# Unwavering Potential

## Optimizing and Estimating the Power Output from a Wave Energy Converter

Mårten Augustsson

Master's thesis
2024:E36

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

**Abstract**

Wave power is a type of renewable energy which holds great potential, but this potential remains largely unused due to difficulties in implementing commercially viable wave energy converters. This study uses Monte Carlo integration to optimize and estimate the electrical power output of a Simulink model based on a real wave energy converter developed by Waves4Power. For 78 sea states defined by their significant wave heights and wave energy periods, we find the optimal choices of which electrical generators to set as active, as well as the optimal generator rotation speeds, using randomly generated wave height time series. These optimal settings are then used on new time series to find estimates of the resulting mean power output and minute-to-minute variance. The optimization is performed with regards to a statistic consisting of the sample mean minus two times the standard deviation, favoring generator settings which give high mean output while simultaneously having little variability. Finally, we fit a logistic regression model to the optimal choices of active generators and linear regression models to the optimal rotation speeds and estimated mean outputs, in all three cases using the significant wave height and wave energy period as predictor variables.

The study finds a mean output of 21.74 kW when taking a weighted average of all sea states by frequency of occurrence, which is a 9% decrease from the mean when optimizing the settings with regards to only the sample mean (23.90 kW). However, the minute-to-minute standard deviation is also reduced by 37% when including standard deviation in the optimization statistic. The efficiency of the wave energy converter, found by dividing the output in each sea state by the theoretical energy potential of that state, also has its mean and standard deviation reduced, though in both cases by less than the unscaled power output. The regression models are successfully fitted and tested on validation data, and in the case of the generator settings give potentially useful decision rules for practical uses of the wave energy converter.

# Populärvetenskaplig sammanfattning

Vågkraft är en energikälla som på många sätt sticker ut från andra förnybara energikällor. Å ena sidan har havsvågor en helt annan regelbundenhet än till exempel vind eller solsken och har därför potentialen att utgöra en mycket mer tillförlitlig energikälla än dessa. Energiinnehållet i världens havsvågor är dessutom oerhört stort och jämförbart med världens totala förbrukning av elektricitet. Å andra sidan är teknologin för vågkraft långt ifrån lika utvecklad som den för sol- eller vindkraft, och än idag finns mycket få exempel på lyckade kommersiella implementeringar av vågkraftverk. Samtidigt som energipotentialen i havsvågor är oerhört stor är det en resurs som är mycket svår att utnyttja effektivt.

Denna masteruppsats i matematisk statistik studerar energiproduktionen i ett vågkraftverk, hur den kan optimeras, vad medelvärdet blir och hur mycket effekten varierar kring detta medelvärde. Studien använder sig av en simulator i MathWorks verktyg Simulink, vilken i sin tur är baserad på ett verkligt vågkraftverk som tidigare använts på en teststation i Runde i Norge och som är på väg att sjösättas igen under 2024. Kraftverket består av en flytande boj med en cylinder som går djupt ner i vattnet, och i vilken en kolv rör sig upp och ner i otakt med bojens guppande på vattenytan. Denna otakt skapar en pumpande rörelse som genom ett hydrauliskt system och två generatorer skapar elektrisk energi som kan skickas in i ett elnät.

Vågor slumpgenereras som tidsserier där de förenklas till våghöjder över tid på den punkt på havsytan där bojen befinner sig. Detta sker för 78 olika så kallade havstillstånd, kategorier som beskriver havets beteende i kraftverkets närhet och ger olika mönster i vågserierna. Genom simulatorn kan vi sedan få fram den elektriska effekten som följer av de givna våghöjderna under ett visst tidsspann, men på grund av den stora komplexiteten i omvandlingsprocessen från våghöjd till effekt är det praktiskt omöjligt att exakt beräkna parametrar såsom den förväntade effekten eller variationen i effekt. Istället görs sådana beräkningar approximativt genom att för varje havstillstånd simulera en mycket lång vågserie, köra den genom simulatorn för att få fram effekten, och använda medelvärdet över tid som en skattning av den ”sanna” förväntade effekten för havstillståndet.

Inledningsvis genomförs dock en optimeringsprocess där effekten av inställningarna på kraftverkets generatorer undersöks. För varje havstillstånd genomförs en maximeringsalgoritm för att se vilket val av generatorer – endast den ena (50 kW), endast den andra (100 kW) eller båda – som ger bäst effekt, och vilket varvtal generatorn eller generatorerna bör hålla för att maximera effekten. Eftersom det finns ett intresse av att effekten inte bara ska ha ett högt medelvärde utan också vara stabil kring medelvärdet utan för mycket variation används inte medelvärdet i sig som det värde som ska maximeras, utan medelvärdet minus två gånger standardavvikelsen. Optimeringen genomförs dock också med endast medelvärdet som måltal, för att resultaten från den huvudsakliga optimeringen ska kunna jämföras med denna.

Efter att ha tagit fram dessa optimala generatorval och varvtal genomförs en skattning (på nya vågserier) för att hitta den skattade medeleffekten och variansen för varje havstillstånd, givet dessa inställningar. Dessutom beräknas ett viktat medelvärde av de olika havstillståndens medeleffekter (och varianser) för att få fram en skattning av den generella medeleffekten när vi inte betingar på

ett särskilt havstillstånd.

Den obetingade medeleffekten, med alla havstillstånd sammanvägda, är knappt 22 kW, och kraftverkets effektivitet (för varje tillstånd beräknad som medeleffekten delad på den teoretiska energipotentialen i det tillståndets vågor) ca 21.5%. Drygt en femtedel av den vågenergi som träffar bojen omvandlas alltså till elektrisk energi. Om generatorval och varvtal optimeras enbart utifrån att medeleffekten ska maximeras ökar dessa siffror till 24 kW respektive 23%, men detta innebär också en markant ökning av variansen i effekt och effektivitet.

Slutligen genomförs regressionsanalyser på såväl generatorval och varvtal som på medeleffekterna för att hitta lämpliga regler för hur dessa värden kan förklaras av havstillståndens karakteristika – med andra ord av vågornas höjd och längd. Dessa analyser leder, i generator- och varvtalsfallen, till regler som potentiellt skulle kunna användas för att i verkligheten anpassa ett vågkraftverk efter de rådande förhållandena.

# Acknowledgements

My first and most profound thanks go to my supervisor, Magnus Wiktorsson, for not only being a wellspring of advice, feedback and ideas during the writing process, but for providing me with the idea for the thesis as well as material and code which has been essential to the project. His guidance and knowledge has been invaluable for this thesis coming into fruition.

I would also like to thank my family – in particular my partner Marina and my parents – for delivering moral support and writing advice when I have needed it the most, and for being so very invested in my work this spring despite no prior knowledge of either the topic or the academic subject. For Marina, I want to add a further thank you for starting long code runs when time was of the essence and I was away from Lund for the weekend.

Thanks are owed to all those friends – "ingen nämnd, ingen glömd" as the Swedish saying goes – who have shown their support and interest. Among them I wish to give a particular thanks to Joel Bertilsson, who sold me the computer on which several weeks' worth of computations have been performed and without which this quite empirical project would have been a practical impossibility.

Lastly, I want to thank all members – past, present and future – of Östgöta nation, for the literal and figurative home the nation has been for me during the seven years since a 20-year old me moved to Lund to study in the political science and economics program. Though the nation has been a smaller part of my life during this spring, for the joy it has given me throughout the long academic journey for which this thesis marks the end, I want to give it a deep and heartfelt final thank you.

*"Ditt minne skall förbliva"*
*– August Lindh*

Mårten Augustsson
Lund, June 12, 2024

# Contents

# Notation and definitions

Meaning of shorthands used in the thesis, unless another local meaning is explicitly stated.

### Non-letters

$\bar{a}$ Arithmetic average of $a$, $\bar{a} = \frac{1}{N}\sum_{i=1}^{N} a_i$.

$\hat{a}$ Estimate of $\hat{a}$.

$a^*$ Complex conjugate of $a$.

### Roman letters

$\mathbb{C}[X, Y]$ Covariance of random variables $X$ and $Y$.

$CI_\theta$ Confidence interval for parameter $\theta$.

$\mathbf{Corr}[X, Y]$ Correlation of random variables $X$ and $Y$.

$\mathbb{E}[X]$ Expected value of random variable $X$.

$e_i$ Residual of observation $i$ in a linear regression, $e_i = y_i - \hat{y}_i$.

$\mathbf{F} = (\vec{F}, \vec{M})$ The force vector enacted upon a body by a wave, in six different modes: Surge, sway, heave, roll, pitch and yaw. The first three components ($\vec{F}$) are translational and the last three ($\vec{M}$) are rotational.

$g$ The gravity acceleration constant.

$\mathbf{H}$ Hat matrix, the matrix linking the predicted and observed response values in a regression analysis.

$H_0$ Null hypothesis.

$H_s$ The significant wave height in meters of a wave state, defined as the average of the highest third of the waves, from trough to crest. Estimated as $4\sqrt{m_0}$ where $m_0$ is the zeroth moment of the wave state's spectral density.

$H(\omega)$ Frequency response function.

$h(t)$ Time-domain impulse response function.

$I$ The intensity of a wave, time-average energy transport per unit time and per unit area in the propagation direction.

$i$ The imaginary unit, $i = \sqrt{-1}$.

$J$ The wave-energy transport, in kW/m. The wave energy transported along a vertical strip parallel to the wave front.

$J_\alpha(x)$ Bessel function of the first kind.

$K$ Number of segments created in a segmentation of a stochastic process, $N = KN_{sub}$.

$m$ Sample mean.

$m_n$ The n:th moment of a spectral density, $m_n = \int_{-\infty}^{\infty} \omega^n S(\omega) d\omega$.

$N$ Total number of observations.

$n$ Number of separate processes, $N = nT$.

$N_{sub}$ Number of observations in a segment, $N = KN_{sub}$.

**OWC** Oscillating Water Column, the type of wave energy converter used in the study.

$p = p(x, z, t)$ Dynamic pressure of a plane harmonic wave.

$P$ The power received by an oscillating body.

$P_{out}$ The power output of a WEC.

$\mathbb{P}(A)$ Probability of event $A$.

$Q_{t,k}$ The volume flow of the water surface oscillating within an OWC type WEC.

$R[X]$ Coefficient of variation for random variable $X$, defined as the ratio between standard deviation and expected value.

$R^2$ Coefficient of determination in a linear regression model, $R^2 = 1 - SSE/SST$ where $SSE$ and $SST$ are the residual and total sums of squares.

$R^2_{adj}$ Adjusted coefficient of determination in a linear regression model, $R^2_{adj} = 1 - \frac{SSE/(N-p)}{SST/(N-1)}$ where $SSE$ and $SST$ are the residual and total sums of squares, while $N$ is the number of observations and $p$ the number of model coefficients including the intercept.

$r(\tau)$ The covariance function of a stationary stochastic process for lag $\tau$.

$r_{opt}$ Optimal generator rotation speed for a sea state.

$\mathbf{Re}\{a\}$ The real part of complex number $a$.

$S(f)$ The spectral density of a stationary stochastic process for frequency $f$, with period $2\pi$.

$S(\omega)$ The corresponding spectral density with period $T_e$, for frequency $\omega$.

$s^2$ Unbiased estimator of the variance $\sigma^2$.

$T$ Number of observations in a single stochastic process, with in total $N = nT$ observations for $n$ processes.

$T_e$ The energy period, in seconds, of a wave state, "the mean wave period with respect to the spectral distribution of energy" (Mollison, 1994, p. 207).

$T_p$ The peak period, in seconds, of a wave state, meaning the time between the peaks of the most dominant waves. $T_p = 2\pi/\omega_p$.

$t_\alpha(f)$ The $\alpha$ quantile of a Student's $t$-distributed random variable $X$ with $f$ degrees of freedom, such that $\mathbb{P}(X \leq t_\alpha(f)) = \alpha$.

$\mathbf{u} = (\vec{U}, \vec{\Omega})$ The velocity vector of a wave for the six different modes: Surge, sway, heave, roll, pitch and yaw. Three components ($\vec{U}$) are translational and three ($\vec{\Omega}$) are rotational.

$\mathbb{V}[X]$ Variance of random variable $X$.

**W** Weight matrix in a weighted least squares regression model.

**WEC** Wave Energy Converter.

### Greek letters

$\boldsymbol{\beta}$ Vector of coefficients corresponding to the $p$ predictors $X_j$ in a regression model, $j = 0, 1, 2, \ldots, p-1$. $\beta_0$ is the model's intercept corresponding to the constant $X_0 = 1$.

$\Gamma(a)$ The gamma function.

$\Gamma(a, b)$ The upper incomplete gamma function.

$\epsilon_i$ The random error of observation $i$ in a regression model, $\epsilon_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$.

$\zeta_j$ The j:th intercept in an ordinal logistic regression model.

$\eta$ Wave elevation.

$\lambda$ The penalty strength on the standard deviation in an optimization statistic which seeks to achieve high mean and low variance.

$\mu$ See $\mathbb{E}[X]$.

$\nu_g$ The group velocity of a water wave.

$\nu_p$ The phase velocity of a water wave.

$\rho$ The mass density of a fluid, specifically sea water.

$\boldsymbol{\Sigma}$ Covariance matrix of multiple random variables.

$\sigma^2$ See $\mathbb{V}[X]$.

$\tau$ Lag between two different time points.

$\phi$ The velocity potential of a wave.

$\phi_\alpha$ The $\alpha$ quantile of a standard Gaussian random variable $Z$, such that $\mathbb{P}(Z \leq \phi_\alpha) = \alpha$.

$\chi^2_\alpha(f)$ The $\alpha$ quantile of a $\chi^2$-distributed random variable $X$ with $f$ degrees of freedom, such that $\mathbb{P}(X \leq \chi^2_\alpha(f)) = \alpha$.

$\psi_X(t)$ Moment generating function of random variable $X$.

$\omega_p$ The peak frequency of a spectral density, $\omega_p = 2\pi/T_p$.

# Chapter 1

# Introduction

Wave power, also known as wave energy, is the conversion of the energy in the ocean's wind-generated waves into electrical power (United States Energy Information Administration, 2022). In a time when the climate crisis amplifies the need for cheap renewable energy (United Nations, 2024), wave power stands out as being a "highly concentrated energy resource" with power density many times higher than that of solar or wind power (CorPowerOcean, 2024). Barstow et al. (2008, p. 93) write that the global energy resource in water waves has an effect of approximately 2 TW, which over a year becomes approximately 18 000 TWh. This is of the same order as the total global electricity consumption, which in 2022 was estimated to 25 530 TWh (Statista, 2024). Even though the same authors also estimate a limit of 10-25% to the proportion of this energy that could be used (Barstow et al., 2008, p. 93), that proportion could still be a great contribution to the world's energy resources. Furthermore, water waves also have much greater consistency and predictability than other renewable energy sources such as wind or solar, potentially enabling much greater stability in an electrical grid using wave energy (Wu, 2023).

However, at present wave power makes up a vanishingly small percentage of the world's energy production (Ritchie and Rosado, 2024). Problems with implementing wave power at a large scale include the challenge of producing conversion systems which can handle extreme wave conditions, and overcoming the currently high costs of the implementation (Burke and Petraviciute, 2021). These costs are further augmented by the need for frequent maintenance of the devices due to exposure to the sea, and there is also the concern that large-scale implementation of wave energy devices would disturb the marine life (Wu, 2023).

Nevertheless, the energy potential in water waves is too great to ignore and wave energy conversion is the subject of intense study and development today (United States Energy Information Administration, 2022; Wu, 2023; Burke and Petraviciute, 2021). The European Commission (2024) furthermore estimates that ocean energy (wave, tidal and others) could make up 10% of the EU's energy production in 2050. The study of the effects of wave energy converters is of great practical interest for the energy sector and, by extension, the world at large.

## 1.1 Aims

This thesis aims to study the case of a particular model of wave energy converter (WEC) of the oscillating water column (OWC) type, and its power output under various marine conditions called sea states. Five fundamental questions are asked, explored and answered:

1. What is the optimum speed for each sea state, in rotations per minute, of the generators which convert hydraulic energy to electrical energy in the WEC?

2. With two available generators in the WEC, for which sea states is it optimal to use one, the other, or both?

3. Given these optimal rotation speeds and generator choices, what is the mean power output of the WEC, for each sea state and overall?

4. Under the same conditions, what is the variance of the power output of the WEC, for each sea state and overall?

5. How can linear and logistic regression be used to explain the generator choice, optimal rotation speed and mean power output, conditionally on the sea state's significant wave height and energy period?

Due to the complexity of the power output as a function of the wave height, questions 1 through 4 will be explored using Monte Carlo methods, and therefore only be answered approximately.

The reader is assumed to know the basics of probability theory and statistics such as expected values, variances and covariances, probability density functions and cumulative density functions, confidence intervals and statistical inference. Basic mathematical knowledge is also expected.

## 1.2 Historical background

### 1.2.1 Wave power

Wave power has been theoretically envisioned since at least 1799, when a first patent application for wave energy technology was made by the Frenchman Pierre-Simon Girard (Li, Sun, and Zhou, 2022, p. 2-3). After the 1799 initiative, the development was at a relative standstill until the early 1960's when the japanese naval commander Yoshio Masuda invented a WEC system (Aderinto and Li, 2019, p. 2). Masuda, by some named "the father of modern wave energy technology" (Falcão, 2010, p. 899) was also the first to develop the concept of an oscillating water column (OWC) type WEC in the 1940's, a type of WEC which is one of the most popular (Alves, 2016, p. 28). After a period of slower development, during the 1970's the oil crisis led to increased research in renewable energy including wave power, and new WEC:s were developed (Li, Sun, and Zhou, 2022, p. 3). However, the early WEC:s based on Masuda's designs still had low power output due to the very limited available knowledge on wave energy absorption (Aderinto and Li, 2019, p. 2). Though the funding of wave energy projects decreased as oil prices dropped in the 1980's, the decade saw the testing of several WEC prototypes in Scotland, India, Japan and Norway, but proper marine energy testing centers were not established until the early 21st century (Li, Sun, and Zhou, 2022, p. 3). The world's first floating WEC to provide electricity to a national grid, the Pelamis 750 device, was implemented in 2004 at Billia Croo on the island of Orkney north of Scotland (European Marine Energy Centre, 2022b; European Marine Energy Centre, 2024). Ahamed, McKee, and Howard (2020, p. 1) describe the research on wave energy as having happened in two periods, first during the 1970's and then during the "present", presumably meaning the 21st century.

Despite estimations that the theoretical energy potential of wave power globally could be as high as 29 500 TWh in a year (Edenhofer et al., 2011, p. 504), according to Our World in Data (Ritchie and Rosado, 2024) the category "other renewables" (which includes both wave and tidal power) only produced 89.81 TWh of electricity in 2023 – 0.304% of the potential above and 0.305% of the electricity produced from all sources globally. Astariz and Iglesias (2015, p. 6) cite the levelized cost of energy for wave power as between 90 and 140 €/MWh for nearshore plants, and between 180 and 490 €/MWh for offshore ones. This is a high cost in comparison to other energy sources, renewable and non-renewable, which generally range from 30 to 70 €/MWh with offshore wind power being an outlier at 101 €/MWh (Astariz and Iglesias, 2015, table 8), thus explaining the current lack of incentive for commercialized wave power.

### 1.2.2 Monte Carlo methods

The statistical methodology that has come to be known as Monte Carlo methods can be traced back to Stanislaw Ulam, a mathematician working at the Los Alamos National Laboratory. In 1946, while playing solitaire, Ulam asked himself what the probability was of a successful solitaire outcome, and considered the notion that it might be easier to play 100 times and check the number of successful outcomes, rather than calculate the probability explicitly through combinatorics (Eckhardt, 1987, p. 131). This notion, as well as the possibility of applying the same type of method to nuclear energy research, was soon communicated to fellow Los Alamos mathematician John von Neumann (Metropolis, 1987, p. 126-127). The idea was made all the more relevant by two events which occurred in 1945: the first successful test detonation of an atomic bomb at Alamogordo, and the building of the first electronic computer, ENIAC (Metropolis, 1987, p. 125).

Two letters between John von Neumann and Robert Richtmyer in March-April 1947 mark the first known written documentation of the method. In the first, von Neumann highlights the usefulness of statistical methods for solving neutron diffusion and multiplication problems using ENIAC for simulation and computing. The idea he expresses is to follow the simulated paths of 100 neutrons and through their empirical behaviour draw conclusions on the general behaviour of neutrons (Neumann and Richtmyer, 1947). The method quickly garnered interest and was, according to Nicholas Metropolis (1987, p. 127), named by himself after the Monte Carlo casino in Monaco which Ulam's uncle used to frequent.

## 1.3 Previous studies

This thesis owes a great deal to the article "Optimizing the Hydraulic Power Take-Off System in a Wave Energy Converter" by Zeinali et al. (forthcoming), which uses the same methods and Simulink model for generating random waves and simulating their conversion into electrical energy through a WEC. The aim of that study is to optimize the hydraulic pressure in the accumulator tank, which it does for 78 different sea states. This thesis aims to use these optimal pressures as initial values in its simulations, and furthermore try to optimize the rotation speed of the generator(s) involved in converting the hydraulic energy into electrical energy. It is also a goal to estimate the expected electrical outputs for each sea state, given that the rotation speed is optimal, where Zeinali et al. concern themselves with the retension of the waves' energy into the WEC:s hydraulic system but not the subsequent conversion into electrical energy.

There are many studies investigating the power efficiency of different WEC:s, defined in (Aderinto and Li, 2019, p. 4) as the ratio between the power absorbed by a device and the power available within the width of its primary wave-interacting part. Aderinto and Li (2019) published a review study in 2019 which evaluated the efficiency of three different types of WEC devices, including

the OWC type. They found efficiencies ranging from 5% to 40% for the OWC devices based on ten studies, with six of them showing efficiencies in the 20-25% range. 20-30% seemed to be the normal efficiency range in most studies, though a few studies of heaving or pitching oscillating body devices reached the range 40-50% (Aderinto and Li, 2019, Fig. 3a). An average efficiency of 24% was reported by Zeinali et al. (forthcoming, p. 8) for the same WEC studied in this thesis, though this varied heavily (between 90 and almost 0%) for different sea states.

Ren, Tan, and Xing (2023) use a Kriging Machine Learning approach to estimate the power output matrix (for different sea state parameters) of an oscillating body WEC, predicting the mean output conditionally on the sea state parameters and reaching very low prediction errors through an active learning approach. The article does not present results regarding the power output or efficiency of its WEC, instead focusing on the methodology and the effectiveness of their algorithm in making accurate predictions.

Babarit, Hals, et al. (2021) take a more empirical approach in a review study on the annual absorbed energy of eight WEC:s of different types, at five different locations. One of the locations is the European Marine Energy Centre on Orkney north of Scotland (Babarit, Hals, et al., 2021, fig. 9), the centre which provides the sea state data for this thesis. The study finds annual mean absorbed power values between 1.6 and 981 kW for the various devices and locations, while also comparing the power relative to the mass of each device, its surface area and its root mean square power take-off force (Babarit, Hals, et al., 2021, fig. 12-19). One of the devices is a floating oscillating water column WEC with a characteristic mass of 1800 tonnes and a characteristic external surface of 6500 $m^2$ (Babarit, Hals, et al., 2021, p. 47). This WEC gave yearly mean values between 147 and 745 kW depending on location, with 262 kW being the value for EMEC (Babarit, Hals, et al., 2021, fig. 19).

Shao et al. (2023) study the same WEC as in this article, comparing two different versions of the device in regards to their power performance and the fatigue enacted upon the buoy's mooring lines. They use both regular harmonic waves with different periods and amplitudes, and irregular waves from three different sea states based on empirical wave data from the Runde test site (Shao et al., 2023, p. 450, 453). For the irregular waves, however, only the strain upon the mooring lines is studied, while the power output is primarily studied for the regular waves. The power is found to be greatest for high wave amplitudes (3 meters from still water line to crest) and wave periods of 4 or 5 seconds for the 3.0 and 4.0 versions of the WEC, respectively (Shao et al., 2023, p. 454, fig. 10) while the hydrodynamic efficiency of the device – how well it utilizes the energy in the waves which pass through it – is clearly highest for those same 4- and 5-second wave periods (around 40%, compared to 0-20% for most other periods), with smaller differences depending on wave amplitude (Shao et al., 2023, p. 454, fig. 11).

# Chapter 2

# Theoretical background

## 2.1 The dynamics of water waves

The waves which we observe on the ocean are generally known as *water waves* or surface waves (Encyclopaedia Britannica, 2023), to distinguish them from other types of waves such as the movement of sound, light and subatomic particles (Encyclopaedia Britannica, 2024). Water waves may be further categorized as either tsunamis, wind surges, *wind waves* or *swell*. The former two are connected to rare and extreme events such as seismic disturbances causing tsunamis or wind surges travelling in front of cyclones, while the latter two are more commonly occurring and both caused by the wind acting upon the ocean surface. Wind waves are the waves directly generated by the wind, while swell is their further gravitational propagation after the wind has abated (Encyclopaedia Britannica, 2023).

A fundamental demarcation of this thesis is that it is not concerned with the modelling of extreme wave events, which may instead be of interest in e.g. risk modelling for the construction of sea walls. Extreme waves are naturally characterized by occurring very rarely and are thus of little interest when the goal is to model the mean energy output of a WEC, though they are highly relevant to the study of its durability and maintenance costs (Falnes and Kurniawan, 2020, p. 1). Extreme waves affect the WEC buoy in completely different ways than linear theory can capture (ITTC, 2002, p. 507-508), and for this reason, tsunamis and wind surges are not of interest and this thesis will only deal with wind waves and swell.

Wind waves are a form of *gravity waves* (Encyclopaedia Britannica, 2023) in the sense that though they are initialized by wind, their further propagation is also caused by gravity trying to restore the disturbed equilibrium of the water surface (Lighthill, 1978, p. 205). In the same way, swell is also a type of gravity waves. The remainder of this subsection deals with the stochastic modelling of wind waves and swell as well as their physical dynamics and energy content.

### 2.1.1 Stationary stochastic processes and Fourier series

The most straightforward and regular way to model water waves is as one-dimensional, one-directional regular harmonic waves which are defined only by a wave height and a wave period. This kind of waves may easily be theoretically envisioned, and might be realistic to observe in a controlled environment (Chakrabarti, 2005, p. 1019-1020). Real ocean waves, however, particularly wind waves,

are stochastic and irregular due to the randomness of the winds that create them (Falnes and Kurniawan, 2020, p. 87). Therefore, a regular and deterministic process such as a simple harmonic wave is not an efficient way of capturing them, though it might be more successful in the case of swell. A more apt way of modelling real wind waves may be as stochastic processes, which may be considered stationary provided that there is reasonably constant wind over some time interval (G. Lindgren, Rootzén, and Sandsten, 2014, p. 88). Using such a model makes modelling much more convenient since it allows for random variation over time, while also having certain short-term consistencies such as time-invariant mean and autocovariance.

A very important representation of a stationary stochastic process is its *spectral density*, which is the inverse Fourier transform of the process's covariance function (provided that said function is integrable) while the covariance function, in turn, is the regular Fourier transform of the spectral density (G. Lindgren, Rootzén, and Sandsten, 2014, p. 82). Expressed mathematically,

$$r(\tau) = \int_{-\infty}^{\infty} e^{i2\pi f\tau} S(f) df \tag{2.1}$$

$$S(f) = \int_{-\infty}^{\infty} e^{-i2\pi f\tau} r(\tau) d\tau \tag{2.2}$$

where $S(f)$ is the spectral density for frequency $f$ and $r(\tau)$ the covariance function for time lag $\tau$ (G. Lindgren, Rootzén, and Sandsten, 2014, p. 82-85).

An alternate way of viewing the wind waves is as a superposition of multiple simple harmonic waves which may each have differing frequencies and (stochastic) phases and amplitudes (Falnes and Kurniawan, 2020, p. 87). In fact, as will be shown, this perspective does not contradict that of waves as stationary stochastic processes since certain such processes may themselves be seen as a superposition of several harmonic processes with random phase and amplitude.

A more complicated periodic function $f(x)$ may be expressed as a weighted sum of simpler periodic functions (sine and cosine functions) in what is known as a *Fourier series*:

$$f(x) = \sum_{n=0}^{\infty} \alpha_n \sin\left(\frac{2\pi n x}{a}\right) + \sum_{m=0}^{\infty} \beta_m \cos\left(\frac{2\pi m x}{a}\right) \tag{2.3}$$

where $a$ is the period of the functions and $\alpha_n, \beta_m$ are the *Fourier coefficients* (Chong, 2021, p. 77). The Fourier series expression for a periodic function is always valid provided that the function is square-integrable, meaning

$$\int_{-a/2}^{a/2} |f(x)|^2 dx \tag{2.4}$$

has to exist and be finite (Chong, 2021, p. 78). Now, uniting the two summations and using Euler's formula, eq. (2.3) may be rewritten to

$$f(x) = \sum_{n=-\infty}^{\infty} e^{i2\pi n x/a} f_n \tag{2.5}$$

where $f_n$ are new, complex Fourier coefficients (Chong, 2021, p. 78). We immediately see how similar this is to a Fourier transform from spectral density to covariance function, becoming identical if we replace $1/a$ with $f\tau$, $f_n$ with $S(f)$ and the summation with an integral. In fact, the Fourier transform is a version of the Fourier series which is valid for functions defined over the entire real line, not just the interval $[-a/2, a/2]$ (Chong, 2021, p. 80). $f(x)$ is then the covariance function $r(\tau)$, which itself is of a periodic nature.

The results above indicate that besides rewriting eq. (2.3) to be an alternate expression for the (deterministic) covariance function, we should also be able to rewrite it to be an expression for

the stochastic process itself. We may introduce stochasticity both through the Fourier coefficients, corresponding to a random amplitude of the sine and cosine functions, and through the sine and cosine arguments with either random frequencies or random phases. Such an approach is used by Zeinali et al. (forthcoming, p. 4), using only the cosine terms while letting the frequency be deterministic and the amplitude and phase random, and having the amplitude also depend on a spectral density. Since it is not possible in practice to have an infinite number of superimposed processes, any empirical realisation of such a superimposition must necessarily be an approximation.

In summation, for a limited time period during which stationarity may be assumed, it seems reasonable to model random wind waves as a periodic stationary stochastic process which may be expressed as the sum of multiple random sine and cosine functions.

### 2.1.2 Sea states and the Bretschneider spectrum

Outside of a limited time period, the behaviour of wind waves may differ significantly in a way that violates the assumption of stationarity and makes a stationary stochastic process unfit to model the wave height. One way to handle this problem is to use *sea states*, a way of classifying the conditions of the sea and its waves so that the sea may shift from one state to another over time and in that way inform the behaviour of the stochastic process. Stationarity may then be assumed during the time that the sea remains in the same state.

A classical description of sea states is the Beaufort scale, which measures wind speeds and corresponding sea states on a scale from 0 to 12, defining the states by wave heights (Service, 2024). However, height is not the only measure of waves. The Sea State Code developed by the World Meteorological Organization uses both the wave height and the length of the *wave energy periods* to assign a state from 0 to 9 (Industry, 2023). Since wave height naturally varies throughout the trajectory of the wave, a common measure is the *significant wave height* ($H_s$), defined as "the average height of the highest one-third of the waves within a wave group or record, measured from the trough to the crest" (Industry, 2023). The energy period ($T_e$), on the other hand, is described by Mollison (1994, p. 207) as "the mean wave period with respect to the spectral distribution of energy". This thesis will use these two measures, significant wave height and wave energy period, as state-defining parameters, but will use a greater number of states with cross-combinations of the two. The combinations used will be taken from the sea states defined by European Marine Energy Centre (2022a), describing the wave conditions at the Billia Croo test site on Orkney, a group of islands to the north-east of Scotland.

To describe the conditional wave behaviour for a certain sea state, the significant wave height and the wave period may be used as parameters of a wave spectral density $S(f)$ or $S(\omega)$, for the frequency $f$ or angular frequency $\omega$ (ITTC, 2002, p. 512) and with $\omega = 2\pi f$ (Wave Generation and Analysis, 1989, p. 795). This spectral density may then be used to generate random waves through Fourier sums. The spectrum also provides an alternate definition of the significant wave height to the one given above, namely by estimating it as $4\sqrt{m_0}$ where $m_0$ is the zeroth moment of the spectrum (Wave Generation and Analysis, 1989, p. 800). A common formulation of the general-case spectral density is

$$S(f) = \frac{A}{f^5} \exp(-B/f^4) \tag{2.6}$$

where $A$ and $B$ are constants (ITTC, 2002, p. 517). This is the form often used for the *Bretschneider spectrum*, also known as the *Generalized Pierson-Moskowitz spectrum*, which may more specifically be written as (Zeinali et al., forthcoming, p. 2, eq. 1)

$$S(\omega) = m_0 \frac{5}{\omega_p} \left( \frac{\omega}{\omega_p} \right)^{-5} e^{-\frac{5}{4}\left(\frac{\omega}{\omega_p}\right)^{-4}} \tag{2.7}$$

where $f = \omega/\omega_p$, $A = m_0\frac{5}{\omega_p}$, $B = 5/4$ and $\omega_p = 2\pi/T_p$. $\omega_p$ is the peak of the spectral density, corresponding to the most dominant frequency of the sea state (Zeinali et al., forthcoming, p. 2). Since the waves of the very highest frequencies oscillate so quickly that the mechanical system may barely react to them, it may also be valid to use a truncated version of this spectrum which sets the spectral density to zero above a certain threshold. Zeinali et al. (forthcoming, p. 2) set this cutoff point to five times the spectral peak $\omega_p$, preserving 99.8% of the spectral energy. The density then becomes (Zeinali et al., forthcoming, p. 2, eq. 1)

$$S(\omega) = \begin{cases} m_0\frac{5}{\omega_p}\left(\frac{\omega}{\omega_p}\right)^{-5} e^{-\frac{5}{4}\left(\frac{\omega}{\omega_p}\right)^{-4}}, & 0 \leq \omega \leq 5\omega_p \\ 0, & \omega > 5\omega_p. \end{cases} \tag{2.8}$$

The moments of the GPM spectrum, as written by Zeinali et al. (forthcoming, p. 2, eq. 2), are

$$m_n = \omega_p^n m_0 \left(\frac{5}{4}\right)^{n/4} \Gamma\left(1 - \frac{n}{4}\right) \tag{2.9}$$

$$m_n^{trunc} = \omega_p^n m_0 \left(\frac{5}{4}\right)^{n/4} \Gamma\left(1 - \frac{n}{4}, 500\right) \tag{2.10}$$

for the standard and truncated spectrum, respectively. $\Gamma(a)$ is the standard gamma function while $\Gamma(a, b)$ is the upper incomplete gamma function (Zeinali et al., forthcoming, p. 2-3).

When using the expression from (Zeinali et al., forthcoming, p. 3, eq. 3) for the energy period $T_e$, this gives

$$T_e = 2\pi\frac{m_{-1}}{m_0} \tag{2.11}$$

$$\begin{aligned} T_e &= 2\pi\frac{\omega_p^{-1} m_0 (5/4)^{-1/4}\Gamma(5/4)}{m_0} \\ &= \frac{2\pi}{\omega_p}(4/5)^{1/4}\Gamma(5/4) \end{aligned} \tag{2.12}$$

$$\begin{aligned} T_e &= \frac{2\pi}{\omega_p} \cdot 0.8572225 \\ &= 0.86 T_p \end{aligned} \tag{2.13}$$

where in the last equality we use that the peak period $T_p = 2\pi/\omega_p$ (Zeinali et al., forthcoming, p. 3).

### 2.1.3 Pressure and energy content of waves

In general, water waves are not propagated on a line or in a space but on a surface. They are therefore two-dimensional unlike, for example, guided electromagnetic waves along a cylinder (one-dimensional) or acoustic waves in a fluid (three-dimensional) (Falnes and Kurniawan, 2020, p. 46). Plane harmonic waves along the water surface have a dynamic pressure $p = p(x, z, t)$ dependent on the horizontal position along which the wave propagates ($x$), as well as the height $z$ and the time $t$. On a surface above deep water, this pressure has the complex amplitude

$$\hat{p} = \hat{p}(x, z) = A(z)e^{-ikx} + B(z)e^{ikx} \tag{2.14}$$

(Falnes and Kurniawan, 2020, p. 48, eq. 3.5). The *phase velocity* of water waves, meaning the rate of propagation of a single wave (Brillouin, 1960, p. 1), may be written as

$$\nu_p \equiv \frac{\omega}{k} = \frac{g}{\omega} = \sqrt{\frac{g}{k}} \tag{2.15}$$

(Falnes and Kurniawan, 2020, p. 48, eq. 3.9) where $g$ is the acceleration of gravity, $k = 2\pi/\lambda$ is the angular repetency and $\lambda$ is the wavelength (Falnes and Kurniawan, 2020, p. 47-48). Superimposing several waves gives a *group velocity* (Brillouin, 1960, p. 2) which is written as

$$\nu_g = \frac{d\omega}{dk} = -\frac{\partial F/\partial k}{\partial F/\partial \omega} \tag{2.16}$$

with $F(\omega, k) = 0$ being the dispersion relationship (Falnes and Kurniawan, 2020, p. 49, eq. 3.10). On deep water, the group velocity becomes $\nu_g = g/(2\omega) = \nu_p/2$ (Falnes and Kurniawan, 2020, p. 49, eq. 3.12).

Having defined the velocity, we move on to the *intensity* of the wave, meaning the "time-average energy transport per unit time and per unit area in the direction of the wave propagation" (Falnes and Kurniawan, 2020, p. 49). For harmonic waves, the intensity is

$$I = I_x = I_x(x, y, z) = \overline{p\nu_x} = \frac{1}{2}\,\mathrm{Re}\{\hat{p}\hat{\nu}_x^*\} \tag{2.17}$$

(Falnes and Kurniawan, 2020, p. 50, eq. 3.16) where $p$ and $\nu_x$ are the dynamic pressure and the $x$-directional component of the particle velocity. For deep water the intensity becomes $I_x = I_0 e^{2kz}$ with $I_0$ being the intensity at the surface level, $z = 0$ (Falnes and Kurniawan, 2020, p. 50).

Finally, we have the *wave-energy transport*

$$J = \int_{-\infty}^{0} I_x(z)dz = I_0 \int_{-\infty}^{0} e^{2kz} = \frac{I_0}{2k} \tag{2.18}$$

defined as the wave energy transported along a vertical strip parallel to the wave front, per unit of time (Falnes and Kurniawan, 2020, p. 50, eq. 3.19). For a superposition of $N$ random harmonic waves, the energy transport becomes

$$\begin{aligned} J &= \rho g \sum_{m=1}^{N} \frac{1}{2}|\eta_i(\omega_m)|^2 \nu_g(\omega_m) = \rho g \sum_{m=1}^{N} \Delta\omega S_\omega(\omega_m)\nu_g(\omega_m) \\ &= \rho g \int_0^\infty S_\omega(\omega)\nu_g(\omega)d\omega = \rho g \int_0^\infty S(f)\nu_g(f)df \end{aligned} \tag{2.19}$$

where $\eta$ is the wave elevation (Falnes and Kurniawan, 2020, p. 67, eq. 4.31). On deep water,

$$J = \frac{\rho g^2}{2} \int_0^\infty S_\omega(\omega)\omega^{-1}d\omega = \frac{\rho g^2}{4\pi} \int_0^\infty S(f)f^{-1}df \tag{2.20}$$

$$= \frac{\rho g^2}{64\pi} T_e H_s^2. \tag{2.21}$$

where $T_e$ is of course the wave energy period, here equivalently defined as

$$T_e = \frac{2\pi \int_0^\infty S_\omega(\omega)\omega^{-1}d\omega}{\int_0^\infty S_\omega(\omega)d\omega} = \frac{\int_0^\infty S(f)f^{-1}df}{\int_0^\infty S(f)df} \tag{2.22}$$

(Falnes and Kurniawan, 2020, p. 90-91, eq. 4.186-190) and $\rho$ above is the mass density of the fluid (Falnes and Kurniawan, 2020, p. 62). $H_s$ is, naturally, the significant wave height, here defined as

$$\overline{\eta^2(x, y, t)} = \int_0^\infty S(f)df = \frac{H_s^2}{16} \tag{2.23}$$

(Falnes and Kurniawan, 2020, p. 88, eq. 4.170), which agrees with our previous definition since $\int_0^\infty S(f)df = m_0$ is the zeroeth moment of the spectral density. The result in eq. (2.21) above is also

used by Zeinali et al. (forthcoming, p. 3, eq. 5), derived in an alternate way through the negative first moment of the GPM spectrum and eq. (2.11) for the energy period:

$$J = \frac{\rho g^2}{2} \int_0^\infty \omega^{-1} S_\omega(\omega) d\omega = \frac{\rho g^2}{2} m^{-1} \tag{2.24}$$

$$= \frac{\rho g^2}{2} \frac{m_0 T_e}{2\pi} = \frac{\rho g^2}{2} \frac{\left(\frac{H_s}{4}\right)^2 T_e}{2\pi} \tag{2.25}$$

$$= \frac{g^2 \rho}{64\pi} T_e H_s^2. \tag{2.26}$$

The unit of $J$ is $W/m$ (Falnes and Kurniawan, 2020, p. 82, eq. 4.132). With all the above, the energy content of a wave has been derived as a function of the gravity acceleration, the fluid's mass density, the wave energy period and the significant wave height. The first of these two parameters are constant given the geographical location, while the latter two may vary between sea states. However, all of this energy is not necessarily possible to absorb.

## 2.2 Wave-body interactions

Having discussed both the stochastic modelling and energy content of wind waves, we move on to the interaction between these waves and an immersed body such as a WEC buoy.

In general, a body immersed in water may oscillate in six different modes. Three of these are translational *along* a particular axis – surge (x), sway (y) and heave (z) – and three are rotational *around* an axis – roll (x), pitch (y) and yaw (z) (Falnes and Kurniawan, 2020, p. 121-122). The six-dimensional velocity vector $\mathbf{u} \equiv (u_1 u_2 u_3 u_4 u_5 u_6) \equiv (U_x, U_y, U_z, \Omega_x, \Omega_y, \Omega_z) = (\vec{U}, \vec{\Omega})$ (Falnes and Kurniawan, 2020, p. 122, eq. 5.3) has components corresponding to these modes, and with a corresponding normal vector $\mathbf{n}$ we may write the derivative of the wave's *velocity potential* $\phi$ – or its complex amplitude $\hat{\phi}$ – as

$$\frac{\partial \phi}{\partial n} = \mathbf{u}^T \mathbf{n} \tag{2.27}$$

$$\frac{\partial \hat{\phi}}{\partial n} = \hat{\mathbf{u}}^T \mathbf{n} \tag{2.28}$$

on the wet body surface (Falnes and Kurniawan, 2020, p. 122-123, eq. 5.7-8). A body's oscillation generates a superposition of radiated waves in the six modes:

$$\hat{\phi}_r = \sum_{j=1}^{6} \varphi_j \hat{u}_j \tag{2.29}$$

with $\varphi_j$ as a complex proportionality coefficient which satisfies

$$\frac{\partial \varphi_j}{\partial n} = n_j \tag{2.30}$$

(Falnes and Kurniawan, 2020, p. 123, eq. 5.9-10).

To define the force enacted upon a body by a wave, we define the force vector $\mathbf{F} \equiv (F_1, F_2, F_3, F_4, F_5, F_6) \equiv (F_x, F_y, F_z, M_x, M_y, M_z) = (\vec{F}, \vec{M})$ with the $F$ components being translational force and the $M$ components rotational force (Falnes and Kurniawan, 2020, p. 125, eq. 5.20). These components and their

complex amplitudes are

$$F_j = -\iint_S p n_j dS \tag{2.31}$$

$$\hat{F}_j = i\omega\rho \iint_S \hat{\phi} n_j dS \tag{2.32}$$

(Falnes and Kurniawan, 2020, p. 125, eq. 5.21-22) and give us the power received by the oscillating body:

$$P(t) = \vec{F}(t) \cdot \vec{U}(t) + \vec{M}(t) \cdot \vec{\Omega}(t) = \sum_{j=1}^{6} F_j u_j \tag{2.33}$$

(Falnes and Kurniawan, 2020, p. 126, eq. 5.24). If the body doesn't oscillate on its own and the velocity potential $\hat{\phi}$ is therefore only the result of an incident wave, the force vector $\mathbf{F}$ is equal to the excitation force vector $\mathbf{F}_e = (\vec{F}_e, \vec{M}_e)$ (Falnes and Kurniawan, 2020, p. 126, eq. 5.25), where

$$\hat{\mathbf{F}}_{e,j} = i\omega\rho \iint_S (\hat{\phi}_0 + \hat{\phi}_d) n_j dS, \tag{2.34}$$

$\hat{\phi}_0$ is the undisturbed incident wave and $\hat{\phi}_d$ the diffracted wave (Falnes and Kurniawan, 2020, p. 126, eq. 5.26).

If the body oscillates, however, specifically in the heave mode, two more forces act upon it: the radiation force $\mathbf{F}_r$ and $\mathbf{F}_b$ and the hydrostatic buoyancy force $\mathbf{F}_b$. Besides this, there may also be a viscous force $\mathbf{F}_v$, a control or load force $\mathbf{F}_u$ and a friction force $\mathbf{F}_f$ (Falnes and Kurniawan, 2020, p. 189-190). In linear theory, the buoyancy force is assumed to be proportional to the displacement of the body from its equilibrium position so that

$$\mathbf{F}_b = -\mathbf{S}_b \mathbf{s} \tag{2.35}$$

where the buoyancy stiffness matrix $\mathbf{S}_b$ is made up of proportionality constants (Falnes and Kurniawan, 2020, p. 189-190, eq. 5.337). In a similar way, the friction and viscosity forces are assumed to be proportional to the differentiated displacement,

$$\mathbf{F}_f = -\mathbf{S}_f \mathbf{u} = -\mathbf{S}_f \frac{d\mathbf{s}}{dt} \tag{2.36}$$

$$\mathbf{F}_\nu = -\mathbf{S}_\nu \mathbf{u} = -\mathbf{S}_\nu \frac{d\mathbf{s}}{dt} \tag{2.37}$$

(Falnes and Kurniawan, 2020, p. 190, eq. 340).

Considering only the heave motion, the matrices and vectors above become scalars, and the heave-mode buoyancy stiffness for small displacement is

$$S_b = \rho g S_w \tag{2.38}$$

where $S_w$ is the area of the submerged body at the point where it crosses the water surface (in equilibirum) (Falnes and Kurniawan, 2020, p. 190).

## 2.3 Oscillating Water Column type WEC:s

Wave Energy Converters often involve some kind of oscillating body in water, but over the years many different specific types of WEC's have been conceptualized. One of the most efficient is the

Oscillating Water Column WEC (OWC), originally envisioned by Yoshio Masuda in the 1940's (Alves, 2016, p. 28). Physically, the classical version of this device is built around a chamber which contains both water and air, the water flowing into the WEC directly from the surrounding sea and the air connected to the outside air through a turbine. As the water level rises and falls, the air pressure in the chamber changes, which then causes air to flow either in or out, passing the turbine which transfers energy to a generator (Alves, 2016, p. 27-28).

If we let $\phi$ be the velocity potential of the wave after meeting with the WEC. This velocity potential can be decomposed as

$$\phi = \phi_0 + \phi_d + \phi_r \tag{2.39}$$

where the terms represent the velocity potentials from the incident wave ($\phi_0$), the diffracted wave ($\phi_d$) and the radiated wave ($\phi_r$) (Falnes and Kurniawan, 2020, p. 243, eq. 7.7). Correspondingly, the volume flow of the water surface oscillating within the WEC may be written and decomposed as

$$\begin{aligned} Q_{t,k} &= \iint_{S_k} \nu_z dS = \iint_{S_k} \frac{\partial \phi}{\partial z} dS \\ &= Q_{e,k} + Q_{r,k} \end{aligned} \tag{2.40}$$

$$Q_{e,k} = \iint_{S_k} \frac{\partial}{\partial z}(\phi_0 + \phi_d) dS \tag{2.41}$$

$$Q_{r,k} = \iint_{S_k} \frac{\partial \phi_r}{\partial z} dS \tag{2.42}$$

where the two terms $Q_{e,k}$ and $Q_{r,k}$ represent excitation and radiation volume flow, and $\nu_z = \frac{\partial \phi}{\partial z}$ is the velocity (Falnes and Kurniawan, 2020, p. 244, eq. 7.15-18).

We now get to the absorbed power of the OWC type WEC. This may be written as the difference between excitation and radiated power (Falnes and Kurniawan, 2020, p. 245-246, eq. 7.25-27):

$$P = P_e - P_r \tag{2.43}$$

$$P_e = \frac{1}{2}\mathrm{Re}\{\hat{p}_k \hat{Q}^*_{e,k}\} = \frac{1}{2}\mathrm{Re}\{\hat{p}_k q^*_{e,k} A^*\} \tag{2.44}$$

$$P_r = \frac{1}{2} G_{kk} |\hat{p}_k|^2 \tag{2.45}$$

where $G_{kk} = \mathrm{Re}\{Y_{kk}\} = \frac{1}{2}(Y_{kk} + Y^*_{kk})$ is the radiation conductance (Falnes and Kurniawan, 2020, p. 245, eq. 7.22), $p_k$ is the air pressure fluctuation (Falnes and Kurniawan, 2020, p. 243), * signifies complex conjugate (Falnes and Kurniawan, 2020, p. 13) and $A$ is "the complex amplitude of the incident wave's elevation at the origin" (Falnes and Kurniawan, 2020, p. 208).

## 2.4 Hydraulic power take off

After the power of the wave has been absorbed into the WEC, it needs to somehow be converted into electrical energy through a power take off (PTO) system. The "most suitable device for generating useable electricity from wave energy" (Ahamed, McKee, and Howard, 2020, p. 3) is the hydraulic motor-based system. This system typically works through a hydraulic cylinder which is set in motion by the waves, and in turn changes the pressure in a working medium such as hydraulic oil. This hydraulic motor can take both translational and rotational energy as input, and in turn drives a generator which outputs electrical energy (Ahamed, McKee, and Howard, 2020, p. 3).

The pressure in the hydraulic system is the pressure in its accumulator tank, which may prevent energy output either if it's too high (in which case the pressure may counteract and cancel out the force of the incident wave) or if it's too low (which may lead to a very low PTO force) (Babarit, Guglielmi, and Clément, 2009, p. 1015). Given a certain wave input, there should be a causal relation between the hydraulic pressure and the rotation speed of the connected generator, since the wave input will determine the medium flow into the accumulator tank while the rotation speed will determine the flow out of it. Since this thesis aims to optimize the generator speed, the hydraulic pressure will not be controlled but simply be allowed to fluctuate as the wave input varies but the rotation speed remains constant. Optimization of the hydraulic pressure has been studied in detail by Zeinali et al. (forthcoming), and the optimal values found there inform the design of the accumulator tank (Zeinali et al., forthcoming, p. 2).

The hydraulic energy converted to electrical energy through the generator is measured in kW, meaning that in order to compare it with the energy potential $J$ from eq. (2.26), we must multiply $J$ both by $10^{-3}$ (to convert from W/h to kW/h) and by the diameter of the WEC at the water surface level.

## 2.5 Monte Carlo integration

With the theory above, it becomes feasible to both generate random waves as stationary stochastic processes and to convert the resulting wave oscillations to electrical energy through transformations in a simulated WEC. It should also be possible to calculate expected values for, for instance, the wave amplitude explicitly given a particular sea state. However, the expected electrical output would, due to the very complex low-pass filtering transformation in the WEC simulator, be all but impossible to calculate analytically. It is for this reason the thesis will instead use Monte Carlo methods, or more specifically Monte Carlo integration, the history of which was presented in section 1.2.2.

The basic premise of Monte Carlo is "the statistical estimation of the value of an integral using evaluations of an integrand at a set of points drawn randomly from a distribution with support over the range of integration" (Givens and Hoeting, 2013, p. 151). This is relevant to our problem since the expected value or variance of any continuous random variable (such as the electrical energy output of a WEC) may be defined as an integral (Gut, 2009, p. 7, eq. 5.1).

If we let $\mathbf{X}$ be a continuous random variable of arbitrary dimension, and define $Y = h(\mathbf{X})$ as a deterministic function of $\mathbf{X}$, we may then express the expectation of $Y$ as

$$\mu_Y = \mathbb{E}[Y] = \mathbb{E}[h(\mathbf{X})] = \int h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \tag{2.46}$$

where $f(\mathbf{x})$ is the probability density function of $\mathbf{X}$. Monte Carlo integration is required when $h(\mathbf{X})$ is such that this integral is not analytically solvable, at least not practically solvable due to the extreme complexity. This method generally uses the distribution of $\mathbf{X}$ to generate $N$ independent random samples $\mathbf{X}_i$, which are then entered into the function $h$. The sample mean $\hat{\mu}_{MC}$ is then used as an estimate for $\mu$. Basic probability theory gives the expected value and variance of $\hat{\mu}_{MC}$:

$$\mathbb{E}[\hat{\mu}_{MC}] = \mathbb{E}\big[\frac{1}{N}\sum_{i=1}^{N} h(\mathbf{X}_i)\big] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[h(\mathbf{X}_i)] = \frac{1}{N}N\mathbb{E}[h(\mathbf{X})] = \mu \tag{2.47}$$

$$\mathbb{V}[\hat{\mu}_{MC}] = \mathbb{V}\big[\frac{1}{N}\sum_{i=1}^{N} h(\mathbf{X}_i)\big] = \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{V}[h(\mathbf{X}_i)] = \frac{N}{N^2}\mathbb{V}[h(\mathbf{X})] = \frac{1}{N}\mathbb{V}[h(\mathbf{X})] \tag{2.48}$$

where in the variance calculation we use that the samples are independent and identically distributed. What the calculations above show is that on average, $\hat{\mu}_{MC}$ will estimate $\mu$ correctly – it is an unbiased

estimator – and its variance will decrease linearly as the sample size $N$ increases. In other words, the estimator is consistent.

While we know that the variance of this estimate will decrease with $N$, the actual value of the variance $\mathbb{V}[h(\mathbf{X})] = \sigma^2$ is unknown to us. A possible estimator of this variance would be the sample variance

$$\hat{\sigma}^2_{MC} = \frac{1}{N-1}\sum_{i=1}^{N}(h(\mathbf{X}_i) - \hat{\mu}_{MC})^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} h^2(\mathbf{X}_i) - \frac{1}{N}\big(\sum_{i=1}^{N} h(\mathbf{X}_i)\big)^2\right) \tag{2.49}$$

which is an unbiased estimator of $\sigma^2$ (Blom et al., 2017, p. 251, thm. 11.2). Furthermore, the central limit theorem gives that the estimator $\hat{\mu}_{MC}$ will converge (in the distribution sense) to a Gaussian distribution $\mathcal{N}(\mu, \sigma^2/N)$ and therefore the transformation $\frac{\hat{\mu}_{MC}-\mu}{\sigma/\sqrt{N}}$ converges to a standard Gaussian distribution (Gut, 2009, p. 162-164). This means that we could, given a large enough sample size, construct a confidence interval of degree $1-\alpha$ using the quantiles $\phi$ of the standard Gaussian distribution:

$$\begin{aligned} CI_\mu &= (CI_{lower}, CI_{upper}) \\ &= \left((\hat{\mu}_{MC} + \phi_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2_{MC}}{N}}), (\hat{\mu}_{MC} + \phi_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2_{MC}}{N}})\right) \end{aligned} \tag{2.50}$$

where $\mathbb{P}(Z > \phi_{\alpha/2}) = \alpha/2$ ($Z$ being standard Gaussian) (Blom et al., 2017, p. 397) and $\mathbb{P}(CI_{lower} < \mu < CI_{upper}) = 1-\alpha$ (Blom et al., 2017, p. 288).

Besides being used to make an approximate confidence interval for the mean estimate, $\hat{\sigma}^2_{MC}$ being an unbiased estimator of $\sigma^2 = \mathbb{V}[h(\mathbf{X})]$ means that we may use it as a point estimate if the variance itself is the parameter of interest. If the variable $h(\mathbf{X})$ is assumed to be Gaussian (not just approximately as $N$ increases), then we know the distribution of $\hat{\sigma}^2_{MC}$ to be $\chi^2(N-1)$, which may then be used to construct an asymmetric confidence interval for $\sigma = \sqrt{\sigma^2}$ (Blom et al., 2017, p. 296). This interval is written (Blom et al., 2017, p. 297)

$$\begin{aligned} CI_\sigma &= \big(k_1\sqrt{\hat{\sigma}^2_{MC}}, k_2\sqrt{\hat{\sigma}^2_{MC}}\big) \\ k_1 &= \sqrt{f/\chi^2_{\alpha/2}(f)} \\ k_2 &= \sqrt{f/\chi^2_{1-\alpha/2}(f)} \end{aligned} \tag{2.51}$$

where $f = N-1$ are the degrees of freedom. Squaring, we get the corresponding confidence interval for the variance:

$$\begin{aligned} CI_{\sigma^2} &= \big(k_1\hat{\sigma}^2_{MC}, k_2\hat{\sigma}^2_{MC}\big) \\ k_1 &= f/\chi^2_{\alpha/2}(f) \\ k_2 &= f/\chi^2_{1-\alpha/2}(f). \end{aligned} \tag{2.52}$$

which agrees with the results in (Blom et al., 2017, p. 296). In both these intervals, we have that $k_1 > k_2$ since the $\chi^2$ quantile is in the denominator, and therefore we have that $k_1$ gives the upper interval limit when multiplied with either $\hat{\sigma}_{MC}$ or $\hat{\sigma}^2_{MC}$, and $k_2$ gives the lower.

If $h(\mathbf{X})$ is not Gaussian, one may use a version of the central limit theorem for the sample variance which is proven by Hannig (2008, p. 44, thm. 5.3.2). The statement is that the asymptotic distribution of the sample variance $s^2$ is Gaussian as the sample size $N$ goes to infinity,

$$s^2 \overset{\text{asymp.}}{\sim} \mathcal{N}(\sigma^2, \sigma^4(\kappa-1)/N), \tag{2.53}$$

where $\kappa = \mu_4/\sigma^4$ and $\mu_4$ is the centralized fourth moment of $h(\mathbf{X})$, i.e.

$$\mu_4 = \mathbb{E}\big[(h(\mathbf{X}) - \mu)^4\big]. \tag{2.54}$$

This agrees with the fact that even for a Gaussian random variable $X$ which has $\chi^2$-distributed sample variance, since the sample variance is a random variable the mean of multiple such variables should be asymptotically Gaussian as the sample size increases, in accordance with the general central limit theorem. For a large enough sample, we should therefore be able to construct confidence intervals for the variance estimate $\hat{\sigma}^2_{MC}$ using either the Gaussian or Chi-square distribution, depending on which distribution the observed variances seem to agree best with. For a sample which is not small enough, however, neither type of confidence interval may be applicable.

It is possible, given a sample from a random variable, to test an empirical distribution's adherence to a normal distribution. Razali and Wah (2011) investigate the power of four different formal tests of normality through Monte Carlo simulation. They find that the Shapiro-Wilk test performs better than the others (Kolmogorov-Smirnov, Anderson-Darling and Lilliefors tests), reaching power 1 (for significance level $\alpha = 5\%$) at 200-300 observations when the true distribution is asymmetrical, and power 0.9937 for 1000 observations when the true distribution is Student's $t$ with seven degrees of freedom (Razali and Wah, 2011, p. 28-30).

The Shapiro-Wilk test uses the test statistic

$$W = \frac{\left(\sum_{i=1}^{N} a_i y_i\right)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{b^2}{S^2} \tag{2.55}$$

where $y_i$ are the ordered random observations (order statistics) and $a' = (a_1, \ldots, a_N) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}$, $m'$ being the vector of expected values for the $N$ order statistics and $V$ being their covariance matrix. If the sample is from a normal distribution, $b^2$ and $S^2$ are both estimates of the variance $\sigma^2$, whereas for non-normal distributions they estimate different things (Shapiro and Wilk, 1965, p. 592-593). The statistic has a value between zero and one, and the null hypothesis of normality is rejected for small values (Razali and Wah, 2011, p. 25).

### 2.5.1 Ergodicity

When estimating a measure of a stochastic process, such as the mean, one option is to simulate multiple iterations of the process, calculate the sample mean of each one and use the average of averages as an estimator for the true mean of the process. Since each iteration of the process should unconditionally have an equal expected trajectory, this would lead to an unbiased estimate (provided that each process is itself unbiased) and increasing the number of iterations would increase the estimate's accuracy – thereby making the estimator consistent. Corresponding approaches should also be valid for other measures such as the covariance structure of the process, though there is no guarantee that one can find a consistent estimator for every conceivable measure.

Under certain conditions it would also be an option to replace the multiple iterations with a single iteration of much greater length. The mean of this process is an unbiased estimator of the mean, but might not increase in accuracy as the length of the process increases.

If for a stochastic variable $X$ the expected value, here called the *ensemble average*, is defined as

$$\mu = \mathbb{E}[X] \tag{2.56}$$

then the law of large numbers tells us that as long as $X$ has finite expectation, the average of $N$

independent samples of $X$ will converge (in probability) to $\mu$ as the sample size increases,

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} x_i \xrightarrow{P} \mu, \text{ as } N \to \infty \tag{2.57}$$

(G. Lindgren, Rootzén, and Sandsten, 2014, p. 47, eq. 2.15; Gut, 2009, p. 149, ex. 1.3). If this is expanded to an entire stochastic process $X(t)$ sampled $N$ times, then the expected value of the process at each time $t$ may be estimated by the mean of the $N$ processes' values at that time, that is,

$$\hat{\mu}_N(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) \xrightarrow{P} \mu(t), \text{ when } N \to \infty \tag{2.58}$$

and if the process is stationary then the means are equally $\mu(1) = \mu(2) = \cdots = \mu(t) = \cdots = \mu$ (G. Lindgren, Rootzén, and Sandsten, 2014, p. 48). This means that an arbitrary observation at any time $t$, or an average of $N$ such observations at identical or different points in time, is an unbiased estimator of $\mu$ since

$$\mathbb{E}\big[\frac{1}{N}\big(x(t_1) + \cdots + x(t_N)\big)\big] = \frac{1}{N}\big(\mathbb{E}[x(t_1)] + \cdots + \mathbb{E}[x(t_N)]\big) = \frac{1}{N}(\mu + \cdots + \mu) = \frac{1}{N} \cdot N \cdot \mu = \mu \tag{2.59}$$

and if all observations used are also from pairwise independent realisations of the stochastic process (but from arbitrary times $t$ for each observation) then the average will also be more accurate for larger $N$ and therefore a consistent estimator.

However, if the observations are taken from the same realisation of the stochastic process then their mean

$$\hat{\mu}_T = \frac{1}{T}\sum_{t=1}^{T} x(t) \tag{2.60}$$

is not sure to be a consistent estimator. If it is, however, the process is said to be *ergodic*, specifically *linearly ergodic* (G. Lindgren, Rootzén, and Sandsten, 2014, p. 48). The definition may be expanded to *second-order ergodicity* if the process's covariance function $r(\tau) = \mathbb{E}[(X(t) - \mu)(X(t+\tau) - \mu)]$ may also be estimated analogously, and *complete ergodicity* if this type of consistent estimation from a single realisation is possible for the expected value of any arbitrary function $g(X_{t_1}, \ldots, X_{t_p})$ (G. Lindgren, Rootzén, and Sandsten, 2014, p. 53). An intuitive explanation of the concept is that all possible outcomes of a process will eventually occur, with the respective probabilities the outcomes would have for independent realisations, as the length of the single realisation approaches infinity (G. Lindgren, Rootzén, and Sandsten, 2014, p. 48).

For a stationary stochastic process, G. Lindgren (2013, p. 67, cor. 2.2) states that the time-mean will converge in quadratic mean to the stationary mean of the process if the average covariance function over lags $\tau$ goes towards zero with increased process lengths. Mathematically,

$$\frac{1}{T}\int_0^T r(\tau)d\tau \to 0, \text{ as } T \to \infty \tag{2.61}$$

implies

$$\frac{1}{T}\int_0^T x(t)dt \xrightarrow{q.m.} \mu. \tag{2.62}$$

Lindgren states this only for $\mu = 0$, but adding a constant to an entire process should not have any effect on its ergodicity.

## 2.6 Regression

### 2.6.1 Linear regression

Regression analysis is described by Olive (2017, p. 17, def. 2.2) as "the study of the conditional distribution $Y|\mathbf{x}$ of the response variable $Y$ given the $p \times 1$ vector of predictors $\mathbf{x} = (x_1, \ldots, x_p)^T$". A response variable $Y$ is assumed to be a function of $p$ predictor variables $X_1, X_2, \ldots, X_p$, in the case of (multiple) linear regression simply as a linear combination,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{2.63}$$

where the coefficients $\beta$ are fixed but unknown and $\epsilon$ is a random error (Olive, 2017, p. 18). We choose to redefine this slightly so that the intercept $\beta_0$ corresponding to $X_0 = 1$ is included in the $p$ parameters, and thus the vector of $p$ coefficients is $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_{p-1}]$.

If the values of the predictor variables and the coefficients are known, the conditional expected value of the response variable may be calculated as

$$\hat{Y}|X_1 = x_1, \ldots, X_{p-1} = x_{p-1} = \mathbb{E}[Y|X_1 = x_1, \ldots, X_{p-1} = x_{p-1}] = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1} \tag{2.64}$$

since the errors are assumed to have expected value zero (Olive, 2017, p. 18). In the *constant variance* multiple linear regression model, other central assumptions are that when multiple observations are made, the conditional variances of the error (and therefore of the observed $Y$ values) are equal, their conditional covariances are pairwise zero and they are independent of the predictor variables $X$ (Olive, 2017, p. 18). Further assumptions may also be made, such as that the errors are from a unimodal distribution without strong skewness, or even that the errors are i.i.d. Gaussian $\mathcal{N}(0, \sigma^2)$ (Olive, 2017, p. 19).

It is well known that the least squares estimate of the matrix $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.65}$$

and that predicted $y$ values are then expressed as

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.66}$$

(Sheather, 2009, p. 132). The relation between the observed values $\mathbf{y}$ and the predicted values $\hat{\mathbf{y}}$ is called the *hat matrix*, $\mathbf{H}$:

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{H}\mathbf{y} \end{aligned} \tag{2.67}$$

$$\mathbf{H} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{2.68}$$

(Olive, 2017, p. 19).

Using the previous assumptions on the variance and covariance of the residuals, their covariance matrix may be written as $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the n-by-n-dimensional identity matrix (Sheather, 2009, p. 134). If this is the case, and if also the residuals are Gaussian, it is easy to construct confidence intervals for the $\beta$ estimates as well as prediction intervals for unobserved values of $y$.

The $\boldsymbol{\beta}$ estimator above has the following expected value and variance:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} \tag{2.69}$$

$$\mathbb{V}[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{2.70}$$

where the error variance $\sigma^2$ may be estimated as

$$MSE = s_e^2 = \frac{SSE}{N-p} = \frac{1}{N-p}\sum_{i=1}^{N} e_i^2 \tag{2.71}$$

(Sheather, 2009, p. 135). If the model assumptions are fulfilled, one may then construct confidence intervals for the coefficients:

$$CI_{\beta_j} = [\hat{\beta}_j + t_{N-p}(\alpha/2) \cdot se(\hat{\beta}_j), \hat{\beta}_j + t_{N-p}(1-\alpha/2) \cdot se(\hat{\beta}_j)] \tag{2.72}$$

where $t_{N-p}(a)$ is the $a$:th quantile of the Student's $t$ distribution with $N-p$ degrees of freedom and standard error $se(\hat{\beta}_j)$ is the $j$:th diagonal element of $s_e^2(\mathbf{X}'\mathbf{X})^{-1}$. This confidence interval follows from the fact that

$$T_j = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{N-p} \tag{2.73}$$

(Sheather, 2009, p. 135) which may then also be used to test whether a particular coefficient is significantly different from zero by comparing $T_j$ to relevant $t$ distribution quantiles. Other tests such as full or partial $F$ tests may be carried out using the regression, residual and total sums of squares,

$$\begin{aligned} SSR &= \sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2 \\ SSE &= \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \\ SST &= \sum_{i=1}^{N}(Y_i - \bar{Y})^2 = SSR + SSE \end{aligned} \tag{2.74}$$

to check whether the model as a whole is significant or whether one model is significantly better than another (Olive, 2017, p. 29-30 and 45-46)(Sheather, 2009, p. 136). From these sums of squares we may derive the coefficient of determination $R^2$, which is between 0 and 1 (Olive, 2017, p. 31, def. 2.15). This coefficient may be interpreted as the proportion of the variation in $Y$ which is explained by the regression model, but Olive (2017, p. 31) points out, among other warnings, that the coefficient is not meaningful unless the residual diagnostics look good and that small $N$ generally lead to overestimation of $R^2$.

It is also important to note that generally, $R^2$ increases as more predictors are added to the model, even if the contribution of the added predictors is very small and they are functionally irrelevant. In the interest of not having the model be unnecessarily complex, one may replace $R^2$ by an adjusted version, which is defined in (Sheather, 2009, p. 137) as

$$R^2_{adj} = 1 - \frac{SSE/(N-p)}{SST/(N-1)} \tag{2.75}$$

and is more appropriate to use in comparisons between models with different numbers of parameters.

### 2.6.2 Residual diagnostics

It is necessary to investigate whether the model assumptions are fulfilled by the data set before drawing conclusions based on any of the model's associated tests. For instance, Olive (2017, p. 21,

rem. 2.1) stresses the importance of making scatterplots of the predicted values $\hat{\mathbf{y}}$ against the observed values $\mathbf{y}$ as well as against the residuals $\mathbf{e}$. The observed values should be scattered randomly above and below the fitted identity line with no clear patterns or large outliers and the residuals should show the same kind of behaviour around zero. These plots could show non-linear patterns in the data, differing residual variance for different levels of $\hat{y}$ (heteroscedasticity) or large outliers (Olive, 2017, p. 22-24), which all in some way imply violations of the model assumptions. One may also check the Gaussianity of the residuals through a normal probability plot which compares the ordered residuals to the expected values of order statistics from a normal distribution (Olive, 2017, p. 29). Since an important assumption is that the errors are also uncorrelated with the predictor variables, the residuals should also be plotted against these.

The residuals used in these plots could be the crude residuals $\mathbf{e} = \hat{\mathbf{y}} - \mathbf{y}$, but it is generally more informative to use either the *standardized residuals*,

$$e_{i,std} = \frac{e_i}{\sqrt{s_e^2}} \tag{2.76}$$

or the (internally) *studentized residuals* (Olive, 2017, p. 130),

$$e_{i,stud} = \frac{e_i}{\sqrt{s_e^2(1-h_{ii})}}, \tag{2.77}$$

where $s_e^2$ is the estimated variance of the residuals (as per eq. (2.71)) and $h_{ii}$ is the i:th diagonal element of the hat matrix $\mathbf{H}$ (as per eq. (2.68)). Both types of transformations divide the crude residuals by an estimate of their standard error, where the studentized residual uses a more accurate estimate since the covariance matrix of the residual vector is

$$\begin{aligned} \mathbb{V}[\mathbf{e}] &= \mathbb{V}[\mathbf{y} - \hat{\mathbf{y}}] = \mathbb{V}[\mathbf{y} - \mathbf{H}\mathbf{y}] \\ &= \mathbb{V}[\mathbf{y}] + \mathbf{H}'\mathbf{H}\mathbb{V}[\mathbf{y}] - 2\mathbf{H}\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{y}](\mathbf{I} - 2\mathbf{H} + \mathbf{H}'\mathbf{H}) \\ &= \mathbb{V}[\mathbf{y}](\mathbf{I} - \mathbf{H}) = \sigma^2(\mathbf{I} - \mathbf{H}) \end{aligned} \tag{2.78}$$

which follows from the fact that

$$\begin{aligned} \mathbf{H}'\mathbf{H} &= \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{H}. \end{aligned} \tag{2.79}$$

This covariance matrix has the i:th diagonal element

$$\mathbb{V}[e_i] = \sigma^2(1 - h_{ii}) \tag{2.80}$$

which is the denominator used, with $\hat{\sigma^2} = MSE$, in the expression for the studentized residuals.

Another potential problem which may be discovered upon residual analysis is that the (studentized) residuals $e_{i,stud}$ are not entirely Gaussian. Since the tests mentioned above for the significance of the $\beta$ coefficients – individually ($t$ test), all together ($F$ test) or a subset (partial $F$ test) – are based on probability distributions derived from the normal distribution (Soch, 2024a; Soch, 2024b), the tests' validity depends on the Gaussianity of the errors $\epsilon_i$. non-Gaussian residuals imply non-Gaussian random errors and therefore invalidate the tests. Other diagnostics such as the coefficient of determination and its adjusted variant should still be useful, howeve, as long as the other model assumptions are fulfilled.

### 2.6.3 Handling violations of the model assumptions

If all investigations of the residuals show good agreement with the model assumptions, no further adjustments are needed and the regression model's associated tests, coefficient of determination,

and predictions are valid. If the assumptions are violated, however, there are different strategies available for handling these violations.

When there are clear signs of heteroscedasticity in the residuals, meaning that different predictor values lead to different variation in the response variable, Sheather (2009, p. 71) suggests either transformations or the weighted least squares method for handling this problem. The transformation method applies a transformation $g$ to the predictors, the response variable, or both. The regression model is then one of the following:

$$g(\mathbf{y}) = \boldsymbol{\beta}\mathbf{X} + \mathbf{e} \tag{2.81}$$

$$\mathbf{y} = \boldsymbol{\beta} \times g(\mathbf{X}) + \mathbf{e} \tag{2.82}$$

$$g(\mathbf{y}) = \boldsymbol{\beta} \times g(\mathbf{X}) + \mathbf{e} \tag{2.83}$$

with two possible transformations being the square root and the logarithm (Sheather, 2009, p. 77-79). Transformations may also be useful to handle non-linearity in the relation between the predictors and the response variable (Sheather, 2009, p. 83).

Weighted least squares regression is another option for handling heteroscedasticity. Here, we modify the model assumptions so that the errors $\epsilon_i$ are still pairwise independent and zero-mean, but have differing variances $\sigma_i^2 = \sigma^2/w_i$ (Sheather, 2009, p. 115). This then also means that $w_i = \sigma^2/\sigma_i^2$, or $1/\sigma_i^2$ if we assume the "standard" variance to be unit. The corresponding $\beta$ coefficients are then found by minimizing the weighted sum of squares:

$$WSSE = \sum_{i=1}^{N} w_i(y_i - \hat{y}_{i,w})^2 = \sum_{i=1}^{N} w_i(y_i - \beta_0 - \beta_1 x_{1,i} - \cdots - \beta_p x_{p,i})^2 \tag{2.84}$$

(Sheather, 2009, p. 115). In matrix form, this becomes

$$\begin{aligned} WSSE &= (\mathbf{y} - \hat{\mathbf{y}}_w)'(\mathbf{y} - \hat{\mathbf{y}}_w) \\ &= (\mathbf{y} - \boldsymbol{\beta}\mathbf{X})'\mathbf{W}(\mathbf{y} - \boldsymbol{\beta}\mathbf{X}) \end{aligned} \tag{2.85}$$

where $\mathbf{W}$ is an N-by-N matrix with the weights $w_i$ on the diagonal and zeros everywhere else. The solution $\hat{\boldsymbol{\beta}}$ that minimizes this sum of squares is

$$\hat{\boldsymbol{\beta}}_{WLS} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \tag{2.86}$$

(Olive, 2017, p. 165, def. 4.8). However, Olive (2017, p. 165) also states that this approach requires the weights to be fully known, and suggests the feasible weighted least squares (FWLS) method as an approach if the weights are unknown, replacing the $w_i$ in the weight matrix by estimates $\hat{w}_i = w_i(\hat{\boldsymbol{\theta}})$, $\boldsymbol{\theta}$ being some unknown vector of parameters on which the weight estimates are based. Regardless, the estimated response values are

$$\hat{\mathbf{y}}_{WLS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{WLS} \tag{2.87}$$

$$\hat{\mathbf{y}}_{FWLS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{FWLS} \tag{2.88}$$

(Olive, 2017, p. 165). This means that we may also define a WLS hat matrix,

$$\mathbf{H}_{WLS} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W} \tag{2.89}$$

which may be used to studentize the residuals resulting from a WLS or FWLS fit. However, in this context standardization should not be done by dividing the residuals by their common variance, since this is not assumed to be equal. Instead, we standardize and studentize as follows:

$$e_{i,std} = \frac{e_i}{\sqrt{s_{e,i}^2}} \tag{2.90}$$

$$e_{i,stud} = \frac{e_i}{\sqrt{s_{e,i}^2(1 - h_{ii})}}, \tag{2.91}$$

where $s_{e,i}^2$ is the estimated variance for that particular observation. If the weights are defined as $w_i = 1/s_{e,i}^2$, this may also be written as

$$e_{i,std} = e_i\sqrt{w_i} \tag{2.92}$$

$$e_{i,stud} = \frac{e_i\sqrt{w_i}}{\sqrt{1-h_{ii}}}. \tag{2.93}$$

### 2.6.4 Logistic regression

Linear regression is fit for modelling continuous numerical variables, but if one wishes to model the outcome of a categorical variable it is more reasonable to use *logistic regression*. In its most common form, logistic regression is binary and used to find a classification rule for two categories (Hosmer, Lemeshow, and Sturdivant, 2013, p. 1). One may also use multinomial logistic regression when there are more than two categories (Hosmer, Lemeshow, and Sturdivant, 2013, p. 269), and ordinal logistic regression when the multiple categories have a distinct order (Hosmer, Lemeshow, and Sturdivant, 2013, p. 289).

Logistic regression, like its linear counterpart, uses the linear combination of several predictor variables $X_j$, each associated with a coefficient $\beta_j$. However, this linear combination is not the predicted value of the response variable $Y$, but rather a log-odds transformation of the predicted probability of $Y$ taking a certain value. For the binary logistic model with outcomes 0 and 1, the equation is (Hosmer, Lemeshow, and Sturdivant, 2013, p. 35, eq. 2.1)

$$\log\left(\frac{\mathbb{P}(Y_i = 1|\mathbf{x}_i)}{1-\mathbb{P}(Y_i = 1|\mathbf{x}_i)}\right) = \mathbf{x}_i\boldsymbol{\beta} \tag{2.94}$$

or equivalently (Hosmer, Lemeshow, and Sturdivant, 2013, p. 35, eq. 2.2),

$$\mathbb{P}(Y_i = 1|\mathbf{x}_i) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}. \tag{2.95}$$

In the ordinal logistic model, specifically the *continuation-ratio logistic model*, the $\beta$ coefficients associated with a predictor are the same for each response category, but the intercept terms $\zeta$ are different. For $q$ response categories and $p-1$ non-intercept predictors, we have

$$\begin{aligned}
\log\frac{\hat{\mathbb{P}}(Y_i = 1)}{\hat{\mathbb{P}}(Y_i > 1)} &= \zeta_1 - \beta_1 x_{1,i} - \cdots - \beta_{p-1}x_{p-1,i} = \zeta_1 - \mathbf{x}_i\boldsymbol{\beta} \\
\log\frac{\hat{\mathbb{P}}(Y_i \leq 2)}{\hat{\mathbb{P}}(Y_i > 2)} &= \zeta_2 - \mathbf{x}_i\boldsymbol{\beta} \\
&\cdots \\
\log\frac{\hat{\mathbb{P}}(Y_i \leq q-1)}{\hat{\mathbb{P}}(Y_i = q)} &= \zeta_{q-1} - \mathbf{x}_i\boldsymbol{\beta}
\end{aligned} \tag{2.96}$$

which requires special software to solve (Hosmer, Lemeshow, and Sturdivant, 2013, p. 291). After fitting the $\zeta$ and $\beta$ values, we may use the equations above to recursively find the predicted probabilities for outcomes $k = 1, \ldots, q$. We rewrite the probabilities in the numerator and denominator and take the exponential function of both sides of the equation to get

$$\frac{\hat{\mathbb{P}}(Y_i \leq k-1)}{\hat{\mathbb{P}}(Y_i > k-1)} = \frac{1-\sum_{j=k}^{q}\hat{\mathbb{P}}(Y_i = j)}{\sum_{j=k}^{q}\hat{\mathbb{P}}(Y_i = j)} = \exp(\hat{\zeta}_{k-1} - \mathbf{x}_i\boldsymbol{\beta}), \tag{2.97}$$

which may then be rewritten as

$$\begin{aligned}\sum_{j=k}^{q} \hat{\mathbb{P}}(Y_i = j) &= \frac{1}{1 + \exp(\hat{\zeta}_{k-1} - \mathbf{x}_i\boldsymbol{\beta})} \\ \hat{\mathbb{P}}(Y_i = k) &= \frac{1}{1 + \exp(\hat{\zeta}_{k-1} - \mathbf{x}_i\boldsymbol{\beta})} - \sum_{j=k+1}^{q} \hat{\mathbb{P}}(Y_i = j),\end{aligned} \tag{2.98}$$

meaning that we first calculate the predicted probability of the highest category $q$ and then continue downwards, eventually calculating the predicted probabilities of the lowest categories by subtracting the other categories' probabilities from 1 (A. Lindgren, 2023, p. 8).

The general assumption in logistic regression is not that the true value of $Y_i$ is equal to the linear combination $\mathbf{x}_i\boldsymbol{\beta}$ plus a Gaussian random error $\epsilon_i$, but rather that the categorical outcome of $Y_i$ is binomially or multinomially distributed with probabilities $\mathbb{P}(Y_i = k)$ depending on the predictor values for observation $i$ (Hosmer, Lemeshow, and Sturdivant, 2013, p. 7). For this reason, the same diagnostics are not valid, but one may instead define a *confusion matrix* which compares the frequencies at which each category is observed (on one axis) and predicted (on the other) (Ting, 2010). This gives an idea of the precision of the model at predicting particular outcomes, which is calculated as the number of *correctly* predicted instances of an outcome divided by the total number of predicted instances of that outcome, as well as its overall accuracy, meaning the overall ratio of correct predictions (Jayaswal, 2020).

In the case of a binary logistic regression, a common visualization tool is the ROC-curve (receiver operating characteristic), which plots the true positive rate (precision for one of the categories) against the false positive rate (one minus the precision for the other category) and investigates how they change when one adjusts the decision threshold for how high the estimated probability for one category must be in order for that category to be predicted (Sahngun Nahm, 2021, p. 25-26).

# Chapter 3

# Methodology

In order to estimate the target values of the project, two main operations need to be performed. Firstly, a great number of random wave data needs to be simulated for each wave state. Secondly, this wave data needs to be entered as input into a model which simulates the conversion process of wave heights to electrical energy.

## 3.1 Wave generation

The wave generation is done through the WAFO (Wave Analysis for Fatigue and Oceanography) toolbox in Matlab, created by Brodtkorb et al. (2000) and freely available online (WAFO-group, 2017).

We generate the waves from 78 different sea states, a subsample of the 215 wave states defined by the European Marine Energy Centre in (European Marine Energy Centre, 2022a). The subsample is the same one used by Zeinali et al. (forthcoming, p. 13, table 2), and covers 92.43% of the state occurrences measured at Billia Croo, Orkney, Scotland. These states, defined by their significant wave heights ($H_s$ or Hs) and energy periods ($T_e$ or Te), are presented in table 3.1. The full data set includes sea states with significant wave heights between 0.25 and 7.25 m, and energy periods between 2.5 and 17.5 s (European Marine Energy Centre, 2022a).

For each of these states, the first step is to calculate the spectral density of the waves. We will use a Bretschneider or Generalized Pierson-Moskowitz spectrum, truncated so that the frequencies above $5\omega_p$ are set to zero. In accordance with results already presented in section 2.1.2, this gives the spectral density (Zeinali et al., forthcoming, p. 2, eq. 1)

$$S(\omega) = \begin{cases} m_0 \frac{5}{\omega_p} \left(\frac{\omega}{\omega_p}\right)^{-5} e^{-\frac{5}{4}\left(\frac{\omega}{\omega_p}\right)^{-4}}, & 0 \leq \omega \leq 5\omega_p \\ 0, & \omega > 5\omega_p \end{cases} \tag{3.1}$$

and the spectral moments (Zeinali et al., forthcoming, p. 2, eq. 2)

$$m_n^{trunc} = \omega_p^n m_0 \left(\frac{5}{4}\right)^{n/4} \Gamma\left(1 - \frac{n}{4}, \frac{1}{500}\right). \tag{3.2}$$

In order to calculate the spectral density for each wave state, we need to find how the significant wave height and energy period values in table 3.1 translate to the parameters $m_0$ and $\omega_p$ in the

| Hs(m)/Te(s) | 5.5 | 6.5 | 7.5 | 8.5 | 9.5 | 10.5 | 11.5 | 12.5 | 13.5 | Sum |
|---|---|---|---|---|---|---|---|---|---|---|
| 6.75 | | | | | | 0.11 | | | | **0.11** |
| 6.25 | | | | | 0.11 | 0.23 | 0.16 | | | **0.50** |
| 5.75 | | | | | 0.24 | 0.36 | 0.16 | | | **0.76** |
| 5.25 | | | | 0.17 | 0.43 | 0.41 | 0.24 | | | **1.25** |
| 4.75 | | | | 0.35 | 0.72 | 0.58 | 0.30 | 0.14 | | **2.09** |
| 4.25 | | | 0.14 | 0.71 | 0.95 | 0.68 | 0.37 | 0.17 | | **3.02** |
| 3.75 | | | 0.43 | 1.27 | 1.24 | 0.80 | 0.42 | 0.20 | | **4.36** |
| 3.25 | | 0.13 | 0.96 | 1.73 | 1.67 | 1.10 | 0.61 | 0.27 | 0.12 | **6.59** |
| 2.75 | | 0.52 | 1.75 | 2.35 | 1.96 | 1.21 | 0.62 | 0.30 | 0.14 | **8.85** |
| 2.25 | 0.16 | 1.43 | 2.62 | 3.10 | 2.57 | 1.43 | 0.76 | 0.36 | 0.15 | **12.58** |
| 1.75 | 0.81 | 2.48 | 3.18 | 3.27 | 2.43 | 1.53 | 0.81 | 0.37 | 0.15 | **15.03** |
| 1.25 | 2.18 | 3.97 | 4.71 | 4.06 | 2.65 | 1.43 | 0.61 | 0.24 | 0.10 | **19.95** |
| 0.75 | 3.16 | 4.65 | 4.49 | 2.77 | 1.40 | 0.64 | 0.23 | | | **17.34** |
| Sum | **6.31** | **13.18** | **18.28** | **19.78** | **16.37** | **10.51** | **5.29** | **2.05** | **0.66** | **92.43** |

Table 3.1: 78 of the 215 wave states defined by European Marine Energy Centre (2022a) with wave heights (Hs) in meters and energy periods (Te) in seconds. The sampled wave states are those used by Zeinali et al. (forthcoming) and have been filled in with their relative frequencies of occurrence (%).

density expressions above. We may first use the relation between spectral moments and the energy period $T_e$ (mentioned in eq. (2.11)),

$$T_e = 2\pi \frac{m_{-1}}{m_0} \tag{3.3}$$

which when using the truncated spectrum becomes

$$\begin{aligned} T_e &= 2\pi \frac{\omega_p^{-1} m_0 (5/4)^{-1/4} \Gamma(5/4, 1/500)}{m_0 \Gamma(1, 1/500)} \\ &= \frac{2\pi}{\omega_p} (4/5)^{1/4} \frac{\Gamma(5/4, 1/500)}{\Gamma(1, 1/500)} \\ &= \frac{2\pi}{\omega_p} \cdot 0.8586184 \\ &= 0.86 \frac{2\pi}{\omega_p}. \end{aligned} \tag{3.4}$$

This differs only slightly from the energy period of the untruncated spectrum in eq. (2.13). From this expression we may easily calculate that

$$\omega_p = \frac{0.86 \cdot 2\pi}{T_e} = \frac{5.395859}{T_e} \tag{3.5}$$

and similarly, since we know that the significant wave height is defined as $H_s = 4\sqrt{m_0}$,

$$m_0 = \frac{H_s^2}{4} \tag{3.6}$$

and have thus calculated both the parameters necessary for the spectral density from the two defining wave state parameters. We may now visualize the spectral density in fig. 3.1 for the most common sea state, with $H_s = 1.25$ meters and $T_e = 7.5$ seconds.

The next step is to use the resulting spectral density to generate the time series of a wave. In accordance with Zeinali et al. (forthcoming, p. 4, eq. 8), we will simulate a linear Gaussian time-series $Z_t$ (representing the height of the sea surface without the buoy interaction) as a Fourier series

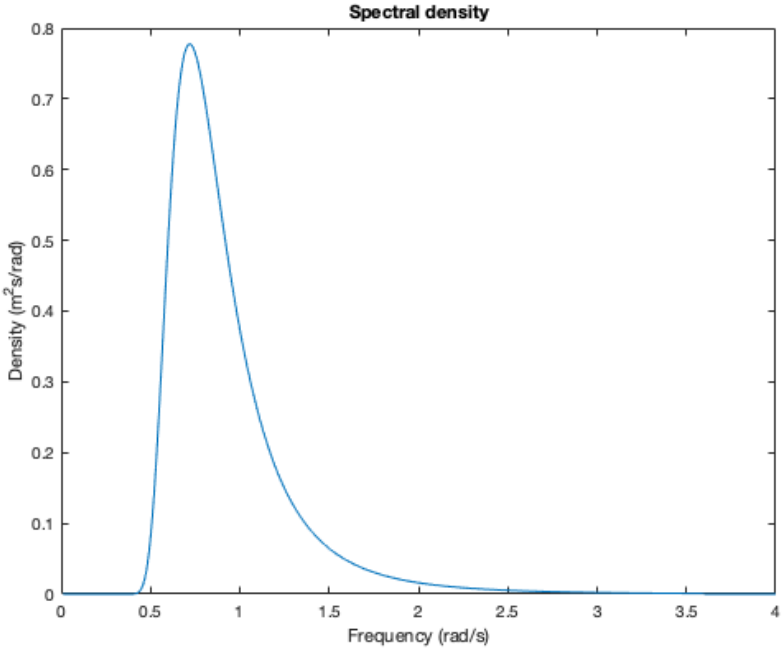


Figure 3.1: The spectral density of the truncated Bretschneider spectrum, following eq. (3.1) with $m_0 = H_s^2/4 = 1.25^2/4 \approx 0.39$ and $\omega_p \approx 5.396/T_e = 5.396/7.5 \approx 0.72$.

(see section 2.1.1)

$$Z_t = \sum_{i=0}^{N-1} \sqrt{S(\omega_i)\Delta_\omega} R_i \cos(\omega_i t + \theta_i) \tag{3.7}$$

where $S(\omega_i)$ is the truncated spectral density, $\Delta_\omega = \frac{2\pi}{N-1}$ is the frequency step size, $\omega_i = i\Delta_\omega$, $\theta_i$ are independent uniform random phases on the interval $(0, 2\pi)$ and $R_i$ are independent Rayleigh distributed random amplitudes.

In sine-cosine form, the Fourier series may be expressed as

$$Z_t = \sum_{i=0}^{N-1} \sqrt{S(\omega_i)\Delta_\omega}\big(X_i \cos(\omega_i t) + Y_i \sin(\omega_i t)\big) \tag{3.8}$$

where $X_i$ and $Y_i$ are independent random variables with expected value $\mathbb{E}[X_i] + \mathbb{E}[Y_i] = 0$ and variance $\mathbb{V}[X_i] + \mathbb{V}[Y_i] = \sigma_i^2$ (G. Lindgren, Rootzén, and Sandsten, 2014, p. 116). Introducing the random amplitude $R_i = \sqrt{X_i^2 + Y_i^2}$ and the random phase $\theta_i$ (such that $\cos\theta_i = X_i/R_i$ and $\sin\theta_i = -Y_i/R_i$ (G. Lindgren, Rootzén, and Sandsten, 2014, p. 116), we see through trigonometric

identities that

$$\begin{aligned}\sum_{i=0}^{N-1}\big(X_i\cos(\omega_i t)+Y_i\sin(\omega_i t)\big) &= \sum_{i=0}^{N-1}\big(\frac{\cos\theta_i}{\cos\theta_i}X_i\cos(\omega_i t)+\frac{\sin\theta_i}{\sin\theta_i}Y_i\sin(\omega_i t)\big)\\ &= \sum_{i=0}^{N-1}\big(\frac{X_i}{X_i/R_i}\cos(\theta_i)\cos(\omega_i t)-\frac{Y_i}{Y_i/R_i}\sin(\theta_i)\sin(\omega_i t)\big)\\ &= \sum_{i=0}^{N-1}R_i\big(\cos(\theta_i)\cos(\omega_i t)-\sin(\theta_i)\sin(\omega_i t)\big)\\ &= \sum_{i=0}^{N-1}R_i\cos(\omega_i t+\theta_i).\end{aligned} \tag{3.9}$$

Since $R_i\cos(\omega_i t+\theta_i)=X_i\cos(\omega_i t)+Y_i\sin(\omega_i t)$, we may then scale both sides by the (non-random) factor $\sqrt{S(\omega_i)\Delta_\omega}$ to get

$$\sum_{i=0}^{N-1}\sqrt{S(\omega_i)\Delta_\omega}\big(X_i\cos(\omega_i t)+Y_i\sin(\omega_i t)\big)=\sum_{i=0}^{N-1}\sqrt{S(\omega_i)\Delta_\omega}R_i\cos(\omega_i t+\theta_i) \tag{3.10}$$

and thus have equality for the two expressions. The proof of the distributions of $R_i$ (Rayleigh) and $\theta_i$ (uniform) is excluded for brevity. As already stated, the wave at a certain time $t$ is a sum of Gaussian random variables, making the wave height Gaussian as well, with expected value zero (since all terms in the sum have zero expected value) and variance

$$\begin{aligned}\mathbb{V}[Z_t] &= \sum_{i=0}^{N-1}S(\omega_i)\Delta_\omega\big(\cos^2(\omega_i t)+\sin^2(\omega_i t)\big)\sigma_i^2\\ &= \Delta_\omega\sum_{i=0}^{N-1}S(\omega_i)\sigma_i^2\end{aligned} \tag{3.11}$$

since all the Gaussian random variables $X_i$ and $Y_i$, $i=1,\ldots,N-1$ are independent, and $X_i$ and $Y_i$ have the same variance $\sigma_i^2$. However, the wave heights for different times $t$ will be highly correlated since they are constructed from the same weighted sum of random Gaussian variables, with only $t$ being different. The covariance between the wave height at times $t$ and $t+\tau$ may be rewritten as:

$$\begin{aligned}\mathbb{C}[Z_t,Z_{t+\tau}] = \mathbb{C}\Big[&\sum_{i=0}^{N-1}\sqrt{S(\omega_i)\Delta_\omega}\big(X_i\cos(\omega_i t)+Y_i\sin(\omega_i t)\big),\\ &\sum_{i=0}^{N-1}\sqrt{S(\omega_i)\Delta_\omega}\big(X_i\cos(\omega_i(t+\tau))+Y_i\sin(\omega_i(t+\tau))\big)\Big]\\ =&\sum_{i=0}^{N-1}S(\omega_i)\Delta_\omega\mathbb{C}\big[X_i\cos(\omega_i t)+Y_i\sin(\omega_i t),X_i\cos(\omega_i(t+\tau))+Y_i\sin(\omega_i(t+\tau))\big]\end{aligned} \tag{3.12}$$

due to the independence of each harmonic wave. Further,

$$\begin{aligned}\mathbb{C}[Z_t,Z_{t+\tau}] = \Delta_\omega\sum_{i=0}^{N-1}S(\omega_i)\Big(&\cos(\omega_i t)\cos(\omega_i(t+\tau))\cdot\mathbb{V}[X_i]+\sin(\omega_i t)\sin(\omega_i(t+\tau))\cdot\mathbb{V}[Y_i]\\ &+\big(\cos(\omega_i t)\sin(\omega_i(t+\tau))+\sin(\omega_i t)\cos(\omega_i(t+\tau))\big)\cdot\mathbb{C}[X_i,Y_i]\Big)\end{aligned}$$

$$\mathbb{C}[Z_t,Z_{t+\tau}]=\Delta_\omega\sum_{i=0}^{N-1}S(\omega_i)\sigma_i^2\big(\cos(\omega_i t)\cos(\omega_i(t+\tau))+\sin(\omega_i t)\sin(\omega_i(t+\tau))\big) \tag{3.13}$$

due to $X$ and $Y$ being uncorrelated with equal variance. Using trigonometric identities,

$$\mathbb{C}[Z_t, Z_{t+\tau}] = \Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \Big( \cos(\omega_i t)\big(\cos(\omega_i t)\cos(\omega_i \tau) - \sin(\omega_i t)\sin(\omega_i \tau)\big) + \sin(\omega_i t)\big(\sin(\omega_i t)\cos(\omega_i \tau) + \sin(\omega_i \tau)\cos(\omega_i t)\big)\Big) \tag{3.14}$$

$$\mathbb{C}[Z_t, Z_{t+\tau}] = \Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \big( \cos^2(\omega_i t)\cos(\omega_i \tau) - \sin(\omega_i t)\cos(\omega_i t)\sin(\omega_i \tau) + \sin^2(\omega_i t)\cos(\omega_i \tau) + \sin(\omega_i t)\cos(\omega_i t)\sin(\omega_i \tau)\big) \tag{3.15}$$

$$\begin{aligned} \mathbb{C}[Z_t, Z_{t+\tau}] &= \Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \cos(\omega_i \tau)\big(\cos^2(\omega_i t) + \sin^2(\omega_i t)\big) \\ &= \Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \cos(\omega_i \tau) \\ &= \Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \cos(i\Delta_\omega \tau) = \frac{2\pi}{N-1} \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \cos(i\frac{2\pi}{N-1}\tau) \end{aligned} \tag{3.16}$$

which as we see depends on the time lag $\tau$ and not the time $t$, confirming the process's stationarity. The normalized covariance, or correlation, is then found by dividing with the variance:

$$\begin{aligned} \mathrm{Corr}[Z_t, Z_{t+\tau}] &= \frac{\mathbb{C}[Z_t, Z_{t+\tau}]}{\mathbb{V}[Z_t]} \\ &= \frac{\Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \cos(i\Delta_\omega \tau)}{\Delta_\omega \sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2} \\ &= \frac{\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2 \cos(i\Delta_\omega \tau)}{\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2}. \end{aligned} \tag{3.17}$$

To see how the covariance structure develops as the time lag increases, we plot the correlation function using $N = 10000$ and $\tau$ from 0 to 30 with step size 0.1. We also assume the variances $\sigma_i^2$ to be equal for all superimposed waves $i = 0, \ldots, N-1$. This plot is seen in fig. 3.2 for four different sea states and we see that the correlation function is harmonic, as expected due to the harmonic nature of the process itself, and that the amplitude of the correlation gets progressively weaker as the lag increases. The speed of the correlation's disappearance seems to depend on the energy period $T_e$ but not on the significant wave height $H_s$, which agrees with our intuition as waves with longer periods should have longer-lasting covariance structures. The absolute value of the autocorrelation stays below 0.05 after lag 18 for energy period 7.5, and after lag 28 for energy period 11.5. The corresponding lag for all energy periods in the sample may be found in table 3.2, and the trend is clearly linear: an increase of 1 second in the energy period increases the lag threshold by approximately 2.5 seconds. The highest threshold for these sea states is 34.1 seconds, or 341 observations.

We may also get the energy content in the waves for each sea state by applying eq. (2.26). First, we need values for the fluid density and the gravity acceleration constant, which we will base on the geographical locations where the wave states were identified and where the WEC is located – that is, Billia Croo on Orkney and Runde in Norway, both adjacent to the northernmost North Sea. The mass density for sea water is between 1020-1030 $kg/m^3$ depending on location in the world, and approximately 1027 around both the Billia Croo and Runde test sites (Webb, 2019, p. 135), and we

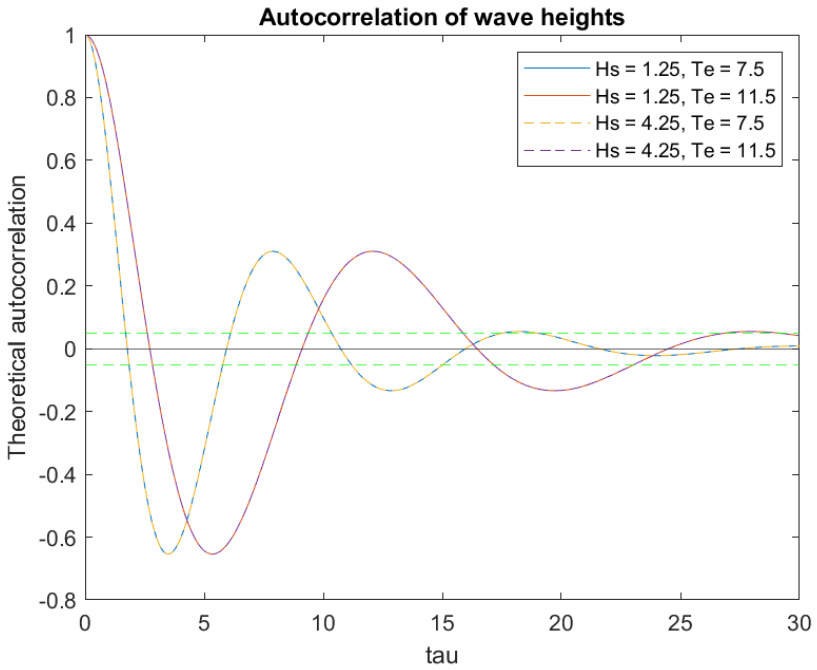


Figure 3.2: Theoretical autocorrelation of the wave heights $Z_t$ and $Z_{t+\tau}$ for time lags $\tau = 0.1, 0.2, 0.3, \ldots, 30$, using sample size $N = 10000$, for four different sea states. Blue and yellow lines ($T_e = 7.5$ s) agree perfectly with each other, as do the red and purple lines ($T_e = 11.5$ s). Dashed green lines are at $\pm 0.05$.

| Energy period ($T_e$) | Lag ($\tau$) |
|---|---|
| 5.5 | 13.9 |
| 6.5 | 16.4 |
| 7.5 | 18.9 |
| 8.5 | 21.4 |
| 9.5 | 24.0 |
| 10.5 | 26.5 |
| 11.5 | 29.1 |
| 12.5 | 31.6 |
| 13.5 | 34.1 |

Table 3.2: Time lag $\tau$ after which the absolute autocorrelation never exceeds 0.05, for each energy period $T_e$.

will thus use that as our $\rho$ value. The global average gravity acceleration is 9.807 $m/s^2$ (Conférence Générale des Poids et Mesures, 1901, p. 68) but the figure at Runde is about 10 mGal greater and thus closer to 9.82 $m/s^2$ (NASA Earth Observatory, 2003), which we will therefore use. This gives an approximate energy content of $J = 0.493 T_e H_s^2$. The contents for the different sea states range from 1.524 $J/(ms)$ ($H_s = 0.75$, $T_e = 5.5$) to 235.65 $J/(ms)$ ($H_s = 6.75$, $T_e = 10.5$).

The lengths of the generated time series will vary between different uses, but regardless we will always add a burn-in period of 500 observations (50 seconds) which are discarded, in order for the WEC:s simulated mechanics to start moving and oscillating from the standstill which they are in before the incident wave. In practice, different sea states will replace one another and the WEC:s components being completely stationary should never occur. Therefore, we discard the period which would reasonably be affected the most by the initial stillness of the Simulink model. The burn-in period is also the same one used by Zeinali et al. (forthcoming).

Having generated waves based on the sea state parameters, we now discuss the interaction between the wave and the WEC buoy.

## 3.2 WEC simulator

The WEC system which forms the basis of the simulink model is the WaveEL buoy, a real OWC-type WEC which is owned by Waves4Power and located near the Runde Environment Center AS in Norway. The previous version of this buoy, the WaveEl 3.0, was placed in water in February 2016 and was connected to the Norwegian electricity grid in June 2017, remaining active for five months until November when two mooring lines were cut off by unknown means. The installation of a new version, the WaveEL 4.0, is currently an ongoing project and is meant to be implemented commercially in 2024 (Waves4Power, 2024a).

The WaveEL 4.0 buoy consists of a large vertical cylindrical tube which extends deep enough in the water (28-38 meters) to make the effect of surface waves upon the internal water tube negligible. The interior of the tube has a diameter of 3.5 meters, but the WEC also has an upper part at the surface level which is 8 meters in diameter (Waves4Power, 2024a). Within the tube, which is hollow throughout the entire structure, a floating water piston is attached to a vertical rod and connects to a hydraulic system. Furthermore, the water piston is equipped with dampers, ensuring that it somewhat follows the movement of the buoy and forcing the water column up and down (Waves4Power, 2024b). In terms of the modes discussed in section 2.2, the buoy oscillates in a heave motion along the z axis. If the dampening force is either zero (the piston always in line with the water surface, moving freely) or maximized (no movement at all), there will be no electricity produced. In other cases, as a wave hits the buoy, the hull and the water piston shift vertically to different degrees and with different phase, causing the piston to move relative to the buoy and thus also moving the hydraulic rod to which it is connected (Waves4Power, 2024b). Effectively, there are two bodies oscillating – the WEC buoy and the water column – and the piston movement is caused by the phase difference between these two, making the water piston a form of pump (Waves4Power, 2024a). The hydraulic piston also has bumpers at each end, to soften the impact of the piston in case the pumping motion causes it to reach either extreme of its oscillation range.

The movement of the hydraulic piston increases the pressure in the accumulator tank, and this pressure also influences the movement of the water piston and water column, while at the same time the hydraulic accumulator tank is connected to a generator which converts the pressure to electrical power (Zeinali et al., forthcoming, p. 5). The flow of hydraulic fluid out of the accumulator tank is what drives the generator, and one of the main goals of this thesis is to investigate what generator rotation speed is optimal for the sea state. A too low speed will underutilize the hydraulic pressure

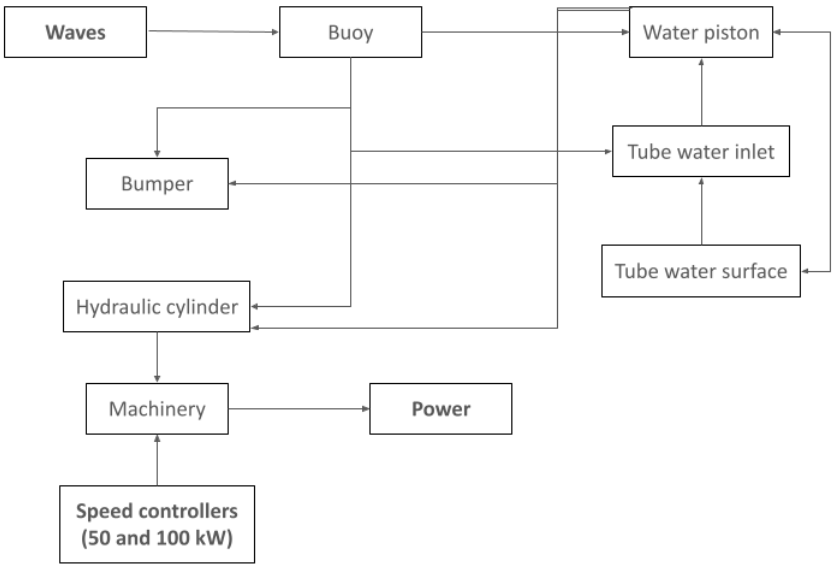


Figure 3.3: Conceptual sketch of the WEC model in Simulink.

and a too high speed leave the accumulator tank too empty and thus decreasing the pressure faster than it can be reaccumulated from the waves' kinetic energy – in other words, there needs to be a balance between the in- and outflows of energy in the tank. The accumulator tank has an initial pressure when the simulation starts, which for shorter wave processes may have a significant impact on the produced power. Therefore, we will use the optimal pressures found by Zeinali et al. (forthcoming) as our initial pressures for each sea state, since these should be close to the pressure levels acquired throughout the simulation. A sketch of the components of the WEC Simulink model is featured in fig. 3.3.

It is of note that the waves generated as above are treated as one-dimensional, only varying vertically. We assume that this represents their amplitude when hitting the buoy, where if the wave's horizontal dimensions are large in comparison to the buoy's diameter, this one-dimensional approximation should be valid. Waves4Power themselves describe the WaveEL WEC as a point absorber due to its small dimensions compared to the passing water waves (Waves4Power, 2024a).

However, Zeinali et al. (forthcoming, p. 4) argue that the shape of the buoy in its entirety needs to be taken into account since it may well affect the dynamics of the wave. They suggest a frequency response function $H(\omega)$ which may be used to transform the wave amplitude $A_\omega$ to a "phantom amplitude" $\tilde{A}_\omega = A_\omega \times H(\omega)$ which would have the property that a wave $\tilde{A}_\omega \cos(\kappa x - \omega t)$ has "the same hydrostatic buoyancy variation at the origin as the original wave gives on average over the ring" (Zeinali et al., forthcoming, p. 4). This frequency response function is (Zeinali et al., forthcoming, p. 4, eq. 10)

$$H(\omega) = \frac{2g}{\omega^2(R_2^2 - R_1^2)} \left( R_2 J_1\left(\frac{R_2\omega^2}{g}\right) - R_1 J_1\left(\frac{R_1\omega^2}{g}\right)\right) \tag{3.18}$$

where $g$ is gravity (Zeinali et al., forthcoming, p. 3), $R_1$ and $R_2$ are the inner and outer radii of the buoy's opening ring and $J$ is a Bessel function of the first kind (Zeinali et al., forthcoming, p. 5),

meaning

$$J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\,\Gamma(m+\alpha+1)} \left(\frac{x}{2}\right)^{2m+\alpha} \tag{3.19}$$

$$J_1\left(\frac{R_j\omega^2}{g}\right) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\,\Gamma(m+2)} \left(\frac{R_j\omega^2}{2g}\right)^{2m+1}, \; j = 1, 2 \tag{3.20}$$

as defined by Abramowitz and Stegun (1964, p. 360).

The time version of this filter, meaning the impulse response, is defined by Zeinali et al. (forthcoming, p. 5, eq. 11-12) as

$$h(t) = \frac{g|t|}{48(R_2^2 - R_1^2)} \big(f(t, R_2) - f(t, R_1)\big) \tag{3.21}$$

where

$$\begin{aligned} f(t,x) =& \left(\frac{t^4 g}{2x} + 8x\right)\left(J_{-\frac{1}{4}}\left(\frac{t^2 g}{8x}\right)^2 - J_{\frac{1}{4}}\left(\frac{t^2 g}{8x}\right)^2\right) \\ &- 2gt^2\left(J_{-\frac{1}{4}}\left(\frac{t^2 g}{8x}\right) J_{\frac{3}{4}}\left(\frac{t^2 g}{8x}\right) + J_{-\frac{1}{4}}\left(\frac{t^2 g}{8x}\right) J_{\frac{3}{4}}\left(\frac{t^2 g}{8x}\right)\right) \\ &+ \frac{t^4 g}{2x}\left(J_{\frac{3}{4}}\left(\frac{t^2 g}{8x}\right)^2 J_{-\frac{3}{4}}\left(\frac{t^2 g}{8x}\right)^2\right). \end{aligned} \tag{3.22}$$

After using the impulse response function to filter the generated waves, we simulate their input into the WEC system. This is done using a Simulink model provided by Zeinali et al. (forthcoming). To find the optimal rotations per minute, a maximization algorithm is run on this simulink model, for each sea state finding the rotation speed which gives the highest mean power output. This is done through an iterative procedure which takes several minutes for each state. We set the tolerance for both the output value (mean power) and the input value (rotation speed) to $10^{-2}$, meaning the algorithm will stop if either $|(r_i - r_{i+1})| < 10^{-2} \cdot (1 + |r_i|)$ or $|(m(r_i) - m(r_{i+1}))| < 10^{-2} \cdot (1 + |m(r_i)|)$ (MathWorks, Inc., 2024). Also of note is that the Simulink model contains adaptive solvers which lead to the output data (including the power) being non-equidistantly sampled. For this reason, linear interpolation is used in the Simulink to produce final output data with the same sampling frequency as the input, i.e. 0.1 s (Zeinali et al., forthcoming, p. 6).

In the Simulink model, the hydraulic system is not connected to just one generator, but to two: one with a working power of 50 kW and another with a working power of 100 kW. Theoretically, the WEC could produce energy using only one of the generators, besides varying the rotation speed (in rotations per minute) used. If both generators are connected, their speed should be the same to maintain balance in the hydraulic flow – two different speeds would mean two different voltages in the output which is difficult to deal with. It is much more convenient to, at least for a brief period, have a constant voltage and therefore a constant rotation speed. The optimization algorithm will therefore explore three options: An optimal rotation speed will be found iteratively for the case with only the 50 kW generator switched on, for the case with only 100 kW, and for the case with both generators switched on.

After optimizing the three rotation speeds, the associated mean outputs are compared to each other to find if the maximum for that sea state is reached through only one generator (and in that case, which one) or both. Finally, the resulting electrical output may be calculated using the optimum rotation speed, the optimum choice of generator(s), and the random wave at hand. This output

may also be compared to the energy potential in the wave in order to measure the WEC:s efficiency. The energy potential is, as stated in eqs. (2.21) and (2.26),

$$J = \frac{g^2\rho}{64\pi} T_e H_s^2 \ J/(ms) \approx 492.56 T_e H_s^2 \ W/m = 0.493 T_e H_s^2 \ kW/m \tag{3.23}$$

using the values $g = 9.82 \ m/s^2$ and $\rho = 1027 \ kg/m^3$. Since this energy potential is per unit width of the wave, we need to scale it by the diameter of the buoy's cylinder, since this is the width of waves interacting with the WEC at the water surface. This width is 8 meters (Shao et al., 2023, p. 448, fig. 2) which gives the energy potential per second

$$J \cdot m \approx 8 \cdot 0.493 T_e H_s^2 \ kW = 3.94 T_e H_s^2 \ kW \tag{3.24}$$

which may readily be compared to the simulated mean power output of each sea state.

The Simulink model is, naturally, a simplification in many respects. For instance, the mooring lines of the WEC buoy are not taken into account, and would in reality reduce the efficiency somewhat. Treating the WEC as a point absorber is also a simplification since the water cylinder still has a diameter of 3.5 meters, and the upper part a diameter of 8 meters. Furthermore, the model does not take possible non-heave movements of the buoy into consideration, even though such movement could well have effects (though likely small ones) on the energy conversion.

### 3.2.1 Choice of statistic for optimization

Having a high mean is not necessarily the only desired property of the power output. We would also like it to have a low variance so that the output is reliable given a particular sea state, and possibly reliable even when the sea states shift, though that is outside the scope of this thesis. Another reason to have a low variance is to have a consistent voltage on the output in the sense that it stays within a reasonably narrow interval. We may implement variance reduction into our computations by not using the mean itself as our target variable to maximise in the rotation speed optimization, but rather the mean minus a rescaling of the power's standard deviation, that is,

$$\hat{\mu}^* = \hat{\mu} - \lambda \cdot \hat{\sigma} \tag{3.25}$$

where $\lambda \geq 0$. An alternative could be to instead use the coefficient of variation $R$, but since $\mu$ may be negative for certain data this may also give a negative coefficient of variation, making it difficult to interpret and not useful in an optimization. Therefore, even though the potentially different magnitudes of variance for each sea state may lead to different effects from increasing the scaling parameter $\lambda$, we will still use the pure standard deviation in our optimization statistic. Another distinction is that we will not calculate the standard deviation based on all observations in the time series, but rather use sub-interval means of the time series as our data, since the short-term variability of the energy output is not of great interest but rather the variability of the mean output over some time interval, which we choose to be one minute or 600 observations. This measure of the standard deviation will be smaller than the one based on individual observations, which means that the $\lambda$ coefficient needs to be larger to have a tangible effect upon the optimization statistic.

To investigate the effect of the hyperparameter $\lambda$ upon the optimal rotation speed as well as the associated mean output and standard deviation, we try running the optimization algorithm on a smaller wave process of length 12 000 (excluding the burn-in of 500 observations), using values $\lambda = 0, 0.5, \ldots, 4.5, 5$ and always using both generators in order to save time. We then estimate the mean output and variance on new processes of length 24 000.

The effect varies between the sea states. In fig. 3.4, we see that for the most common sea state the effect is quite irregular. The rotation speed slowly decreases from 633 to around 500 rpm when

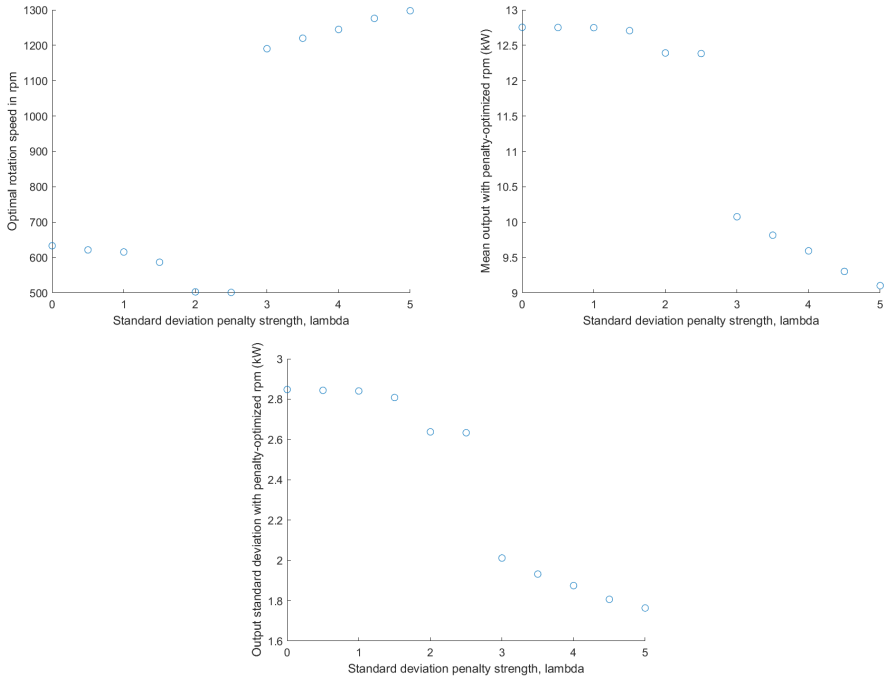


Figure 3.4: The optimized rotation speed (upper left) as well as corresponding estimated mean power (upper right) and standard deviation (lower) for various values of the hyperparameter $\lambda$ in eq. (3.25), using wave processes of length 12 000 for optimization and 24 000 for estimation from the most common sea state ($H_s = 1.25$ m, $T_e = 7.5$ s).

$\lambda$ goes from 0 to 2.5, then suddenly jumps up to 1190 rpm whereafter it increases approximately linearly to 1297 rpm. The mean power starts out with a corresponding slow decrease from around 12.75 to 12.38 kW, but then decreases rapidly to 10.07 and onwards down to 9.10 kW. The standard deviation, finally, follows an almost identical pattern to the mean. The total decrease in standard deviation is smaller than the one in mean (1.08 kW compared to 3.65 kW), but is somewhat larger in relative terms (38% and 29%) when dividing the change by the respective values when $\lambda = 0$.

In fig. 3.5, the corresponding results for the second most common sea state are shown, and are similar in some respects and different in others. The optimal rotation speed initially increases slowly, rather than decreasing, from 500 to 903 rpm. Then, just as with the most common sea state, there is a sudden large increase, though this one occurs for $\lambda = 3.5$ rather than $\lambda = 3$. The increase is also even more extreme, going up to values between 2400 and 2900 rpm. The mean decreases slowly from below 7 to around 4.5 kW, but then goes to around zero after the threshold. The standard deviation decreases more evenly, and up to $\lambda = 3$ has a larger relative decrease than the mean (60% compared to 34%), but its decrease at $\lambda = 3.5$ is less extreme than that of the mean. This pattern likely follows from the second sea state generally providing less energy, meaning that as $\lambda$ increases it becomes increasingly difficult to get a high optimization statistic.

A similar visual examination of other sea states (excluded for brevity) shows that they follow their own patterns, but in general still have some things in common: Firstly, as $\lambda$ increases, both the mean and the standard deviation of the output effect are reduced. Secondly, for most states there are multiple "jumps" in both rotation speed, mean and standard deviation (generally simultaneously), increases in $\lambda$ which have very large effects. The sizes and thresholds of these sudden increases or decreases vary between the states, however. A reason for the instability of the optima may be the higher variability of the sample variance, which tends to take more differing values depending on the sample data than the mean does. This makes the optima more uncertain when the variance is

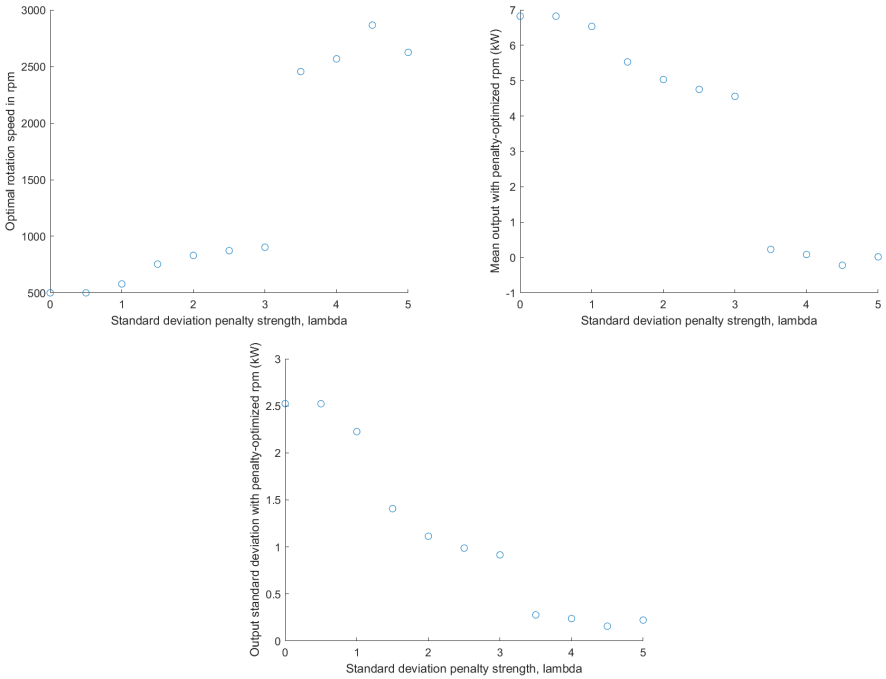


Figure 3.5: The optimized rotation speed (upper left) as well as corresponding estimated mean power (upper right) and standard deviation (lower) for various values of the hyperparameter $\lambda$ in eq. (3.25), using wave processes of length 12 000 for optimization and 24 000 for estimation from the second most common sea state ($H_s = 0.75$ m, $T_e = 6.5$ s).

included in the maximization statistic.

To get an aggregated overview of the effect of $\lambda$ upon the energy output without having to show all 78 sea states separately, we take the weighted average output of the sea states using their relative frequencies of occurrence from table 3.1 to get an estimate of the total mean power output, and also calculate the aggregated standard deviation. We calculate these values, with the probability of occurrence being $\mathbb{P}(k)$ for sea state $k$, as

$$\mu_{total} = \sum_{k=1}^{N} \frac{\mathbb{P}(k)}{\sum_{j=1}^{N} \mathbb{P}(j)} \mu_k \tag{3.26}$$

$$\sigma_{total} = \sqrt{\sum_{k=1}^{N} \left( \frac{\mathbb{P}(k)}{\sum_{j=1}^{N} \mathbb{P}(j)} \right)^2 \sigma_k^2} \tag{3.27}$$

The total mean and standard deviation values resulting from different $\lambda$ values are featured in fig. 3.6, while in fig. 3.7 we find the absolute and relative changes for each $\lambda$ value (compared to when $\lambda = 0$). The absolute changes are

$$\Delta_{\mu,abs}(\lambda) = \mu_{total}(\lambda) - \mu_{total}(0) \tag{3.28}$$

$$\Delta_{\sigma,abs}(\lambda) = \sigma_{total}(\lambda) - \sigma_{total}(0) \tag{3.29}$$

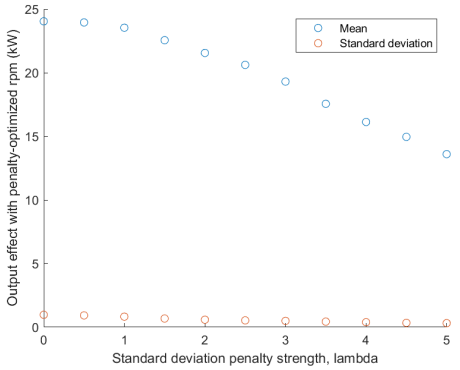


Figure 3.6: Estimated mean and standard deviation for the energy output as a weighted average of all 78 sea states.

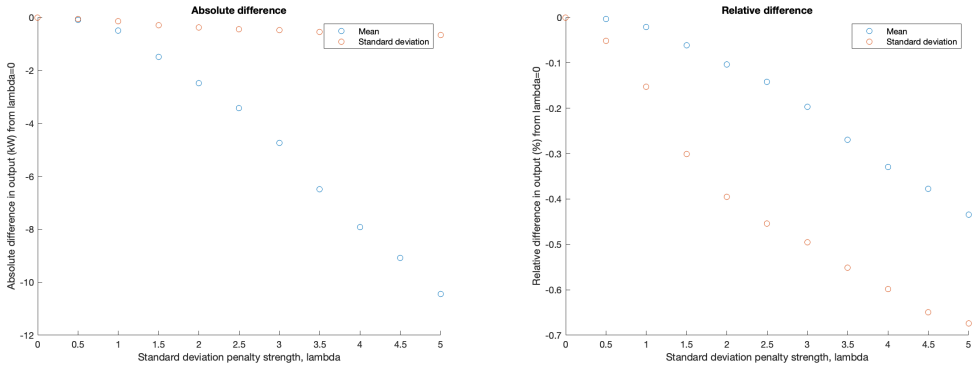


Figure 3.7: Absolute and relative differences (compared to the case when $\lambda = 0$) for the energy output as a weighted average of all 78 sea states.

while the relative changes are

$$\Delta_{\mu,rel}(\lambda) = \frac{\Delta_{\mu,abs}(\lambda)}{\mu_{total}(0)} = \frac{\mu_{total}(\lambda) - \mu_{total}(0)}{\mu_{total}(0)} \tag{3.30}$$

$$\Delta_{\sigma,rel}(\lambda) = \frac{\Delta_{\sigma,abs}(\lambda)}{\sigma_{total}(0)} = \frac{\sigma_{total}(\lambda) - \sigma_{total}(0)}{\sigma_{total}(0)}. \tag{3.31}$$

As was initially seen for the individual sea states, both the mean and the standard deviation decrease, though quite slowly, as we increase $\lambda$. The changes are now much smoother than they were for the individual sea states, without large jumps in the measures, since the jumps happen at different $\lambda$ for different sea states. Interestingly enough, the absolute changes are much larger for the mean, likely due to the fact that the standard deviation decreases across all $\lambda$ when we take the weighted average. The relative change, however, is much larger for the standard deviation.

Overall, increasing $\lambda$ all the way from 0 to 5 decreases the estimated mean output from 24.04 to 13.60 kW while the standard deviation decreases from 0.9774 to 0.3182 kW. The absolute reduction is clearly much larger for the mean and the contrast becomes starker the higher our $\lambda$. The reduction in standard deviation is only 0.6592 kW even for $\lambda = 5$, while the corresponding reduction for the mean is more than 10 kW. However, the relative reduction is instead larger for the standard deviation, particularly for $\lambda$ above 1.5. With high enough $\lambda$ the standard deviation may be reduced by more than 67%, though this is also associated with the mean being reduced by 43%.

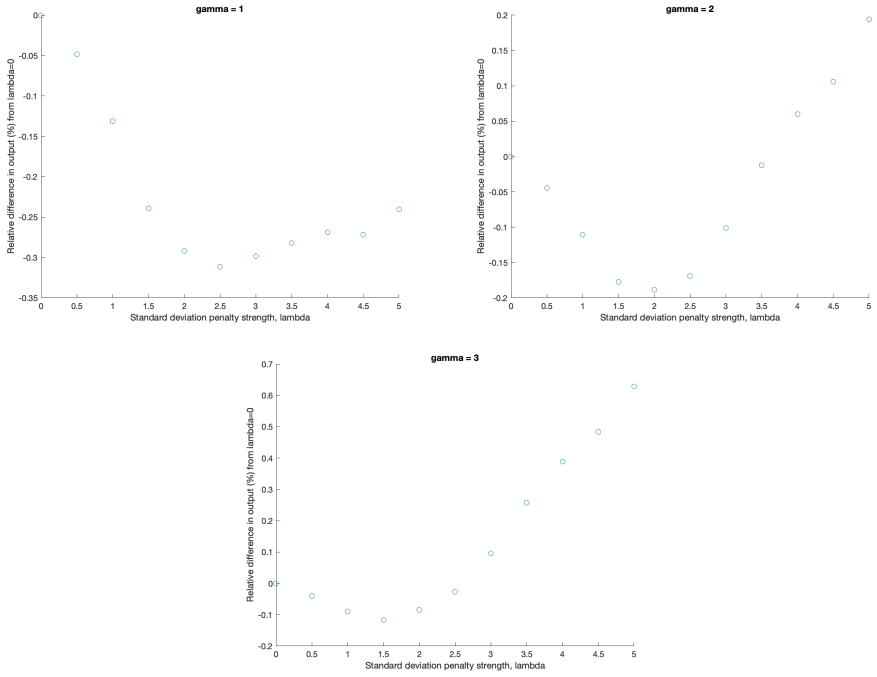


Figure 3.8: Weighted difference (in percentage points) between the relative changes in mean and standard deviation for different values of $\lambda$, according to eq. (3.32).

To select a value for $\lambda$, we investigate one more measure: The difference in relative change between the mean and the standard deviation, with a possible weighting due to the fact that we would possibly be more averse to reduction in mean than we are positive to reduction in variance. We therefore calculate the following difference:

$$D = \Delta_{\sigma,rel} - \gamma\Delta_{\mu,rel} \tag{3.32}$$

where $\gamma$ increases above 1 if we want to prioritize maintaining a high mean more than reducing variance. In fig. 3.8, we plot the value of $D$ for $\lambda = 0, 0.5, \ldots, 5$, using $\gamma = 1, 2, 3$. Since both $\Delta$ values are negative, a low $D$ implies a strong relative reduction of $\sigma$ without making the reduction of $\mu$ too strong.

All three plots give clear minima, and as is to be expected the minimum is at a different $\lambda$ value depending on $\gamma$, lower the higher we prioritize low mean reduction over high variance reduction: $\lambda = 2.5$ for $\gamma = 1$, $\lambda = 2$ for $\gamma = 2$ and $\lambda = 1.5$ for $\gamma = 3$. We settle for $\lambda = 2$, which according to fig. 3.7 gives a mean reduction of approximately 10% (2.488 kW) and a standard deviation reduction of 40% (0.3864 kW). This seems like a fair balance between reducing the variance while still maintaining a high mean. Therefore, the maximization statistic used in the algorithm will be

$$\hat{\mu}^* = \hat{\mu} - 2\hat{\sigma}. \tag{3.33}$$

However, the main results will be compared to the case when $\lambda = 0$, in order to see the true overall impact of the coefficient.

## 3.3 Monte Carlo estimation

Since the optimal rotation speed and the resulting electrical output are both dependent on the generated wave, they are random variables and the accuracy of a single sample is unknown to us. What we would, ideally, like to know are the expected value and variance of the output given a particular sea state. However, these values would be calculated as integrals (in accordance with eq. (2.46))

$$\mu_{P_{out,t}} = \mathbb{E}[h_t(Z)] = \int h_t(z) f(z) d\mathbf{z} \tag{3.34}$$

$$\sigma^2_{P_{out,t}} = \mathbb{V}[h_t(Z)] = \int (h_t(z) - \mu_{P_t}) f(z) d\mathbf{z} \tag{3.35}$$

which we may not evaluate explicitly due to the complexity of the function $h_t(Z)$, here meaning the electrical output as a deterministic function of the wave height $Z$ not only at a certain time $t$ but of the wave heights at multiple adjacent time points. Thus, we will use Monte Carlo integration, as described in section 2.5 to improve the estimates both of the optimal rotation speed and of the expected output, as well as to get an idea of the variability of these estimates.

One issue that must be taken into consideration is the computing time of these operations. One option for optimizing the rotation speed and estimating the expected output (for each sea state) would be to generate $n_1$ independent wave processes of a length $T$, optimizing the rotation speed for each process and using the average optimum as a fixed parameter in $n_2$ more independent processes of the same length. The mean outputs of these new processes are then averaged to get the point estimate for the sea state's expected power output. Since the computing time for the optimization is significantly longer than for the time required for simply using the Simulink model with a fixed rotation speed, it seems reasonable to settle for a lower $n_1$ and a higher $n_2$. The law of large numbers dictates that this would give a decent approximation of the optimum rotation speed, and a better approximation of the expected output given this speed.

An alternative would be to replace the ensemble of processes with only two processes for each state: One of length $N_1 = n_1 \cdot T$ and one of length $N_2 = n_2 \cdot T$. This would still give the same number of total observations for the tasks of optimization and estimation, respectively, as if we had multiple independent processes. Though the longer process takes more time to generate than $n$ shorter processes due to each observation being a sum of $N = nT$ terms (see eq. (3.10)) and the computation time thus increasing non-linearly, this should be made up for by only having to run the optimization algorithm once. This should be much faster than running it $n$ times, even though the process has many more observations. The difference in computation time was confirmed by empirically testing the computing times for the operations for different process lengths.

These methods do not necessarily give equivalent results, however. Firstly, in order for the time-average of $N = nT$ observations to be equivalent in accuracy to an ensemble average of $n$ processes of length $T$, the process must have some degree of linear ergodicity. If not, the time-average will not be a consistent estimator. However, as stated in section 2.5.1 a stationary process will be ergodic in the quadratic-mean sense if the average of its covariances converges to zero for large $N$, which should occur if the covariance function goes to zero for large lags $\tau$ and positive and negative covariances occur approximately equally. Since the wave processes are stationary, we may investigate the behaviour of their covariance functions to evaluate their degree of ergodicity. Similarly, even though ergodic wave processes do not necessarily imply ergodic power output due to the complexity of the Simulink model, we may investigate the ergodicity of the output through the empirical covariance structure.

Having only a singular process also leaves us without any measure on the estimates' accuracy. To accommodate this, one solution might be to save the individual observations of the process rather

than just their mean, and simply calculate the variance of these observations, but this carries with it the problem that the individual observations are correlated to others in their proximity since they are all based on the same $N$ superimposed harmonic processes. If the observations are correlated, the variance of their mean will not be proportional to the variance of the observations, and the sample variance is then not appropriate to use when constructing a confidence interval for the mean. Furthermore, as already stated, the exact fluctuations of the power are not of interest but rather its mean over a time period of a set length.

Instead of using the individual observations, we may divide the process into $K$ intervals or segments of length $N_{sub}$, and use the interval means as our saved observations, rather than the global mean or the original wave heights. This is similar to the division into sub-intervals of length 600 which we used in the optimizations for different $\lambda$ coefficients, but for a different purpose. As long as the process is stationary, the expected value and covariance structure should be equal in each segment and in this regard it should then be equivalent to simulating $n = K$ processes. Also, as the interval length increases, the covariance between the intervals should decrease until it is negligible, enabling us to use the mean and variance of the interval means to construct a confidence interval for the expected power output. If the process is stationary and sufficiently ergodic, and the covariance between segments is negligible, the global mean will have the properties (stated for the general case in eqs. (2.47) and (2.48))

$$\mathbb{E}[\bar{Z}_{Global}] = \mathbb{E}\Big[\frac{1}{K}\sum_{i=1}^{K}\bar{Z}_i\Big] = \frac{1}{K}\cdot K\cdot\mathbb{E}[\bar{Z}] = \mathbb{E}[\bar{Z}] \tag{3.36}$$

$$\mathbb{V}[\bar{Z}_{Global}] = \mathbb{V}\Big[\frac{1}{K}\sum_{i=1}^{K}\bar{Z}_i\Big] \tag{3.37}$$

$$\begin{aligned} &= \frac{1}{K^2}\sum_{i=1}^{K}\Big(\mathbb{V}[\bar{Z}_i] + \sum_{j\neq i}\mathbb{C}[\bar{Z}_i,\bar{Z}_j]\Big) \\ &\approx \frac{1}{K^2}\sum_{i=1}^{K}\mathbb{V}[\bar{Z}_i] = \frac{1}{K}\mathbb{V}[\bar{Z}] \end{aligned} \tag{3.38}$$

making the global mean a consistent estimator. The above relations for the wave height $Z$ also hold for the power output $P_{out}$ and its averages (local and global), as long as the power output too is stationary and has negligible inter-segment covariance. If this is the case, the interval means for the power output may be used to construct a confidence interval for the global mean using eq. (2.50):

$$CI_{\mu_{P_{out}}(k)} = \left(\big(\hat{\mu}_{P_{out}}(k) + \phi_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2_{P_{out}}(k)}{K}}\big), \big(\hat{\mu}_{P_{out}}(k) + \phi_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2_{P_{out}}(k)}{K}}\big)\right) \tag{3.39}$$

where $k$ is the sea state, $\hat{\mu}_{P_{out}}(k)$ is the global sample mean and $\hat{\sigma}^2_{P_{out}}(k)$ is the sample variance among the interval means.

### 3.3.1 Covariance and ergodicity of wave heights

It is then important to investigate what segment length is required in order for the covariance to be negligible, as well as whether the processes are ergodic. The covariance between the average wave

height of two adjacent segments (the first starting at observation $m$) may be calculated as

$$\begin{aligned}\mathbb{C}[\bar{Z}_1, \bar{Z}_2] &= \mathbb{C}\left(\frac{1}{N_{sub}}\sum_{k=m}^{m+N_{sub}-1} Z_k, \frac{1}{N_{sub}}\sum_{j=m+N_{sub}}^{m+2N_{sub}-1} Z_j\right) \\ &= \frac{1}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{j=m+N_{sub}}^{m+2N_{sub}-1}\mathbb{C}[Z_k, Z_j] \\ &= \frac{1}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=N_{sub}-k+2}^{2N_{sub}-k+1}\mathbb{C}[Z_t, Z_{t+\tau}]\end{aligned} \tag{3.40}$$

where we may then use eq. (3.16) to get

$$\begin{aligned}\mathbb{C}[\bar{Z}_1, \bar{Z}_2] &= \frac{1}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=N_{sub}-k+2}^{2N_{sub}-k+1}\Delta_\omega\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau) \\ &= \frac{\Delta_\omega}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=N_{sub}-k+2}^{2N_{sub}-k+1}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau) \\ &= \frac{2\pi}{N_{sub}^2(N-1)}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=N_{sub}-k+2}^{2N_{sub}-k+1}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\frac{2\pi}{N-1}\tau).\end{aligned} \tag{3.41}$$

The variance for a single segment mean would be

$$\begin{aligned}\mathbb{V}[\bar{Z}_1] &= \mathbb{V}\left(\frac{1}{N_{sub}}\sum_{k=m}^{m+N_{sub}-1} Z_k\right) = \frac{1}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{j=m}^{m+N_{sub}-1}\mathbb{C}[Z_j, Z_k] \\ &= \frac{1}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=0}^{N_{sub}-k}\mathbb{C}[Z_t, Z_{t+\tau}] \\ &= \frac{1}{N_{sub}^2}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=0}^{N_{sub}-k}\Delta_\omega\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau) \\ &= \frac{2\pi}{N_{sub}^2(N-1)}\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=0}^{N_{sub}-k}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau),\end{aligned} \tag{3.42}$$

which then gives us a correlation equal to

$$\begin{aligned}\mathrm{Corr}[\bar{Z}_1, \bar{Z}_2] &= \frac{\mathbb{C}[\bar{Z}_1, \bar{Z}_2]}{\mathbb{V}[\bar{Z}_1]} \\ &= \frac{\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=N_{sub}-k+2}^{2N_{sub}-k+1}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau)}{\sum_{k=m}^{m+N_{sub}-1}\sum_{\tau=0}^{N_{sub}-k}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau)}.\end{aligned} \tag{3.43}$$

The index $m$, signifying the time of the first wave height in the first segment, should not affect either the variance or the covariance as long as the process is stationary, since the two segments should then have the same properties regardless of their starting point. We may therefore, for simplicity, set $m = 1$, having the segments contain the first $2N_{sub}$ observations of the process. The correlation is then expressed as

$$\mathrm{Corr}[\bar{Z}_1, \bar{Z}_2] = \frac{\sum_{k=1}^{N_{sub}}\sum_{\tau=N_{sub}-k+2}^{2N_{sub}-k+1}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau)}{\sum_{k=1}^{N_{sub}}\sum_{\tau=0}^{N_{sub}-k}\sum_{i=0}^{N-1} S(\omega_i)\sigma_i^2\cos(i\Delta_\omega\tau)}. \tag{3.44}$$

with the indices of the outermost summations changed. The strength of this correlation for different choices of segment length $N_{sub}$ may be calculated numerically for a given sea state and sample size

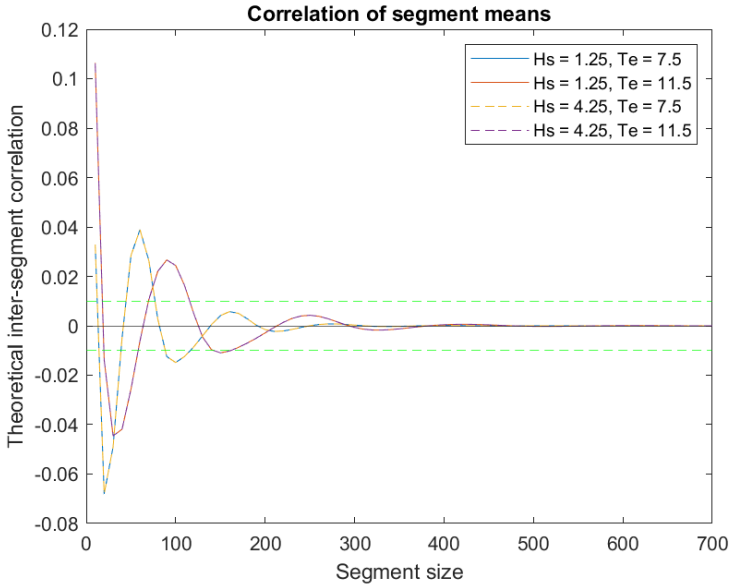


Figure 3.9: Theoretical correlation of the segment average wave heights $\bar{Z}_1$ and $\bar{Z}_1$ for adjacent segments with length $N_{sub} = 10, 20, \ldots, 690, 700$ and sample size $N = 5000$. Dashed green lines are at $\pm 0.01$.

$N$. The theoretical correlations for $N_{sub} = 10, 20, \ldots, 690, 700$ are found in fig. 3.9, with $N = 5000$ and the same sea states as in fig. 3.2.

As expected, the inter-segment correlation is not affected by the significant wave height but is slightly affected by the energy period. For energy period $T_e = 7.5$, the correlation starts slightly above 0.03 for $N_{sub} = 10$, grows as strong as $-0.06$ for $N_{sub} = 30$, and then starts decreasing periodically. For segment lengths above 120, the absolute correlation is well below 0.01, with length $N_{sub} = 500$ giving a correlation very close to zero. With energy period 11.5, the correlations degenerate slightly more slowly, starting out above 0.1 (for $N_{sub} = 10$), remaining below 0.01 for $N_{sub}$ of approximately 160 and upwards, and not being quite as close to zero at $N_{sub} = 500$ as when using the shorter energy period. However, it is still very close to zero with values of order $10^{-4}$. Even for the longest energy period, $T_e = 13.5$ (presented in fig. 3.10), the inter-segment correlation is of order $10^{-4}$ for segment length 500.

The covariances are also used to investigate the ergodicity of the wave process. As stated in eqs. (2.61) and (2.62), for a stationary stochastic process we have second-order ergodicity if the cumulative average of the covariance function converges to zero as we add covariances for longer and longer time lags $\tau$. To investigate this, we calculate theoretical covariances using eq. (3.16) for $\tau = 0, 0.1, \ldots, T_{max}$ and calculate the $T_{max}$ averages $\bar{r}_T(\tau) = \frac{1}{T}\sum_{j=1}^{T} r(0.1 \cdot j)$. This is plotted in fig. 3.11 for the same sea states as in fig. 3.9, and we may clearly see that all four approach zero, though closer analysis reveals that they do not quite reach it even for $T_{max} = 100$. Investigating similar plots for higher $T_{max}$ (excluded for brevity) reveals that the convergence towards zero continues, albeit slowly. We therefore conclude that the wave height time series are indeed second-order ergodic, making it reasonable to use only one time series each for the optimization and estimation on a given sea state.

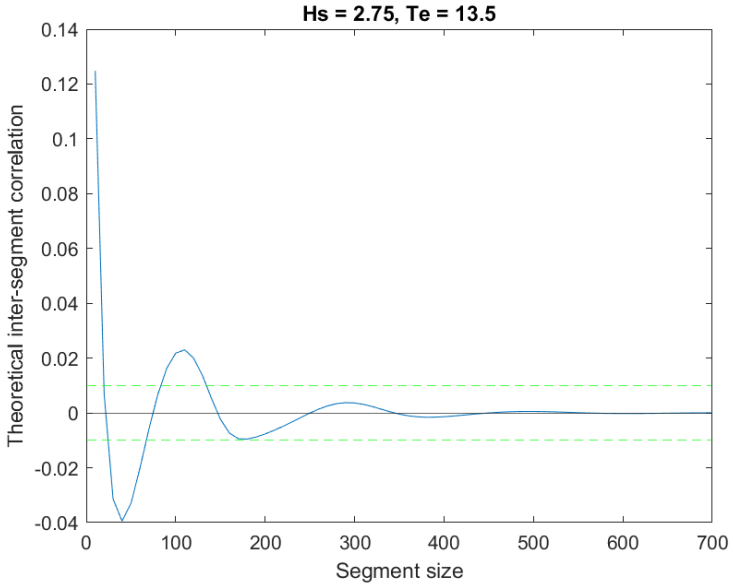


Figure 3.10: Theoretical correlation of the segment average wave heights $\bar{Z}_1$ and $\bar{Z}_1$ for adjacent segments with length $N_{sub} = 10, 20, \ldots, 690, 700$ and sample size $N = 5000$. Significant wave height $H_s = 1.25$, energy period $T_e = 13.5$ (the highest of any wave state according to the EMEC classification (European Marine Energy Centre, 2022b)). Dashed green lines are at $\pm 0.01$.

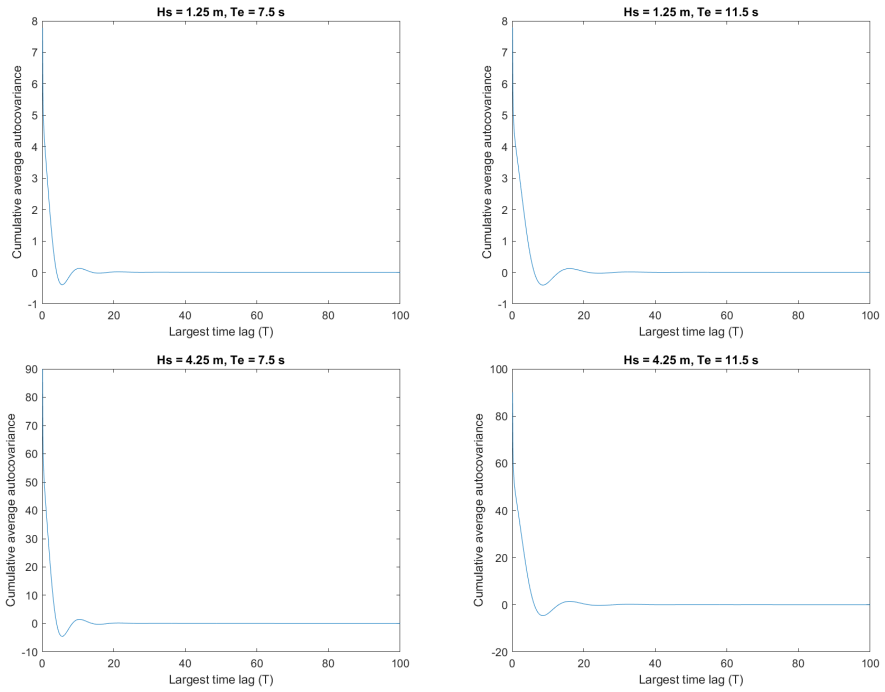


Figure 3.11: Theoretical cumulative average covariance of the wave heights, $\bar{r}_T(\tau) = \frac{1}{T}\sum_{j=1}^{T} r(0.1 \cdot j)$, with sample size $N = 5000$. The sea states used have significant wave height $H_s$ equal to 1.25 for the top row and 4.25 for the bottom, while the energy period $T_e$ is 7.5 for the left column and 11.5 for the right.

### 3.3.2 Empirical covariance of the power output

We need to consider the risk that even though the correlation between the average wave height segments quickly reaches a small order of magnitude, we have no way of explicitly calculating how this translates to the covariance, variance or correlation of the segments' mean power output – if we did, using Monte Carlo integration for this problem would be meaningless. It can be assumed that this correlation, too, should generally decrease for larger $n$, but the degree of this reduction is unknown. Since the mean power output for a segment not only depends on the mean wave height but on other, non-linear patterns among the wave heights, it is very possible that $N_{sub} = 500$ is not a sufficient segment length.

There is also the concern that even though the incident waves are Gaussian, due to the complicated transformations of the WEC Simulink the power output could very well be non-Gaussian. Thus, in order to construct a Gaussian (or rather, Student's t) confidence interval we would need the central limit theorem, which would also require the number of sub-intervals to be large in order for the global mean to approach Gaussianity. This is, of course, not an issue if the interval means themselves are sufficiently Gaussian, which they should be provided that $N_{sub}$ is large enough. Since the estimator of interest is the global mean power output, that is, the average over the entire process regardless of how many segments were used, its Gaussianity should in fact not at all depend on the relative sizes of $K$ and $N_{sub}$. However, we could use the $K$ interval means to get a visualization of how well the mean output approaches a Gaussian distribution with only $N_{sub}$ observations. If the interval means are close to Gaussian, the global mean should be even closer.

We first investigate the power output's ergodicity through the cumulative average of the sample covariances for individual observations, to find whether it seems to converge to zero. These empirical cumulative average autocorrelations are found in fig. 3.12 for four different sea states. For all four of the states, the average autocorrelation quickly decreases towards zero, and for three of them then fluctuate around zero to different degrees, showing a manner of convergence besides a random fluctuation. For the sea state with $H_s = 1.25$ m, $T_e = 11.5$ s, we see an increase in the average autocorrelation as the maximum lag approaches 20 000. However, when investigating this sea state further (figure excluded for brevity) we see that it too fluctuates around zero, albeit much more slowly. Other sea states show similar patterns, with the fluctuations being slower and larger for some states (primarily those with low significant heights and long energy periods) but small and fast for most. We conclude that at least for the majority of sea states the process is highly ergodic, and we continue using the time-mean in place of an ensemble mean.

We move on to investigating the correlation between segment mean power outputs. The correlation plots in figs. 3.9 and 3.10 suggest that a segment length of 500 should be enough to get almost zero correlation between mean wave heights, but as already stated this does not mean that the correlation between mean power outputs will be as small. Theoretically, if we reach such $N_{sub}$ that the segment means are completely uncorrelated then the observed means will be as if they from a white noise process with non-zero mean $\mu$ and variance $\sigma^2$, which should also be asymptotically Gaussian given a large enough $K$. Under such conditions, Jakobsson (2021, p. 48) states that the empirical autocorrelation function of the segments, $\hat{r}(\tau^*)$ will be Gaussian with mean 0 and variance $1/K$, and that we may therefore classify autocorrelations with absolute value above $\phi_{1-\alpha/2} \cdot \sqrt{1/K}$ (where $\phi$ is a normal distribution quantile) as significant. We could therefore simulate data and check the empirical autocorrelation using different segment lengths. However, for this to be valid $n$ also needs to be large enough for the interval means to be approximately Gaussian.

In order to check the segment length necessary to make the segment means uncorrelated (and thus effectively equivalent to independent observations), we simulate data from the four sea states used in figs. 3.2 and 3.9 and optimize generator choice and rotation speed for each one. We then try out different lengths $N_{sub}$ and make both an autocorrelation function (acf) plot and a normal probability

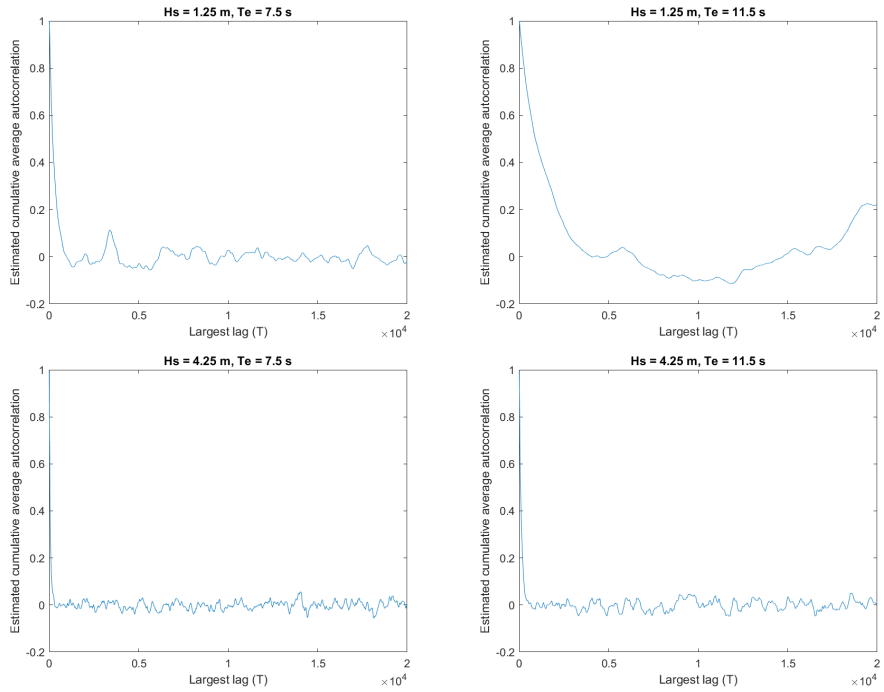


Figure 3.12: Empirical cumulative average covariance of the momentary power output, $\bar{r}_T(\tau) = \frac{1}{T}\sum_{j=1}^{T} r(0.1 \cdot j)$, with sample size $N = 240000$. The sea states used have significant wave height $H_s$ equal to 1.25 for the top row and 4.25 for the bottom, while the energy period $T_e$ is 7.5 for the left column and 11.5 for the right.

plot, comparing the empirical quantiles of the segment means to the theoretical ones from a normal distribution. The resulting plots for segment length $N_{sub} = 1000$ are found in fig. 3.13, and the ones for $N_{sub} = 4000$ in fig. 3.14. Other segment lengths were also tried.

The ACF plots show a clear difference between interval lengths 1000 and 4000. In fig. 3.13, three out of four ACF plots show significant correlations. In one case there is only one out of 20, which is not unreasonable since even in a white noise process 5% of the observed autocorrelations should be outside the 95% confidence interval bounds, but in the others there are two. In the case of the second sea state, where $H_s = 1.25$ and $T_e = 11.5$, the correlation for lag 1 is as strong as 0.6, and there are also signs of periodicity in the ACF. For segment length 4000, there is only one observed significant autocorrelation, which as stated is not unreasonable. There are also much fewer signs of periodic behaviour in the ACF of the second sea state. It is important to note, however, that fig. 3.14 has wider confidence intervals due to the lower number of segments, thus making its tolerance for observed correlations higher. Nevertheless, increasing the interval length does seem to make the segment means less correlated, and $N_{sub} = 4000$ seems to give segment means which are close enough to zero correlation. The reader should note that segments of length 4000 are used only for the estimation using already optimized rotation speeds. For the optimization of the rotation speed, the covariance between segments is not of concern and we use the segment length 600 (one minute).

Regarding the normal probability plots, the pattern is less clear. For three out of the four investigated sea states, there seems to be good adherence to a normal distribution regardless of interval length, though the deviation in the tails appears slightly stronger for $N_{sub} = 1000$. For the second sea state, however, there are once again more problems. There is a clear positive skewness with the highest means much higher than would be expected were the means Gaussian. This problem almost seems to be accentuated for $N_{sub} = 4000$, but this could well be due to the smaller number of means in

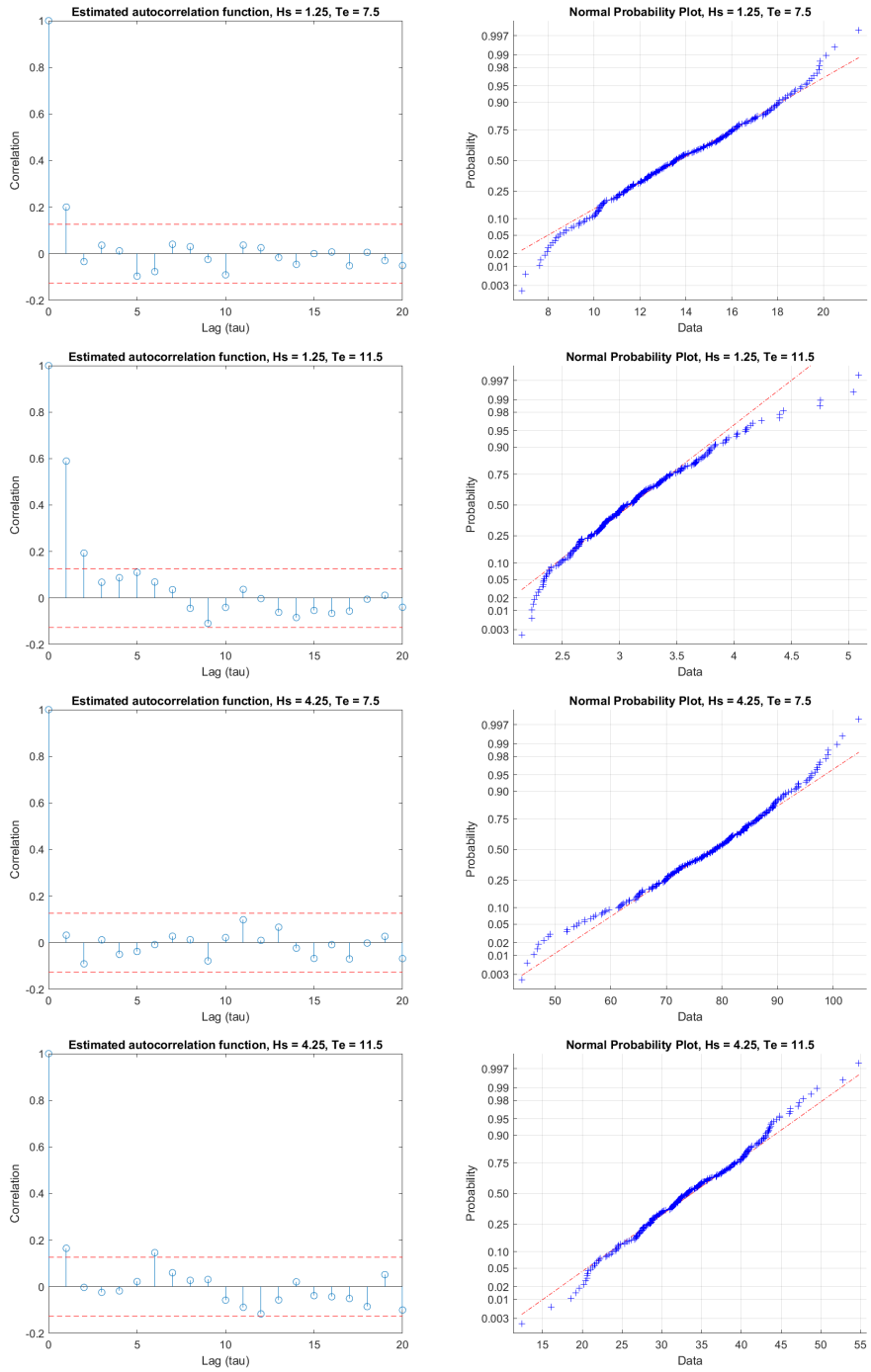


Figure 3.13: Empirical autocorrelation function for lag $\tau$ (left), and normal probability plot (right), for the series of $K = 240$ interval means, with interval length $N_{sub} = 1000$. Sea states are: Row 1 - $H_s = 1.25$, $T_e = 7.5$. Row 2 - $H_s = 1.25$, $T_e = 11.5$. Row 3 - $H_s = 4.25$, $T_e = 7.5$. Row 4 - $H_s = 4.25$, $T_e = 11.5$

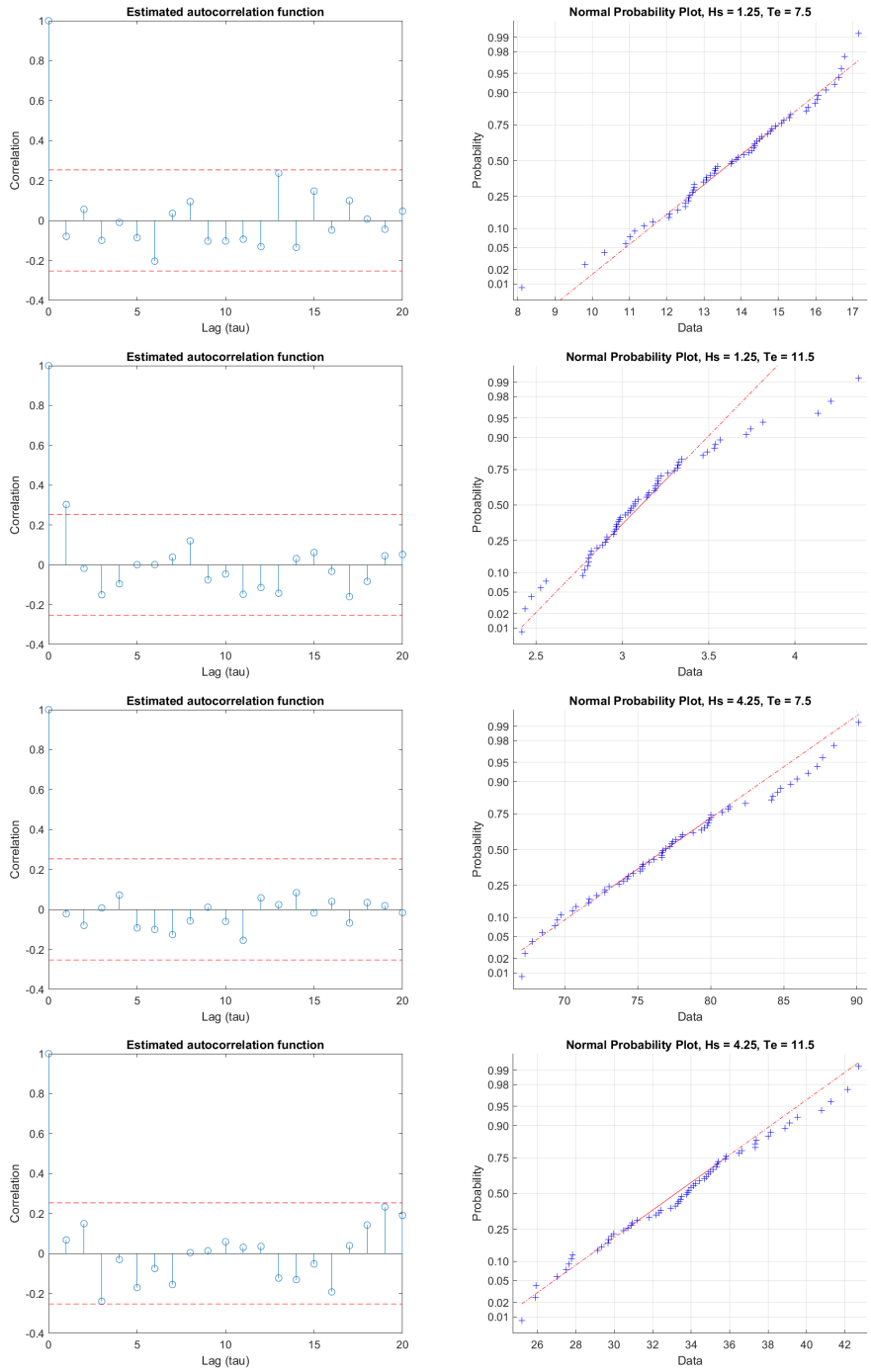


Figure 3.14: Empirical autocorrelation function for lag $\tau$ (left), and normal probability plot (right), for the series of $K = 60$ interval means, with interval length $N_{sub} = 4000$. Sea states are: Row 1 - $H_s = 1.25$, $T_e = 7.5$. Row 2 - $H_s = 1.25$, $T_e = 11.5$. Row 3 - $H_s = 4.25$, $T_e = 7.5$. Row 4 - $H_s = 4.25$, $T_e = 11.5$

the plot. Nevertheless, the $K$ segment means are not quite Gaussian, which casts doubt on whether the global mean could be.

Empirical investigations of other sea states show that the issues of non-Gaussian interval means and lingering autocorrelations are more prevalent when the significant wave height is low and the energy period long. The autocorrelation problem in particular seems worst for the most extreme sea state in this regard ($H_s = 1.25$, $T_e = 13.5$), but becomes less prevalent as we increase the significant wave height or shorten the energy period. The non-Gaussian means follow a less clear pattern, appearing to in general increase for longer energy periods but not showing an increase or decrease as the significant wave height changes. Unfortunately, the interval means do not seem to become more Gaussian for these sea states as the interval length increases. Nevertheless, due to the very large number of observations included in the mean of the full process, we will trust that in accordance with the central limit theorem even the means for these sea states will converge sufficiently.

We choose to use a simulated process of 120 000 observations (excluding the burn-in) to optimize the rotation speed using the maximization statistic in eq. (3.25) and calculating the sample standard deviation from the means of 600-observation segments. This optimization process also checks whether the WEC should use one generator or both in the sea state at hand, based on which choice leads to the greatest optimization statistic for the process at hand. Secondly, the optimal speed is taken as a fixed parameter for a new process of 240 000 observations (again, excluding the burn-in). In a real-life application, any optimized rotation speed will be derived from a random sample of waves which may not be representative for the sea state's average behaviour. The real waves which the rotation speed is used on would then not be the same as those used to optimize the value. Estimating the mean output using the same data would, in fact, lead to overestimation since the rotation speed has been fit to that particular data set, and would likely perform slightly less optimally on new data. In order to get more correct estimates, we therefore use a new stochastic process for the mean power output estimation.

This new process and its sub-interval means will be used to construct a confidence interval for the mean power output. Expressed mathematically (with $k$ being the sea state),

$$\begin{aligned}\hat{r}_{opt}(k) &= \arg\max_r \left(\hat{\mu}_{P_{out}}(r, \mathbf{X}(k)) - \lambda \cdot \hat{\sigma}_{P_{out},600}(r, \mathbf{X}(k))\right) \\ &= \arg\max_r \left( \frac{1}{120000} \sum_{i=1}^{120000} P_{out}(r, \mathbf{X}_i(k)) \right. \\ &\quad \left. - \lambda \sqrt{\frac{1}{199} \sum_{j=1}^{200} \left( \frac{1}{600} \sum_{i=1}^{600} P_{out}(r, \mathbf{X}_{600(j-1)+i}(k)) - \frac{1}{120000} \sum_{i=1}^{120000} P_{out}(r, \mathbf{X}_i(k)) \right)^2} \right)\end{aligned} \tag{3.45}$$

$$m_{segmented}(j) = \frac{1}{240000/K} \sum_{i=1}^{240000/K} P_{out}(\hat{r}_{opt}(k), \mathbf{X}_i(k)) \tag{3.46}$$

$$\hat{\mu}_{P_{out}}(k) = \frac{1}{K} \sum_{j=1}^{K} m_{segmented}(j). \tag{3.47}$$

This estimator has expected value

$$\begin{aligned}\mathbb{E}[\hat{\mu}_{P_{out}}(k)] &= \frac{1}{K}\sum_{j=1}^{K}\mathbb{E}[m_{segmented}(j)] = \mathbb{E}[m_{segmented}(j)] \\ &= \frac{1}{240000/K}\sum_{i=1}^{240000/K} P_{out}(\hat{r}_{opt}(k), \mathbf{X}_i(k)) \\ &= \mathbb{E}[P_{out}(k)] = \mu_{P_{out}}(k)\end{aligned} \tag{3.48}$$

and if we assume that the covariance between sub-intervals becomes negligible as the interval length increases, we get the asymptotic result

$$\begin{aligned}\mathbb{V}[\hat{\mu}_{P_{out}}(k)] &\approx \frac{1}{K^2}\sum_{j=1}^{K}\mathbb{V}[m_{segmented}(j)] \approx \frac{1}{K}\mathbb{V}[m_{segmented}(j)] \\ &= \frac{1}{K}\frac{1}{(240000/K)^2}\mathbb{V}\left[\sum_{i=1}^{240000/K} P_{out}(\hat{r}_{opt}(k), \mathbf{X}_i(k))\right] \\ &= \frac{K}{240000^2}\mathbb{V}\left[\sum_{i=1}^{240000/K} P_{out}(\hat{r}_{opt}(k), \mathbf{X}_i(k))\right]\end{aligned} \tag{3.49}$$

where the variance of a sub-interval cannot be further evaluated without fully exploring the covariance structure of its $N_2/K$ terms. Instead, this variance will be estimated as

$$\begin{aligned}\hat{\mathbb{V}}[\hat{\mu}_{P_{out}}(k)] &= \frac{1}{K}\hat{\mathbb{V}}[m_{segmented}(j)] \\ &= \frac{1}{K}\sum_{j=1}^{K}(m_{segmented}(j) - \hat{\mathbb{E}}[m_{segmented}])^2 \\ &= \frac{1}{K}\sum_{j=1}^{K}(m_{segmented}(j) - \hat{\mu}_{P_{out}}(k))^2 \\ &= s^2_{segmented}\end{aligned} \tag{3.50}$$

and we get the confidence intervals

$$CI_{\mu_{P_{out}}(k)} = \left((\hat{\mu}_{P_{out}}(k) - \lambda_{\alpha/2}\sqrt{\frac{s^2_{segmented}(k)}{K}}), (\hat{\mu}_{P_{out}}(k) + \lambda_{\alpha/2}\sqrt{\frac{s^2_{segmented}(k)}{K}})\right). \tag{3.51}$$

It should be noted that a sample of 240 000 observations is equivalent to 24 000 seconds or 400 minutes, which is an unrealistically long duration for a sea state. However, this is not relevant to the problem since our goal is to use these samples for better estimation accuracy of the optimum rotations per minute and the time-average power output for each state. If we want the expected output over a period of time, we would then take a weighted average (by the probabilities of each state) of the outputs, and multiply this by the time period.

The overall mean power output, provided that only these 78 sea states may occur and that they do so with probabilities proportional to the ones given in table 3.1 ($\mathbb{P}(k)$), is

$$\hat{\mu}_{P_{out}} = \sum_{k=1}^{78}\hat{\mu}_{P_{out}}(k)\mathbb{P}^*(k) = \sum_{k=1}^{78}\hat{\mu}_{P_{out}}(k)\frac{\mathbb{P}(k)}{\sum_{k=1}^{78}\mathbb{P}(k)} = \sum_{k=1}^{78}\hat{\mu}_{P_{out}}(k)\frac{\mathbb{P}(k)}{0.9243} \tag{3.52}$$

though since the 48 excluded sea states have not been chosen at random, this estimator will be slightly biased. Its variance is

$$\begin{aligned}\mathbb{V}[\hat{\mu}_{P_{Out}}] &= \sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 \mathbb{V}[\hat{\mu}_{P_{Out}}(k)] \\ &= \sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 \frac{K}{N_2^2} \mathbb{V}\Big[\sum_{i=1}^{N_2/K} P_{out}(\hat{r}_{opt}(k), \mathbf{X}_i(k))\Big]\end{aligned} \tag{3.53}$$

since the data for each sea state is sampled independently of the other states. The variance is estimated as

$$\begin{aligned}\hat{\mathbb{V}}[\hat{\mu}_{P_{Out}}] &= \sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 \hat{\mathbb{V}}[\hat{\mu}_{P_{out}}(k)] \\ &= \sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 s^2_{segmented}(k)\end{aligned} \tag{3.54}$$

which gives the confidence interval

$$CI_{\mu_{P_{out}}} = \Bigg( \big(\hat{\mu}_{P_{out}} - \lambda_{\alpha/2} \sqrt{\sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 \frac{s^2_{segmented}(k)}{78K}}\big), \big(\hat{\mu}_{P_{out}} + \lambda_{\alpha/2} \sqrt{\sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 \frac{s^2_{segmented}(k)}{78K}}\big) \Bigg). \tag{3.55}$$

The corresponding estimates for the energy efficiency are

$$\hat{\mu}_{eff} = \sum_{k=1}^{78} \frac{1000 \hat{\mu}_{P_{out}}(k)}{8J(k)} \frac{\mathbb{P}(k)}{0.9243} \tag{3.56}$$

$$\hat{\mathbb{V}}[\hat{\mu}_{eff}] = \sum_{k=1}^{78} \big(\mathbb{P}^*(k)\big)^2 \frac{10^6 s^2_{segmented}(k)}{64J^2(k)} \tag{3.57}$$

with confidence intervals completely analogous to those for the mean power. The factors 1000 and $10^6$ are due to the output being measured in kW and the energy potential in W.

Variance estimates and confidence intervals for the rotation speeds may not be calculated since we only have one observation available for each sea state.

### 3.3.3 Variance estimation

It is also of interest to measure the variance of the power output for each sea state, given the optimal rotation speed. As already stated, the variance of the wave process segments is not proportional to the variance of the individual observations due to their significant autocorrelation. Since shorter intervals likely also have correlations, we may not use the longer 4000-observation segments as our data if we are interested in the variance between shorter segments. Instead, we will use the same interval length as when optimizing the mean with a variance penalty – the variance between 600-observation or 60-second intervals. Since a sea state usually lasts around 20 minutes, what we will measure is the variance between 20 adjacent one minute long segments.

The time series of 240 000 observations will then give 400 means for each sea state, which may be used to estimate the variance between 20 such segments, 20 times each. The 20 variance estimates

for each state may then be averaged for a final variance estimate for that state. This should be a consistent estimator as long as the process is stationary and second-order ergodic. If the interval means are Gaussian we may also construct confidence intervals using the $\chi^2$ distribution as in eq. (2.52). However, this is only valid if the segment means are indeed Gaussian. The estimator is mathematically written as (cf. eq. (2.49))

$$\hat{\sigma}^2_{minute} = \frac{1}{20}\sum_{l=1}^{20}\frac{1}{19}\sum_{j=1}^{20}\left(\frac{1}{600}\sum_{i=1}^{600} x_{i+600(j-1)+12000(l-1)}(k) - \frac{1}{12000}\sum_{i=1}^{12000} x_{i+12000(l-1)}(k)\right)^2. \tag{3.58}$$

If the segment means are found to be Gaussian but the variance estimates seem to agree better with a Gaussian distribution than with a Chi-square one, we instead construct Gaussian confidence intervals. If the means are not found to be Gaussian, we do not construct confidence intervals since we are not sure that the average sample variance will converge sufficiently to either distribution.

To investigate the Gaussianity of the power output, we will use the 400 segment means for each sea state. In case the mean power is found to be reasonably Gaussian, we will also use them to construct confidence intervals for the variance. The normality test used is the Shapiro-Wilk test, using the Matlab function provided by BenSaïda (2024). Since the sample size for the test will be 60, we expect it to have a reasonably high power for an individual sea state in accordance with the results of (Razali and Wah, 2011). However, 78 sea states are tested. A nominal significance level of $\alpha = 5\%$ gives a multiple-testing significance level of

$$\begin{aligned}
&\mathbb{P}(H_0 \text{ rejected for at least one sea state} | H_0 \text{ true for all sea states}) \\
&= 1 - \mathbb{P}(H_0 \text{ rejected for zero sea states} | H_0 \text{ true for all sea states}) \\
&= 1 - \left(\mathbb{P}(H_0 \text{ not rejected for single sea state} | H_0 \text{ true for single sea state})\right)^{78} \\
&= 1 - (1-\alpha)^{78} \\
&= 0.9817
\end{aligned} \tag{3.59}$$

meaning there is an overwhelming risk of falsely classifying the power output distribution of at least one state as non-Gaussian, even if all 78 are in reality Gaussian. Nevertheless, the desired result is for the null hypothesis *not* to be rejected, since this is what would validate using $\chi^2$-distribution confidence intervals for the variance. This means that we will actually be more conservative with our application of this confidence interval method, being likely to apply it to fewer sea states than we could have. In a situation where all sea states give Gaussian power output distributions, the probability of misclassifying $n$ of them as non-Gaussian is given by table 3.3. The probability of misclassifying more than five distributions is 0.195, and the probability of misclassifying more than ten is only 0.002. If the power outputs are generally Gaussian, we should be able to produce confidence intervals for the variance of the vast majority of them.

| $n$ | $\mathbb{P}(N = n)$ | $\mathbb{P}(N \leq n)$ |
|---|---|---|
| 0 | 0.018 | 0.018 |
| 1 | 0.075 | 0.093 |
| 2 | 0.152 | 0.246 |
| 3 | 0.203 | 0.449 |
| 4 | 0.200 | 0.649 |
| 5 | 0.156 | 0.805 |
| 6 | 0.100 | 0.905 |
| 7 | 0.054 | 0.956 |
| 8 | 0.025 | 0.984 |
| 9 | 0.010 | 0.994 |
| 10 | 0.004 | 0.998 |

Table 3.3: Probabilities (pointwise and cumulative) of $n$ distributions being misclassified as non-Gaussian in the Shapiro-Wilk test given that all 78 are Gaussian.

What also needs to be considered is the multiple-testing power, meaning the probability of rejecting the null hypothesis of Gaussianity for all sea states provided that they are all non-Gaussian. This cannot be calculated exactly since the power of the test is not known to us. Given a power $1 - \beta_k$ for sea state $k$, we have

$$\begin{aligned} &\mathbb{P}(H_0 \text{ rejected for all sea states} | H_0 \text{ false for all sea states}) \\ &= \Pi_{k=1}^{78}(1 - \beta_k) \end{aligned} \tag{3.60}$$

which is unknown. However, Razali and Wah (2011) found a power equal to 1 for the Shapiro-Wilk test when the true distribution was asymmetrical, for sample sizes much smaller than 1000. A single-test power of $1-\beta_k = 1$ for all sea states still gives a group power of 1, and for (homogeneous) single-test power $1 - \beta_k = 0.999$ the group power is $0.999^{78} = 0.925$, which on average would give six sea states mistakenly classified as Gaussian. This number would decrease as the individual tests' powers approach 1. When Razali and Wah used the Shapiro Wilk test on a symmetric distribution, namely Student's $t(7)$, the power was only 0.9937 for 1000 samples (Razali and Wah, 2011, p. 28), which would give a group power of 0.611 and 30 expected misclassifications. Since the $t$ distribution is quite similar to a Gaussian distribution, misclassifying between the two might also not be as serious as misclassifying a heavily skewed distribution as Gaussian.

For sea states where the power output is found to be Gaussian, we will construct confidence intervals for the variances either as

$$\begin{aligned} CI_{\sigma^2(k)} &= \left(k_1 \hat{\sigma}^2_{MC}(k), k_2 \hat{\sigma}^2_{MC}(k)\right) \\ k_1 &= 399/\chi^2_{\alpha/2}(399) \\ k_2 &= 399/\chi^2_{1-\alpha/2}(399) \end{aligned} \tag{3.61}$$

in accordance with eq. (2.52) or as Gaussian intervals

$$\begin{aligned} CI_{\sigma^2(k)} &= \left(s^2(k) + \phi_{\alpha/2}\sqrt{s^4(k) \cdot (\hat{\kappa}(k) - 1)/N)}, s^2(k) + \phi_{1-\alpha/2}\sqrt{s^4(k) \cdot (\hat{\kappa}(k) - 1)/N)}\right) \\ \hat{\kappa} &= \frac{\hat{\mu}^4}{s^4} \end{aligned} \tag{3.62}$$

using the result in eq. (2.53), where $\hat{\kappa}$ is an appropriate estimator of the kurtosis.

## 3.4 Regression using the sea state parameters as predictors

Besides estimating the optimal rotation speed and its associated expected power output for each sea state individually, it is of interest to find a general relations between these values and the parameters of the sea states (the significant wave height and the energy period). There are three variables whose relation to the sea state parameters are of particular interest.

- The choice of using the 50 kW generator, the 100 kW generator, or both.
- The optimal rotation speed for the generator(s), in rpm.
- The mean power output given the optimal rotation speed.

The first two of these relations are immediately relevant in the sense that they give easy ways to, given information regarding the current sea state, change the settings of the generators to maximize the power output, until the sea shifts to another state which then causes the generators to change settings again. The third relation is meaningful for explaining the general effect of the sea state on the potential power production of the WEC, but also for predicting the mean power output of other sea states not included in this study.

The fact that we cannot calculate the optimal rotation speed or the expected power output explicitly, even after conditioning on the sea state, makes a (linear) regression analysis appropriate for finding this relation. If analytical and exact values of the $\hat{r}_{opt}$ and $\hat{\mu}_{P_{out}}$ were available, we would be able to find a deterministic rule, though possibly a very complicated one, for how the optimal rotation speed and expected power relate to the sea, and there would be no random errors. In this situation, however, the values we have may be assumed to include a random error term,

$$\hat{r}_{opt} = r_{opt} + \epsilon_r \tag{3.63}$$

$$\hat{\mu}_{P_{out}} = \mu_{P_{out}} + \epsilon_m \tag{3.64}$$

and for the segment means,

$$m_{segmented}(i) = \mu_{P_{out}} + \epsilon_{m,i}. \tag{3.65}$$

Let a first model for sea state $k$ (with observation index $i$) be expressed as

$$\hat{\mu}_{P_{out},i}(k) = \beta_0 + \beta_1 H_{s,i}(k) + \beta_2 T_{e,i}(k) + \epsilon_i(k) \tag{3.66}$$

where we assume that the errors $\epsilon_i$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ for each observation $i$. We then have $\mathbb{E}[\hat{\mu}_{P_{out},i}(k)] = \beta_0 + \beta_1 H_{s,i} + \beta_2 T_{e,i}$ and $\mathbb{V}[\hat{\mu}_{P_{out},i}(k)] = \sigma^2$. As described in section section 2.6.1, estimates of the $\beta$ coefficients may be found easily using the least squares method. A completely analogous approach may be utilized to find a model for the optimal rotation speed.

However, it seems reasonable that increasing the significant wave height may have a different effect on the power (or the rotation speed) for different lengths of the energy period, or vice versa. For this reason, we will expand the regression model to include an interaction term,

$$\mu_{P_{out},i}(k) = \beta_0 + \beta_1 H_{s,i}(k) + \beta_2 T_{e,i}(k) + \beta_{1,2} H_{s,i}(k) T_{e,i}(k) + \epsilon_i. \tag{3.67}$$

To fit this regression model, we use the $K$ interval means for each of the 78 sea states. Thereafter, we run diagnostics on the residuals to check their Gaussianity, if there are any non-linear patterns, if they seem correlated with the predicted response values or the predictor variables, and if the variances

seem to differ. If necessary, we will redo the regression analysis with for instance transformed variables or weighted least squares, until the model assumptions appear to be fulfilled. In case of a weighted least squares approach, we will use the variances of the observations from each sea state to determine estimated state-wise weights $\hat{w}_i$, using the fact that we have discrete categories available for the predictor variables.

If we encounter several models which seem to fit the data well, we will compare them through their respective adjusted coefficients of determination ($R^2_{adj}$).

The model for the rotation speeds has the complication that some of the sea states might have rotation speeds optimized for only one generator rather than two, making the values incomparable. Therefore, we will fit separate linear regression models for those sea states using only the 50 kW generator, the 100 kW generator, or both, respectively. It will, however, be much more difficult to fit this regression model satisfactorily due to there only being one observation available for each sea state. It will also not be possible to set aside a random sample of these observations as validation data. For this reason, we will create new validation data using time series of 12 000 observations for optimization, and 48 000 observations for output estimation (in both cases the number is without the 500-observation burn-in period). These new data will be used to test our regression models of all kinds and avoid overfitting to the training data.

Furthermore, the choice of active generators is a categorical variable with three possible values, meaning a linear regression is not useful to model it. Rather, since the values are discrete but also clearly ordered – $50kW < 100kW < 50 + 100kW$ – an ordinal logistic regression model seems most reasonable. Here, the same selection of predictors are available as for the linear regressions, except that they are used for predicting probabilities of the three generator choices in accordance with eqs. (2.96) to (2.98). For the ordinal logistic regression model, we use the confusion matrix and its associated statistics for evaluating the model, both on the training and the validation data. We also plot the accuracy for different decision thresholds, as well as a ROC-curve in the case of a binary logistic regression.

# Chapter 4

# Results

## 4.1 Estimation and regression analysis of the optimal rotation speeds

The first result to be presented in this section is the one that is acquired first, namely the choice of generator(s) for each sea state and the optimal speed in rotations per minute (rpm). The rotation speed values, optimized over stochastic processes of 120 000 observations (after a 500-observation burn-in), are found in table 4.1 and also plotted as a heat map in fig. 4.1 for a better overview. 64 out of the 78 investigated sea states get the highest optimization statistic value when using both generators, while 13 perform best with only the 100 kW generator (marked with $^{**}$ in the table) and only one performs best with only the 50 kW generator (marked $^{*}$). A few patterns may be seen in the table: Firstly, among the sea states using both generators, the optimal speed increases for higher significant wave heights $H_s$ but decreases for longer energy periods $T_e$, though the size of these effects varies. The highest observed optimal rotation speeds are reached for $[H_s, T_e]$ values of $[6.25, 9.5]$, $[4.25, 7.5]$ and $[5.25, 8.5]$. One generator alone giving better results than two generally occurs for the lower significant wave heights and longer energy periods, and the shift from two generators to one (and from 100 kW to 50 kW) seems to occur when the optimal rotation speed would otherwise pass below 500, which is the minimum number of rotations per minute. The algorithm chooses the 50 kW generator only for one sea state, which has the longest included energy period for the lowest significant wave height. However, the patterns in the table are irregular and we sometimes see an increase in the rotation speed or a switch from one to two generators despite a decrease in $H_s$ or an increase in $T_e$. The starkest irregularity occurs for $H_s = 2.75$ m when the energy period increases from 12.5 to 13.5 seconds: The algorithm goes from choosing only the 100 kW generator to choosing to use both, and also increases the rotation speed from 526 to 918 rpm.

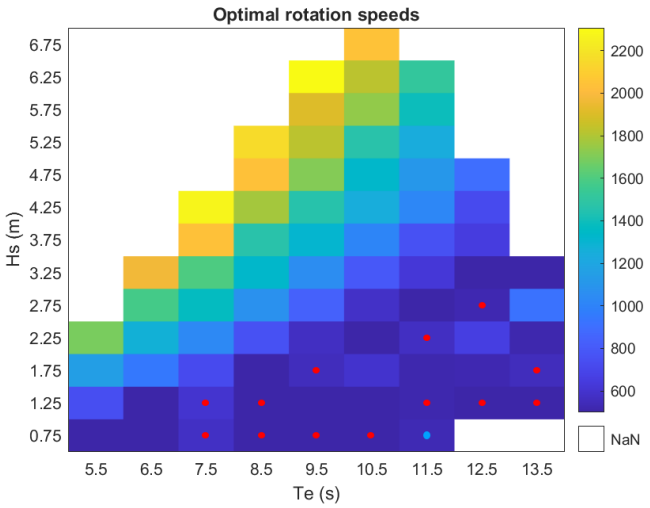


Figure 4.1: Heat map of the optimal rotation speed for each sea state, using the maximization statistic eq. (3.33). Sea states marked with a red dot use only the 100 kW generator, the sea state marked with a blue dot uses only the 50 kW generator, and all other states use both.

| Hs(m)/Te(s) | 5.5 | 6.5 | 7.5 | 8.5 | 9.5 | 10.5 | 11.5 | 12.5 | 13.5 |
|---|---|---|---|---|---|---|---|---|---|
| 6.75 | | | | | | 2 040.7 | | | |
| 6.25 | | | | | 2 306.8 | 1 824.8 | 1 513.2 | | |
| 5.75 | | | | | 1 904.6 | 1 735.8 | 1 392.7 | | |
| 5.25 | | | | 2 156.4 | 1 823.3 | 1 462.6 | 1 236.8 | | |
| 4.75 | | | | 2 035.9 | 1 710.7 | 1 327.2 | 1 112.4 | 899.9 | |
| 4.25 | | | 2 271.9 | 1 773.0 | 1 455.0 | 1 245.5 | 1 015.6 | 720.8 | |
| 3.75 | | | 2 037.6 | 1 466.5 | 1 308.2 | 1 013.3 | 760.7 | 650.5 | |
| 3.25 | | 1 968.7 | 1 600.6 | 1 327.6 | 1 052.2 | 789.3 | 622.2 | 501.7 | 500.1 |
| 2.75 | | 1 577.1 | 1 374.5 | 1 077.1 | 834.1 | 580.2 | 501.7 | 525.9** | 918.0 |
| 2.25 | 1 691.1 | 1 270.1 | 1 030.1 | 760.9 | 564.4 | 500.5 | 567.3** | 656.1 | 519.0 |
| 1.75 | 1 149.9 | 948.2 | 725.1 | 500.5 | 550.6** | 596.7 | 516.2 | 525.6 | 551.8** |
| 1.25 | 745.1 | 501.0 | 602.6** | 503.7** | 500.4 | 500.3 | 519.6** | 500.1** | 500.2** |
| 0.75 | 500.7 | 500.1 | 571.2** | 501.2** | 500.4** | 500.1** | 532.2* | | |

Table 4.1: The optimal generator rotation speeds (in rpm) for each of the 78 sea states, optimized from stochastic processes of length 120 000 with step size 0.1 seconds, and using eq. (3.33) as a maximization target statistic. Unmarked values indicate that the optimization chose to use both generators for the sea state, while those marked with * and ** (both in red) are chosen to use only the 50 kW or only the 100 kW generator, respectively.

No confidence intervals are produced for these optimum values, since we have no measure of their variance. Getting such a variance measure would require multiple optimizations for each sea state, which is very computationally heavy. We will, however, fit a linear regression model to the rotation

| Parameter | Estimate | Standard Error | $p$ value |
|---|---|---|---|
| $\zeta_{100}$ | -6.35 | 2.33 | 0.00645 |
| $\zeta_{150}$ | -1.94 | 1.82 | 0.288 |
| $\beta_{Hs}$ | -3.06 | 0.881 | 0.000513 |
| $\beta_{Te}$ | 0.635 | 0.225 | 0.00485 |

Table 4.2: Parameter estimates for the ordinal logistic regression model using only the main effects of significant wave height and energy period.

| True/Predicted generator choice | 50 kW | 100 kW | 50+100 kW | Sum |
|---|---|---|---|---|
| 50 kW | 0 | 1 | 0 | 1 |
| 100 kW | 0 | 8 | 5 | 13 |
| 50+100 kW | 0 | 3 | 61 | 64 |
| Sum | 0 | 12 | 66 | 78 |

Table 4.3: Confusion matrix for the ordinal logistic regression model for choice of generators.

speed values. Checking the relative occurrence frequencies of the single-generator sea states in table 3.1, we find that they together make up 30.46% of the occurrences of all 126 sea states, or 24.76% of the occurrences of the 78 states included in this thesis. This is clearly not negligible and we need to take the generator choice into account in our analysis.

### 4.1.1 Logistic regression models

To predict the choice of generator(s), we fit an ordinal logistic regression model and find that using only the main effects gives a confusion matrix with an overall accuracy of 88.46%, or 69 out of 78 states correctly classified. The fit parameters are featured in table 4.2. The confusion matrix is featured in table 4.3, and we may see there that due to there only being one sea state where the 50 kW generator alone is recommended, this case is never predicted by the model but instead the 100 kW generator choice is predicted. The precisions for correctly predicting the 100 kW and double-generator choices are 61.54% and 95.31%, respectively.

We also try adding the interaction term, but this led to a decrease rather than an increase in the model's overall accuracy (to 87.18%), predicting one more of the double-generator observations as only 100 kW. Adding further terms such as the squared predictors gives the exact same results.

Since the model fails to predict the 50 kW case entirely, an option might be to disregard that case – in other words, treating it as a 100 kW case – and use a binary logistic model to predict when to use only the 100 kW generator and when to use both. We fit this model and get an accuracy of 89.74%, only a slight improvement which follows entirely from the 50 kW case no longer counting as misclassified. However, this model is easier to interpret and therefore easier to use for practical decision making. It is also, as we see in the following paragraphs, easier to tune for improved prediction. The fitted parameters are featured in table 4.4.

| Associated predictor | Parameter | Estimate | Standard Error | $p$ value |
|---|---|---|---|---|
| Intercept | $\beta_0$ | 1.90 | 1.83 | 0.297 |
| Significant wave height | $\beta_{Hs}$ | 2.80 | 0.831 | $7.38 \cdot 10^{-4}$ |
| Wave energy period | $\beta_{Te}$ | -0.589 | 0.221 | 0.00779 |

Table 4.4: Parameter estimates for the binary logistic regression model using only the main effects of significant wave height and energy period.

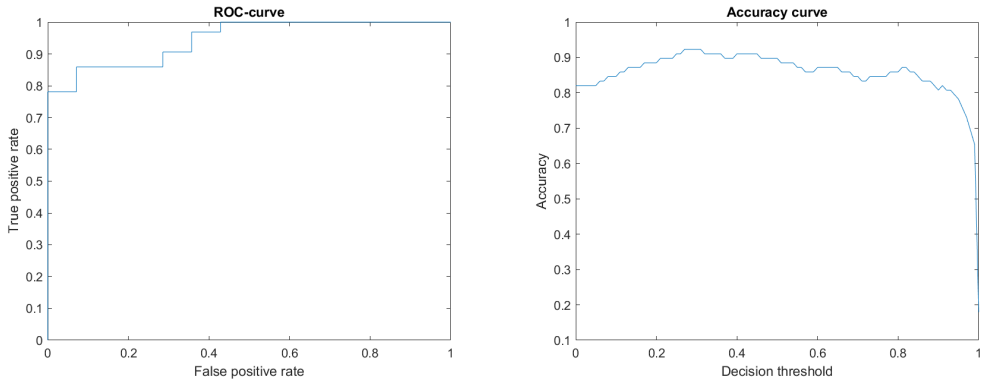


Figure 4.2: ROC and accuracy curves for the binary logistic model, using the training data and decision threshold values $0, 0.01, \ldots, 1$.

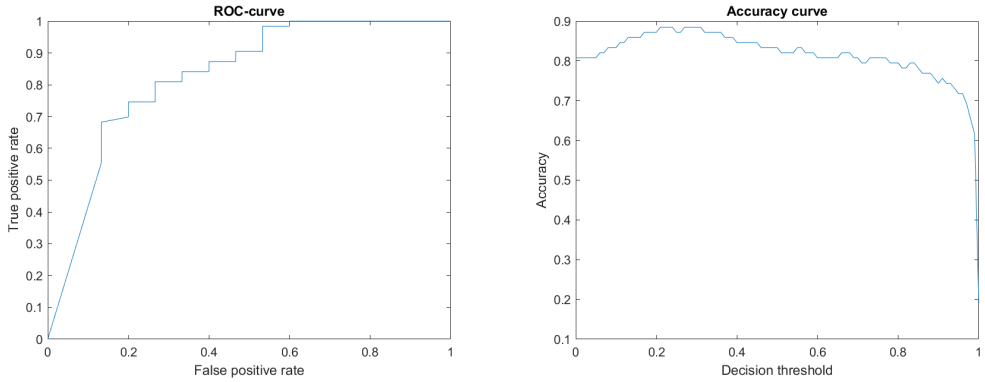


Figure 4.3: ROC and accuracy curves for the binary logistic model, using the validation data and decision threshold values $0, 0.01, \ldots, 1$.

The model's accuracy is based on the assumption that we predict the use of two generators if the predicted probability of two generators is above 0.5, and one (100 kW) generator otherwise. However, choosing a different threshold and thus being more or less inclined to predict one of the outcomes might well increase the accuracy. To investigate this, we create predicted outcomes based on threshold values $0, 0.01, \ldots, 1$ and investigate the performances. In fig. 4.2, we show both the ROC-curve for the different thresholds (false positive rate vs true positive rate) and the development of the accuracy. The latter plot suggests that a threshold at or slightly below 0.3 gives the highest accuracy, which is 92.31% on the training data.

In order to avoid overfitting, we also check the model's performance on the validation data and see that the threshold 0.3 is still optimal and gives an accuracy of 88.46%. The ROC and accuracy curves are plotted in fig. 4.3.

We may now derive a decision rule using the threshold 0.3 and the parameter estimates in table 4.4:

- If the model gives a probability of two generators greater than 0.3, use both generators.
- The probability is above 0.5 if the odds are greater than 0.429, which occurs when the log-odds – i.e. $\boldsymbol{\beta}\mathbf{X}$ – is above -0.368.
- $\beta_0 + \beta_{Hs} H_s + \beta_{Te} T_e > -0.368$ when $\beta_{Hs} H_s + \beta_{Te} T_e > -\beta_0$, i.e. when $2.80 H_s - 0.589 T_e > -2.27$.

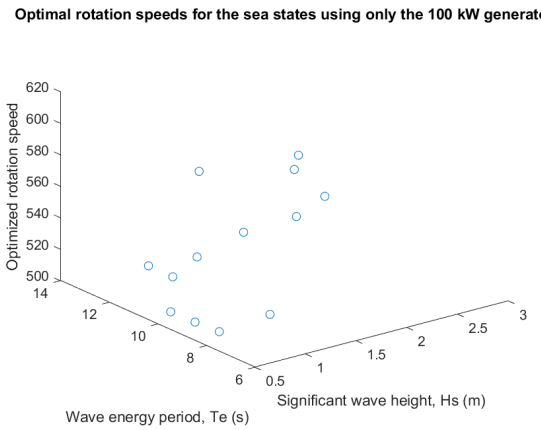


Figure 4.4: The observed optimal rotation speeds for the sea states using only the 100 kW generator.

- If $2.80H_s(k) - 0.589T_e(k) > -2.27$ for sea state $k$, use both generators for that sea state. Otherwise, use only the 100 kW generator.

This rule gives the most optimal decision in approximately 88.46% of cases, judging from the validation data.

### 4.1.2 Linear regression models for rotation speed

Having done this, we move on to the linear regression models for the optimal rotation speed, conditionally on the generator choice. For the single sea state where only one generator is active, it is meaningless to fit a regression model since there is only one observation. Furthermore, since the decision rule established in the previous section disregards the 50 kW cases entirely and treats them as 100 kW cases, any subsequent rule for setting the rotation speed will not be relevant for 50 kW cases.

For only the 100 kW generator, there are 13 observations, still making it difficult to fit a model properly, particularly since they do not vary much – they range from 500.07 to 602.64 rpm while the ones using both generators go as high as 2306.82 rpm. The values are plotted in fig. 4.4 and we note that the pattern looks very irregular and that it may therefore be difficult to fit a regression model to it. Unsurprisingly, it is difficult to find a model which is at all significantly better than simply using the mean of the observations. The only model which is barely significant is one which uses the significant wave height, the energy period and the squared energy period, but this model does not give satisfactory residual diagnostics. Thus, for the sea states using only the 100 kW generator we simply use the average optimal rotation speed, which is 530.50 rpm if we also include the one sea state where the 50 kW generator was recommended.

Finally, we have the 64 observations from the sea states where it was deemed optimal to use both generators. For these data, the best model we are able to fit given the restriction of only having one observation per sea state is one which uses the main effects of the two predictors, their interaction, the squared energy period and the interaction between energy period and squared significant wave height. The estimated parameters of this model are in table 4.5, and the residual diagnostics are featured in fig. 4.5. While the model assumptions do not seem entirely fulfilled – particularly, there is a large outlier for $H_s = 2.75$ m, $T_e = 13.5$ s – we are not able to find a better model given the limited data. This model still gives an adjusted coefficient of determination of 0.969, and all parameters are highly significant. We will therefore move on to evaluating this model on validation

| Associated predictor | Parameter | Estimate | Standard Error | $p$ value |
|---|---|---|---|---|
| Intercept | $\beta_0$ | 1364.88 | 224.8 | $1.04 \cdot 10^{-7}$ |
| Significant wave height | $\beta_{Hs}$ | 1298.19 | 57.58 | $2.20 \cdot 10^{-30}$ |
| Wave energy period | $\beta_{Te}$ | -404.91 | 52.02 | $1.41 \cdot 10^{-10}$ |
| Interaction | $\beta_{Hs,Te}$ | -110.27 | 8.28 | $2.67 \cdot 10^{-19}$ |
| Squared energy period | $\beta_{Te2}$ | 27.69 | 2.97 | $3.81 \cdot 10^{-13}$ |
| Interaction between period and squared height | $\beta_{Hs2,Te}$ | 2.03 | 0.586 | 0.00102 |

Table 4.5: Parameter estimates for the linear regression model predicting rotation speed in the cases using both generators.

data, to check whether it has good predictive accuracy.

On the test data, the model gives an adjusted coefficient of determination of 0.827, a clear decrease but still a high number. The validation data diagnostics are featured in detail in appendix A. These residuals actually seem to behave better than they did for the training data, with the exception of a large outlier for the sea state $H_s = 6.25$ m, $T_e = 11.5$ s which has predicted rotation speed 1471 rpm (and observed speed 1513 rpm in the training data), but 2548 rpm observed in the validation data.

We then have a model for determining the rotation speed to set for any sea conditions when it is deemed appropriate to use both generators, using the information on significant wave height and wave energy period:

$$\hat{r}_{opt} = 1364.88 + 1298.19H_s - 404.91T_e - 110.27H_sT_e + 27.69T_e^2 + 2.03H_s^2T_e \tag{4.1}$$

## 4.2 Estimation and regression analysis of the mean power output

Using the optimal rotation speeds and generator choices from table 4.1 on 78 new stochastic processes of 240 000 observations each (excluding the burn-in), we get power output time series for each sea state. These are presented for four sea states, and for observations 192 000-240 000 (the last fifth of the time series, cropped for visibility) in fig. 4.6. Clearly, the power is higher for sea states with higher significant wave height and shorter energy period, but the variability also seems to increase significantly for these states, as seen in particular in the bottom row of the figure ($H_s = 4.25$ m).

We move on to the estimation of mean power over the entire 240 000-observation process. The point estimates and confidence intervals of the mean are presented in table 4.6, while the point estimates along with the corresponding efficiencies are also plotted in fig. 4.7. Each confidence interval is based on the variance of the means of 60 segments with a length of 4000 observations each. The estimates and their confidence intervals are also seen in fig. 4.8, ordered by the sea states' frequency of occurrence (descending). A three-dimensional plot of the same estimates and intervals against the sea state parameters is featured in fig. 4.9. We also compare these intervals to the mean power values acquired for the original processes of length 120 000, and find that the optimization data's mean power was outside the estimation data confidence intervals for 23 out of 78 sea states, or almost 29.5%. This is reduced to 14 states, or 18%, if we widen the confidence intervals to be based on 120 000 observations rather than 240 000, but this is still more observations outside the confidence interval bounds than we would expect. This implies that the confidence intervals are either made too narrow, or that the optimization data systematically deviate from the estimation data. Seven of the 14 observations are above the confidence interval, and seven are below.

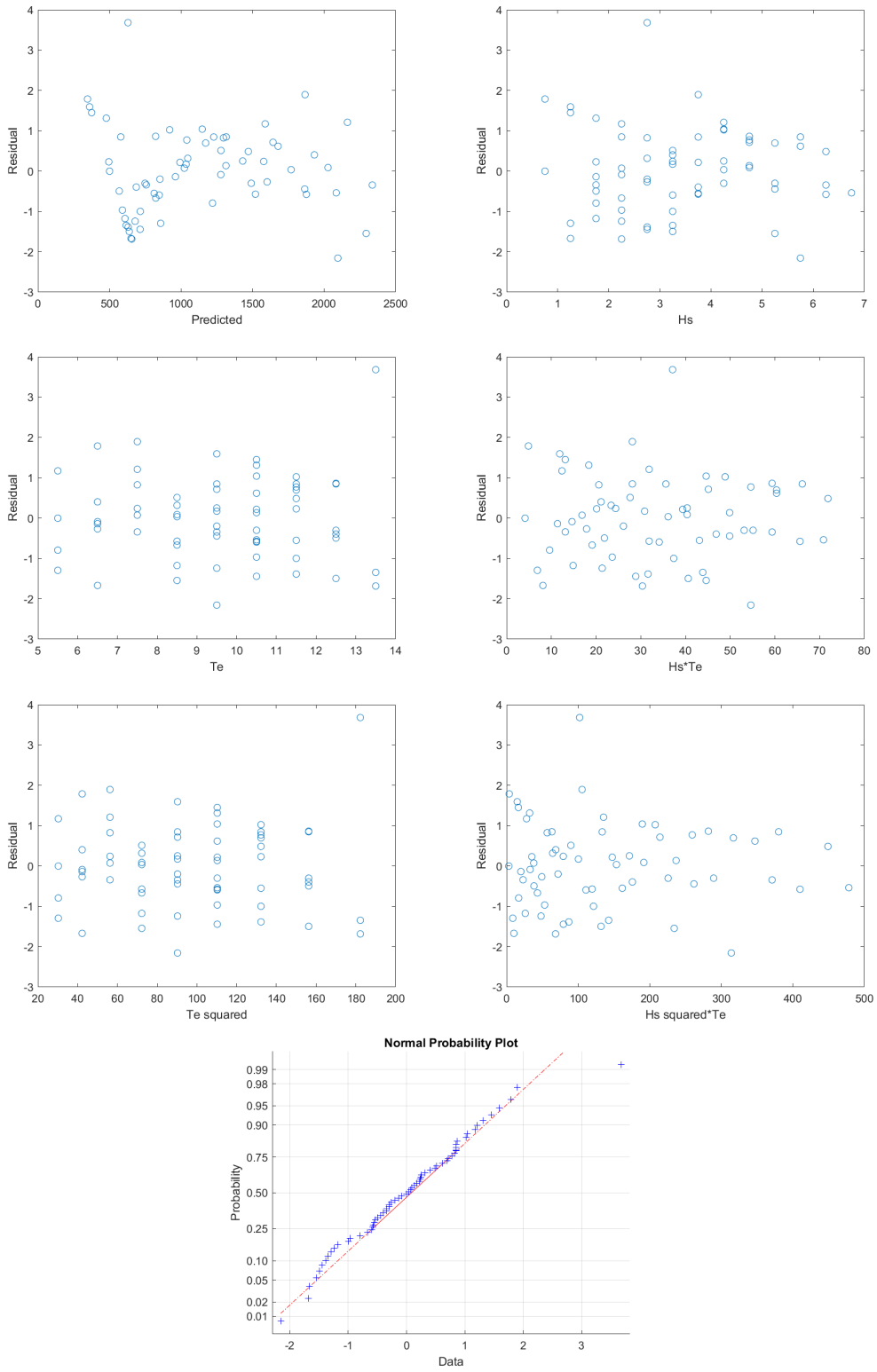


Figure 4.5: Residual diagnostics for the model predicting the rotation speed, for the 64 sea states using both generators.

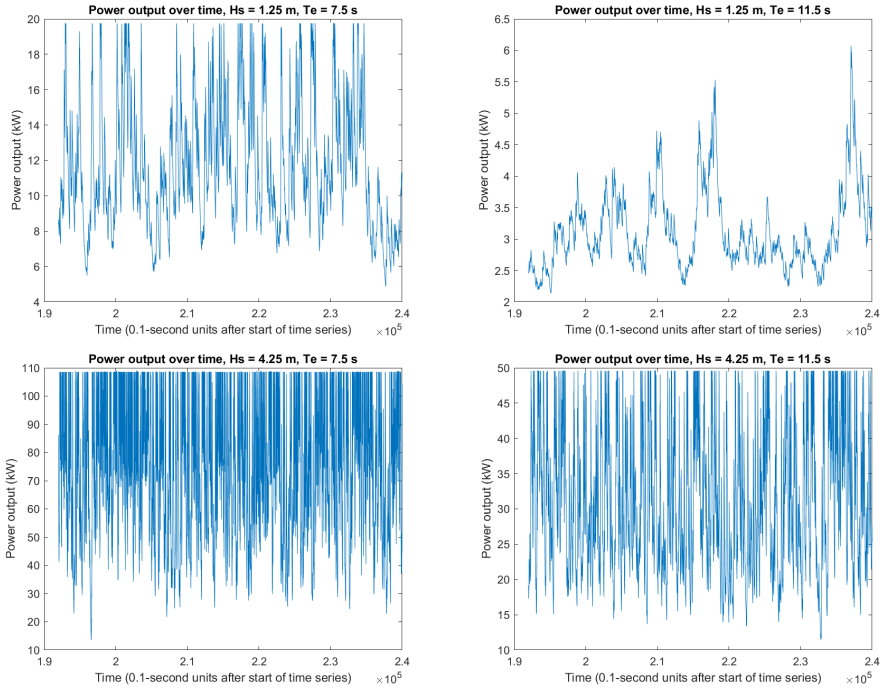


Figure 4.6: Observations 192 000 through 240 000 of the power time series for four sea states, using the optimized generator choices and rotation speeds. The sea states used have significant wave height $H_s$ equal to 1.25 for the top row and 4.25 for the bottom, while the energy period $T_e$ is 7.5 for the left column and 11.5 for the right.

| Hs(m)/Te(s) | 5.5 | 6.5 | 7.5 | 8.5 | 9.5 | 10.5 | 11.5 | 12.5 | 13.5 |
|---|---|---|---|---|---|---|---|---|---|
| 6.75 | | | | | | 65.7 [64.9, 66.6] | | | |
| 6.25 | | | | | 73.0 [72.2, 73.8] | 59.7 [58.9, 60.6] | 49.5 [48.7, 50.3] | | |
| 5.75 | | | | | 64.6 [63.9, 65.4] | 56.1 [55.2, 57.1] | 45.7 [45.1, 46.4] | | |
| 5.25 | | | | 73.0 [72.0, 74.0] | 59.8 [58.7, 60.8] | 48.7 [48.0, 49.4] | 40.0 [39.5, 40.6] | | |
| 4.75 | | | | 66.7 [65.7, 67.6] | 55.0 [54.1, 55.9] | 43.0 [42.2, 43.7] | 36.0 [35.2, 36.7] | 28.6 [28.1, 29.1] | |
| 4.25 | | | 74.1 [73.1, 75.2] | 57.8 [56.9, 58.7] | 47.4 [46.6, 48.3] | 39.9 [39.1, 40.6] | 31.7 [31.1, 32.3] | 23.3 [22.9, 23.6] | |
| 3.75 | | | 65.1 [64.1, 66.1] | 50.4 [49.7, 51.0] | 42.0 [41.3, 42.6] | 32.4 [31.8, 33.0] | 24.4 [23.9, 24.9] | 19.9 [19.5, 20.3] | |
| 3.25 | | 66.5 [65.3, 67.6] | 54.2 [53.3, 55.1] | 43.4 [42.7, 44.1] | 34.1 [33.4, 34.7] | 25.4 [24.9, 25.9] | 19.3 [18.9, 19.7] | 14.8 [14.4, 15.1] | 12.4 [12.1, 12.8] |
| 2.75 | | 54.1 [53.4, 54.9] | 44.0 [43.2, 44.8] | 34.5 [33.9, 35.2] | 26.3 [25.7, 26.8] | 18.5 [18.1, 18.9] | 14.7 [14.4, 15.0] | 10.2 [9.9, 10.4] | 8.2 [7.93, 8.54] |
| 2.25 | 54.4 [53.5, 55.3] | 43.2 [42.4, 43.9] | 33.6 [33.0, 34.3] | 24.5 [24.1, 25.0] | 18.2 [17.8, 18.7] | 14.0 [13.6, 14.3] | 10.1 [9.8, 10.4] | 7.87 [7.60, 8.13] | 6.39 [6.19, 6.59] |
| 1.75 | 37.8 [37.0, 38.6] | 31.2 [30.6, 31.8] | 22.9 [22.5, 23.4] | 15.9 [15.6, 16.3] | 11.0 [10.7, 11.3] | 9.17 [8.81, 9.52] | 6.72 [6.46, 6.98] | 4.86 [4.69, 5.03] | 3.61 [3.50, 3.72] |
| 1.25 | 22.8 [22.2, 23.4] | 16.5 [16.2, 16.9] | 11.9 [11.6, 12.2] | 8.69 [8.45, 8.93] | 6.49 [6.27, 6.71] | 4.55 [4.40, 4.71] | 3.07 [2.98, 3.17] | 2.32 [2.25, 2.39] | 1.79 [1.74, 1.84] |
| 0.75 | 9.85 [9.47, 10.23] | 6.88 [6.59, 7.18] | 4.77 [4.59, 4.95] | 3.31 [3.20, 3.43] | 2.04 [1.98, 2.11] | 1.40 [1.36, 1.43] | 1.05 [1.02, 1.08] | | |

Table 4.6: The estimated mean power output in kW for each of the 78 wave states, using the optimal rotation speeds and generator choices from table 4.1, plus corresponding 95% confidence intervals.

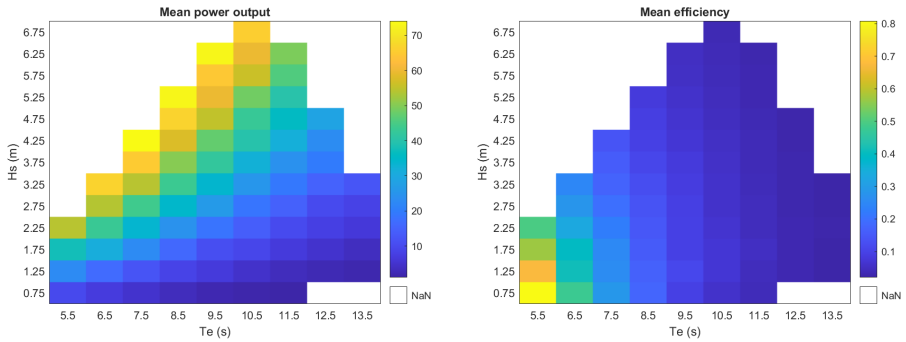


Figure 4.7: Heat maps of the mean power output and the mean efficiency for each sea state, using the rotation speeds and generator choices from table 4.1.

Overall, the mean power output follows a very similar pattern to the optimal rotation speed and the highest values are found for short energy periods and high significant wave heights. Increasing the energy period has a negative effect on the power whereas increasing the significant wave height has a positive effect, with a mean power point estimate range from 1.05 ($H_s = 0.75$, $T_e = 11.5$) to 74.13 ($H_s = 4.25$, $T_e = 7.5$).

Next, we calculate the wave-energy transport for each state using the expression in eq. (2.26). As stated previously, we use the fluid density 1027 $kg/m^3$ and the gravity acceleration constant 9.82 $m/s^2$. These energy potentials are not presented in detail, but range between 1.52 kW/m ($H_s = 0.75$ m, $T_e = 5.5$ s) and 235.65 kW/m ($H_s = 6.75$ m, $T_e = 10.5$ s). The power efficiency, estimated mean power divided by energy potential times width of the wave front, may then be calculated and ranges between 2.05% ($H_s = 2.75$ m, $T_e = 13.5$ s) and 80.82% ($H_s = 0.75$ m, $T_e = 5.5$ s). The efficiency is seen in fig. 4.7 to be highest for the states with short energy periods and *low* significant wave heights, reaching around 80% for $H_s = 0.75$ m, $T_e = 5.5$ and then quickly dropping down to around 50-60% for the adjacent sea states. For most states, the mean efficiency seems to be below 20% and even close to zero for the states with very long and/or high waves. This is likely due to the drastically increased energy potential in these waves, which the WEC is only capable of capturing to a very small degree.

The total, weighted average mean power output is estimated to 21.74 with a variance of 0.07. This gives us the confidence interval

$$I_{\mu_{P_{out}}} = [21.68, 21.81] \tag{4.2}$$

though, as stated before, this estimate has a slight bias due to the sea states excluded from the study. The weighted average efficiency is estimated to 21.50%, with a variance of $0.004699\%^2$. Finally, we get the following confidence interval for the efficiency:

$$I_{eff_{P_{out}}} = [0.2132, 0.2168]. \tag{4.3}$$

The calculations above use the rotation speeds optimized on 120 000 observations, but we also investigate if there is any difference in results if we use the generator choices and rotation speeds predicted by the regression models in section 4.1. We therefore try using these regression-derived generator choices and rotation speeds on time series of length 48 000 observations, and find the total weighted average power output. Interestingly, when estimating the output using the optimized rotation speeds and the validation data time series, we get an estimated mean output of 21.64 kW, a variance (between 400-second segments) of 0.0654, and a confidence interval of [21.48, 21.81] – slightly lower than when using the training data but not significantly different.

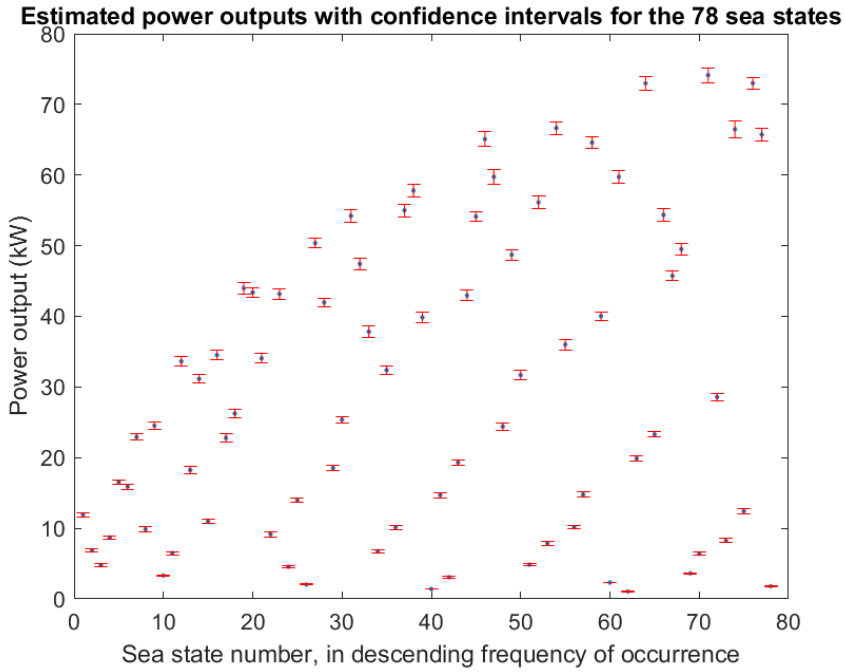


Figure 4.8: Point estimates and confidence intervals for the mean power output of all the 78 included sea states, sorted by frequency of occurrence.

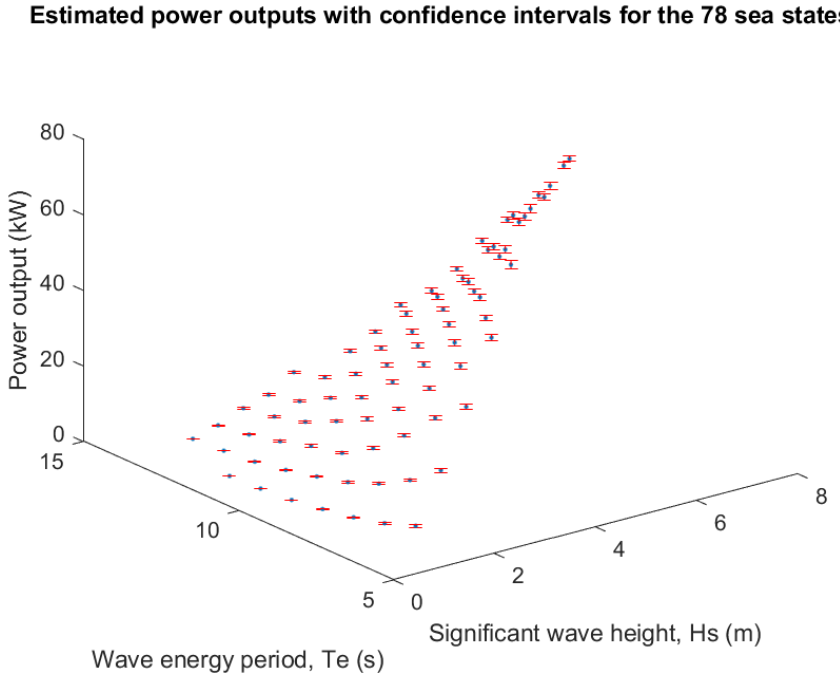


Figure 4.9: Point estimates and confidence intervals for the mean power output of all the 78 included sea states, by significant wave height and energy period.

When using the rotation speeds and generator choices given by the regression models, however, the output actually seems to increase: The point estimate is 21.98 kW, the variance (between 400-second segments) 0.0740 MW and the confidence interval [21.80, 22.15] kW. This is a significant increase from the mean output when using the optimized rotation speeds, regardless of whether we look at the training data or the validation data.

Another important comparison to make is between the above results and the ones acquired when we set $\lambda = 0$ in eq. (3.25) and thus only maximize the mean of the power output without regard for the variance, while using the same time series for optimization and estimation as when $\lambda = 2$. This process led to a mean output of 23.90 kW, meaning that the inclusion of the variance in our optimization led to an estimated decrease in the mean of $23.90 - 21.74 = 2.16$ kW, or 9%. The estimated 400-second variance is 0.1886 MW, giving a confidence interval of [23.79, 24.01] kW and a highly significant difference.

The efficiency when using $\lambda = 0$ is estimated to 23.04%, with a 400-second variance of $0.00655\%^2$ and a confidence interval of [0.2283, 0.2325]. Here, we have a decrease by $23.04 - 21.50 = 1.54$ percentage points, or 6.68%.

### 4.2.1 Ordinary linear regression model

Having the data of the segments' mean power, we now set to construct a regression model for explaining the power with the significant wave height and energy period as predictor variables. We use a random sample of the segment means as validation data, taking out 10 means for each sea state and keeping 50 as training data.

Since we have established that both parameters seem to have an impact on the mean power, the first model will include both parameters rather than only one of them. This model is thus

$$\hat{\mu}_{P_{out}} = \beta_0 + \beta_{Hs} H_s + \beta_{Te} T_e \tag{4.4}$$

and fitting it with least squares to gives the coefficients $\hat{\beta}_0 = 46.4$, $\hat{\beta}_{Hs} = 12.2$ and $\hat{\beta}_{Te} = -5.49$, all highly significant. The values also agree with our previous observations regarding the general effects of the two predictors, and the model has an adjusted $R^2$ value of 0.897. However, the residuals for this model show several problems. The residuals for this model are featured in appendix A and excluded here for brevity, but overall show problems such as non-linear residual patterns, heteroscedasticity and non-Gaussianity, meaning the model assumptions are not fulfilled.

Due to these problems with the initial model, and also due to the non-linear pattern observed in fig. 4.9, it seems reasonable to also include an interaction term between the two predictors as well as squared versions of each main predictor. We also improve the model further by adding an interaction between the *squared* significant wave height and the energy period. This gives the model

$$\hat{\mu}_{P_{out}} = \beta_0 + \beta_{Hs} H_s + \beta_{Te} T_e + \beta_{Hs2} H_s^2 + \beta_{Te2} T_e^2 + \beta_{Hs,Te} H_s T_e + \beta_{Hs2Te} H_s^2 T_e \tag{4.5}$$

which when fit with least squares has coefficient estimates $\hat{\beta}_0 = 16.9$, $\hat{\beta}_{Hs} = 55.7$, $\hat{\beta}_{Te} = -8.51$, $\hat{\beta}_{Hs,Te} = -4.23$, $\hat{\beta}_{Hs2} = -3.18$, $\hat{\beta}_{Te2} = 0.563$, and $\hat{\beta}_{Hs2,Te} = 0.306$. All coefficients are highly significant and the adjusted $R^2$ value has risen from 0.897 to 0.984. This model, however, still has problems in its residuals similar to the ones in the main effects model, as may be seen in detail in appendix A.

As discussed previously, these problems of heteroscedasticity and non-Gaussianity could potentially be amended by transforming either the response variable, the predictors, or both. The transforma-

| Associated predictor | Parameter | Estimate | Standard Error | $p$ value |
|---|---|---|---|---|
| Intercept | $\beta_0$ | 3.77 | 0.0760 | 0 |
| Significant wave height | $\beta_{Hs}$ | 5.45 | 0.0474 | 0 |
| Wave energy period | $\beta_{Te}$ | -0.761 | 0.0135 | 0 |
| Interaction | $\beta_{Hs,Te}$ | -0.328 | 0.00476 | 0 |
| Squared wave height | $\beta_{Hs2}$ | -0.513 | 0.00826 | 0 |
| Squared energy period | $\beta_{Te2}$ | 0.0364 | 0.000685 | 0 |
| Squared height and energy period interaction | $\beta_{Hs2,Te}$ | 0.0376 | 0.000811 | 0 |

Table 4.7: Parameter estimates for the linear regression model predicting the square root of the mean power using the two predictors, the interaction, the squared predictors and an interaction between squared height and period. Studentized residuals plotted against predicted values (top left), significant wave height $H_s$ (top right), wave energy period $T_e$ (second row left), interaction term $H_s \cdot T_e$ (second row right), $H_s$ squared (third row left), $T_e$ squared (third row right) and $H_s$ squared times $T_e$ (bottom left), as well as a normal Q-Q plot of the residuals (bottom right).

tion we use is taking the square root of the response $\mu_{P_{out}}$ to get the model

$$\sqrt{\hat{\mu}_{P_{out}}} = \beta_0 + \beta_{Hs}H_s + \beta_{Te}T_e + \beta_{Hs,Te}H_sT_e + \beta_{Hs2}H_s^2 + \beta_{Te2}T_e^2 + \beta_{Hs2,Te}H_s^2T_e \tag{4.6}$$

or, equivalently,

$$\hat{\mu}_{P_{out}} = (\beta_0 + \beta_{Hs}H_s + \beta_{Te}T_e + \beta_{Hs,Te}H_sT_e + \beta_{Hs2}H_s^2 + \beta_{Te2}T_e^2 + \beta_{Hs2,Te}H_s^2T_e)^2. \tag{4.7}$$

The coefficient estimates are now the ones featured in table 4.7, and the adjusted $R^2$ has risen even further to 0.991. The model assumptions are still not entirely fulfilled, though they are closer than the previous models. As may be seen in detail in appendix A, certain sea states still have slightly different residual variances.

The model seems to provide a reasonably good fit to the data when we make a 3D plot of the observed mean powers for each sea state, their confidence intervals and the plane representing the model's predictions. Such a plot is featured in fig. 4.10, and shows good agreement between the Monte Carlo estimations and the fitted regression model. However, we may note that for 38 of the 78 sea states the regression model estimations fall outside the Monte Carlo confidence intervals, 19 above and 19 below.

### 4.2.2 Weighted least squares model

As a final model, we perform a weighted least squares fit to the data. The weights are the inverse of the estimated variance of each sea state, but otherwise the model is identical to the one in eq. (4.6), using the square root of the mean power as response variable and multiple interaction and squared terms of the predictors. The residuals from this model are studentized using the square root of the weights, in accordance with eq. (2.93). These studentized residuals are plotted against the predictions as well as the predictor variables in fig. 4.11, together with a normal Q-Q plot.

For this model the residuals seem to agree with the model assumptions: The heteroscedasticity is almost gone and the residuals seem Gaussian. The adjusted coefficient of determination is 0.987, slightly lower than for the ordinary least squares model, but since that model did not quite fulfill the model assumptions we choose the WLS one. Its coefficients are found in table 4.8. We may also visually inspect the agreement of the predicted values to the Monte Carlo estimations and their corresponding confidence intervals. Here, 40 of the predicted values are outside the confidence intervals.

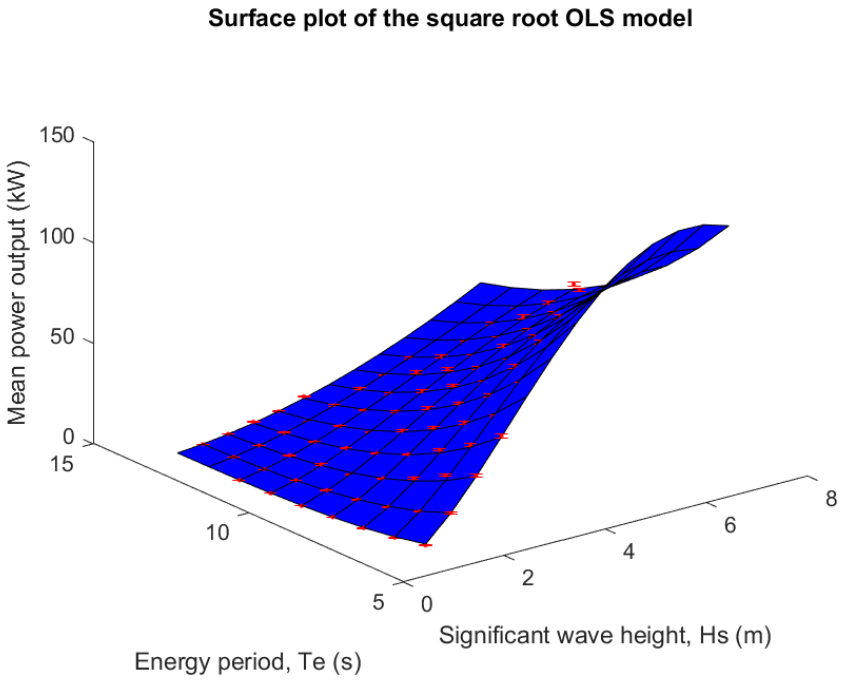


Figure 4.10: Surface plot (blue) of the mean power output values estimated by the function in eq. (4.6), together with Monte Carlo estimated mean values for each state (red dots) and corresponding 95% confidence intervals (red lines).

| Associated predictor | Parameter | Estimate | Standard Error | $p$ value |
|---|---|---|---|---|
| Intercept | $\beta_0$ | 3.91 | 0.0656 | 0 |
| Significant wave height | $\beta_{Hs}$ | 5.38 | 0.0408 | 0 |
| Wave energy period | $\beta_{Te}$ | -0.785 | 0.0108 | 0 |
| Interaction | $\beta_{Hs,Te}$ | -0.319 | 0.00391 | 0 |
| Squared wave height | $\beta_{Hs2}$ | -0.496 | 0.00737 | 0 |
| Squared energy period | $\beta_{Te2}$ | 0.0370 | 0.000511 | 0 |
| Squared height and energy period interaction | $\beta_{Hs2,Te}$ | 0.0358 | 0.000706 | 0 |

Table 4.8: Parameter estimates for the linear regression model using the weighted least squares method.

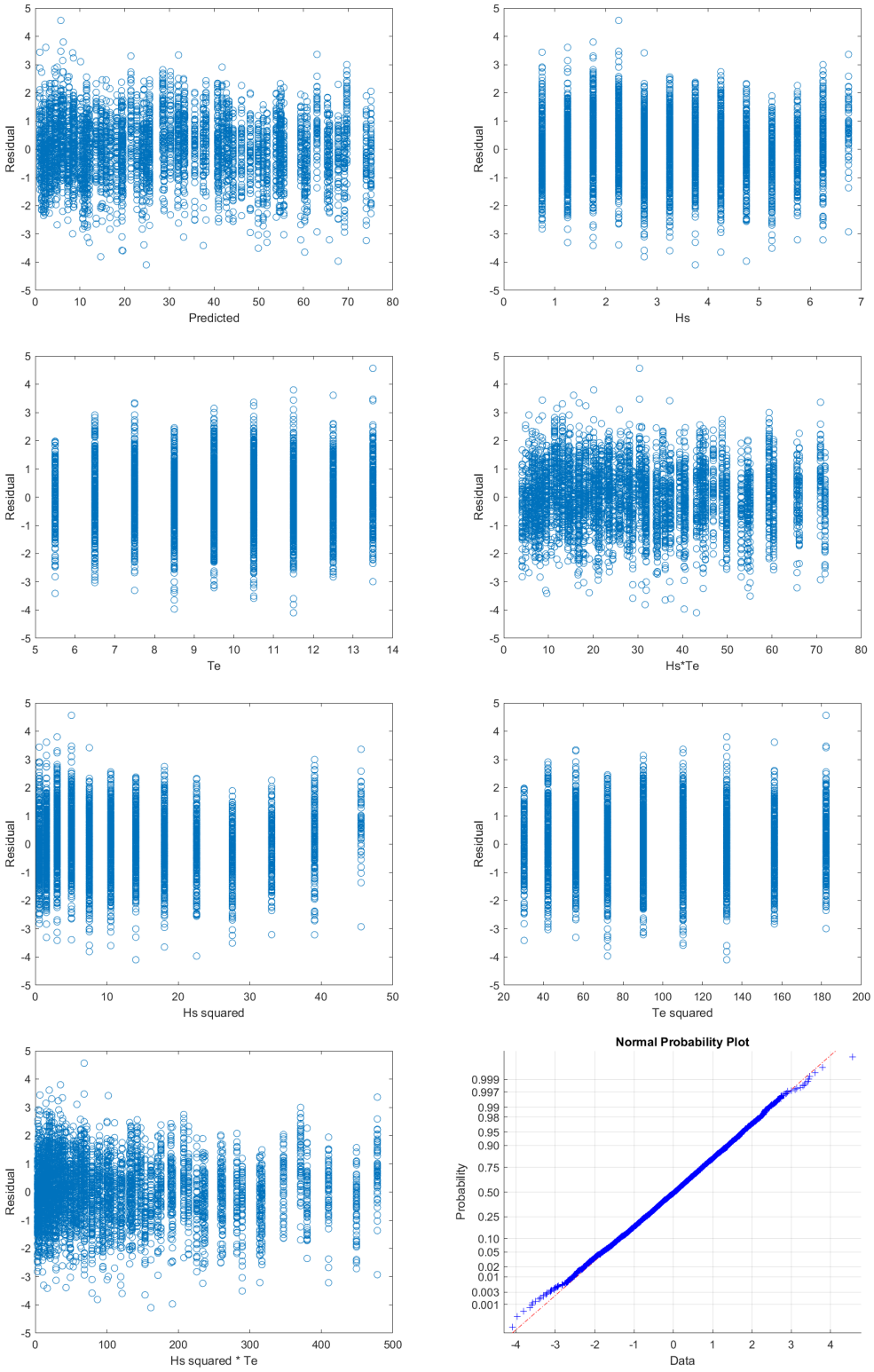


Figure 4.11: Residual diagnostics for the weighted least squares model. Studentized residuals plotted against predicted values (top left), significant wave height $H_s$ (top right), wave energy period $T_e$ (second row left), interaction term $H_s \cdot T_e$ (second row right), $H_s$ squared (third row left), $T_e$ squared (third row right) and $H_s$ squared times $T_e$ (bottom left), as well as a normal Q-Q plot of the residuals (bottom right).

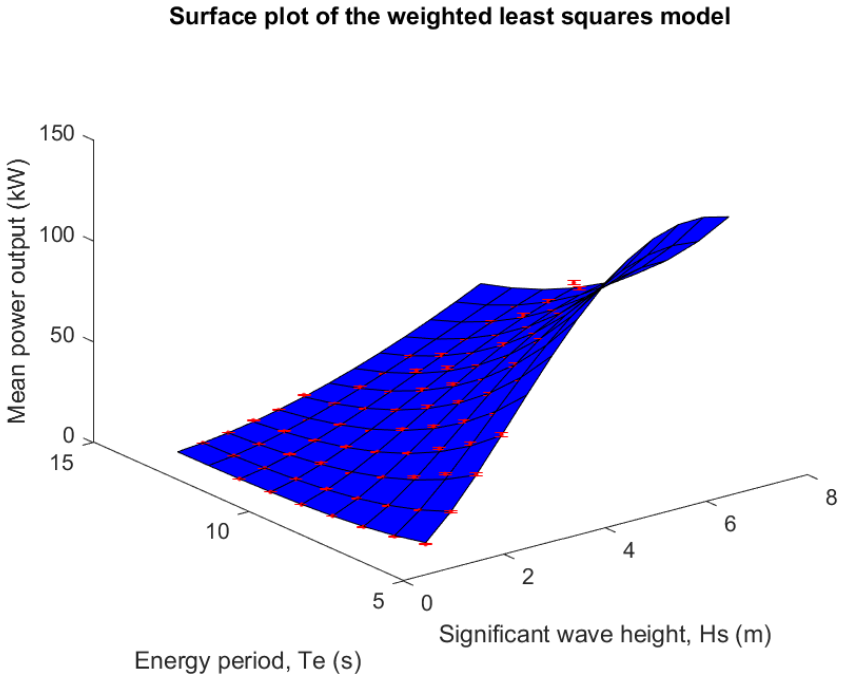


Figure 4.12: Surface plot (blue) of the mean power output values estimated by weighted least squares to the function in eq. (4.6), together with Monte Carlo estimated mean values for each state (red dots) and corresponding 95% confidence intervals (red lines).

To get a final evaluation of the model, we test its predictive power on the validation data and find that the model gives an adjusted coefficient of determination of 0.986, only marginally lower than the corresponding statistic for the training data. We also check the residuals on the validation data (shown in appendix A) and find that they are still satisfactory. We thus conclude that this model is well suited to predicting the mean power, and give the prediction rule

$$\hat{\mu}_{P_{out}} = (3.91 + 5.38H_s - 0.785T_e - 0.319H_sT_e - 0.496H_s^2 + 0.0370T_e^2 + 0.0358Hs^2T_e)^2. \quad (4.8)$$

## 4.3 Estimation of the power output variance

Finally, we present the estimates of the minute-to-minute variance in mean power output over a 20-minute period, found through averaging 20 initial estimates.

The Shapiro-Wilk test, run at significance level $\alpha = 0.05$, rejects the null hypothesis of Gaussianity for 77 of the 78 sea states' interval means, where the one non-rejection may well be due to chance. We conclude that the one-minute means are not Gaussian, and since we do not know whether the average of 20 such means will converge sufficiently this casts doubt on the validity of both Chi-square and Gaussian confidence intervals for the variances. We therefore refrain from writing out any such intervals and settle with providing point estimates. These are found in table 4.9 and also plotted in fig. 4.13 along with the corresponding variances in efficiency. The patterns in the variance estimates are similar to those of the mean estimates: The power variance is highest for sea states with high waves and short energy periods, reaching estimated values as high as 108 MW. The variance is clearly largest for the sea states with waves of short height and short energy periods, having a maximum value around 4% but ending up below 0.5% for the vast majority of sea states.

| Hs(m)/Te(s) | 5.5 | 6.5 | 7.5 | 8.5 | 9.5 | 10.5 | 11.5 | 12.5 | 13.5 |
|---|---|---|---|---|---|---|---|---|---|
| 6.75 | | | | | | 86.88 | | | |
| | | | | | | 9.32 | | | |
| 6.25 | | | | | 108.04 | 67.42 | 48.30 | | |
| | | | | | 10.39 | 8.21 | 6.95 | | |
| 5.75 | | | | | 63.29 | 73.84 | 43.44 | | |
| | | | | | 7.96 | 8.59 | 6.59 | | |
| 5.25 | | | | 89.63 | 69.10 | 48.04 | 35.13 | | |
| | | | | 9.47 | 8.31 | 6.93 | 5.93 | | |
| 4.75 | | | | 80.26 | 70.23 | 45.84 | 35.31 | 22.87 | |
| | | | | 8.96 | 8.38 | 6.77 | 5.94 | 4.78 | |
| 4.25 | | | 106.46 | 65.65 | 61.34 | 46.81 | 30.79 | 16.27 | |
| | | | 10.32 | 8.10 | 7.83 | 6.84 | 5.55 | 4.03 | |
| 3.75 | | | 93.23 | 44.41 | 47.06 | 32.90 | 20.78 | 15.01 | |
| | | | 9.66 | 6.66 | 6.86 | 5.74 | 4.56 | 3.87 | |
| 3.25 | | 87.70 | 57.93 | 47.70 | 34.96 | 21.79 | 14.46 | 9.11 | 8.62 |
| | | 9.36 | 7.61 | 6.91 | 5.91 | 4.67 | 3.80 | 3.02 | 2.94 |
| 2.75 | | 57.18 | 63.82 | 33.30 | 24.80 | 12.26 | 9.38 | 4.56 | 3.35 |
| | | 7.56 | 7.99 | 5.77 | 4.98 | 3.50 | 3.06 | 2.14 | 1.83 |
| 2.25 | 77.64 | 42.76 | 28.93 | 19.27 | 12.50 | 9.26 | 5.10 | 3.23 | 1.94 |
| | 8.81 | 6.54 | 5.38 | 4.39 | 3.54 | 3.04 | 2.26 | 1.80 | 1.39 |
| 1.75 | 35.18 | 26.47 | 19.26 | 9.57 | 4.94 | 5.56 | 2.40 | 0.870 | 0.323 |
| | 5.93 | 5.15 | 4.39 | 3.09 | 2.22 | 2.36 | 1.55 | 0.933 | 0.568 |
| 1.25 | 22.59 | 9.64 | 6.65 | 3.72 | 1.86 | 0.640 | 0.232 | 0.0928 | 0.0482 |
| | 4.75 | 3.10 | 2.58 | 1.93 | 1.36 | 0.800 | 0.482 | 0.305 | 0.220 |
| 0.75 | 6.24 | 2.77 | 0.862 | 0.315 | 0.0758 | 0.0216 | 0.0120 | | |
| | 2.50 | 1.67 | 0.928 | 0.561 | 0.275 | 0.147 | 0.110 | | |

Table 4.9: The estimated variances (upper rows) of the power output in MW, using the optimal rotation speeds and generator choices from table 4.1, and corresponding standard deviations (lower rows) in kW.

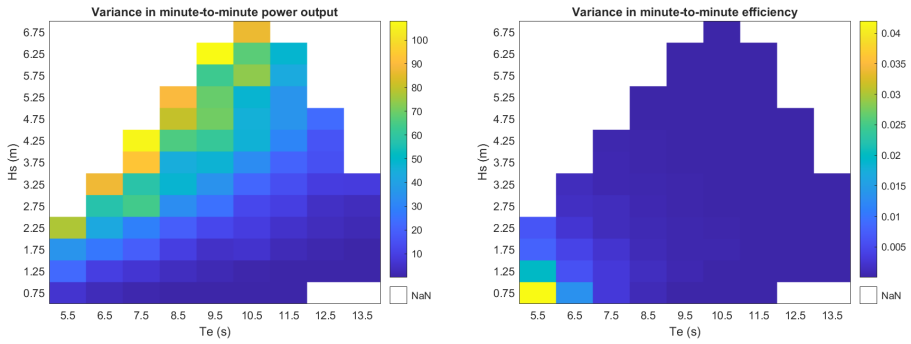


Figure 4.13: Heat maps of the variance in power output (MW) and efficiency for each sea state, using the rotation speeds and generator choices from table 4.1.

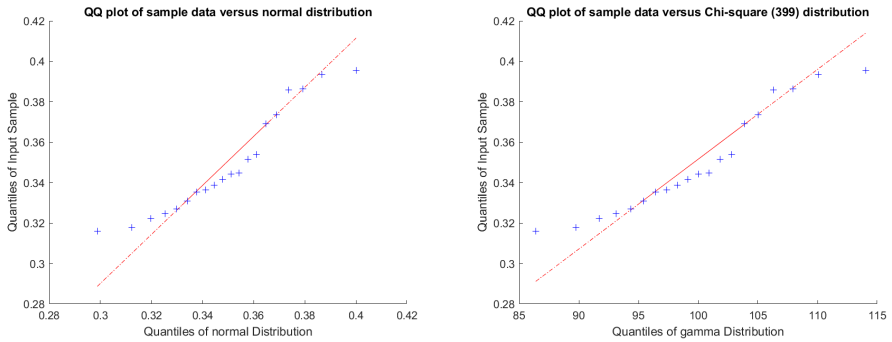


Figure 4.14: Q-Q plots for the 20 estimates of total power variance weighted across all sea states, comparing them the Gaussian and Chi-square (399) distributions.

The total (weighted average) variance has a point estimate of 0.349 MW. If we take the weighted average of the segment means across the sea states, the 400 resulting averages are found by the Shapiro-Wilk test to not deviate significantly from the Gaussian distribution ($p$-value 0.102). It may therefore be valid to construct a confidence interval for this variance using either the Gaussian or Chi-square distribution. If we take the weighted averages across all sea states to get 20 estimates of the total variance, we may check the empirical distribution of these 20 estimates. In fig. 4.14, we show Q-Q plots comparing it to both a Gaussian distribution and a Chi-square distribution with 399 degrees of freedom (since the estimates are based on the weighted averages of 400 segment means).

Neither distribution aligns perfectly with the observations, and there is essentially no difference between the fit. This is logical since by the central limit theorem, a Chi-square distribution with as many as 399 degrees of freedom should be very similar to a Gaussian distribution. Ultimately, we choose to construct a Chi-square confidence interval for the point estimate 0.349 MW, using eq. (2.52):

$$I_{\sigma^2_{P_{out}}} = [0.306, 0.404]. \tag{4.9}$$

The corresponding estimate of the standard deviation is 0.591 kW, with a confidence interval of

$$I_{\sigma_{P_{out}}} = [0.553, 0.635]. \tag{4.10}$$

Using Gaussian confidence intervals changes the interval limits by amounts of order $10^{-3}$ for both the variance and the standard deviation.

Rescaling the variance to efficiency as

$$\hat{\sigma}^2_{eff} = \frac{10^6 \cdot \hat{\sigma}^2_{P_{out}}}{8^2 \cdot J^2}, \tag{4.11}$$

we get the point estimate $1.37 \cdot 10^{-4}$, with corresponding standard deviation 0.0117. Unlike with the variance in power output, here the 400 weighted averages are found to be non-Gaussian ($p$-value around $10^{-5}$) and we refrain from constructing confidence intervals just as for the individual sea state variances.

When we instead use the generator settings (generator choices and rotation speeds) optimized on only the mean output, the total weighted average variance becomes 0.877 MW, with standard deviation 0.936 kW. For the efficiency, we have variance $2.08 \cdot 10^{-4}$ and standard deviation 0.0144.

The confidence intervals are found to be

$$I_{\sigma^2_{P_{out}}} = [0.767, 1.01] \tag{4.12}$$

$$I_{\sigma_{P_{out}}} = [0.876, 1.01] \tag{4.13}$$

$$I_{\sigma^2_{eff}} = [1.82 \cdot 10^{-4}, 2.40 \cdot 10^{-4}] \tag{4.14}$$

$$I_{\sigma_{eff}} = [0.0135, 0.0155] \tag{4.15}$$

and since the confidence intervals do not overlap with their equivalent intervals above which result from optimization on the statistic in eq. (3.33), we conclude that the differences are significant in all four cases and that the variability is clearly reduced when we penalize it in the optimization algorithm. The estimated decrease in standard deviation for the power output is 0.345 kW, or in relative numbers 36.87%. For the efficiency, the corresponding decrease is 0.272 percentage points, or 18.91%.

# Chapter 5

# Discussion

## 5.1 Conclusions

### 5.1.1 WEC generator settings

The first set of conclusions in this thesis concerns the settings of the WEC:s generators for potential practical use, giving rules for the use of either one or two generators as well as a way to determine the rotation speed from the current significant wave height and wave energy period.

Optimizing the rotation speeds and generator choices based on the minimization statistic $\mu^* = \mu - 2\sigma$ has given us 78 estimated optima and as many associated estimated mean outputs and variances. The results show that all three tend to decrease as we lessen the significant wave height $H_s$ or increase the wave energy period $T_e$. For the rotation speeds, however, the changes are much more irregular, frequently showing points where the optimum increases even when, for instance, the energy period increases. It is possible that this occurs due to local minima in the optimization algorithm. These irregularities do not propagate to the mean output estimates, but do make it more complicated to fit regression models to the rotation speeds and generator choices. They also do not consider the increased wear and tear upon the system resulting from using both generators rather than just one, which we will discuss later in this chapter.

Since the ordinal logistic regression model we initially tried to fit to the generator choices failed to ever predict the 50 kW case, we settle for the binomial logistic regression model, which after checking the accuracies following from various decision thresholds on both training and validation data, gave us the following decision rule which could potentially be relevant for practical implementation:

- Use both generators, 50 and 100 kW, if $2.80H_s - 0.589T_e > -2.27$, where $H_s$ is the significant wave height and $T_e$ the wave energy period. Otherwise, use only the 100 kW generator.

Furthermore, since the rotation speeds for the single-generator cases did not have a high degree of variation and no clear patterns which a regression model could capture, we found it more reasonable to simply use the average of all these rotation speeds for the cases where a single generator is chosen, namely 531.88 rpm.

- If the previous rule led to only the 100 kW generator being used, set the rotation speed to

530.50 rpm.

For the double-generator cases, it was more feasible to fit a linear regression model using the two predictors $H_s$ and $T_e$ as well as squared versions of them and interactions terms. The model found to give the most well-behaved residuals, as well as a high coefficient of determination for both training and validation data, was the following, also usable as a decision rule:

- If the generator decision rule led to both generators being used, set the rotation speed for both to $\hat{r} = 1398.38 + 1295.77H_s - 411.87T_e - 110.06H_sT_e + 28.05T_e^2 + 2.01H_s^2T_e$.

It is of note that the mean output resulting from using these settings were significantly better than those resulting from using the optimized settings, with point estimates of 21.98 and 21.64 kW on the validation data, respectively.

### 5.1.2 Energy output and efficiency

Secondly, we have results on the power output from the WEC. Like the rotation speed, both the mean and the variance of the output generally increase with higher wave heights and shorter energy periods, though with the variance there are some minor exceptions to this pattern. Taking a weighted average of the sea states by relative frequency of occurrence, we get point estimates and confidence intervals for the overall power output:

- The mean power output is estimated to $\hat{\mu}_{P_{out}} = 21.74$ kW, with 95% confidence interval $I_{\mu_{P_{out}}} = [21.68, 21.81]$.
- The variance among 20 one-minute averages in power output is estimated to $\hat{\sigma}^2_{P_{out}} = 0.349$ MW, with confidence interval $I_{\sigma^2_{P_{out}}} = [0.306, 0.404]$.
- The corresponding standard deviation is then estimated to $\hat{\sigma}_{P_{out}} = 0.591$ kW, with confidence interval $I_{\sigma_{P_{out}}} = [0.553, 0.635]$.

The estimated coefficient of variation is $0.591/21.73 = 0.0272$.

We may compare these numbers to those found when optimizing the generators with regards to only the mean output, and find that the inclusion of the variance in the optimization statistic through $\lambda = 2$ led to a significant decrease in both mean output and variance. The estimated decrease in mean is 2.16 kW (from 23.90 to 21.74 kW), or 9% of the $\lambda = 0$ result. For the standard deviation, the decrease is estimated to 0.345 kW (from 0.936 to 0.591 kW), or 37%. We may also calculate the coefficient of variation to 0.0392, meaning the reduction is 0.0120 percentage points or 30%. These results agree with the preliminary ones found in section 3.2.1 and specifically fig. 3.7, where on the shorter time series of length 12 000 we found that $\lambda = 2$ led to a decrease of 10% in mean and 40% in standard deviation. It is of note, though, that the decrease for both statistics is slightly smaller for the longer time series. This might be in part due to the fact that for the preliminary results, we used both generators for all sea states to save time. Using optimized generator choices might well lead to slightly higher mean output, and possibly also slightly higher variance.

The efficiency, meaning the power output relative to the energy potential of the relevant sea state, is estimated to 21.51%, with a 95% confidence interval of [0.2133, 0.2169]. This is slightly lower than the efficiency achieved by Zeinali et al. (forthcoming, p. 8), 24%. However, this efficiency is calculated on the mechanical energy entering the hydraulic system, it stands to reason that some

efficiency would be lost upon the conversion to electrical energy. A better statistic for comparison would be the efficiency achieved when setting $\lambda = 0$ in eq. (3.25), which is estimated to 23.04% with a confidence interval of [0.2283, 0.2325]. This is still slightly lower than the efficiency found by Zeinali et al. (forthcoming), likely due to the additional energy conversion, but is significantly higher than when using $\lambda = 2$. The decrease in efficiency is estimated to 1.54 percentage points, or 6.7%. The standard deviation in efficiency, on the other hand, had an estimated decrease of 0.272 percentage points (from 1.44% to 1.17%) or 19%. In other words, the relative decrease is smaller for the efficiency than for the power output, in terms of both mean and variance but particularly variance. This suggests that setting $\lambda = 2$ reduces the mean and variance more for those sea states where the output is high in absolute numbers but not in relation to the energy potential. Here, the coefficient of variation is reduced by 0.0081 or 13%.

The difference in reduction between power output and efficiency is logical since the optimization algorithm targets the output itself, with no regard for the energy potential, thus prioritizing the reduction of variance (and through this also slightly reducing the mean) in sea states with high output without considering whether the energy content of those states is even higher. An alternative approach could indeed be to use the efficiency, rather than the pure power output, in the optimization.

The mean output, too, was fit to a regression model. In this case, we settled for a weighted least squares linear regression model which gave residuals that seemed to fulfill the model assumptions and also explained a great deal of the variation in mean power among sea states – $R^2_{adj} = 0.987$ for the training data and 0.986 for the valdiation data. The model's regression equation may be written as

$$\hat{\mu}_{P_{out}} = (3.90 + 5.41H_s - 0.786T_e - 0.322H_sT_e - 0.503H_s^2 + 0.0372T_e^2 + 0.0365Hs^2T_e)^2. \quad (5.1)$$

## 5.2 Further research

This study has led to conclusions regarding the output of this particular wave energy converter, and thus joins the ranks of other case studies examining such performances by various types of WEC devices. Due to the great variety of solutions available in wave power, it is a challenge to draw general conclusions applying not only for the particular device of study. What this study shows for its own particular case is that the energy output is largest, but also has a greater variance, for higher and shorter – that is, steeper – waves. Through changing the optimization criteria used for choosing rotation speed of the WEC:s generator(s), one may reduce the variance of the output significantly while not decreasing the mean to the same degree. Through this, the time-average output weighted over all included sea states becomes very stable around its mean with a coefficient of variation no larger than 0.0271. The variance is measured as that of twenty one-minute interval means and on a shorter time-scale, the variance will naturally be higher.

An issue not covered in this thesis is the practical consideration of how the wear and tear upon the WEC is affected by the generator choices and rotation speeds. Using both generators simultaneously naturally leads to shorter lifespans for both of them, when one might get only a slightly lower effect using only the 100 kW generator but then over time reducing the cost of maintenance for the 50 kW one. A possible approach could be to set a threshold for how much higher the mean effect with double generators must be in order for that option to be chosen. Furthermore, a higher rotation speed itself should wear down the generator faster, making it interesting to in the future study the maintenance costs of the WEC in general and include these in an optimization algorithm.

One must also consider the fact that in reality, all sea states will not occur by their respective frequencies even on a middle-length time horizon. Sea states will last a number of minutes, seconds

or hours each and even in an entire day some states will likely be better represented due to weather variation. For this reason, an interesting field of further study would be the patterns of when the sea shifts from one state to another – when do we rule that the state has shifted, how often does this occur, and what states are more likely to follow after certain other states? Studying this as a Markov process, for instance, would give a better idea of the rate at which the time-average converges to the total mean output, and of the stability of the output over different time horizons. This naturally requires further study of the behaviour of the waves themselves, not only the WEC.

Another natural follow-up would be to implement the generator setting rules established in an actual WEC and investigating the results to see whether they align with the simulated results in this study, though this would naturally be much more costly in time and money.

# Bibliography


Abramowitz, M. and I.A. Stegun (1964). *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*. United States Department of Commerce, National Bureau of Standards.

Aderinto, T. and H. Li (2019). "Review on Power Performance and Efficiency of Wave Energy Converters". In: *Energies* 12 (22). DOI: 10.3390/en12224329.

Ahamed, R., K. McKee, and I. Howard (2020). "Advancements of wave energy converters based on power take off (PTO) systems: A review". In: *Ocean Engineering* 204. DOI: 10.1016/j.oceaneng.2020.107248.

Alves, M. (2016). "Numerical Modelling of Wave Energy Converters". In: ed. by Matt Folley. Academic Press. Chap. Frequency-Domain Models, pp. 11–30.

Astariz, S. and G. Iglesias (2015). "The economics of wave energy: A review". In: *Renewable and Sustainable Energy Reviews* 45, pp. 397–408. DOI: 10.1016/j.rser.2015.01.061.

Babarit, A., M. Guglielmi, and A. Clément (2009). "Declutching control of a wave energy converter". In: *Ocean Engineering* 36 (12-13), pp. 1015–1024. DOI: 10.1016/j.oceaneng.2009.05.006.

Babarit, A., J. Hals, et al. (2021). "Numerical benchmarking study of a selection of Wave Energy Converters". In: *Renewable Energy* 41, pp. 44–63. DOI: 10.1016/j.renene.2011.10.002.

Barstow, S. et al. (2008). "Ocean Wave Energy – Current Status and Future Perspectives". In: ed. by J. Cruz. Springer. Chap. The Wave Energy Resource, pp. 93–132.

BenSaïda, A. (2024). *Shapiro-Wilk and Shapiro-Francia normality tests*. https://se.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests. (Visited on 02/08/2024).

Blom, G. et al. (2017). *Sannolikhetsteori och statistikteori med tillämpningar*. Seventh. Studentlitteratur.

Brillouin, L. (1960). *Wave Propagation and Group Velocity*. Academic Press Inc.

Brodtkorb, P.A. et al. (2000). "WAFO - a Matlab Toolbox for the Analysis of Random Waves and Loads". In: *Proc. 10'th Int. Offshore and Polar Eng. Conf., ISOPE, Seattle, USA*. Vol. 3, pp. 343–350.

Burke, M. and E. Petraviciute (2021). *The Potential of Wave Power*. https://www.jdsupra.com/legalnews/the-potential-of-wave-power-9588054/. (Visited on 05/17/2024).

Chakrabarti, S.K. (2005). *Handbook of Offshore Engineering*. Elsevier.

Chong, Y.D. (2021). "Complex Methods for the Sciences". In: Nanyang Technological University. Chap. Fourier Series and Fourier Transforms. (Visited on 02/23/2024).

Conférence Générale des Poids et Mesures (1901). "Déclaration 2 de la 3e CGPM (1901)". In.

CorPowerOcean (2024). *Wave Energy*. https://corpowerocean.com/wave-energy/. (Visited on 05/20/2024).

Eckhardt, R. (1987). "Stan Ulam, John von Neumann and the Monte Carlo Method". In: *Los Alamos Science* (15), pp. 131–143. (Visited on 02/14/2024).

Edenhofer, O. et al. (2011). *Renewable Energy Sources and Climate Change Mitigation*. Tech. rep. IPCC.

Encyclopaedia Britannica (Dec. 30, 2023). *Wave — Properties, Characteristics Effects*. https://www.britannica.com/science/wave-water. (Visited on 03/19/2024).

Encyclopaedia Britannica (Mar. 7, 2024). *Wave — Behavior, Definition Types*. https://www.britannica.com/science/wave-physics. (Visited on 03/19/2024).

European Commission (2024). *Ocean energy*. https://research-and-innovation.ec.europa.eu/research-area/energy/ocean-energy_en. (Visited on 05/20/2024).

European Marine Energy Centre (2022a). *Billia Croo wave data base*.

— (2022b). *Our History*. https://www.emec.org.uk/about-us/emec-history/. (Visited on 02/09/2024).

— (2024). *Pelamis Wave Centre*. https://www.emec.org.uk/about-us/wave-clients/pelamis-wave-power/. (Visited on 05/07/2024).

Falcão, A. (2010). "Wave energy utilization: A review of the technologies". In: *Renewable and Sustainable Energy Reviews* 14 (3), pp. 899–918. DOI: 10.1016/j.rser.2009.11.003.

Falnes, J. and A. Kurniawan (2020). *Ocean Waves and Oscillating Systems*. Second. Cambridge University Press.

Givens, G. and J. Hoeting (2013). *Computational Statistics – Integration and Simulation*. John Wiley Sons, Inc.

Gut, A. (2009). *An Intermediate Course in Probability*. Springer.

Hannig, J. (2008). *Handout on Asymptotic Distribution of Sample Variance*. https://hannig.cloudapps.unc.edu/STOR655/stor655_handouts.html. Lecture handout. (Visited on 05/21/2024).

Hosmer, D., S. Lemeshow, and R. Sturdivant (2013). *Applied Logistic Regression*. 3rd. Wiley.

Industry, OUCO (2023). *Understanding Sea State: A Comprehensive Guide*. https://ouco-industry.com/understanding-sea-state-a-comprehensive-guide/. (Visited on 05/20/2024).

ITTC (2002). "The specialists committee on waves, final report and recommendations to the 23rd ITTC". In: *Proceedings of the 23rd ITTC*, pp. 505–736.

Jakobsson, A. (2021). *An Introduction to Time Series Modeling*. Fourth. Studentlitteratur.

Jayaswal, V. (Sept. 14, 2020). *Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score*. https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262. (Visited on 04/29/2024).

*Wave energy: history, implementations, environmental impacts, and economics* (2022). SPIE,the international society for optics and photonics. DOI: 10.1117/12.2646119. (Visited on 05/17/2024).

Lighthill, J. (1978). *Waves In Fluids*. Cambridge University Press.

Lindgren, A. (2023). *MASM22/FMSN30: Linear and Logistic Regression: Lecture 11, spring 2023, Multinomial and ordinal logistic regression, Quantile regression*. https://canvas.education.lu.se/courses/23000/pages/lecture-10-12-other-regressions?module_item_id=823560. Lecture notes. (Visited on 05/07/2024).

Lindgren, G. (2013). *Stationary Stochastic Processes – Theory and Appilcations*. CRC Press.

Lindgren, G., H. Rootzén, and M. Sandsten (2014). *Stationary Stochastic Processes for Scientists and Engineers*. CRC Press.

MathWorks, Inc. (2024). *Set Optimization Options*. https://se.mathworks.com/help/matlab/math/setting-options.html#bt00189-1.

Metropolis, N. (1987). "The Beginning of the Monte Carlo Method". In: *Los Alamos Science* (15), pp. 125–130. (Visited on 02/14/2024).

Mollison, D. (1994). "Statistics for the Environment 2: Water Related Issues". In: ed. by V. Barnett and K. Feridun Turkman. John Wiley and Sons Ltd. Chap. Assessing the Wave Energy Resource, pp. 205–222.

NASA Earth Observatory (2003). *Earth's Gravity Field*. https://earthobservatory.nasa.gov/images/3666/earths-gravity-field. (Visited on 03/12/2024).

Neumann, J. von and R. Richtmyer (1947). *Statistical methods in neutron diffusion*. Tech. rep. Los Alamos Scientific Laboratory report LAMS-551. Los Alamos Scientific Laboratory.

Olive, D. (2017). *Linear Regression*. Springer.

Razali, N.M. and Y.B. Wah (2011). "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling". In: *Journal of Statistical Modeling and Analytics* 2 (1), pp. 21–33. (Visited on 05/17/2024).

Ren, C., J. Tan, and Y. Xing (2023). "ALK-PE: An efficient active learning Kriging approach for wave energy converter power matrix estimation". In: *Ocean Engineering* 286. DOI: 10.1016/j.oceaneng.2023.115566.

Ritchie, H. and P. Rosado (2024). *Electricity Mix*. https://ourworldindata.org/electricity-mix. (Visited on 05/17/2024).

Sahngun Nahm, F. (2021). "Receiver operating characteristic curve: overview and practical use for clinicians". In: *Korean Journal of Anesthesiology*, pp. 25–36. DOI: 10.4097/kja.21209.

Service, National Weather (2024). *Beaufort Scale*. https://www.weather.gov/pqr/beaufort. (Visited on 05/20/2024).

Shao, X. et al. (2023). "A comparison of two wave energy converters' power performance and mooring fatigue characteristics – One WEC vs many WECs in a wave park with interaction effects". In: *Journal of Ocean Engineering and Science* 8, pp. 446–460. DOI: 10.1016/j.joes.2023.07.007.

Shapiro, S.S. and M.B. Wilk (1965). "An Analysis of Variance Test for Normality (complete samples)". In: *Biometrika* 52 (3 and 4), pp. 591–611. DOI: 10.2307/2333709.

Sheather, S. (2009). *A Modern Approach to Regression with R*. Springer.

Soch, J. (2024a). *Definition: Chi-squared distribution*. https://statproofbook.github.io/D/chi2. DOI: 10.5281/ZENODO.4305949. (Visited on 05/20/2024).

— (2024b). *Definition: F-distribution*. https://statproofbook.github.io/D/f. DOI: 10.5281/ZENODO.4305949. (Visited on 05/20/2024).

Statista (2024). *Net electricity consumption worldwide in select years from 1980 to 2022 (in terawatt-hours)*. https://www.statista.com/statistics/280704/world-power-consumption/. (Visited on 05/06/2024).

Ting, K.M. (2010). "Encyclopedia of Machine Learning". In: ed. by C. Sammut and G. Webb. Springer. Chap. Confusion Matrix, p. 209. (Visited on 05/17/2024).

United Nations (2024). *What is renewable energy?* https://www.un.org/en/climatechange/what-is-renewable-energy. (Visited on 05/20/2024).

United States Energy Information Administration (Aug. 10, 2022). *Wave power*. https://www.eia.gov/energyexplained/hydropower/wave-power.php. (Visited on 05/20/2024).

WAFO-group (2017). *Wafo – a Matlab toolbox for analysis of random waves and loads*. https://github.com/wafo-project. (Visited on 01/30/2024).

Wave Generation, IAHR Working Group on and Analysis (1989). "List of Sea-State Parameters". In: *Journal of Waterway, Port, Coastal, and Ocean Engineering* 115 (6), pp. 727–849. DOI: 10.1061/(ASCE)0733-950X(1989)115:6(793).

Waves4Power (2024a). *The WaveEL System*. https://www.waves4power.com/waveel/. (Visited on 05/16/2024).

— (2024b). *Waves4Power – The Solution*. https://www.waves4power.com/waves4power-the-solution/. Website changed and specific page removed, accessible through Wayback Machine. (Visited on 02/12/2024).

Webb, Paul (2019). *Introduction to Oceanography*. Roger Williams University.

Wu, T.L. (2023). *The Pros and Cons of Wave Energy*. https://earth.org/the-pros-and-cons-of-wave-energy/. (Visited on 05/17/2024).

Zeinali, S. et al. (forthcoming). "Optimizing the Hydraulic Power Take-Off System in a Wave Energy Converter". In: *Ocean Engineering*. Submitted 2024-04-15.

# Appendix A

# Additional residual diagnostics

This appendix contains the residual diagnostics for the regression models presented in chapter 4 but not deemed relevant enough to include in the main text.

## A.1 Regression models for rotation speed

Fig. A.1 shows the residuals on validation data for the linear regression model predicting the optimal rotation speed for the sea states using both generators. The predictors are the main effects (significant wave height and wave energy period), an interaction term between them, the squared wave energy period, and the interaction between squared significant wave height and non-squared wave energy period. The fitted coefficients for this model are $\hat{\beta}_0 = 1364.88$, $\hat{\beta}_{Hs} = 1298.19$, $\hat{\beta}_{Te} = -404.91$, $\hat{\beta}_{Hs,Te} = -110.27$, $\hat{\beta}_{Te2} = 27.69$ and $\hat{\beta}_{Hs2,Te} = 2.03$. The residuals seem to agree with the model assumptions except for a large outlier for the sea state $H_s = 6.25m, T_e = 11.5s$, which has predicted rotation speed 1471 rpm (and observed speed 1513 rpm in the training data), but 2548 rpm observed in the validation data.

## A.2 Regression models for mean power output

Fig. A.2 contains the residuals for the main effects model, using only the pure predictors (significant wave height and wave energy period) without any interactions or polynomial functions. The model has the fitted coefficients $\hat{\beta}_0 = 46.4$, $\hat{\beta}_{Hs} = 12.2$ and $\hat{\beta}_{Te} = -5.49$. We observe that there is a clear non-linear pattern in the residuals both for different levels of the predicted response and for different predictor values. There are also strong signs of heteroscedasticity, and the residuals do not seem entirely Gaussian but rather somewhat positively skewed.

Fig. A.3 shows the corresponding residual diagnostics for the regression model which includes, besides the main effects of the model above, the interaction term between the two predictors, the squared versions of them and the interaction between squared significant wave height and non-squared wave energy period. The coefficients in this model are fit to $\hat{\beta}_0 = 16.9$, $\hat{\beta}_{Hs} = 55.7$, $\hat{\beta}_{Te} = -8.51$, $\hat{\beta}_{Hs,Te} = -4.23$, $\hat{\beta}_{Hs2} = -3.18$, $\hat{\beta}_{Te2} = 0.563$ and $\hat{\beta}_{Hs2,Te} = 0.306$. The non-linearity in the residuals is not as strong as in fig. A.2, but there is a very large difference in the variance of the residuals depending on the level of the predicted response. The residuals are also still not quite

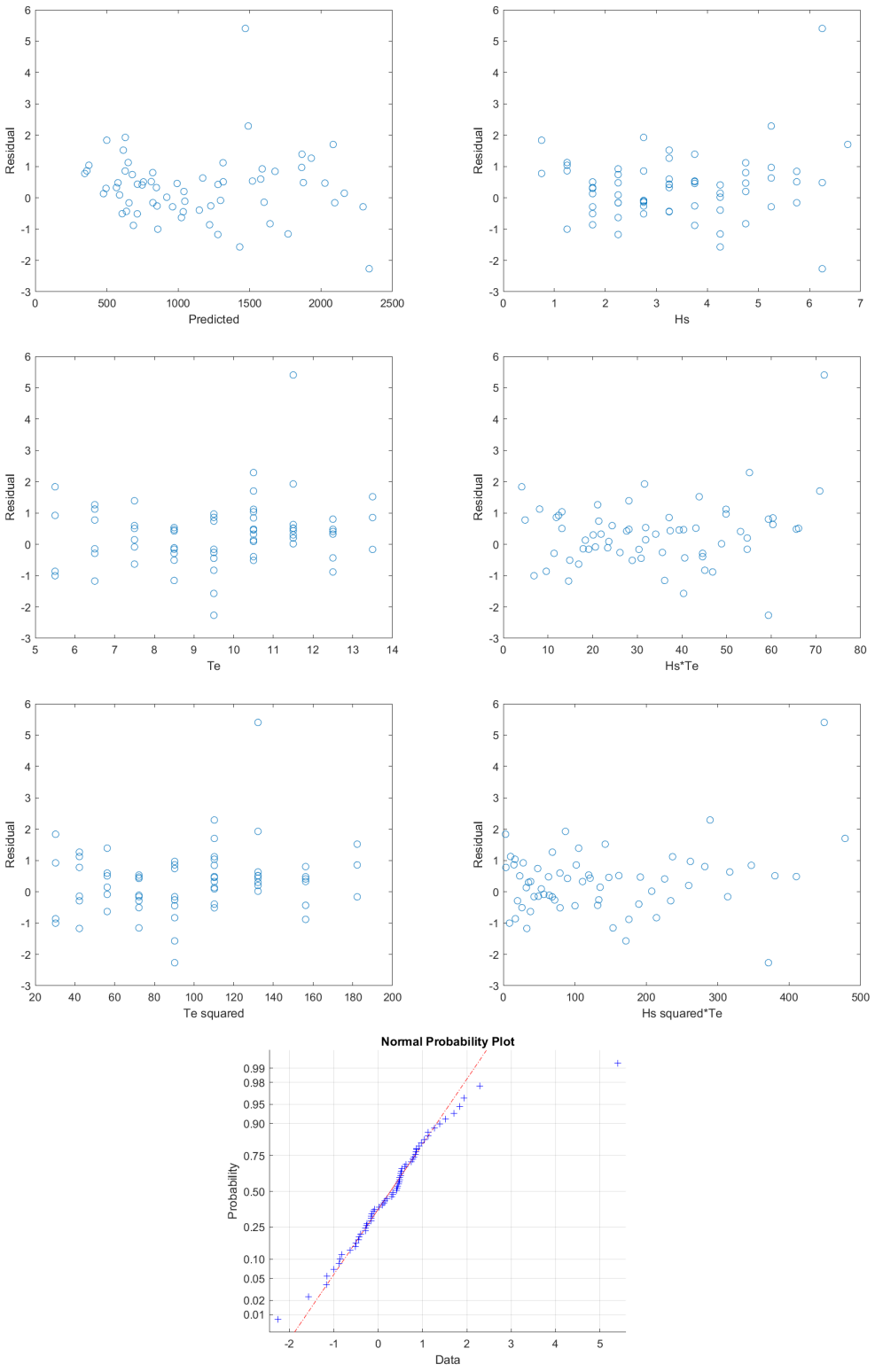


Figure A.1: Residual diagnostics on the validation data for the model predicting the rotation speed, for the 64 sea states using both generators.

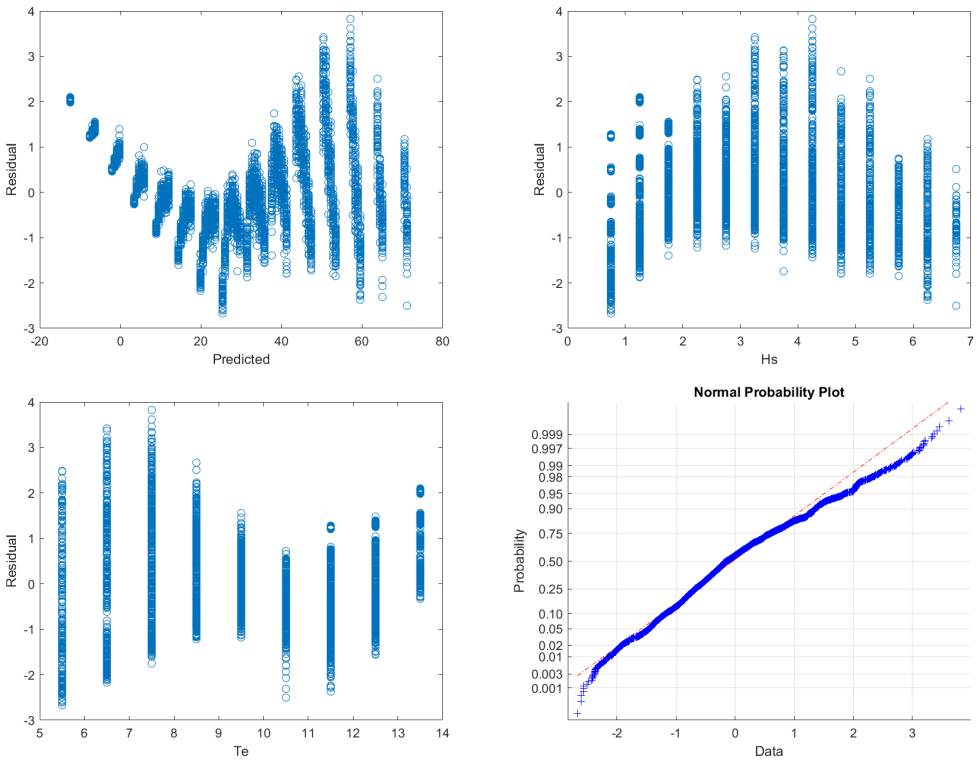


Figure A.2: Residual diagnostics for the model with only main effects from the two predictors. Studentized residuals plotted against predicted values (top left), against significant wave height (top right) and against wave energy period (bottom left), as well as a normal Q-Q plot of the residuals (bottom right).

Gaussian, now having heavy tails on both sides.

Fig. A.4 shows the residuals when using the same predictors as above to predict the square root of the mean power output. The coefficients for this model are estimated to $\hat{\beta}_0 = 3.77$, $\hat{\beta}_{Hs} = 5.45$, $\hat{\beta}_{Te} = -0.761$, $\hat{\beta}_{Hs,Te} = -0.328$, $\hat{\beta}_{Hs2} = -0.513$, $\hat{\beta}_{Te2} = 0.0364$ and $\hat{\beta}_{Hs2,Te} = 0.0376$. The residuals now seem even more Gaussian with only slightly heavy tails, and the heteroscedasticity has been much reduced. There is still a difference in the residual variance, however, perhaps most clearly for $T_e = 13.5$, $H_s = 0.75$ and generally the observations with small predicted values.

Fig. A.5 shows the residuals for the weighted least squares model on the validation data. The corresponding plots for the training data are shown in the main text, in chapter 4, and the fitted coefficients are $\hat{\beta}_0 = 3.91$, $\hat{\beta}_{Hs} = 5.38$, $\hat{\beta}_{Te} = -0.785$, $\hat{\beta}_{Hs,Te} = -0.319$, $\hat{\beta}_{Hs2} = -0.496$, $\hat{\beta}_{Te2} = 0.0370$ and $\hat{\beta}_{Hs2,Te} = 0.0358$. The residuals, like the ones from the training data, seem to agree well with the model assumptions.

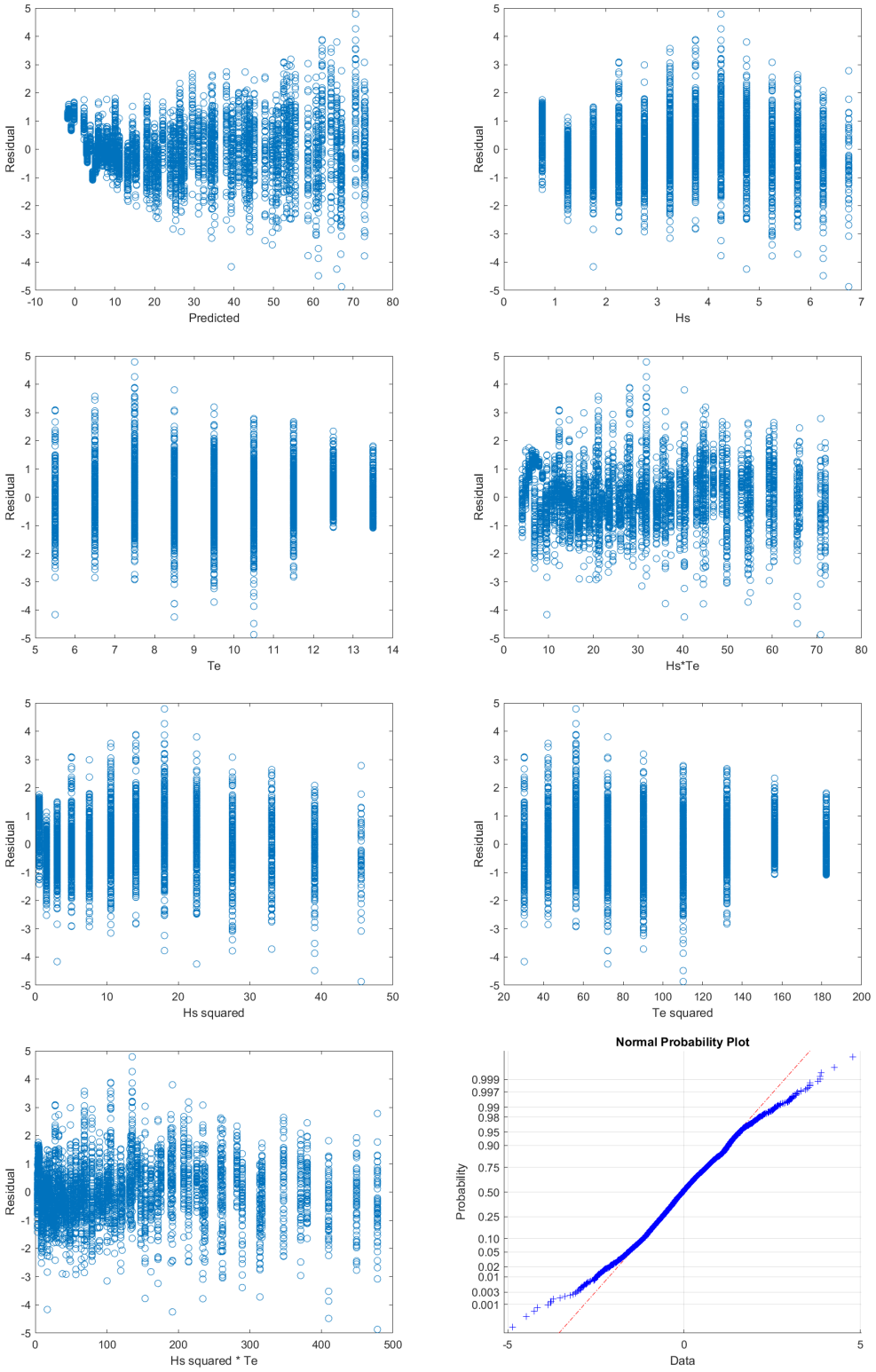


Figure A.3: Residual diagnostics for the model with main effects, interaction term and squared predictors. Studentized residuals plotted against predicted values (top left), significant wave height $H_s$ (top right), wave energy period $T_e$ (second row left), interaction term $H_s \cdot T_e$ (second row right), $H_s$ squared (third row left), $T_e$ squared (third row right) and $H_s$ squared times $T_e$ (bottom left), as well as a normal Q-Q plot of the residuals (bottom right).

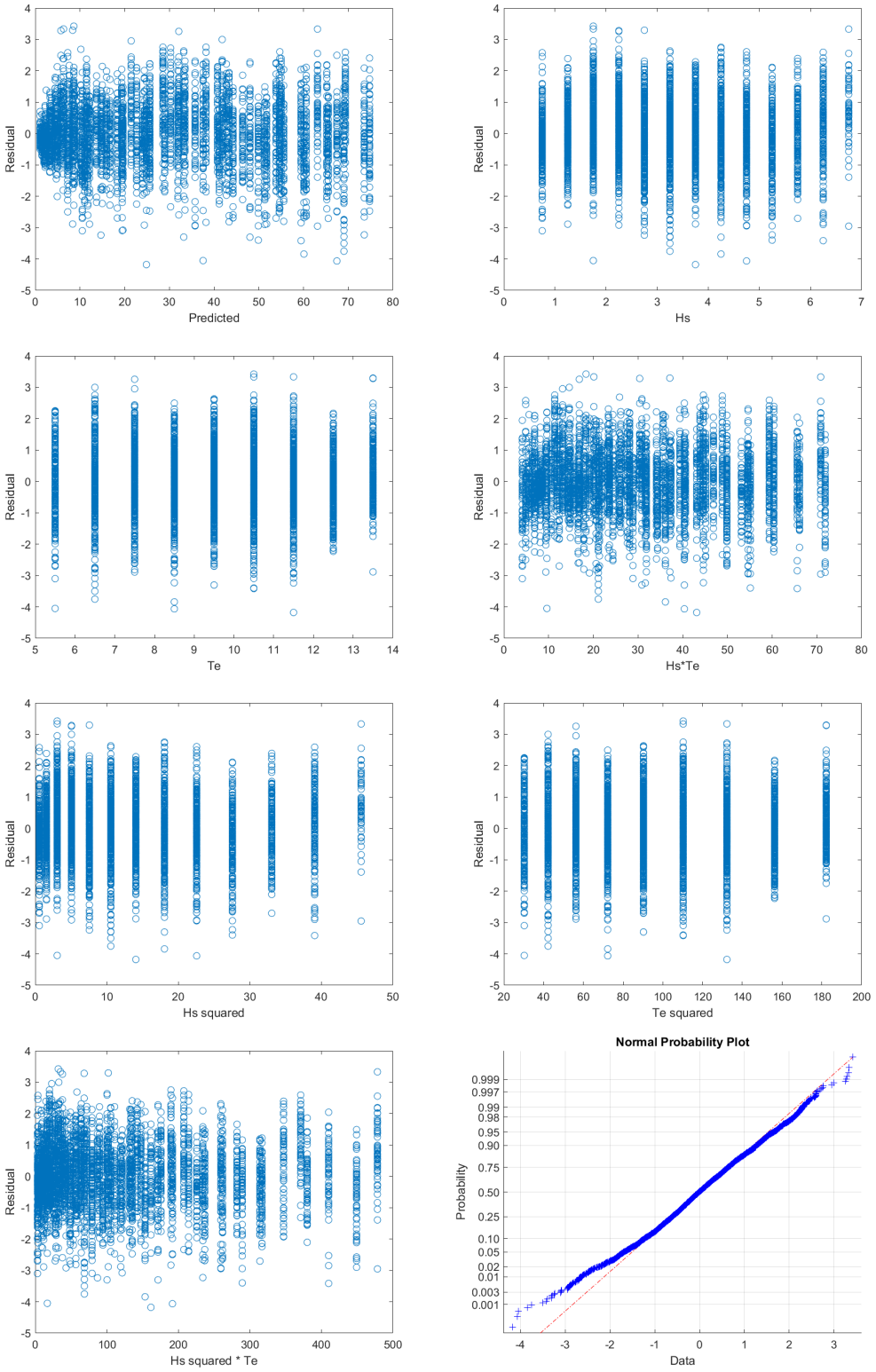


Figure A.4: Residual diagnostics for the model with main effects, interaction term and squared predictors, where the response variable is the square root of the mean power. Studentized residuals plotted against predicted values (top left), significant wave height $H_s$ (top right), wave energy period $T_e$ (second row left), interaction term $H_s \cdot T_e$ (second row right), $H_s$ squared (third row left), $T_e$ squared (third row right) and $H_s$ squared times $T_e$ (bottom left), as well as a normal Q-Q plot of the residuals (bottom right).

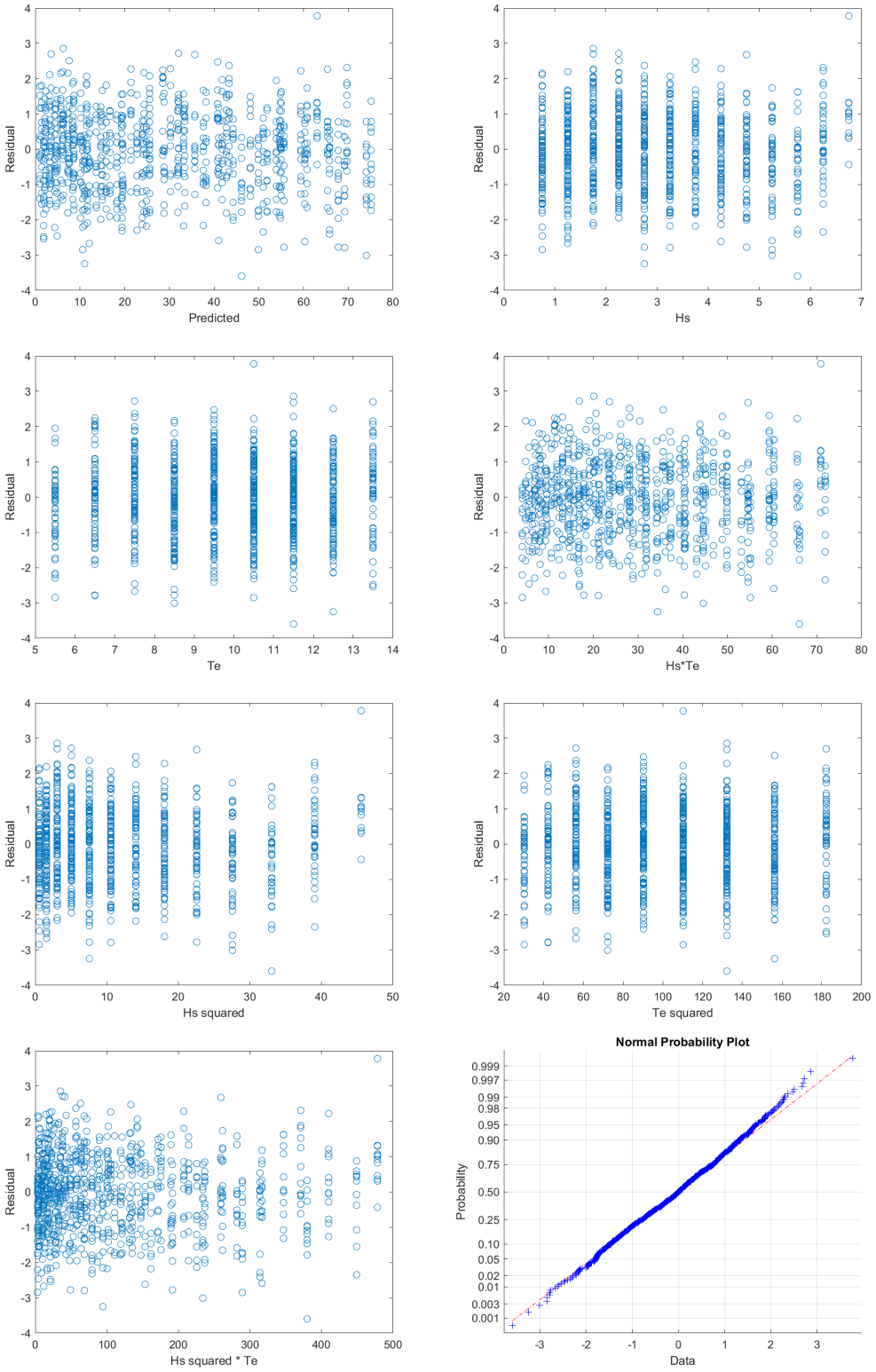


Figure A.5: Residual diagnostics for the weighted least squares model. Studentized residuals plotted against predicted values (top left), significant wave height $H_s$ (top right), wave energy period $T_e$ (second row left), interaction term $H_s \cdot T_e$ (second row right), $H_s$ squared (third row left), $T_e$ squared (third row right) and $H_s$ squared times $T_e$ (bottom left), as well as a normal Q-Q plot of the residuals (bottom right).

Master's Theses in Mathematical Sciences 2024:E36
ISSN 1404-6342

LUNFMS-3128-2024

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/