2} \tag{2.10}$$

However, in this formula, the fourth-order moment m_4 is required to calculate the spectral bandwidth, which is sensitive to high frequency noises and cut off frequency. Hence, this bandwidth parameter should be used with great care, especially for field measurements,

Another spectral width parameter proposed by Longuet–Higgins (1975) is defined as:

$$\nu = \left(\frac{m_0 m_2}{m_1^2} - 1\right)^{1/2}. \tag{2.11}$$

Although this parameter avoids calculating higher order spectral bandwidth, recent studies show this parameter is still sensitive to cut off frequency, especially for heavy high-frequency tailed wave spectra, such as JONSWAP spectrum.

Goda (2000) suggested another dimensionless parameter that describes the spec-

tral bandwidth as:

$$Q_p = \frac{2}{{m_0}^2} \int_0^\infty f S^2(f) \mathrm{d}f, \tag{2.12}$$

where $S(f)$ is the variance density spectrum. This parameter Q_p has less sensitivity to the high-frequency tail of the spectrum (and cut-off frequency) than other bandwidth metrics (Serio et al. 2005).

2.1.3 Wave skewness

Skewness, a parameter that describes the asymmetry of the probability distribution is defined by the third moment of the free surface and the zeroth moment of the wave spectrum:

$$skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(\eta_i - \overline{\eta})^3}{m_0^{(3/2)}}. \tag{2.13}$$

Skewness is a useful parameter in measuring the second order contributions to the main wave package, which will be explained in section 2.4.2.

2.1.4 Wave kurtosis

Another important statistical parameter, kurtosis, is defined by the fourth moment of the free surface and the zeroth moment of the wave spectrum as:

$$kur = \frac{\frac{1}{n}\sum_{i=1}^{n}(\eta_i - \overline{\eta})^4}{m_0^2}, \tag{2.14}$$

where n is the sample size of the elevation time series, η_i is the i^{th} elevation data in the time series and $\overline{\eta}$ is the mean water level. According to Janssen (2003), the excess kurtosis is closely related to third order interactions and modulational instabilities.

Kurtosis is defined as:

$$C_4 = \frac{\frac{1}{n}\sum_{i=1}^{n}(\eta_i - \overline{\eta})^4}{3 \cdot m_0^2} - 1. \tag{2.15}$$

As kurtosis is a measure of how heavy the tail of a probability distribution is, in wave statistics, this is mostly associated with extreme deviations from the mean surface i.e. rogue events. Hence, a wave record with higher kurtosis means that it is more likely to observe rogue waves within that record (Mori & Janssen 2006).

2.2 The hydrodynamic problem

To model the wave behaviour analytically, some simplifications are necessary. In this work, water is assumed to be incompressible, irrotational and has no viscosity. A function named velocity potential $\Phi(x, y, z)$ is introduced to approach the hydrodynamic problem, which is defined as a function and the spatial derivatives of which are equal to the velocity vector in the flow field. Given the fact that the ideal water is incompressible and inviscid, mass conservation requires the velocity potential to satisfy Laplace's equation for the whole fluid domain, which gives:

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} = 0. \tag{2.16}$$

A Cartesian coordinate system is established with the origin at the still water surface and the z-axis pointing upwards. $\mathbf{x} = [x, y]$ is the horizontal vector and t is the time. Two boundary conditions are required to define the air and the water interface $\eta(\mathbf{x}, t)$. The first kinematic boundary condition restrains the fluid particles on the surface, which is given by:

$$\left.\frac{\partial \Phi}{\partial z}\right|_{z=\eta} = \frac{\partial \eta}{\partial t} + \frac{\partial \Phi}{\partial x}\frac{\partial \eta}{\partial x} + \frac{\partial \Phi}{\partial y}\frac{\partial \eta}{\partial y} \qquad \text{on } z = \eta. \tag{2.17}$$

The second dynamic boundary condition requires the Bernoulli's equation to be satisfied at the surface. This gives:

$$\frac{\partial \Phi}{\partial t} + \frac{1}{2}|\nabla \Phi|^2 + g\eta = 0 \qquad \text{on } z = \eta, \tag{2.18}$$

where g is the gravitational acceleration. The seabed, which is assumed to be rigid, is impermeable with a profile of $z = -d(x, y)$. This gives the last kinematic boundary condition at the seabed, which requires no flow normal to the seabed.

$$\frac{\partial \Phi}{\partial z} + \nabla d \cdot \nabla \Phi = 0, \qquad \text{on } z = -d, \tag{2.19}$$

where ∇d is the gradient of bottom profile. As the two boundary conditions at surface elevation (Equations 2.17 and 2.18) are nonlinear, they must be solved at the free surface, which is, however, unknown. This gives a significant increase in difficulty for the attempts on solving the governing equation (Equation 2.16) analytically or numerically, especially since the ocean waves are random with directional spreading.

2.3 Linear waves

2.3.1 Linear wave theory

Linear theory is the most basic wave theory originated with the work of Stokes (1847). This theory linearizes the boundary conditions at the free surface by evaluating them at the mean surface level and removing the higher order terms. There are two additional important consequences of these assumptions: the Gaussian sea and the

principles of superposition. Although these are not true for real water, linear theory is still a robust approximation for small amplitude waves in deep water.

A unidirectional regular wave with a single frequency, which satisfies the Laplace equation, can be described as a cosine with amplitude A, frequency ω and phase φ_θ. Therefore, the observed elevation at the time of t and position of x can be described as:

$$\eta_{regular}(x,t) = A\cos(\omega t - kx + \varphi_\theta), \tag{2.20}$$

where the wavenumber k is obtained by the linear dispersion equation:

$$\omega^2 = gk\tanh(kd), \tag{2.21}$$

where g is the gravitational acceleration and d is the water depth.

However, real water waves are not just of a single frequency but have an infinite number of frequency components. The linear theory assumes these waves with different frequencies can propagate freely and can be added linearly. Therefore, the dynamics of each component is governed by Equation 2.21.

Another important assumption from the linear theory is that the different frequency components for real water waves are drawn from a Gaussian process. This gives an expression for the linear surface elevation as

$$\eta_{linear}(\mathbf{x},t) = \sum_{i=1}^{\infty}[a_i\cos(\mathbf{k}_i\mathbf{x} - \omega_i t) + b_i\sin(\mathbf{k}_i\mathbf{x} - \omega_i t)], \tag{2.22}$$

where $\mathbf{k}_i$ is vector wavenumber, $\mathbf{x}$ is position vector, which has the same direction as the wave propagation and its magnitude is given by Equation 2.21. a_i and b_i are the independent random variables drawn from a normal distribution $N(0,\sigma_i^2)$. The variance of each wave component σ_i^2 is defined by the spectrum of the sea-state as

$\sigma_i^2 = Sp(\omega_i)\Delta\omega$.

Equation 2.22 is also widely used for generating random linear time series (Tucker et al. 1984). The amplitude for each component is determined as $A_i = \sqrt{a_i^2 + b_i^2}$. A random phase with a uniform distribution from $(-\pi, \pi)$ is also induced to avoid repetition in a long random time series simulation.

2.3.2 NewWave theory

Based on the assumption of a Gaussian process, Lindgren (1970) and Boccotti (1983) suggest that the most probable shape of the average profiles of extreme events is the scaled auto-correlation function. This concept is then widely used in offshore engineering, which is now known as the NewWave model. In this model, the average shape of a large linear wave with unit amplitude is given by:

$$\eta(t) = \frac{1}{std^2} \int_0^\infty Sp(f) \cos(2\pi f t)\ df, \tag{2.23}$$

where *std* is the standard deviation of the wave record.

Walker et al. (2004) also proposed higher order corrections to this model with narrow band approximations, which will be covered in section 2.4. This model has been validated by many researchers, such as Jonathan & Taylor (1997), where the NewWave model is consistent with the large deep water waves in the North Sea.

The NewWave is closely linked to the related approach of quasi-determinism which is preferred by the oceanographic community. This thesis uses both terms at different points.

2.3.3 Linear wave crest distributions

Another important conclusion from the Gaussian sea assumption is the model for wave height and crest height distributions. The Rayleigh distribution predicts the

wave height and crest height behaviour based on the linear theory. If the wave spectra are assumed to be narrow banded and the phases are uniformly distributed, the probability density function of crests is Longuet-Higgins (1952)

$$P_{\eta_{crest}}(\eta) = \frac{\eta}{m_0} \exp\Big(- \frac{\eta^2}{2m_0} \Big). \tag{2.24}$$

Although the Rayleigh distribution might provide a rather accurate prediction for the wave height of real water waves, it underestimates the largest waves and it is not suitable for extreme experimental conditions (see results in Figure 4.5). However, the Rayleigh distribution provides a base line for measurements to be compared with, which clearly describes how real water waves depart from linear theory.

2.4 Second order waves

2.4.1 Second order interactions

One of the most significant differences between real water waves and the random waves generated based on linear theory is that real water waves tend to have taller and spikier crests and more rounded and flat troughs. This arises from the non-linearity of the water wave problem. The interaction between two wave trains leads to a resultant wave component at the sum of the two wavenumbers ('sum term') and a wave component at the difference of the two wavenumbers ('difference term'). One important characteristic of these terms is that these are bound waves, which means the motions of these waves are governed by the primary components (i.e. they travel with these and not under their own dynamics). Thus the motion of these bound harmonics will not affect the dynamics of the primary wave trains.

There are several methods to determine the second order terms from the linear components, such as the exact solution provided by Dalzell (1999). However, due to

this method being computationally demanding, a narrow banded approximation of second order terms proposed by Walker et al. (2004) is commonly used for processing large datasets in the real ocean.

The approximated second order sum term for irregular waves from linear signal η_L is

$$\eta^{2+}(t) = \frac{S_{22}}{d}(\eta_L^2 - \eta_{LH}^2), \tag{2.25}$$

where η_{LH} is the Hilbert transform of the linear signal and S_{22} is defined as a function of relative water depth as

$$S_{22} = k_0 d \frac{\coth(k_0 d)(1 + 2 \cdot \operatorname{sech}(2k_0 d))}{2(1 - \operatorname{sech}(2k_0 d))}. \tag{2.26}$$

However, as real water is directionally spread and often associated with a broad spectrum, this approach usually overestimates the second order sum term. To obtain a better estimate with some relaxations on the narrow banded constrain, Toffoli et al. (2010) provide an estimate of second order sum term with bandwidth corrections. This gives

$$\eta^{2+}(t_0) = \frac{1}{2}A_c^2 + iA_c\frac{\partial A_c}{\partial t_0} - \frac{1}{2}A_c\frac{\partial^2 A_c}{\partial t_0^2}, \tag{2.27}$$

where A_c is the complex envelope of the signal and the t_0 is nondimensionalized by the peak frequency. This method provides an almost identical estimate to the exact value from Dalzell (1999) of the second order sum terms.

2.4.2 Wave linearization methods

As the second order bound harmonics are always 'slaved' to the main wave package yet not affecting the nonlinear evolution, examining the nonlinear effects in real

water waves directly can be quite challenging. Additionally, these bound harmonics also exist in some numerical simulations, which solve the fully nonlinear equations. Therefore, it can be helpful to 'linearize' the data to remove the second order bound harmonics. This will be important in the thesis with field data, experiments and fully nonlinear simulations, so it is discussed in detail here.

The approximate removal of the second order difference term is relatively straightforward: a high pass filter is designed to filter out the low frequency components. The effect of cutoff frequency for the low pass filter is discussed in detail by Toffoli et al. (2007). However, as high frequency cutoff is not practical for real time series. This is due to the linear components also having significant amount of energy, which overlaps with second order harmonics in the frequency domain. Several linearization methods are evaluated below.

Minimizing skewness

For a perfect balance between accuracy and computational speed, the second order coefficient in Equation 2.25 is computed by minimizing skewness since the second order sun term induces positive skewness into the measured time series whereas a linear signal should have zero skewness. From Figure 2.3, this method reaches almost the same accuracy as the complex envelope method but with less computational effort. Additionally, for the difference between crests and the corresponding troughs after the linearization process in Figure A.1 (b) in Appendix A, the sorted crest-trough pairs roughly follow the guideline except for several of the largest pairs. Thus the Stokes expansion in Equation 2.25 is still valid for extreme cases and this linearization process is acceptable for real water waves.

Phase inversion method

Another useful linearization technique is based on the phase inversion method. The odd and even components of a regular wave with Stokes expansion can be obtained via phase manipulation (Fitzgerald et al. 2014):

$$\eta_{odd} = S_{11} A \cos\theta + O(A^3)... = \frac{\eta - \eta_{inv}}{2}, \tag{2.28}$$

$$\eta_{even} = S_{20} A^2 + S_{22} A^2 \cos 2\theta + O(A^4)... = \frac{\eta + \eta_{inv}}{2}, \tag{2.29}$$

where A is the linear amplitude, η_{inv} is elevation time series with 180 degree out of phase compared to the original time series η, and θ is the phase. This method is convenient for numerical simulations and in a physical laboratory as it is usually straightforward to manipulate the initial conditions or paddle signal. However, applying this approach to field data requires averaging across extreme events when the inverted signal is not available. The 'two-phase' approach here can be extended to 'four-phases' (Fitzgerald et al. 2014) and beyond to improve separation between components.

2.5 Higher order waves

Higher order interactions (i.e. beyond second order) can cause energy transfer between frequencies and potentially increase the rogue wave density by creating correlation between components and modify the shape of extreme events.

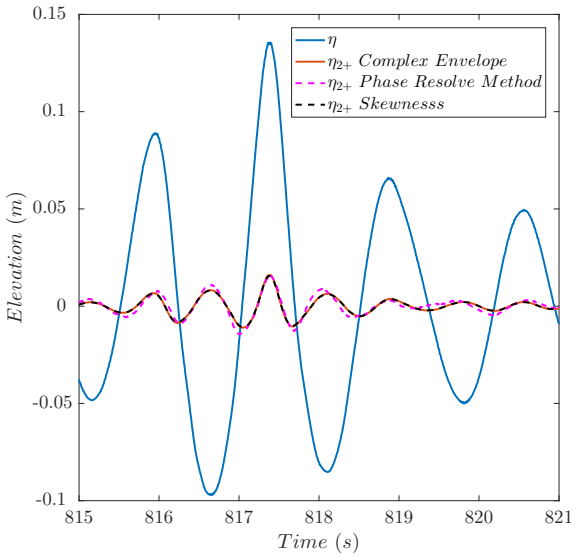

Figure 2.3: Comparison between different methods for a record at the Lake George.

2.5.1 Higher order effects

The impact of higher order resonant and near resonant interactions on extreme events has been studied for decades. A key point is that these interactions are fundamentally different in unidirectional waves (1D) compared to the directional spread waves found in ocean (2D). The studies of extreme events in field data link the averaged shape of these events with the NewWave model (see section 1.3.2 for details). Based on this finding, Gibbs & Taylor (2005) investigated the modifications of a focused NewWave group under nonlinear evolution and predicted three main differences between the linear and nonlinear model:

(i) a contraction in the mean wave direction,

(ii) a movement of the peak crest to the front of the wave-group,

(iii) an expansion in the lateral direction.

Adcock et al. (2012) provided an analytical approximation to the contraction in mean wave direction and the expansion in the lateral direction described above. Additionally, all three phenomena are observed by averaging the shape of the largest

events in a numerically simulated random sea (Adcock et al. 2015), which further confirms that these nonlinear modifications of focused NewWave group may also happen in real ocean.

Although some of the asymmetry properties of the wave-group in 2D have been explored by Lo & Mei (1985) in 1985 and the amplification of extremes in 2D has been found by Adcock & Taylor (2016*c*) in numerical simulations, Adcock & Taylor (2014, 2016*b*) suggested that the nonlinear physics is fundamentally changed by the directional spread of the different components and all naturally wind generated waves are directionally spread in the open ocean. Apart from 2D simulations being incapable of predicting (iii), the simulated results indicate that the increase in directional spread can reduce the extra nonlinear amplification and also decrease (ii), which may suggest that (i) and (ii) can be relatively small for open ocean waves. However, the expansion in the lateral direction, which is also referred to as the 'wall of water' in Gibbs (2004), can be significant with various steepness, bandwidths and directional spreads (Adcock et al. 2016). This suggests that the expansion in the lateral direction for extreme events could be more common in the open ocean and might have a significant impact on offshore structures.

Although directional information can provide more useful details for understanding the sea-state, the degree of spreading is extremely hard to obtain from a single point measurement (McAllister et al. 2018). It is relatively expensive to install an array of point measurement devices to monitor the directional spread of ocean waves, so there is little directional information available from most field data. Therefore, for field measurements, the recordings are usually processed with an omni-directional spectrum.

2.5.2 Higher order near-resonant effect: Benjamin-Feir instability.

In addition to the fact that the averaged shape of extreme events is modified by higher order resonant and near resonant effects, the probability of the occurrence of these events also departs from the Rayleigh distribution due to other nonlinear processes, such as the Benjamin-Feir instability. This instability modifies a regular wave train into modulations of long wavelength and was first systematically described by Benjamin & Feir (1967).

Although this instability usually occurs when two side-band disturbances are within a narrow region around the primary wave in terms of wavenumber, it results in an increase of the two side-band components at the cost of the primary wave. This is further confirmed by the experimental results in Chapter 5, where the survival function further departs from the Rayleigh distribution when the sea state became steeper and more narrow banded in unidirectional waves. Further investigations such as the flume experiments performed by Onorato et al. (2006) and Onorato, Cavaleri, Fouques, Gramstad, Janssen, Monbaliu, Osborne, Pakozdi, Serio, Stansberg, Toffoli & Trulsen (2009) suggest that the Benjamin-Feir instability may also influence the wave statistics in random waves and although the amplification is very much reduced, the Benjamin-Feir instability may be important in directionally spread waves (see for instance the major experimental study by Latheef & Swan (2013)).

2.6 OceanWave3D

The OceanWave3D model has been developed at the Technical University of Denmark since 2006 as a three dimensional fully nonlinear model (i.e. 2D surface plus full representation of the kinematics under waves). This model is based on the flexible-order finite difference discretisations to determine the exact potential flow solutions

(Engsig-Karup et al. 2009). This numerical potential flow solver has two versions: the CPU version is single threaded and coded in Fortran 90 and the GPU version is written in C/C++. This model provides a computationally efficient way to simulate steep nonlinear waves and shoaling problems. Another advantage of the OceanWave3D model is that the internal kinematics do not require further computation since the potential flow equations are solved within the fluid (unlike other fully non-linear solvers such as BST (Bateman et al. 2001), HOS (Ducrozet et al. 2011)).

OceanWave3D solves the potential flow equations listed in section 1.3.1. The same Cartesian coordinate system is used. Following the equation established by Zakharov (1968), the kinematic and dynamic free surface boundary conditions are expressed in terms of velocity potential $\widetilde{\Phi} = \Phi(\mathbf{x}, \eta, t)$ and the vertical velocity at the free surface $\widetilde{w} = \frac{\partial \Phi}{\partial z}|_{z=\eta}$ using the chain rule from Equation 2.17 and 2.18:

$$\eta_t = -\nabla\eta \cdot \nabla\widetilde{\Phi} + \widetilde{w}(1 + \nabla\eta \cdot \nabla\eta), \tag{2.30}$$

$$\widetilde{\Phi}_t = -g\eta - \frac{1}{2}\nabla\widetilde{\Phi} \cdot \nabla\widetilde{\Phi} + \frac{1}{2}\widetilde{w}^2(1 + \nabla\eta \cdot \nabla\eta), \tag{2.31}$$

where $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ is the horizontal gradient operator, g is the gravitational acceleration and partial differentiation is noted as the subscripts. To obtain the vertical velocity component $\widetilde{w}$ the Laplace equation (Equation 2.16) along with the same kinematic bottom condition (Equation 2.19) are used. To obtain the direct solution of the Laplace problem, the ψ transform is applied, where

$$\psi(\mathbf{x}, z, t) = \frac{z + d(\mathbf{x})}{\eta(\mathbf{x}, t) + d(\mathbf{x})} = \frac{z + d(\mathbf{x})}{d^{total}(\mathbf{x}, t)}, \tag{2.32}$$

where $d^{total} = \eta + d$ is the thickness of the fluid layer. This transform scales the z-axis and simplifies the finite difference scheme. After solving the Laplace equation and the kinematic bottom condition for the potential $\Phi(\mathbf{x}, \psi)$, the vertical velocity component

is given by

$$\widetilde{w} = \frac{1}{d}\Phi_{\psi}|_{\psi=1}. \tag{2.33}$$

This allows the Equation 2.30 and 2.31 to be stepped forward in time domain.

Chapter 3

Field measurement of non-linear changes to large gravity wave-groups

3.1 Abstract

The dynamics of large gravity waves are known to be modified from the linear model by non-linear physics. In this paper we analyse Eulerian surface elevation time histories measured from two sites, Lake George and the North Sea, to examine how weak non-linearity has modified the shape of extreme wave-groups relative to linear theory. We analyse the asymmetry of the extreme wave-groups and find that, on average, the wave in front of an extreme wave is smaller than the wave following it. We also observe a contraction in the envelope width of the wave-group relative to linear theory. The departures from linear theory are strongly correlated with the steepness of the underlying sea-state and are generally consistent with theoretical expectations providing strong evidence that such non-linear phenomena arise in naturally occurring water waves.

3.2 Introduction

Water wave propagation is non-linear due to the nature of the free surface boundary condition and the kinetic energy term in the governing equations. There is much interest in whether this 'weak' non-linear physics (as opposed to the "strong" non-linearity associated with wave breaking) can lead to abnormal waves in the ocean – sometimes called freak or rogue waves (Kharif & Pelinovsky (2003), Dysthe et al. (2008)). In unidirectional waves, non-linear instabilities, first described in the pioneering work of Benjamin & Feir (1967), with important developments by Janssen (2003) and many others, will lead to waves of far higher amplitude than that expected in a linear model (Mori et al. 2007, Onorato et al. 2006). However, most waves found in nature, and certainly those in severe ocean storms, are not unidirectional but have a significant directional distribution of energy. This directional distribution fundamentally changes the nature of the non-linear interactions and the process by which extreme events form in random seas (see discussion in Adcock & Taylor (2014)).

In real directionally spread waves, non-linear physics gives relatively little extra amplitude above that expected by a model based on linear propagation (with corrections for bound harmonics) (Onorato et al. 2001, Socquet-Juglard et al. 2005, Fedele et al. 2016). This appears to agree with the most extensive studies of wave statistics in the ocean (Christou & Ewans 2014). However, non-linear physics is predicted to have the potential for making dramatic changes to the shape of the largest waves on deep water in directional spread seas relative to those predicted by linear theory. In linear theory, the expected shape of extreme waves is given by NewWave – the group is symmetrical in space and time with the shape described by the autocorrelation function (Lindgren 1970, Boccotti 1983, Tromans et al. 1991). Numerical (Gibbs & Taylor 2005, Adcock et al. 2015) and analytical (Adcock et al. 2012) work predicts that non-linear physics would modify this so that on average:

1. The largest wave moves to the front of the wave-group;

2. The group expands in the lateral direction relative to linear theory;

3. The group contracts in the mean wave direction relative to linear theory.

Experimental work at Imperial College (Latheef et al. 2017) has found evidence of the second prediction in real water. However, it has proved difficult to test these predictions against ocean measurements partly due to the limited amount of wave by wave data available in deep water, and also because these changes (particularly (i) and (iii)) only occur in very steep and relatively narrow-banded seas (Adcock et al. 2016). Set against this, Fujimoto et al. (2019) found relatively small effects in their numerical simulations when spectra which were broader and closer to equilibrium (and more realistic) than in the aforementioned studies.

The first study to search explicitly for (i) in field measurements was by Gemmrich & Thomson (2017). They analysed buoy data in the Pacific Ocean as well as pressure measurements. The analysis methods they used are somewhat different to those used herein. Unfortunately they provide no general information on the period or steepness of the sea-states analysed. Our expectation, based on usual metocean behaviour, is that the sea-states would be rather less steep than we consider in this study. If this is correct then theory would predict little wave-group asymmetry. Gemmrich and Thomson state they found 'no evidence for asymmetric wave envelopes for large waves'. However, when they consider steep waves (their fig. 3(d)) there does appear to be a clear trend which they note is consistent with the work of Adcock et al. (2015). As such, we agree with their conclusion that if you consider all groups – which are mainly not very steep – you will not see asymmetry. However, given the information available to us, their data appears to us to agree very well with theory although this is not clear from their conclusions.

In this study, we analyse datasets from Lake George in Australia and the North

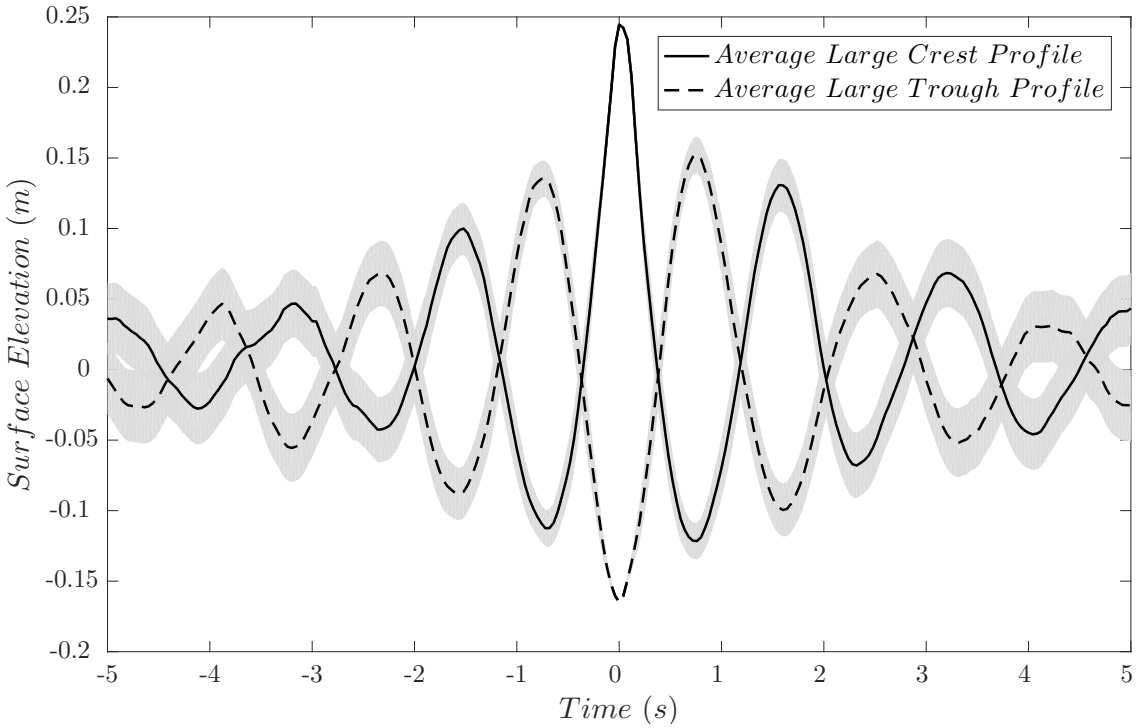

Figure 3.1: Average largest crest and trough profiles form the Lake George dataset with Lindgren Variance shaded following the approach of Santo et al. (2013).

Sea. The former are not open ocean measurements but these are records of environmentally generated real waves. Due to the fetch-limited nature of the waves, many of the 'sea-states' are steep and relatively narrow-banded, thus making non-linear changes easier to identify. Unfortunately, we have an insufficient directional resolution for Lake George to draw any conclusion on the lateral expansion of the group, and no directional information is available for the North Sea. However, we are able to clearly identify the other non-linear changes in the shape of extreme wave relative to the linear theory described above. Figure 3.1 presents an overview of the average shape of the top twenty largest crests and troughs from a typical elevation record considered in this study – the difference between the waves in front of and behind the crest is clear. At the end of the paper we discuss possible sources of error, consider other possible physical mechanisms which could produce these results, and examine some implications of the observations.

3.3 Data

In the paper, we examine extreme waves from two distinct datasets: Lake George and the North Sea. We present some background to these datasets in subsections 2(a) and 2(b) before considering the types of sea-states present in each in 2(c).

3.3.1 Description of the Lake George dataset

Surface elevation measurements taken from the Lake George field experiment site is a part of the integrated datasets obtained by Young et al. (2005). The measurement campaign lasted from September 1997 to May 1999. The geographical layout of Lake George is shown in Figure 3.2, indicating a large lake with a fairly flat bed. Previous experiments at Lake George successfully determined the spectral evolution of wind-generated waves (Young & Verhagen 1996) and the directional spectrum with the implementation of wave arrays using Maximum Likelihood Method (MLM) (Young 1994, Young et al. 1995). Based on prior experience, a project was established to investigate the physics behind the steep and narrow-banded waves at a new experimental site shown in Figure 3.2. During this project, a spatial array consisting of 8 capacitance probes were deployed to determine the surface elevation at a resolution of 25 Hz (Young et al. 2005). The detailed configuration is shown in Figure 3.3: five probes were separated evenly in a 30 cm diameter circle and another three probes were placed in the middle. These probes measured high-quality wave records with values of peak frequency f_p ranging from 0.33 to 0.5 Hz and values of significant wave height H_s varying from 0.08 to 0.5 m.

As a part of the integrated dataset, the wave elevation records have been used for many purposes. Babanin et al. (2001) applied the data together with video records and acoustic signals to obtain the breaking probability at finite depth. He also used the spectral properties of wave elevation data as well as wind data to address the

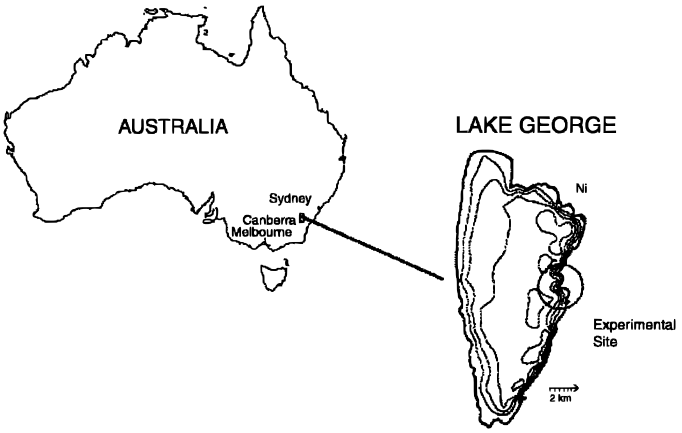

Figure 3.2: Location of the Lake George site. (taken from Babanin et al. (2001)).

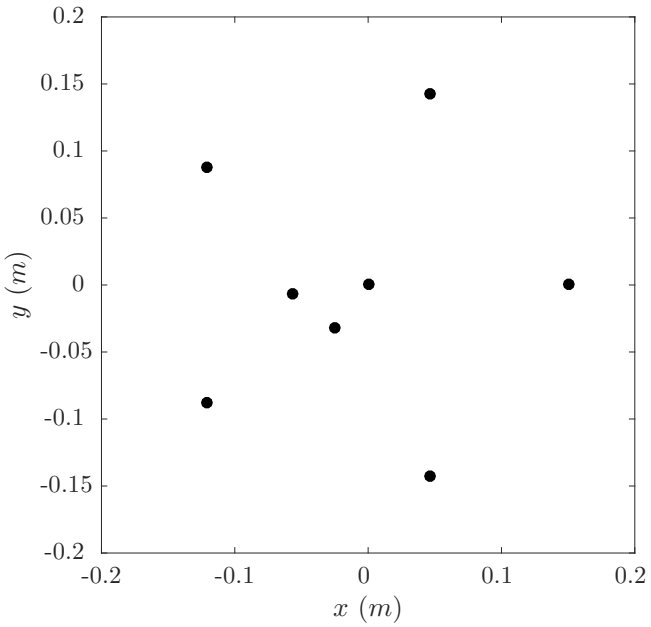

Figure 3.3: Configuration of the capacitance array.

dependencies of drag coefficient on the wind speed, sea-state and gustiness (Babanin & Makin 2008). In addition, Toffoli et al. (2007) also analysed the statistical properties of wave crests to validate the numerical simulation of second-order wave theory.

During the measurement period, as part of natural hydrology cycles, the lake was drying out gradually with the water depth dropping from 1.1 m in 1997 to 0.4 m in 1999. In this study, we use a pre-filtering process to exclude the time series with H_s less than 0.1 m or the non-dimensional water depth ($k_p d$, where k_p is the peak wavenumber and d is the water depth) less than 1.6. This confines the data to be more representative of developed wind-generated waves at deep and intermediate water depths. Additionally, a rigorous data quality control procedure is applied to each dataset to remove instrumental errors and to produce a reliable dataset following the approach of Christou & Ewans (2014).

3.3.2 Description of the North Sea dataset

The North Sea dataset contains a large amount of wave and wind data taken from several oil platforms in the North Sea (see Figure 3.4 for detailed locations). The measurement period started on November 2013 and ran for four consecutive months. The collected data are then divided into 30-minute time series, and over 43000 wave records are analysed in the project. The large amount of wave records provide an insight into wave behaviour in the North Sea with several storms observed.

The data were measured by downward looking SAAB radars installed at the side of the platforms. All the data were from fixed jacket structures and not from buoys. Measuring waves in the harsh maritime environment is very difficult. The accuracy of different measuring methods is not known definitively but studies such as Forristall et al. (2004) suggests that radar measurements are not as accurate as a laser or a wave staff but that the measurements from radars are still useful. Ewans et al. (2013) have looked in detail at the performances of wave radars and their recent work Ewans

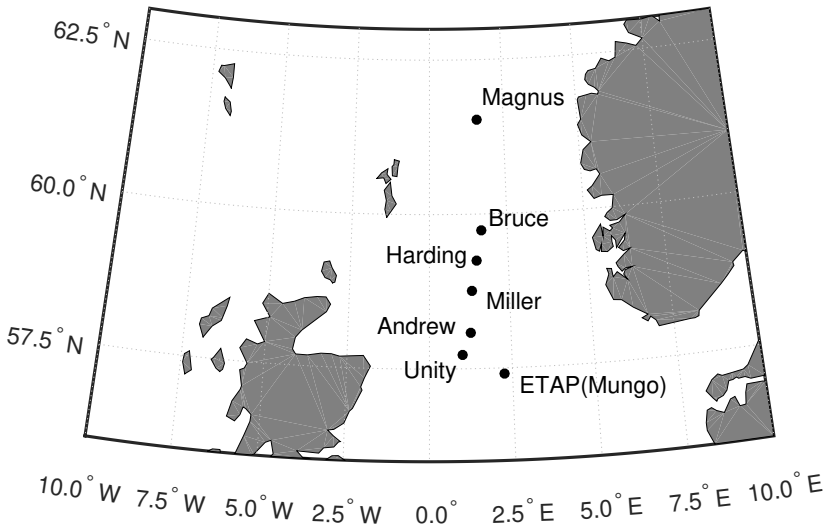

Figure 3.4: North Sea data locations.

et al. (2014) shows that the SAAB radars in the North Sea have no obvious bias.

Field measurement made by wave radars have been used by numerous authors including Christou & Ewans (2014) major study of rogue waves. Other examples of using radars for the analysis of extreme waves include Bell et al. (2017) and Taylor & Williams (2004) who examined the average properties of extreme waves. Whilst we think measurements made with wave radar are accurate enough for analysis, particularly where an average is taken over many records, some caution should be applied to any calculations drawn from wave radar data.

The measuring frequency is 2 Hz. Since there is no exact description of the rig for each platform, all the time series within the North Sea dataset are assumed to be free-field measurements, although wave-structure interactions can be significant for some platforms. Possible sources of error are considered further in the discussions.

After the initial inspection for missing entries, all the time series are filtered to remove any background noise over 4 rad/s. The filtered time series are then processed using the same quality check approach as the one used for the Lake George dataset to obtain reliable wave records. The North Sea records have H_s varying from 0.38 m

to 13.33 m and the zero crossing frequency, f_z, between 0.54 rad/s and 1.95 rad/s.

3.3.3 Comparison of two datasets

Due to the location of the measurement and the instrumentation, there are several significant differences between the wave elevation records from the two datasets. The location of the Lake George dataset allows the deployment of high-quality elevation probes, which provides high-quality data with higher sampling frequency and less disturbance from the rig. Additionally, there is no swell or tidal variation for the Lake George dataset. However, the total amount of available records is significantly less than the North Sea dataset, leading to more statistical variation in our analysis.

For comparison purposes, some basic parameters from both datasets are presented. In Figure 3.5 the zero crossing period T_z is plotted against significant wave height H_s from both datasets. Additionally, lines of constant mean steepness of the sea-state are also presented for comparison, as the steepness is a key measurement of non-linearity. Steepness for deep water waves is given by

$$Steepness = \frac{2\pi}{g}\frac{H_s}{T_z^2}, \tag{3.1}$$

where g is the gravitational acceleration. Although the wave height of the Lake George data is relatively small, these records are exceptionally steep. This is primarily because most of the waves in the Lake George data are generated by strong and localised wind (see Young et al. (2005) for details).

As well as steepness, another key parameter measuring the non-linearity of wave recordings is the Benjamin-Feir Index (BFI) proposed by Janssen (2003). In this paper, the approach of Serio et al. (2005) is adapted to compute BFI from the time series:

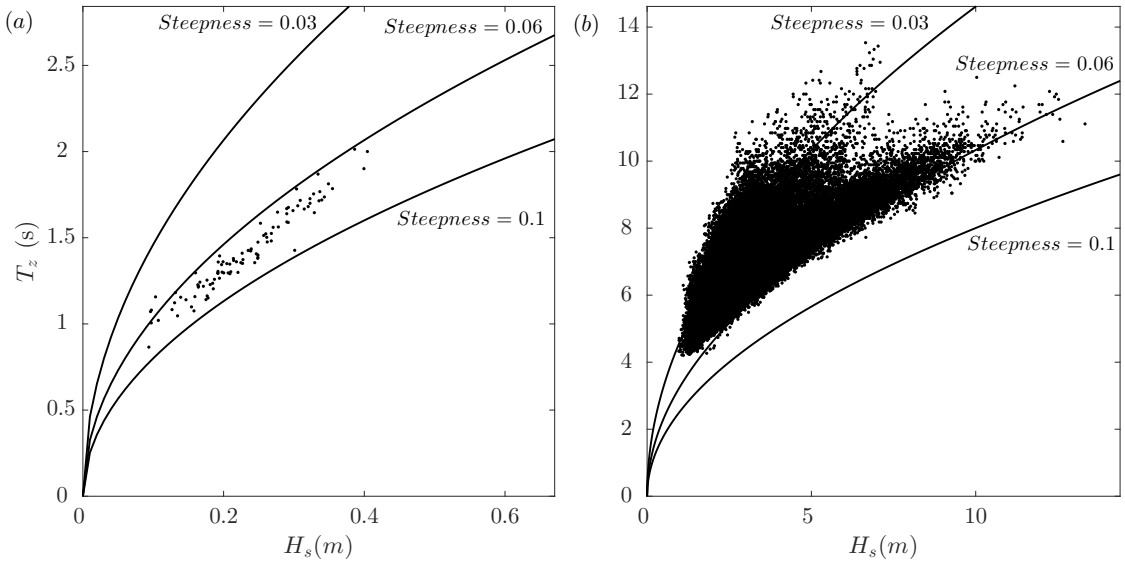

Figure 3.5: Scatter plot for H_s and T_z from (a): Lake George, (b): North Sea

$$BFI = \sqrt{m_0} k_p Q_p \kappa \sqrt{2\pi}, \tag{3.2}$$

where $m_0 = H_s^2/16$ is the zeroth moment of the energy spectrum, κ is a depth correction factor, which is unity in deep water. κ can be computed as:

$$\kappa = \mu \sqrt{\frac{|\tau|}{\chi}}, \tag{3.3}$$

where μ, τ and χ are all dimensionless coefficients depending on relative water depth $k_p d$. The general forms of μ, τ and χ are reported (see Mei (1989) for detailed derivation):

$$\mu = 1 + 2\frac{k_p d}{\sinh(2k_p d)}, \tag{3.4}$$

$$\tau = -\mu^2 + 2 + 8(k_p d)^2 \frac{\cosh(2k_p d)}{\sinh^2(2k_p d)}, \tag{3.5}$$

$$\chi = \frac{\cosh\left(4k_p d\right) + 8 - 2\tanh^2\left(k_p d\right)}{8\sinh^4\left(k_p d\right)} - \frac{\left(2\cosh^2\left(k_p d\right) + 0.5\mu\right)^2}{\sinh^2\left(2k_p d\right)\left[\frac{k_p d}{\tanh\left(k_p d\right)} - \left(\frac{\mu}{2}\right)^2\right]}. \tag{3.6}$$

Q_p is the quality factor introduced by Goda (2000). Q_p is a dimensionless parameter, which describes the spectral bandwidth. It has less sensitivity to the high frequency tail of the spectrum (and cut-off frequency) than other bandwidth metrics (Serio et al. 2005, Prasada Rao 1988). Q_p is given by:

$$Q_p = \frac{2}{{m_0}^2}\int_0^\infty fS^2(f)\mathrm{d}f, \tag{3.7}$$

where $S(f)$ is the wave spectral density function.

Hence, the quality factor with depth correction and the steepness for each dataset are computed in Figure 3.6 with constant lines of BFI. Although containing fewer records, the majority of the Lake George dataset have higher BFI value than the North Sea dataset, mainly because of the high wave steepness. The average value of the depth corrected quality factor are close for both datasets but the waves in the Lake George data are generally more narrow banded (see Figure 3.7 for detailed comparison). Although the BFI is not a perfect parameter for describing non-linearity (it does not contain any information on directional spreading, which is known to be important), it is remarkable that several values are close to the critical BFI = 1 line yet no data exceeds this with the depth correction factor.

3.3.4 Typical spectrum

The wave spectrum is the basic tool used to analyse a sea-state. A sea-state varies over time, so each time series has a different spectrum. Nonetheless, it is useful to present a typical omnidirectional spectrum for a wave record taken from the Lake George

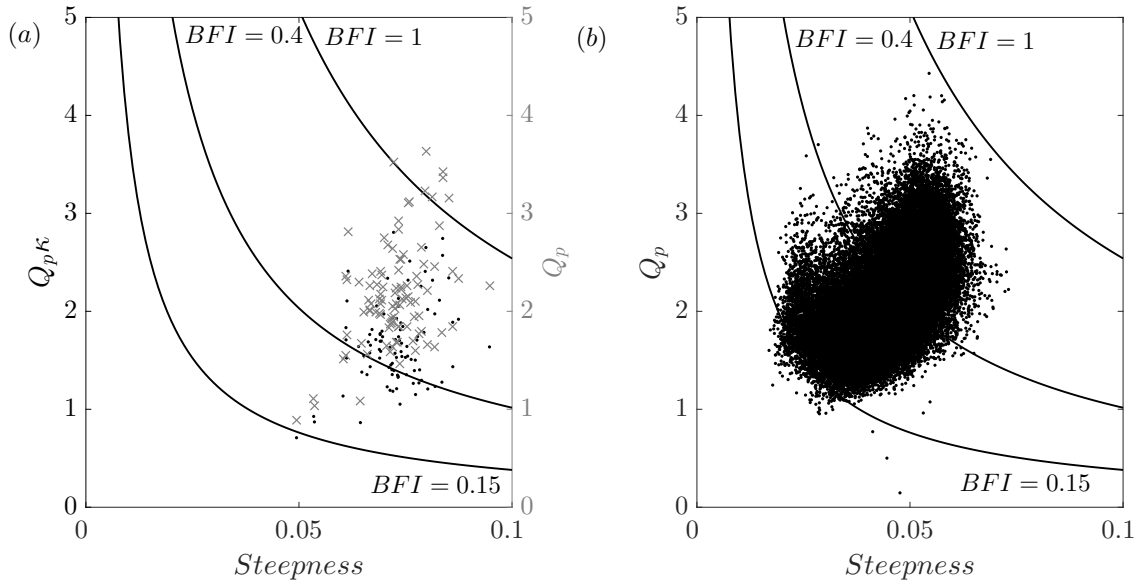

Figure 3.6: Scatter plot for steepness and quality factor from (a): Lake George, the secondary axis with 'x' marker shows the quality factor without depth correction, (b): North Sea. (depth correction does not affect the quality factor in deep water).

dataset and the North Sea dataset respectively, which is shown in Figure 3.7. A Tukey window is applied before the Fast Fourier Transform to prevent spectral leakage. The spectra are generated based on the average of 30 non-overlapping segments, which are then normalised by m_0 and f_z of the wave record for comparison.

One significant difference between two spectra is that the Lake George dataset is more narrow banded, which is probably due to the records in the North Sea dataset being well developed whereas there are quite a few relatively young waves in the Lake George dataset. This is consistent with the Figure 3.6, where the Lake George dataset tends to have a higher quality factor (over 5% in the mean value) without the depth correction κ (*i.e.* low spectral bandwidth) on the secondary axis. However, the depth correction factor does not modify the quality factor in the North Sea dataset as the water is rather deep at these oil platforms. Along with the steepness, spectral bandwidth and degree of directional spreading are key factors, that are expected to influence the non-linear physics investigated by us in this paper (Adcock & Taylor 2016*b*).

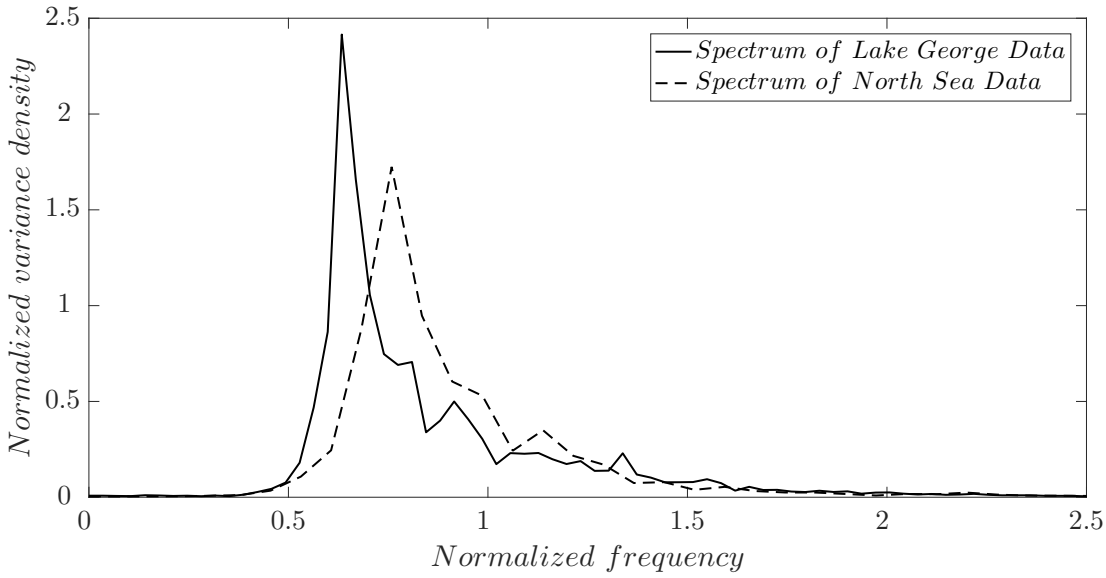

Figure 3.7: Comparison of typical Lake George spectrum with North Sea data normalized by the zeroth moment m_0 and zero crossing frequency f_z of the individual wave record.

The directional spectrum for the Lake George dataset is also obtained using the MLM (Young 1994). The spreading parameters of these spectra indicate that the spreading angle is generally less than a broadly spread wind sea. However, we found our estimates of directional spread to be noisy and therefore these numbers are not reported in detail. Unfortunately, there was no directional spreading information available for North Sea dataset, which consists of Eulerian point measurements from which it is challenging to estimate the exact directional spreading (Adcock & Taylor 2009*a*). Nevertheless, we would expect directional spreads to be mainly 'following' sea-states with a small number of crossing seas. See for instance the North Sea data reported in McAllister et al. (2017).

3.4 Results

For the Lake George dataset, a total of 256 sets of data were collected in a fairly controlled environment (Young et al. 2005), and 98 of them are within the scope of

this paper after performing pre-filtering and data quality check. Each set of data contains eight twenty-minute time series, which were captured by eight capacitance probes (see Figure 3.3) simultaneously. Due to the probes in the array being quite close to each other compared to the typical wave length (see Figure 3.3), the measurements from one probe with the highest consistency (*i.e.* no significant departure in wave statistics from the measurements obtained by the probes nearby) are presented here. Meanwhile, the North Sea dataset has 43000 usable 30-min time series. In the following subsections, the data will be processed to investigate horizontal asymmetry using three different parameters as each gives a slightly different insight to the problem: the elevation difference between three successive large crests, the height difference between either side of the envelope, and the change in envelope width due to the sea-state variation.

3.4.1 Analysis of raw time series

We start by analysing the relatively raw (although quality controlled) time series for both datasets. Our aim here is to demonstrate that the key phenomenon observed in the study does not result from some of the post-processing techniques used to obtain greater insight into the data in subsequent sections.

To assess the horizontal asymmetry properties of the raw wave records from the Lake George and the North Sea dataset, a robust data processing method is applied to find the difference between the preceding crest height and the following crest height of the largest waves. We analyse the largest five crests/waves in each time series – an example is shown in Figure 3.8. This process is also restricted by a rule that the largest waves must be separated by a minimum distance of two zero crossing periods to prevent selecting multiple waves from the same wave-group. Afterwards, a parameter α is defined to measure the relative ratio of adjacent crests to the largest one, which is similar to the measurement of unexpectedness established by Gemmrich

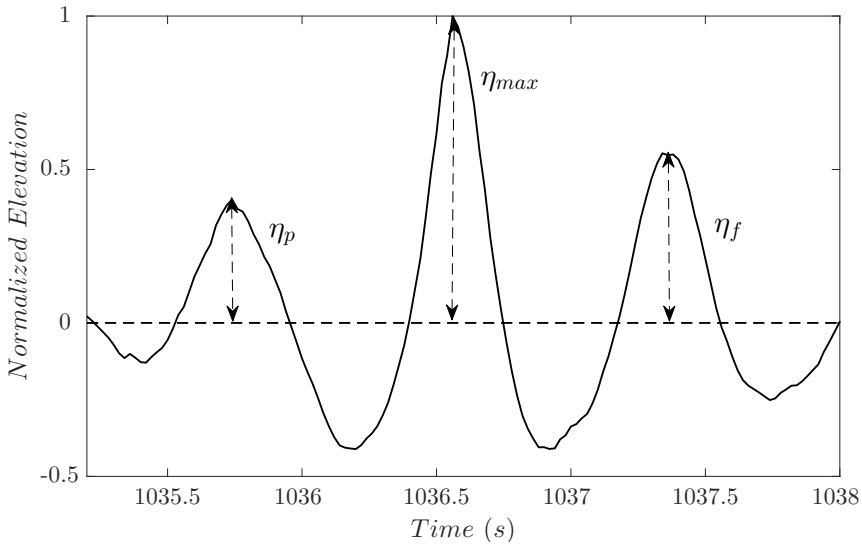

Figure 3.8: Illustration of the relative height of preceding and following crests.

& Thomson (2017). Here, α is measured by the preceding wave crest η_p or following wave crest η_f in relation to the largest crest η_{max} in this wave-group (Figure 3.8):

$$\alpha_1 = \frac{\eta_p}{\eta_{max}}, \ \alpha_2 = \frac{\eta_f}{\eta_{max}}. \tag{3.8}$$

After mass processing each wave record from the two pre-filtered datasets, the mean values of α_1 and α_2 are calculated by averaging the top five largest crests. The mean values are then categorised into different bins based on two sea-state parameters: mean steepness, and the ratio between maximum crests η_{max} in each time series and significant wave height (H_s). The mean steepness is a key measurement for non-linearity, whereas the ratio η_{max}/H_s shows how big the largest waves are relative to the underlying sea-state. The average mean values of α_1 and α_2 from each bin are presented in Figure 3.9. A total of 490 wave profiles from the Lake George dataset and over 200,000 wave profiles from the North Sea dataset are analysed herein. To get a measure of the statistical validity, we use bootstrapping to give a 90% confidence interval on the mean value of each bin (Efron & Tibshirani 1994).

In Figure 3.9 (a, c), the α values are divided into small bins based on the η_{max}/H_s

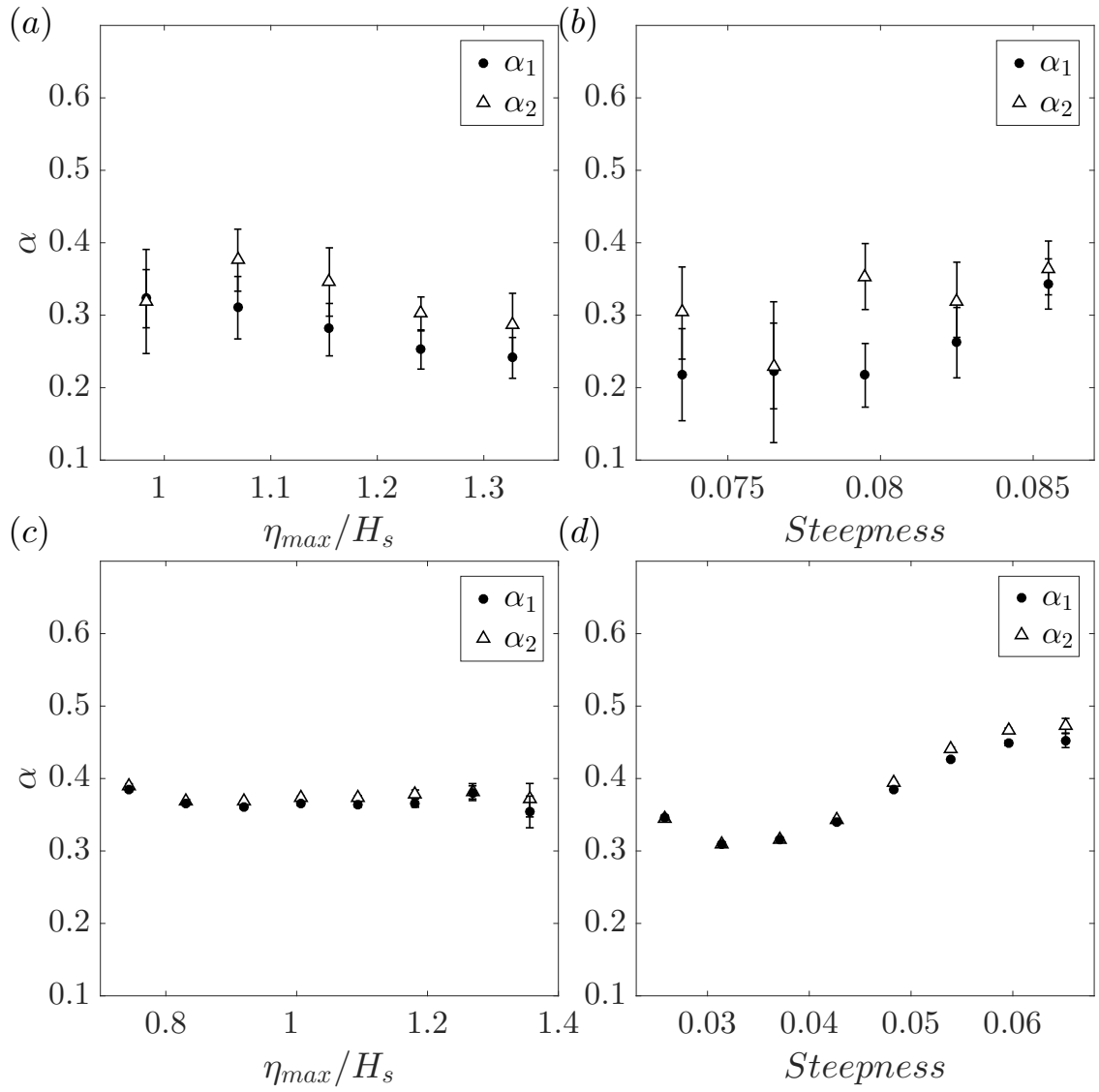

Figure 3.9: Relative height of preceding and following crests from: (a, b): Lake George data, (c, d): North Sea data, for different (a, c): normalized maximum elevation, (b, d): mean steepness.

ratio, which measures the significance of a maximum crest compared to the background sea-state. There is a clear separation between α_1 and α_2 for the Lake George dataset. This indicates that the largest crests in the Lake George dataset tend to have a relatively smaller crest at the front and a larger crest at the back, which is consistent with the numerical simulation proposed by Lo & Mei (1985) in unidirectional numerical wave tank and Adcock et al. (2015) for directional spread waves. However, for the open ocean data (Figure 3.9 (c)), α_1 and α_2 seem to be coincident with each other for a relatively small η_{max}/H_s ratio but the difference becomes more evident for the wave records with the presence of rogue waves (defined as $\eta/H_s \geq 1.25$).

Additionally, both α_1 and α_2 seem to have a decreasing trend for the Lake George dataset, when the records contain larger waves. This phenomenon could be attributed to a non-linear contraction to the wave-group but could also be accounted for in different ways. For instance, bound harmonics would increase the size of the largest wave relative to the wave on either side. Based on the same reasons, it is not surprising to find that there is no clear overall trend for the North Sea dataset.

In Figure 3.9 (b, d), the α values are categorised by the steepness of the underlying sea-state, which also shows a clear difference between the preceding crest height and the following crest height for the Lake George dataset. Although the separation shown in Figure 3.9 (d) for the North Sea dataset is not as significant as the previous one, this phenomenon can still be observed in the open ocean, especially for time series with a higher steepness value, which further confirms that there is some horizontal asymmetry in both datasets. However, a more detailed analysis is required to understand the trends involved.

3.4.2 Analysis of envelope

To further analyse the horizontal asymmetry presented in the section $3(a)$, a more delicate investigation into the envelope shape is conducted to demonstrate the hori-

zontal asymmetry properties of the time series. The envelope $|U|$ is evaluated from the phase-resolved linearised free surface elevation to avoid the influence of bound harmonics (see Appendix A for details) using:

$$|U| = \sqrt{\eta_L^2 + \eta_{LH}^2}, \tag{3.9}$$

where η_{LH} is the Hilbert transform of the linearised elevation record.

To examine the asymmetrical properties of the envelope, a parameter β, which is similar to the ζ introduced by Adcock et al. (2015), is defined (see Figure 3.10 for details) as the ratio between the envelope height ($|U_p|$ and $|U_f|$ respectively) at half zero crossing period T_z before and after the maximum envelope height $|U_{max}|$:

$$\beta_1 = \frac{|U_p|}{|U_{max}|}, \ \beta_2 = \frac{|U_f|}{|U_{max}|}. \tag{3.10}$$

For a linear random time series, the value of β should be dependent on the spectral width. Indeed, in the linear model, the expected shape of an extreme event with unity amplitude is given by unit NewWave (Lindgren 1970, Boccotti 1983, Tromans et al. 1991):

$$\eta(t) = \frac{1}{m_0} \int_0^\infty S(f) \cos(2\pi f t) df, \tag{3.11}$$

To ensure that any observed trend is not due to the correlations between steepness and bandwidth, we have also analysed the shape of NewWave groups derived from the underlying spectra of the entire linearised background sea-state. Figure 3.11 (a) and (b) show the variation of β values for different η_{max}/H_s ratios, in terms of linearised real time series and the corresponding NewWave profile. As for the real data, which are described in Figure 3.11 (a), the relative envelope height before the envelope maximum (β_1) is significantly less than the one after the peak (β_2). Moreover, the difference between β_1 and β_2 seems to be more significant when an extreme wave

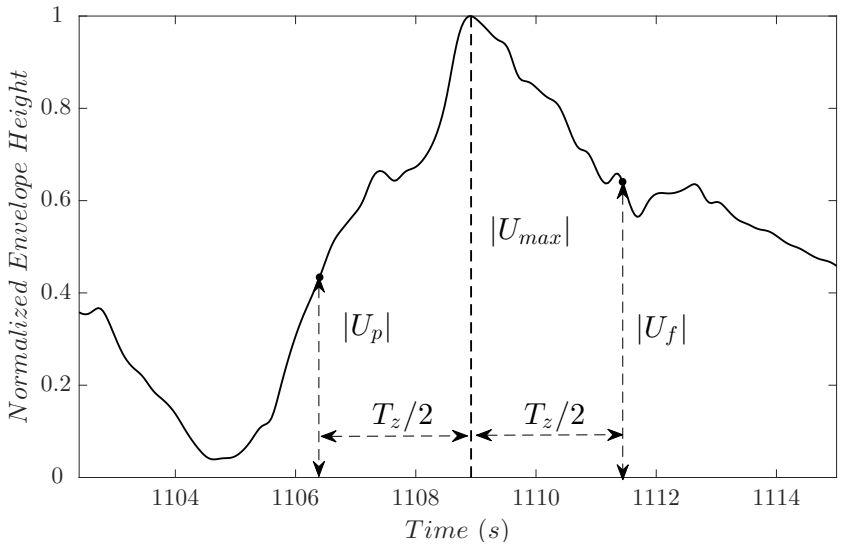

Figure 3.10: Illustration of the relative envelope height at $T_z/2$ away from the envelope maximum.

occurs within the wave record, which is captured by a higher η_{max}/H_s ratio. This will also lead to a decrease in both β_1 and β_2, which we suggest is caused by the non-linear contraction of the wave-group, as predicted by theory (Adcock et al. 2012). However, these three main trends for real linearised time series are not found in NewWave (see Figure 3.11 (b)). The distinct difference between the real-time series and the NewWave indicates that the apparent contraction of wave-group is not a linear effect but due to some non-linear changes during the extreme events.

However, the overall tendency of the β_1 and β_2 for the North Sea dataset is not obvious as the spectral bandwidth varies significantly with the change in both the η_{max}/H_s ratio and the steepness. Hence, a new parameter B, which is defined as the ratio between $\beta_{measured}$ and $\beta_{NewWave}$, is established to describe the relative percentage of NewWave envelope taken by the envelope of measured data:

$$B_1 = \frac{\beta_{1,measured}}{\beta_{1,NewWave}}, \; B_2 = \frac{\beta_{2,measured}}{\beta_{2,NewWave}}. \tag{3.12}$$

Although this parameter is perfect for describing the difference between the measured envelope and NewWave envelope, during data processing, the division also in-

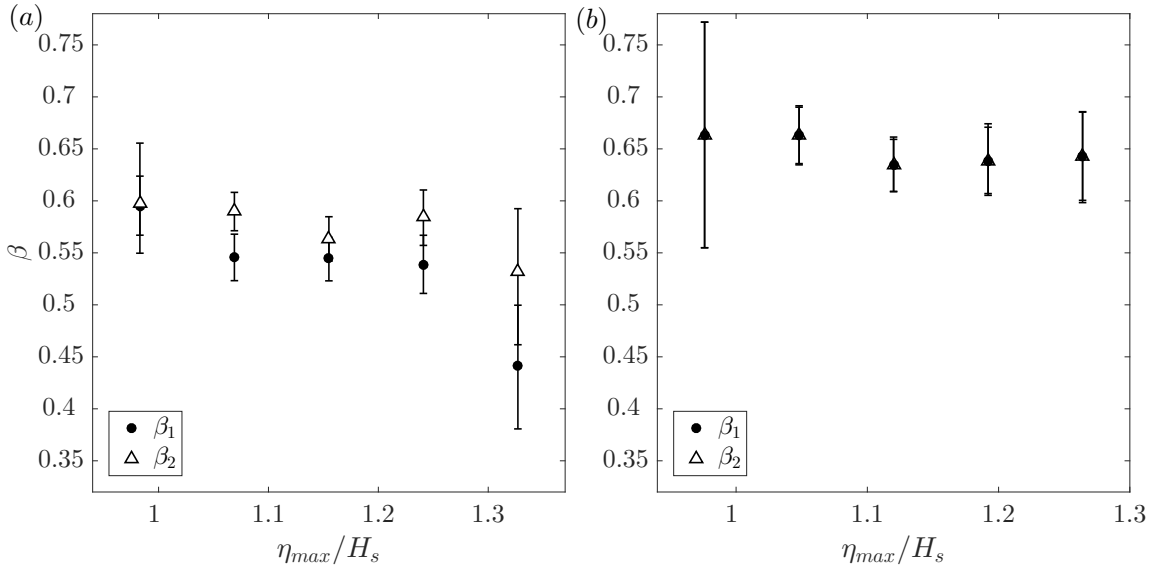

Figure 3.11: Envelope height half period away from peak at different normalized maximum elevation from Lake George dataset for different (a): linearised data, (b): NewWave.

troduces some statistical uncertainty, which results in this parameter being unsuitable for a relatively small sized dataset *i.e.* the Lake George dataset.

In contrast, the confidence interval of North Sea results (over 43000 usable time series) is still acceptable, which is shown in Figure 3.12. In Figure 3.12 (a), the relationship between relative envelope ratio B and the η_{max}/H_s ratio is established. Similar to the Lake George dataset, compared with the NewWave envelope, the measured relative envelope height at a half period before the maximum (B_1) is smaller than that after the maximum (B_2), which indicates there may be some horizontal asymmetry even in the open ocean. Moreover, the overall decreasing trend is also found in the North Sea dataset, which is believed to be caused by the non-linear group contraction in mean wave direction. We note that compared to the Lake George dataset, which is exceptionally steep, the horizontal asymmetry observed in North Sea dataset is relatively small for most records.

In Figure 3.12 (b), relative envelope ratio B is categorised based on the mean steepness, which presents two almost identical trends in Figure 3.12 (a): B_1 and B_2

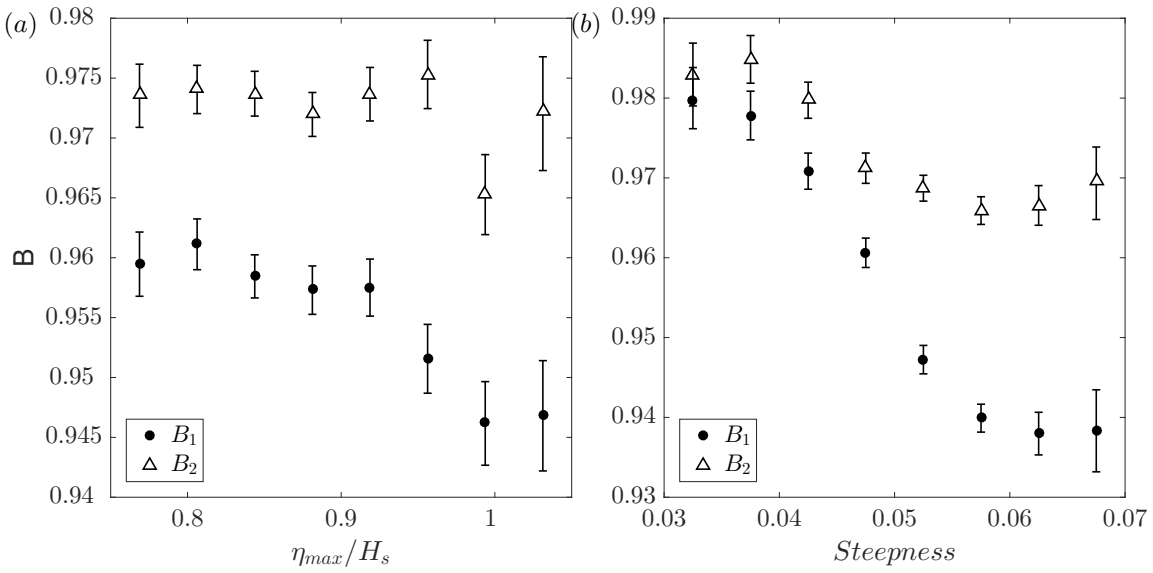

Figure 3.12: NewWave based relative envelope height half period away from peak from North Sea dataset for different (a): normalized maximum elevation, (b): mean steepness.

are well separated especially for sea-states with higher steepness value, and both B_1 and B_2 decrease for steeper sea-states. This indicates the envelope tends to have a steeper front and a relatively flat tail, and the wave envelope tends to contract for steeper sea-states. This further confirms that the non-linear physics can also alter the shape of the envelope in the ocean. However, for relatively low steepness, both ratios B_1 and B_2 are quite close to 1, which suggests that the NewWave is a good approximation for most extreme events in the open ocean. Other physical explanations which might lead to this phenomenon will be discussed in Section 3.5.

As an alternative approach to studying the shape of the wave package, a parameter σ is established to measure the envelope width, which is similar to the group bandwidth presented by Adcock et al. (2015). The parameter is defined as the envelope width (σ_1 and σ_2 respectively) on both sides between the envelope peak and the point, where envelope height is 80% of the maximum of the envelope. Additionally, the total envelope width σ_{total} is also presented for a better description of envelope width (see Figure 3.13 for a detailed illustration).

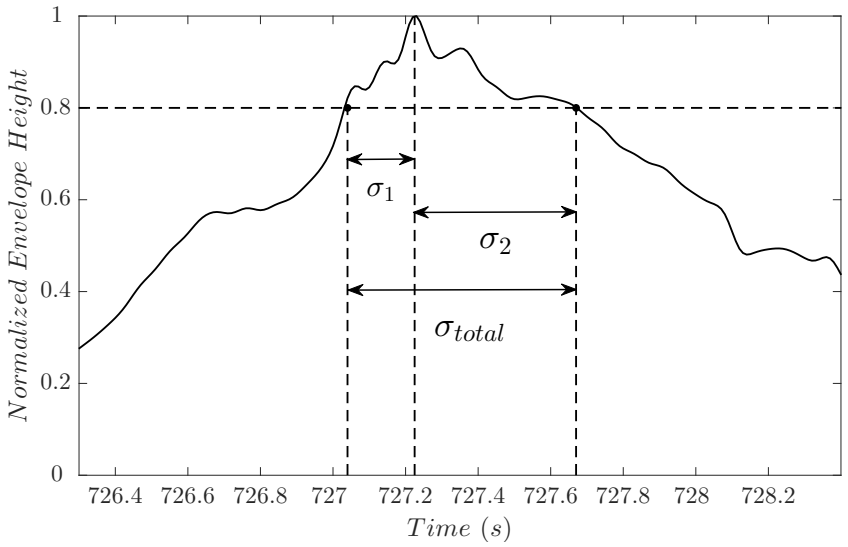

Figure 3.13: Illustration of the envelope width at 80% of maximum peak height of the envelope.

The correlation between σ and steepness for the Lake George dataset is plotted in Figure 3.14 for both linearised real time series and NewWave derived from the spectrum. It is obvious that the envelope width before the envelope peak is less than that after the maximum for relatively steep wave records, which indicates stronger envelope horizontal asymmetry for steeper sea-states. Additionally, with the increase in steepness of the sea-state, the envelope becomes much narrower around the crest, which is consistent with the previous numerical simulation by Adcock et al. (2016) as well as some experiments (Shemer et al. 1998) albeit in unidirectional waves. Additionally, the same check on NewWave is conducted to ensure this phenomenon is not caused by the background spectrum. The envelope width of NewWave is much broader than measured data, which is also consistent with non-linear physics causing a contraction of the wave-group.

Figure 3.15 presents the results from analysing the envelope width from the North Sea dataset with the same data processing method. Due to the fact that the envelope width in the linear model is highly dependent on the spectral bandwidth, width of a NewWave group, which accounts for the variation in bandwidth is used for

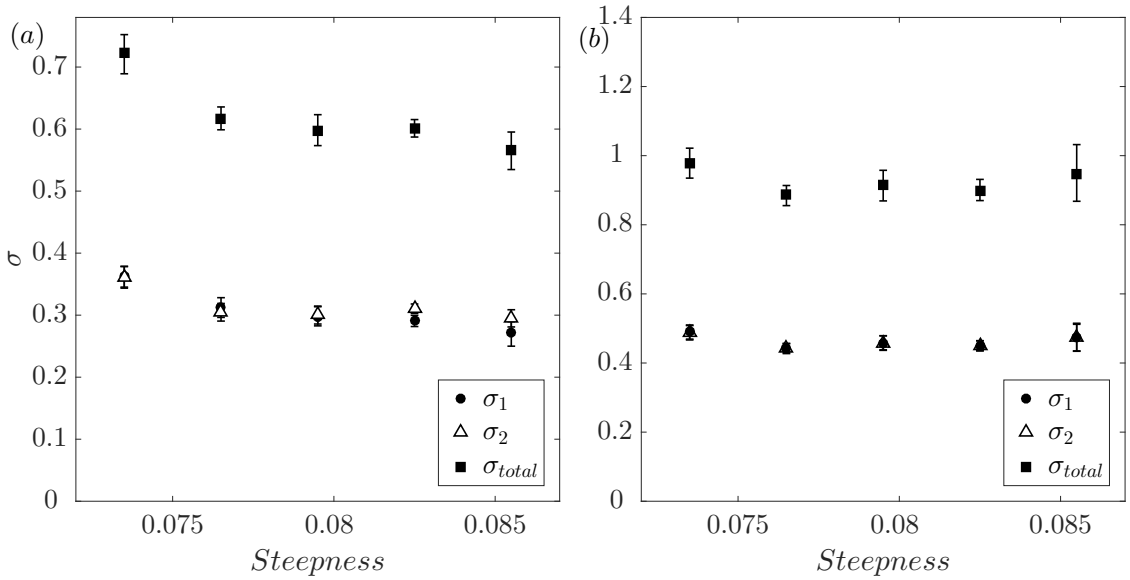

Figure 3.14: Envelope width at 80% of maximum peak height from Lake George dataset at different steepness for different (a): linearised data, (b): NewWave.

normalization. Therefore, a new parameter ξ defined as the ratio of measured envelope width $\sigma_{measured}$ to the $\sigma_{NewWave}$ is computed as below:

$$\xi_1 = \frac{\sigma_{1,measured}}{\sigma_{1,NewWave}},\ \xi_2 = \frac{\sigma_{2,measured}}{\sigma_{2,NewWave}}. \tag{3.13}$$

The variation of this parameter in terms of the η_{max}/H_s ratio and steepness is also presented in Figure 3.15 (a) and (b) respectively. Overall the result is very similar to Figure 3.12, which indicates that there is a subtle but clear separation between ξ_1 and ξ_2 with respect to both the η_{max}/H_s ratio and steepness. For small steepness, the ξ value is also quite close to 1 indicating NewWave is still an excellent description for the extreme events when non-linear physics is not significant.

3.5 Discussions

In this paper, we have analysed the asymmetry in field measurements of large waves and have found that, on average, the wave in front of a large wave is smaller than

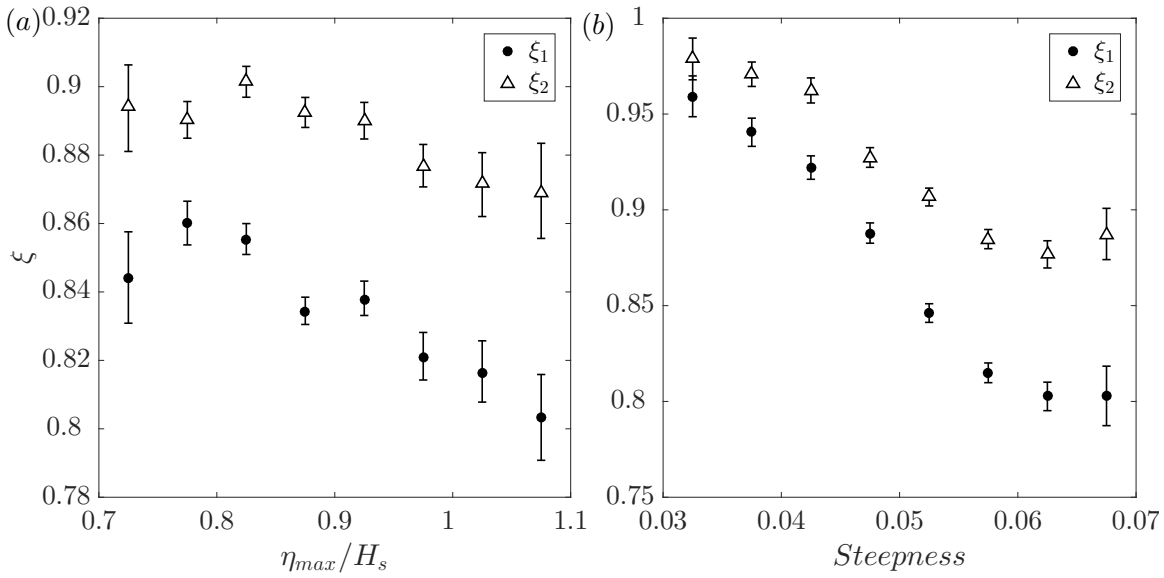

Figure 3.15: NewWave based relative envelope width at 80% of maximum peak height from North Sea dataset for different (a): normalized maximum elevation, (b): mean steepness.

the wave which comes after it. We believe the physical mechanism is due to non-linear dispersion where large waves will tend to travel faster than smaller waves and thus move to the front of the wave-group. Taking both datasets, it is clear that this trend is dependent on the steepness of the underlying sea-state, which is consistent with the explanation that the asymmetry is due to non-linear physics. Unfortunately, either because of lack of information, or the complexity of the available data, it has been difficult to study the influence of bandwidth on this phenomenon. The influence of bandwidth was considered by Gemmrich & Thomson (2017) but their results do not account for the correlation between the expected group shape and the spectral bandwidth predicted by linear theory (Lindgren 1970).

We have also looked for the contraction of the wave-group predicted by non-linear theory. This is again present in both datasets and is strongly correlated with steepness, which is consistent with analytical and numerical predictions.

Unfortunately, the expansion of wave-group in the lateral direction cannot be directly examined based on these datasets because of the limited information available

to fully describe the sea surface elevation. The experimental work at Imperial College has already shown that this occurs in real water but it may prove to be difficult to observe this with *in situ* measurements of the real ocean without specially designing an instrument array to detect it (Latheef et al. 2017). This is unfortunate since it is probably the most significant of the non-linear changes predicted by theory since it (i) appears to occur at lower steepnesses than the other two changes; (ii) increases the inline kinematics which is important for loading on fixed structures.

We should consider causes other than weakly non-linear physics that could produce the results found in this paper.

One possibility is the local wind-wave interactions – a mechanism that has been explored as a possible mechanism for causing waves to deviate from the underlying Gaussian distribution (Kharif et al. 2008, Toffoli et al. 2017). Indeed, Agnon et al. (2005) have looked at very localised asymmetry properties and connected these with wind in the Lake George dataset. One problem, as noted by Adcock & Taylor (2014), is that steep and narrow-banded conditions will tend to be associated with strong wind, so it is often difficult to attribute the cause of any unusual observations. However, based on our results, the overall trend of local changes in the averaged shape of largest events for both datasets have strong correlations with the wave steepness, which is a direct measure of wave nonlinearity. Hence, in the present study, whilst we cannot rule out local wind effects, the clear trends in the data seem to fit better with the cause be non-linear physics than with local wind/wave interactions.

One of the reviewers suggested that wave breaking, which is also associated with large steepness, could account for the results observed in this paper. Wave breaking will introduce asymmetry but this will tend to be very localised to the extreme crest (see Myrhaug & Kjeldsen (1986), Babanin et al. (2007)). Studies such as Melville & Rapp (1988) track the envelope of breaking waves (for highly idealised laboratory conditions) – we have found it difficult to draw any general conclusions on the in-

fluence of breaking on asymmetry or group shape change from their work. Recent numerical simulations and experiments show that the tallest crest in a breaking water wave-group travels slower than expected (Banner et al. 2014, Barthelemy et al. 2018), which suggests that wave breaking might lead to deceleration of the largest crest. This could lead to opposite result to that observed in this work. Let us consider what happens as the large wave in a group breaks but remains large enough to be one of the largest five waves and so enter our analysis. The peak of the wave might move forward slightly, which would give some asymmetry consistent with our findings. This is plausible. But, because the size of the largest wave is reduced, the relative length of the group would expand rather than contract as we have observed in the field data (see Figure 3.14 and Figure 3.15). Thus, we think weak non-linear physics is a best explanation of the results as all the data appears consistent with this well-established theory.

According to Young et al. (2005), there is almost no disturbance from the experimental rig on the waves for the Lake George dataset. However, some of the North Sea rigs are substantial structures and diffraction could cause the wave measured following a giant wave to be bigger than it would have been otherwise. It is not straightforward to eliminate this from the analysis and so caution should be applied to the North Sea results. A further source of uncertainty is that the North Sea data are measured with wave radars, which will not perfectly reproduce the free surface. However, wave-structure interaction would not obviously explain the observed contraction in the width of the group relative to that predicted by linear theory and the asymmetry results seem consistent with those from Lake George. Therefore, we are confident in these findings.

Finally, we should comment on implications of the results of this paper. We feel there is clear evidence that the non-linear changes predicted by theory can occur in nature. Although, the horizontal asymmetries and wave-group contractions found in

these datasets provide an excellent insight of how non-linear physics would change the average shape of the largest events in directional spread real water, the Lake George data are much steeper and more narrow-banded than most storms in the open ocean. To further investigate these phenomena in open ocean, we applied almost the same techniques to the North Sea dataset. Although we saw the same effects in the North Sea, the changes are relatively small and only identifiable because we have a large dataset available. Thus the changes analysed in this paper, relative to that predicted by linear theory, will often be small in practice. Our conclusion here appears to be consistent with the previous study by Gemmrich & Thomson (2017). However, although the changes are small, we do think these are robust features of real ocean waves. The changes are perhaps not as dramatic as previously observed in numerical simulations probably due to the broader spectrum and increased directionality of real ocean waves. As noted above, the broadening of the crest – what Gibbs & Taylor called the 'wall of water' as reported by many mariners – is predicted to occur for lower steepnesses (Adcock et al. 2016), so may be more common in the ocean than that asymmetry and group contraction analysed here.

Chapter 4

Comparison of two versions of the MNLS with the full water wave equations

4.1 Abstract

Versions of the non-linear Schrödinger equation are frequently used for modelling the non-linear propagation of water waves. In this paper, we compare two models against the results of fully non-linear numerical simulations. We consider uni-directional versions of the non-linear Schrödinger equation of Dysthe (1979) with the hybrid model of Trulsen et al. (2000). The model of Trulsen et al. (2000) is shown to have clear advantages in all situations considered including modelling wave crest statistics for highly non-linear cases. However, for very broad bandwidths this model does start to break down, presumably due to the inherent limitation of the envelope representation of water waves. This in turn leads to a small, non-physical, leakage of energy in non-linear simulations, although, this leakage is much smaller than for the version with 5th order linear dispersion relationship.

4.2 Introduction

The propagation of steep ocean waves in deep water is weakly non-linear. Engineers and scientists need to understand this propagation as it alters the statistics of waves and ultimately the loading on structures.

Versions of the non-linear Schrödinger (NLS) equation have been widely used for investigating the non-linear propagation of gravity waves in deep water. This equation models the evolution of the complex wave envelope. The history of this equation is described in detail in Dysthe & Trulsen (2001). These equations are based on the expansion of two small parameters: the relative bandwidth and the wave steepness. The simplest form of these equations is the cubic NLS (see Shabat & Zakharov (1972), Benney & Newell (1967), Zakharov (1968) for details). This is very restrictive in bandwidth and was extended first by Dysthe (1979) and then by Trulsen & Dysthe (1996). Numerous other modifications have been made to include finite depth, multi-directional waves, surface tension, etc. The basic form of these equations is a linear part, which essentially models linear dispersion, and non-linear terms which account for finite amplitude effects. Trulsen et al. (2000) proposed a hybrid model where the linear dispersion part of the equation was solved 'exactly' (rather than by a perturbation expansion which has poor convergence properties). The purpose of this paper is to understand the practical differences between Trulsen & Dysthe (1996) and Trulsen et al. (2000) by comparing these to a fully non-linear model.

Different non-linear Schrödinger equations have been used extensively for studying the non-linear evolution of ocean waves. The equations are considerably simpler than the full water wave equations and have therefore been used as the basis for analytical studies by numerous authors (Mori & Janssen 2006, Adcock & Taylor 2009*b*, Adcock et al. 2012, Fedele 2015*b*, Janssen & Janssen 2019). Similarly, the relatively quick computational times have led to NLS equations being used to investigate waves

numerically (Lo & Mei 1985, Onorato et al. 2001, Socquet-Juglard et al. 2005, Toffoli et al. 2010, Adcock et al. 2015, Cousins et al. 2019). These modified NLS equations are also widely used for solitary groups and modulationally unstable patterns and compared with primitive water equations and laboratory experiments by Slunyaev and his colleagues (Slunyaev 2009, Slunyaev et al. 2013, Slunyaev & Shrira 2013, Slunyaev & Pelinovsky 2019).

However, the approach of Trulsen et al. (2000) appears to be rarely used. Relatively few studies have directly compared the NLS equations to solutions of the fully non-linear equations. Without these, it is difficult to assess how accurate the modified non-linear Schrödinger (MNLS) equations are for understanding waves with the sorts of bandwidths observed in the open ocean. An early effort in this direction was the work of Henderson et al. (1999). Adcock & Taylor (2009*b*) recently compared numerical solutions to the exact water wave equations with the cubic and NLS with fifth order linear dispersion for focusing wave-groups. Both uni-directional (Adcock & Taylor 2016*c*) and directional (Adcock & Taylor 2016*a*) results were reported. The key conclusion from these studies was that the bandwidth limitation is more significant than some authors realise as the 'fast' spectral changes around a large wave lead to an effective increase in the bandwidth. The present study is restricted to uni-directional waves but extends these studies to consider the model of Trulsen et al. (2000) as well as presenting some new comparisons for random waves.

We should stress that in the open ocean waves are not uni-directional and this makes a fundamental difference to their non-linear dynamics (for instance see arguments in Adcock & Taylor (2014)). Nevertheless, uni-directional waves are widely studied in the laboratory and numerically and form the starting point for investigating real ocean waves. Additionally, recent studies suggest that in the case of directional spread sea, there was a nonlinear evolution towards a wall of water, approaching the case of one-dimensional propagation (see Adcock et al. (2015), Barratt et al. (2021)

for details).

4.3 Numerical models

4.3.1 Fully non-linear model

In the present paper we use OceanWave3D (see Engsig-Karup et al. (2009) for details) to solve the standard fully non-linear water wave equations. We take considerable care to ensure the numerical accuracy of the solution following Barratt, Bingham & Adcock (2020). This numerical method solves the governing equations of potential flow throughout the domain with fully non-linear boundary conditions using finite difference method. There is a built-in local smoothing filter, which activates when the vertical acceleration is larger than 0.4 m/s^2. This filter is only applied for the irregular wave simulations (mainly in section 5), as we do not report any case with breaking for the deterministic wave groups. We use a water depth of 500 m (corresponding to $k_0 d = 14$). We use such a high value to ensure that the return current under the wave-group is not significantly affected by the water depth (see van den Bremer & Taylor (2015)) as the relevant length scale for this flow is that of the wave group and not the individual waves. We use a resolution of 30 points per wavelength and 22 points vertically in the water column, which are not uniformly distributed (see Engsig-Karup et al. (2009) for detailed guidance on the convergence of the vertical resolution). The time-step is 0.1 seconds (120 per period). For the deterministic groups we 'correct' initial conditions using the second order theory of Sharma & Dean (1981) and also add in an approximate third order correction following Barratt, Bingham & Adcock (2020).

In this paper we assume that these results represent the exact evolution of the waves and use it as a base case to compare approximate models with.

4.3.2 MNLS equation

In the present paper we consider numerical solutions to both Trulsen & Dysthe (1996) and Trulsen et al. (2000). We describe these as the 'T1996' or 'T2000' respectively.

These equations model the complex wave envelope, A, of the waves. We calculate the free surface from the complex envelope using the following:

$$\eta = \Re(\eta_{linear} + \eta_{2-} + \eta_{2+} + \eta_3), \tag{4.1}$$

where

$$\eta_{2-} = \frac{1}{2\omega}\frac{\partial\phi}{\partial x} - \frac{1}{16k}\frac{\partial^2 |A|^2}{\partial x^2}, \tag{4.2}$$

$$\eta_{linear} = A\exp(i(kx - \omega t)), \tag{4.3}$$

$$\eta_{2+} = \left[\frac{kA^2}{2} - \frac{iA}{2}\frac{\partial A}{\partial x}\right]\exp(2i(kx - \omega t)), \tag{4.4}$$

$$\eta_3 = \frac{3}{8}k^2A^3\exp(3i(kx - \omega t)). \tag{4.5}$$

The MNLS equation with 5th order linear dispersion relationship is given by

$$\begin{aligned}\frac{\partial A}{\partial t} + \frac{\omega}{2k}\frac{\partial A}{\partial x} + i\frac{\omega}{8k^2}\frac{\partial^2 A}{\partial x^2} - \frac{\omega}{16k^3}\frac{\partial^3 A}{\partial x^3} \\ - i\frac{5\omega}{128k^4}\frac{\partial^4 A}{\partial x^4} + \frac{7\omega}{256k^5}\frac{\partial^5 A}{\partial x^5} = -\frac{i\omega k^2}{2}A|A|^2 - \\ \frac{3}{2}\omega k|A|^2\frac{\partial A}{\partial x} - \frac{1}{4}\omega kA^2\frac{\partial A^*}{\partial x} - ikA\left.\frac{\partial\phi}{\partial x}\right|_{z=0},\end{aligned} \tag{4.6}$$

where k is the carrier wavenumber here taken as identical to k_0, A^* is the complex conjugate of the envelope. The linear dispersion relation relates ω to k_0 . The left hand side of equation 4.6 is a narrowbanded approximation of linear dispersion. The right hand gives a non-linear modification to linear dispersion. The influence of the return current on the wave envelope is given by the final term in equation 4.6. This is most easily found using the Hilbert transform method described in Janssen (1983).

The potential ϕ is given by

$$\left.\frac{\partial \phi}{\partial z}\right|_{z=0} = \frac{\omega}{2}\frac{\partial |A|^2}{\partial x}, \tag{4.7}$$

and the potential satisfies Laplace's equation in the fluid, thus

$$\nabla^2 \phi = 0, \tag{4.8}$$

with ϕ tending to zero in the limit as $z \to -\infty$. Although we do not present details in the present paper it is possible to extend the non-linear terms to higher orders of steepness (Debsarma & Das 2005). However, for practical values of steepness these extra terms appear to be negligible and so are not analysed in detail here.

The linear part of the fifth order MNLS is solved using a pseudo-spectral scheme. The non-linear terms (right-hand part of equation 4.6) are evaluated in the spatial domain. A spatial resolution 23 points per wavelength is used. The hybrid scheme proposed by Trulsen et al. (2000) replaces the linear part of this solution with an 'exact' linear solver, which is solved spectrally. The exact linear operator used to solve the linear part (left-hand part of equation 4.6) can be written as an integral in the physics space:

$$\frac{\partial A}{\partial t} + \frac{1}{2\pi}\int\int_{-\infty}^{+\infty} i[\omega(k_0+\lambda)-\omega_0]e^{i\lambda\dot{(}x-y)}A(y,t)dy\ d\lambda = 0, \tag{4.9}$$

where λ is the modulation wave vector relative to k_0. As noted by Trulsen et al. (2000), this scheme is no more computationally demanding than solving the T1996. There are significant practical differences between the two models which are discussed below in this paper. However, we note here that for the same case and with the same numerical resolution we find that the T2000 model typically conserves 'energy' (i.e. the integral over the domain of $|A|^2$) at least an order of magnitude better than the

T1996.

4.4 Linear dispersion

We start by briefly considering linear dispersion without any non-linear physics. Solving for exact linear dispersion is straightforward and this can be used as the basis for comparison. We choose to analyse the propagation of a Gaussian wave group which has the form at the focus point

$$\eta(x) = A_{max} e^{-\frac{1}{2}S_x^2 x^2} \cos(k_0 x) \tag{4.10}$$

where, S_x is the bandwidth of the group. This wave group has a underlying Gaussian spectrum in the form of:

$$Sp(k) = \exp\Big(\frac{-(k-k_0)^2}{2S_x^2}\Big), \tag{4.11}$$

where the bandwidth of the Gaussian spectrum S_x shares the same value as the bandwidth of the group. In this case, for linear evolution, the amplitude, A_{max} is not important. We start with a perfectly 'focused' group shown in equation 4.10 as initial condition and allow this to propagate for 100 periods. We then simply evaluate the difference between the waves given by exact linear dispersion and the waves being predicted by both T1996 model and T2000 model with linear terms using the linear regression. The coefficient of determination is defined as:

$$R = 1 - \frac{\sum_{i=1}^{n}(\eta_{exact,linear,i} - \eta_{simulated,linear,i})^2}{\sum_{i=1}^{n}\eta_{exact,linear,i}^2}, \tag{4.12}$$

where η_{exact} is the exact linear surface elevation profile predicted by linear theory and $\eta_{simulated}$ is the simulated linear surface elevation profile.

Figure 4.1 presents the correlations for the linear part of the wave groups with different bandwidths after 100 periods of evolution. Simulations beyond 100 periods

have no further impact on the overall trend shown herein. For narrow bandwidths both T1996 and T2000 models give almost identical results to the exact results. As bandwidth increases a discrepancy develops between the T1996 solution and exact linear dispersion and the error appears to grow rapidly. This is unsurprising as the T1996 result is derived assuming a narrow bandwidth and is known to have poor convergence properties for larger bandwidths. Perhaps more surprisingly, for larger bandwidths we also find a discrepancy emerging between T2000 and the exact results. This result is not dependant on numerical resolution used. We suggest this comes from a fundamental limit to the envelope representation of a wave. When the bandwidth is a comparable value to the carrier wavenumber the envelope representation breaks down and hence so does the T2000 linear dispersion model. Despite this, it is clear that T2000 gives a far more accurate representation of linear dispersion than the T1996 model.

4.5 Non-linear wave group simulations

To investigate non-linear dispersion we first investigate the focusing of deterministic wave-groups. This approach has been used by numerous authors as a way of examining the physics during the formation of large waves (Lo & Mei 1985, Baldock et al. 1996, Gibson & Swan 2006, Gibbs & Taylor 2005). We use NewWave type groups which, under linear evolution, are the average shape of an extreme event in the open ocean (Lindgren 1970, Boccotti 1983, Tromans et al. 1991, Tang, Tromans & Adcock 2019). Following past studies we conceptually start with a group which, under linear evolution, would form a perfectly focused NewWave group 80 periods later. This focusing is modified in our simulations by the non-linear physics. We continue the simulations past focus to capture the de-focusing process. Thus we capture essentially the same 'fast' non-linear physics as occurs as a large wave forms from

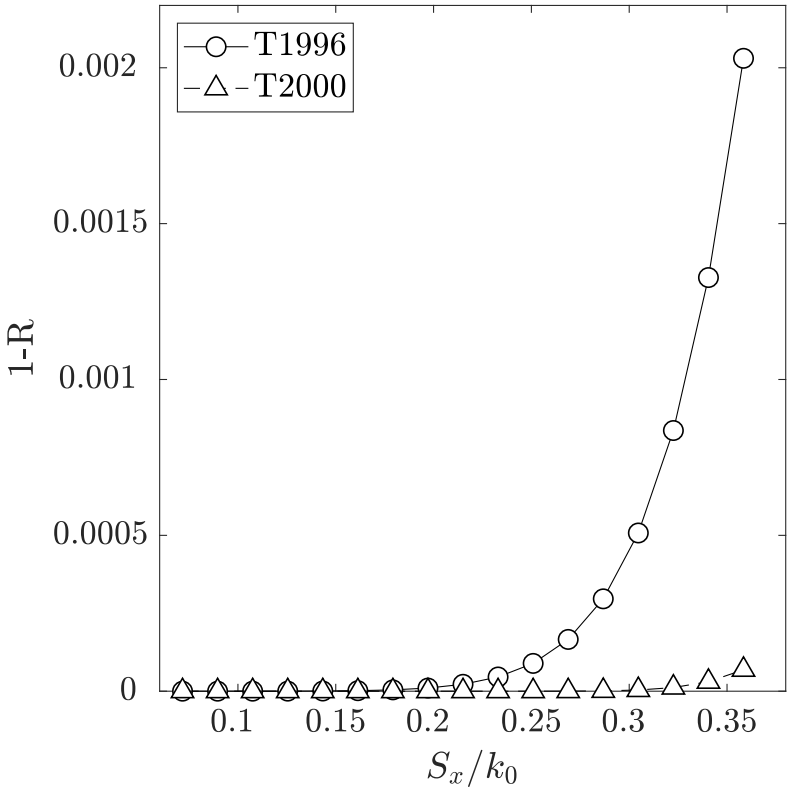

Figure 4.1: Linear regression coefficient comparison between MNLS with fifth order approximated linear dispersion and exact linear dispersion with only linear parts considered after the wave group having propagated for 100 periods.

a random background. We classify the simulations, following Gibbs (2004) by the steepness ($A_{max}k_0$) that the group would have had at focus under linear evolution. See Baldock et al. (1996), Gibbs (2004), Adcock & Taylor (2016*c*) for more details of this approach and the physics. We use a Gaussian spectrum with Sx/k_0 = 0.165 (following Gibbs (2004)). The peak period of the wave groups presnted in this section is 12 seconds. This spectrum is a narrowbanded approximation to the peak of a JONSWAP spectrum with γ = 3.3.

We are primarily interested in the free waves in this section. The MNLS gives these directly (approximately see Gramstad & Trulsen (2011)) without employing equation 4.2, 4.4 and 4.5. However, the full non-linear scheme has bound harmonics. To remove these, we run four simulations using 4 different phases 0, 90, 180 and 270 of the waves in the wave-group. Addition and subtraction is then used to isolate the free waves. See the in depth discussion of the method in Adcock et al. (2019). We note that previous work along the same lines by Gibbs (2004), Adcock & Taylor (2016*c*) only used 2 phases to extract free waves – as part of this work we found the spectrum of the largest cases was too broad for this, suggesting minor inaccuracy in these previous studies.

4.5.1 Spatial profiles

Figure 4.2 presents the spatial shape of the envelope of the linearised wave groups at 'focus' (i.e. when the maximum point of the envelope is at its largest).

Two cases are presented. For a group with $A_{max}k_0$ = 0.12 the significant change in the shape, and increase in elevation of the wave-group caused by the non-linear physics is captured well by all three non-linear models. For the more non-linear case $A_{max}k_0$ = 0.18, the non-linear changes are more dramatic. The T2000 agrees very closely with the fully non-linear simulations, although there is a small discrepancy at the crest. There is no obvious feature which is unphysical. However, the T1996

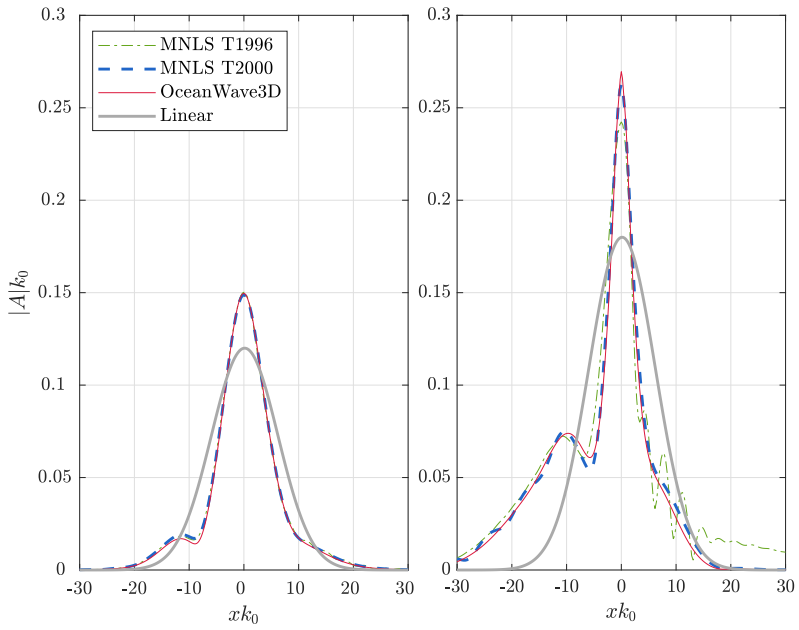

Figure 4.2: Wave envelope comparison between MNLS with fifth order approximated linear dispersion, exact linear dispersion, fully nonlinear OceanWave3D and envelope shape under linear evolution for $A_{max}k_0 = 0.12$ (left) and $A_{max}k_0 = 0.18$ (right).

significantly underestimates the peak of the wave envelope. It also has wiggles at the front of the wave-group which are clearly unphyiscal. These are discussed in detail in Adcock & Taylor (2016*c*).

4.5.2 Spectral changes

These changes can be explored in more detail by looking at the spectrum of the surface elevation profile of the wave-groups. Figure 4.3 presents the initial spectrum, the spectrum at 'focus' and the spectrum at the end of the run after the group has de-focused. Under linear evolution these would remain unchanged.

For the $A_{max}k_0 = 0.12$ all the non-linear models produce consistent results with very minor differences. On a linear scale there are no obviously unphysical changes in the location of the energy. For $A_{max}k_0 = 0.18$ there are significant differences. The differences are particularly obvious after focusing where there is clear structure not captured by either models. However, the T2000 model does significantly better than

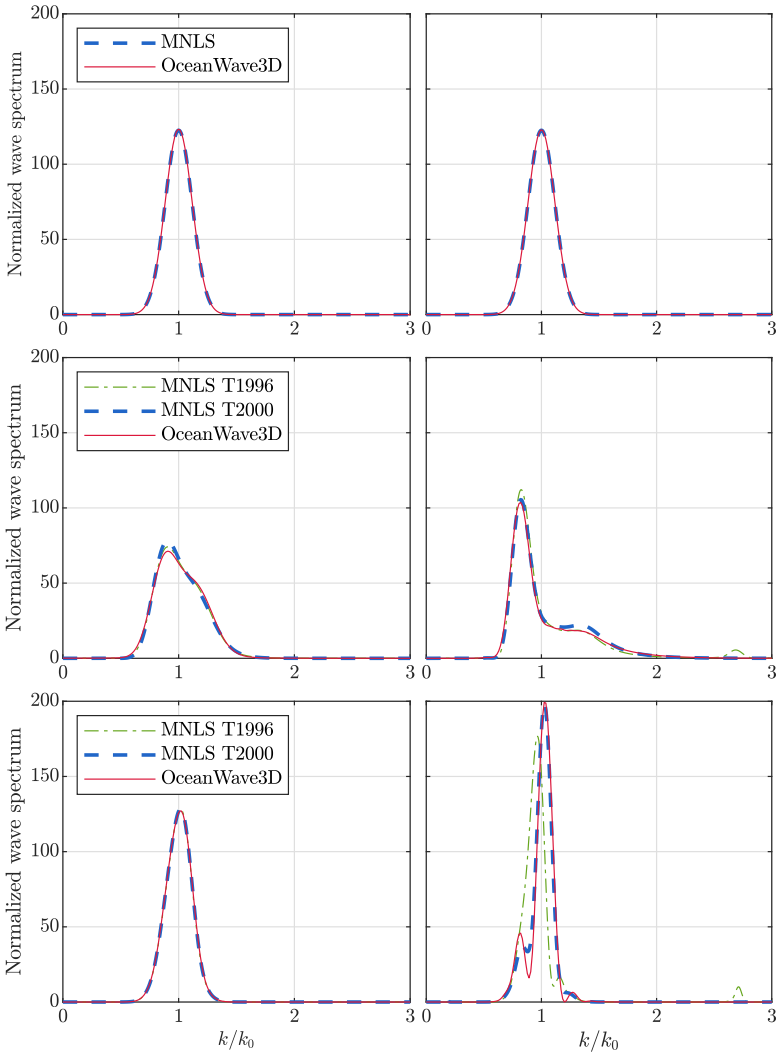

Figure 4.3: Wave variance density spectrum comparison at start (top), nonlinear focus (middle) and after $80T_p$ (bottom) between MNLS with fifth order approximated linear dispersion, exact linear dispersion and fully nonlinear OceanWave3D for $A_{max}k_0$ = 0.12 (left) and $A_{max}k_0$ = 0.18 (right).Wave spectra are normalised by first moment and k_0.

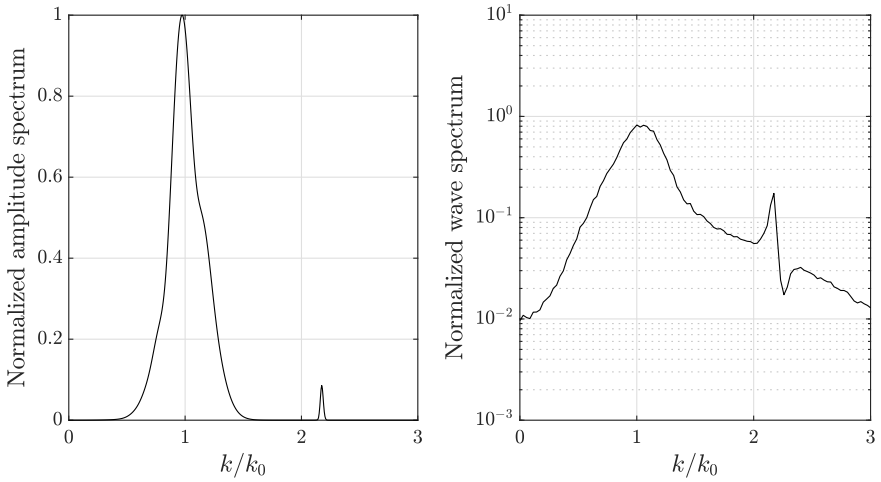

Figure 4.4: Normalised wave amplitude spectrum based on T2000 model for Gaussian wave group with $Ak_0 = 0.16$ at $80T_0$ (left) and wave spectrum for irregular waves after $300T_0$ for case 3 (right). Wave spectrum is normalised by first moment and k_0.

the T1996, particularly for the upshift in the spectral peak. There are also unphysical transfers of energy. From Fig. 4.3, there is unphysical energy transfer for T1996 model, which give rise to the extra peak around 3 times of the peak wavenumber especially for $A_{max}k_0 = 0.18$. This is likely due to well-known leakage of energy described in the works of Martin & Yuen (1980), Yuen & Ferguson (1978). Although not clear on the linear scale, there is also a much smaller energy transfer to $2.1k_0$ for the T2000 results. This is presumably due to interactions between the non-linear terms and the small inaccuracy in the linear dispersion term. We demonstrate this energy leakage in Fig. 4.4 where we plot the amplitude spectrum for the non-linear evolution of a wave-group as well as a power spectrum for random waves (described in detail in the next section). We should emphasise two points. Firstly, this energy leakage is not a numerical artefact – the magnitude does not change when the resolution is changed. Secondly, although T2000 still has energy leakage, this is an order of magnitude smaller than for the T1996.

4.6 Wave crest statistics

Finally we consider random wave simulations. We choose to focus on wave-crest statistics which are perhaps the most easily understood and are important in ocean engineering practice. We calculate the free surface from the MNLS using equation (4.1).

We consider the three cases based on the work of Onorato et al. (2001). The initial parameters for the three cases considered are given in Table 4.1. The wave steepness is obtained as:

$$Steepness = \frac{2\pi}{g}\frac{H_s}{T_z^2}, \tag{4.13}$$

where g is the gravitational acceleration, T_z is the zero-crossing period and H_s is significant wave height. The initial spectrum in all cases is a JONSWAP spectrum. The peak period for all the cases presented in this section is 1.5 seconds. In the fully non-linear simulations, waves are created using a relaxation zone at one end of the domain. The MNLS simulations are run somewhat differently. A periodic domain is used and initial waves are specified in space rather than time. The spatial periodicity for MNLS simulations assumes energy conservation and spatial homogeneity in the wave field as opposed to the time homogeneity for fully nonlinear simulations. The latter assumption restricts the initial evolution of wave statistics in time domain. However, both assumptions should have minor impacts on the wave statistics at steady-state. In both cases the random waves are generated following Tucker et al. (1984). There is an initial transient in the wave statistics as correlations form between components and the spectrum changes before an equilibrium appears to form. We choose to analyse statistics after this equilibrium develops. We note that there is a small amount of wave breaking in the fully non-linear simulation, which is not modelled in the MNLS simulations. This does not appear to have a major impact on

Table 4.1: Parameters of the three test cases

Cases	**JONSWAP γ**	$\boldsymbol{H_s}$ **(m)**
1	1.0	0.11
2	3.3	0.14
3	6.0	0.16

the results.

Figure 4.5 presents wave crest exceedance statistics for the different models (with the Rayleigh distribution included for reference). For both MNLS models we run sufficient simulations that the error bars are very small and so we neglect these for clarity. For all three cases, although Case 3 is exceptionally non-linear, there is excellent agreement between T2000 and the fully non-linear simulations. However, the T1996 significantly underestimates the size of the extreme waves presumably due to the same process as observed in the deterministic wave-groups studied above.

4.7 Conclusions

In this study we have examined the effects which lead to inaccuracy of two NLS based models for water wave propagation against a fully non-linear solver. Our study is confined to uni-directional waves. Using the 'exact' linear dispersion relationship presented in Trulsen et al. (2000), we find that the T2000 model can predict the linear part of focused wave group significantly more accurate for broad banded cases. This enables T2000 model to capture the envelope shape at the non-linear focus up to $A_{max}k_0$ = 0.18, whereas the T1996 tends to under estimate the envelope peak. The T2000 model can also capture the spectral evolution precisely, when compared with the fully nonlinear solver. In terms of the wave crest statistics, T2000 model shows excellent agreement with the fully non-linear solver, whereas the T1996 seems to underestimate the probability of extreme waves especially for cases with steepness over 0.046. These numerical tests indicate that T2000 is significantly more accurate than

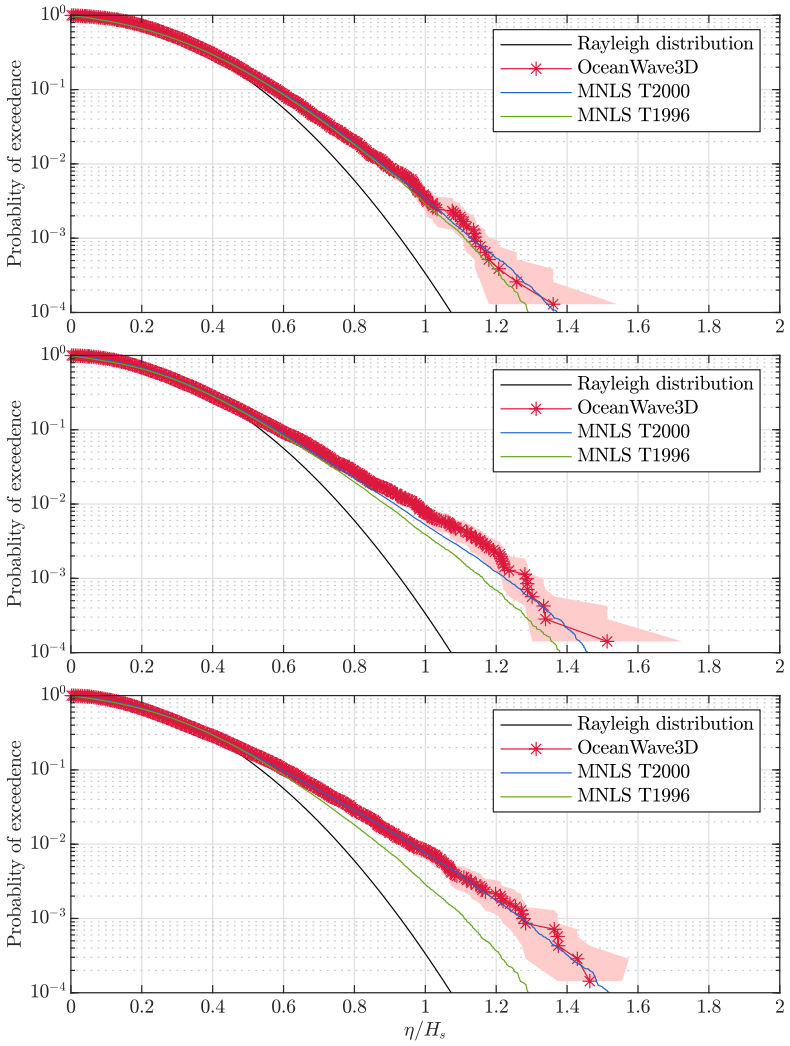

Figure 4.5: Wave crest exceedance comparison at equilibrium state between MNLS with fifth order approximated linear dispersion and exact linear dispersion and fully nonlinear OceanWave3D for Case 1 (top), Case 2 (middle) and Case 3 (bottom). Shades represent the 95% confidence interval based on bootstraping. We ran 185 simulations with different random seeds for each case with MNLS code, the confidence interval is small at the scale we are looking at, so it is removed for clarity.

the T1996. We attribute this to bandwidth limitations in the latter. We highlight conclusions from past studies (Adcock & Taylor 2016*a*,*c*) that it is misleading to use the bandwidth of any input spectrum. This is because the non-linear physics causes the wave-group around an extreme wave to contract in x which is analogous to a localised increase in the spectral width.

Although T2000 is significantly more accurate than the T1996, it does still suffer from an energy leakage to high wavenumbers. This is an order of magnitude smaller than that from the T1996. This leakage is presumably due the the inherent bandwidth limitation of any envelope model which means linear dispersion cannot be captured exactly as the length scale of the envelope approaches that of the carrier wave.

As Trulsen et al. (2000) note, their model is no more computationally expensive than the T1996 model. As such, it is strongly recommended to use this model over T1996.

Chapter 5

Spatial evolution of the kurtosis of steep unidirectional random waves

5.1 Abstract

We study the evolution of unidirectional water waves from a randomly forced input condition with uncorrelated Fourier components. We examine the kurtosis of the linearised free surface as a convenient proxy for the probability of a rogue wave. We repeat the laboratory experiments of Onorato *et. al.* (2004, *Phys. Rev. E* **70** (6), 067302), both experimentally and numerically, and extend the parameter space in our numerical simulations. We consider numerical simulations based on the Modified Nonlinear Schrödinger equation (MNLS) and the fully nonlinear water wave equations, which are in good agreement. For low steepness, existing analytical models based on the Nonlinear Schrödinger equation (NLS) are found to be accurate. For cases which are steep or have very narrow bandwidths, these analytical models over-predict the rate at which excess kurtosis develops. In these steep cases, the kurtosis in both our experiments and numerical simulations peaks before returning to an equilibrium level. Such transient maxima are not predicted by NLS-based

analytical models. Above a certain threshold of steepness, the steady-state value of kurtosis is primarily dependent on the spectral bandwidth. We also examine how the average shape of extreme events is modified by nonlinearity over the evolution distance, showing significant asymmetry during the initial evolution, which is greatly reduced once the spectrum has reached equilibrium. The locations of the maxima in asymmetry coincide approximately with the locations of the maxima in kurtosis.

5.2 Introduction

The simplest model of free surface gravity waves assumes linear dynamics. In this model, and if the free surface is modelled as a Gaussian random process, wave amplitudes are well approximated by a Rayleigh distribution (Longuet-Higgins 1952), and the average shape of large waves is given by the scaled autocorrelation function (Lindgren 1970, Boccotti 1983). Nonlinearity leads to modifications in the wave statistics and to the shape of extreme wave events. In particular, when the fluid is deep and the waves are unidirectional, the Benjamin-Feir instability (Benjamin & Feir 1967) leads to more large waves than would be expected from the linear model as correlations develop between Fourier components. A convenient parameter to describe the increased number of large waves is the kurtosis (or excess kurtosis) of the free surface (Mori & Janssen 2006) (see also the experimental results in Onorato et al. (2004), where the rogue wave density is found to have a strong correlation with the kurtosis.) Here, we examine how the kurtosis and average shape of an extreme wave evolve in space from a Gaussian random input condition without correlation between components. The over-arching objective is to improve our understanding of the nature of the non-linear physics of surface gravity waves and its impact on wave statistics.

Analysis of this problem originates from Janssen (2003). In this paper the wave field is assumed to be sufficiently weakly nonlinear that the sea surface is in a near-

to-Gaussian state, meaning that the kurtosis can be expressed in terms of lower-order moments, following the approach of Hasselmann (1962). Ensemble averaging of the Zakharov equation and the assumption of spatial homogeneity provides closed-form expressions for the fourth cumulant and the evolution of kurtosis (see Eq. (20) and Eq. (28) of Janssen (2003)). Mori & Janssen (2006) additionally invoked the assumption of narrow bandwidth, which is consistent with the cubic Nonlinear Schrödinger equation (NLS), and assume an underlying one-dimensional Gaussian spectrum.

For unidirectional waves, the result by Mori & Janssen (2006) implies an excess kurtosis that increases with a time scale dependent on the bandwidth of the initial spectrum and the dominant wave period (Janssen & Herbers 2009, Mori et al. 2011). Following monotonic increase from zero, the dynamic excess kurtosis then levels off tending to a value of $\pi/(3\sqrt{3})$ times the Benjamin-Feir Index (BFI) squared, where the BFI is a measure of the ratio of the significant wave steepness to the bandwidth of the waves (Mori & Janssen 2006). Fedele (2014) obtained an equation for the evolution of excess kurtosis using the one-dimensional (1D) cDZ equation of Fedele & Dutykh (2012), based on Dyachenko & Zakharov (2011). The cDZ equation is valid for weakly nonlinear four-wave interactions like the NLS, but does not have any constraints on the spectral bandwidth unlike the NLS (Fedele 2014). The excess kurtosis based on the cDZ equation is generally less than predicted by the equivalent NLS expression, and the reduction is greater for greater bandwidth (see Eq. (D16) of Fedele (2014)).

It is important to distinguish conceptually, on the one hand random wave fields that are homogeneous in space and evolving and thus non-stationary in time and on the other hand those that are stationary in time and evolving and thus inhomogeneous in space. The aforementioned authors (Janssen (2003), Mori & Janssen (2006), Janssen & Herbers (2009) and Mori et al. (2011) and Fedele (2014)) all examined the evolution of spatially homogeneous fields in time. Using analogous premises

to Mori & Janssen (2006), an equation for the spatial evolution of temporally homogeneous waves has been derived by Fedele et al. (2010). In the narrow-banded limit, the results of Mori & Janssen (2006) and Fedele et al. (2010) are equivalent, and mapping between space and time takes place using the group velocity (see also Chabchoub & Grimshaw (2016)), noting that the bandwidth in frequency is half the bandwidth in wavenumber (Fedele et al. 2010). For the broad-banded cDZ, mapping between space and time is not straightforward, and we are unaware of any authors describing the evolution of kurtosis in space for temporally homogeneous waves based on the cDZ, although a spatial of the Zakharov equation exists (Shemer et al. 2001, Kit & Shemer 2002).

Laboratory measurements are almost exclusively made in the time domain at a finite number of wave gauges (Fedele et al. 2010, Zhang et al. 2014, 2016, Onorato et al. 2004, 2006, Kokorina & Slunyaev 2019). Apart from the time-domain experiments, spatio-temporal measurements are beginning to be successfully made in the laboratory using stereo-imaging techniques (Zavadsky et al. 2017) commonly applied in the field (e.g. Fedele et al. (2013)). A restriction of many of the validation studies in the laboratory is the length of the experiments, meaning that only the initial stages of the evolution have been compared to theory. We focus herein on the laboratory experiments of Onorato et al. (2004) (henceforth O04), which were compared to the theoretical results of Mori & Janssen (2006) by Mori et al. (2007) and to numerical simulations of the Dysthe equation (Dysthe 1979) by Onorato et al. (2005). In doing so, we consider evolution of temporally homogeneous (or stationary) waves in space.

Previously, Shemer & Sergeeva (2009) presented experimental results for the spatial evolution of wave statistics along the wave tank for unidirectional random waves, and their observed short-term probability distributions were well predicted by the third-order model of Tayfun & Fedele (2007). Shemer, Sergeeva & Slunyaev (2010) compared these experimental results with the cubic NLS and the MNLS, and MNLS

could provide satisfactory predictions of individual groups in the time domain as well as statistical parameters. Shemer, Sergeeva & Liberzon (2010) further examined the impact of the initial spectral width on the evolution of the wave spectrum, and wave statistics. Slunyaev & Sergeeva (2012) examined the phase correlation during the initial stage of evolution and its connection with the evolution of wave statistics.

In nature, storm waves are directionally spread, and this makes a fundamental difference to the nonlinear physics. Janssen & Herbers (2009) first showed that kurtosis is generally reduced by directional spreading. Extending the results by Mori & Janssen (2006) based on the unidirectional NLS, Fedele (2015*b*) then showed analytically that the normalised excess kurtosis of directionally spread waves reaches a maximum and eventually tends monotonically to zero as the wave field reaches a quasi-equilibrium, finding good agreement with the experimental data of Onorato, Waseda, Toffoli, Cavaleri, Gramstad, Janssen, Kinoshita, Monbaliu, Mori, Osborne et al. (2009) and numerical simulations by Toffoli et al. (2010) (see also Xiao et al. (2013), Annenkov & Shrira (2009)). Based on the 2D+1 NLS, the result for directionally spread waves by Fedele (2015*b*) thus paints a very different picture with large values of kurtosis and associated rogue waves being transient (see Janssen & Janssen (2019) for further discussion of the asymptotics).

In this paper, we perform both new laboratory experiments and numerical simulations to examine how the kurtosis of unidirectional water waves evolves over relatively long distances. We compare our experimental and numerical results with the experiments of O04 and with the theoretical solutions for the evolution of the kurtosis of unidirectional random waves based on the NLS by Mori & Janssen (2006) and based on the cDZ by Fedele (2014). Throughout, we cite Mori & Janssen (2006), although full details of the evolution are developed in Janssen & Herbers (2009) and Mori et al. (2011) and we actually compare to the closed-form solution given in Fedele et al. (2010) (see appendix B). These theoretical solutions, with the exception of

Fedele et al. (2010), are valid for waves evolving in time, and we convert these to waves evolving in space using relationships only valid in the narrow-banded limit (Fedele et al. 2010). Our numerical simulations are performed using two methods: OceanWave3D (Engsig-Karup et al. 2009), which solves the fully nonlinear (potential-flow) water wave equations, and the Modified Nonlinear Schrödinger (MNLS) model of Trulsen et al. (2000).

For the steep cases, both our experiments and our numerical simulations (using both methods) show that the kurtosis peaks before returning to an equilibrium level. Such transient maxima are not predicted by the NLS-based and cDZ-based analytical models of Mori & Janssen (2006) and Fedele (2014). We note that these transient maxima in kurtosis can also be observed in the experiments of O04 and analogous simulations of the Dysthe equation (Dysthe 1979) by Onorato et al. (2005), where maxima are noted but not explored in detail. Peaks in kurtosis are also found in numerical simulations of the NLS by Onorato et al. (2016), where the evolution of kurtosis is linked to the evolution of the spectral bandwidth. Using the MNLS, we study how the properties of these transient maxima for unidirectional waves depend on input steepness and bandwidth. We note that this pattern of the kurtosis reaching a peak is superficially similar to what is predicted and observed for directionally spread behaviour (Onorato, Waseda, Toffoli, Cavaleri, Gramstad, Janssen, Kinoshita, Monbaliu, Mori, Osborne et al. 2009, Toffoli et al. 2010, Fedele 2015*b*), except there the kurtosis is predicted to return to zero at long distances (Janssen & Janssen 2019), although this may not always be observed (Xiao et al. 2013). We also examine how the expected shape of an extreme wave group evolves over the same distance. To our knowledge, this has not previously been examined for this problem. We find significant asymmetry during the initial evolution. Locations of maxima in asymmetry coincide approximately with the locations of the maxima in kurtosis. We emphasise our paper is confined to unidirectional waves, so that its findings cannot readily be

extended to realistic ocean waves. Nevertheless, investigating unidirectional waves as a limiting case of directionally spread waves is useful for three reasons. First, it will help elucidate the nonlinear physics at work, especially over longer distances. Second, it will inform offshore engineering models tests, which are still often conducted in unidirectional waves. Third, some extreme events can be quite similar in behaviour to the unidirectional limit (e.g. Adcock et al. (2015)). Furthermore, there are analogues between 1D waves and those in other media such as optical fibres (Dudley et al. 2019).

5.3 Methods

5.3.1 Experimental set-up

The experiments were carried out in the Multifunction Towing Tank at Shanghai Jiao Tong University. The tank is 300 m by 16 m and has a flat bed with a water depth of 7.5 m, giving a non-dimensional water depth for our experiments of $k_0 d = 13.4$ based on the spectral peak. There are 40 hinged-flap type wavemakers at one end of the flume. Linear wave generation theory was applied, and the impact of second-order error waves on the overall wave statistics was analysed carefully and found not to significantly affect the results. There was a parabolic beach at the far end of the flume opposite the wavemakers. Reflection analysis suggests that less than 3% of the energy is reflected. The wave surface elevation was measured by 10 capacitance probes at 100 Hz with excellent calibration characteristics. However, the wave probes could only be installed on a movable carriage. To track the wave evolution over a wider range, the experiments were repeated with different carriage positions. Irregular wave repeatability tests showed very consistent wave statistics at the same position over five repeats.

5.3.2 Numerical methods

We use two numerical modelling approaches in this paper. First, we solve the fully nonlinear potential flow equations for water waves using OceanWave3D (Engsig-Karup et al. 2009). The numerical wave tank length in the wave propagation direction is 778 m, which we discretise with 10242 nodes giving a spatial resolution of 0.076 m. The water depth of the wave tank is 7.5 m covered by 15 clustered nodes. Care has been taken, following the approach of Barratt, Bingham & Adcock (2020), to ensure sufficient resolution to accurately capture the nonlinear physics. The simulation time is 1920 s – identical to the experiments. Waves are generated using a relaxation zone at the start of the domain and are absorbed by a damping zone at the end of the flume. A wave breaking model is applied, which is triggered by downward Lagrangian particle accelerations on the free surface that are larger than $0.4g$. We followed experiments results in Ochi & Tsai (1983) in terms of the wave breaking criterion. After determining these breaking events, a filter is applied to the local free surface region to remove energy from the waves until the downward particle accelerations are below the threshold.

We also use a faster code which solves the Modified Nonlinear Schrödinger (MNLS) model of Trulsen et al. (2000). Using this MNLS model, which unlike our fully nonlinear model does not capture wave breaking, should give us confidence that the physics we are observing is not influenced by wave breaking. The fast computation times also mean that the MNLS results can be run repeatedly, reducing the uncertainty in the estimated kurtosis sufficiently that confidence bands associated with statistical variability are not required. The model set-up is similar to that above with waves generated using a relaxation zone and absorbed at the far end of the domain. We use a very high (for an envelope model) spatial discretion of 23 points per wavelength. In addition to standard numerical checks, we can also test this code for energy conservation by studying the related problem of the evolution of a 'sea state' covering the

entire domain with wrap-around boundary conditions at the end. For this problem with the same initial spectrum, energy loss over 100 wave periods is less than 0.5%.

5.4 Results: kurtosis

In this section, we first compare the experimental results and numerical simulations with both analytical results for the evolution of normalised dynamic excess kurtosis based on the NLS (Mori & Janssen 2006) and the cDZ (Fedele 2014) equations and the experiments of O04. To broaden the parameter space, we further examine a range of input 'sea-state' parameters using the MNLS with Gaussian input spectra.

5.4.1 Comparison with the experiments of O04

'Sea-state' parameters

We choose to study cases based on the seminal experiments of O04. These are summarised in table 5.1. The cases are based on the JONSWAP spectrum with different peak enhancement factors. The Benjamin-Feir Index (BFI) is computed following the method recommended in Serio et al. (2005), who base their recommendation on a review of different methods:

$$\text{BFI} = \sqrt{m_0} k_0 Q_p \sqrt{2\pi}, \tag{5.1}$$

where $m_0 = H_s^2/16$ is the zeroth moment of the energy spectrum with H_s the significant wave height, k_0 is the peak wavenumber, and Q_p is a dimensionless parameter that describes the spectral bandwidth (Goda 2000). The parameter Q_p has less sensitivity to the high-frequency tail of the spectrum (and cut-off frequency) than other

Table 5.1: 'Sea-state' parameters of the three test cases measured at the first probe, with T_0 the peak period, $\nu = \sqrt{m_0 m_2/m_1^2 - 1}$ the bandwidth parameter, where m_n are nth-order spectral moments of the variance density spectrum $S(\omega)$ in angular frequency ω. Also shown are the 'sea-state' parameters of the experiments of O04 calculated with the method used herein (in round brackets) and the 'sea-state' parameters of the experiments of O04 as reported in their table 1 [in square brackets].

Case	γ	H_s [m]	$\varepsilon = k_0 H_s/2$	ν	Q_p	BFI
1	1.0	0.125 (0.116) [0.11]	0.113 (0.103) [0.098]	0.14 (0.15)	3.87 (3.88)	0.6 (0.5) [0.2]
2	3.3	0.162 (0.143) [0.14]	0.144 (0.129) [0.125]	0.09 (0.13)	6.71 (5.92)	1.3 (0.9) [0.9]
3	6.0	0.182 (0.168) [0.16]	0.160 (0.150) [0.142]	0.08 (0.12)	8.68 (7.69)	1.9 (1.3) [1.2]

bandwidth metrics (Serio et al. 2005), and is given by:

$$Q_p = \frac{2}{{m_0}^2} \int_0^\infty f S^2(f) \mathrm{d}f, \tag{5.2}$$

where $S(f)$ is the variance density spectrum. We note that the BFI is not the most robust numerical parameter, as its precise value is strongly dependent on the way the bandwidth is calculated. The waves we generated during the experiment were slightly steeper and considerably more narrow-banded than those in the experiments of O04 (see table 5.1). For completeness, we report in table 5.1 the values of the parameters corresponding to our experiments, to the experiments of O04 calculated with the method used in our paper (in round brackets) and to the experiments of O04 as reported in its table 1 [in square brackets]. We emphasise the considerably higher values of BFI we obtain in comparison those reported in table 1 of O04. Whilst the waves made numerically agree very well in terms of the spectral shape with those desired (the input conditions in O04), we found the waves created experimentally to be somewhat larger in terms of significant wave height and considerably more narrow-banded (see table 5.1). This should be allowed for in the comparisons that follow.

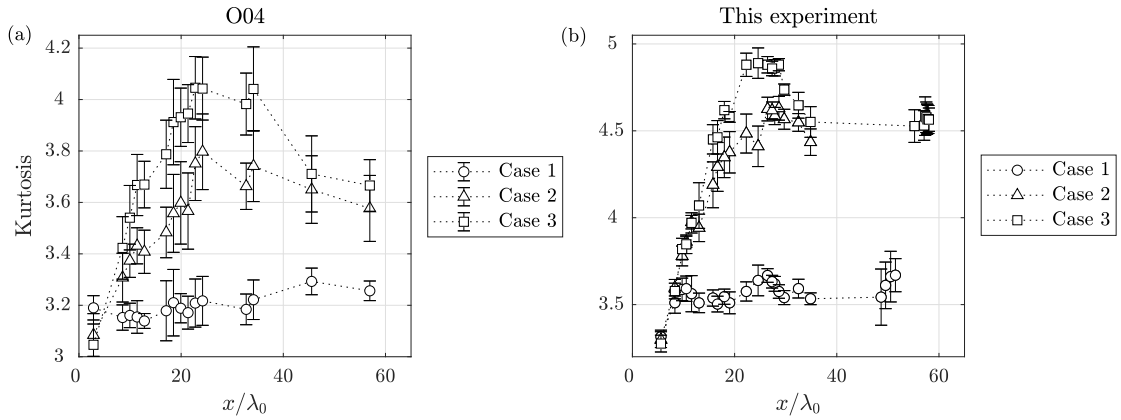

Figure 5.1: Spatial evolution of kurtosis: (a) results from Onorato et al. (2004), (b) our results. Error bars show the 95% of confidence interval based on standard deviation. The distance from the wavemaker is denoted by x, and λ_0 is the peak wave length.

Spatial evolution of kurtosis

We start by comparing our experimental results with those of O04. Figure 5.1 shows the evolution of kurtosis for the three cases. Despite the slight mismatch in the size of the waves, there is good basic agreement in the shape of the curves between the different experiments. With more probes positioned around the kurtosis peak, we are able to provide a better insight of where the kurtosis reaches its peak, which is most clearly observable for case 3.

Figure 5.2 presents the spatial evolution of the normalised excess kurtosis for the three cases using the different approaches (experiments, OceanWave3D and MNLS). The excess kurtosis, C_4^d, is corrected for the presence of bound waves using the method of Tayfun (1980). We have ensemble averaged the value of excess kurtosis to provide a clearer overall trend. The fluctuations that remain, especially for the experiments and OceanWave3D simulations, are the result of a relatively small number of ensemble members owing to their significant (computational) costs. Only 5 ensemble members were used for the experiments and 9 for the OceanWave3D simulations. Hence, we have added confidence intervals for the experimental and OceanWave3D results. Both numerics and experiments show the same general trends. In the simu-

lations presented, the MNLS captures the overall trend in all cases. Figure 5.2 also shows the theoretically predicted solution of excess kurtosis of Mori & Janssen (2006) based on the NLS and Fedele (2014) based on the cDZ (see appendix B for the exact equations we use). Both the theoretical predictions assume Gaussian input spectrum, which is different from the JONSWAP spectrum for experiments and simulations. It is worth mentioning that both solutions are given for spatially homogeneous waves evolving in the time domain in the original papers. We convert these to the temporally homogeneous waves evolving in space we study using the group velocity, which is valid in the context of the already narrow-bandwidth restricted NLS (Fedele et al. 2010) but only valid in the narrow-bandwidth limit of the cDZ. What we label cDZ in figure 5.2 is only the leading-order correction for broad bandwidth from Fedele (2014) (see appendix B).

The agreement of the two theoretical solutions, which give comparable predictions, with the experiments is very good over the whole course of the evolution for case 1 (the least steep and most broad-banded case). The experiments and theoretically predicted solutions start with an excess kurtosis of zero for all cases. Asymptotic analysis (see Janssen & Herbers (2009) and 5.4.2) of the analytical solution based on the NLS suggests that the initial growth rate of kurtosis should be quadratic in space. In cases 2 and 3 (the steeper cases), there appears to be a delay in the onset of the kurtosis increasing in the experiments compared to both theoretical solutions, as also observed in Fedele et al. (2010). After different non-dimensional length scales, cases 2 and 3 both depart entirely from the theoretical solutions. Both cases peak without exceeding the theoretical value. The excess kurtosis then slowly decreases until it reaches a steady-state value. Before we interpret the results in figure 5.2, we note that they depend strongly on the initial value of BFI, since its square acts as the normalising factor on the vertical axis. We also present a version of this figure in which excess kurtosis is normalised by steady-state BFI in appendix B.2. Normalisation on

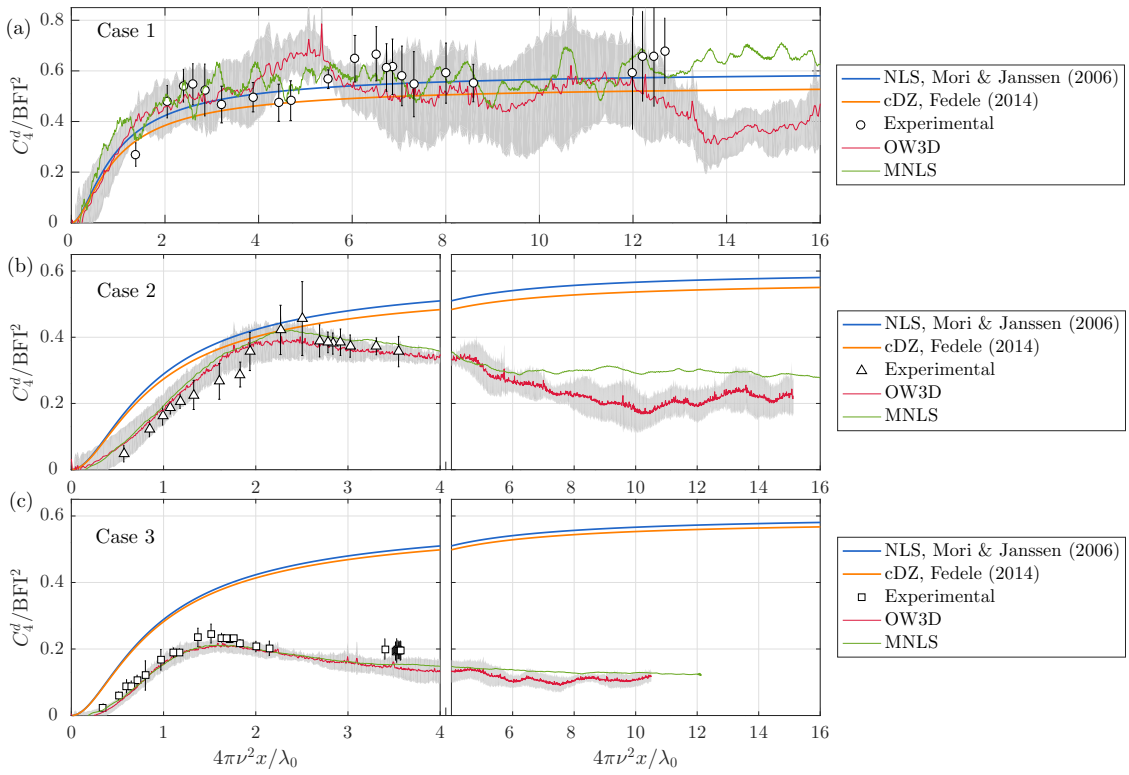

Figure 5.2: Evolution of normalised dynamic excess kurtosis at different distances from the wave generator: (a) case 1, (b) case 2 and (c) case 3. Shading represents the 95% confidence intervals for OceanWave3D simulation with 8 different random seeds. A total of 120 different random seeds are used in MNLS simulations. Consequently, for the MNLS, the error bars are negligible and have been omitted for clarity. The parameter C_4^d is dynamic excess kurtosis, BFI is the Benjamin-Feir Index, ν is the input bandwidth and λ_0 is the peak wave length.

BFI at different position also yields different evolution curves as the BFI value also varies along the tank. Although the normalised excess kurtosis is sensitive to the BFI value, the overall trend in the evolution of kurtosis is consistent in this study.

To interpret these results, we note that the theoretical solutions based on the NLS and the cDZ are based on small-steepness and narrow-bandwidth approximations (with the cDZ accounting for a wider bandwidth than the NLS, cf. appendix B). Thus, we would expect these solutions to work best in the region where the approximation is most valid, i.e. for the lower steepness waves with narrower bandwidths. Taking into account the leading-order balance between these two effects, the theoretical solution should work best when their ratio, as captured by the BFI, is smallest. This is consistent with our finding in the present paper, since the theory works best for the low-BFI case where the bandwidth of the spectrum is largest. Although the cDZ does explain a small reduction in kurtosis predictions compared to the NLS, it does not predict a maximum. The bandwidth correction for cDZ predictions on case 3 is limited as the wave spectrum is quite narrowbanded.

Hypothesising what happens at the level of a wave group, a spatial contraction of the wave group takes place around a large wave, which can be thought of as a local expansion of bandwidth. This effect is dependent on the Benjamin-Feir index of the 'sea state' (see for instance Eq. (3.12) in Adcock & Taylor (2009*b*)). It is important when considering the limitations of NLS-type equations not to base the bandwidth limitation on the input spectrum but on the local extremes which occur within the simulation.

Spatial evolution of the spectrum

Figure 5.3 shows the spatial evolution of the spectrum for case 3. It can be seen that all the spectral change takes place during the initial phase of the simulation before the kurtosis reaches its peak (see also Shemer, Sergeeva & Liberzon (2010)).

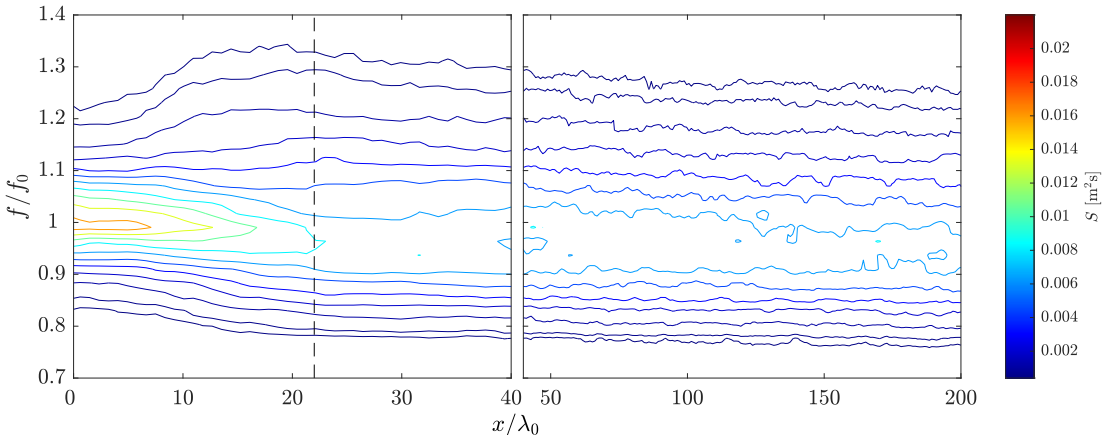

Figure 5.3: Spatial evolution of the frequency spectrum of case 3 averaging over multiple fully nonlinear simulations using OceanWave3D. The dashed vertical line shows the location of peak kurtosis.

To examine this further, we consider what happens if, during the evolution, we randomise the phase of the simulations removing all the correlations between components. We do this only using the MNLS as phase randomisation is more straightforward when bound waves are not directly simulated. Figure 5.4 presents the results of these simulations. Where phase randomisation occurs, the excess kurtosis is zero (i.e. at x/λ_0 = 168, 380). The energy of the system is kept as constant at each randomisation. Following the randomisation, the kurtosis increases rapidly although over a slightly longer length scale at each randomisation, presumably due to the slight broader spectrum at each successive randomisation. After each successive randomisation, the kurtosis reaches a peak in kurtosis before slowly settling back to a steady-state value. We find that after each randomisation the maximum value and the steady-state value are slightly lower. This demonstrates that prescribed uncorrelated initial random phase distribution is different from the correlated phase distribution at steady state as a result of nonlinear evolution (see also Slunyaev & Sergeeva (2012)). Comparing figure 5.4 (a) and (b), it is evident that the peaks in kurtosis go hand in hand with an overshoot in the broadening of the spectrum.

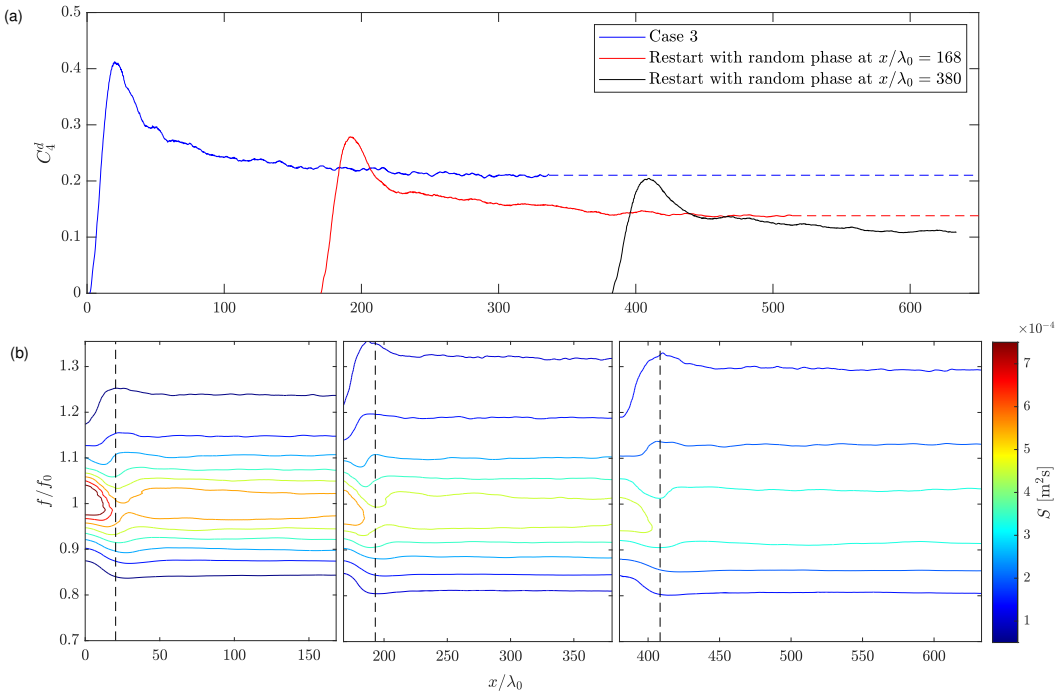

Figure 5.4: (a): Spatial evolution of the dynamic excess kurtosis for case 3 with phases randomised at x/λ_0 = 168, 380. The dashed lines indicate the steady-state excess kurtosis value. (b): Corresponding spectral evolution at different locations. The dashed vertical line shows the location of peak kurtosis.

Wave breaking

In the above discussion, we have ignored the effect of wave breaking. Wave breaking is modelled in the fully nonlinear simulations and is, of course, present in the experiments. Wave breaking is most active in the steepest and most nonlinear cases. Although this may play some role in our results, we suspect that breaking is of relatively minor importance, as our MNLS results, which do not include breaking, show very similar general results. Of course, over very long distances wave breaking will dissipate energy from the system. This takes place on a length scale much greater than we have considered.

5.4.2 Gaussian input spectra

To explore the parameter space further, we consider a second set of simulations where we just utilise the MNLS model. We do this to explore a significantly wider parameter space. We use Gaussian input spectra as it is more straightforward to define a bandwidth for such spectra than for spectra with an algebraic frequency tail, such as the JONSWAP spectrum used in the experiments of O04. Furthermore, the theoretical solutions of Mori & Janssen (2006) assume a narrow-banded Gaussian input spectrum. The Gaussian input spectrum is defined as:

$$S(k) = \left(\frac{H_s}{4}\right)^2 \frac{1}{\Delta f\sqrt{2\pi}} \exp\left(-(f - f_0)^2/(2\Delta f^2)\right), \tag{5.3}$$

where f_0 is the peak frequency and Δf controls the bandwidth of the spectra. We vary input steepness and bandwidth, but note our test matrix is not uniformly spaced in steepness and bandwidth. This irregular spacing was to explore as wide a range of values as possible without running cases that were too nonlinear or would take very long distances to reach a steady state. Since the leading-order broad-bandwidth correction from the cDZ (Fedele 2014) is small, we only compare to NLS-based (Mori

& Janssen 2006) theory here.

We will start by considering the initial rate of increase of the kurtosis. For spatially homogeneous waves that evolve in time, Janssen & Herbers (2009) showed that the theoretical solutions of Mori & Janssen (2006) reduce to a quadratic initial evolution of normalised excess kurtosis with respect to time (see also Mori et al. (2011) and Fedele (2015*b*)). Noting that for these narrow-banded spectra evolution in time can simply be expressed as evolution in space using the group velocity (Fedele et al. 2010), this quadratic evolution in time corresponds to the following initial quadratic evolution in space of normalised excess kurtosis, C_d^4/BFI^2, for the temporally homogeneous, unidirectional waves we study:

$$\frac{C_d^4}{\mathrm{BFI}^2} = \frac{1}{2}\xi^2 \quad \text{for } \xi \ll 1, \tag{5.4}$$

where the intrinsic dimensionless length $\xi \equiv 4\pi\nu^2 x/\lambda_0$, x the dimensional length, and λ_0 the peak wave length. If we express ξ as the ratio of distance from the wavemaker x and what we call the initial growth length scale L_s, we can express (5.4) as $C_d^4/\mathrm{BFI}^2 = (1/2)(x/L_s)^2$. We can now compare the theoretical prediction of the length scale $L_{s\xi} \equiv \lambda_0/(4\pi\nu^2)$ to an estimate from our data, which we obtain from the inverse of the rate of change of the square root of normalised excess kurtosis for $\xi < 0.5$:

$$L_s = \left(\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{\frac{2C_d^4}{\mathrm{BFI}^2}}\right)^{-1} \quad \text{for } x \ll \frac{\lambda_0}{4\pi\nu^2}. \tag{5.5}$$

Figure 5.5a presents the comparison of our estimate of the initial growth length scale from simulations of the MNLS L_s to its theoretically predicted counterpart $L_{s\xi}$. For the most narrow-banded cases, the growth length scale is quite close to the theoretical values. However, theory seems to over-predict the initial growth length scale for broad-banded cases. Steepness has only a small role.

Figure 5.5b shows the normalised peak kurtosis $C_d^4|_{\max}/\mathrm{BFI}^2$ reached during each

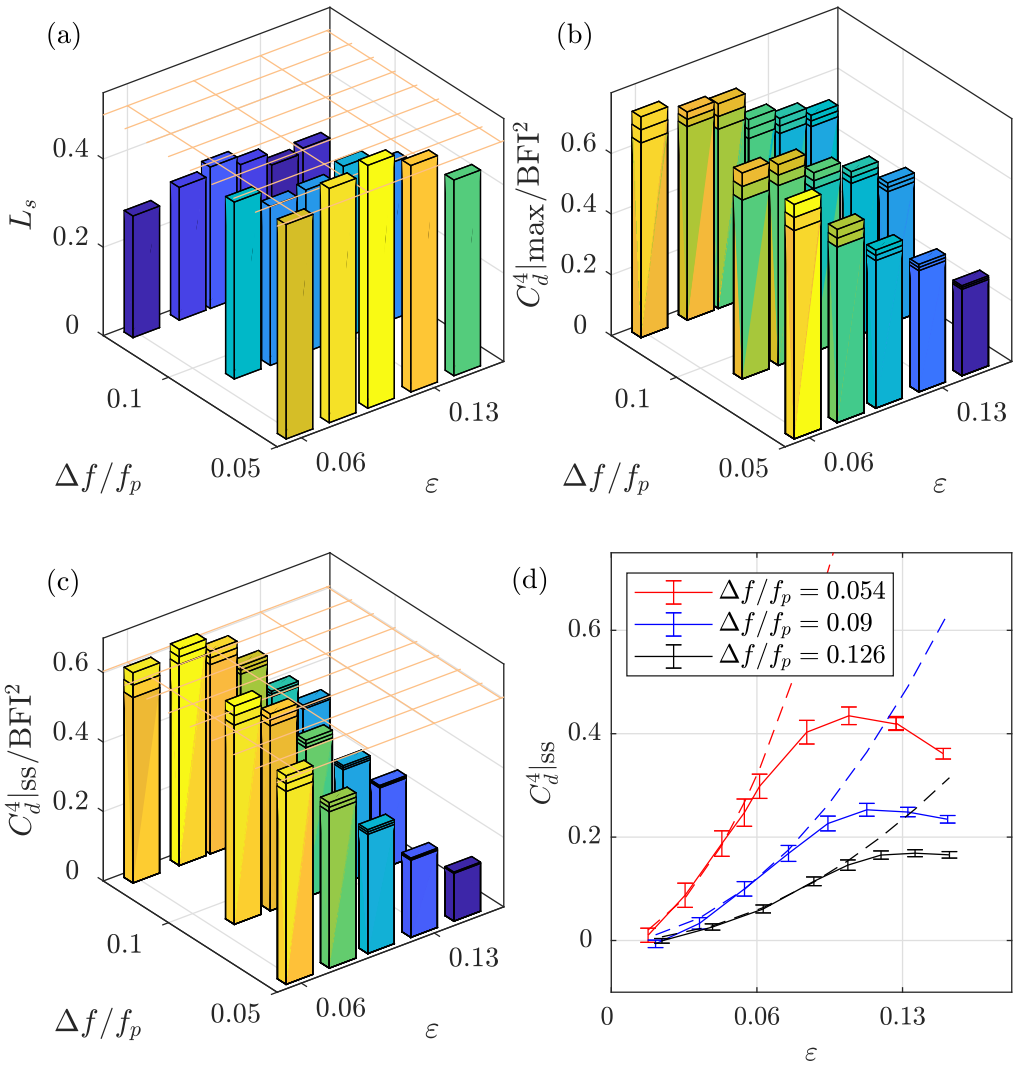

Figure 5.5: Analysis of kurtosis of random waves from an initially Gaussian spectrum: (a) ratio between the initial growth length scales of normalised excess kurtosis predicted by the simulation L_s and theory $L_{s\xi}$, (b) maximum value of kurtosis $C_d^4|_{\max}$ reached during the simulation, (c) steady-state value of normalised excess kurtosis $C_d^4|_{\mathrm{ss}}/\mathrm{BFI}^2$, (d) steady-state value of kurtosis $C_d^4|_{\mathrm{ss}}$ as a function of the initial steepness (ε) for three different initial bandwidths ($\Delta f/f_0$). The grid in (a), (c) and the dashed lines in (d) shows the theoretical prediction based on Mori & Janssen (2006). The error bars in (b), (c) and (d) show 95% confidence interval based on the standard deviation.

simulation. The NLS-based theory of Mori & Janssen (2006) predicts the steady-state value to be at $\pi/(3\sqrt{3}) \sim 0.604$. This is close to the value reached by the least nonlinear case we considered. However, as either steepness increases, or bandwidth decreases, the peak value of kurtosis normalised by the value of the Benjamin-Feir Index squared decreases. This behaviour, and the discussion of the physical reasons for it, is consistent with that presented above for the cases based on the JONSWAP spectrum.

Finally, we consider the steady-state value that is reached at the end of the simulations. Reaching a steady state takes a different distance for different simulations. We determine steady-state kurtosis $C_d^4|_{\mathrm{ss}}$ as the averaged kurtosis over a distance of $\Delta x/\lambda_0 = 50$, where the maximum variation in the excess kurtosis over this distance is less than 10% of the maximum excess kurtosis. Figure 5.5c presents the steady-state value of normalised steady-state kurtosis for a range of input conditions. The trends are very similar to those for peak kurtosis. Essentially, the higher the starting BFI, which results in a larger denominator for the normalised kurtosis, the smaller is the steady-state normalised kurtosis.

Additional insight into the steady-state values of kurtosis is given by considering steady-state kurtosis without normalisation for cases with different steepnesses but the same bandwidths. Figure 5.5d shows the results for three different bandwidths, where we have also added the theoretical predictions based on Mori & Janssen (2006). For sufficiently small steepness, we observe an increase in the steady-state kurtosis with steepness, as expected. The values of kurtosis measured from simulations agree well with the theoretical predictions. However, above a certain initial steepness, which depends on the initial bandwidth, the steady-state kurtosis appears to flatten off as steepness is increased further. This flattening also causes the measured kurtosis at steady state to depart from the theoretical predictions. This figure further demonstrates the discrepancy between numerical simulations and theoretical predictions for

the steady-state values of normalised kurtosis in figure 5.2.

5.5 Results: group shape

In a linear model of wave evolution, the theory of quasi-determinism can be used to describe the shape of extreme waves (Lindgren 1970, Boccotti 2000, Tromans et al. 1991). As part of this theory, the average shape of an extreme crest is given by the scaled auto-correlation function (Boccotti 1983) and the shape of the wave with the largest crest-to-trough height by the scaled difference of two time-shifted auto-correlation functions (Boccotti 1989). In the time domain, the expected shape of an extreme crest $\eta = \eta_{\max}$ at any location x (with the largest crest shifted to occur at $t = 0$) is given by:

$$\eta(t,x) = \frac{\eta_{\max}}{m_0} \int_0^\infty S(f,x)\cos(2\pi f t)\mathrm{d}f, \tag{5.6}$$

where $S(f,x)$ is the power spectral density function at the location of the measurement x, $m_0 = H_s^2/16$ is the zeroth moment of the energy spectrum at that location. Nonlinear physics might be expected to modify this. For deep-water waves, the effect of nonlinear physics on the shape of a nonlinear event has mainly been studied for wave groups (e.g. Baldock et al. (1996)). For a unidirectional wave group, analytical results based on the NLS predict that the group would contract spatially, and this is dependent on the amplitude-to-width ratio of the group (analogous to the BFI but for a group) (Adcock & Taylor 2009*b*). Little attention has been paid to unidirectional random waves with the study of Lo & Mei (1985) and the recent work of Dematteis et al. (2019) being exceptions. In this section, we consider the shape (in the time-domain) of extreme events throughout the spatial evolution using the MNLS with input Gaussian spectra.

Figure 5.6 (a) presents the measured average shape of the 20 largest crests at different locations in the numerical tank. The bandwidths at the different locations

are $\nu = \{0.12, 0.15, 0.18, 0.18, 0.18, 0.17\}$ corresponding to $x/\lambda_0 = \{0, 10, 20, 30, 40, 50\}$. Figure 5.6 (b) presents the linear predictions of these events at the same position based on the theory of quasi-determinism. The second-order bound harmonics are excluded in the figure, as the MNLS code computes the free wave directly without second-order bound harmonics. When generated, the group is symmetric and consistent with linear theory. As the waves evolve in space, the shape of the wave groups is modified by nonlinear physics. The main characteristics of this modification are a movement to the front of the wave group of the largest wave and a contraction of the wave group. We have also presented the shape of the crests profiles predicted by the theory of quasi-determinism shown in figure 5.6 (b). The changes in the predicted profiles in figure 6 (b) are only due to changes in the wave spectrum with x (cf. ($S(f, x)$ in (5.6)). Additionally, we present the averaged shape of the wave with the largest crest-to-trough height in appendix B.3.

We quantify the changes to group shape using two parameters (following Tang, Tromans & Adcock (2019)). The first is a measure of the width of the group and the second a measure of the asymmetry. We begin by defining σ_1 and σ_2 as the durations when the envelope height exceeds 80% of the peak height of the envelope obtained by averaging the envelope corresponding to the 20 largest waves, where the envelope is obtained using a Hilbert transform. As illustrated in figure 5.7, σ_1 denotes the duration of the front flank of the group (passing the observer at earlier time) and σ_2 the duration of the rear flank (passing the observer at later time). We now define the parameters B_1 and B_2 to be the durations σ_1 and σ_2 as fractions of what would be predicted by linear theory from the local spectrum (i.e. from the theory of quasi-determinism):

$$B_1 = \frac{\sigma_{1,\text{measured}}}{\sigma_{1,\text{quasi-determinism}}}, \quad B_2 = \frac{\sigma_{2,\text{measured}}}{\sigma_{2,\text{quasi-determinism}}}. \tag{5.7}$$

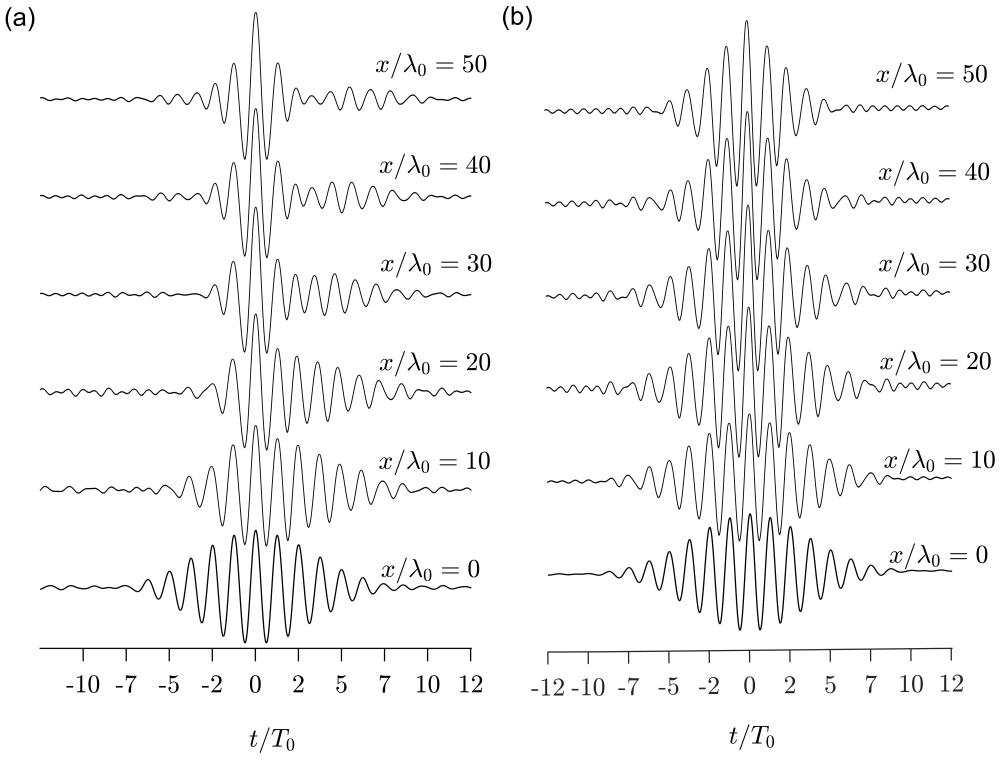

Figure 5.6: Normalised average of: (a) the 20 largest crest profiles out of over 6400 waves, (b) predicted crest profiles based on the theory of quasi-determinism (right) at $x/\lambda_0 = 0, 10, 20, 30, 40, 50$ for random waves with a Gaussian input spectrum with $\Delta f/f_0 = 0.054$ and $\epsilon = 0.044$.

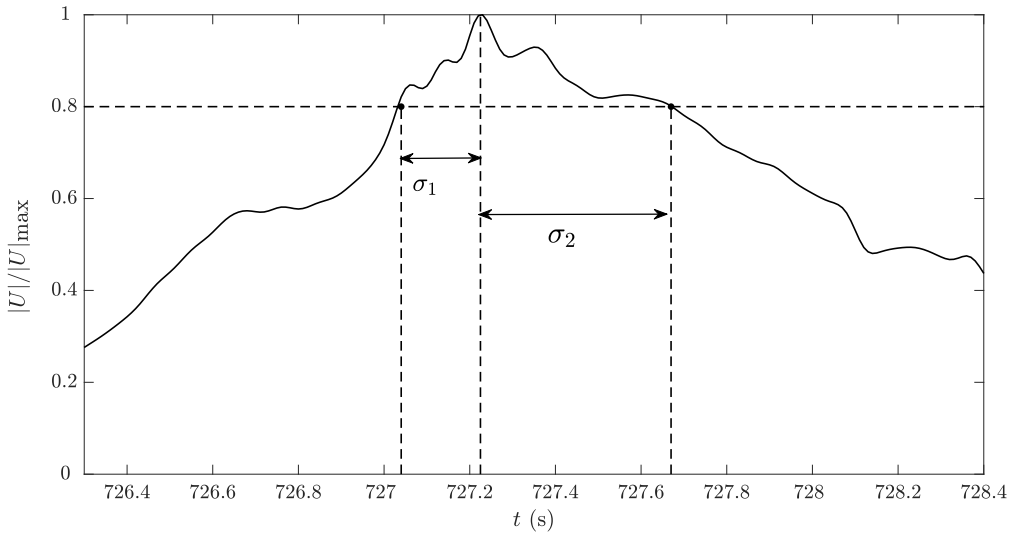

Figure 5.7: Illustration of the envelope duration when the normalised envelope height $|U|/|U|_{\max}$ exceeds 80% of its peak height.

The ratios B_1 and B_2 quantify the nonlinear modifications to both sides of the measured envelope shape compared to linear theory. We then define B_{mean} and ΔB as parameters respectively measuring changes in width and asymmetry relative to linear evolution:

$$B_{\text{mean}} = \frac{B_1 + B_2}{2}, \quad \Delta B = B_2 - B_1. \tag{5.8}$$

Thus, a positive value of $1 - B_{\text{mean}}$ implies the group has contracted relative to the shape expected under linear evolution. ΔB can be seen as the difference of envelope width between the front and rare part, which is then normalised by the half envelope width predicted by the linear theory ($\sigma_{1,\text{quasi-determinism}}$). Hence, a positive value of ΔB implies the largest wave has moved towards the front of the group.

Figure 5.8 presents the evolution of the parameters for asymmetry and change in group duration. Different initial bandwidth values are given in different panels with different lines for different steepnesses. In all cases, we observe a positive asymmetry, that is, the largest wave moves to the front of the group and arrives earlier (for individual extreme events, ΔB is not guaranteed to be positive, however, we show the average value for over 3,000 large waves herein.). Spatially, the pattern is similar to that of kurtosis. For most cases, there is a peak in the asymmetry parameter during the initial phase before settling down to a smaller near-zero equilibrium value. As for kurtosis, the length scale for equilibrium to be reached is faster for narrower bandwidth. The degree of contraction of the wave-group behaves differently. It increases steadily until it flattens off as equilibrium is achieved. The contraction occurs on a different, much longer, spatial scale than the transient asymmetry. The nonlinear dispersion relationship may lead to horizontal asymmetry. However, the physical expiation of transient behaviour of the horizontal asymmetry and the contraction of wave envelope for random waves remains unknown (see the prior theoretical work for wave groups in Adcock et al. (2012)).

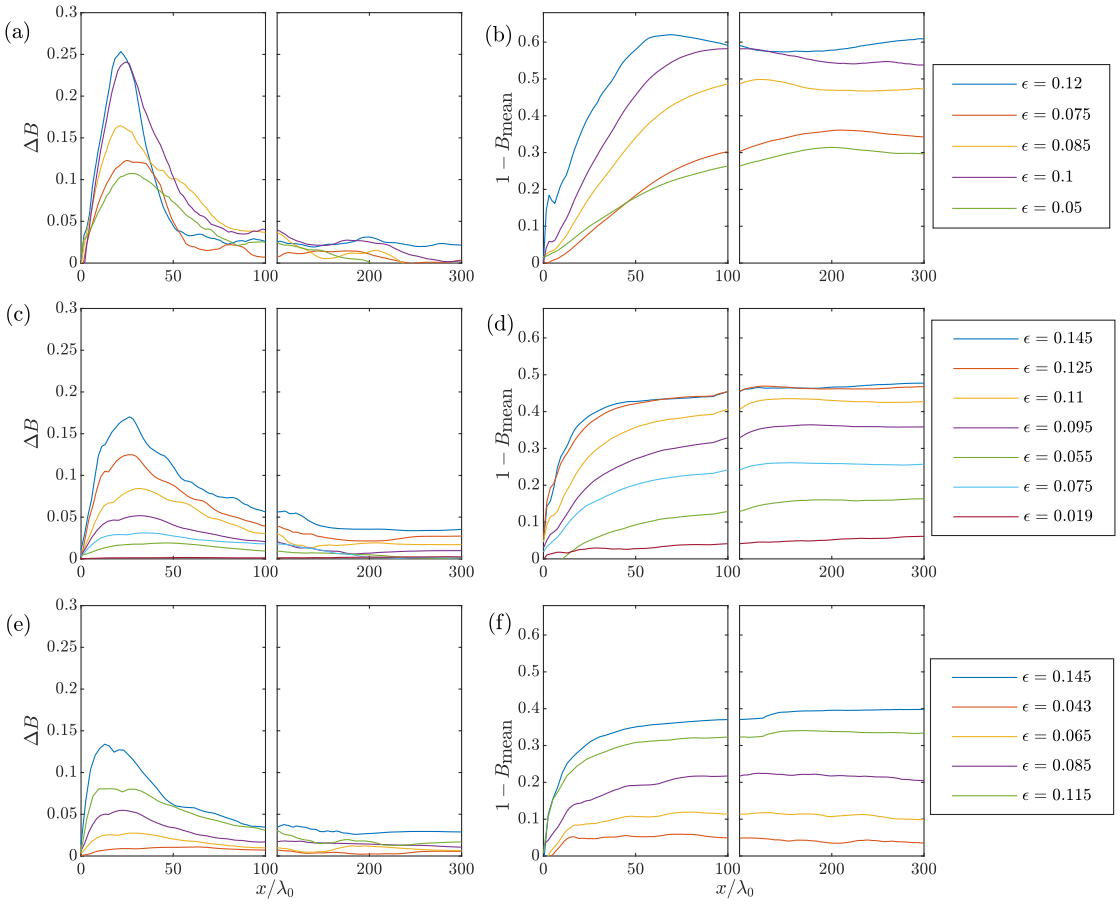

Figure 5.8: Group shape during nonlinear evolution in space: (a,c,e) envelope asymmetry (ΔB) and (b,d,f) nonlinear change in the duration when the envelope height exceeds 80% of its peak height ($1 - B_{\text{mean}}$) with different initial bandwidths, (a,b) $\Delta f/f_0 = 0.054$ (c,d) $\Delta f/f_0 = 0.09$ and (e,f) $\Delta f/f_0 = 0.126$.

5.6 Discussion and conclusions

This paper has examined experimentally and numerically how the kurtosis and the shape of large waves, evolves over relatively long distances for unidirectional surface gravity waves. In doing so, we have revisited the seminal unidirectional laboratory experiments of Onorato et al. (2004), which we repeat, extend and complement with numerical solutions of the fully nonlinear water wave equations and the MNLS equation of Trulsen et al. (2000). We concentrate on the spatial evolution from a random boundary condition without correlation between components. Following nonlinear evolution, the kurtosis must eventually settle down to a steady state for unidirectional waves (without higher-order nonlinearity), in which changes to the spectrum no longer occur, and the kurtosis has a fixed value greater than that of a Gaussian 'sea state'. We have investigated the transition between these two states and the dependence of the final state on the input conditions.

The picture we find is consistent across the different models and the different cases investigated. For the unidirectional cases studied here, we do not find any significant difference in overall behaviour between JONSWAP and Gaussian spectra, although we note that there is significant ambiguity when calculating the spectral bandwidth for the JONSWAP spectrum. For an input spectrum which is not too narrow-banded or steep (low BFI), theory based on either the NLS (Mori & Janssen 2006) or cDZ (Fedele & Dutykh 2012) are excellent models matching both experiments and simulations. However, if the input spectrum is steep or narrow, as captured by a high value of the BFI, the evolution of the excess kurtosis departs from these models. Significant changes occur to the spectrum over relatively short distances. Over the same short distances, excess kurtosis rapidly increases and peaks. The overshoot can be interpreted as being driven by the initial departure of phases from their equilibrium distribution (as can be seen by randomising the phases once steady state has been achieved). Over longer distances excess kurtosis then drops until it reaches a steady-

state value with no further evolution of the spectrum. The peak and hence the final value are at a lower kurtosis than that predicted theoretically based on the NLS and cDZ. For cases which are sufficiently nonlinear to depart from the theoretical curve, the kurtosis at steady state appears to be primarily dependent on the initial bandwidth of the spectrum (rather than the steepness).

We note that Janssen (2003) was already aware of the absence of an overshoot in his theoretical calculations based on the kinetic approach; these calculations based on the kinetic approach ultimately led to the result by Mori & Janssen (2006) we have compared to herein. Janssen's (2003) numerical simulations for the time evolution of spectral width show a similar overshoot to what we have observed for kurtosis (cf. his figure 3). This overshoot is also present in numerical simulations based on the NLS in Onorato et al. (2016) (their figure 1, where the overshoot is also present in bandwidth). This overshoot, Janssen (2003) notes, is likely ignored in his theoretical calculations owing to the assumption that the the action density varies slowly, an assumption which was later relaxed in the Generalized Kinetic Equation of Annenkov & Shrira (2006). In this paper we have examined a scenario where the spectrum undergoes a transient broadening at the location of the peaks in kurtosis. We thus envisage that calculations based on the Generalized Kinetic Equation of Annenkov & Shrira (2006) will predict the behaviour of kurtosis for steep and narrow-bandwidth spectra identified herein.

Examining simultaneously the effect of nonlinear evolution on shape of the largest waves, there is a tendency for the largest wave in a packet to move towards the front of the group. This asymmetry follows a similar path to the kurtosis, showing a clear peak during the early stage of evolution before reducing to become relatively small once equilibrium is reached. There is also a reduction in the width of an extreme event making the extreme event appear more transient. This follows a different evolution as it increases monotonically before reaching equilibrium on a longer spatial scale than

spatial scale associated with the asymmetry. The locations of the maxima of kurtosis coincide approximately with the locations of the maxima of asymmetry of the average shape of extreme events.

We conclude by emphasising our paper is confined to unidirectional waves, so that its findings cannot directly be extended to real-world ocean waves, which are directionally spread. For directionally spread ocean waves, peaks in kurtosis also occur (Onorato, Waseda, Toffoli, Cavaleri, Gramstad, Janssen, Kinoshita, Monbaliu, Mori, Osborne et al. 2009, Toffoli et al. 2010, Xiao et al. 2013, Fedele 2015*b*, Annenkov & Shrira 2009, 2018), yet as a result of very different underlying physics.

Chapter 6

The influence of finite depth on the evolution of extreme wave statistics in numerical wave tanks

6.1 Abstract

Model tests with unidirectional random wave fields are common for coastal engineering purposes. In this paper, we investigate the finite water depth effect on both the probability of extreme events and the averaged shape of them in such experimental conditions. We simulate some typical model test conditions with a numerical tool, which solves the fully non-linear wave equations. We find the analytical solution based on Nonlinear Schrödinger equation (NLS) to be accurate for low steepness and relatively deep water cases. However, the analytical solution underestimates the kurtosis at the steady state for high steepness cases in shallow water, and also gives zero value of kurtosis at critical water depth, whereas a small but non-zero value of kurtosis is observed in the numerical tank. We also investigate the averaged shape of the extreme events, which are modified by nonlinear physics over the distance. The

horizontal asymmetry is significant initially but is greatly reduced to a steady state, and the contraction of the wave groups grows monotonically until the steady state is reached along the wave tank. Finite water depth has limited effects on the averaged shape of large events, especially for the extremely steep waves.

6.2 Introduction

Coastal engineers often need to perform model tests in unidirectional wave flumes. To simulate the short-term wave variability in nature they will do tests using random waves with a predefined spectrum. Typically, waves will be generated by randomly assigning a phase to different Fourier components, and using this as a basic model to drive the paddle (possibly including second order physics see Orszaghova et al. (2014)). However, in unidirectional waves, non-linear interactions lead to correlations forming between components, increasing the number of large waves and modifying the shape of the average event away from that predicted in a linear model. This contrasts to waves in the real ocean which are directionally spread. In such realistic seas, where the wave field is in 'equilibrium', models based on linear dynamics are expected to be good predictors of wave height (Christou & Ewans 2014, Fedele 2015*b*) and of group shape for all but the most extreme waves (Tang, Tromans & Adcock 2019, Karmpadakis et al. 2020).

It is important for testing models that this departure from linear statistics is understood. For instance, if the tester simulates a design sea state at the paddle, the loads predicted on the object being tested will be over-estimated if the non-linear physics has led to more large waves. Further, the shape of the extreme wave-group can be important in determining forces. A clear example of this is in structures which have a dynamics response such as the 'ringing' effect which may impact the design of offshore wind turbines (Grue 2011, Zang et al. 2010). Thus, model testers need to be

able to assess the importance of these change in the shape of extremes in their work.

The theoretical work on this problem perhaps originates with the pioneering work of Benjamin & Feir (1967). However, analysis of the problem looking at the evolution of wave from an initially random initial condition stems from the work of Janssen (2003) with major theoretical contributions from Mori & Janssen (2006), Fedele (2015*b*), Fedele & Dutykh (2012), Fedele et al. (2010), Janssen & Janssen (2019). Experiments and numerics such as Xiao et al. (2013), Onorato et al. (2006), Onorato, Waseda, Toffoli, Cavaleri, Gramstad, Janssen, Kinoshita, Monbaliu, Mori, Osborne et al. (2009), Toffoli et al. (2010) examine the multidimensional random wave fields. Most of this work has focussed on deep water and has focussed on the amplitude distribution of waves rather than the shape of extreme wave groups.

In finite water depth, the wave-induced current stabilise the nonlinear focusing for unidirectional wave trains at sufficiently small relative water depths below a critical value of $k_0 d = 1.363$ (Janssen & Onorato 2007, Mori & Yasuda 2002, Whitham 1974) and the cubic nonlinearity vanishes (Kakutani & Michihiro 1983) in nonlinear Schrödinger equation (NLS) type solutions. However, nonlinear focusing can still be triggered with directional disturbances in finite water depth (see numerical simulations in Fernandez et al. (2014), Slunyaev et al. (2002) and experiments in Toffoli et al. (2013)). The role of directionality in finite water depth in wave statistics of random wave fields have been examined by Toffoli et al. (2009) with experimental results and numerical simulations (see also Fernandez et al. (2016) and experiments in Karmpadakis et al. (2019)).

Kurtosis of the free surface is widely used as a proxy for the distribution of extreme waves (see for instance Mori & Janssen (2006)). Unfortunately, most of the closed-form analytical predictions on the evolution of kurtosis focus on deep water waves, which excludes the suppression of nonlinear wave focusing in finite water depth. In the present paper, we apply a simple depth correction to the existing analytical model

and investigate its validity when including finite depth effects. We numerically explore what happens to steep sea-states as they evolve from uncorrelated initial conditions to an equilibrium state and how the shape of the largest waves varies along the flume using a fully non-linear potential flow solver.

6.2.1 The model

To investigate the wave evolution in different water depths, OceanWave3D is used to solve the standard fully non-linear wave equations (see Engsig-Karup et al. (2009) for details). This numerical scheme solves the potential flow governing equations over the whole domain with fully non-linear boundary conditions at the free surface using a finite difference method.

The numerical wave tank has a length of 320 m in the wave propagation direction with 4098 nodes, which provides a spatial resolution of 0.078 m (approximately 23 nodes per peak wavelength). Fifteen clustered nodes are used vertically in the water column. We have been careful to mitigate the impact of different grid resolutions on our results and have ensured the numerical accuracy of the results following the approach of Barratt, Bingham & Adcock (2020). The total simulation time is 1920 s with a time step of 0.02 s (75 per period).

Waves are generated with a relaxation zone in the first 30 m of the numerical wave tank and are absorbed by a 35 m long pressure damping zone at the end. Care has been taken to ensure the second order error waves and wave reflection effects are negligible. There is a local smoothing filter to model the wave breaking effects. This local filter activates when the vertical Lagrangian acceleration on the free surface is greater than 0.4 m/s^2. When this filter is triggered, a small amount of energy is removed from the breaking waves locally until the vertical particle acceleration is below the threshold. For the case of deep water, the model agrees very well the experimental results (Tang, Xu, Barratt, Bingham, Li, Taylor, van den Bremer &

Table 6.1: Sea state parameters for both cases.

No.	T_0 [s]	k_0 [m^{-1}]	d [m]	$k_0 d$	H_s [m]	ϵ
Case 1 No. 1	1.5	2.06	0.65	1.33	0.14	0.04
Case 1 No. 2	1.5	1.97	0.77	1.5	0.14	0.04
Case 1 No. 3	1.5	1.82	1.33	2.42	0.14	0.04
Case 1 No. 4	1.5	1.80	1.75	3.14	0.14	0.04
Case 2 No. 1	1.5	2.06	0.65	1.33	0.11	0.03
Case 2 No. 2	1.5	1.97	0.77	1.5	0.11	0.03
Case 2 No. 3	1.5	1.82	1.33	2.42	0.11	0.03
Case 2 No. 4	1.5	1.80	1.75	3.14	0.11	0.03

Adcock 2020).

6.2.2 Test conditions

The input wave spectrum for the tests conditions, is a TMA transform as proposed by Hughes (1984):

$$S(f)_D = \frac{S(f)_\infty}{2n} \tanh^2 k_0 d, \tag{6.1}$$

where $S(f)_D$ is the transformed finite water depth spectrum, $S(f)_\infty$ is the spectrum for a deep water conditions, n is the ratio between group velocity and phase velocity at water depth of d and k_0 is the dominant wave number. This transform retains the spectral shape in deep water but accounts for the notional influence of linear shoaling on all the components. For $S(f)_\infty$, we used a JONSWAP spectrum with $\gamma = 3.3$. A recent study shows this TMA transformed JONSWAP spectrum predicts the averaged shape of extreme events well in intermediate water depth (Whittaker et al. 2016). We consider two cases in this paper. The main difference between them is the significant wave height. Case 1 has a higher significant wave height of 0.14 m (steepness $(\epsilon) = 2\pi H_s/(gT_0^2) = 0.04$), and case 2 have a significant wave height of 0.11 m (steepness $(\epsilon) = 0.03$). We summarise the sea state parameters in Table 6.1:

6.3 Results

In this section, we aim at investigating the evolution of the probability of the extreme events and the averaged shape of these. We first compare the numerical results with analytical results for the evolution of dynamic excess kurtosis based on the NLS (Fedele et al. 2010) in different water depths. We further investigate the average shape of largest events by quantifying the horizontal asymmetry and the contraction of the wave envelope in the mean wave direction at different water depths.

6.3.1 Evolution of dynamic kurtosis

We start by comparing our numerical results with modified theoretical predictions based on Fedele et al. (2010). We compute the kurtosis C_4^d, which is suggested by Janssen (2003) as a parameter for estimating rogue wave density:

$$C_4^d = \frac{\langle \eta_L^4 \rangle}{3\mathrm{std}^4} - 1, \tag{6.2}$$

where η_L is the linearised free surface elevation with respect to the mean water level, std is the standard deviation of the free surface, and angle brackets denote a statistical average. We 'correct' the simulated free surface for bound waves by removing all the harmonics that do not follow the dispersion relationship in the frequency domain (see Barratt, Bingham, van den Bremer & Adcock (2020) for details). The influence of bound harmonics is discussed in section 6.3.2.

Mori & Janssen (2006) present the evolution of excess kurtosis for unidirectional waves in deep water with a three-dimensional integral evolution equation (their Eq. (14)) and a simplified version in the narrow-bandwidth and large-time limit for Gaussian spectra (their Eq. (28)). Without invoking the large-time assumptions, the integral can be evaluated in closed-form (Fedele et al. 2010) (see also Janssen & Janssen

(2019))

$$\frac{C_4^d(x)}{\mathrm{BFI}^2} = \frac{\pi}{3\sqrt{3}}\Big[1 - \frac{6}{\pi}\mathrm{Im}\Big(i \arcsin\frac{1+2i\alpha}{2}\Big)\Big], \tag{6.3}$$

where C_4^d is the dynamic excess kurtosis, Im is the imaginary part, $\alpha = 2\nu^2 x/\lambda_0$ and ν is the bandwidth of the wave spectrum.

However, Equation 6.3 is derived for deep water waves. To obtain the finite water corrections, we simplistically follow Serio's method of computing Benjamin-Feir Index (BFI) in finite water depth (Serio et al. 2005):

$$BFI = \sqrt{m_0} k_0 Q_p \kappa \sqrt{2\pi}, \tag{6.4}$$

where $m_0 = H_s^2/16$ is the zeroth moment of the energy spectrum, κ is a depth correction factor, which is unity in deep water. κ can be computed as:

$$\kappa = \mu \sqrt{\frac{|\tau|}{\chi}}, \tag{6.5}$$

where μ, τ and χ are all dimensionless coefficients depending on relative water depth $k_0 d$. The general forms of μ, τ and χ are (see Mei (1989) for detailed derivation):

$$\mu = 1 + 2\frac{k_0 d}{\sinh(2k_0 d)}, \tag{6.6}$$

$$\tau = -\mu^2 + 2 + 8(k_0 d)^2 \frac{\cosh(2k_0 d)}{\sinh^2(2k_0 d)}, \tag{6.7}$$

$$\chi = \frac{\cosh(4k_0 d) + 8 - 2\tanh^2(k_0 d)}{8\sinh^4(k_0 d)} - \frac{(2\cosh^2(k_0 d) + 0.5\mu)^2}{\sinh^2(2k_0 d)\Big[\frac{k_0 d}{\tanh(k_0 d)} - \Big(\frac{\mu}{2}\Big)^2\Big]}. \tag{6.8}$$

Q_p is the quality factor introduced by Goda (2000). Q_p is a dimensionless parameter,

which describes the spectral bandwidth. It has less sensitivity to the high frequency tail of the spectrum (and cut-off frequency) than other bandwidth metrics (Serio et al. 2005, Prasada Rao 1988). Q_p is given by:

$$Q_p = \frac{2}{{m_0}^2} \int_0^\infty f S^2(f) \mathrm{d}f, \tag{6.9}$$

where $S(f)$ is the wave spectral density function.

Figure 6.1 shows the evolution of dynamic excess kurtosis (C_4^d) at different water depths. From Figure 6.1, both numerical simulations and the analytical solution based on cubic nonlinear Schrödinger equation (NLS) start with excess kurtosis of zero. The excess kurtosis reduced noticeably as the water depth decreases towards the critical value, which agrees well with the existing literature on the suppression of nonlinear focusing at finite water depth (Janssen & Onorato 2007, Mori & Yasuda 2002, Whitham 1974). The asymptotic analysis captured the overall trends except for the shallowest case ($k_0 d = 1.33$). Near the critical water depth of $k_0 d = 1.36$ where the cubic nonlinearity vanishes (Kakutani & Michihiro 1983) meaning the analytical solution predicts that the free surface remains Gaussian (excess kurtosis is always zero). However, we still observe a small value of dynamic excess kurtosis at steady state. This could be due to the imperfections in correcting for bound waves, but is also likely to be due to nonlinearity beyond the cubic nonlinearity contributing to the excess kurtosis. This could also explain the small gap between the theoretical predictions and the simulated results for the case with $k_0 d = 1.5$. For relative deep water cases ($k_0 d = 2.41$ and $k_0 d = 3.14$), the numerical results agree well with the analytical solutions, which suggests the asymptotic analysis can provide an accurate estimation of kurtosis for low steepness and relatively deep water wave fields. The evolution of wave kurtosis in deeper water depth with higher wave nonlinearity is presented in Chapter 5.

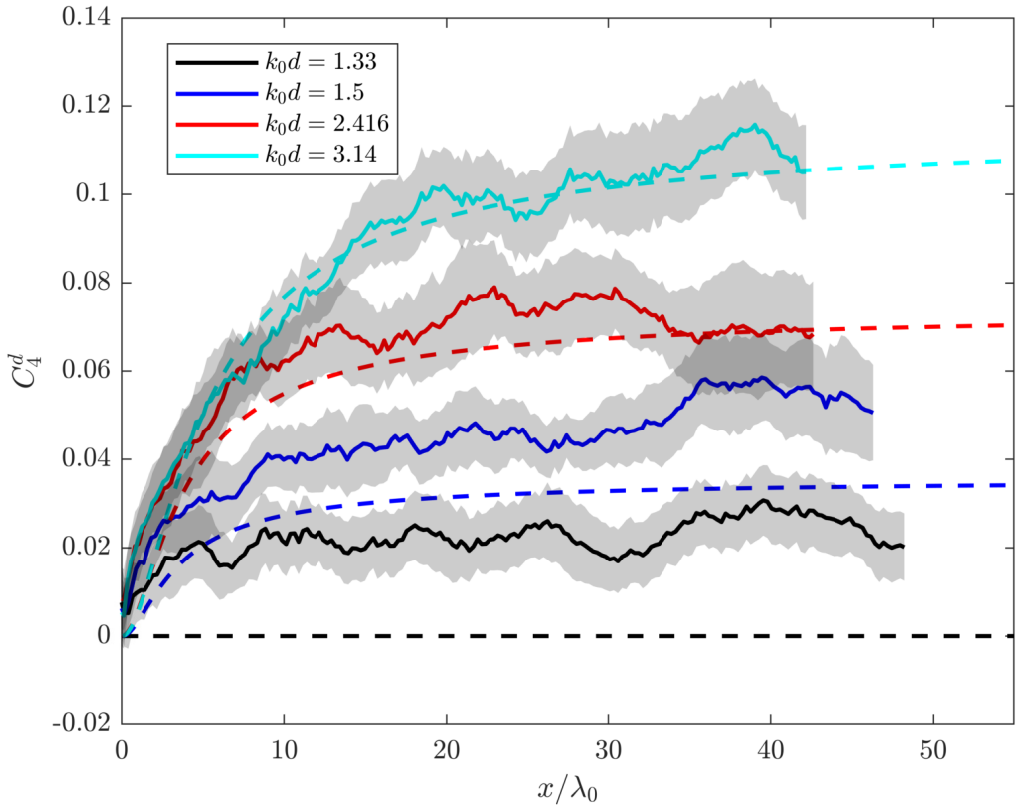

Figure 6.1: Evolution of dynamic excess kurtosis for case 2 along the numerical flume with different water depth. The dashed line shows the corresponding theoretical predictions from Fedele et al. (2010) for each water depth.

We further investigate the evolution of kurtosis at the same water depths but with higher wave steepness in Figure 6.2. For the relatively deep water case ($k_0 d$ = 3.14), the analytical solution predicts the initial growth of the excess kurtosis well, but after the peak, numerical results depart from the theoretical solutions with a decreasing trend. This could be due to the inherent limitations of the NLS typed approximations, which assume a narrowbanded Gaussian spectrum. However, at the initial stage of wave propagation, the spectrum broadens as the correlations between initially uncorrelated components develop (see Onorato et al. (2004) for example). This introduces permanent changes to the wave spectrum. The assumption that kurtosis can be expressed in terms of lower-order moments (Eq. (16) of Janssen (2003)) may also become invalid if the waves are too steep. The discrepancies in excess kurtosis, for relatively shallow water case ($k_0 d$ = 1.36) increase for higher steepness. This is expected as the contribution from higher order nonlinearity should become more significant for high steepness cases.

6.3.2 Contribution of bound waves

In the above we focus on the free waves as a surrogate measure of the correlation between freely propagating components in our numerical simulations. We quantify the contribution from nonlinear interactions between free waves by examining the spatial evolution of dynamic excess kurtosis (C_4^d). However, for real water waves, the contribution to kurtosis induced by the bound waves can be significant, especially at finite water depth (Toffoli et al. 2007, 2008, Mori & Janssen 2006). The bound wave contribution to kurtosis, C_4^b, can be approximated by subtracting dynamic excess kurtosis from the excess kurtosis of original surface elevation:

$$C_4^b = \frac{\langle \eta^4 \rangle}{3\mathrm{std}^4} - 1 - C_4^d, \tag{6.10}$$

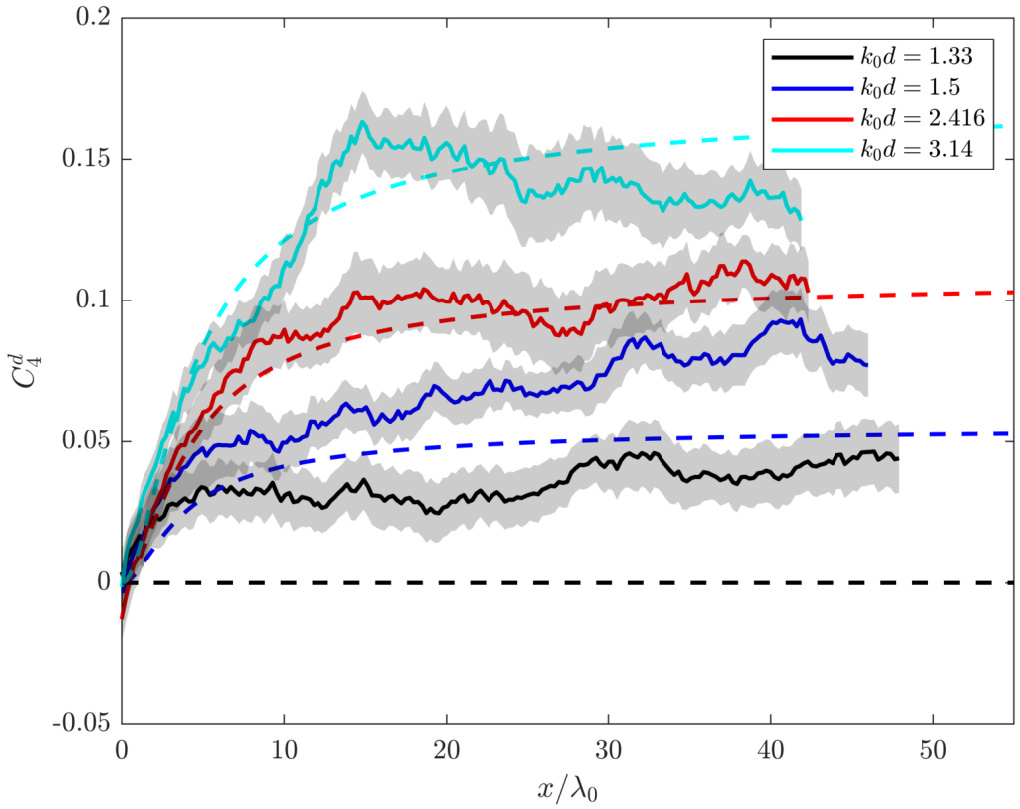

Figure 6.2: Evolution of dynamic excess kurtosis for case 1 along the numerical flume with different water depth. The dashed line shows the corresponding theoretical predictions from Fedele et al. (2010) for each water depth.

Janssen & Herbers (2009) and Fedele (2015*b*) show that for deep water waves, the bound wave contribution to kurtosis, as an independent value of space, can be approximated as:

$$C_4^b = C_4^b|_{ss} = 6(k_0\ std)^2. \tag{6.11}$$

In this section, we examine the steady-state value of bound wave contribution to kurtosis ($C_4^b|_{ss}$), which is approximated as the average value of last ten wave lengths in the numerical tank. From Figure 6.3, in relatively shallow water, the bound wave contribution to kurtosis increases significantly. This agrees well with the previous studies (Toffoli et al. 2007). On the deep water side, the theoretical predictions agrees extremely well with the simulated results for Case 2. The discrepancy between these for Case 1 could be due to the dynamic kurtosis departing from the theoretical predictions (see Figure 6.2). Unsurprisingly, Wave steepness also has a significant impact on the bound wave contribution as the contribution ratio for Case 1 is generally higher than Case 2.

6.3.3 Averaged shape of the extreme waves

The linear wave theory suggests the average shape of the largest event is given by the scaled autocorrelation function (Lindgren 1970, Boccotti 1983, Tromans et al. 1991). This approach has been referred to as NewWave (primarily in the engineering community) and as the theory of quasi-determinism (primarily in the physics community). The expected shape of an extreme event with the amplitude of the peak (at $t = 0$) scaled to unity is given by:

$$\eta(t) = \frac{1}{m_0}\int_0^\infty S(f)\cos(2\pi f t)\mathrm{d}f, \tag{6.12}$$

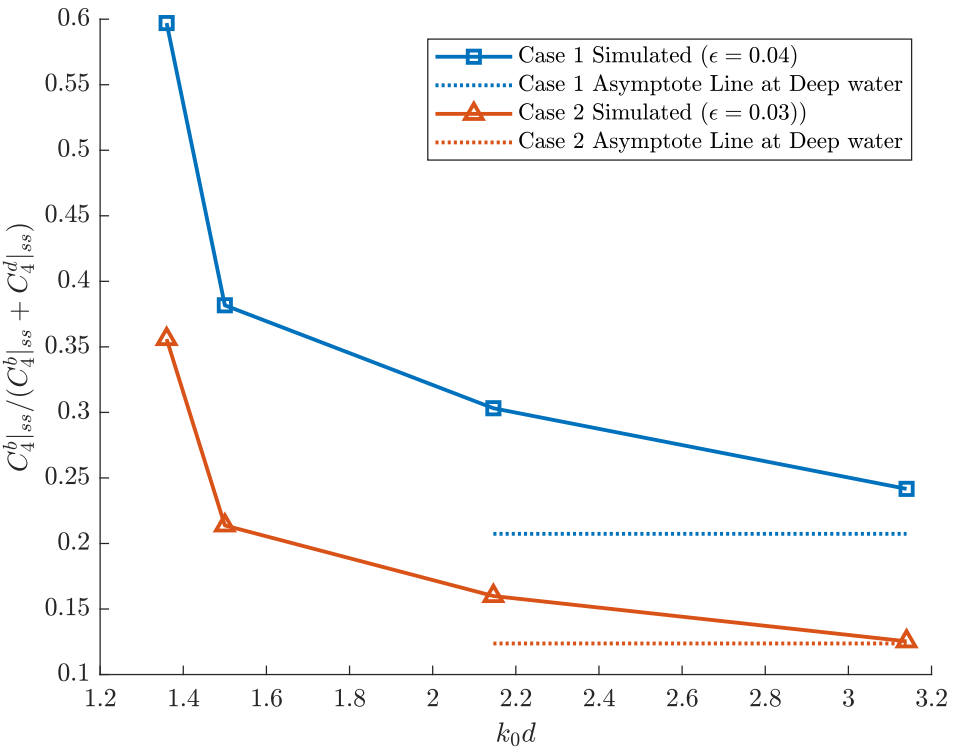

Figure 6.3: Contribution ratio from bound waves to kurtosis at steady state for different relative water depths for both Case 1 and Case 2. Solid lines shows the simulation results and dashed lines indicates the theoretical value obtained from Equation 6.11.

where $S(f)$ is the wave spectrum, $m_0 = H_s^2/16$ is the zeroth moment of the wave spectrum. However, the nonlinear physics might modify the shape of these extreme events. For deep-water waves, nonlinear changes on the shape of extreme events have been investigated for wave groups and random seas (see Baldock et al. (1996)). Adcock & Taylor (2009*b*) shows that the group would contract spatially according to the analytical results based on NLS. For unidirectional random waves, the work of Lo & Mei (1985) and Dematteis et al. (2019) looked at the averaged shapes of extreme events in deep water and the recent work of Karmpadakis et al. (2020) investigates these extreme events in finite water depth.

We herein present the change in the averaged shape of the extreme events (in the time-domain) and look at how this changes as the waves evolve spatially along the wave tank. Figure 6.4 shows the averaged free surface of the largest events at different positions in the numerical tank. When the waves are being generated, the wave group is symmetric horizontally as predicted by the linear theory. Nonlinear physics modifies the shape of these largest events as the wave propagates along the wave tank. From Figure 6.4, two main modifications from nonlinear physics are that the largest wave tends to move to the front of the wave group (horizontal asymmetry) and the contraction of the wave group. The finite water depth seems to have limited impact on the averaged shape of extreme events. The results presented here have had the bound harmonics removed to facilitate easier comparison with linear theory and between water depths.

We quantify the change of the extreme events with two parameters (see Tang, Tromans & Adcock (2019) for details). The first parameter measures the width of the wave group and the second parameter measures the wave group asymmetry. From Figure 6.5, σ_1 and σ_2 is defined as the duration when the envelope height exceeds 80% of the maximum height of the envelope averaged from top 20 or 5 largest extreme events from a 32 minutes simulations. The wave envelope is calculated using a Hilbert

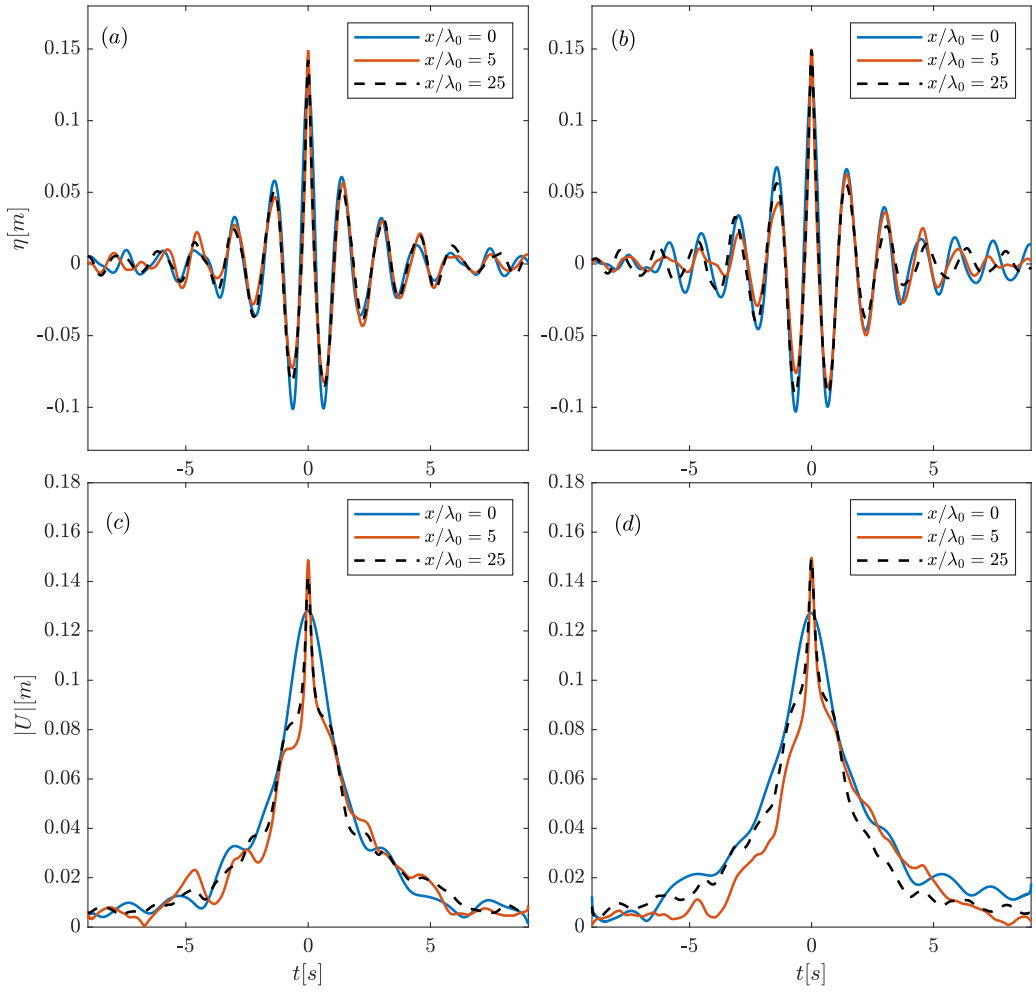

Figure 6.4: Average shape of the 5 largest crest profiles (a, b) and the corresponding envelope profiles (c, d) out of over 1280 waves at $x/\lambda_0 = 0, 5, 25$ for random waves with a, c: $k_0 d = 1.33$ and b, c: $k_0 d = 3.14$.

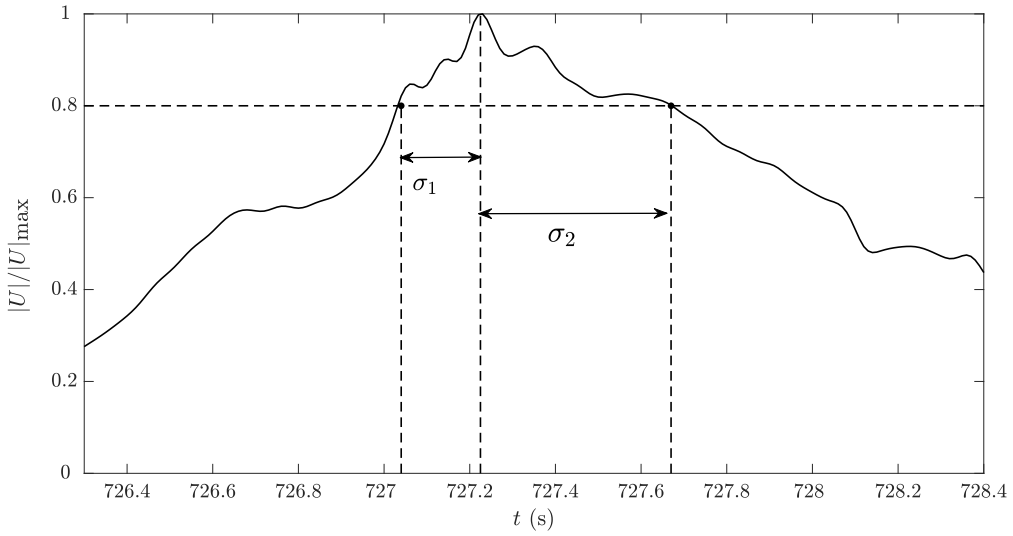

Figure 6.5: Illustration of the envelope duration when the normalised envelope height $|U|/|U|_{\max}$ exceeds 80% of its peak height.

transform:

$$|U| = \sqrt{\eta^2 + \eta_H^2}, \tag{6.13}$$

where $|U|$ is the wave envelope and η_H is the Hilbert transformed surface elevation.

We define the parameters B_1 and B_2 to be the ratios between the measured averaged duration σ_1 and σ_2 and the duration that would be predicted by the linear theory based on the local spectrum:

$$B_1 = \frac{\sigma_{1,\text{measured}}}{\sigma_{1,\text{quasi-determinism}}}, \quad B_2 = \frac{\sigma_{2,\text{measured}}}{\sigma_{2,\text{quasi-determinism}}}. \tag{6.14}$$

We then establish B_{mean} and ΔB as parameters describing changes in the group width and the group asymmetry when compared to the linear theory:

$$B_{\text{mean}} = \frac{B_1 + B_2}{2}, \quad \Delta B = B_2 - B_1, \tag{6.15}$$

A positive value of $1 - B_{\text{mean}}$ suggests the group tends to contract relative to the

shape expected under linear evolution. A positive value of ΔB implies the largest wave tends to move to the front of the group.

Figure 6.6 shows the evolution of envelope asymmetry for the top 20 largest events in 32 min simulations (approximately 1280 waves) of case 2 in different water depths. Among all of the water depths, we observe a positive value of ΔB, which suggests the largest wave moves to the front of the wave group. Despite the waves propagating in different water depths, the general trends of the evolution of horizontal asymmetry of largest events are similar. ΔB starts at zero as the generated waves are linear random waves. Soon after the wave departs from the relaxation zone, strong wave asymmetry is observed for all cases. This is where the sea-state is furthest from its equilibrium state. The horizontal asymmetry then peaks at slightly different positions for different water depths and settles down to a smaller near-zero steady state value. Waves propagating in shallower water depths seem to have slightly higher asymmetry value at the peak, especially for the shallowest water case ($k_0 d$ = 1.33). At steady state, there seems to be a trend that waves have slightly higher asymmetry value in shallower water. Unfortunately, due to the asymmetry value being generally small at steady state and sensitive to the inherent randomness of the wave field, we can not draw a general conclusion. The length scales of the initial increase for different cases seems to be similar for all the water depths.

Figure 6.7 shows the average contraction of the wave group around an extreme event given by $1 - B_{\text{mean}}$. We observe a positive value of $1 - B_{\text{mean}}$ for all the cases, which indicates the wave groups tend to contract compared to the evolution under linear theory. Different from the evolution of horizontal asymmetry, the contraction of the wave group increases initially before flattening off when equilibrium is achieved. The contraction occurs on a slightly different spatial scale compared to the horizontal asymmetry, but the spatial scale of initial contraction increase for all different water depths seem to be similar. There seems to be a slight decreasing trend in the envelope

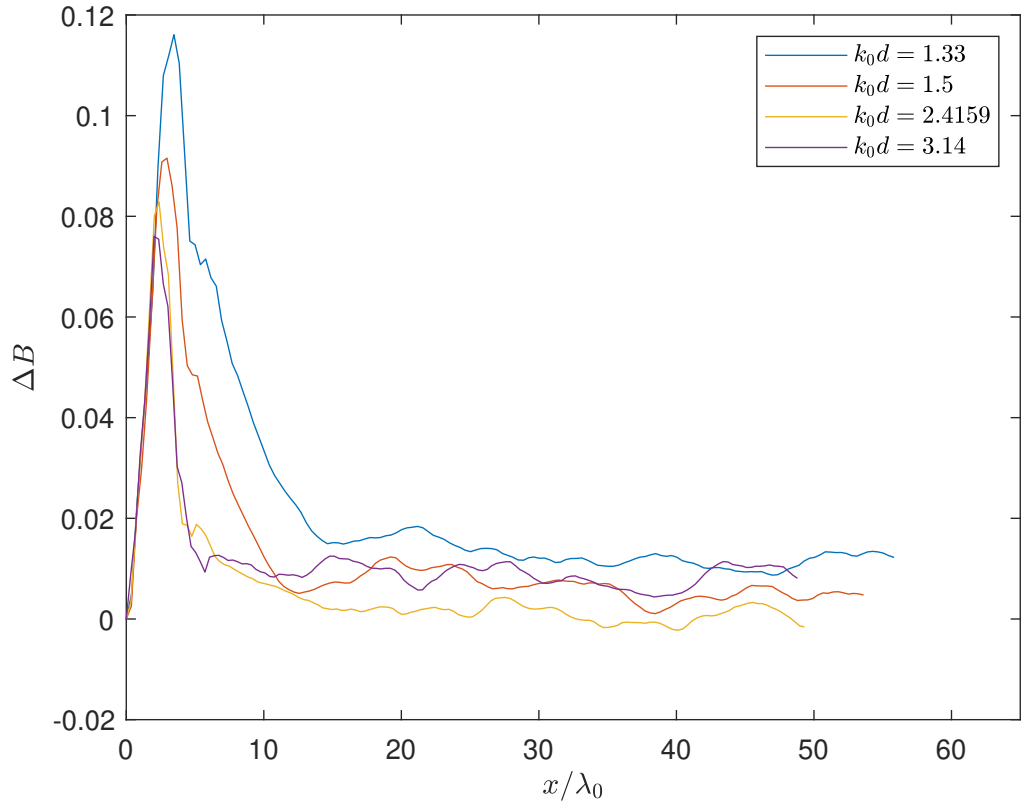

Figure 6.6: Change in envelope asymmetry (ΔB) for top 20 largest events in each realisation of case 2 with different water depth.

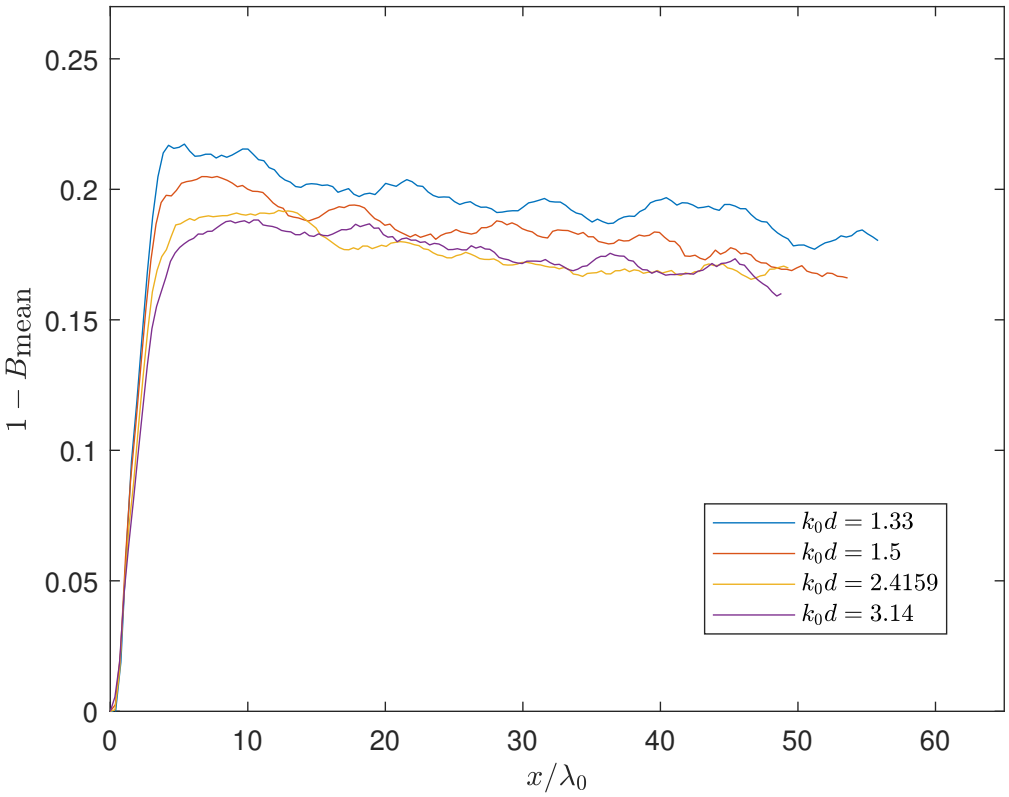

Figure 6.7: Nonlinear change in the duration when the envelope height of top 20 largest events exceeds 80% of its peak height ($1 - B_{\mathrm{mean}}$) in each realisation of case 2 with different water depth.

contraction for all the cases at different water depths. However, the length scale of this decreasing trend is too large to be considered in this study.

In the recent work of Karmpadakis et al. (2020), comparison of the horizontal asymmetry is made between the deep water case and the shallow water case (their Figure 14). Their results show that for medium and relatively large waves, cases with shallower water depths show more horizontal asymmetry, which agrees well with our results in Figure 6.6. They also find that, particularly for deep water wave, there is a greater level of horizontal asymmetry for the extremely large waves. To further investigate the wave depth effects on extremely large waves, we performed the same data analysis method but only for the top 5 largest waves in a 32 min time series.

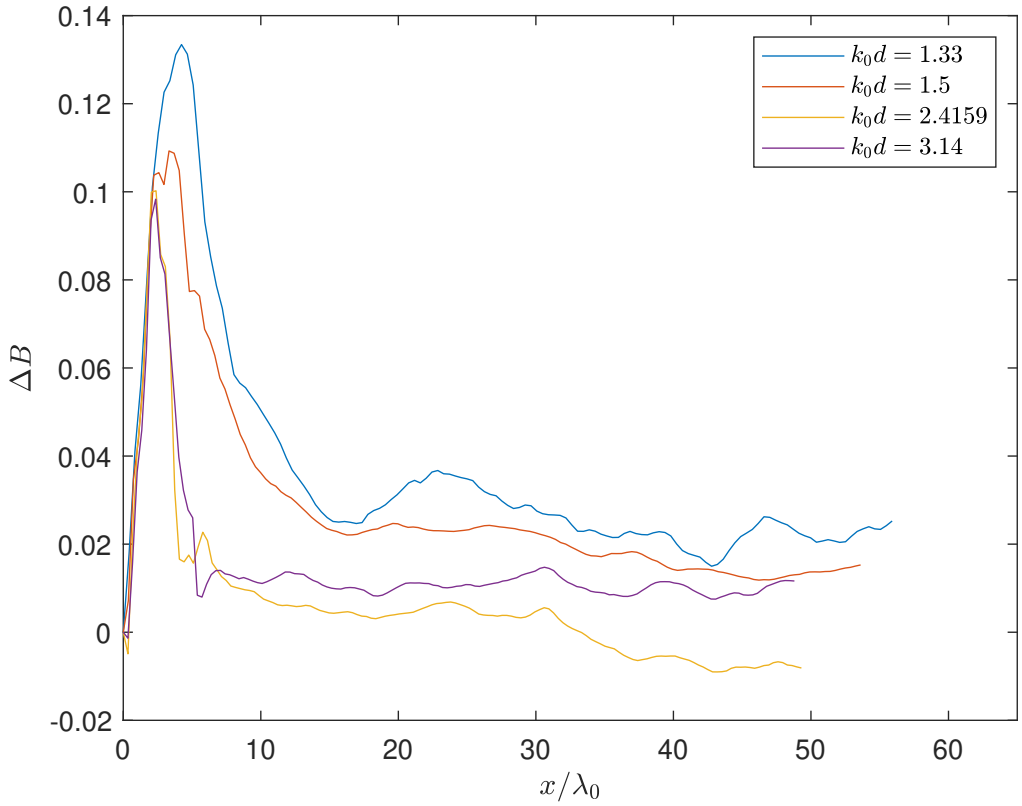

Figure 6.8: Change in envelope asymmetry (ΔB) for the top 5 largest events in each realisation of case 2 with different water depth.

This will give us the measure of the averaged shape of extreme events with exceedance probability abound 10^{-4}.

Figure 6.8 shows the evolution of envelope asymmetry for the top 5 largest events in different water depth. From Figure 6.8, the overall trends and the spatial scale of the averaged envelope asymmetry of the top 5 largest events are quite similar with that of the top 20 largest events shown in Figure 6.6. However, the top 5 largest events generally show more significant horizontal asymmetry at the peak and at the steady state. The effect of water depth on the envelope asymmetry is greatly reduced. This observation agrees well with the results in Karmpadakis et al. (2020).

In addition to the similar horizontal asymmetry results as Karmpadakis et al. (2020), we further investigate the contraction of the envelope for top 5 largest events

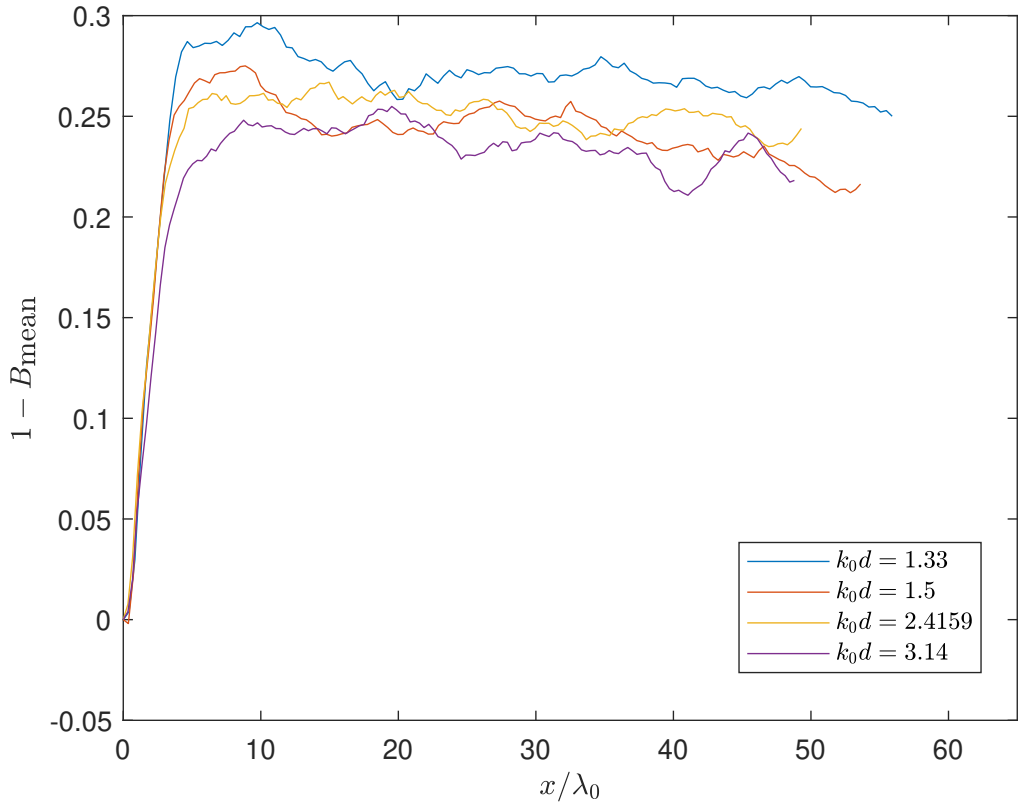

Figure 6.9: Nonlinear change in the duration when the envelope height of top 5 largest events exceeds 80% of its peak height ($1 - B_{\text{mean}}$) in each realisation of case 2 with different water depth.

at different water depth in Figure 6.9. The general trends are similar to the top 20 largest events shown in Figure 6.7, except the top 5 largest events have a greater contraction when compared to the top 20 largest events. As such, our results are consistent with the study of Karmpadakis et al. (2020)

6.4 Discussion and conclusions

This paper studies how the proprerties of extreme waves evolve in wave-tanks. We examine both the amplitude of the waves but also the changes to the average shape of an extreme wave group. We use fully non-linear numerical simulations to study these across a variety of depths in intermediate and deep water. For all these parameters,

the initial state is not in equilibrium and so these go through a transient period before settling down to an equilibrium state.

We use the kurtosis of the free surface as a proxy for the number of extreme waves. At the paddle the kurtosis of the linearised signal is 3. In all cases we see an increase in kurtosis down the tank although this is significantly supressed by finite depth effects. This suppression of nonlinear wave focussing agrees well with the previous literature (Janssen & Onorato 2007, Mori & Yasuda 2002, Whitham 1974) and experiments (Toffoli et al. 2009). This may not be desirable for engineers engaged in model testing in unidirectional flumes since this increase in the number of large waves would not be expected in the open ocean. We have adapted an analytical model to account for finite depth which captures the leading order physics. We propose this as a simple model which could be used by engineers to understand whether an undesirably large number of big waves will occur in their experiments. The effects described in this paper are driven by non-linearity and, unsurprisingly, will be more significant in steeper sea-states. For model testing in directional spread flumes, increase in the number of large waves due to oblique perturbations should also be considered (see results in Fernandez et al. (2014), Slunyaev et al. (2002), Toffoli et al. (2013, 2009)).

The width and duration of a large wavegroup behaves consistently across all depths. There is a monotonic transition from the initial state to the equilibrium state. There appears to be very little variation in this with depth which is interesting suggesting that the cubic term in the NLS may not be driving this. The greatest contraction appears to be for the shallowest case. This result goes somewhat against the results of Adcock & Yan (2010) who studied isolated wave groups and found the contraction was inhibited by depth although not by as much as would be expected using predictions from the cubic NLS. However, our results show that the departure from the shape expected in linear theory is transient and average group shape returns to one close to that predicted by linear theory. Thus, if the group shape is critical in

a test, it is suggested the structure to be tested is placed after this transient period (roughly 20 wavelengths away from the paddle).

Our results here are intended as a guide to those undertaking experimental work with random waves although we hope to have made a contribution to understanding the physics of these processes.

Chapter 7

Data driven analysis on the extreme wave statistics over an area

7.1 Abstract

In this paper we analyse ocean wave crest statistics over different sized areas using data driven methods. We use second order numerical simulations to generate extreme crest data. We consider a simplistic Gumbel distribution fit as well as using a Random Forest Model to map the sea-state parameters to extreme crest values. Our simulations are compared with the existing distributions in the literature. We find that existing distributions perform well for more straightforward cases but that as more parameters are introduced the data science approach can capture features other methods cannot. Our approach also highlights the importance of different parameters such as steepness or length in the mean wave direction. We conclude that machine learning model is promising approach to predicting wave crest distributions in complex scenarios.

7.2 Introduction

Short-term wave crest distributions are essential for the design of offshore infrastructure. For example, an important criterion in the offshore platform design is to leave enough air gap between the mean sea level and the deck, which requires an accurate estimation of the crest distributions based on the sea state parameters. Linear theory predicts the short-term crest distribution closely follows the Rayleigh distribution. However, real water waves are nonlinear and this physics modifies the crest statistics.

The water waves which occur naturally in the open ocean are directionally spread. This makes a fundamental difference in the nonlinear wave interactions and largely inhibits the build up of correlations between freely propagating components (Fedele et al. 2016, 2013, Fedele 2015*b*). This indicates the extra amplification is usually small when compared to statistics predicted by second order models provided the spectrum is in equilibrium. This has been shown in experimental and numerical simulations in Xiao et al. (2013), Toffoli et al. (2010), Onorato et al. (2006), Annenkov & Shrira (2009) as well as by field data (Christou & Ewans 2014). Thus, in the present paper we model the waves as second order which implies that the underlying dynamics are linear.

Second order harmonics make the crests higher than predicted by linear theory, which is particularly important in steep sea states. Forristall (2000) proposed a Weibull distribution with empirical fitted parameters to capture the additional amplitude from second order contribution. His distribution is widely used in engineering practice. Fedele & Tayfun (2009) replaced the empirical fitting with a rigorous theoretical framework. The existing crest distribution models predicts wave statistics well when compared to both experimental and field measurements (Onorato et al. 2004, Fedele & Tayfun 2009, Cherneva et al. 2009, Forristall 2007). Additionally. higher order wave-wave interactions can further modify the crests statistics (Janssen 2003). Tayfun & Fedele (2007) have established short term crest distributions accounting

for the additional crests amplitude from these higher order nonlinear interactions (see also some further development of statistics of oceanic crests in Fedele & Tayfun (2009), Fedele et al. (2016, 2017)).

These models are developed and verified against the waves measured at a single Eulerian point. For large engineering structures using point statistics can underestimate the magnitude of the wave crests (Forristall 2006). Hence, the wave statistics over an area becomes more relevant for engineering designs (Forristall 2006, Dysthe et al. 2008, Fedele et al. 2013). The extra amplitude arises from the fact that a point measurement will tend to miss the true crest of the wave (which dominates for small areas) as well as there simply being more waves sampled (which dominates at larger areas) (Fedele 2012, Forristall 2011, Fedele et al. 2013, Benetazzo et al. 2015).

Instead of the short term crest distributions, we study the maximum crest distributions, which only considers one maximum crest in a space-time ensemble. The maximum crest distributions are more stable, and can be predicted with several existing theoretical models. Adler (1981) and Piterbarg's theorem (Piterbarg 1996) provide estimations of the impact of the area to the maximum crest distribution for Gaussian multidimensional random wave fields. Piterbarg's theorem was extended to second order by Socquet-Juglard et al. (2005) with a scaling parameter for second order contributions and Forristall (2006) with empirical fitting accounting for second order effects. Fedele (2012) presented a model based on the Adler and Taylor's Euler characteristics approach (Adler 1981, Adler & Taylor 2007, Fedele et al. 2012), which is further extended in Fedele et al. (2012) to account for second order physics. These theoretical model have been found to be reasonably accurate when compared to field measurements and numerical simulations (Benetazzo et al. 2012, Forristall 2015, Benetazzo et al. 2015).

In the present paper we have taken an alternative approach. Rather than fitting data with a parameterised distribution we take a data science approach to the prob-

lem. There has been a rapid expansion of the application of these data driven methods in recent years as a results of both the significant improvement in the methodology of data driven methods and the boom in computational hardware. The training and validating of these data driven methods are more accessible than ever before. As a result, data driven methods are starting to be used in ocean environment studies (Ali & Prasad 2019, Callens et al. 2020, Brunton et al. 2020, Sapsis 2020*b*, Ragone & Bouchet 2019), and predicting extreme statistics with active sequential sampling methods (Gramstad et al. 2020, Mohamad & Sapsis 2018, Sapsis 2020*a*, Ragone & Bouchet 2019, Dripta & Dutykh 2020) and probabilistic decomposition method (Mohamad & Sapsis 2016, Cousins & Sapsis 2014, Farazmand & Sapsis 2017, Mohamad et al. 2016). The object of this paper is primarily to demonstrate the capability of taking a machine learning approach to the problem of predicting short-term crest statistics from a given input spectrum. We also compare the existing distributions against our numerical data.

In this paper, we first briefly introduce our simulation method in section 7.3.1, which forms the database for model training and comparison. The detailed dataset information is given in section 7.3.3. We also present a step-wise prediction routine in section 7.3.4 to clarify the methodology we used to compare different wave statistical models. We describe the theoretical models in section 7.4.1 and data driven models in section 7.4.3 and 7.4.4. We start with second order maximum crest distributions at a single point as a limiting case in section 7.5.1. We first aim to examine the feasibility of data driven methods under a simplified condition. We further extended the simulation into waves over an area, and compared the performance of different models in section 7.5.2 for linear waves and in section 7.5.3 for second order waves. In these sections, we increase the complexity of the simulated ocean environment by including second order effects to further explore the potential of data driven methods in predicting real wave statistics in the open ocean. We summarise the performance

of different models in section 7.5.4, and present an analysis of the relative importance of input parameters in section 7.5.5.

7.3 Methodology

7.3.1 Numerical simulations

To investigate the extreme wave statistics over an area, we have developed an envelope based code to simulate directional spread waves over an area at finite water depth with optimal speed. For the linear part, the wave surface elevation can be obtained as:

$$\eta_{linear}(\mathbf{x},t) = \sum_{i=1}^{\infty}[a_i \cos(\mathbf{k}_i \cdot \mathbf{x} - \omega_i t) + b_i \sin(\mathbf{k}_i \cdot \mathbf{x} - \omega_i t)], \tag{7.1}$$

where $\mathbf{x}$ is the position vector, $\mathbf{k}_i$ is wavenumber vector. a_i and b_i are the independent random variables drawn from an normal distribution $N(0,\sigma_i^2)$. The variance of each wave component σ_i^2 is defined by the spectrum of the sea-state as $\sigma_i^2 = S(\omega_i)\Delta\omega$. Two dimensional inverse Fast Fourier Transform is applied to improve the speed of the simulations on the linear part.

To further optimise this simulation code for the speed, we also applied a multi-scale time-space frame to obtain the local maximum crest height over a given area and duration. The first simulation is run on a coarse time step at 0.75 seconds and a relatively coarse spatial step (0.3 meters). The time when the local maximum of the envelope occurs is recorded. The second simulation then starts from two periods before that maximum event to two periods after. This simulation is run on a fine time step of 0.05 seconds and a fine space step of 0.1 meters to obtain the accurate value of this maximum crest height. Rigorous tests have been performed to confirm this multi-scale time frame method can capture the maximum events accurately.

All the second order corrections were computed with a hybrid envelope method. The leading order coefficients of the sum terms and difference terms are corrected according to Slunyaev (2005). The coefficients for the broad banded corrections are interpolated from the exact second order simulations (Dalzell 1999) (see C for details). Despite the narrowbanded approximation used the results agree closely with simulations using Sharma & Dean (1981) type double summation.

7.3.2 Input conditions

In this study, we simulate waves using a directional JONSWAP spectrum $S_{dir}(f, \theta_s)$ with $\gamma = 3.3$ (Hasselmann et al. 1973). We used a normal directional spreading function $D(\theta_s)$ with a spreading of $\mu_\theta = 22°$ to specify the energy distribution, which is given by:

$$D(\theta_s) = \frac{1}{\sqrt{2\pi}\mu_\theta} \exp\Big(- \frac{\theta^2}{2\mu_\theta^2}\Big), \tag{7.2}$$

where θ is the angle deviated from the mean wave direction and μ_θ is the directional spread. For all the simulated cases, the peak wave period is 1.5 seconds, and we simulate 150 periods for all the realisations in this study.

In this study, we have covered seven input conditions, which are summarised in Table 7.1

We choose the test condition for sections from 7.5.1 to sections 7.5.2 to be a single sample within the input range. This provides details in the difference between the shape of maximum crest distributions from different models. However, we choose the relative error in the expected value of maximum crest to be the test matrix because of its significance in ocean engineering applications.

Table 7.1: Input and test conditions for all the cases in this study. $H_s k_p$ is the wave steepness parameter, x and y are the side length of sampling area along mean wave direction and transverse direction respectively, and $k_p d$ is the relative water depth parameter.

Section	**Sampling**	**Simulation**	**Input range**	**Test condition**
7.5.1	Point	2^{nd} order	$H_s k_p$ = 0.0358 to 0.25	$H_s k_p$ = 0.1788
7.5.1	Point	2^{nd} order	$k_p d$ = 1.4 to 7	$k_p d$ = 2.5
7.5.1	Point	2^{nd} order	$H_s k_p$ = 0.0358 to 0.25 $k_p d$ = 1.4 to 7	$H_s k_p$ = 0.178, $k_p d$ = 2.5
7.5.2	Square	Linear	$x = y$ = 0.5λ to 20λ	$x = y$ = 3.5λ
7.5.2	Rectangular	Linear	x, y = 2.5λ to 20λ	x, y = 4λ to 18λ
7.5.3	Rectangular	2^{nd} order	x, y = 2.5λ to 20λ	x, y = 4λ to 18λ
7.5.3	Rectangular	2^{nd} order	x, y = 2.5λ to 20λ, $H_s k_p$ = 0.075 to 0.25	x, y = 4λ to 20λ, $H_s k_p$ = 0.1 to 0.25

Table 7.2: Summary of different types of dataset used in this paper

Name	**Usage**	**Realizations per Case**	**Domain Coverage**	**Note**
Fitting dataset	Empirical Fitting	2000	Grid-based	To obtain stable statistical results
Training dataset	Random Forest	500/2000	Random	RF favours more simulation cases for better coverage
Test dataset	Performance evaluation	5000/10000	Grid-based	To provide the most accurate results

7.3.3 Datasets types

In this study, for each input condition, we simulated three different types of datasets for model predictions and performance assessment, which are summarised in Table 7.2. For all the test cases, a rigorous check on the statistical variance of distribution parameters is performed, additional realisations are simulated to reduce the confidence interval.

7.3.4 Prediction routine

In this study, we applied following prediction routine to examine the performance of different wave statistical models as:

1. Based on the input parameters listed in Table 7.1, numerical simulations are performed to obtain maximum crest distributions.

2. Gumbel distribution is fitted to parameterise the maximum crest distributions for fitting, training and test dataset (see Table 7.2 for dataset details). Figure 7.1 shows a typical parameter space of coefficient A when only one parameter (H_s) is varied. Figure 7.7 shows a typical parameter space of coefficient A when two parameters (H_s and $k_0 d$) are varied simultaneously.

3. Wave statistical models are applied based on the input sea state parameters to predict the maximum crest distributions.

4. Models predictions are compared with the test dataset to determine the error.

5. Step 1-4 are repeated for all the different input conditions listed in Table 7.1.

7.4 Wave statistical models

7.4.1 Theoretical models

Forristall2006 model

For short-term wave distribution, linear waves crests measured at a given point closely follow the Rayleigh distribution. For second order waves, Forristall (2000) proposed a two parameter Weibull distribution:

$$P(\eta) = \exp\Big[- \Big(\frac{\eta}{\alpha_F H_s}\Big)^{\beta_F} \Big]. \tag{7.3}$$

For deep water waves, α_F and β_F were found by Forristall to be:

$$\alpha_F = \sqrt{1/8} + 0.2568 S_1 \quad \text{and} \quad \beta_F = 2 - 1.7912 S_1, \tag{7.4}$$

where S_1 is the steepness of the sea state.

However, for most offshore structures, the design of these platforms should be able to survive a wave crest reaching any part of the platform area. Hence, it is vital for the designer to be able to predict the maximum crests distribution over a given area during a certain period.

As for maximum crests distribution over an area, the linear statistics closely follow the Gumbel distribution according to the Piterbarg theorem (Piterbarg 1996). For second order waves, Forristall (2006) proposed a two-parameter Gumbel distribution:

$$P(\eta_{max} > s\sigma) = \exp\{-\exp[-(-B + s)/A]\}, \tag{7.5}$$

where parameter A and B are linked with α_F and β_F:

$$A = (\beta_F/4\alpha_F)(\mathrm{Log}N)^{1-1/\beta_F} \quad \text{and} \quad B = 4\alpha_F(\mathrm{Log}N)^{1/\beta_F}, \tag{7.6}$$

and N is the equivalent number of waves, which can be estimated as follows: For a point measurement,

$$N = T/T_z. \tag{7.7}$$

For relatively large areas,

$$N = \frac{2\pi XYT}{1.25\lambda_x\lambda_y T_z}. \tag{7.8}$$

For relatively small areas,

$$N = \frac{2L_x T}{\lambda_x T_z}. \tag{7.9}$$

For this theoretical model, the value of α_F and β_F are obtained by following

equation 7.4. The value of A and B are then computed following equation 7.6 and 7.7. The calculated coefficients are used to generate the curve referred as 'Forristall2006'.

7.4.2 Forristall2015 model

Forristall provides an updated fit for linear waves over an squared area and recommends a simple second order correction with a simple multiplication in Forristall (2015). The updated fit suggests that the effects from the duration and side length of the sampling area is addictive, the expected value of maximum crest height for linear waves can be estimated as:

$$E\left(\eta_{\max} \mid \text{ area }\right)/\sigma = E\left(\eta_{\max} \mid \text{ point }\right)/\sigma + f(L/\lambda), \tag{7.10}$$

where L is the side length of the squared area, the expected value of for a given Gumbel distribution is

$$E\left(\eta_{\max}\right) = (B + 0.5772/A)\sigma, \tag{7.11}$$

and the fitting function $f(L/\lambda)$ of side length is given as:

$$f(L/\lambda) = 0.9829 + 04170\ln(L/\lambda) + 0.0427\ln^2(L/\lambda) \tag{7.12}$$

Unfortunately, although this updated fit can provide accurate estimation on the expect value of maximum crest height, the performance of this model is not fully examined in this study due to the lack of distribution curve information.

Tromans and Vanderschuren fitting equation

This is a fit which follows the report OTC 7683 and the extreme crests distribution can be expressed as:

$$P(\eta_{max} > s\sigma) = \exp\{-\exp[-\mathrm{Log}(N_p)((-b_p + s)^2 - 1)]\}, \tag{7.13}$$

We note that the equation 7.13 is originally intended for maximum wave heights distributions with varying wave spectra during a storm at North Sea, which is different from the cases studied in this study. As this modified Gumbel distribution shows excellent performance in the original report, we used a modified Gumbel distribution in similar format for maximum crest distributions. There are no theoretical expressions for the value of N_p and b_p in the original report. In this paper, the value of N_p and b_p are obtained by fitting the results from the simulations.

Fedele2012 model

Fedele (2012) presented a model based on the Adler and Taylor's Euler characteristics approach (Adler 1981, Adler & Taylor 2007, Fedele et al. 2012) for predicting the probabilities of the highest crest within the space–time ensemble, which gives:

$$\begin{aligned} &P_{F2012,max}\{\eta_{max}/\sigma > z\} \approx \\ &\exp\Big\{-\exp\Big[-(z-h)\Big(h-\frac{2N_V h + N_S}{N_V h^2 + N_S h + N_B}\Big)\Big]\Big\}, \end{aligned} \tag{7.14}$$

where the modal value h satisfies:

$$(N_V h^2 + N_S h + N_B)\exp(-h^2/2) = 1, \tag{7.15}$$

where N_V is the average number of waves within the volume, N_S for the waves on the boundary surfaces and N_B accounts the waves along the perimeter (see Fig.1 in Fedele (2012) for details). The formula for N_V, N_S, N_B are given as:

$$\begin{aligned} N_V &= 2\pi\frac{XYD}{L_x L_y T_m}\sqrt{1-\alpha_{xt}^2-\alpha_{yt}^2-\alpha_{xy}^2+2\alpha_{xt}\alpha_{yt}\alpha_{xy}}, \\ N_S &= \sqrt{2\pi}\Big(\frac{XD}{L_x T_m}\sqrt{1-\alpha_{xt}^2}+\frac{XY}{L_x L_y}\sqrt{1-\alpha_{xy}^2}+\frac{YD}{L_y T_m}\sqrt{1-\alpha_{yt}^2}\Big), \\ N_B &= \frac{X}{L_x}+\frac{Y}{L_y}+\frac{D}{T_m}, \end{aligned} \tag{7.16}$$

where T_m is the mean wave period, L_x is the averaged wave length in the mean wave direction and L_y is the averaged wave length in the lateral direction. α_{xy}, α_{xt}, α_{yt} are the irregularity parameters of the sea state. Both wave parameters and irregularity parameters can be calculated from the moments of the directional wave spectrum $S(f,\theta)$:

$$T_m = \sqrt{\frac{m_{000}}{m_{002}}}, L_x = 2\pi\sqrt{\frac{m_{000}}{m_{200}}}, L_y = 2\pi\sqrt{\frac{m_{000}}{m_{020}}}, \tag{7.17}$$

$$\alpha_{xt} = \frac{m_{101}}{\sqrt{m_{200}m_{002}}}, \alpha_{yt} = \frac{m_{011}}{\sqrt{m_{020}m_{002}}}, \alpha_{xy} = \frac{m_{110}}{\sqrt{m_{200}m_{020}}} \tag{7.18}$$

where $m_{i,j,l}$ is the moments of the directional spectrum, which can be computed as:

$$m_{i,j,l} = \int\int k_x^i k_y^j f^l S(f,\theta) df d\theta, \tag{7.19}$$

where the k_x and k_y are the wave number vector component in x and y directions respectively. The expected value of the maximum crest height for linear waves is given as:

$$E_{max}/\sigma \approx h + \frac{\gamma}{h - \frac{2N_V h + N_S}{N_V h^2 + N_S h + NB}}. \tag{7.20}$$

Based on the results in Fedele & Tayfun (2009), Fedele further extend extended the theory with second order corrections in Fedele et al. (2013) and compare the model perdition with field data. The extended version of space-time maximum crest distributions is futher compared with field data in Benetazzo et al. (2015) and found to be relatively accurate. In the second order version of Fedele2012 model, the probabilities of the highest crest within the space–time ensemble corrected to second order

within the narrow banded limit is given as:

$$P_{F2012\text{-}2,max}\{\eta_{max}/\sigma > \xi\} \approx \\ \exp\Big\{-\exp\Big[-\frac{1}{1+\mu h}(\xi - h - \frac{\mu^2}{2}h^2)\Big(h - \frac{2N_V h + N_S}{N_V h^2 + N_S h + N_B}\Big)\Big]\Big\}, \tag{7.21}$$

where μ is wave steepness, which can be computed as:

$$\mu = \mu_m(1 - \nu + \nu^2), \tag{7.22}$$

where $\mu_m = \sigma(2\pi m_{001}/m_{000})^2/g$ is an integral measure of wave steepness corrected with spectral bandwidth $\nu = \sqrt{m_{000}m_{002}/m_{001^2} - 1}$.

In Fedele2012 model, scale dimension parameter β is introduced as a measure of the relative scale of the wave when compared to the volume size. For linear waves with $\beta = 1$, the crest distribution closely follows Rayleigh distribution, which can be interpreted as waves at a point with varying time. The limiting case of $\beta = 2$ describes the waves with either length in x or length in y or time duration is null. For example, $\beta = 2$ can be attained for waves over an area at an instantaneous time. The maximum value of $\beta = 3$ suggests a sufficient number of waves can be observed on all three dimensions.

The β can be computed as:

$$\beta = \frac{4N_S\xi + 2N_B}{16N_V\xi^2 + 4N_S\xi + N_B}, \tag{7.23}$$

where ξ can be obtained through:

$$(16N_V\xi^2 + 4N_S\xi + N_B)\exp(-8\xi^2) = 1. \tag{7.24}$$

This Fedele2012 model is further extended to include third order nonlinearities by Fedele in Fedele (2015*a*). The extended model has been used in the analysis of the

sinking of the El Faro and has been compared with Higher Order Spectral simulations (Fedele et al. 2017). In the present work we do not examine the performance of this updated version as our numerical simulation is only to second order accurate. In this study, we include the data driven model, which requires a large amount of training data. Hence, we optimised our simulation for the speed of simulation to explore a wider range of parameter space.

7.4.3 Empirical Fitting model

Followed by Piterbarg's theorem (Piterbarg 1996) and other theoretical models (Fedele 2012, Forristall 2006, Krogstad et al. 2004), the maximum crest distributions over an area can be described as a Gumbel distribution with two parameters A and B shown in Equation 7.5. Polynomial fitting $f_{\text{fit}}(x)$ can be applied to predict the correlation between these two parameters and input parameters (i.e. length in x, length in y etc.) as:

$$\begin{aligned} A &= f_{\text{fit},A}(In_1, In_2...) \\ B &= f_{\text{fit},B}(In_1, In_2...), \end{aligned} \tag{7.25}$$

where $In_1, In_2...$ are input parameters. Herein we present a simple example for illustration purposes. A polynomial fit is used to interpolate the A and B at desired input values (see Figure 7.1). The interpolated coefficients are used to generate the curve labelled as "direct fit A and B".

Alternatively, the value of α_F and β_F can be obtained by fitting the crest exceedance distributions, following Equation 7.3:

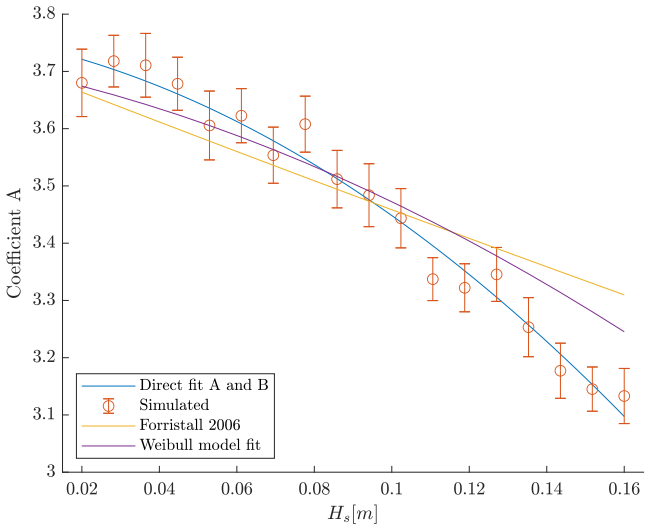

Figure 7.1: Comparison of the prediction curve from different models for A (equation 7.6) for different H_s with a heavy fitting weight towards the tail. The error bar gives 95% of confidence interval based on bootstrapping method. The duration of each simulation lasts for 150 periods with a peak period of fixed 1,5 seconds. All the cases used in this figure is simulated in deep water. Each simulated point is obtained by fitting the maximum crest of 1000 realisations.

$$\begin{aligned} \alpha &= f_{\text{fit},\alpha}(In_1, In_2...) \\ \beta &= f_{\text{fit},\beta}(In_1, In_2...). \end{aligned} \tag{7.26}$$

A polynomial fit is used to interpolate the α_F and β_F at specific input values and the value of A and B are then computed following Equation 7.6 and 7.7. The calculated coefficients are used to generate the curve labelled as "Weibull model fit".

However, the difficulty of obtaining a proper fit can be significantly increased when the number of inputs is large. Hence, for second order problems, we proposed a two step fitting method, which greatly reduces the difficulty of fitting:

$$
\begin{aligned}
A &= f_{\text{fit,linear},A}(In_1, In_2...) + f_{\text{fit,2nd},A}(\mu), \\
B &= f_{\text{fit,linear},B}(In_1, In_2...) + f_{\text{fit,2nd},B}(\mu),
\end{aligned}
\tag{7.27}
$$

where μ is the wave steepness. In this simplified model, the extra crest heights from the second order contributions are only fitted with the wave steepness and hence reduced the number of data points required.

7.4.4 Random Forest model

The Random Forest model is an ensemble learning method that predicts the output by constructing multiple decision trees (see Figure 7.2 as an example). Each tree is built based on a different bootstrapped sample of the original training dataset, which is referred to as bootstrap aggregating or bagging. Additionally, at each leaf node, only a specific number of randomly sampled features are selected as the candidates of the split. The best split point can only be selected within these nominated features. This characteristic is usually referred to as the random selection of features or feature sampling. Both bagging and feature sampling are the core features of the Random Forest models, which greatly reduce the correlations between different individual trees. In this study, we are not limiting the depth of tree splitting and minimum number of samples for each split is 2. We choose this as this provides more accurate regression performance and our results shows no over-fitting when checked against out-of-bag samples. Hence, the output value of each individual decision tree can be treated as an independent prediction, and the average of these independent predictions is the output as the predicted value of the Random Forest model (see Breiman (2001) for details).

To minimise the bias introduced during the random split of the training and validation data, we applied the k-fold cross validation process (Rodriguez et al. 2009)

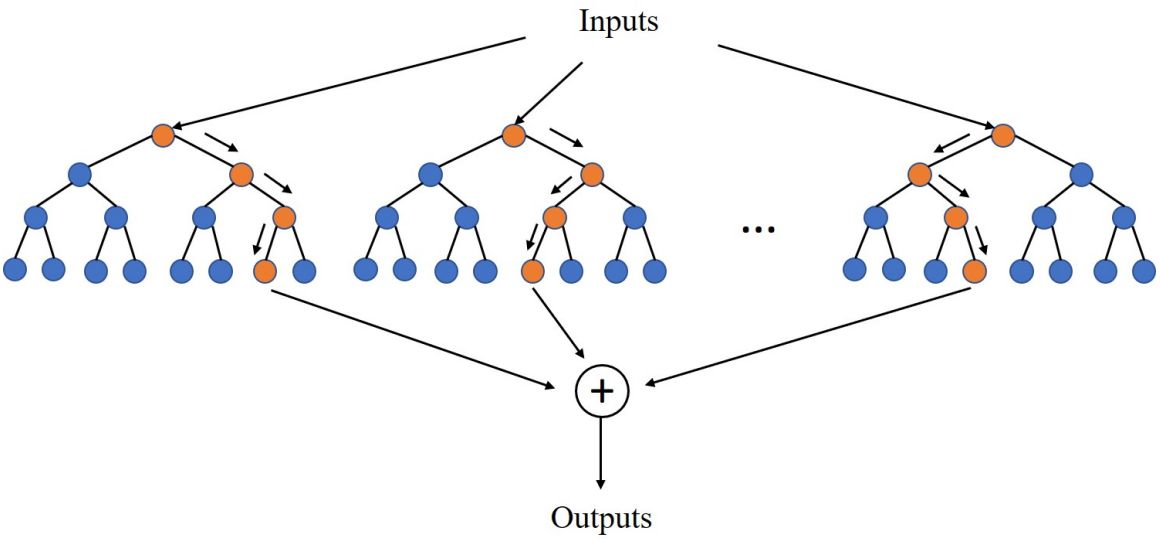

Figure 7.2: General structure of a Random Forest.

to obtain more accurate predictions based on the Random Forest model. The key idea of the k-fold cross validation process is to improve the validity and accuracy of the model by splitting the dataset into subsets and evaluating the performance of the model on each subset. For example, during a k-fold cross validation process, the dataset is uniformly divided into k subsets. The model is then trained, based on $k-1$ uniform-sized subsets. The trained model is validated against the left-out subset as the out-of-bag samples. The training and the validating process are repeated k times with a different subset as the out-of-bag samples. The performance measurements for each subset is calculated, and the average of them is treated as the final prediction accuracy.

Apart from the splitting between training and validation data, hyper-parameters also have a significant impact on the training quality of the Random Forest model and hence will influence the accuracy of the predictions. For the Random Forest model, there are two main hyper-parameters: the number of trees in the forest and minimum required samples at a leaf node. In this study, we tune all the hyper-parameters within the k-fold cross validation process. When building a Random Forest for each

fold, we also search for the optimal combinations of hyper-parameters. We use either grid search as an uninformed method for simple models or Bayesian optimisation algorithm (Snoek et al. 2012) as informed tuning method for complicated methods. Although an informed method generally takes fewer iterations and may achieve better final tuning results (Bergstra et al. 2011), a grid search method has an advantage in parallel computing for simple models.

In this study, different Random Forest models are trained for different test conditions. Although the inputs for these models vary, the output from these models are the two coefficients shown in Equation 7.5. Hence, the Random Forest model can be simplified as:

$$[A, B] = f_{\mathrm{RF}}(In_1, In_2...). \tag{7.28}$$

Based on these two predicted coefficients, the Equation 7.5 gives a probability of exceedance of the maximum crest height and the expected value of maximum crest height under given input conditions.

7.5 Results

7.5.1 Extreme second order wave statistics at a point

We start our investigation on the statistics from a point measurement as a limiting case to examine the performance of different models in a simplified test. We summarised the prediction models used in this section in Table 7.3.

Table 7.3: Summary of prediction models

Model Names	**Model Type**	**Primary Equations**	**Input Type**
Weibull model fit	Empirical Fitting	Equation 7.3 & 7.26	$P(\eta)$
TV model fit	Empirical Fitting	Equation 7.13 & 7.25	$P(\eta_{max})$
Direct fit A and B	Empirical Fitting	Equation 7.5 & 7.25	$P(\eta_{max})$
Forristall2006	Theoritical	Equation 7.5	$P(\eta_{max})$
Random Forest	Data Driven	Equation 7.5 & 7.28	$P(\eta_{max})$

Point measurement models including steepness effects

In this section, we examine the performance of different models, when steepness varies from $H_s k_p$ = 0.0358 to $H_s k_p$ = 0.25. When the wave steepness is the only parameter, the performance of different models at $H_s k_p$ = 0.1788 is shown in Figure 7.3, which is representative of other test cases with different wave steepness values. The model with the Tromans and Vanderschuren fit to the data and Random Forest model seems to predict the extreme crests distribution accurately except for a few extreme cases. For all the models following Equation 7.5, the accuracy for the intermediate values of maximum crest height are relatively poor. We also found that the direct fit a and b model outperforms other models based on the Equation 7.5. The general performance trend is similar at other wave steepness.

Apart from the exceedance plot, which focuses on the tail of the distribution, the probability density function of the extreme crests distributions is shown in Figure 7.4. This illustrates the performance around the probability density peak, which is more important for offshore engineering design as the expected maximum crest height is commonly used to describe the extreme wave conditions. All Forristall models tends to under predict the probability of relative small extreme values and over predict the probability of intermediate and extreme values. Tromans and Vanderschuren fit to the data seems to agree better with the simulated probability density function.

The weight function can also affect the fitting results and hence affects the final performance of different models. Figure 7.5 shows the results with a normal fitting

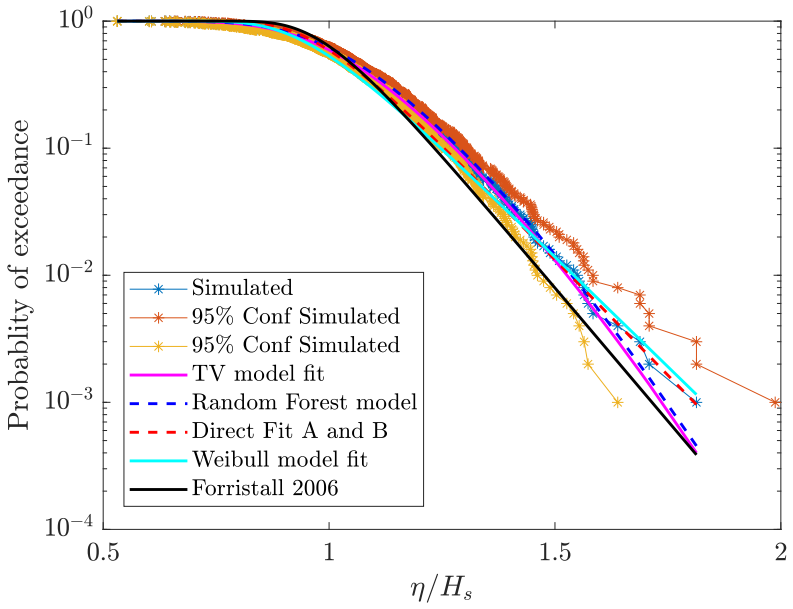

Figure 7.3: Comparison between simulated maximum crest distribution for test condition with steepness of $H_s k_p$ = 0.1788 against the predictions from different point measurement statistical models, which only consider wave steepness effects. η is the maximum crest elevation from a realisation of simulation.

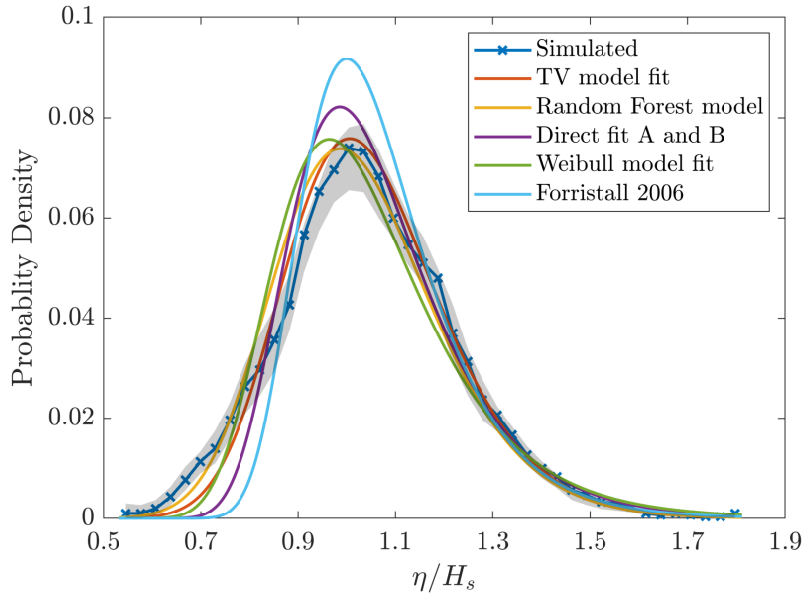

Figure 7.4: Comparison between simulated maximum crest distribution for test condition at steepness of $H_s k_p$ = 0.1788 against the predictions from different point measurement statistical models, which only consider wave steepness effects.The shaded area represents 95% confidence interval. η is the maximum crest elevation from a realisation of simulation.

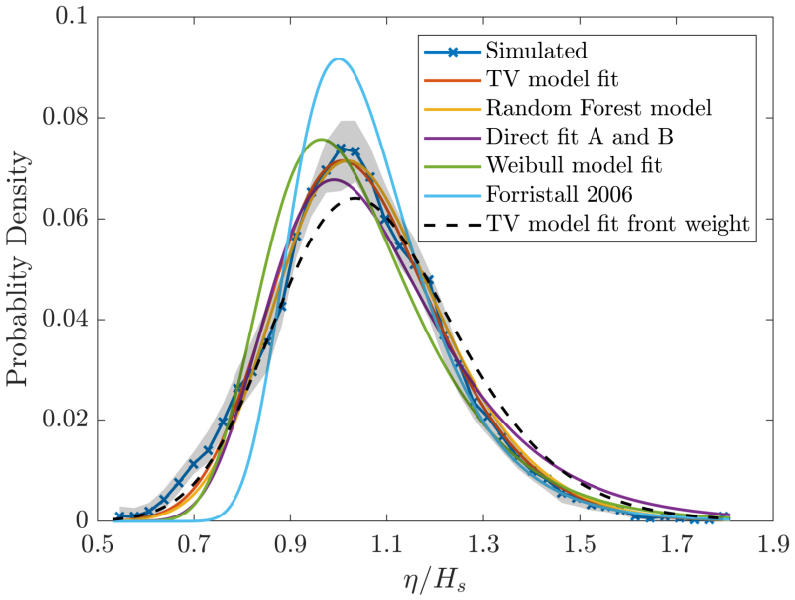

Figure 7.5: Comparison between simulated maximum crest distribution for test condition at steepness of $H_s k_p$ = 0.1788 against the predictions from different point measurement statistical models, which only consider wave steepness effects with a normal fitting weight during Gumbel fitting process (step 2 in section 7.3.4). η is the maximum crest elevation from a realisation of simulation.

weight. From the figure, the general performance of different models around the peak of the probability density function is improved but at the cost of the loss of accuracy towards the tail. As the expect value of the maximum crest height has more significance in engineering practice, we used the normal fitting weight function to examine the performance of different space-time models.

Point measurement models including water depth effects

In this subsection, we fix the wave steepness but vary the relative water depth from $k_p d$ = 1.4 to $k_p d$ = 7 to explore the impact of water depth on second order wave statistics.

Figure 7.6 compares the accuracy of different models at a fixed wave steepness of $H_s k_p$ = 0.178, when the water depth varies. We choose the water depth of test dataset to be $k_p d$ = 2.5, which represents well the model performance over the test domain.

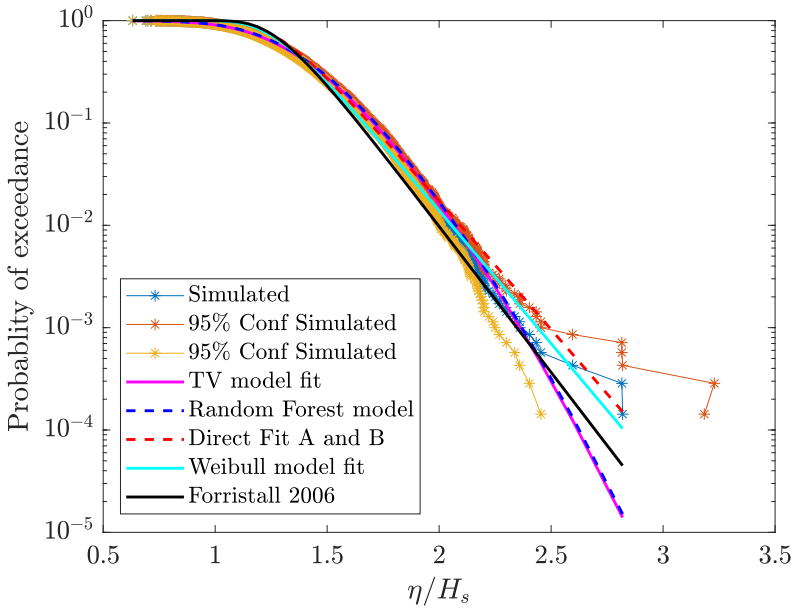

Figure 7.6: Comparison between simulated maximum crest distribution for test condition at water depth of k_pd = 2.5 against the predictions from different point measurement statistical models, which only consider finite water depth effects. η is the maximum crest elevation from a realisation of simulation for 150 wave periods.

The Tromans and Vanderschuren fit to the data works best for the majority of the cases. The Random Forest model gives almost the same results as the polynomial fit. All of the models following equation 7.5 seem to over predict the probability at the probability peak.

Point measurement models including steepness and water depth co-effects

We now analyse the performance of different models when water depth and wave steepness vary simultaneously. In this section, the relative water depth varies from k_pd = 1.4 to k_pd = 7 and the wave steepness varies from H_sk_p = 0.040 to H_sk_p = 0.25 simultaneously.

For empirical fitting models, we used a total of 49 different combinations of input parameters as a grid. All the coefficients are averaged from 5 simulations to reduce the statistical variability. This training dataset is then fitted with a third order polynomial to get the prediction curves (see Figure 7.7 as an example).

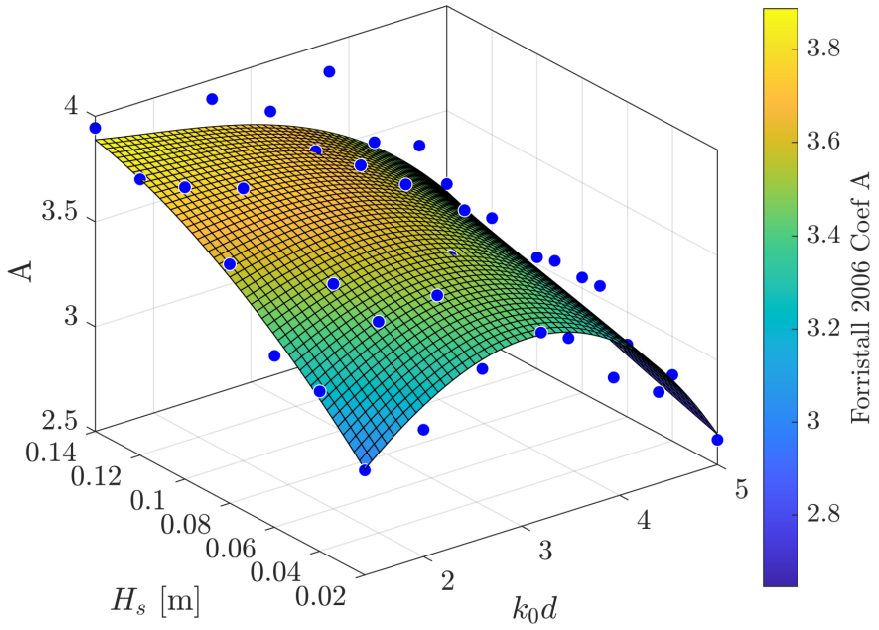

Figure 7.7: Prediction curve from Fit A and B model for coefficient A values for different significant wave heights and water depths.

A simple Random Forest model is used here to reduce the demands on the training datasets. We present a typical test case with relative water depth of $k_p d$ = 2.5 and wave steepness of $H_s k_p$ = 0.178. From Figure 7.8, Random Forest seems to have an edge when comparing to traditional fitting methods when two inputs vary at the same time. All the theoretical models have worse performance when compared to the cases with only one variable, and also tend to underestimate the position of the probability peak.

We present the maximum crest statistics at a single point as a simplified example to investigate the feasibility of data driven methods. For a single input parameter from the sea state, all the models work well. However, when a more complicated problem is considered, the Random Forest model demonstrates advantages when compared to other methods. Although the Random Forest model can provide an accurate prediction in this particular example, we still consider this model as an early stage prototype instead of the state of the art.

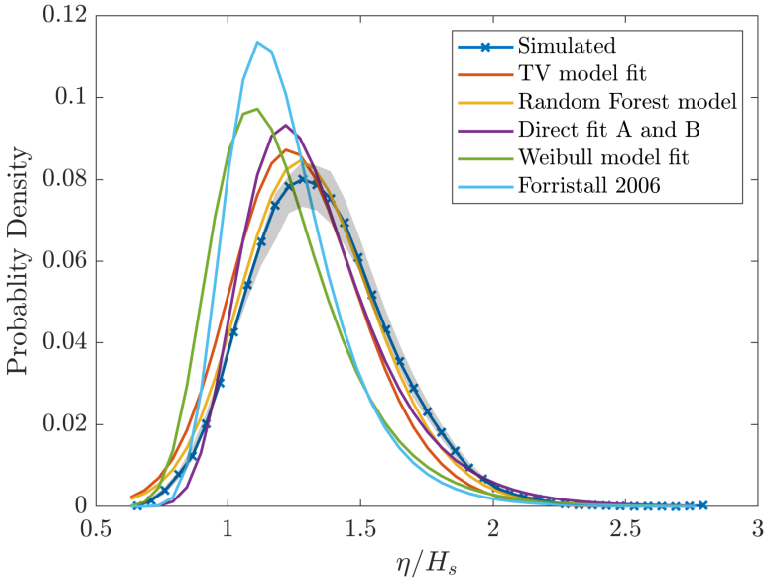

Figure 7.8: Comparison between simulated maximum crest distribution for test condition at water depth of $k_p d$ = 2.5 and wave steepness of $H_s k_p$ = 0.178 against the predictions from different point measurement statistical models, which consider both wave steepness and finite water depth effects. η is the maximum crest elevation from a realisation of simulation for 150 wave periods.

7.5.2 Linear waves over an area

Wave maximum crest distributions over an area differs from that measured at a single point particularly for relatively large areas. In this section, we examine maximum crests distributions of linear waves over an area. We start with the linear waves over a squared area, and the only variable is the side length of the area. We then further investigate the maximum crests height over a rectangular area, where the two side lengths vary simultaneously.

Linear space-time models for a square area

In this section, we will first examine the performance of different models for a square area. We only change the length of the side from 0.5λ to 20λ. The prediction curve of the fitting method is based on a 25 point polynomial fit. Simulations with random side lengths are used for Random Forest training to obtain a proper prediction on

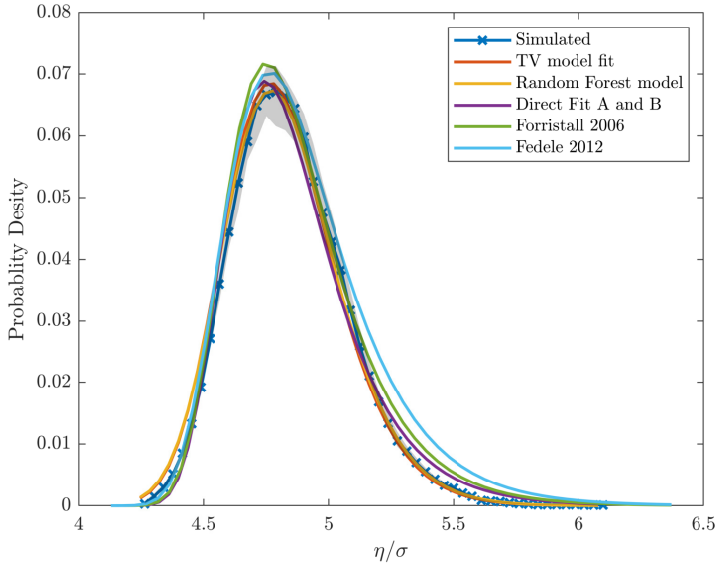

Figure 7.9: Comparison between linear simulated maximum crest distribution over a squared area with side length of 3.5λ against the predictions from different linear space-time statistical models, which consider only the side length effects. η is the maximum crest elevation from a realisation of simulation for 150 wave periods.

coefficients of the Gumbel distribution.

To examine the performance of different models, we have presented a test case with a side length of 3.5λ. We have chosen this test case as it represents the general trends well for other test cases within the test domain. Figure 7.9 shows the predictions of different models when compared to the test case. The shade shows the 95% confidence interval. In general, all the theoretical, fitting and Random Forest models predict the probability density function well. Random Forest model tends to over predict at the lower end. Both Forristall2006 model and Fedele2012 model tend to slightly over predict at the higher end.

Linear space-time models for a rectangular area with side lengths effects

We further examine the performance of different models when the area is rectangular, and both the length in x and the length in y varies. Figure 7.10 shows the percentage error of expected maximum crest height predicted from different models at different

combinations of lengths in x and y directions.

From Figure 7.10, both theoretical models can predict the maximum crest height over a near square sized area accurately. However, when side length in x direction is small compare to the length in y, both models tend to over predict the extreme crest height, particularly for the Forristall2006 model. For small lengths in y direction, both theoretical models tend to under predict the maximum crest height. Theoretical models are expected to perform well for near square areas because of their linear Gaussian framework. Fedele2012 model should perform better for rectangular areas when compared to Forristall2006 model as the latter model is based on empirical fitting through near squared areas simulations. However, both fitting model and Random Forest model tend to provide accurate predictions for areas with extreme aspect ratios. Random Forest model slightly outperforms the fitting model with less random errors in the middle of the test domain.

7.5.3 Second order waves over an area

Waves in open oceans are modified by nonlinear physics. This will make the crests higher than those predicted by linear theory, particularly in steep sea states. In this section, we simulate waves with second order corrections to further examine the accuracy of different models. We first look at rectangular areas with fixed wave steepness in 7.5.3, and we further extend this with varying wave steepness in 7.5.3.

For consistency we match the order of theoretical model with the order of accuracy in the numerical simulations. In this section, as the second order waves being simulated, we applied the second order version for both Forristall2006 and Fedele2012 model.

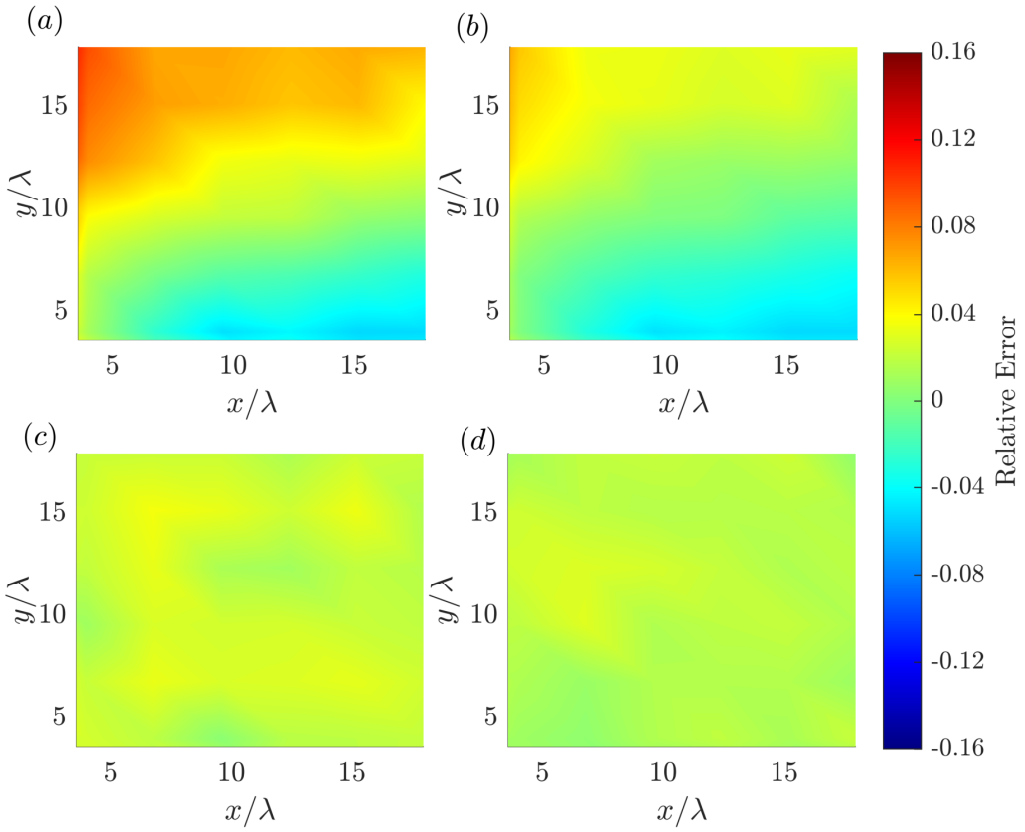

Figure 7.10: Comparison of the relative error in the expected value of maximum linear wave crest value for 150 wave periods from (a): Forristall 2006, (b) Fedele 2012, (c) Fit A and B and (d) Random Forest model at different area sizes for linear simulation.

Second order space-time models for a rectangular area with side lengths effects

We examine the performance of different models for second order waves over a rectangular area. In this subsection, the wave steepness is fixed at $H_s k_p$ = 0.178. The relative error in the expected value of maximum crests height for second order waves at different combination of length in x and length in y is shown in Figure 7.11. When compared to the performance in linear simulations in Figure 7.10, the relative error of two theoretical models is significantly increased in second order simulations. For Forristall2006 model, inaccurate predictions of the extra crest height due to second order harmonics leads to a significant bias, which also leads to large errors (over 14%) at large length in x. Fedele2012 model maintains its performance when the aspect ratio is close to 1. However, the relative error is increased for extreme aspect ratio cases. The increase in relative error is likely due to the magnification effect from the extra crest height in second order wave field. The fitting method and Random Forest method maintains their high performance as these two models can easily include the extra crest height from second order harmonics at a fixed wave steepness.

Second order space-time models for a rectangular area with steepness and side lengths co-effects

We first examine the impact of both area size and the wave steepness on the two coefficients of the Gumbel distribution. In Figure 7.12, we present the scattering of two coefficients A and B, when we change the area size and the wave steepness simultaneously. From Figure 7.12 (c), coefficient A seems to have a strong correlation with the area size, which indicates the area size will affect mostly on the shape of the curve. Coefficient B shows strong correlation with the wave steepness as the wave steepness mainly has an impact on the mean value of maximum crest distribution.

Finally, the performance of all the models will be examined by predicting the

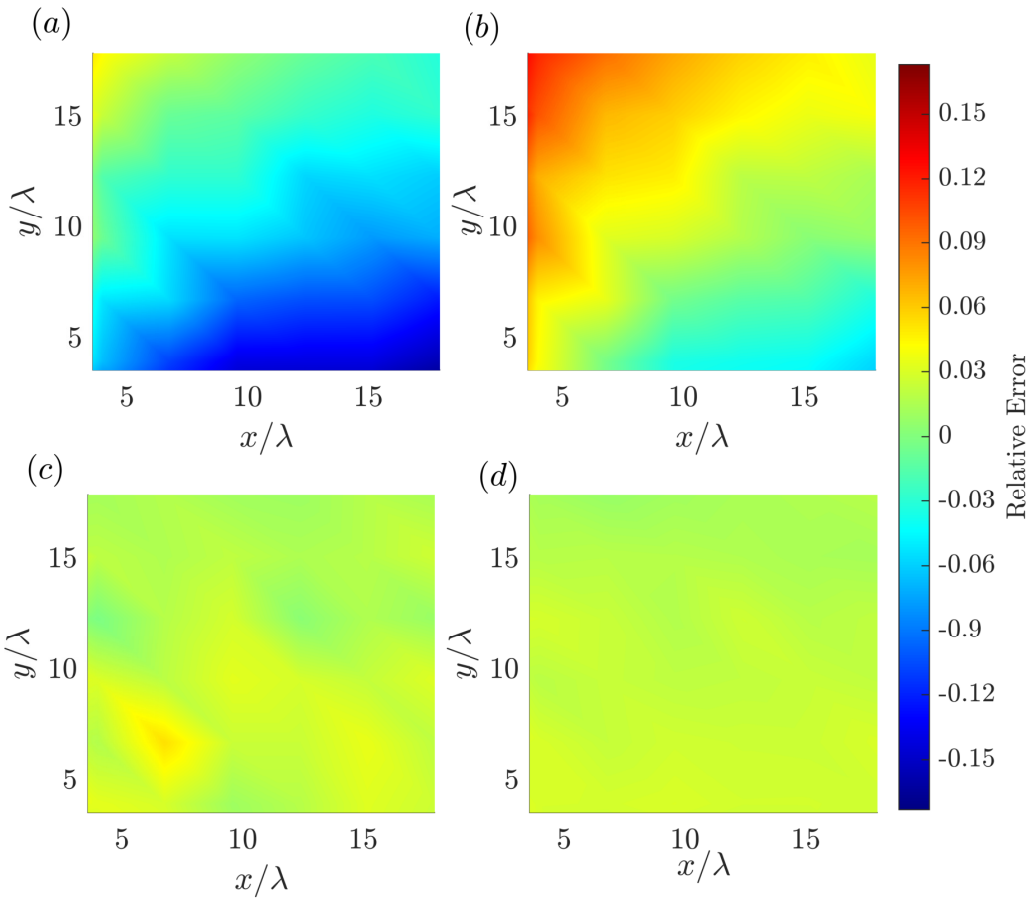

Figure 7.11: Comparison of the relative error in the expected value of maximum crest value for 150 wave periods from (a): Forristall 2006, (b) Fedele 2012, (c) Fit A and B and (d) Random Forest model at different area sizes for second order simulation at fixed wave steepness.

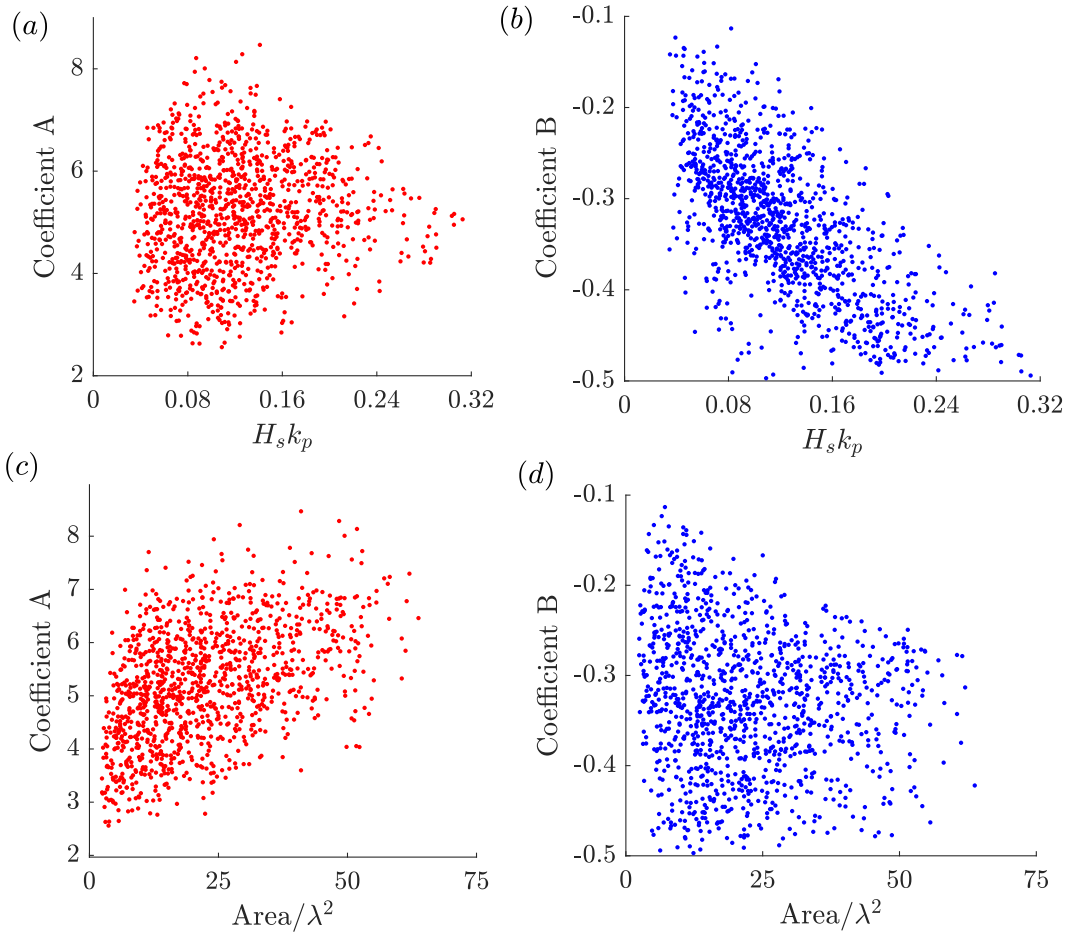

Figure 7.12: Scatter plot of coefficients A and B against (a), (b): wave steepness, (c), (d) area size.

maximum crest height over a rectangular area with varying wave steepness. In Figure 7.13, we present the relative error at different cross planes of the total test domain. For both theoretical models, the horizontal cross section at $H_s k_p = 0.178$ is the same as the error plot shown in Figure 7.11. The results at the lowest wave steepness $H_s k_p = 0.089$ is similar to the error plot shown in Figure 7.10 with linear simulation. This similarity is primarily because the second effect is not significant in low steepness sea states. However, for high steepness sea states, both theoretical models tend to give large errors, especially for Forristall2006 model, which underestimates the maximum crest height. This increasing relative error from low steepness to high steepness for the Forristall2006 model indicates there is a steepness correlated offset, which is most likely due to the underestimation of extra crest height from the second order corrections. However, the Fedele2012 model generally provides a reasonably accurate estimation of second order effects for a near squared area even at high steepness. The increased relative error for high steepness cases is likely due to the amplification effect from the extra crest height in second order effects.

The fitting method tends to overestimate the maximum crest height, especially for areas with a short length in x direction. This overestimation is probably due to the large change rate in both coefficients A and B in that region. The third order polynomial fitting method seems to struggle in this fast-changing region.

The Random Forest model gives the best overall performance when compared to the other three models. The relative error is comparatively small in the middle of the training domain, but the Random Forest model tends to give a large error at the boundary of the training domain, which shows at the lowest steepness with a short length in x. This is primarily because the nature of the Random Forest model as there is generally less training data available at the boundary of the training domain. The same trend applies to the relatively large error at high wave steepness with short lengths in y.

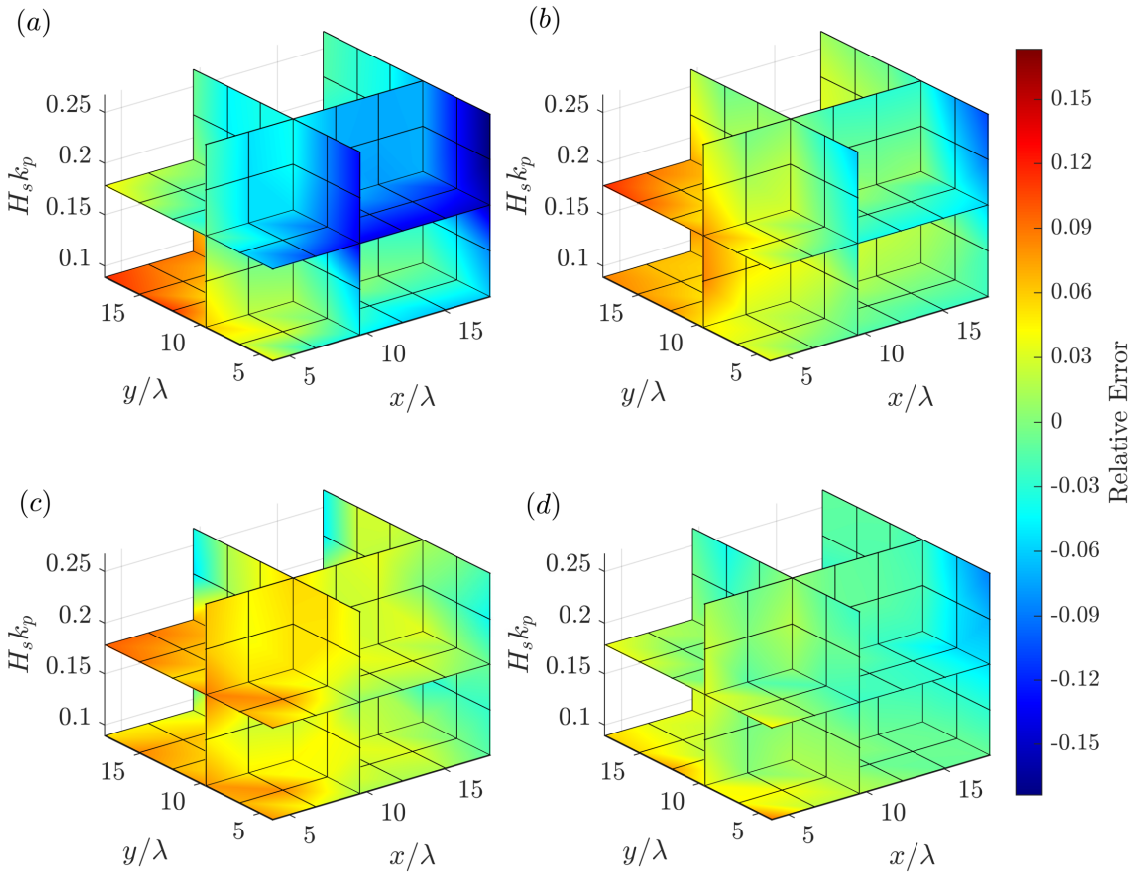

Figure 7.13: Comparison of the relative error in the expected value of maximum crest value for 150 wave periods from (a): Forristall 2006, (b) Fedele 2012, (c) Fit A and B and (d) Random Forest model at different area sizes for second order simulations with different wave steepnesses.

7.5.4 Space-time model performance comparison

We summarise the performance of different models under different wave conditions as the averaged absolute relative error and relative error range. The former parameter provides a measure of the general performance of the model, and the latter parameter shows the worst performance of the model within the test range.

Figure 7.14 shows the averaged absolute relative error for all the models with linear waves, second order waves with fixed wave steepness and second order waves with varying wave steepness. In general, for linear waves, all the models gives less overall errors, and the Random Forest model has the best performance. For second order waves with fixed wave steepness, both fitting methods and the Random Forest model provide relatively accurate predictions on the maximum crest height. When wave steepness is introduced as an additional parameter, the fitting method requires more training data and hence leads to increased relative error. The Random Forest model still has the best performance among all the models tested herein. The performance of fitting and the Random Forest model can also be further improved with better fitting distributions. This study is restricted with the Gumbel distribution as this is the most straight forward distribution. More sophisticated fitting distribution can be developed to obtain better results.

Figure 7.15 shows the maximum error of different models in the test domain. For second order waves with varying wave steepness cases, we examine the error range at the wave steepness of $H_s k_p$ = 0.178. In general, the fitting method and the Random Forest model tend to provide relatively small error ranges for both linear and second order waves, and all models generally produce less error for linear cases. For second order waves, the error ranges of both theoretical models increase significantly. Both fitting method and the Random Forest model tend to slightly over predict the results for the second order waves with fixed wave steepness. When wave steepness is introduced as an additional parameter, the fitting method and Random Forest models

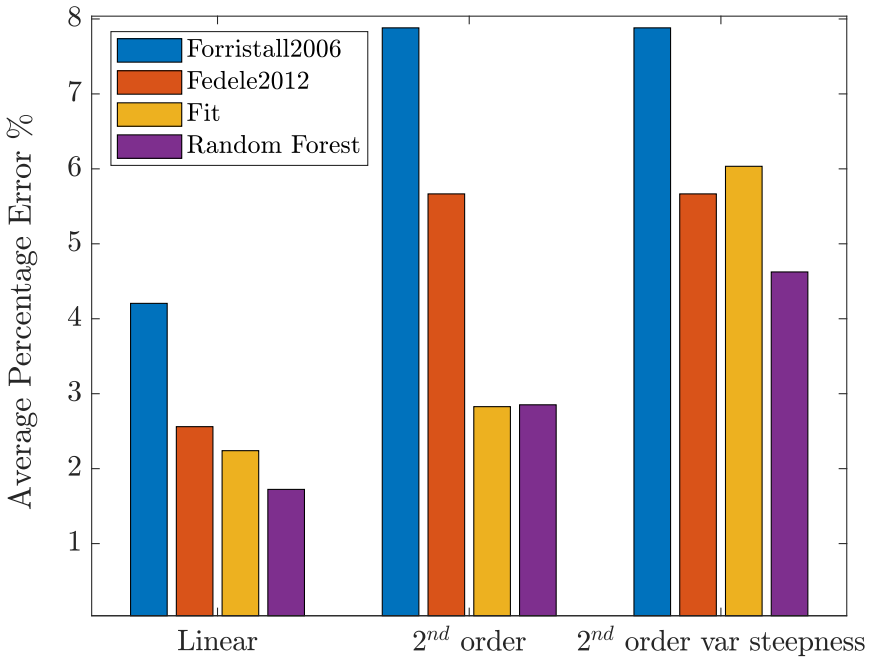

Figure 7.14: Comparison of the absolute average relative error in the expected value of maximum crest value from different models.

tends to have larger error range, and the fitting method continues to over predict the maximum crest height.

We apply the data driven methods in space-time wave statistics as the second example in this paper. For relatively simple cases with few input parameters, such as linear waves over a square area, all the theoretical, fitting and the Random Forest model works well. However, as the situation becomes more complicated, additional input parameter increases the complexity of the prediction models. Prediction errors increase for all the models, but the Random Forest model starts to show its potential in handling complicated problems.

7.5.5 Importance of the parameters

Apart from the providing predictions on the maximum crest distributions, the Random Forest model can also provide an advanced variable importance measure as 'permutation accuracy importance'. This measurement indicates the relative impor-

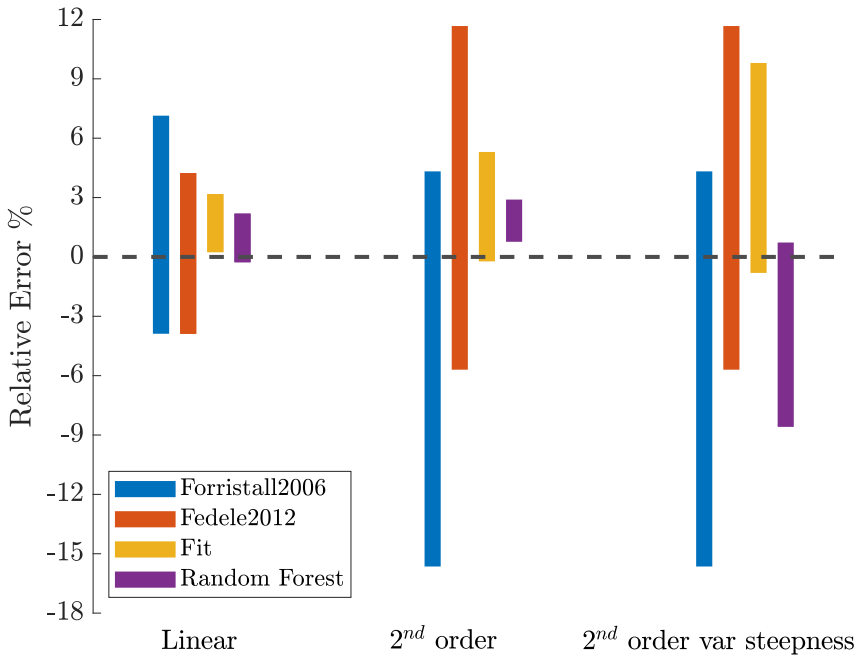

Figure 7.15: Comparison of the maximum error range in the expected value of maximum crest value from different models.

tance of different input parameters based on the prediction trees. This importance estimate parameter can help both in eliminating the irrelevant inputs to increase the training speed and accuracy of the model, and also can provide a brief insight into the nature of maximum crest statistics.

Figure 7.16 shows the relative importance parameter when the test matrix is divided into four subgroups based on the wave steepness and area size. The general trend of the importance for all of the inputs is similar for both high and low steepness models. Wave steepness become more critical with high steepness cases as the second order effect tends to have more impact in a steep sea state. This double confirmation further demonstrates that these importance estimates from the Random Forest model can provide extra insights into the nature of the maximum crest distributions.

The relative importance of different inputs changes dramatically when the area size changes. For small areas, the scale dimension coefficient β seems to be much more important when comparing to large area cases. This is primarily due to the

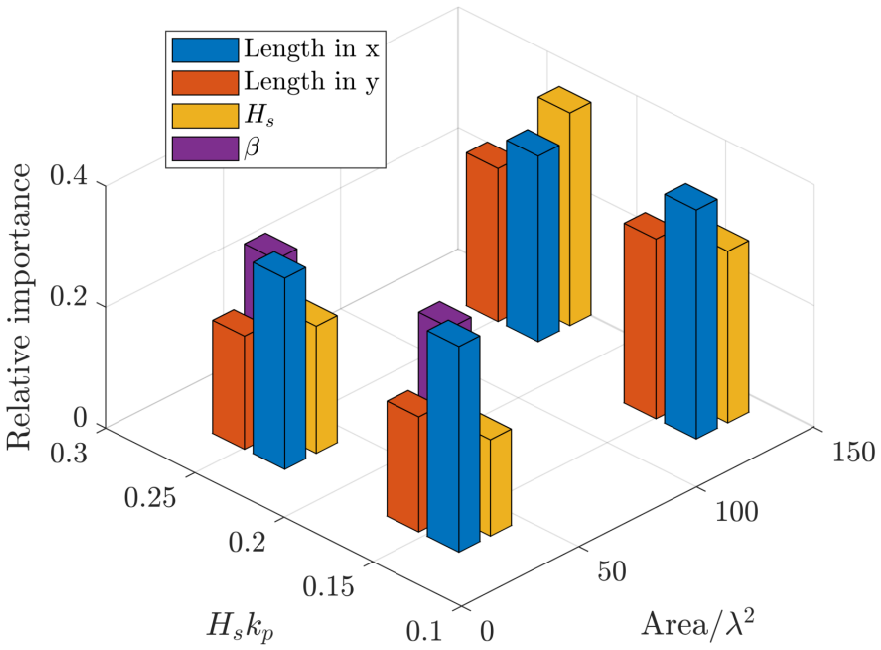

Figure 7.16: Relative importance estimated from the Random Forest model based on the input range of area size and wave steepness for waves with second order correction.

scale dimension coefficient β which quantifies the statistical change when a 3D wave reduces to the 2D or even 1D wave at small areas. The importance of wave steepness increases for large areas, which could be because for relatively large areas, increase area size is not as effective as increase wave steepness. The importance of length in y direction is also increased for large areas, which could be because length in the y direction is more effective for 3D waves. When an area is small, the difference in the peak wavelength in x and y direction will lead to differences in the number of waves in 2D waves, which agrees well with Fedele's theory (Fedele 2012). However, for relatively large areas, the effect from the difference in the peak wavelength tends to be less significant, which agrees well with the predictions from Random Forest model.

7.6 Discussions and conclusions

This paper reviewed existing theoretical models for maximum crest distributions over an area and also proposed two data driven models to estimate the crest distributions based on the sea state parameters without any simulations required during the prediction phase. We use second order numerical simulations to study the accuracy of both the theoretical models and data driven models under various conditions.

In this study, we present two examples of applying data driven methods to predict wave statistics. In the first example, we explore the second order maximum crest distributions at a single point, where the Random Forest model shows its advantages in accuracy for the most complicated scenario. In the second example, we extend our research into the maximum crest distributions over an area. Of two theoretical models, the Fedele2012 model (Fedele 2012) tends to provide more accurate predictions, particularly for second order waves. The polynomial fitting method works equally well with the Random Forest model for the linear cases. The error in the empirical fitting method increased significantly for the second order case, which primarily because of the fitting difficulty in a complex problem with many input parameters. However, the Random Forest model continues to perform well in the most complicated case and thus shows significant potential for modelling complex situations.

The two examples studied in this paper are intended as practicability tests of applying data driven method to predict maximum crest distributions under various conditions. The complexity of the Random Forest model presented herein is restricted by the number of inputs, which ultimately depends on the time constraint and accessible computational power. The underlying physics for these two examples examined herein are relatively straightforward with proper assumptions. Hence, two theoretical models can provide relatively accurate predictions particularly for the areas close to a square,

However, real water wave statistics in the open ocean are modified by week and

strong nonlinear physics. For complicated situations such as wave-current interactions, wind wave interactions, wave shoaling effects, non-equilibrium spectrum or wave breaking effects, it is very difficult to derive satisfactory analytical solutions to account for all these phenomena. This makes design difficult.

For relatively complicated problems, the traditional polynomial fitting methods requires large amount of fitting data yet with large errors for multiple inputs cases. However, the Random Forest model requires less training data and provides more accurate predictions even for two simplified situations examined herein. For engineering applications, the Random Forest model can be used just as the traditional fitting equations. Given the input from the sea state parameters (i.e. steepness, sampling area dimensions), the Random Forest model can make predictions without requiring much computational power once it is properly trained and validated. However, cares need to be taken when applying these data driven methods directly as random error peak could occur especially at the boundary of the training domain.

Based on the performance of two examples examined, data driven approaches look promising in terms of predicting space-time wave statistics. This type of statistical model has substantial potential in predicting maximum crest height distributions under intricate situations, which can also be used for engineering purposes with proper validation.

Chapter 8

Discussions and Conclusions

Gaining more knowledge of rogue wave events is important for the safety of offshore structures and ships. Especially, engineers are interested in the probability of extreme wave events and the most probable shape of these events in the open ocean as they are crucial to the design of offshore structures. Physicists and mathematicians are motivated to understand formation mechanisms of rogue waves, which offer feasible paths towards more accurate models for the prediction of rogue waves. In this study, we investigated three main aspects of extreme waves, which would potentially contribute to conquer challenges in the design of offshore structures as well as to provide insights of the formation mechanisms of rogue waves. The three aspects are explained as follows.

(i) the averaged shape of rogue wave events,

(ii) spatial evolution of the rogue wave density of steep unidirectional random waves,

(iii) space-time wave statistics model for maximum crest distribution.

The main achievements in this study on different aspects of rogue wave events are summarised separately in the following sections in which physical interpretations are also presented.

8.1 Averaged shape of rogue wave events

In chapter 3, we explored the averaged shape of extreme events in naturally occurring water waves with two field datasets from Lake George and the North sea. The former dataset provides measurements of relative steep wind-driven water waves and the latter dataset represents the waves in open ocean. We examined the horizontal asymmetry of the wave group and the contraction of the wave envelope with both datasets. We find that the wave in front of an extreme wave tends to be smaller than the wave following it. The wave group tends to contract in the mean wave direction relative to the linear theory. Both horizontal asymmetry and wave group contraction are strongly correlated with the wave steepness, which suggests such nonlinear modifications of extreme events can be observed in the open ocean.

In chapter 5, we further investigated the nonlinear properties of unidirectional surface waves including horizontal asymmetry and wave group contraction using numerical simulations for waves at deep water. We in particular examined the spatial evolution of these nonlinear properties for random waves . It is found that the horizontal asymmetry is significant during the initial evolution but is greatly reduced after the spectrum reached equilibrium. The level of asymmetry during the transient state and the residual level of asymmetry at steady state seems to be correlated with wave steepness, which indicates these changes originate from the nonlinear interactions. Additionally, the spatial locations of maximum horizontal asymmetry seems to be coincident with the maximum point in kurtosis.

This coincidence, as pointed out by one of the reviewers of the paper, could be inherently linked with the initial evolution of kurtosis. Unfortunately, we have very limited number of theories to further verify this suggestion. The state-of-art theoretical models for the averaged shape of extreme waves such as NewWave theory or the quasi-determinism theory are based on the assumption of linear Gaussian random seas and the higher order bound harmonics based on Stokes theory. Some

nonlinear physics would lead to this phenomenon such as the nonlinear dispersion relation will cause the largest crest move to the front of the wave group, yet no theory provides the full details of the reason why the horizontal asymmetry is greatly reduced at the steady state.

The averaged shape of largest events also tends to contract, relative to the averaged shape predicted using a linear theory, in the mean wave direction during the wave evolution. This suggests that nonlinear physics also modifies the envelope width of extreme events. We interpret this contraction of envelope as a local expansion of the spectrum of the extreme events on a much faster scale as opposed to the increase in bandwidth of the sea state. Additionally, due to the local modification from nonlinear physics being significant only for relatively steep waves, the theoretical predictions based on the sea state can not predict these localised changes.

This contraction of the envelope from nonlinear modifications increases monotonically to a steady state value. The evolution of contraction of the envelope seems to be associated with the asymmetry as both reach their steady state roughly at the same location. Additionally, both modifications happen on a faster time scale and are only apparent locally for the extreme events. However, work remains to connect all these findings.

In chapter 6, we further explored the averaged shape of extreme waves on finite water depth. In general, the finite water depth tends to slightly amplify the nonlinear effects of the averaged shape of the largest events. However, for extremely large events, this amplification is less significant.

Although significant departure from linear physics is found with unidirectional numerical simulations, the available theories can predict the averaged shape in open ocean relatively accurately (see results in Santo et al. (2013)), even in relative shallow water (Taylor & Williams 2004). This is partially due to the horizontal asymmetry being mostly a transient behaviour, but most of the waves in open ocean are generated

with wind through long distances and are thus in seas which are close to equilibrium. Additionally, real water waves are directionally spread, which leads to fundamental difference in nonlinear physics. Wave in open ocean is also less steep when compared to our numerical simulations, which can reduce the nonlinear modifications significantly. This also agrees well with the observations in chapter 3, where we find significant nonlinear modifications for the waves in Lake George with steep sea state but these modifications are greatly reduced in the North Sea.

8.2 Spatial evolution of kurtosis

In chapter 5, we study the evolution of kurtosis of unidirectional random waves on deep water based on the measurements obtained from experiments. As kurtosis is a convenient measure of rogue wave density, we aim to explore the physical mechanisms that drive the wave statistics departing from a linear Gaussian event. We also extend the parameter with two numerical simulation methods, which agree well with the experiments. For a low steepness, the evolution of kurtosis is well predicted by the existing analytical models based on Nonlinear Schrödinger equation. However, for cases where waves are steeper, the analytical models tend to overpredict the steady state value of the kurtosis. Additionally, in those cases, the kurtosis in experiments and numerical simulations reaches a maximum value before reaching the steady value, which is not captured by the analytical models.

Different nonlinear physics could potentially lead to the overshoot of kurtosis, which is not being captured by the theory based on Nonlinear Schrödinger equation. Rapid evolution of wave spectrum of the sea state is found at the initial stage of wave evolution. This is associated with the initial departure of phases from their equilibrium distribution. We superimpose the forced input as initial conditions with a random phase assigned to each Fourier component. However, certain phase corre-

lations can be found for random waves in an equilibrium state. The development of correlations between initially uncorrelated components seems to be connected with the rapid evolution of wave spectrum. This spectral evolution could lead to the theories ignoring the peaking behaviour of kurtosis as the theories based on Nonlinear Schrödinger equation or compact Zakharov equation assume that the action density varies slowly. Additionally, these theories based on slow varying action density assumptions cannot capture the local spectral expansion associated with the envelope contraction of extreme events. This local spectral expansion may also invalidate the narrow bandwidth assumptions of the theory based on Nonlinear Schrödinger equation.

Additionally, the analytical models overpredict the steady state value of the normalised kurtosis for steep sea states. This overprodiction is primarily due to the evolution of the Benjamin-Feir index (BFI). Analytical theories based on slow varying spectral density assumptions expect the BFI value to be a constant value spatially. However, from both experimental results and numerical simulations, the value of BFI at steady state can even be half of the initial BFI value. This significant reduction in BFI value is primarily due to the broadening of the spectrum during the transient stage. As one of the reviewers suggested, normalising the kurtosis by the BFI value at steady state will significantly reduce the overestimation. However, the value of BFI at steady state is unknown prior to the experiments or numerical simulations. Additionally, the steady state kurtosis value depends more on the initial bandwidth of the spectrum above a certain threshold of steepness. This behaviour explains the large discrepancy between analytical models and numerical simulations for steep sea states.

In chapter 6, we further study the evolution of kurtosis at finite water depth. For high steepness cases in shallow water, the kurtosis at the steady state is underestimated by analytical models, which also give zero value of kurtosis at critical water

depth. However, we find a small but non-zero value of kurtosis in the numerical simulations. This suggests that nonlinearity beyond the cubic nonlinearity contributes to the kurtosis in finite water depth.

8.3 Space-time wave statistics

In chapter 7, we study the space-time maximum crest statistics with linear and second order simulations. We have applied two data driven methods to establish predictive models based on the numerical simulation results. We also examine the performance of two theoretical models with second order numerical simulation results. Although the Forristall2006 model (Forristall 2006) can predict the crest distributions of linear waves over areas that are close to a square, the large error in predicting the second order component leads to a notable bias for second order statistics. Fedele2012 model (Fedele 2012) can provide accurate predictions on second order waves over areas close to squares but will give large errors for areas with large aspect ratios.

The empirical fitting method can give precise predictions with linear inputs, but unsurprisingly, errors increase significantly for the second order cases. This sharp increase in error is due to the dramatic rise in fitting difficulty for a complex problem with many input parameters. Random forest model tends to provide the most accurate predictions for both linear and second order predictions.

The importance estimates from the random forest model provide a brief insight into wave physics. The scale dimension parameter has great importance when the area is relatively small which agrees well with Fedele's study in Fedele (2012). Additionally, for small areas, the length in y is less important than the length in x as the difference in characteristic wavelength in both directions will lead to a different number of waves. However, for large areas, the importance of length in y increases significantly as the number of waves depends further on the total area size for large areas. Wave steepness

seems to be more critical at large areas, which indicates that the extra amplitude from second order harmonics has more impact on the crest distributions for large areas.

Overall, the random forest model shows excellent performance in predicting wave statistics over an area in this simplified condition. With additional input parameters and more nonlinear physics introduced, data driven methods have huge potential in engineering applications.

8.4 Recommendations for future work

8.4.1 Space-time wave measurements

Nonlinear interactions are fundamentally different in directionally spread seas, compared to unidirectional cases. Space-time wave measurements would help to gain a better understanding of the mechanisms leading to rogue waves in the open ocean. In this study, we explore the space-time wave statistics and horizontal asymmetry and contraction of wave groups numerically. Numerical simulations suggest the wave group also tends to expand in the lateral direction (Adcock et al. 2015), which requires the spatial profile of wave group. Additionally, space-time wave statistics in the open ocean, frequency and wavenumber distributions are all of great importance to offshore design. Further exploration of space-time wave measurements may lead to new insights on nonlinear wave behaviour in the open ocean.

Space-time evolution of wave fields can be simulated straightforwardly with numerical tools. However, very few wave observation techniques could provide space-time information of real occurring waves during experiments and field measurement campaign. Instrumentation, like radars, laser scanners, and stereo cameras can provide spatial information of sea surface elevations, which could be applied in the future (see some prior work Bergamasco et al. (2017), Alvise et al. (2017)).

8.4.2 Wave breaking effects

Extreme events are often associated with wave breaking, which dissipates energy and limits the maximum crest height. Wave breaking, as a 'strong' nonlinear effect, modifies both the shape of extreme events and the rogue wave density (Babanin et al. 2001, 2007, Banner et al. 2014), which is particularly important in spread sea states (McAllister et al. 2018).

Unfortunately, many challenges exist in order to examine the wave breaking effects with traditional methods. A conventional potential flow solver struggles at solving the overturning of the fluid domain and it also fails to capture the dissipation of energy as it is obviously beyond its regime of applicability. Recent work shows that Higher Order Spectral method can predict the breaking position, but the simulation become unstable after the wave breaking event (Seiffert & Ducrozet 2018). The built-in chaos and random nature of wave breaking also hinders attempts to use experiments as the repeatability is in question (Barthelemy et al. 2018, Babanin et al. 2007). The identification of individual breaking wave is demanding for practical wave observation techniques.

Examining the breaking occurring in wave groups would be an excellent starting point. Numerical simulations with wave group are less computationally demanding and would allow methods such as CFD to be deployed (Deike et al. 2015). Experiments with waves groups rule out the inherent randomness of wave fields and could yield repeatable results.

Machine Learning methods are also suitable for modelling wave breaking effects. Kouvaras & Dhanak (2018) applied Machine Learning techniques to classify the types of breaking waves. Physics based Machine Learning models would be perfect candidates to predict wave breaking effects within the conventional potential flow framework. Similar approaches have been widely used in other fluid mechanics problems (Wan & Sapsis 2018, Wan et al. 2018).

8.4.3 Data driven methods in wave statistics

In finite depth, some additional nonlinear effects have to be considered (Kakutani & Michihiro 1983, Sergeeva et al. 2014, Janssen & Janssen 2019, Fedele 2015*b*, Shemer et al. 1998). Changes in water depth, especially from deep water into intermediate depth can significantly alter the local wave statistics. Wave current interactions can also lead to non-Gaussian wave statistics (Janssen & Herbers 2009), which are particularly important in coastal zones. However, partially due to experimental difficulties, this area has received relatively little attention particularly for wave crest distributions.

In this study, data driven methods have shown great potential in predicting space-time wave statistics at finite water depth, which could be extended to coastal zones with seafloor topography and background current.

We have only explored Random Forest model as the only machine learning model in this study. Alternative data science techniques, such as gradient boost trees or Gaussian processes can be applied for comparison.

We have used Monte Carlo simulation in this study to obtain the wave statistical distributions. This requires a significant number of simulations with a large number of computational resources, and hence reduces the complexity of problem we studied. Recently, sequential sampling strategy with machine learning as introduced by Gramstad et al. (2020), Mohamad & Sapsis (2018), Sapsis (2020*a*), could potentially reduce the number of samples required for complex extreme value problems.

References

Adcock, T. A. A., Feng, X., Tang, T., Day, A. H., Dai, S., Li, Y., Lin, Z., Xu, W. & Taylor, P. H. (2019), 'Application of phase decomposition to the analysis of random time series from wave basin tests', *38th Int. Ocean Offshore Arct. Eng. Conf., Glasgow UK* .

Adcock, T. A. A., Gibbs, R. H. & Taylor, P. H. (2012), 'The nonlinear evolution and approximate scaling of directionally spread wave groups on deep water', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **468**(2145), 2704–2721.

Adcock, T. A. A. & Taylor, P. H. (2009*a*), 'Estimating ocean wave directional spreading from an Eulerian surface elevation time history', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **465**(2111), 3361–3381.

Adcock, T. A. A. & Taylor, P. H. (2009*b*), 'Focusing of unidirectional wave groups on deep water: an approximate nonlinear Schrödinger equation-based model', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **465**(2110), 3083 LP – 3102.

Adcock, T. A. A. & Taylor, P. H. (2014), 'The physics of anomalous ('rogue') ocean waves', *Reports Prog. Phys.* **77**(10), 105901.

Adcock, T. A. A. & Taylor, P. H. (2016*a*), 'Fast and local non-linear evolution of steep wave-groups on deep water: A comparison of approximate models to fully non-linear simulations', *Phys. Fluids* **28**(1), 16601.

Adcock, T. A. A. & Taylor, P. H. (2016*b*), 'Non-linear evolution of large waves in deep water – The influence of directional spreading and spectral bandwidth', *26th Int. Ocean Polar Eng. Conf.* .

Adcock, T. A. A. & Taylor, P. H. (2016*c*), 'Non-linear evolution of uni-directional focussed wave-groups on a deep water: A comparison of models', *Appl. Ocean Res.* **59**, 147–152.

Adcock, T. A. A., Taylor, P. H. & Draper, S. (2015), 'Nonlinear dynamics of wave-groups in random seas: unexpected walls of water in the open ocean', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **471**(2184).

Adcock, T. A. A., Taylor, P. H. & Draper, S. (2016), 'On the shape of large wave-groups on deep water—The influence of bandwidth and spreading', *Phys. Fluids* **28**(10), 106601.

Adcock, T. A. A. & Yan, S. (2010), 'The focusing of uni-directional Gaussian wave-groups in finite depth: an approximate NLSE based approach', *29th Int. Ocean Offshore Arct. Eng. Conf., Shanghai, China* **49125**, 569–576.

Adler, J. E. & Taylor, R. J. (2007), *Random Fields and Geometry*, New York: Springer-Verlag New York.

Adler, R. J. (1981), *The Geometry of Random Fields*, Vol. 62, SIAM.

Agnon, Y., Babanin, A. V., Young, I. R. & Chalikov, D. (2005), 'Fine scale inhomogeneity of wind-wave energy input, skewness, and asymmetry', *Geophys. Res. Lett.* **32**(12).

Ali, M. & Prasad, R. (2019), 'Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition', *Renew. Sust. Energ. Rev.* **104**, 281–295.

Alvise, B., Francesco, B., Filippo, B., Sandro, C. & Mauro, S. (2017), 'Space-time extreme wind waves: Analysis and prediction of shape and height', *Ocean Model.* **113**, 201–216.

Annenkov, S. Y. & Shrira, V. I. (2006), 'Role of non-resonant interactions in the evolution of nonlinear random water wave fields.', *J. Fluid Mech.* **561**, 181–207.

Annenkov, S. Y. & Shrira, V. I. (2009), 'Evolution of kurtosis for wind waves', *Geophys. Res. Lett.* **36**(13).

Annenkov, S. Y. & Shrira, V. I. (2018), 'Spectral evolution of weakly nonlinear random waves: kinetic description versus direct numerical simulations', *J. Fluid Mech.* **844**, 766–795.

Babanin, A. V., Chalikov, D., Young, I. R. & Savelyev, I. (2007), 'Predicting the breaking onset of surface water waves', *Geophys. Res. Lett.* **34**(7).

Babanin, A. V. & Makin, V. K. (2008), 'Effects of wind trend and gustiness on the sea drag: Lake George study', *J. Geophys. Res.* **113**(2), C02015.

Babanin, A. V., Young, I. R. & Banner, M. L. (2001), 'Breaking probabilities for dominant surface waves on water of finite constant depth', *J. Geophys. Res.* **106**(C6), 11659.

Baldock, T. E., Swan, C. & Taylor, P. H. (1996), 'A laboratory study of nonlinear surface waves on water', *Philos. Trans. Royal Soc. A Math. Phys. Eng. Sci.* **354**(1707), 649–676.

Banner, M. L., Barthelemy, X., Fedele, F., Allis, M., Benetazzo, A., Dias, F. & Peirson, W. L. (2014), 'Linking reduced breaking crest speeds to unsteady nonlinear water wave group behavior', *Phys. Rev. Lett.* **112**, 114502.

Barratt, D., Bingham, H. B. & Adcock, T. A. A. (2020), 'Nonlinear evolution of a steep, focusing wave group in deep water simulated with OceanWave3D', *J. Offshore Mech. Arct. Eng.* **142**(2).

Barratt, D., Bingham, H. B., Taylor, P. H., Van Den Bremer, T. S. & Adcock, T. A. A. (2021), 'Rapid spectral evolution of steep surface wave groups with directional spreading', *J. Fluid Mech.* **907**.

Barratt, D., Bingham, H. B., van den Bremer, T. S. & Adcock, T. A. A. (2020), 'Lineariation of the wave spectrum: A comparison of methods', *39th Int. Ocean Offshore Arct. Eng. Conf., Florida, USA* .

Barthelemy, X., Banner, M. L., Peirson, W. L., Fedele, F., Allis, M. & Dias, F. (2018), 'On a unified breaking onset threshold for gravity waves in deep and intermediate depth water', *J. Fluid Mech.* **841**, 463–488.

Bateman, W. J. D., Swan, C. & Taylor, P. H. (2001), 'On the Efficient Numerical Simulation of Directionally Spread Surface Water Waves', *J. Comput. Phys.* **174**(1), 277–305.

Bell, R. J., Gray, S. L. & P. Jones O (2017), 'North Atlantic storm driving of extreme wave heights in the North Sea', *J. Geophys. Res. Ocean.* **122**(4), 3253–3268.

Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S. & Sclavo, M. (2015), 'Observation of extreme sea waves in a space-time ensemble', *J. Phys. Oceanogr.* **45**(9), 2261–2275.

Benetazzo, A., Fedele, F., Gallego, G., Shih, P.-C. & Yezzi, A. (2012), 'Offshore stereo measurements of gravity waves', *Coast. Eng.* **64**, 127–138.

Benjamin, T. B. & Feir, J. E. (1967), 'The disintegration of wave trains on deep water Part 1. Theory', *J. Fluid Mech.* **27**(3), 417–430.

Benney, D. J. & Newell, A. C. (1967), 'The propagation of nonlinear wave envelopes', *J. Math. Phys.* **46**(1-4), 133–139.

Bergamasco, F., Torsello, A., Sclavo, M., Barbariol, F. & Benetazzo, A. (2017), 'Wass: An open-source pipeline for 3d stereo reconstruction of ocean waves', *Comput. & Geosci.* **107**, 28–36.

Bergstra, J. S., Bardenet, R., Bengio, Y. & Kégl, B. (2011), 'Algorithms for hyper-parameter optimization', *Adv. Neural Inf. Process Syst.* pp. 2546–2554.

Boccotti, P. (1983), 'Some new results on statistical properties of wind waves', *Appl. Ocean Res.* **5**(3), 134–140.

Boccotti, P. (1989), 'On mechanics of irregular gravity waves', *Atti Accademia Nazionale dei Lincei, Memorie viii* pp. 111–170.

Boccotti, P. (2000), *Wave mechanics for ocean engineering*, Elsevier.

Bouws, E., Ephraums, J., Ewing, J., Francis, P., Günther, H., Janssen, P., Komen, G., Rosenthal, W. & de Voogt, W. (1985), Shallow water intercomparison of wave models—Part I Three different concepts to model surface waves in finite water depth, *in* 'The Ocean Surface', Springer, pp. 201–205.

Breiman, L. (2001), 'Random forests', *Mach. Learn* **45**(1), 5–32.

Brunton, S. L., Noack, B. R. & Koumoutsakos, P. (2020), 'Machine learning for fluid mechanics', *Annu. Rev. Fluid Mech.* **52**, 477–508.

Callens, A., Morichon, D., Abadie, S., Delpey, M. & Liquet, B. (2020), 'Using random forest and gradient boosting trees to improve wave forecast at a specific location', *Appl. Ocean Res.* **104**, 102339.

Cartwright, D. E. & Longuet-Higgins, M. S. (1956), 'The statistical distribution of the maxima of a random function', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **237**(1209), 212–232.

Chabchoub, A. & Grimshaw, R. (2016), 'The hydrodynamic nonlinear Schrödinger equation: Space and time', *Fluids* **1**(3), 23.

Cherneva, Z., Tayfun, M. & Guedes Soares, C. (2009), 'Statistics of nonlinear waves generated in an offshore wave basin', *J. Geophys. Res. Oceans* **114**(C8).

Christou, M. & Ewans, K. (2014), 'Field measurements of rogue water waves', *J. Phys. Oceanogr.* **44**(9), 2317–2335.

Cousins, W., Onorato, M., Chabchoub, A. & Sapsis, T. P. (2019), 'Predicting ocean rogue waves from point measurements: An experimental study for unidirectional waves', *Phys. Rev. E* **99**(3), 032201.

Cousins, W. & Sapsis, T. P. (2014), 'Quantification and prediction of extreme events in a one-dimensional nonlinear dispersive wave model', *Physica D* **280**, 48–58.

Dalzell, J. F. (1999), 'A note on finite depth second-order wave-wave interactions', *Appl. Ocean Res.* **21**(3), 105–111.

Dean, R. G. & Sharma, J. N. (1981), 'Simulation of wave systems due to nonlinear directional spectra', *Proc. Int. Symp. Hydrodyn. Oc. Engrg.* pp. 1211–1222.

Debsarma, S. & Das, K. P. (2005), 'A higher-order nonlinear evolution equation for broader bandwidth gravity waves in deep water', *Phys. Fluids* **17**(10), 104101.

Deike, L., Popinet, S. & Melville, W. K. (2015), 'Capillary effects on wave breaking', *J. Fluid Mech.* **769**, 541–569.

Dematteis, G., Grafke, T., Onorato, M. & Vanden-Eijnden, E. (2019), 'Experimental evidence of hydrodynamic instantons: the universal route to rogue waves', *Phys. Rev. X* **9**(4), 041057.

Dripta, M. J. & Dutykh, D. (2020), 'Learning extreme wave run-up conditions', *Appl. Ocean Res.* **105**, 102400.

Ducrozet, G., Bingham, H. B., Engsig-Karup, A. P., Bonnefoy, F. & Ferrant, P. (2011), 'A comparative study of two fast nonlinear free-surface water wave models', *Int. J. Numer. Methods Fluids* **69**(11).

Dudley, J. M., Genty, G., Mussot, A., Chabchoub, A. & Dias, F. (2019), 'Rogue waves and analogies in optics and oceanography', *Nat. Rev. Phys.* **1**(11), 675–689.

Dyachenko, A. I. & Zakharov, V. E. (2011), 'Compact equation for gravity waves on deep water', *JETP Lett.* **93**(12), 701.

Dysthe, K. B. (1979), 'Note on a modification to the nonlinear Schrödinger equation for application to deep water waves', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **369**(1736), 105–114.

Dysthe, K. B., Krogstad, H. E. & Müller, P. (2008), 'Oceanic rogue waves', *Annu. Rev. Fluid Mech.* **40**(1), 287–310.

Dysthe, K. B. & Trulsen, K. (2001), The evolution of an evolution equation, *in* 'Progress in nonlinear science, dedicated to the 100th anniversary of A.A. Andronov, Nizhny Novgorod'.

Efron, B. & Tibshirani, R. J. (1994), *An introduction to the bootstrap*, CRC press.

Engsig-Karup, A. P., Bingham, H. B. & Lindberg, O. (2009), 'An efficient flexible-order model for 3D nonlinear water waves', *J. Comput. Phys.* **228**(6), 2100–2118.

Ewans, K., Feld, G. & Jonathan, P. (2014), 'On wave radar measurement', *Ocean Dyn.* **64**, 1281–1303.

Ewans, K., Jonathan, P. & Feld, G. (2013), 'What does a wave radar actually measure', *13th Int. Work. Wave Hindcasting 4th Coast. Hazards Symp. Banff, Alberta, Canada* .

Farazmand, M. & Sapsis, T. P. (2017), 'Reduced-order prediction of rogue waves in two-dimensional deep-water waves', *J. Comput. Phys.* **340**, 418–434.

Fedele, F. (2012), 'Space–time extremes in short-crested storm seas', *J. Phys. Oceanogr.* **42**(9), 1601–1615.

Fedele, F. (2014), 'On certain properties of the compact Zakharov equation', *J. Fluid Mech.* **748**, 692–711.

Fedele, F. (2015*a*), 'On oceanic rogue waves', *arXiv preprint arXiv:1501.03370* .

Fedele, F. (2015*b*), 'On the kurtosis of deep-water gravity waves', *J. Fluid Mech.* **782**, 25–36.

Fedele, F., Benetazzo, A., Gallego, G., Shih, P. C., Yezzi, A., Barbariol, F. & Ardhuin, F. (2013), 'Space–time measurements of oceanic sea states.', *Ocean Model.* **70**, 103–115.

Fedele, F., Brennan, J., Ponce de León, S., Dudley, J. & Dias, F. (2016), 'Real world ocean rogue waves explained without the modulational instability', *Sci. Rep.* **6**, 27715.

Fedele, F., Cherneva, Z., Tayfun, M. A. & Guedes Soares, C. (2010), 'Nonlinear Schrödinger invariants and wave statistics', *Phys. Fluids* **22**(3), 036601.

Fedele, F. & Dutykh, D. (2012), 'Special solutions to a compact equation for deep-water gravity waves', *J. Fluid Mech.* **712**, 646–660.

Fedele, F., Gallego, G., Yezzi, A., Benetazzo, A., Cavaleri, L., Sclavo, M. & Bastianini, M. (2012), 'Euler characteristics of oceanic sea states', *Math Comput. Simul.* **82**(6), 1102–1111.

Fedele, F., Lugni, C. & Chawla, A. (2017), 'The sinking of the el faro: predicting real world rogue waves during hurricane joaquin', *Sci. Rep.* **7**(1), 1–15.

Fedele, F. & Tayfun, M. A. (2009), 'On nonlinear wave groups and crest statistics', *J. Fluid Mech.* **620**, 221.

Fernandez, L., Onorato, M., Monbaliu, J. & Toffoli, A. (2014), 'Modulational instability and wave amplification in finite water depth', *Nat. Hazards Earth Syst. Sci.* **14**(3), 705–711.

Fernandez, L., Onorato, M., Monbaliu, J. & Toffoli, A. (2016), Occurrence of extreme waves in finite water depth, *in* 'Extreme ocean waves', Springer, pp. 45–62.

Fitzgerald, C. J., Taylor, P. H., Eatock Taylor, R., Grice, J. & Zang, J. (2014), 'Phase manipulation and the harmonic components of ringing forces on a surface-piercing column', *Proc. R. Soc. A Math. Phys. Eng. Sci.* .

Forristall, G. Z. (2000), 'Wave crest distributions: Observations and second-order theory', *J. Phys. Oceanogr.* **30**(8), 1931–1943.

Forristall, G. Z. (2006), 'Maximum crest heights over an area and the air gap problem', *25th Int. Ocean Offshore Arct. Eng. Conf., Hamburg, Germany* **47489**, 11–15.

Forristall, G. Z. (2007), 'Comparing hindcasts with wave measurements from hurricanes lili, ivan, katrina and rita', *Proc. 10th Int. Workshop Wave Hindcasting Forecasting Coast. Hazards Symp., North Shore, Oahu* pp. 11–16.

Forristall, G. Z. (2011), 'Maximum crest heights under a model TLP deck', *30th Int. Ocean Offshore Arct. Eng. Conf., Rotterdam, The Netherlands* **44342**, 571–577.

Forristall, G. Z. (2015), 'Maximum crest heights over an area: laboratory measurements compared to theory', *30th Int. Ocean Offshore Arct. Eng. Conf., Rotterdam, The Netherlands* **56499**.

Forristall, G. Z., Barstow, S. F., Krogstad, H. E., Prevosto, M., Taylor, P. H. & Tromans, P. S. (2004), 'Wave crest sensor intercomparison study: An overview of WACSIS', *J. Offshore Mech. Arct. Eng.* **126**(1), 26–34.

Fujimoto, W., Waseda, T. & Webb, A. (2019), 'Impact of the four-wave quasi-resonance on freak wave shapes in the ocean', *Ocean Dyn.* pp. 1–21.

Gemmrich, J. & Thomson, J. (2017), 'Observations of the shape and group dynamics of rogue waves', *Geophys. Res. Lett.* **44**(4), 1823–1830.

Gibbs, R. H. (2004), Walls of water on the open ocean, PhD thesis, University of Oxford.

Gibbs, R. H. & Taylor, P. H. (2005), 'Formation of walls of water in 'fully' nonlinear simulations', *Appl. Ocean Res.* **27**(3), 142–157.

Gibson, R. S. & Swan, C. (2006), 'The evolution of large ocean waves: the role of local and rapid spectral changes', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **463**(2077), 21–48.

Goda, Y. (2000), *Random seas and design of maritime structures*, Vol. 15, World Sci. Publ. Co. Pte. Ltd.

Gramstad, O., Agrell, C., Bitner-Gregersen, E., Guo, B., Ruth, E. & Vanem, E. (2020), 'Sequential sampling method using gaussian process regression for estimating extreme structural response', *Mar. Struct.* **72**, 102780.

Gramstad, O. & Trulsen, K. (2011), 'Hamiltonian form of the modified nonlinear schrödinger equation for gravity waves on arbitrary depth', *J. Fluid Mech.* **670**, 404–426.

Grue, J. (2011), 'Two phenomena: Honji instability, and ringing of offshore structures', *Theor. App. Mech. Lett.* **1**(6), 062001.

Hasselmann, K. (1962), 'On the non-linear energy transfer in a gravity-wave spectrum. Part 1: General theory', *J. Fluid Mech.* **12**, 481.

Hasselmann, K., Barnett, T. P., Bouws, E., Carlson, H., Cartwright, D. E., Enke, K., Ewing, J. A., Gienapp, H., Hasselmann, D. E. & Kruseman, P. (1973), 'Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP)', *Ergänzungsh. 8-12* .

Haver, S. (2004), 'A possible freak wave event measured at the Draupner Jacket January 1 1995', *Rogue waves* **2004**, 1–8.

Henderson, K. L., Peregrine, D. H. & Dold, J. W. (1999), 'Unsteady water wave modulations: fully nonlinear solutions and comparison with the nonlinear Schrödinger equation', *Wave motion* **29**(4), 341–361.

Hughes, S. A. (1984), The TMA shallow-water spectrum description and applications, Technical report, Coastal Engineering Research Center (US).

Janssen, P. A. E. M. (1983), 'On a fourth-order envelope equation for deep-water waves', *J. Fluid Mech.* **126**, 1–11.

Janssen, P. A. E. M. (2003), 'Nonlinear four-wave interactions and freak waves', *J. Phys. Oceanogr.* **33**(4), 863–884.

Janssen, P. A. E. M. & Janssen, A. J. E. M. (2019), 'Asymptotics for the long-time evolution of kurtosis of narrow-band ocean waves', *J. Fluid Mech.* **859**, 790–818.

Janssen, P. A. E. M. & Onorato, M. (2007), 'The intermediate water depth limit of the Zakharov equation and consequences for wave prediction', *J. Phys. Oceanogr.* **37**(10), 2389–2400.

Janssen, T. T. & Herbers, T. H. C. (2009), 'Nonlinear wave statistics in a focal zone', *J. Phys. Oceanogr.* **39**(8), 1948–1964.

Jonathan, P. & Taylor, P. H. (1997), 'On irregular, nonlinear waves in a spread sea', *J. Offshore Mech. Arct. Eng.* **119**(1), 37–41.

Kakutani, T. & Michihiro, K. (1983), 'Marginal state of modulational instability—note on Benjamin-Feir instability', *J. Phys. Soc. Japan* **52**(12), 4129–4137.

Karmpadakis, I., Swan, C. & Christou, M. (2019), 'Laboratory investigation of crest height statistics in intermediate water depths', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **475**(2229), 20190183.

Karmpadakis, I., Swan, C. & Christou, M. (2020), 'Assessment of wave height distributions using an extensive field database', *Coast. Eng.* **157**, 103630.

Kharif, C., Giovanangeli, J.-P., Touboul, J., Grare, L. & Pelinovsky, E. (2008), 'Influence of wind on extreme wave events experimental and numerical approaches', *J. Fluid Mech.* **594**, 209–247.

Kharif, C. & Pelinovsky, E. (2003), 'Physical mechanisms of the rogue wave phenomenon', *Eur. J. Mech.* **22**(6), 603–634.

Kit, E. & Shemer, L. (2002), 'Spatial versions of the Zakharov and Dysthe evolution equations for deep-water gravity waves', *J. Fluid Mech.* **450**, 201–205.

Knowles, C. (1982), 'On the effects of finite depth on wind-wave spectra: 1. a comparison with deep-water equilibrium-range slope and other spectral parameters', *J. Phys. Oceanogr.* **12**(6), 556–568.

Kokorina, A. & Slunyaev, A. V. (2019), 'Lifetimes of rogue wave events in direct numerical simulations of deep-water irregular sea waves', *Fluids* **4**(2), 70.

Kouvaras, N. & Dhanak, M. R. (2018), 'Machine learning based prediction of wave breaking over a fringing reef', *Ocean Eng.* **147**, 181–194.

Krogstad, H. E., Liu, J., Socquet-Juglard, H., Dysthe, K. B. & Trulsen, K. (2004), 'Spatial extreme value analysis of nonlinear simulations of random surface waves', *23th Int. Ocean Offshore Arct. Eng. Conf., Vancouver, Canada* **37440**, 285–295.

Latheef, M. & Swan, C. (2013), 'A laboratory study of wave crest statistics and the role of directional spreading', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **469**.

Latheef, M., Swan, C. & Spinneken, J. (2017), 'A laboratory study of nonlinear changes in the directionality of extreme seas', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **473**(2199), 20160290.

Lindgren, G. (1970), 'Some Properties of a Normal Process Near a Local Maximum', *Ann. Math. Stat.* **41**(6), 1870–1883.

Lo, E. & Mei, C. C. (1985), 'A numerical study of water-wave modulation based on a higher-order nonlinear Schrödinger equation', *J. Fluid Mech.* **150**, 395–416.

Longuet-Higgins, M. S. (1952), 'On the statistical distribution of the height of sea waves', *J. Mar. Res.* **11**, 245–266.

Longuet–Higgins, M. S. (1975), 'On the joint distribution of the periods and amplitudes of sea waves', *J. Geophys. Res.* **80**(18), 2688–2694.

Martin, D. U. & Yuen, H. C. (1980), 'Quasi-recurring energy leakage in the two-space-dimensional nonlinear Schrödinger equation', *Phys. Fluids* **23**(5), 881–883.

McAllister, M. L., Adcock, T. A. A., Taylor, P. H. & van den Bremer, T. S. (2018), 'The set-down and set-up of directionally spread and crossing surface gravity wave groups', *J. Fluid Mech.* **835**, 131–169.

McAllister, M. L., Venugopal, V. & Borthwick, A. G. L. (2017), 'Wave directional spreading from point field measurements', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **473**(2200), 20160781.

Mei, C. C. (1989), *The Applied Dynamics of Ocean Surface Waves*, volume: 1 edn, World Scientific Publishing Co. Pte. Ltd, New York.

Melville, W. K. & Rapp, R. J. (1988), 'The surface velocity field in steep and breaking waves', *J. Fluid Mech.* **189**, 1–22.

Mohamad, M. A., Cousins, W. & Sapsis, T. P. (2016), 'A probabilistic decomposition synthesis method for the quantification of rare events due to internal instabilities', *J. Comput. Phys.* **322**, 288–308.

Mohamad, M. A. & Sapsis, T. P. (2016), 'Probabilistic response and rare events in mathieu' s equation under correlated parametric excitation', *Ocean Eng.* **120**, 289–297.

Mohamad, M. A. & Sapsis, T. P. (2018), 'Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems', *Proc. Natl. Acad. Sci. U. S. A.* **115**(44), 11138–11143.

Mori, N. & Janssen, P. A. E. M. (2006), 'On kurtosis and occurrence probability of freak waves', *J. Phys. Oceanogr.* **36**(7), 1471–1483.

Mori, N., Onorato, M. & Janssen, P. A. E. M. (2011), 'On the estimation of the kurtosis in directional sea states for freak wave forecasting.', *J. Phys. Oceanogr.* **41**, 1484–1497.

Mori, N., Onorato, M., Janssen, P. A. E. M., Osborne, A. R. & Serio, M. (2007), 'On the extreme statistics of long-crested deep water waves: Theory and experiments', *J. Geophys. Res.* **112**(C9), C09011.

Mori, N. & Yasuda, T. (2002), 'Effects of high-order nonlinear interactions on unidirectional wave trains', *Ocean Eng.* **29**(10), 1233–1245.

Myrhaug, D. & Kjeldsen, S. P. (1986), 'Steepness and asymmetry of extreme waves and the highest waves in deep water', *Ocean Eng.* **13**(6), 549–568.

Ochi, M. K. & Tsai, C. H. (1983), 'Prediction of occurrence of breaking waves in deep water', *J. Phys. Oceanogr.* **13**(11), 2008–2019.

Onorato, M., Cavaleri, L., Fouques, S., Gramstad, O., Janssen, P. A. E. M., Monbaliu, J., Osborne, A. R., Pakozdi, C., Serio, M., Stansberg, C. T., Toffoli, A. & Trulsen, K. (2009), 'Statistical properties of mechanically generated surface gravity waves: a laboratory experiment in a three-dimensional wave basin', *J. Fluid Mech.* **627**, 235.

Onorato, M., Osborne, A. R., Serio, M. & Bertone, S. (2001), 'Freak Waves in Random Oceanic Sea States', *Phys. Rev. Lett.* **86**(25), 5831–5834.

Onorato, M., Osborne, A. R., Serio, M. & Cavaleri, L. (2005), 'Modulational instability and non-Gaussian statistics in experimental random water-wave trains.', *Phys. Fluids* **17**, 078101.

Onorato, M., Osborne, A. R., Serio, M., Cavaleri, L., Brandini, C. & Stansberg, C. T. (2004), 'Observation of strongly non-Gaussian statistics for random sea surface gravity waves in wave flume experiments', *Phys. Rev. E* **70**(6), 067302.

Onorato, M., Osborne, A. R., Serio, M., Cavaleri, L., Brandini, C. & Stansberg, C. T. (2006), 'Extreme waves, modulational instability and second order theory: wave flume experiments on irregular waves', *Eur. J. Mech. - B/Fluids* **25**(5), 586–601.

Onorato, M., Proment, D., El, G., Randoux, S. & Suret, P. (2016), 'On the origin of heavy-tail statistics in equations of the Nonlinear Schrödinger type', *Phys. Lett. A* **380**(39), 3173 – 3177.

Onorato, M., Waseda, T., Toffoli, A., Cavaleri, L., Gramstad, O., Janssen, P. A. E. M., Kinoshita, T., Monbaliu, J., Mori, N., Osborne, A. R. et al. (2009), 'Statistical properties of directional ocean waves: the role of the modulational instability in the formation of extreme events', *Phys. Rev. Lett,* **102**(11), 114502.

Orszaghova, J., Taylor, P. H., Borthwick, A. G. L. & Raby, A. C. (2014), 'Importance of second-order wave generation for focused wave group run-up and overtopping', *Coast. Eng.* **94**, 63–79.

Piterbarg, V. I. (1996), *Asymptotic methods in the theory of Gaussian processes and fields*, Vol. 148, American Mathematical Society.

Prasada Rao, C. x. K. (1988), 'Spectral width parameter for wind-generated ocean waves', *Proc. Indian Acad. Sci. - Earth Planet. Sci.* **97**(2), 173.

Ragone, F. & Bouchet, F. (2019), 'Computation of extreme values of time averaged observables in climate models with large deviation techniques', *J. Stat. Phys.* pp. 1–29.

Rodriguez, J. D., Perez, A. & Lozano, J. A. (2009), 'Sensitivity analysis of k-fold cross validation in prediction error estimation', *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(3), 569–575.

Santo, H., Taylor, P. H., Eatock Taylor, R. & Choo, Y. S. (2013), 'Average properties of the largest waves in Hurricane Camille', *J. Offshore Mech. Arct. Eng.* **135**(1), 11602–11607.

Sapsis, T. P. (2020*a*), 'Output-weighted optimal sampling for bayesian regression and rare event statistics using few samples', *Proc. R. Soc. A Math. Phys. Eng. Sci.* **476**(2234), 20190834.

Sapsis, T. P. (2020*b*), 'Statistics of extreme events in fluid flows and waves', *Annu. Rev. Fluid Mech.* **53**.

Seiffert, B. R. & Ducrozet, G. (2018), 'Simulation of breaking waves using the high-order spectral method with laboratory experiments: wave-breaking energy dissipation', *Ocean Dyn.* **68**(1), 65–89.

Sergeeva, A. V., Slunyaev, A. V., Pelinovsky, E., Talipova, T. & Doong, D.-J. (2014), 'Numerical modeling of rogue waves in coastal waters', *Nat. Hazards Earth Syst. Sci.* **14**(4), 861.

Serio, M., Onorato, M., Osborne, A. R. & Janssen, P. A. E. M. (2005), 'On the computation of the Benjamin-Feir Index', *Nuovo Cim. della Soc. Ital. di Fis. C* **28**(6), 893–903.

Shabat, A. & Zakharov, V. (1972), 'Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media', *Sov. Phys. JETP* **34**(1), 62.

Sharma, J. N. & Dean, R. G. (1981), 'Second-order directional seas and associated wave forces', *Soc. Pet. Eng. J.* **21**(01), 129–140.

Shemer, L., Jiao, H., Kit, E. & Agnon, A. (2001), 'Evolution of a nonlinear wave field along a tank: experiments and numerical simulations based on the spatial Zakharov equation', *J. Fluid Mech.* **427**, 107–129.

Shemer, L., Kit, E., Jiao, H. & Eitan, O. (1998), 'Experiments on nonlinear

wave groups in intermediate water depth', *J. Waterw. Port, Coastal, Ocean Eng.* **124**(6), 320–327.

Shemer, L. & Sergeeva, A. V. (2009), 'An experimental study of spatial evolution of statistical parameters in a unidirectional narrow-banded random wavefield', *J. Geophys. Res: Oceans* **114**(C1).

Shemer, L., Sergeeva, A. V. & Liberzon, D. (2010), 'Effect of the initial spectrum on the spatial evolution of statistics of unidirectional nonlinear random waves', *J. Geophys. Res. Oceans* **115**(C12).

Shemer, L., Sergeeva, A. V. & Slunyaev, A. V. (2010), 'Applicability of envelope model equations for simulation of narrow-spectrum unidirectional random wave field evolution: Experimental validation', *Phys. Fluids* **22**(1), 016601.

Slunyaev, A., Kharif, C., Pelinovsky, E. & Talipova, T. (2002), 'Nonlinear wave focusing on water of finite depth', *Physica D* **173**(1-2), 77–96.

Slunyaev, A. V. (2005), 'A high-order nonlinear envelope equation for gravity waves in finite-depth water', *J. Exp. Theor.* **101**(5), 926–941.

Slunyaev, A. V. (2009), 'Numerical simulation of "limiting" envelope solitons of gravity waves on deep water', *J. Exp. Theor.* **109**(4), 676.

Slunyaev, A. V. & Pelinovsky, E. (2019), 'Numerical simulations of modulated waves in a higher-order Dysthe equation', *Water Waves* pp. 1–19.

Slunyaev, A. V., Pelinovsky, E., Sergeeva, A. V., Chabchoub, A., Hoffmann, N., Onorato, M. & Akhmediev, N. (2013), 'Super-rogue waves in simulations based on weakly nonlinear and fully nonlinear hydrodynamic equations', *Phys. Rev. E* **88**(1), 012909.

Slunyaev, A. V. & Sergeeva, A. V. (2012), 'Stochastic simulation of unidirectional intense waves in deep water applied to rogue waves', *JETP Lett.* **94**(10), 779–786.

Slunyaev, A. V. & Shrira, V. I. (2013), 'On the highest non-breaking wave in a group: fully nonlinear water wave breathers versus weakly nonlinear theory', *J. Fluid Mech.* **735**, 203–248.

Smith, C. B., ed. (2006), *Extreme Waves*, The National Academies Press, Washington, DC.

Snoek, J., Larochelle, H. & Adams, R. P. (2012), 'Practical bayesian optimization of machine learning algorithms', *Adv. Neural Inf. Process Syst.* **25**, 2951–2959.

Socquet-Juglard, H., Dysthe, K., Trulsen, K., Krogstad, H. E. & Liu, J. (2005), 'Probability distributions of surface gravity waves during spectral changes', *J. Fluid Mech.* **542**(-1), 195–216.

Stokes, G. G. (1847), 'On the theory of oscillatory waves', *Camb. Trans.* **8**, 441.

Takeishi, N., Kawahara, Y. & Yairi, T. (2017), 'Learning Koopman invariant subspaces for dynamic mode decomposition', *Adv. Neural Inf. Process Syst.* pp. 1130–1140.

Tang, T., Li, Y., Bingham, H. B. & Adcock, T. A. A. (2020), 'Comparison of two versions of the MNLS with the full water wave equations', *39th Int. Ocean Offshore Arct. Eng. Conf., Florida, USA* .

Tang, T., Tromans, P. S. & Adcock, T. A. A. (2019), 'Field measurement of nonlinear changes to large gravity wave groups', *J. Fluid Mech.* **873**, 1158–1178.

Tang, T., Xu, W., Barratt, D., Bingham, H. B., Li, Y., Taylor, P. H., van den Bremer, T. S. & Adcock, T. A. A. (2020), 'Spatial evolution of the kurtosis of steep unidirectional random waves', *J. Fluid Mech.* **908**.

Tang, T., Yelland, M. J. & Adcock, T. A. A. (2019), 'The average shape of large waves in the Norwegian sea: Is non-linear physics important?', *38th Int. Ocean Offshore Arct. Eng. Conf., Glasgow, UK* **58851**, V07BT06A058.

Tayfun, M. A. (1980), 'Narrow-band nonlinear sea waves', *J. Geophys. Res: Oceans* **85**(C3), 1548–1552.

Tayfun, M. A. & Fedele, F. (2007), 'Wave-height distributions and nonlinear effects', *Ocean Eng.* **34**(11-12), 1631–1649.

Taylor, P. H. & Williams, B. A. (2004), 'Wave statistics for intermediate depth water NewWaves and symmetry', *J. Offshore Mech. Arct. Eng.* **126**(1).

Toffoli, A., Benoit, M., Onorato, M. & Bitner-Gregersen, E. M. (2009), 'The effect of third-order nonlinearity on statistical properties of random directional waves in finite depth', *Nonlinear Process Geophys.* **16**(1), 131–139.

Toffoli, A., Fernandez, L., Monbaliu, J., Benoit, M., Gagnaire-Renou, E. M., Lefevre, J. M., Cavaleri, L., Proment, D., Pakozdi, C., Stansberg, C. T., Waseda, T. & Onorato, M. (2013), 'Experimental evidence of the modulation of a plane wave to oblique perturbations and generation of rogue waves in finite water depth', *Phys. Fluids* **25**(9), 091701.

Toffoli, A., Gramstad, O., Trulsen, K., Monbaliu, J., Bitner-Gregersen, E. & Onorato, M. (2010), 'Evolution of weakly nonlinear random directional waves: Laboratory experiments and numerical simulations', *J. Fluid Mech.* **664**, 313.

Toffoli, A., Monbaliu, J., Onorato, M., Osborne, A. R., Babanin, A. V. & Bitner-Gregersen, E. (2007), 'Second-order theory and setup in surface gravity waves: A comparison with experimental data', *J. Phys. Oceanogr.* **37**(11), 2726–2739.

Toffoli, A., Onorato, M., Bitner-Gregersen, E., Osborne, A. R. & Babanin, A. (2008), 'Surface gravity waves from direct numerical simulations of the euler equations: a comparison with second-order theory', *Ocean Eng.* **35**(3-4), 367–379.

Toffoli, A., Proment, D., Salman, H., Monbaliu, J., Frascoli, F., Dafilis, M., Stramignoni, E., Forza, R., Manfrin, M. & Onorato, M. (2017), 'Wind generated rogue waves in an annular wave flume', *Phys. Rev. Lett.* **118**(14), 144503.

Tromans, P. S., Anatruk, A. R. & Hagemeijer, P. (1991), 'New model for the kinematics of large ocean waves application as a design wave', *Proc. First Int. Offshore Polar Eng. Conf.* **8**, 64–71.

Trulsen, K. & Dysthe, K. B. (1996), 'A modified nonlinear Schrödinger equation for broader bandwidth gravity waves on deep water', *Wave motion* **24**(3), 281–289.

Trulsen, K., Kliakhandler, I., Dysthe, K. B. & Velarde, M. G. (2000), 'On weakly nonlinear modulation of waves on deep water', *Phys. Fluids* **12**(10), 2432–2437.

Tucker, M., Challenor, P. & Carter, D. J. T. (1984), 'Numerical simulation of a random sea: a common error and its effect upon wave group statistics', *Appl. Ocean Res.* **6**(2), 118–122.

Tucker, M. & Pitt, E. (2001), *Waves in Ocean Engineering (Elsevier Ocean Engineering)*, Elsevier ocean engineering book series, Elsevier.

van den Bremer, T. S. & Taylor, P. H. (2015), 'Estimates of Lagrangian transport by surface gravity wave groups: The effects of finite depth and directionality', *J. Geophys. Res. Oceans* **120**(4), 2701–2722.

Walker, D. A. G., Taylor, P. H. & Eatock Taylor, R. (2004), 'The shape of large surface waves on the open sea and the Draupner New Year wave', *Appl. Ocean Res.* **26**(3-4), 73–83.

Wan, Z. Y. & Sapsis, T. P. (2018), 'Machine learning the kinematics of spherical particles in fluid flows', *J. Fluid Mech.* **857**.

Wan, Z. Y., Vlachas, P., Koumoutsakos, P. & Sapsis, T. P. (2018), 'Data-assisted reduced-order modeling of extreme events in complex dynamical systems', *PloS one* **13**(5), e0197704.

Whitham, G. (1974), 'Linear and nonlinear waves', *John Wiley & Sons* .

Whittaker, C. N., Raby, A. C., Fitzgerald, C. J. & Taylor, P. H. (2016), 'The average shape of large waves in the coastal zone', *Coast. Eng.* **114**, 253 – 264.

Williams, M. O., Rowley, C. W. & Kevrekidis, I. G. (2014), 'A kernel-based approach to data-driven koopman spectral analysis', *Comput. Dyn.* **2**(2158-2491), 247.

Xiao, W., Liu, Y., Wu, G. & Yue, D. K. P. (2013), 'Rogue wave occurrence and dynamics by direct simulations of nonlinear wave-field evolution', *J. Fluid Mech.* **720**, 357–392.

Young, I. R. (1994), 'On the measurement of directional wave spectra', *Appl. Ocean Res.* **16**(5), 283–294.

Young, I. R., Banner, M. L., Donelan, M. A., McCormick, C., Babanin, A. V., Melville, W. K. & Veron, F. (2005), 'An Integrated System for the Study of Wind-Wave Source Terms in Finite-Depth Water', *J. Atmos. Ocean. Technol.* **22**(7), 814–831.

Young, I. R. & Verhagen, L. A. (1996), 'The growth of fetch limited waves in water of finite depth. Part 1. Total energy and peak frequency', *Coast. Eng.* **29**(1-2), 47–78.

Young, I. R., Verhagen, L. A. & Banner, M. L. (1995), 'A note on the bimodal directional spreading of fetch-limited wind waves', *J. Geophys. Res.* **100**(C1), 773.

Yuen, H. C. & Ferguson, W. E. (1978), 'Relationship between Benjamin–Feir instability and recurrence in the nonlinear Schrödinger equation', *Phys. Fluids* **21**(8), 1275–1278.

Zakharov, V. E. (1968), 'Stability of periodic waves of finite amplitude on the surface of a deep fluid', *J. Appl. Mech. Tech. Phys.* **9**(2), 190–194.

Zang, J., Taylor, P. H., Morgan, G., Stringer, R., Orszaghova, J., Grice, J. & Tello, M. (2010), 'Steep wave and breaking wave impact on offshore wind turbine foundations–ringing re-visited', *25th Int. Workshop Water Waves Floating Bodies, Harbin, China* pp. 9–12.

Zavadsky, A., Benetazzo, A. & Shemer, L. (2017), 'On the two-dimensional structure of short gravity waves in a wind wave tank.', *Phys. Fluids* **29**, 016601.

Zhang, H. D., Guedes Soares, C., Chalikov, D. & Toffoli, A. (2016), 'Modeling the spatial evolutions of nonlinear unidirectional surface gravity waves with fully nonlinear numerical method', *Ocean Eng.* **125**, 60–69.

Zhang, H. D., Guedes Soares, C. & Onorato, M. (2014), 'Modelling of the spatial evolution of extreme laboratory wave heights with the nonlinear Schrödinger and Dysthe equations', *Ocean Eng.* **89**, 1–9.

Appendix A

Linearisation theory

To remove the influence of bound harmonics from the data, we carry out a 'linearisation' process. Instead of using the exact second order interaction kernel established by Dean & Sharma (1981) and Dalzell (1999), a narrow-banded approximation following Walker's Stokes-type approximations (Walker et al. 2004) are used to estimate the size of the second order sum terms.

The Stokes regular wave expansion up to the second order term can be written as:

$$\eta(t) = a\cos\phi + \frac{\S_{22}}{d}a^2\cos 2\phi + O\Big(\frac{a^3}{d^2}\Big), \tag{A.1}$$

where a is the linear wave amplitude, $\S_{22}$ is the second order coefficient, d is the water depth and ϕ is the phase.

Our approximate linearisation starts with the calculation of second order contribution from a linear record η_L and its Hilbert transform η_{LH}:

$$\eta_L = a\cos\phi,\ \eta_{LH} = a\sin\phi. \tag{A.2}$$

The double frequency contribution η_2 can be approximated as:

$$\eta_2 = a^2 \cos 2\phi = (\eta_L^2 - \eta_{LH}^2). \tag{A.3}$$

where η_L and its Hilbert transform η_{LH} are then approximated by filtering the second order difference term out of the fully non-linear record through a high pass filter. Hence, the linear component of a record can be calculated from the approximation:

$$\eta_L \approx \eta - \frac{\S_{22}}{d}(\eta^2 - \eta_H^2), \tag{A.4}$$

Lastly, the second order coefficient can be obtained by finding the value of $\S_{22}$, for which the skewness of η_L is zero.

This approach has been used before for different dataset with different sea-states (*e.g.* Walker et al. (2004), Adcock & Taylor (2009*a*), Santo et al. (2013)). However, the Lake George data are exceptionally steep, increasing the size of the bound harmonics. Thus we present a short analysis of this method to show that it gives a good approximation to the linear signal even for exceptionally steep sea-states.

A.1 Validation

Compared to the North Sea dataset, most of the data records in the Lake George dataset are exceptionally steep, which may lead to some difficulties in linearisation. To validate the processes, one simple technique is to investigate the difference between sorted measurements of peak elevation and trough depression, since the vertical asymmetry properties are dominated by the second order contributions in Stokes expansions. Therefore, the local maximum between up and down crossings and the local minimum between down and up crossings are sorted based on their own rankings and the n^{th} largest trough and n^{th} largest crest are paired and plotted in Figure A.1 (a). There is no temporal relationship between the crest and trough in each pair. A

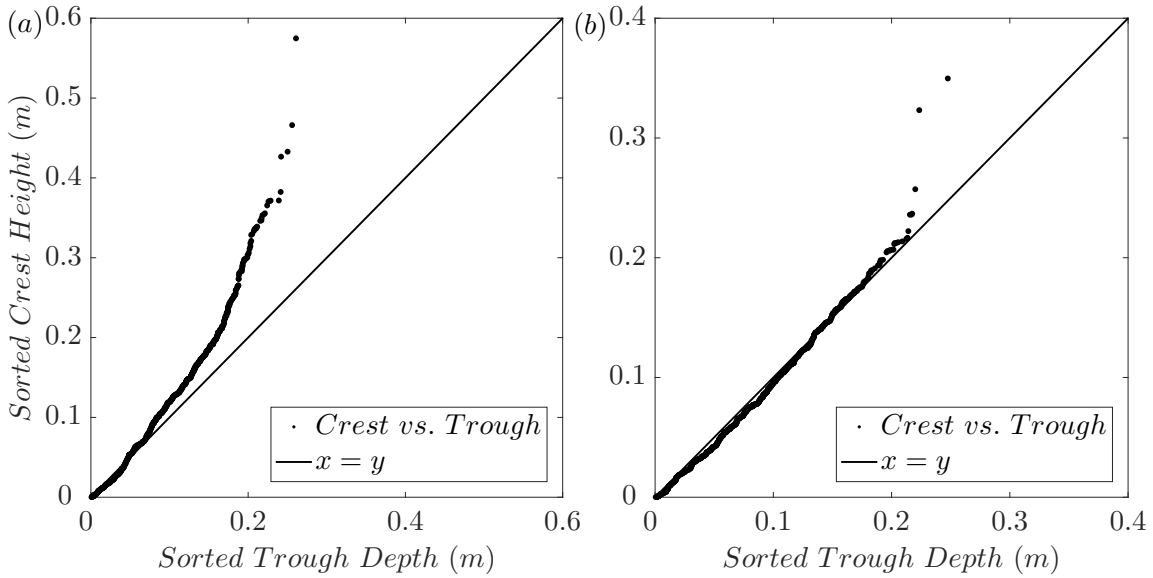

Figure A.1: Order statistics for crest and troughs for (a): one original time series, (b): linearised time series.

guide line is also plotted to represent the record without any vertical asymmetry as would be expected for linear random waves.

It is clear that there is a strong vertical asymmetry in this sample time series (for instance stronger than the one presented by Taylor & Williams (2004)) as the crest-trough pairs depart the red line very soon from the beginning and there is a considerable difference between the largest crests and troughs as well. This also suggests that the Lake George dataset contains many records with very high non-linearity.

As for the difference between crests and the troughs after the linearisation process in Figure A.1 (b), the sorted crest-trough pairs roughly follow the guideline except for several of largest pairs. Thus the Stokes expansion A.1 is still valid for extreme cases and the linearisation processes described in the previous subsection are acceptable.

Appendix B

The kurtosis of unidirectional random waves

B.1 Evolution of excess kurtosis

In this appendix, we explicitly provide the equations we use to evaluate the analytical solutions shown in figure 5.2

B.1.1 The NLS (Mori & Janssen 2006)

Mori & Janssen (2006) present the evolution of excess kurtosis for the time evolution of spatially homogeneous unidirectional waves that are initially normally distributed in the form of a three-dimensional integral (their Eq. (14)), which is subsequently evaluated in the narrow-bandwidth and large-time limit for Gaussian spectra (their (28)). Without taking the large-time limit, but invoking the other two solutions, this integral can be evaluated in closed-form (Fedele et al. 2010) (see also Fedele (2015*b*), Janssen & Janssen (2019)):

$$\frac{C^{d}_{4,\mathrm{NLS}}(x)}{\mathrm{BFI}^2} = \frac{\pi}{3\sqrt{3}}\Big[1 - \frac{6}{\pi}\mathrm{Im}\Big(i \arcsin\frac{1+2i\alpha}{2}\Big)\Big], \tag{B.1}$$

where $C^d_{4,\mathrm{NLS}}$ is the dynamic excess kurtosis based on NLS, Im is the imaginary part, $\alpha = 2\nu^2 x/\lambda_0$ and ν is the bandwidth of the frequency spectrum. To obtain (B.1), we have also converted from time to space (see Fedele et al. (2010)).

B.1.2 The cDZ (Fedele 2014)

Fedele (2014) also proposes a correction of $C^d_{4,\mathrm{NLS}}$ based on the compact Zakharov (cDZ) equation:

$$C^d_{4,\mathrm{cDZ}} = C^d_{4,\mathrm{NLS}}(1 - 0.4\nu_k^2). \tag{B.2}$$

Here, $\nu_k \approx 2\nu$ is the spectral bandwidth of the wavenumber spectrum $S(k)$ and ν that of the frequency spectrum $S(f)$. Then,

$$C^d_{4,\mathrm{cDZ}} = C^d_{4,\mathrm{NLS}}(1 - 1.6\nu^2), \tag{B.3}$$

for narrow-band waves (small ν).

We do not use the full model derived in Fedele (2014), which is valid for waves evolving in time, but only the leading-order correction in bandwidth given here.

B.2 Figure with BFI at steady state

In figure B.1 we show the same results as in figure 5.2 but with excess kurtosis normalised by at steady state $\mathrm{BFI}^2_{\mathrm{ss}}$. We thus obtain better agreement with theoretical results at large distances but worse agreement at small distances. Evidently, neither normalisation captures the occurrence of a maximum.

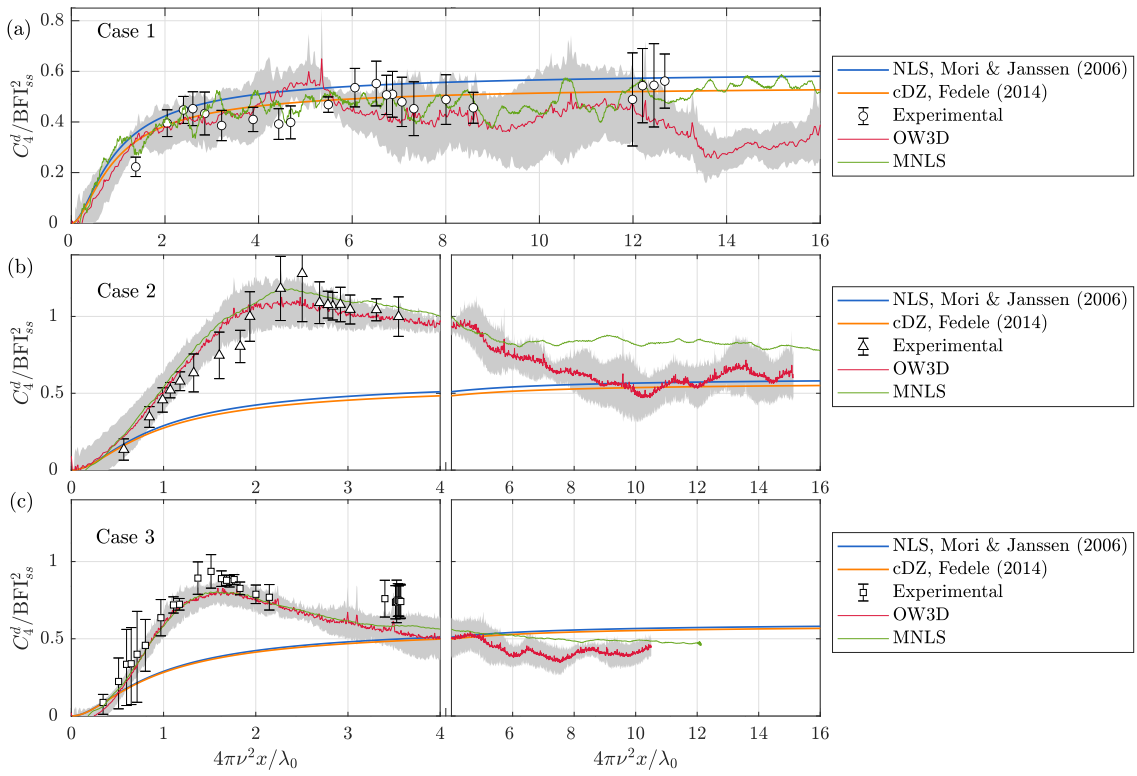

Figure B.1: Evolution of normalised dynamic excess kurtosis at different distances from the wave generator: (a) case 1, (b) case 2 and (c) case 3. Shading represents the 95% confidence intervals for OceanWave3D simulation with 8 different random seeds. A total of 120 different random seeds are used in MNLS simulations. Consequently, for the MNLS, the error bars are negligible and have been omitted for clarity. The parameter C_4^d is dynamic excess kurtosis, BFI_{ss} is the Benjamin-Feir Index at steady state, ν is the input bandwidth and λ_0 is the peak wave length. This figure is equivalent to figure 5.2 except for the normalisation by the steady-state (this figure) rather than the input BFI (figure 5.2).

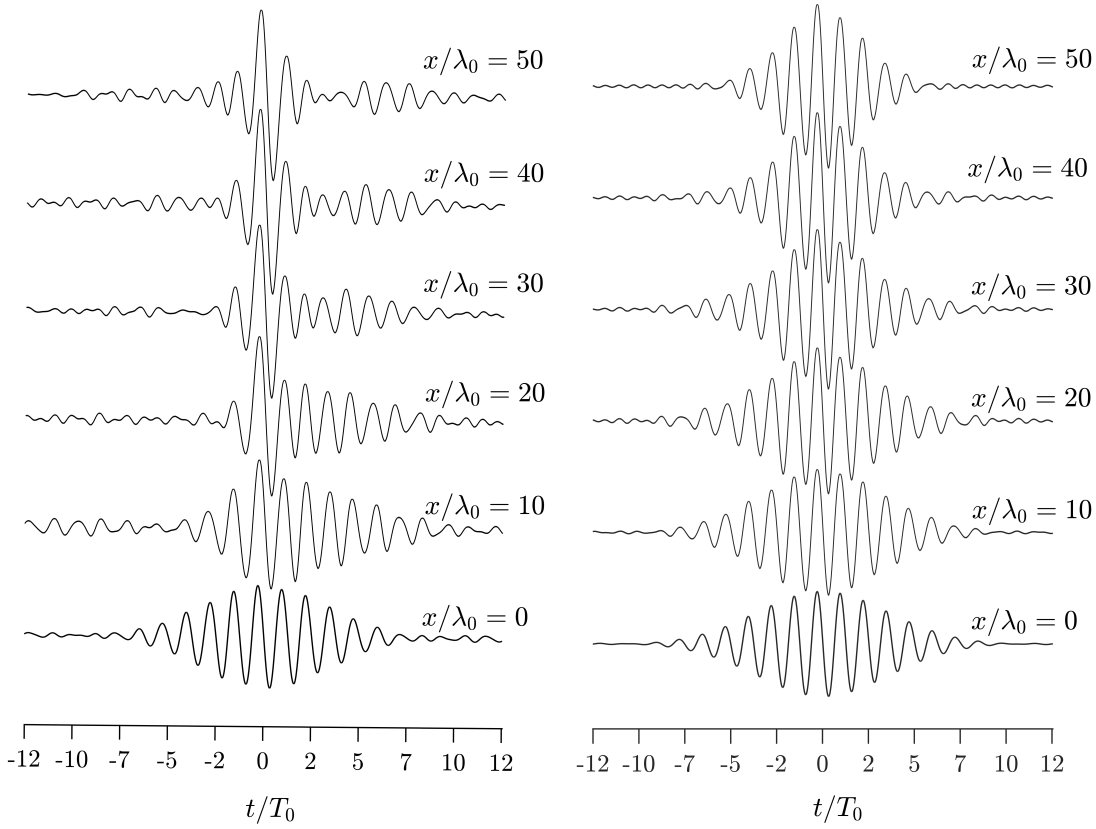

Figure B.2: Average shape of the 5 largest crest-to-trough wave height profiles out of over 6400 waves profiles (left) and wave profiles predicted by the theory of quasi-determinism (right) at $x/\lambda_0 = 0, 10, 20, 30, 40, 50$ for random waves with a Gaussian input spectrum with $\Delta f/f_0 = 0.054$ and $\epsilon = 0.044$.

B.3 Averaged shape of the wave with the largest crest-to-trough height

The theory of quasi-determinism predicts the shape of the wave with the largest crest-to-trough height as the scaled difference of two time-shifted autocovariance functions (Boccotti 1989, 2000).

Figure B.2 (a) presents the average shape of the 5 largest crest-to-trough wave profiles at different locations. The general trend of the shape evolution is very similar to that of the largest crest events presented in figure 5.6. We also observe a movement of the largest crest to the front of the wave group and a contraction of the wave group

for extreme crest-to-trough events. Figure B.2 (b) presents the predictions from the (linear) theory quasi-determinism given in Boccotti (1989, 2000). All the changes to the group shape in figure B.2 (b) are due to changes in the wave spectrum.

Appendix C

Envelope based second order corrections

For the second order corrections, an envelope method is used to compute the sum and difference terms:

$$\eta = \Re(\eta_{linear} + \eta_{2-} + \eta_{2+}). \tag{C.1}$$

For the linear part η_{linear}, we follow the standard envelope equation:

$$\eta_{linear} = U \exp(i(kx - \omega t)), \tag{C.2}$$

where k is the dominant wave number, U is the linear complex envelope, and ω is the wave frequency.

For second order sum terms with bandwidth correction can be expressed as:

$$\eta_{2+} = \left[\text{Coef0}\ \frac{kU^2}{2} - \text{Coef1}\ iU\frac{\partial U}{\partial x} + \text{Coef2}\ U\frac{\partial^2 U}{\partial y^2} + \text{Coef3}\ \frac{1}{k}\left(\frac{\partial U}{\partial y}\right)^2\right] \exp(2i(kx - \omega t)), \tag{C.3}$$

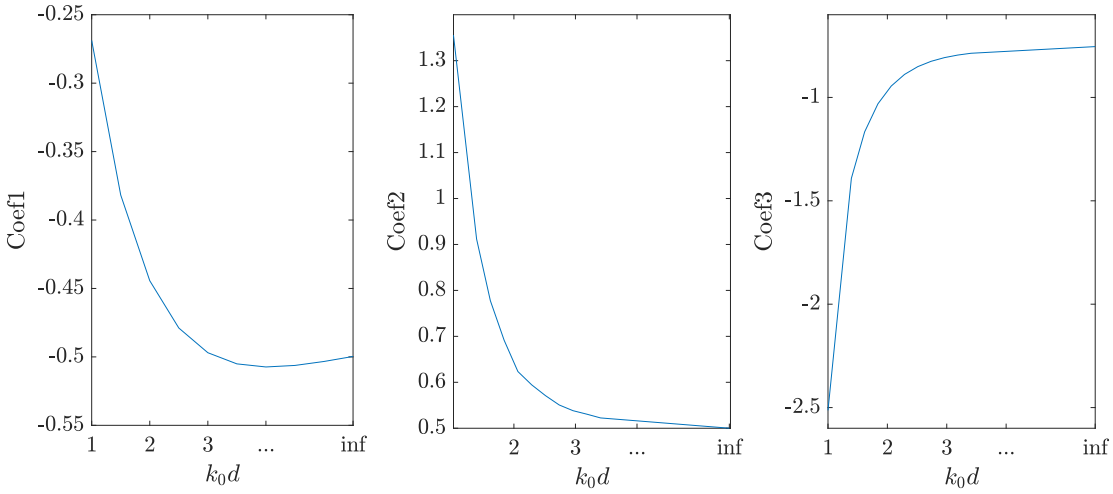

Figure C.1: Bandwidth correction coefficients at different relative water depths.

where Coef0 is the leading order coefficient, which is given in Slunyaev (2005) for finite water depth. For the bandwidth correction coefficients Coef1, Coef1, Coef3, Trulsen & Dysthe (1996) and Toffoli et al. (2010) give the values in the deep water:

$$\text{Coef1} = -\frac{1}{2}, \qquad \text{Coef2} = \frac{1}{2}, \qquad \text{and} \qquad \text{Coef3} = -\frac{3}{4}. \tag{C.4}$$

To obtain the bandwidth correction coefficients at finite water depth, we used a global optimisation code to find the best combination of all three bandwidth correction coefficients. The target of the optimisation code is the exact solution presented in Dalzell (1999). This process is repeated at different water depths to obtain the bandwidth correction coefficient values at different the water depths.

The obtained bandwidth correction coefficients at different relative water depths is shown in Figure C.1. All the bandwidth correction coefficients converges to the deep water value shown in Equations C.4 when the relative water depth tends to infinity. This suggests that the global optimisation method seems to successfully find the optimal combinations to provide the best bandwidth corrections for the second order sum term.

For the second order difference term, we followed Trulsen & Dysthe (1996) and Toffoli et al. (2010) for deep water cases:

$$\eta_{2-} = \frac{1}{2\omega}\frac{\partial\phi}{\partial x} - \frac{1}{16k}\frac{\partial^2|U|^2}{\partial x^2} - \frac{1}{8k}\frac{\partial^2|U|^2}{\partial y^2}, \tag{C.5}$$

where ϕ is the velocity potential. For the second order difference term in finite water depth, since we are only interested at the peak of the largest crest, we followed the an simplified equation used by McAllister et al. (2018) as:

$$\eta_{2-,c} = -\frac{-|U|^2}{4d}\frac{1}{1+R}, \tag{C.6}$$

where $\eta_{2-,c}$ is the difference term at the envelope peak, d is the water depth, and R is the spatial aspect ratio of the wave group. We examine the accuracy of the difference term by comparing the results against the exact solution presented in Dalzell (1999) at different water depths. Figure C.2 presents the percentage difference at different water depth. There is about 3% of difference in terms of the trough of the difference term at the shallow water end and everything is accurate on the deep water side.

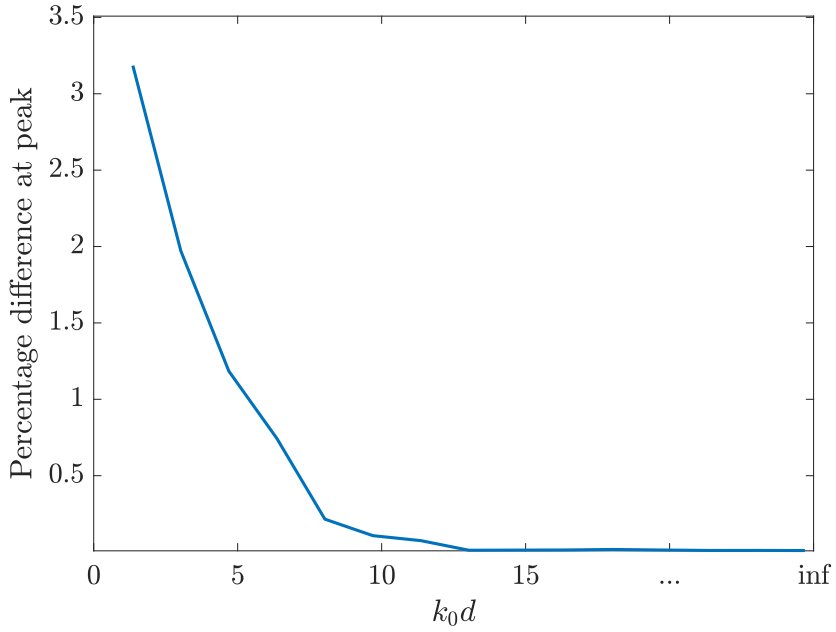

Figure C.2: Percentage difference of second order difference term at different water depths.