

Spectral Time Series Analysis of Ocean Wave Buoy Measurements

Jake P. Grainger, B.Sc.(Hons.), M.Res



Submitted for the degree of Doctor of
Philosophy at Lancaster University.

September 2022

Abstract

Waves in the ocean can be as dangerous as they are impressive. In order to study the behaviour of such waves, buoys are commonly deployed to collect recordings of the ocean surface over time. This results in large quantities of high-frequency multivariate time series data. The statistical analysis of such data is of great importance in a variety of engineering and scientific contexts, from the design of coastal flood defences to offshore structures.

We develop methodology for analysing such buoy data, investigating two key questions. Firstly, how should we perform parameter inference for models of the frequency domain behaviour of the surface, given recorded buoy data? Secondly, how can we detect statistically significant non-linearities present in these time series?

For parameter inference, we find that pseudo-likelihood approaches greatly outperform state-of-the-art methodologies. As a result, not only can we obtain more reliable parameter estimates, but we can also perform inference for more complicated models, allowing for a more intricate description of the waves. Due to the improved performance of such estimates, we are able to see the evolution of these parameters throughout storm events, using recorded buoy data from both California and the North Sea.

For detecting non-linearities, we develop a robust testing procedure by evaluating

the bispectrum of the observed time series against the bispectrum of bootstrap simulated Gaussian processes with similar characteristics. We explore the performance of this technique in simulation studies, and apply the approach to buoy data from California.

Acknowledgements

I want to begin by thanking Adam and Phil. I could not have asked for better supervisors. You have always been there, whenever I needed anything, and for that I am truly grateful. Your constant help, support and advice has made my PhD a very enjoyable experience. It would not have been possible without you. Adam, thank you for your encouragement, positivity and kindness — it makes such a difference. Phil, mae eich brwdfrydedd yn heintus iawn, byddai cyfarfod am hanner awr wedi wyth o’r gloch yn y bore wedi bod yn llawer anoddach hebdo!¹ I don’t think any of us expected to spend a good chunk of the past few years meeting from our bedrooms², but we seem to have managed regardless.

I would also like to thank Kevin, who, despite not being my supervisor, has been there since day one (albeit remotely); and Hans, whose knowledge of the data in particular was incredibly helpful. Without our Thursday morning meetings (or evening, depending on your perspective), my thesis would simply not be the same. The explanations of both oceanography and data proved invaluable to me throughout my time working on this thesis, something for which I am extremely grateful. I would also like to thank Rob, Fay and all the people at JBA for your support and the interesting problems we have been able to work on. I am also grateful to Allan

¹Thank you Robbie for the translation, and the loan of multiple maths books I have yet to return!

²or in Phil’s case, the attic

and TotalEnergies for providing access to the North Sea dataset and for allowing an anonymised portion of the data to be made available.

To my Mum and Dad, my sisters Beth and Millie, my Grandparents, and Jonathan, Laura, and all of my family. Without your encouragement, help and advice I do not think I would have pursued a PhD, much less managed to finish writing a thesis. Growing up, you created an environment in which I was always encouraged to be curious, something which has been vital in my time as a PhD student. Thank you.

Finally I would like to thank my friends. I have always been lucky to have been surrounded by kind, funny and interesting people. To everyone in STOR-i, particularly my year group, thank you for making the experience so memorable. In particular, Drupad, George, Graham and Luke, from overly competitive table football to throwing burritos at each other in the kitchen, I have loved every minute of our time at STOR-i. It seems to have gone so fast!

This thesis is dedicated to my late Grandma, Betty Smith

Love you lots

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

A version of Chapter 2 has been published as Grainger, J. P., Sykulski, A. M., Jonathan, P., and Ewans, K. (2021). Estimating the parameters of ocean wave spectra. *Ocean Engineering*, 229:108934.

A version of Chapter 3 has been accepted for publication pending minor revisions as Grainger, J. P., Sykulski, A. M., Ewans, K., Hansen, H. F., and Jonathan, P. (2022). A multivariate pseudo-likelihood approach to estimating directional ocean wave models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.


This thesis is approximately 67,000 words.

Jake P. Grainger

Contents

Abstract	I
Acknowledgements	III
Declaration	VI
Contents	XII
List of figures	XVIII
List of tables	XIX
List of abbreviations	XX
List of symbols	XXI
Introduction	XXIV
Thesis outline	XXIX
I Parametric modelling of ocean wave spectra	1
1 Spectral time series analysis and ocean waves	2
1.1 The spectral density function	3

1.1.1	The univariate case	3
1.1.2	The multivariate case	5
1.1.3	Non-parametric estimation	8
1.2	The frequency-direction spectrum	11
1.2.1	Definition	11
1.2.2	Relation to other processes of interest	14
1.2.3	Non-parametric estimation	15
1.3	Parametric models for ocean wave spectra	18
1.3.1	Models for the marginal spectral density function	19
1.3.2	Models for the spreading function	21
1.3.3	Parametric estimation of ocean wave spectra	23
1.4	Simulating Gaussian processes	25
1.4.1	Random phase methods	26
1.4.2	Exact method	28
1.5	Conclusion	29
2	Estimating the parameters of ocean wave spectra	30
2.1	Introduction	31
2.2	Background	35
2.2.1	Non-parametric spectral density estimators	37
2.2.2	Models for the spectral density function	39
2.3	Fitting parametric spectral density functions	41
2.3.1	Least squares	41
2.3.2	Maximum likelihood	43
2.3.3	Spectral likelihood	44
2.3.4	Comparison	45
2.3.5	Differencing	49
2.4	Practical concerns for implementation with the generalised JONSWAP	50



2.4.1	Frequency selection	52
2.4.2	Numerical estimation of the autocovariance	53
2.5	Simulation study	56
2.5.1	Method	56
2.5.2	A canonical sea state	58
2.5.3	Robustness of results	63
2.5.4	Differencing for high sampling frequencies	65
2.6	Quantifying estimation uncertainty	66
2.7	Humboldt Bay data	70
2.8	Discussion and conclusion	74
3	A multivariate pseudo-likelihood approach to estimating directional ocean wave models	78
3.1	Introduction	78
3.2	Background	82
3.2.1	Ocean waves and frequency-direction spectra	82
3.2.2	Example data	85
3.2.3	Models for wind-sea	88
3.3	Modelling process	91
3.3.1	Model fitting	91
3.3.2	Simulation study	94
3.4	Modelling the example data set	99
3.4.1	Specification of low- and high-frequency thresholds for inference	99
3.4.2	Parameter estimates	102
3.5	Discussion and conclusions	105
4	Conclusions of Part I	108
4.1	Further work	109

4.1.1	Modelling the behaviour of spectral parameters over time . . .	110
4.1.2	Modelling an array of wave staffs	114

II Detecting departures from Gaussianity in records of ocean waves 117


5 Higher-order spectra 118

5.1	Background	119
5.1.1	Stationarity, Gaussianity and linearity	119
5.1.2	Cumulants	120
5.2	Higher-order spectra	122
5.2.1	The third-order case	122
5.2.2	The general case	126
5.3	Estimating higher-order spectra	132
5.3.1	The third-order case	132
5.3.2	The general case	135
5.3.3	Practical considerations	140
5.4	Normalisation	141
5.4.1	The third-order case	142
5.4.2	The general case	145
5.4.3	Multivariate squared skewness	147
5.5	Global tests for linearity and Gaussianity	150
5.5.1	Univariate bootstrap tests	153
5.6	Conclusion	157

6 Testing for non-Gaussianity in ocean wave records 158

6.1	Background	159
6.1.1	The bispectrum	160

6.1.2	Normalisation	161
6.1.3	Estimation	162
6.1.4	Local testing for non-Gaussianity	163
6.1.5	Choosing the number of surrogate realisations	167
6.2	Toy example	168
6.2.1	Univariate	169
6.2.2	Multivariate	170
6.3	Monte-Carlo simulation studies	171
6.3.1	Univariate simulation study	172
6.3.2	Multivariate	175
6.3.3	Bias correction	178
6.4	Application to ocean wave data	179
6.5	Conclusion	182
7	Conclusions of Part II	185
7.1	Further work	186
	Epilogue	188
	Appendices to Parts I and II	189
A	Appendix to Chapter 2	190
A.1	Derivatives of ocean wave spectra	190
A.2	Computing the variance of the first derivative	194
A.2.1	Approximating $Q(t)$	197
A.3	Assumptions for debiased Whittle	197
B	Appendix to Chapter 3	201



B.1	Deriving the spectral density matrix function from the frequency-direction spectrum	201
B.2	Finite water depth correction	205
B.3	Optimisation and gradient calculation	206
B.4	Confidence intervals of the multivariate debiased Whittle likelihood .	206
B.5	Proof of the approximation of cross-spectral density function for two wave staffs	211
C	Appendix to Chapter 6	215
C.1	Circulant embedding	215
C.1.1	Univariate circulant embedding	215
C.1.2	Multivariate circulant embedding	218
C.1.3	Circulant embedding from non-parametric estimates	219
C.2	Derivation of univariate model	221
C.3	Additional simulation studies	222
	Bibliography	225

List of Figures

I	Illustration of the spectral density function of a process with both swell and wind-sea.	XXVI
2.3.1	Image plots of the correlation matrix of the periodogram of a wind-generated wave process for different values of Δ and the corresponding images for the differenced process, for a generalised JONSWAP with parameters $\alpha = 0.7$, $\omega_p = 0.7$, $\gamma = 3.3$ and $r = 5$	51
2.5.1	Boxplots of parameter estimates and time taken for six different fitting routines with true parameters denoted by the dashed red lines. Each row displays the results for a given fitting technique, as well as the log of the time taken to perform the optimisation, recorded in seconds. The fitting was performed on simulations of 1.28Hz observations recorded for half an hour (2304 observations). The process was repeated 1000 times.	61
2.5.2	Scatter plot of the debiased Whittle estimates obtained from half hour records shown in Figure 2.5.1.	62
2.5.3	Boxplots of 1000 parameter estimates using three different fitting routines on the same simulations using the same true parameters as Figure 2.5.1 and same sampling interval but simulating 3 hour records.	63

2.5.4	Boxplots of 1000 parameter estimates using three different fitting routines with true parameters $\alpha = 0.7$, $\omega_p = 0.7$, $\gamma = 1$ and $r = 5$, estimated from simulated half hour records sampled at 1.28Hz.	66
2.5.5	Boxplots of 1000 parameter estimates for 1Hz and 4Hz data. LS, BLS, DW denote least squares, Bartlett least squares and debiased Whittle respectively, and LSd, BLSd and DWd denote the differenced versions.	67
2.7.1	Plots of H_s , the spectrogram on a decibel scale and the estimated parameters over time. Note that some of the plots in (a) have not been truncated, but that upper bounds were set on the optimisation at 10 for γ and 8 for r	72
2.7.2	Plot of periodograms and fitted models on normal and decibel scale, alongside diagnostic Q-Q plots. The times of each record are given at the top of (a). Colours and line styles are used to denote different fitting techniques in the same manner as Figure 2.7.1a. The symbols “o”, “ Δ ” and “ ∇ ” denote quantiles from the debiased Whittle likelihood, least squares estimates, and Bartlett least squares estimates, respectively. Note that the plots in (c) have been truncated in the y axis so that they have a 1:1 aspect ratio, so some quantiles are not shown. . .	73
3.2.1	Example frequency-direction spectra. The left hand panel shows the frequency-direction spectrum corresponding to a single wind-sea, the middle shows a wind-sea and single swell and the right shows a wind-sea and two swells. Direction here is the direction the waves are travelling <i>from</i> . The polar plots have direction from north (rad) on the angular axis and angular frequency (rad Hz) on the radial axis.	82

3.2.2	The effect of varying the mean direction of a wind-sea model on both the frequency-direction spectrum ($S(\omega, \phi)$, left) and the corresponding spectral density matrix function ($f(\omega)$, right). The elements of $f(\omega)$ are shown as a “matrix of functions”, plotting the real part in the lower triangle and imaginary part in the upper triangle. Due to the conjugate symmetry of $f(\omega)$, this representation contains all of the information present.	85
3.2.3	Summary of the storm data, showing significant wave height, wind speed, a spectrogram on the decibel scale, mean wave direction over time and frequency (direction the waves are travelling from, in radians clockwise from North) and the wind direction over time (direction the wind is travelling from, in radians clockwise from North) recorded at a nearby platform. The x -axis labels are at the start of the day, e.g. day 1 denotes the start of day 1.	87
3.2.4	Example of the decomposition of a frequency-direction spectrum, showing the frequency-direction spectrum (left), marginal spectral density function (middle) and spreading function (right). Plots are given using Cartesian coordinates as this makes the arms of the spreading function easier to visualise.	88
3.3.1	Frequency-direction spectra for Scenario 1 (left), Scenario 2 (middle) and Scenario 3 (right), as defined in in Table 3.3.1.	94
3.3.2	Boxplots of the parameter estimates from the simulation study for parameters of the marginal spectral density function, with the true values indicated by red dashed lines. Marginal parameters estimated using least squares (LS), univariate debiased Whittle (DW uni) and full multivariate debiased Whittle (DW) are shown.	97

3.3.3	Boxplots of parameter estimates from the simulation study for parameters of the spreading function, with the true values indicated by red dashed lines. Spreading parameters estimated using least squares with MLM (LS mlm), least squares with MEM (LS mem), the moments-matching approach (moment) and multivariate debiased Whittle (DW) are shown.	98
3.4.1	Heatmap of $\hat{R}(\omega)$ for each half hour period in the example data set, computed using multitapering.	100
3.4.2	Spectrogram of the example data set on the decibel scale, with the period used in the fitting delimited by solid vertical lines, and the choice of frequency range over the period of interest shown by the dotted lines.	102
3.4.3	Parameter estimates using debiased Whittle likelihood inference over the period in question with approximate 95% confidence intervals. The two panels in the second row also include the spectrogram and wind direction for context. In order left to right then down, the panels show: \hat{H}_s ; $\hat{\alpha}$; $\hat{\omega}_p$ over the spectrogram on the decibel scale; $\hat{\phi}_m$ and wind direction; $\hat{\gamma}$; $\hat{\beta}$ and $\hat{\nu}$; \hat{r} ; and $\hat{\sigma}_l$ and $\hat{\sigma}_r$, with approximate 95% confidence intervals.	104
5.2.1	Symmetries of the bispectrum (left) and principle domain (right). The principle domain corresponds to the triangles highlighted in the darker blue and red in the plot of the symmetries (left).	125
6.1.1	Evaluation frequencies for uniform smoothing over rectangles (left) and hexagons (right). Minor ticks are at each of the Fourier frequencies and the filled orange triangle is the principle domain.	163

6.1.2	The spectral density function of the residuals of AR(p) processes fitted to Gaussian processes with a JONSWAP spectral density function ($\alpha = 0.7, \omega_p = 0.8, \gamma = 3.3, r = 5$).	165
6.2.1	Example model spectral density function (left), bispectrum (middle) and squared skewness function (right).	170
6.3.1	The spectral density functions (top), bispectra (middle) and squared skewness functions (bottom) of Cases 1, 2 and 3, the models used in the univariate simulation study. Here a is equal to zero (left), one (middle) and four (right).	172
6.3.2	Proportion of rejections for different models using both circulant embedding and an AR(1) process to generate surrogate data for Cases 1, 2 and 3 (the univariate cases).	173
6.3.3	Proportion of rejections with different correction levels using circulant embedding to generate under the null for Cases 1, 2 and 3.	174
6.3.4	Definition of a true null, top row showing the true bispectra for each process on the estimation grid and bottom row showing which hypothesis is considered true for each frequency pair (black for H_0 and white for H_1). Here we use a definition of $\Gamma(\omega_1, \omega_2) < 0.1$ for the null.	176
6.3.5	Squared skewness function of Cases 4, 5 and 6, the models used in the multivariate simulation study.	177
6.3.6	Proportion of rejections using multivariate circulant embedding and marginal univariate circulant embedding to generate under the null for Cases 4, 5 and 6 (the multivariate cases).	178
6.3.7	Average bias of estimated skewness function for Cases 4, 5 and 6 of the multivariate simulation, from 5000 Monte Carlo replications. Top row shows the bias of the standard estimates, and bottom row shows the bias of the corrected estimates.	180

6.4.1	The squared skewness function of the X , Y and Z processes. If a p -value was classed as significant the border of the hexagon is highlighted with a colour depending on the correction method. In particular, in purple (no correction), red (first FDR method) and orange (second FDR method).	182
6.4.2	The squared skewness function of the X , Y and Z processes after applying various filters. The filters are fitted MA models with $q = 5, 10, 15$ and 20 for rows from top to bottom. Test results are highlighted analogously to Figure 6.4.1.	183
6.4.3	Estimate of the multivariate squared skewness function of the CDIP data, both with no filter and marginal MA models as applied in Figure 6.4.2.	183
C.3.1	Squared skewness function for univariate processes with $a = 1$ but with varying γ parameters.	223
C.3.2	Results of simulation study with the models given in Figure C.3.1. . .	223
C.3.3	Proportion of rejections with different correction levels using circulant embedding to generate under the null for multivariate processes.	224

List of Tables

2.5.1 Average percentage bias, standard deviation and root mean squared error across 24 different parameter choices calculated from 1000 repetitions per parameter choice for least squares (LS), Bartlett least squares (BLS) and debiased Whittle likelihood (DW). Parameter choices were all the combinations of $\alpha = 0.7$; $\omega_p = 0.7, 0.9, 1.2$; $\gamma = 1, 2, 3, 3, 5$ and $r = 4, 5$. Simulated records were half an hour long and sampled at 1.28Hz. The bottom row shows the average of these quantities over all the parameters in the model.	64
3.2.1 Parameter descriptions for the JONSWAP spectrum and bimodal wrapped Gaussian spreading function.	90
3.3.1 Table showing the parameters for each scenario in the simulation study.	94
6.3.1 Parameters used to generate the models for the simulation study. . . .	172
6.3.2 FDR and power for various corrections for multiple testing. These results are for the univariate simulation study using circulant embedding to generate surrogate data. We give 95% confidence intervals in brackets, so that the significance of any differences can be ascertained.	177
6.3.3 Parameters used to generate the models for the multivariate simulation study (Cases 4, 5 and 6).	177

List of abbreviations

MLM	Maximum Likelihood Method
MEM	Maximum Entropy Method
JONSWAP	Joint North Sea Wave Project
FWER	Family-wise Error Rate
FDW	False Discovery Rate
i.i.d	Independent and Identically Distributed
MCMC	Markov Chain Monte Carlo
RAS	Random Amplitude Scheme
DAS	Deterministic Amplitude Scheme
CDF	Cumulative Distribution Function

List of symbols

a, A	Scalars.
$\underline{a}, \underline{A}$	Vectors.
\mathbf{a}, \mathbf{A}	Matrices.
i	The imaginary unit, $\sqrt{-1}$.
\mathcal{R}	The real part of a complex number.
\mathcal{I}	The imaginary part of a complex number.
$\bar{a}, \bar{\underline{a}}, \bar{\mathbf{a}}$	The conjugate of a scalar, vector or matrix.
$\underline{\mathbf{A}}^T, \mathbf{A}^T$	Transpose of a vector or matrix.
$\underline{\mathbf{A}}^H, \mathbf{A}^H$	Conjugate transpose of a vector or matrix.
$f * g$	Convolution of functions f and g .
$\mathbb{E}[X]$	Expectation of a random variable X .
$\text{var}(X)$	Variance of a random variable X .
$\text{cov}(X, Y)$	Covariance between random variables X and Y .
$\mathcal{C}[X_1, \dots, X_k]$	Cumulant of random variables X_1, \dots, X_k .
Δ	Sampling interval.
X	Continuous-time univariate stochastic process.
X_Δ	Discrete-time univariate stochastic process.
$\underline{\mathbf{X}}$	Continuous-time multivariate stochastic process.
$\underline{\mathbf{X}}_\Delta$	Discrete-time multivariate stochastic process.

d	The dimension of a multivariate stochastic process.
$X(t), \underline{\mathbf{X}}(t)$	Stochastic process at time t .
$\zeta_X(\omega)$	Associated spectral process.
τ	Lag in time.
$c_X(\tau)$	Autocovariance of a process X .
$c_{X_i X_j}(\tau)$	Cross-covariance between processes X_i and X_j .
$\mathbf{c}_X(\tau)$	Autocovariance of a multivariate process $\underline{\mathbf{X}}$.
ω	Angular frequency.
ϕ	Direction in radians.
$\delta(\cdot)$	Dirac delta function.
$\delta_\Delta(\cdot)$	Aliased Dirac delta function (periodic with period $2\pi/\Delta$).
$f_X(\omega)$	Spectral density function of a process X .
$f_X^{(\Delta)}(\omega)$	Aliased spectral density function of a process X_Δ .
$f_{X_i X_j}(\omega)$	Cross-spectral density function between X_i and X_j .
$\mathbf{f}_X(\omega)$	Spectral density matrix function of a process $\underline{\mathbf{X}}$.
$S(\omega, \phi)$	Frequency-direction spectrum.
$D(\omega, \phi)$	Spreading function.
n	Length of recorded time series.
$\mathcal{F}_n(\omega)$	Fejér kernel.
$v(t)$	A taper.
$J_v^{(n)}(\omega)$	Tapered discrete Fourier transform.
$I_v^{(n)}(\omega)$	Tapered periodogram.
$\underline{\theta}$	A vector of parameters.
α	Scaling parameter.
ω_p	Peak frequency.
γ	Peak enhancement factor.
r	Spectral tail decay index.

ϕ_m	Mean direction.
β	Limiting peak separation.
ν	Peak separation shape.
σ_l	Limiting angular width.
σ_r	Angular width shape.
$f_{a_1 a_2 a_3}(\omega_1, \omega_2)$	Bispectrum.
$\Gamma_X(\omega_1, \omega_2)$	Squared skewness function.
$\tilde{c}_X(t_1, \dots, t_k)$	k^{th} -order covariance function (k lag).
$c_X(\tau_1, \dots, \tau_{k-1})$	k^{th} -order covariance function ($k - 1$ lag).
$f_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1})$	k^{th} -order cumulant spectral density function.
$g_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1})$	k^{th} -order moment spectral density function.
$\underline{f}_X(\omega_1, \dots, \omega_{k-1})$	k^{th} -order spectral density vector function
$\underline{I}_{v, a}(\omega_1, \dots, \omega_{k-1})$	k^{th} -order tapered polyperiodogram.
$\hat{f}(\omega_1, \dots, \omega_{k-1})$	General estimator of the spectral density function.
$A \otimes B$	Kronecker product of A and B .
$\prod_{j=1}^{\otimes k} A_j$	Repeated Kronecker product of A_1, \dots, A_k .
$A^{\otimes k}$	Kronecker power of A .
\mathcal{P}_k	Set of all partitions of $\{1, \dots, k\}$.
q	Significance level of a hypothesis test.
$p_{(i)}$	i^{th} order statistic.

Introduction

Waves in the ocean rank amongst the most impressive of natural phenomena. For all of recorded history, the ocean (and its waves) have played an important role in many aspects of human activity from coastal fishing and trade to exploration and discovery. Risk to human life and infrastructure has always been a central concern of man's interaction with the ocean's surface. From the design of coastal flood defences (Committee on Climate Change, 2018) to the design of ships and electricity generators from offshore wind and wave installations (Tucker and Pitt, 2001), ocean waves have a major impact on many design and safety assessments.

Ocean waves come in a variety of forms, from short capillary waves generated by pressure differences over the ocean's surface, to tides generated by the gravity of celestial bodies (Holthuijsen, 2007). In this thesis, we will focus on wind-generated surface-gravity waves, which are waves generated by the wind, whose restoring force is gravity. Wind-generated surface-gravity waves in turn can be separated into two types: wind-sea waves and swell waves. Wind-sea waves are generated locally, by the wind blowing over the surface of the ocean. In contrast, swell waves were once wind-sea waves, but have propagated away from the site of their generation, and are now observed elsewhere. Typically, swell waves are longer (lower frequency) than wind-sea waves, and tend to be more regular (with energy that is more focused in both frequency and direction).

Such waves can be recorded in a number of different ways, from in-situ devices such as buoys to remote sensing such as plane mounted radar (Hwang et al., 1998), or even satellites (Douglas and Cheney, 1990). The two most common in-situ measurement techniques are wave staffs and buoys (see Holthuijsen, 2007, for an overview of other kinds of recording devices). In some circumstances, wave staffs can be used to obtain Eulerian measurements of the surface of the water over both space and time, e.g. in small lakes (Donelan et al., 1985; Young et al., 1996) or wave tanks (Forristall, 2015; Schubert et al., 2020). Such setups would be ideal in theory, but, in deeper water, are not possible in practice.³ In contrast, buoys in the water provide recordings of their Lagrangian motion. Whilst some buoys only record their vertical displacement, many record other aspects of their motion. Such measurements can provide information about the directional characteristics of the waves. For example, a heave-pitch-roll buoy records its vertical displacement (heave) and its rotation about its horizontal axes (pitch and roll), see e.g. Longuet-Higgins et al. (1963). More modern buoys, such as a Datawell Waverider MkIII buoy (Datawell, 2006), record their displacement in three orthogonal directions (vertical, Northwards and Eastwards). Summary statistics of these time series, such as significant wave height and peak period, are of great importance, especially their extremes (e.g. Forristall, 2004; Wadsworth and Tawn, 2012; Northrop et al., 2017; Tawn et al., 2018; Shooter et al., 2022).⁴ However, the focus of this thesis is on the buoy displacement time series themselves, a higher resolution phenomenon.

Typically, such time series are analysed in the frequency domain. Heuristically, the recorded time series is treated as a linear combination of regular waves of differing frequencies with random amplitudes and phases (Longuet-Higgins, 1957). Un-

³At least not with existing technology.

⁴Such studies are typically performed on hindcast predictions of significant wave height, generated from physical models, as opposed to from the time series directly. But such hindcasts are inherently predicting a description of the aforementioned time series.

der linear wave theory⁵, these amplitudes and phases are independent of one another, and the process is Gaussian. In some sense, the spectral density function can be thought of as describing the variance of these amplitudes.⁶ Similarly, the ocean surface is viewed as a linear combination of regular waves with differing frequencies and directions, whose amplitudes have a variance specified by the frequency-direction spectrum. The specific shape of these spectra provide an important description of the behaviour of such waves. For example, Figure I shows the spectral density function of a process where both wind-sea and swell waves are present. The presence of two components is clear to see in the frequency domain, but would be much less clear in the time domain (e.g. from looking at autocovariance). In summary, we are interested in the spectral density function, a decomposition of variability in frequency; and the frequency-direction spectrum, a decomposition of variability in *both* frequency and direction. These concepts are defined formally in Sections 1.1 and 1.2 respectively.

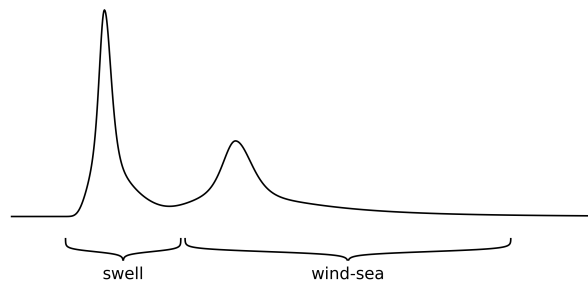


Figure I: Illustration of the spectral density function of a process with both swell and wind-sea.

Often, parametric models are proposed for the shape of the spectral density function and frequency-direction spectrum (one of the most popular models being the JONSWAP spectrum proposed by Hasselmann et al., 1973). The parameters of these

⁵Linear wave theory arises from linearising the equations and boundary conditions that describe the motion of water waves (see Holthuijsen, 2007, for example). The solutions to such equations are often referred to as Airy waves (Airy, 1845).

⁶This is expressed more formally in Chapter 1.

models are then estimated from recorded time series. However, it is known that estimates of these parameters can be unreliable (Ewans and McConochie, 2018). Therefore, in Part I of this thesis, we focus on improving the estimation of such parameters. In particular, we show how, using pseudo-likelihood approaches to directly model the time series, we can get major improvements in both bias and variance of the parameter estimators. Both for parameters of the spectral density function, and of the frequency-direction spectrum.

Whilst Part I focuses on the linear properties of ocean waves, deviations from linearity are also of interest (e.g. Ewans et al., 2021). From the perspective of the time series analysis of ocean waves, this means deviations from Gaussianity. In the frequency domain, this corresponds to the amplitudes (and phases) being uncorrelated⁷ at different frequencies, but not necessarily independent. This dependence can be described by higher-order joint cumulants, determined by the higher-order spectra of the process.

To explore such non-linearities, many authors look for peaks in the higher-order spectra (which would be identically zero in the case of a Gaussian process). Whilst many studies have looked into this problem in the context of ocean waves, both in the third-order case (e.g. Hasselmann, 1962; Elgar and Guza, 1985; Cherneva and Soares, 2007) and fourth-order case (e.g. Chandran et al., 1994; Elgar et al., 1995; Ewans et al., 2021), they only consider univariate processes, with significance levels based on often-unrealisable asymptotics. Therefore, in Part II of this thesis, we focus on higher-order spectra, and their application to ocean wave time series. In particular, we discuss the normalisation problem (see Kim and Powers, 1979; Hinich and Wolinsky, 2005), and local significance testing for bispectra (the third-order spectral density function). The bispectrum (and indeed spectra of higher order) are impor-

⁷Note that this is only true in the stationary case.

tant in a wide range of applications, including plasma physics (Greb and Rusbridge, 1988; Riggs et al., 2021); seismology (Haubrich, 1965); medicine (He and Thomson, 2010) and of course ocean waves. Therefore, we hope that the methodology developed in Chapter 6 will be applicable beyond time series of ocean waves.

Thesis outline

There are two main themes in this thesis, and therefore it is presented in two parts. Part I concerns parametric modelling of ocean wave spectra and is comprised of Chapters 1, 2 and 3, with concluding remarks given in Chapter 4. Part II focuses on higher-order spectra, and is comprised of Chapters 5 and 6, with conclusions given in Chapter 7.

Outline for Part I

In Chapter 1, we review background theory on frequency domain time series analysis and spectral modelling of ocean wind-generated surface-gravity waves (henceforth “ocean waves”). In particular, we give definitions and introduce notation for fundamental quantities including the spectral density function and the frequency-direction spectrum. We introduce a variety of parametric models used to describe the spectral density function and the frequency-direction spectrum for ocean waves. We discuss non-parametric and parametric techniques used in the oceanography literature, and briefly review simulation of (Gaussian) ocean wave time series.

Chapter 2 considers the univariate problem of estimating the parameters of a given model for the spectral density function from ocean wave time series. We discuss techniques used in the oceanographic and statistical literature. We demonstrate, through numerical simulation, that the debiased Whittle likelihood outperforms

competitors, such as least squares fitting, in terms of bias and variance of the recovered parameters. We also discuss uncertainty quantification of parameter estimates, and provide an algorithm for speeding up the approach of Sykulski et al. (2019). We perform an example analysis for data recorded at Humboldt Bay offshore California, to illustrate practical concerns that arise when applying the methodology.

Chapter 3 considers estimating the parameters of models for the frequency-direction spectrum, which decomposes spatiotemporal variability by both frequency and direction. We focus on modelling anchored buoy data, in the form of high frequency multivariate time series. State-of-the-art methods for estimating the parameters of such models do not make use of the full spatiotemporal content of the buoy observations due to various historical unnecessary assumptions and smoothing steps. We explain how the multivariate debiased Whittle likelihood is used to estimate jointly all model parameters directly from the recorded time series. When applied to North Sea buoy data, debiased Whittle likelihood inference reveals a smooth evolution of model parameters over time. Finally, we discuss challenging practical issues including model misspecification, and provide guidelines for future application of the method.

Outline for Part II

Chapter 5 introduces higher-order spectra, providing definitions and basic properties, and discussing a variety of estimation techniques. We also give a detailed discussion of different commonly used normalisations for the bispectrum, reiterating the concerns of Hinich and Wolinsky (2005), and then describe a broad class of normalisations motivated by invariance to certain linear operations. Finally, we discuss global tests for Gaussianity and linearity proposed in the literature.

Chapter 6 considers testing for bispectral non-linearities⁸ locally (at specific frequencies). We discuss the approach of Schulte (2016), propose an improvement based on circulant embedding, and an extension to the multivariate case. We investigate the performance of different approaches in simulation studies, quantifying power and false detection rate. We then apply these methods to buoy data from Humboldt Bay offshore California.

⁸In the oceanographic sense.

Part I

Parametric modelling of ocean wave spectra

“It was like looking at the ocean: some days, you could tell what mood it was in. Most days, though, it was unreadable, mysterious.”

— RICK RIORDAN, *THE LIGHTNING THIEF*

Chapter 1

Spectral time series analysis and ocean waves

The frequency domain analysis of time series revolves around the spectral density function. Heuristically, this describes the contribution to the variance¹ of the process from regular sinusoids of different frequencies. In the theory of ocean waves, this is extended to the idea of a frequency-direction spectrum, which decomposes the variance not only over frequency, but also over direction. In this chapter, we shall formally introduce these quantities, and discuss some aspects of the parametric second-order² modelling of ocean wave time series. Note that Chapters 2 and 3 are standalone, but in this chapter we provide more detailed background.

This chapter is structured as follows. We begin in Section 1.1 with a discussion of the spectral density function and the multivariate generalisation, which will be required for describing multivariate time series. We then give a detailed description of the frequency direction spectra in Section 1.2. Existing parametric techniques are discussed

¹Or, from the physical perspective, energy.

²In the time series sense.

in Section 1.3. Then in Section 1.4 we cover simulation of Gaussian processes, which is required for the simulation studies in Chapters 2 and 3. In this thesis, time is typically measured in seconds, space in metres and angular frequency in rad/s.

1.1 The spectral density function

In this section, we define some of the fundamental quantities of this thesis, including the spectral density function and its multivariate counterpart, the spectral density matrix function. Additionally, we discuss some non-parametric estimation techniques for the spectral density function in Section 1.1.3, though this is less important for understanding Chapters 2 and 3 than the definitions in Sections 1.1.1 and 1.1.2.

1.1.1 The univariate case

Consider a continuous-time stochastic process $X = \{X(t)\}_{t \in \mathbb{R}}$. The process X is said to be second-order stationary if, for all $t, \tau \in \mathbb{R}$,

1. $\mathbb{E}[X(t)] = \mathbb{E}[X(0)]$,
2. $\text{cov}(X(t + \tau), X(t)) = \text{cov}(X(\tau), X(0))$,
3. $\text{var}(X(t)) < \infty$.

In other words, the first and second order cumulants of X exist and are invariant to time shifts. Typically, we will assume without loss of generality that processes we deal with are mean-zero, that is $\mathbb{E}[X(t)] = 0$ for all $t \in \mathbb{R}$.

Let $X_\Delta = \{X(t)\}_{t \in \Delta\mathbb{Z}}$ be the stochastic process which arises from sampling X regularly with a sampling interval of Δ seconds, where $\Delta A = \{\Delta a \mid a \in A\}$ for some

set A . Note that if the process X is second-order stationary, then clearly X_Δ is also second-order stationary.

The autocovariance function of a stationary continuous-time process X is defined as $c_X(\tau) = \text{cov}(X(\tau), X(0))$ for all $\tau \in \mathbb{R}$. The autocovariance sequence³ of the sampled discrete-time process X_Δ is $c_X^{(\Delta)}(\tau) = \text{cov}(X(\tau), X(0))$ and is only defined for $\tau \in \Delta\mathbb{Z}$. By definition, $c_X^{(\Delta)}(\tau) = c_X(\tau)$, for all $\tau \in \Delta\mathbb{Z}$. Provided that the autocovariance of the process is absolutely integrable (absolutely summable in the discrete-time case), the spectral density function (sometimes referred to as the frequency spectrum) of a process is defined as the Fourier transform of its autocovariance.⁴ More specifically, the spectral density function of X is $f_X : \mathbb{R} \rightarrow \mathbb{R}$ and of X_Δ is $f_X^{(\Delta)} : [-\pi/\Delta, \pi/\Delta] \rightarrow \mathbb{R}$ such that

$$f_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_X(\tau) e^{-i\tau\omega} d\tau,$$

$$f_X^{(\Delta)}(\omega) = \frac{\Delta}{2\pi} \sum_{\tau \in \Delta\mathbb{Z}} c_X^{(\Delta)}(\tau) e^{-i\tau\omega},$$

where $i = \sqrt{-1}$. We will often refer to $f_X^{(\Delta)}$ as the aliased spectral density function, because we have the relation

$$f_X^{(\Delta)}(\omega) = \sum_{k \in \mathbb{Z}} f_X(\omega + 2\pi k/\Delta).$$

In other words, $f_X(\omega + 2\pi k/\Delta)$ for all $k \in \mathbb{Z}$ contributes to $f_X^{(\Delta)}$ at frequency ω , so from a discrete-time process, we can only tell that the contribution to variance is from frequencies $\omega \bmod 2\pi/\Delta$. Essentially, the true density at a given frequency in \mathbb{R} has been assigned an ‘‘alias’’ in the interval $[-\pi/\Delta, \pi/\Delta]$. Therefore the phenomena is known as aliasing, and we say the spectral density function of X_Δ has been

³The unusual notation for a sequence is used to avoid overwhelming subscripts later in this thesis.

⁴The spectral density function can be defined in more general settings, but this is sufficient for our purpose. See Cramer and Leadbetter (1967), for example.

aliased.

Cramer and Leadbetter (1967) show that for a mean-zero stationary process X , it is possible to find an associated process ζ_X such that

$$X(t) = \int_{\mathbb{R}} e^{it\omega} d\zeta_X(\omega).$$

The process ζ_X is often referred to as the spectral or amplitude process associated with X . This is related to the spectral density function by

$$\mathbb{E} [d\zeta_X(\omega_1)d\zeta_X(\omega_2)] = \delta(\omega_1 + \omega_2)f_X(\omega_1)d\omega_1d\omega_2 \quad (1.1.1)$$

where δ is the Dirac delta function (Brillinger, 1974).

In other words, the process ζ_X has orthogonal increments and a variance determined by the spectral density function. It is this representation that gives rise to the common interpretation that a stationary stochastic process can be represented as a linear combination of multiple uncorrelated sinusoids, with variance determined by the spectral density function. A spectral representation can also be found for a discrete-time process. In particular, for $t \in \Delta\mathbb{Z}$,

$$X(t) = \int_{-\pi/\Delta}^{\pi/\Delta} e^{it\omega} d\zeta_X^{(\Delta)}(\omega),$$

where the relation to the aliased spectral density function is analogous to (1.1.1).

1.1.2 The multivariate case

Let $\underline{X} = \{\underline{X}(t)\}_{t \in \mathbb{R}}$ be a d -dimensional real-valued stochastic process. Such a process is said to be second-order stationary if, for all $t, \tau \in \mathbb{R}$,

1. $\mathbb{E} [\underline{\mathbf{X}}(t)] = \mathbb{E} [\underline{\mathbf{X}}(0)],$
2. $\text{cov} (\underline{\mathbf{X}}(t + \tau), \underline{\mathbf{X}}(t)) = \text{cov} (\underline{\mathbf{X}}(\tau), \underline{\mathbf{X}}(0)),$
3. $\text{tr}(\text{var} (\underline{\mathbf{X}}(t))) < \infty.$

Again, $\underline{\mathbf{X}}$ will be said to be mean zero if $\mathbb{E} [\underline{\mathbf{X}}(t)] = \mathbf{0}$ for all $t \in \mathbb{R}$. As in the univariate case, let the discretely sampled process be $\underline{\mathbf{X}}_\Delta = \{\underline{\mathbf{X}}(t)\}_{t \in \Delta\mathbb{Z}}$.

The autocovariance function of $\underline{\mathbf{X}}$ is $c_{\underline{\mathbf{X}}} : \mathbb{R} \rightarrow \mathbb{R}^{d \times d}$ and the autocovariance sequence of $\underline{\mathbf{X}}_\Delta$ is $c_{\underline{\mathbf{X}}}^{(\Delta)} : \Delta\mathbb{Z} \rightarrow \mathbb{R}^{d \times d}$ such that

$$\begin{aligned} c_{\underline{\mathbf{X}}}(\tau) &= \text{cov} (\underline{\mathbf{X}}(\tau), \underline{\mathbf{X}}(0)), \\ c_{\underline{\mathbf{X}}}^{(\Delta)}(\tau) &= \text{cov} (\underline{\mathbf{X}}(\tau), \underline{\mathbf{X}}(0)). \end{aligned}$$

Assuming that $c_{\underline{\mathbf{X}}}$ is absolutely integrable and $c_{\underline{\mathbf{X}}}^{(\Delta)}$ is absolutely summable, the spectral density matrix functions of the processes $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}_\Delta$ are $f_{\underline{\mathbf{X}}} : \mathbb{R} \rightarrow \mathbb{C}^{d \times d}$ and $f_{\underline{\mathbf{X}}}^{(\Delta)} : [-\pi/\Delta, \pi/\Delta] \rightarrow \mathbb{C}^{d \times d}$ where

$$\begin{aligned} f_{\underline{\mathbf{X}}}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} c_{\underline{\mathbf{X}}}(\tau) e^{-i\tau\omega} d\tau, \\ f_{\underline{\mathbf{X}}}^{(\Delta)}(\omega) &= \frac{\Delta}{2\pi} \sum_{\tau \in \Delta\mathbb{Z}} c_{\underline{\mathbf{X}}}^{(\Delta)}(\tau) e^{-i\tau\omega}. \end{aligned}$$

Unlike in the univariate case, the spectral density matrix function can be complex valued (though the diagonals are real valued). In the univariate case, the spectral density function is real valued because the autocovariance function is symmetric about zero. This is not true in the multivariate case, where autocovariance is Hermitian symmetric, i.e. $c_{\underline{\mathbf{X}}}(\tau) = c_{\underline{\mathbf{X}}}(-\tau)^H$, where z^H denotes the conjugate transpose of a matrix z .

The autocovariance of a multivariate process is a matrix valued function (or se-

quence). The functions formed from the elements of these matrices are referred to as autocovariance when they are on the diagonal, and cross-covariance when they are off the diagonal. In other words, we often write

$$c_{\underline{X}}(\tau) = \left[c_{X_i X_j}(\tau) \right]_{1 \leq i, j \leq d}$$

where $c_{X_i X_i}(\tau) = c_{X_i}(\tau)$ is the autocovariance of the univariate process X_i and $c_{X_i X_j}(\tau) = \text{cov}(X_i(\tau), X_j(0))$ is the cross-covariance between processes X_i and X_j . Similarly, the spectral density matrix function can be regarded as a matrix of functions

$$f_{\underline{X}}(\omega) = \left[f_{X_i X_j}(\omega) \right]_{1 \leq i, j \leq d}$$

where $f_{X_i X_i}(\omega) = f_{X_i}(\omega)$ is the spectral density function of X_i and $f_{X_i X_j}(\omega)$ is the Fourier transform of the cross-covariance between X_i and X_j , referred to as the cross-spectral density function.

The coherency between X_i and X_j is defined analogously to correlation:

$$r_{X_i X_j}(\omega) = \frac{f_{X_i X_j}(\omega)}{\sqrt{f_{X_i}(\omega) f_{X_j}(\omega)'}}$$

defined to be zero when the denominator is zero (the numerator is bounded above by the denominator, so is also zero in this case). The coherence and group delay are defined to be the absolute value and argument of the coherency respectively. All of these quantities are defined analogously for discrete-time processes.

1.1.3 Non-parametric estimation

In this section, we introduce estimators for the spectral density matrix function in the multivariate case. Consider a regularly sampled d -dimensional times series, denoted $\underline{\mathbf{X}}(0), \underline{\mathbf{X}}(\Delta), \dots, \underline{\mathbf{X}}([n-1]\Delta)$, of some finite length n . The discrete Fourier transform of this time series is $\underline{\mathbf{J}}^{(n)} : [-\pi/\Delta, \pi/\Delta] \rightarrow \mathbb{C}^d$ such that

$$\underline{\mathbf{J}}^{(n)}(\omega) = \sum_{t=0}^{n-1} \underline{\mathbf{X}}(t\Delta) e^{-it\Delta\omega}$$

and the periodogram is $\mathbf{I}^{(n)} : [-\pi/\Delta, \pi/\Delta] \rightarrow \mathbb{C}^{d \times d}$ such that

$$\mathbf{I}^{(n)}(\omega) = \frac{\Delta}{2\pi n} \underline{\mathbf{J}}^{(n)}(\omega)^H \underline{\mathbf{J}}^{(n)}(\omega),$$

where $\underline{\mathbf{z}}^H$ denotes the conjugate transpose of a vector (or matrix) $\underline{\mathbf{z}}$. For finite samples, the periodogram is a biased estimator of the spectral density function (Percival and Walden, 1993). In particular, for $\omega \in [-\pi/\Delta, \pi/\Delta]$, the expected value is given by

$$\mathbb{E} \left[\mathbf{I}^{(n)}(\omega) \right] = [\mathcal{F}_n * \mathbf{f}_{\underline{\mathbf{X}}}^{(\Delta)}](\omega)$$

where $*$ denotes the convolution operator, and \mathcal{F}_n is the Fejér kernel

$$\mathcal{F}_n(\omega) = \frac{\Delta}{2\pi n} \left(\frac{\sin(\omega n/2)}{\sin(\omega/2)} \right)^2.$$

This can result in substantial bias in small samples. Therefore, it is common to apply a taper (a concept introduced by Blackman and Tukey, 1958). The tapered discrete

Fourier transform of the series is

$$\underline{J}_v^{(n)}(\omega) = \sum_{t=1}^n v(t) \underline{X}(t\Delta) e^{-it\Delta\omega}$$

where v is the taper. The tapered periodogram is then

$$\mathbf{I}_v^{(n)}(\omega) = \frac{\Delta}{2\pi} \underline{J}_v^{(n)}(\omega)^H \underline{J}_v^{(n)}(\omega).$$

Here it is assumed that the taper is normalised so that $\sum_{t=1}^n v(t)^2 = 1$. For the tapered periodogram the bias is different (Percival and Walden, 1993), with the expected periodogram at $\omega \in [-\pi/\Delta, \pi/\Delta]$ given by

$$\mathbb{E} \left[\mathbf{I}_v^{(n)}(\omega) \right] = [\mathcal{V}_n * \mathbf{f}_{\underline{X}}^{(\Delta)}](\omega)$$

where $\mathcal{V}_n(\omega) = \frac{\Delta}{2\pi} |V_n(\omega)|^2$ and

$$V_n(\omega) = \sum_{t=1}^n v(t) e^{-it\Delta\omega}.$$

Typically tapers reduce the bias of the periodogram at the cost of increasing narrow-band correlation, i.e. correlation between frequencies that are similar (Percival and Walden, 1993). Whilst tapering the periodogram reduces bias, both the regular and tapered periodograms are inconsistent. As such, some form of smoothing is required to achieve consistency.

Welch's overlapped segment averaging method (Welch, 1967) is commonly used in many practical applications, including oceanography (e.g. Ewans and McConochie, 2018). Welch's method splits the data into segments of length n_l with an overlap of $n_l - n_s$ points (shifting by the window by n_s each time), computes tapered periodograms and then averages across all of the segments. A total of $B = \lfloor (n -$

$n_l)/n_s] + 1$ blocks are used. In particular, the estimate of the spectral density function takes the form

$$\hat{f}_v^{(n_l, B, n_s)}(\omega) = \frac{1}{B} \sum_{b=0}^{B-1} \mathbf{I}_v^{(bn_s:bn_s+n_l-1)}(\omega)$$

where

$$\mathbf{I}_v^{(r:s)}(\omega) = \frac{\Delta}{2\pi} \mathbf{J}_v^{(r:s)}(\omega)^H \mathbf{J}_v^{(r:s)}(\omega)$$

is the periodogram of the b^{th} segment and

$$\mathbf{J}_v^{(r:s)}(\omega) = \sum_{t=0}^{s-r} v(t) \mathbf{X}([t+r]\Delta) e^{-it\Delta\omega}$$

is the discrete Fourier transform of the b^{th} segment with taper v .

Multitapering, introduced by Thomson (1982), computes an average of multiple periodograms with different tapers. In particular, the basic version of a multitaper estimate using K tapers is

$$\hat{f}_K^{(n)}(\omega) = \frac{1}{K} \sum_{j=1}^K \mathbf{I}_{v_j}^{(n)}(\omega)$$

where the tapers v_j are orthogonal, i.e. $\sum_{t=1}^n v_j(t)v_k(t) = 0$ when $j \neq k$. A common choice of taper is the discrete prolate spheroidal sequence (dpss) introduced by Slepian (1978). Typically, more sophisticated weightings are used, details of which can be found in Percival and Walden (1993), for example.

Welch's method and multitapering are consistent estimators (under the appropriate conditions), but this is by no means a complete list of available consistent estimation techniques. Many more can be found in Brillinger (1974); Percival (1992); Brockwell

and Davis (2006) and many other references.

1.2 The frequency-direction spectrum

In this section, we provide a formal definition for the frequency-direction spectrum, discuss its relation to other processes of interest, and introduce techniques for non-parametric estimation. The discussion of non-parametric estimation is important for understanding the existing parametric techniques described in Section 1.3.3, but Chapter 3 could be read without this discussion.

1.2.1 Definition

The frequency-direction spectrum gives a description not only of the behaviour of the waves over frequency but also over direction. Fundamentally, it is not a description of a time series as we have been discussing previously, but a description of a spatio-temporal stochastic random field. Denote such a random field by $\eta = \{\eta(t, x, y)\}_{t, x, y \in \mathbb{R}}$. Then assuming the random field is stationary (homogeneous) and mean-zero, we can write the covariance function

$$c_\eta(\tau, l_1, l_2) = \mathbb{E} [\eta(\tau, l_1, l_2)\eta(0, 0, 0)].$$

Assuming that $c(\tau, l_1, l_2)$ is absolutely integrable, can write

$$\begin{aligned} c_\eta(\tau, l_1, l_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_\eta(\omega, k_1, k_2) e^{-i(\omega\tau + k_1 l_1 + k_2 l_2)} d\omega dk_1 dk_2 \\ &= \int_0^{2\pi} \int_0^{\infty} \int_{-\infty}^{\infty} k f_\eta(\omega, k \cos \phi, k \sin \phi) e^{-i(\omega\tau + k \cos \phi l_1 + k \sin \phi l_2)} d\omega dk d\phi \\ &= \int_0^{2\pi} \int_0^{\infty} \int_{-\infty}^{\infty} \tilde{f}_\eta(\omega, k, \phi) e^{-i(\omega\tau + k \cos \phi l_1 + k \sin \phi l_2)} d\omega dk d\phi. \end{aligned}$$

We call $f_\eta(\omega, k_1, k_2)$ the frequency-wavevector spectral density function of η and $\tilde{f}_\eta(\omega, k, \phi) = kf_\eta(\omega, k \cos \phi, k \sin \phi)$ the frequency-wavenumber-direction spectral density function of η , incorporating a change of variable from Cartesian to polar coordinates.

Under linear wave theory (Holthuijsen, 2007), there is a dispersion relation between frequency and wavenumber

$$\omega^2 = kg \tanh(kh)$$

(where h is the water depth) so that the spectral-density is only supported on a subset of the spectral domain (Barstow et al., 2005). Therefore, we have

$$\tilde{f}_\eta(\omega, k, \phi) = S_\eta(\omega, \phi) \delta(\omega^2 - kg \tanh(kh)),$$

where $S_\eta : \mathbb{R} \times [0, 2\pi] \rightarrow \mathbb{R}$ is the frequency-direction spectral density function, or simply the frequency-direction spectrum. In other words,

$$\begin{aligned} c_\eta(\tau, l_1, l_2) &= \int_0^{2\pi} \int_0^\infty \int_{-\infty}^\infty \tilde{f}(\omega, k, \phi) e^{-i(\omega\tau + k \cos(\phi)l_1 + k \sin(\phi)l_2)} d\omega dk d\phi, \\ &= \int_0^{2\pi} \int_{-\infty}^\infty S_\eta(\omega, \phi) e^{-i(\omega\tau + \kappa(\omega) \cos(\phi)l_1 + \kappa(\omega) \sin(\phi)l_2)} d\omega d\phi, \end{aligned}$$

where $\kappa(\omega)$ satisfies $\kappa(\omega) \tanh(\kappa(\omega)h) = \omega^2$. From now on, we will use the simpler notation $S = S_\eta$ for the frequency-direction spectrum, dropping the subscript.

Note that in this definition, ϕ is the direction the wave is travelling *from*. Some authors use the direction the wave is travelling *towards* (e.g. Barstow et al., 2005). “Direction from” is the same convention used for wind, whilst “direction towards” is the same convention as currents. Since we compare with wind data, we use the convention direction *from*. Aside from some implementation details, it does not mat-

ter which definition is used. Heuristically, we can interpret the frequency-direction spectrum as a measure of the contribution to the variance from waves of a given frequency travelling from a given direction.

Let $Z(t) = \eta(t, x, y)$ for all $t \in \mathbb{R}$ at a fixed location⁵ $x, y \in \mathbb{R}$. We have $c_Z(\tau) = c_\eta(\tau, 0, 0)$, and therefore

$$\int_{-\infty}^{\infty} f_Z(\omega) d\omega = c_Z(\tau) = c_\eta(\tau, 0, 0) = \int_{-\infty}^{\infty} \int_0^{2\pi} S(\omega, \phi) e^{-i\omega\tau} d\phi d\omega.$$

Because the f_Z and S are continuous, we have

$$f_Z(\omega) = \int_0^{2\pi} S(\omega, \phi) e^{-i\omega\tau} d\phi, \quad (1.2.1)$$

for all $\omega \in \mathbb{R}$. The relation in (1.2.1) gives rise to the standard decomposition of the frequency-direction spectrum as

$$S(\omega, \phi) = f_Z(\omega) D(\omega, \phi)$$

where $D(\omega, \phi)$ is known as the spreading function, and f_Z is often referred to as the marginal spectral density function. From (1.2.1), we have

$$\int_0^{2\pi} D(\omega, \phi) d\phi = 1.$$

Therefore, for fixed frequency, the function $D(\omega, \cdot) : [0, 2\pi] \rightarrow \mathbb{R}$ at some fixed $\omega \in \mathbb{R}$ is essentially a probability density function with support on $[0, 2\pi]$. From a physical perspective, we expect $D(\omega, \cdot)$ to be 2π -periodic and continuous. Wrapped distributions provide a convenient class of functions which can be used to model $D(\omega, \cdot)$, see Section 1.3.2.

⁵The specific location does not matter as the process is stationary.

In the same way that the frequency-direction spectrum describes the contribution to the variance from waves of a given frequency and direction, the marginal spectral density function describes the total contribution to variance from waves of a given frequency over all directions. The spreading function can be interpreted as describing the contribution to the variance from waves of a given frequency and direction, *relative* to the contribution of all waves with that frequency.

1.2.2 Relation to other processes of interest

Under linear wave theory (see Holthuijsen, 2007, for example), many processes of interest (say \underline{X}) can be related to the frequency-direction spectrum via a transfer function \underline{G} as

$$f_{\underline{X}}(\omega) = \int_0^{2\pi} \underline{G}(\omega, \phi) \underline{G}(\omega, \phi)^H S(\omega, \phi) d\phi. \quad (1.2.2)$$

For the displacement of a particle in deep water $\underline{G}(\omega, \phi) = [1, i \cos \phi, i \sin \phi]^T$ (Isobe et al., 1984). A table of different processes and their transfer functions can be found in Benoit et al. (1997).

As in Chapter 3, we write $Z = \{Z(t)\}_{t \in \Delta Z}$, $X = \{X(t)\}_{t \in \Delta Z}$ and $Y = \{Y(t)\}_{t \in \Delta Z}$, for the vertical, northwards and eastwards displacements of the particle respectively. Additionally, we will write $\underline{P} = \{\underline{P}(t) = [Z(t), X(t), Y(t)]^T\}_{t \in \mathbb{R}}$ to be the multivariate process describing constituted by the full three-dimensional motion of the particle. Of course, a buoy is not a particle in the water, but it is assumed to approximately behave as such (see Section 3.4.1 for an example of when it does not).

1.2.3 Non-parametric estimation

In an ideal world, we would estimate the frequency-direction spectrum from a record of the spatio-temporal process η . However, recording such processes in deep water is not (currently) possible in practice. Therefore, we record other related processes (which are typically multivariate time series, such as the displacement of a buoy) and use those to *attempt* to estimate the frequency-direction spectrum. Before we discuss some non-parametric estimation techniques, it should be noted that this problem is hard. Without the assistance of a model, such estimates can be unreliable.

Estimating the frequency-direction spectrum is equivalent to estimating both the marginal spectral density function and the spreading function. We have already discussed estimating the former from a time series in Section 1.1.3. However, estimating the spreading function is challenging. Assuming that $D(\omega, \cdot)$ is 2π -periodic, we can write it as a Fourier series, i.e.

$$D(\omega, \phi) = \frac{1}{\pi} \left(\frac{1}{2} + \sum_{j=1}^{\infty} a_j(\omega) \cos(j\phi) + b_j(\omega) \sin(j\phi) \right), \quad (1.2.3)$$

for $\phi \in [0, 2\pi]$. We can see from (1.2.2) that for displacement buoys

$$\begin{aligned} a_1(\omega) &= \frac{\mathcal{I}(f_{XZ}(\omega))}{f_{ZZ}(\omega)}, & b_1(\omega) &= \frac{\mathcal{I}(f_{YZ}(\omega))}{f_{ZZ}(\omega)}, \\ a_2(\omega) &= \frac{f_{XX}(\omega) - f_{YY}(\omega)}{f_{ZZ}(\omega)}, & b_2(\omega) &= \frac{2f_{XY}(\omega)}{f_{ZZ}(\omega)}. \end{aligned}$$

where $\mathcal{I}(z)$ denotes the imaginary part of a complex number z . In the case that $f_{ZZ}(\omega) = 0$, these quantities are defined to be zero. Unfortunately, we cannot directly recover the remaining coefficients. Therefore, to obtain the remaining Fourier coefficients, we must make some assumption about their behaviour. The two most common techniques are the Maximum Likelihood method⁶ (Isobe et al., 1984; Capon

⁶Not to be confused with maximum likelihood inference.

et al., 1967) and Maximum Entropy method (Lygre and Krogstad, 1986) (referred to as MLM and MEM respectively). Note that we are not yet estimating the spreading function. Rather, we are attempting to invert the relation given in (1.2.2), which is in general not invertible.

We shall begin by defining the MLM spreading function, which we denote D_{MLM} . The MLM spreading function is constructed to be a linear combination of the spectral density matrix function (Isobe et al., 1984; Benoit et al., 1997), i.e.

$$D_{MLM}(\omega, \phi) = \frac{1}{\mathbf{f}(\omega)} \alpha(\omega, \phi)^H \mathbf{f}_{\underline{\mathbf{X}}}(\omega) \alpha(\omega, \phi)$$

where $\alpha(\omega, \phi)$ is a vector of weights. To choose these weights, Isobe et al. (1984) first write

$$D_{MLM}(\omega, \phi) = \int_0^{2\pi} D(\omega, \phi) W(\omega, \psi, \phi) d\psi, \quad (1.2.4)$$

where

$$W(\omega, \psi, \phi) = |\alpha(\omega, \phi)^H G(\omega, \psi)|^2$$

plays the role of a window function. From (1.2.4) it can be seen that the closer $W(\omega, \phi, \psi)$ is to $\delta(\psi - \phi)$, the closer the MLM spreading function, D_{MLM} , will be to the actual spreading function, D . Therefore, the weighting should be chosen such that this is achieved as closely as possible (Benoit et al., 1997). The optimal choice of weighting function then leads to the following estimate of the spreading function (Isobe et al., 1984)

$$D_{MLM}(\omega, \phi) = \frac{\kappa}{\underline{\mathbf{G}}(\omega, \phi)^H \{ \mathbf{f}_{\underline{\mathbf{X}}}(\omega) \}^{-1} \underline{\mathbf{G}}(\omega, \phi) '}$$

where κ is a normalising constant. The MLM spreading function can be estimated from an observed buoy record by first estimating the spectral density matrix function, and then applying the above, replacing the spectral density with an estimate. MLM estimates are widely used in the literature and are relatively easy and quick to implement.

The MEM spreading function is another widely used transformation of the spreading function, which can also be estimated from buoy measurements. The approach attempts to maximise the entropy of the spreading function such that the first K Fourier coefficients of the spreading function match those derived directly from $f_{\underline{x}}$. More formally, $D_{MLM}(\omega, \phi)$ is defined to satisfy

$$\begin{aligned} \max \quad & - \int_0^{2\pi} \log(D_{MEM}(\omega, \phi)) d\phi, \\ \text{s.t.} \quad & \int_0^{2\pi} D_{MEM}(\omega, \phi) e^{-ij\phi} d\phi = c_j(\omega) \quad \text{for } |j| < K, \end{aligned}$$

for $\omega \in \mathbb{R}$ and $\phi \in [0, 2\pi]$. Where $c_j(\omega) = a_j(\omega) + ib_j(\omega)$ for $j > 0$, $c_j(\omega) = a_j(\omega) - ib_j(\omega)$ for $j < 0$ and $c_0(\omega) = 1$. For the case of a displacement buoy we have $K = 2$ and Lygre and Krogstad (1986) show that

$$D_{MEM}(\omega, \phi) = \frac{1}{2\pi} \frac{1 - F_1(\omega)\overline{c_1(\omega)} - F_2(\omega)\overline{c_2(\omega)}}{|1 - F_1(\omega)e^{-i\phi} - F_2(\omega)e^{-i2\phi}|^2},$$

where \bar{z} denotes the conjugate of a complex number z and

$$\begin{aligned} F_1(\omega) &= \frac{c_1(\omega) - c_2(\omega)\overline{c_1(\omega)}}{1 - |c_1(\omega)|^2}, \\ F_2(\omega) &= c_2(\omega) - c_1(\omega)F_1(\omega). \end{aligned}$$

Notice that D_{MEM} is merely a transformation of the spectral density matrix function (which itself is a transformation of the frequency-direction spectrum, under the as-

sumption of linear wave theory). The MEM estimate can be obtained from observed buoy data by estimating the spectral density matrix function.

It is useful at this stage to note that there are actually two different objects referred to as MLM estimates.⁷ The first is the transformation of the spreading function (a property of the stochastic random field), which we refer to as the MLM spreading function.⁸ The second is the estimate of the MLM spreading function (there is possibly a third meaning if you also include the estimator). We make this distinction as it is important to note that applying these methods to observed buoy data will not produce estimates of the spreading function, but rather will produce estimates of the MLM spreading function, essentially resulting in bias. Unlike in the estimation of the marginal spectral density function, this bias is independent of the sample size.

In both the MLM and MEM cases, the corresponding MLM (or MEM) frequency-direction spectrum can be obtained by multiplying by the marginal spectral density function (or by an estimate of it). Again it should be stressed that this is not the same as the actual frequency-direction spectrum of the stochastic field: the MLM and MEM frequency-direction spectrum are other properties of the stochastic process that are related to, but not the same as, the frequency-direction spectrum.

1.3 Parametric models for ocean wave spectra

Chapters 2 and 3 focus on the parametric estimation of ocean wave spectra. To prepare the ground for these chapters, here we introduce some of the models which have been proposed for such spectra. In this section, we begin by reviewing some

⁷This also applies to MEM, but we state it for MLM only for brevity.

⁸Probably the name maximum likelihood spreading function is more natural than maximum likelihood method spreading function, but we use the latter as the acronym MLM is in keeping with the oceanography literature.

of the popular parametric models used to describe the (marginal) spectral density function of ocean wave time series (in Section 1.3.1). We then discuss models for the spreading function in Section 1.3.2. Whilst we discuss more than one model, each of Sections 1.3.1 and 1.3.2 ends with descriptions of the model we use in Chapters 2 and 3. We then describe the existing parametric estimation techniques in Section 1.3.3. In this section, we present models for wind-sea waves. Models for swell waves can be found in Olagnon et al. (2013) for the marginal spectral density function and Ewans (2002) for the spreading function, for example. Though in this thesis we focus on wind-sea waves specifically, the techniques we develop can be applied to models for swell waves, or models for more complex situations where mixtures of wind-sea and swell waves are present, e.g. by using additive models such as those suggested by van Zutphen et al. (2008).

1.3.1 Models for the marginal spectral density function

One of the earliest proposed parametric forms for the spectral density function of wind-sea waves is the Pierson-Moskowitz spectrum (Pierson and Moskowitz, 1964). In what follows, we will describe the spectral density functions for non-negative frequencies, noting that they are all zero at the zero frequency and are symmetric about zero.⁹

The Pierson-Moskowitz spectrum has a single parameter ω_p , and takes the form

$$f_{PM}(\omega; \omega_p) = \frac{1}{2} \alpha \omega^{-5} \exp \left\{ -\frac{5}{4} \left(\frac{\omega_p}{\omega} \right)^4 \right\}.$$

for $\omega > 0$, where $\alpha = \alpha_c g^2$ and $\alpha_c = 8.1 \times 10^{-3}$ is known as the Phillips constant (Phillips, 1958).

⁹Furthermore, we state the models in the two-sided form, hence the factor of one half.

Following a large study of ocean wave spectra recorded in the North Sea, Hasselmann et al. (1973) proposed a modification known as the JONSWAP spectrum (named after an acronym of Joint North Sea Wave Project). They propose modifying the Pierson-Moskowitz spectrum to have an enhanced peak by multiplying by a Gaussian kernel centred on the peak frequency. In particular, they propose a model of the form

$$f_J(\omega; \omega_p) = \frac{1}{2} \alpha \omega^{-5} \exp \left\{ -\frac{5}{4} \left(\frac{\omega_p}{\omega} \right)^4 \right\} \gamma^{\delta(\omega; \omega_p)}.$$

where

$$\delta(\omega; \omega_p) = \exp \left\{ \frac{(\omega - \omega_p)^2}{2\sigma(\omega; \omega_p)^2 \omega_p^2} \right\}$$

and $\sigma(\omega; \omega_p) = 0.07 + 0.021 \mathbb{1}_{\omega > \omega_p}$. (Note that δ here is not the Dirac delta function used elsewhere in this work). Typically α is the same as in the Pierson-Moskowitz spectrum and $\gamma = 3.3$.

Mackay (2016) expresses the JONSWAP function in a more general form as

$$f_G(\omega; \underline{\theta}) = \frac{1}{2} \alpha \omega^{-r} \exp \left\{ -\frac{r}{4} \left(\frac{\omega}{\omega_p} \right)^{-4} \right\} \gamma^{\delta(\omega; \underline{\theta})}, \quad (1.3.1)$$

where

$$\delta(\omega; \underline{\theta}) = \exp \left\{ -\frac{1}{2\sigma(\omega; \underline{\theta})^2} \left(\frac{\omega}{\omega_p} - 1 \right)^2 \right\},$$

and $\sigma(\omega; \underline{\theta}) = 0.07 + 0.021 \mathbb{1}_{\omega > \omega_p}$. Now $\underline{\theta} = [\alpha, \omega_p, \gamma, r]^T$ is a vector of parameters, where α is a scaling parameter, ω_p is the peak frequency, γ is the peak enhancement factor and r is the tail decay index.¹⁰ Without the peak enhancement term, (1.3.1) is

¹⁰This expression in (1.3.1) is actually slightly less general than the form in Mackay (2016), but

essentially just a scaled Gamma probability density function.

1.3.2 Models for the spreading function

A wide variety of models for the spreading function have been proposed in the literature. We will briefly discuss some of the more popular models, but when applying our methodology in Chapter 3 we use the bimodal model proposed Ewans (1998) due to its flexibility. More complete summaries of available models can be found in Ewans (1998) and Barstow et al. (2005) for example.

One of the oldest models, proposed by Longuet-Higgins et al. (1963), is the so called cos-2s model

$$D(\omega, \phi) = \frac{\Gamma(s(\omega) + 1)}{2\sqrt{\pi}\Gamma(s(\omega) + 1/2)} \cos^{2s(\omega)} \left(\frac{\phi - \phi_{m1}(\omega)}{2} \right),$$

where the functions $\phi_{m1}(\omega)$ and $s(\omega)$ are usually written as functions of summary statistics of the process (e.g. mean direction, peak frequency, spreading etc). See Mitsuyasu et al. (1975); Hasselmann et al. (1980) for example.

Donelan et al. (1985) present evidence that the cos-2s spreading functions described above tend to overestimate the energy in the peak direction. In response to this they propose using a truncated sech² distribution

$$D(\omega, \phi) = \frac{\beta(\omega)}{2\tanh(\pi\beta(\omega))} \operatorname{sech}^2(\beta(\omega) \cdot \{\phi - \phi_1(\omega)\}), \quad (1.3.2)$$

where again $\beta(\omega)$ is some fairly complicated function of peak frequency (with a slightly modified form proposed by Banner, 1990). It should be noted that the spreading function given by (1.3.2) is not 2π -periodic, though this is easily fixed by using a

we use this 4 parameter form as the other parameters are of less interest and make estimation more difficult.

wrapped sech^2 instead (Barstow et al., 2005).

Young et al. (1995) demonstrate the spreading functions of ocean wave records are likely to be bimodal at high frequencies, something that the parametric models above do not describe. To handle this, Ewans (1998) suggests a bimodal spreading function based on a wrapped bimodal Gaussian:

$$D(\omega, \phi) = \frac{1}{\sqrt{8\pi}\sigma(\omega)} \sum_{k=-\infty}^{\infty} \sum_{j=1}^2 \exp \left\{ -\frac{1}{2} \left(\frac{\phi - \phi_{mj}(\omega) - 2\pi k}{\sigma(\omega)} \right)^2 \right\},$$

where $\sigma(\omega)$ is the standard deviation of the Gaussians (called angular width by Ewans, 1998) and $\phi_{m1}(\omega), \phi_{m2}(\omega)$ are the peak directions.

Typically, the angular width and peak direction functions are themselves parameterised. Ewans (1998) gives a parameterisation with fixed values based on observed buoy data, with a single location parameter to determine the mean direction. We shall use a less restrictive description by adding parameters for the shape of the spreading function. A similar parameterisation was used by van Zutphen et al. (2008), but we use slightly fewer parameters as some of the parameters in van Zutphen et al. (2008) have little effect on the frequency-direction spectrum.¹¹ In particular, we write

$$\begin{aligned} \phi_{m1}(\omega; \underline{\theta}) &= \phi_m + \phi_s(\omega; \underline{\theta})/2, \\ \phi_{m2}(\omega; \underline{\theta}) &= \phi_m - \phi_s(\omega; \underline{\theta})/2, \\ \phi_s(\omega; \underline{\theta}) &= \begin{cases} \beta \exp(-v\omega_p/|\omega|) & \text{for } |\omega| > \omega_p, \\ \beta \exp(-v) & \text{otherwise,} \end{cases} \\ \sigma(\omega; \underline{\theta}) &= \sigma_l - \frac{\sigma_r}{3} \left(4 \left(\frac{\omega_p}{|\omega|} \right)^2 - \left(\frac{\omega_p}{|\omega|} \right)^8 \right). \end{aligned}$$

¹¹van Zutphen et al. (2008) parameterise the exponents in $\sigma(\omega; \underline{\theta})$ in addition to the parameters we use.

where ω_p is the peak frequency, as in the JONSWAP spectrum (1.3.1). This adds an additional 5 parameters, namely $\phi_m, \beta, \nu, \sigma_l, \sigma_r$ to the parameters already required for the marginal spectral density function. We prefer this model to the cos-2s and sech² models as it is able to account for potential bimodal behaviour in the spreading function (though unimodal behaviour can be obtained when $\beta = 0$).

1.3.3 Parametric estimation of ocean wave spectra

When estimating the parameters of frequency-direction spectra, a two stage approach is usually taken. Firstly, the parameters of the spectral density function of the vertical displacement are estimated using least squares and then the parameters of the spreading function are estimated separately. We discuss the least squares estimation in detail in Chapter 2, and so we will focus on estimation of parameters of the spreading function in this section. Such estimation is usually performed in one of two ways: a moments-matching approach (Ewans, 1998, for example); or by producing a non-parametric estimate of the spreading function, then fitting using least squares, as in van Zutphen et al. (2008).

Least squares fitting to estimates of the spreading function

A commonly used technique involves fitting the model spreading function to a non-parametric estimate of the spreading function using least squares. In other words, given $\hat{D}(\omega, \phi)$, an estimate of the spreading function (see Section 1.2.3), the parameters, $\underline{\theta}$, are obtained by solving

$$\operatorname{argmin}_{\underline{\theta}} \sum_{\omega \in \Omega} \sum_{\phi \in \Phi} (D(\omega, \phi; \underline{\theta}) - \hat{D}(\omega, \phi))^2,$$

where $\Omega \subseteq [-\pi/\Delta, \pi/\Delta]$ and $\Phi \subseteq [0, 2\pi]$ are finite sets of frequencies and directions respectively. This approach essentially assumes that the estimator used for

the spreading function is unbiased, normally distributed, homoscedastic and that, at different pairs of frequency and direction estimates are uncorrelated. However, in practice, none of these are satisfied. In particular, correlation across frequency and direction is high for both MLM and MEM estimates, and bias is substantial. As a result, estimation of anything other than location parameters using this technique performs poorly, as we will see in Chapter 3.

Moments-matching approach

Early approaches to fitting parametric spreading functions to data from buoys, such as Mitsuyasu et al. (1975), match the Fourier coefficients estimated from the buoy to the theoretical Fourier coefficients from the model (under the relevant transfer function). Begin by estimating the properties of the spreading function at each frequency, e.g. $\phi_{m1}(\omega)$. Importantly, these are different from the model parameters we eventually wish to estimate. Then use regression to estimate the parameters of the model for the behaviour of the spreading function over frequency. In our case, following Ewans (1998), at each frequency we estimate $\underline{\theta}(\omega) = [\phi_{m1}(\omega), \phi_{m2}(\omega), \sigma(\omega)]^T$ using

$$\hat{\underline{\theta}}(\omega) = \underset{\underline{\theta}(\omega)}{\operatorname{argmin}} |c_1(\omega; \underline{\theta}(\omega)) - \hat{c}_1(\omega)|^2 + |c_2(\omega; \underline{\theta}(\omega)) - \hat{c}_2(\omega)|^2$$

where, $c_j(\omega; \underline{\theta}(\omega)) = a_j(\omega; \underline{\theta}(\omega)) + ib_j(\omega; \underline{\theta}(\omega))$ and $\hat{c}_j(\omega)$ is an estimate for $c_j(\omega)$ obtained by plugging estimates for the relevant cross-spectral density functions into (1.2.3). These are estimated for ω in some finite set $\Omega \subseteq [-\pi/\Delta, \pi/\Delta]$, yielding $\hat{\phi}_{m1}(\omega)$, $\hat{\phi}_{m2}(\omega)$, and $\hat{\sigma}(\omega)$.

The parameters of interest $\underline{\theta} = [\phi_m, \beta, \nu, \sigma_l, \sigma_r]^T$ are then estimated by

$$\hat{\underline{\theta}} = \underset{\underline{\theta}}{\operatorname{argmin}} \sum_{\omega \in \Omega} \left[(\phi_{m1}(\omega; \underline{\theta}) - \hat{\phi}_{m1}(\omega))^2 + (\phi_{m2}(\omega; \underline{\theta}) - \hat{\phi}_{m2}(\omega))^2 + (\sigma(\omega; \underline{\theta}) - \hat{\sigma}(\omega))^2 \right].$$

Recall that the other parameters, $\alpha, \omega_p, \gamma, r$, are estimated marginally prior to this. Such a technique is usually not applied to a single sea state, but instead is applied to multiple sea states with the view to fixing the parameters of the spreading function (except the mean direction). As we show in Chapter 3, this approach performs poorly in application to a single sea state. However, it should be remembered that this technique can still be useful for getting a general idea of the shape different aspects of the spreading function can take, as in [Ewans \(1998\)](#), but it is not useful for estimating the parameters of a single sea state.

1.4 Simulating Gaussian processes

At various points in this thesis, it will be useful to be able to simulate processes with a given spectral density function, in order to test methodology via Monte Carlo experiments. In this section, we review some of the common simulation approaches from the oceanography literature, discuss some of their draw backs, and detail exact methods which can be used instead. In particular, we wish to simulate a Gaussian process with known spectral density function (or equivalently autocovariance) at a regular sampling interval of Δ , for n observations. In other words, we want to generate a realisation of $X(0), X(\Delta), \dots, X([n-1]\Delta)$ where the process X is a Gaussian process.

1.4.1 Random phase methods

The most common approach to simulating Gaussian ocean wave time series is an approximate frequency domain technique, due to Thompson (1973) and Tucker et al. (1984). Choose some $m \geq n$, and for $0 \leq k \leq \lfloor m/2 \rfloor$ draw independent

$$R_k \sim \text{Rayleigh}(f_X^{(\Delta)}(2\pi k/\Delta m)^{1/2}),$$

$$U_k \sim \text{Unif}(0, 2\pi),$$

and form

$$Z_k = \begin{cases} R_k & \text{for } k = 0 \text{ or } k = m/2, \\ R_k e^{iU_k} / \sqrt{2} & \text{for } 1 \leq k \leq \lfloor m/2 \rfloor - 1, \\ \overline{Z_{m-k}} & \text{otherwise.} \end{cases}$$

Note that the $k = m/2$ case only applies if m is even. We then take the inverse discrete Fourier transform of Z_k to form a realisation of the Gaussian process. In other words,

$$\tilde{X}(t) = \sqrt{\frac{2\pi}{m\Delta}} \sum_{k=0}^{m-1} Z_k \exp\{itk/\Delta m\}$$

for $t = \Delta, \dots, n\Delta$. Using an FFT we can quickly obtain a time series of m observations spaced equally with a sampling interval of Δ between them, and simply discard observations beyond the n^{th} . Adopting the terminology of Mérigaud and Ringwood (2018), we shall refer to this approach as the random amplitude scheme (RAS).

Since we require n observations spaced at an interval of Δ , Tucker et al. (1984) set $m = n$. Unfortunately, even asymptotically this does not actually simulate from the process of interest. There are three main issues that arise. Firstly, as pointed out by

Méridaud and Ringwood (2018), the resultant process will be periodic with period $n\Delta$, meaning that $\tilde{X}(0) = \tilde{X}(n\Delta)$ and so the correlation between the process at the first time point and at the second time point, is the same as the correlation between the process at the first time point and at the last time point, i.e.

$$\begin{aligned} \text{cov}(X(0), X([n-1]\Delta)) &= \text{cov}(X(n\Delta), X([n-1]\Delta)) \\ &= \text{cov}(X(0), X(\Delta)). \end{aligned}$$

Clearly this should not be the case. Secondly, the expectation of the periodogram of \tilde{X} is equal to the spectral density function at the Fourier frequencies (Percival, 1992). Furthermore, the periodogram of \tilde{X} is exponentially distributed. In other words, the periodogram of \tilde{X} has a *finite* sample distribution equal to the *asymptotic* distribution of the periodogram of a process with the desired spectral density function. The third issue is a somewhat related one, which is that the periodogram ordinates at the Fourier frequencies are, by construction, uncorrelated. Thus the *finite* joint distribution of the periodogram at all of the Fourier frequencies is exactly the *asymptotic* joint distribution of the periodogram of the process we wanted to simulate from.

As a result, if we are using such a simulation method to analyse the performance of techniques that make approximations based on the asymptotic distribution of the periodogram (which we will be doing), then we will see much better performance on series simulated in this manner than series simulated with the correct finite sample properties. This in turn would result in overconfidence in the performance of our estimates.

Deterministic amplitude scheme

A commonly used modification of RAS is the deterministic amplitude scheme (DAS). As the name suggests, this approach is identical to RAS, except that the ampli-

tudes are fixed to be equal to the spectral density function, i.e. replacing R_k by $\tilde{R}_k = f_X^{(\Delta)}(2\pi k/\Delta m)^{1/2}$. Note that this is not equivalent to setting the amplitudes to their expectation, as $\mathbb{E}[R_k] = \sqrt{\pi/2} f_X^{(\Delta)}(2\pi k/\Delta m)^{1/2}$, but rather it fixes the power (amplitude squared) to its expectation.

Not only do we now have periodograms with completely the wrong distribution (a point mass at the spectral density function), but the resultant process is only asymptotically Gaussian (asymptotically in m). Clearly RAS is preferable over DAS, but neither are especially satisfactory. Of course, simulating stationary Gaussian processes (or any Gaussian process for that matter) can be done exactly¹², as we discuss in the next section. This should be preferred over both RAS and DAS for the aforementioned reasons.

1.4.2 Exact method

We begin by discussing an approach to simulate exactly from the Gaussian time series using standard multivariate Gaussian theory. Say we wish to simulate an n -dimensional Gaussian random variable with mean $\underline{\mu} \in \mathbb{R}^n$ and covariance matrix $\underline{\Sigma} \in \mathbb{R}^{n \times n}$. Note that, if $\underline{L} \in \mathbb{R}^{n \times n}$ is such that $\underline{L}\underline{L}^T = \underline{\Sigma}$ (\underline{L} is a matrix square root of $\underline{\Sigma}$) and $\underline{Z} \sim \mathcal{N}(\underline{\mathbf{0}}, \underline{I}_n)$, then setting $\underline{X}_n = \underline{\mu} + \underline{L}\underline{Z}$ we have $\underline{X}_n \sim \mathcal{N}(\underline{\mu}, \underline{\Sigma})$. Therefore we can simulate from a mean-zero stationary Gaussian process by performing the following steps:

1. Construct the covariance matrix $\underline{\Sigma} = [c_X([r-s]\Delta)]_{1 \leq r, s \leq n}$.
2. Find a matrix square root \underline{L} of $\underline{\Sigma}$, e.g. using a Cholesky decomposition.
3. Draw \underline{W} from a $\mathcal{N}(\underline{\mathbf{0}}, \underline{I}_n)$.

¹²Of course, in practice the simulation is only exact up to the quality of random number generators and floating point operations.

4. Compute $\underline{X}_n = \underline{LW}$.

Here $[a_{rs}]_{1 \leq r, s \leq n}$ denotes the matrix with elements a_{rs} and $c_X(\tau)$ is the autocovariance of the process we wish to simulate from. Thus we can simulate exactly from a Gaussian time series, though it is slow compared to RAS (since finding a Cholesky square root in general requires $O(n^3)$ operations). However, when we have a stationary process this can be improved to $O(n \log(n))$ operations, using circulant embedding (Davies and Harte, 1987; Wood and Chan, 1994). This is presented in more detail in Section C.1. The exact approach is easily extended to the multivariate case, as is circulant embedding (Helgason et al., 2011).

1.5 Conclusion

In this chapter, we introduced many of the key ideas used throughout Part I of this thesis. In particular, Chapter 2 focuses on univariate parameter estimation of the spectral density function, introduced in Section 1.1, for the corresponding parametric models described in Section 1.3.1. The extension to frequency-direction spectrum described in Section 1.2 and the models from Section 1.3.2 are utilised in Chapter 3, where we discuss the estimating the parameters of models for the frequency-direction spectrum from buoy data. Additionally, we summarised the differences between some approximate simulation techniques commonly used in the oceanography literature, giving reasons why exact techniques should be preferred over the techniques currently employed, especially in the context of this thesis.

Chapter 2

Estimating the parameters of ocean wave spectra

The content of this chapter is a published paper (Grainger et al., 2021), with the exception of Section 2.7, where we now analyse data recorded at Humboldt Bay, California. In the original work we considered much older data from New Zealand, recorded for 20 minutes at the start of each 3 hour period. The Humboldt Bay data is recorded continuously (at 1.28Hz), and provides an interesting new analysis. The rest of the content is the same, with small notational changes made for consistency with other thesis chapters. We motivate the chapter in Section 2.1, and describe the structure of the chapter at the end of Section 2.1. For simplicity, we will often drop the some of the subscripts introduced in Chapter 1. For example, we will often write f and f_{Δ} as opposed to f_X and $f_X^{(\Delta)}$ respectively. Note that when we refer to X in this chapter, this is the same process Z in Chapter 3, the vertical displacement (not to be confused with the horizontal displacement).

2.1 Introduction

Due to the random nature of wind-generated surface-gravity waves, it is common to treat them as stochastic processes. There is particular interest in the spectral density function of such wave processes. For this reason, it is important that we are able to construct good spectral density estimators. Using such an estimator, estimates of the spectral density function can be obtained from observed wave records. Broadly speaking, there are two approaches for obtaining such an estimator: non-parametric and parametric. The most basic non-parametric spectral density estimator is the periodogram — the Fourier transform of the sample autocovariance. However, the periodogram is a noisy estimator. Therefore many less noisy estimators have been developed (see Section 1.1.3). The second approach is to use a parametric spectral density estimator. Here we assume that the spectral density function follows a parametric form, meaning that the inference task becomes estimation of the parameters of this form. In general, parametric estimators are often preferable because they result in smoother estimates and more concise representations of the spectral density function—and the parameters themselves provide physical interpretation of the nature of the wave process.

Many such parametric forms have been developed in the oceanography literature. Phillips (1958) gave theoretical arguments for the tail behaviour of the spectral density function for wind-generated wave processes. Based on this, Pierson and Moskowitz (1964) established a parametric form that characterised the spectral density function of a fully developed sea, describing both the spectral tail and peak behaviour. This was later extended by Hasselmann et al. (1973), so that the parametric form could encompass a wider variety of spectral density functions, including those associated with fetch limited wave processes. This widely used parametric form is usually known as the JONSWAP spectral form. It should be noted that we use JON-

SWAP to refer to the original formulation given by Hasselmann et al. (1973), with a tail decay of $O(\omega^{-5})$ as $\omega \rightarrow \infty$ (where ω denotes angular frequency).

Despite general acceptance of the JONSWAP spectral form amongst practitioners, there is debate concerning the values of the tail decay index and peak enhancement factor. Arguments for an ω^{-5} tail decay, made by Phillips (1958), were called into question by Toba (1973) and later by Phillips (1985), who argued that an ω^{-4} tail had a stronger theoretical basis. Experimental work such as Hasselmann et al. (1973) and Battjes et al. (1987) found evidence for both ω^{-4} and ω^{-5} tail decays, while Hwang et al. (2017) could not find evidence for either, further suggesting that the tail decay index should be treated as a free parameter. In addition, there is a large literature speculating on other tail behaviours, such as the occurrence of a transition frequency from ω^{-4} to ω^{-5} (Forristall, 1981; Ewans and Kibblewhite, 1986; Babanin, 2010, for example). It is also common to fix the peak enhancement factor to 3.3; however, there is little evidence for using precisely this value. In this work, we adopt a more general version of the JONSWAP spectral form, which treats both the tail decay index and peak enhancement factor as free parameters (though our methods also apply to the special cases mentioned, in terms of estimating the remaining parameters of interest).¹ Many authors (Rodríguez and Soares, 1999; Ewans and McConochie, 2018, for example) have found that both the tail decay index and peak enhancement factor are hard to estimate accurately, using current techniques. However, both of these parameters are important for determining the properties of a given sea state. Our contention is that current techniques are not sufficiently accurate or precise to allow strong statements to be made concerning the true values of the tail decay index or peak enhancement factor, from typical data sets. Indeed, in Section 2.5 we demonstrate with simulated half hour records that estimates for the tail decay (us-

¹To avoid potential confusion, it should be noted that we are interested in estimating the parameters of assumed parametric forms for the spectral density function (such as the generalised JONSWAP) and not spectral parameters of a sea state such as significant wave height (H_s).

ing current estimation techniques) range from $O(\omega^{-3})$ to $O(\omega^{-6})$, when the true tail decay is known to be $O(\omega^{-4})$ as $\omega \rightarrow \infty$. Because there is too much variability in the estimates, it is impossible to determine from such half hour data if the true tail decay is $O(\omega^{-4})$ or $O(\omega^{-5})$ as $\omega \rightarrow \infty$. In this work we present an alternative technique that is capable of obtaining these parameters more accurately and precisely, and show in simulated data that this technique can distinguish $O(\omega^{-4})$ and $O(\omega^{-5})$ high-frequency tail decays, even from short records.

The standard approach for estimating parameters of a stochastic model from data is by using maximum likelihood inference. When an analytical form for the likelihood function is known, such parameters can be optimally estimated using maximum likelihood (Pawitan, 2001). For finite sample sizes, maximum likelihood inference typically results in estimators with smaller RMSE compared to alternative techniques, such as method of moments or least squares. However, in the case of wind-generated wave processes, the exact probability distribution is unknown. Though it is possible to make the simplifying assumption that the wave process is Gaussian, for many sea states this assumption will not be reasonable. For this reason, it has become common for oceanographers to use a non-parametric estimator of the spectral density function, and obtain parameters by fitting a parametric form in the least squares sense. However, such least squares estimators will in general be sub-optimal when compared to full maximum likelihood (Constable, 1988).

We therefore turn to frequency domain likelihoods, which are widely used in both time series analysis and spatial statistics (Nordman and Lahiri, 2006; Fuentes, 2007, for example). The canonical approach is to use an approximation to maximum likelihood known as the Whittle likelihood (Whittle, 1953b). The Whittle likelihood can be computed quickly using Fast Fourier Transforms and does not require Gaussianity (Dzhaparidze and Yaglom, 1983). However, the Whittle likelihood has been shown

to produce biased estimates for small sample sizes (Dahlhaus, 1988; Velasco and Robinson, 2000). Sykulski et al. (2019) developed a debiased version of the Whittle likelihood that corrects for this bias, without sacrificing the computational speed or making extra distributional assumptions. In Section 2.3.4, we will provide some intuition as to why we would expect the debiased Whittle likelihood to perform better than least squares, both in terms of accuracy (bias) and precision (variance). Then in Section 2.5, we evidence this claim using numerical simulations.

The contributions of this chapter are as follows. Firstly, we introduce the debiased Whittle likelihood estimator for use on wind-generated wave processes (Section 2.3). Secondly, we detail practical concerns regarding the implementation of the estimation procedure for wind-generated ocean wave processes (Section 2.4), with accompanying MATLAB code provided on GitHub (Grainger, 2021).² This includes an important generalisation of the Sykulski et al. (2019) procedure to allow parameters to be fitted directly to the proposed spectral form *without* having to posit an analytical form for the time-domain theoretical autocovariance sequence—as required in Sykulski et al. (2019), but unavailable for ocean wave spectral forms. Thirdly, we present a novel reformulation of the variance of the debiased Whittle likelihood estimator which can be used to quantify the uncertainty of parameter estimates (Section 2.6). Finally, we perform a detailed simulation and field data study comparing the performance of different parametric spectral density estimators for wind-generated wave processes (Section 2.5 and Section 2.7).

²Note that this toolbox has been superseded by the Julia packages `WhittleLikelihoodInference.jl` (Grainger, 2022c) and `OceanWaveSpectralFitting.jl` (Grainger, 2022b).

2.2 Background

So far we have used the word “wave” loosely to describe the shifting nature of the sea surface. In truth, we are actually interested in modelling the displacement of the sea surface from the resting surface. Of course, in reality this is a 3-dimensional phenomena, but in this chapter we shall consider the vertical displacement of the surface over time at a specific location in space. We can think of the displacement at a given time as being a *random variable* with some distribution. Therefore we can describe the displacement over time by a *stochastic process*, an indexed family of random variables, which we shall denote $X = \{X(t)\}_{t \in \mathbb{R}}$. Note that this is a family of random variables indexed over continuous time, as the actual physical process is constantly changing. However, since we cannot actually record data continuously in time, we must instead settle for recording the process at discrete points in time. We assume that the data are being sampled regularly and denote the sampling interval Δ and the process that arises from sampling X every Δ seconds we shall call $X_\Delta = \{X(t)\}_{t \in \Delta\mathbb{Z}}$.

For the duration of a given record, observations of the sea surface are usually assumed to be from an underlying process X that is *second-order stationary*. This means X satisfies all of the following conditions:

1. $\mathbb{E}[X(t)] = \mathbb{E}[X(0)],$
2. $\mathbb{E}[|X(t)|^2] < \infty,$
3. $\mathbb{E}[X(t)X(s)] = \mathbb{E}[X(t-s)X(0)],$

for all $t, s \in \mathbb{R}$. However, the sea surface is not actually stationary: it evolves over time. One way to circumvent this is to notice that whilst the conditions at sea do evolve over time, they do so relatively slowly if we sample frequently. Therefore, we

treat the sea surface as being stationary over short time intervals, sometimes known as sea states. This is essentially the same approach as locally stationary modelling in time series analysis (Dahlhaus, 2012). Ideally we would make this sea state as short as possible. However, we must balance this with another fundamental statement: the more observations we have, the more confident we can be in our inferences. The question of the correct time interval to use will not be covered here; though, it is useful to keep in mind that improving the precision of parameter estimates will mean that we could use shorter sea states in our analysis. This would allow us to track the evolution of certain meteorological processes, such as tropical cyclones, at a higher precision and resolution. To summarise, the underlying wind-generated wave process is not second-order stationary; however, for short enough time windows, this is a widely used working assumption that allows some inference to be made about the process in question.

The analysis of second-order stationary stochastic processes usually involves two important characteristics: the *autocovariance* and the *spectral density function*. The autocovariance of a process at a given lag τ , is just the covariance of a process with itself τ time-steps later. More formally, the autocovariance is $c(\tau) = \text{cov}(X(\tau), X(0))$. For our purposes, we assume that $\mathbb{E}[X(t)] = 0$ for all $t \in \mathbb{R}$; noting that if this is not the case, then by first centring the data by subtracting its sample mean we can obtain a process with the desired property. Therefore, the autocovariance simplifies to $c(\tau) = \mathbb{E}[X(\tau)X(0)]$. The spectral density function is a frequency domain analogue of the autocovariance, which for the stochastic processes X and X_Δ we shall denote $f(\omega)$ and $f_\Delta(\omega)$ respectively. For the discrete time process, assuming that $\sum_{\tau \in \Delta\mathbb{Z}} |c(\tau)| < \infty$,

$$f_\Delta(\omega) = \frac{\Delta}{2\pi} \sum_{\tau \in \Delta\mathbb{Z}} c(\tau) e^{-i\tau\omega},$$

for $\omega \in [-\pi/\Delta, \pi/\Delta]$, where π/Δ is the *Nyquist frequency* and is the highest observable frequency of the sampled process. Similarly for the continuous time process, assuming $\int_{-\infty}^{\infty} |c(\tau)| d\tau < \infty$,

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(\tau) e^{-i\tau\omega} d\tau, \quad (2.2.1)$$

for $\omega \in \mathbb{R}$.³ Similarly, the inverse relations are

$$c(\tau) = \int_{-\pi/\Delta}^{\pi/\Delta} f_{\Delta}(\omega) e^{i\tau\omega} d\omega, \quad (2.2.2)$$

for $\tau \in \Delta\mathbb{Z}$ and

$$c(\tau) = \int_{-\infty}^{\infty} f(\omega) e^{i\tau\omega} d\omega, \quad (2.2.3)$$

for $\tau \in \mathbb{R}$. The spectral density of the discrete time process, $f_{\Delta}(\omega)$, can be thought of as an aliased version of the continuous time spectral density function $f(\omega)$. More formally, we have the following relation:

$$f_{\Delta}(\omega) = \sum_{k \in \mathbb{Z}} f\left(\omega + \frac{2\pi k}{\Delta}\right), \quad (2.2.4)$$

for $\omega \in [-\pi/\Delta, \pi/\Delta]$ (Percival and Walden, 1993, Chapter 4). In Section 2.5, we demonstrate that aliasing can cause bias in parameter estimation, which is why it is important to define both $f(\omega)$ and $f_{\Delta}(\omega)$ and understand their relationship.

2.2.1 Non-parametric spectral density estimators

Though our purpose is the analysis of parametric spectral density estimators, it is also pertinent to define some of the non-parametric spectral density estimators that

³Note that we are working with angular frequency here, and for all examples in this thesis this is measured in units of rad s^{-1} .

are used throughout this chapter. There are two important properties that should be considered when choosing an estimator. The first of these is *bias*, which is the expectation of the estimator minus the true value. Ideally we would want to choose an estimator that is unbiased, i.e. has a bias of zero. This is often not possible, but the weaker condition of asymptotically unbiased is often achievable. An estimator is said to be asymptotically unbiased if, as the number of observations increases, the bias tends to zero. The second important property is *consistency*. For an estimator to be consistent it must converge in probability to the true parameter as the number of observations tends to infinity. More formally, denote the true parameter by θ_0 and an estimator from a series of n observations by $\hat{\theta}_n$. Then $\hat{\theta}_n$ is a consistent estimator if, for all $\epsilon > 0$, $\mathbb{P}(\|\hat{\theta}_n - \theta_0\| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$.

The most basic non-parametric estimator for the spectral density function of a discrete time process is the *periodogram*. Let $X_{\Delta,n} = \{X(0), X(\Delta), \dots, X(\Delta[n-1])\}$ be a series of n consecutive random variables from X_Δ , then the periodogram is defined as

$$I(\omega) = \frac{\Delta}{2\pi n} \left| \sum_{t=0}^{n-1} X(\Delta t) e^{-it\Delta\omega} \right|^2,$$

for $\omega \in \mathbb{R}$. In practice, the periodogram is typically only evaluated at the Fourier frequencies $\omega = 2\pi j/\Delta n$ using the FFT procedure, where $j = -\lceil n/2 \rceil + 1, \dots, \lfloor n/2 \rfloor$. For convenience, we shall write $\Omega_{n,\Delta}$ for the set of these frequencies. It should also be noted that the periodogram is an estimator for the spectral density of the discrete time process $f_\Delta(\omega)$, not the spectral density of the continuous time process $f(\omega)$. The periodogram can be shown to be an asymptotically unbiased estimator for $f_\Delta(\omega)$, but the periodogram is not consistent.

For this reason, modified versions of the periodogram, which are consistent, are usually used as an alternative to the periodogram. One such modified periodogram,

suggested by Welch (1967), splits the series into smaller segments, applies a taper, calculates the periodogram of each segment and then averages these modified periodograms at each frequency. In practice, Welch's method results in an estimate that is less noisy than a standard periodogram, but has lost resolution in frequency and may be more biased. A subset of such methods is known as *Bartlett's method* (Bartlett, 1948). This approach uses non-overlapping segments with no window function. In other words, Bartlett's estimator is

$$\hat{f}_B(\omega) = \frac{\Delta}{2\pi B n_l} \sum_{b=0}^{B-1} \left| \sum_{t=b n_l}^{b n_l + n_l - 1} X(\Delta t) e^{-it\Delta\omega} \right|^2,$$

where B is the number of segments and n_l is the number of observations in each segment (with $B n_l \leq n$).

2.2.2 Models for the spectral density function

When describing the sea surface, models are often expressed in terms of the spectral density function. Many different spectral density functions have been developed for ocean waves, perhaps most notably the JONSWAP spectrum, developed by Hasselmann et al. (1973). We shall consider a more general model, which encompasses many of the other waves models that have been developed. Following Mackay (2016), we use the following parametrisation:

$$S_G(\omega; \underline{\theta}) = \alpha \omega^{-r} \exp \left\{ -\frac{r}{s} \left(\frac{\omega}{\omega_p} \right)^{-s} \right\} \gamma^{\delta(\omega; \underline{\theta})}, \quad (2.2.5)$$

where

$$\delta(\omega; \underline{\theta}) = \exp \left\{ -\frac{1}{2\sigma(\omega; \underline{\theta})^2} \left(\frac{\omega}{\omega_p} - 1 \right)^2 \right\},$$

and

$$\sigma(\omega; \underline{\theta}) = \begin{cases} \sigma_1 & \text{for } \omega \leq \omega_p, \\ \sigma_2 & \text{for } \omega > \omega_p, \end{cases}$$

for $\omega > 0$; where $\alpha, \omega_p, s > 0$, $\gamma \geq 1$, $r > 1$ ⁴ and $\underline{\theta}$ denotes the vector of parameters. Typically, and for the remainder of this chapter, σ_1, σ_2 and s are set to 0.07, 0.09, and 4 respectively (Mackay, 2016). In this case, the parameter vector is $\underline{\theta} = [\alpha, \omega_p, \gamma, r]^T$. Also let Θ denote the *parameter space* — the set of possible values that $\underline{\theta}$ can take. Then for this general model, the parameter space is $\Theta = (0, \infty) \times (0, \infty) \times [1, \infty) \times (1, \infty) \subseteq \mathbb{R}^4$. Note that (2.2.5) is a one sided spectral density, and is not defined at $\omega = 0$. We shall work with the two sided version as this fits in with the way we have defined the spectral density function, the way techniques are described in the statistical literature, and the way Fast Fourier Transforms are implemented on a computer. Therefore, we define $f_G : \mathbb{R} \times \Theta \rightarrow [0, \infty)$,

$$f_G(\omega; \underline{\theta}) = \begin{cases} S_G(\omega; \underline{\theta})/2 & \text{for } \omega > 0, \\ 0 & \text{for } \omega = 0, \\ S_G(-\omega; \underline{\theta})/2 & \text{for } \omega < 0. \end{cases} \quad (2.2.6)$$

We shall refer to the function defined by (2.2.6) as the generalised JONSWAP spectral form. In this formulation, α is measured in units of $\text{m}^2 \text{s}^{1-r} \text{rad}^{r-1}$, ω_p in rad s^{-1} and γ and r are dimensionless. For convenience, we omit the units in future references.

⁴Mackay (2016) gives the condition that $r > 0$. However, for the spectral density to be integrable (such that the stochastic process has finite variance), we require that $r > 1$.

2.3 Fitting parametric spectral density functions

The process of fitting a parametric spectral density function to observations can be thought of as estimating the parameters of a statistical model, which we denote $\underline{\theta}$. The techniques discussed in this section are applicable to a broad class of spectral density functions. As such, we consider the general case and shall write $f(\omega; \underline{\theta})$ for the spectral density function of the continuous time process, given some choice of parameters $\underline{\theta}$. We shall also write $f_{\Delta}(\omega; \underline{\theta})$ and $c(\tau; \underline{\theta})$ for the spectral density function of the discrete time process and the autocovariance function respectively. For convenience, we shall sometimes refer to the spectral density function of the continuous time process as the spectral density function, and the spectral density function of the discrete time process as the aliased spectral density function. We also write $\Sigma_{\underline{\theta}}$ for the covariance matrix of the multivariate random variable corresponding to n consecutive random variables from X_{Δ} . We now describe each of the fitting methods discussed in this chapter.

2.3.1 Least squares

Current approaches to estimating parameters of spectral density functions used in the ocean waves literature, such as the approaches described by [Ewans and McConochie \(2018\)](#), usually involves two key steps. Firstly, a non-parametric estimator of the spectral density function is constructed. Secondly, a curve fitting algorithm is used so that the corresponding parametric form is a good fit for the observed data. Typically this involves minimising the square distance between the parametric form and non-parametric spectral density estimator. Therefore, we shall refer to such approaches as least squares fitting techniques.

For the purpose of this section, we let $\hat{f}(\omega)$ denote a general non-parametric spectral density estimator (this could be the periodogram, $I(\omega)$, Bartlett estimator, $\hat{f}_B(\omega)$,

or some other non-parametric spectral density estimator). The second part of this fitting routine involves fitting the parametric form to the obtained non-parametric estimator. Typically, this is done by minimising the Euclidean distance between the non-parametric estimator and the parametric spectral density function. We therefore must minimise the objective function given by

$$\ell_{LS}(\underline{\theta} \mid X_{\Delta,n}) = \sum_{\omega \in \Omega} \left(f(\omega; \underline{\theta}) - \hat{f}(\omega) \right)^2, \quad (2.3.1)$$

where $\Omega \subseteq \Omega_{n,\Delta}$ (the choice of Ω is discussed in Section 2.4.1). In other words, the least squares estimator for $\underline{\theta}$ is defined as $\hat{\underline{\theta}}_{LS} = \operatorname{argmin}_{\underline{\theta} \in \Theta} \ell_{LS}(\underline{\theta} \mid X_{\Delta,n})$. This approach could be adapted to account for aliasing by replacing $f(\omega; \underline{\theta})$ with $f_{\Delta}(\omega; \underline{\theta})$, the aliased spectral density function; however, such a modified approach is not currently used in the ocean waves literature and therefore we shall use the form given by (2.3.1) in our simulation study.

Part of the reason that least squares performs poorly is that the variance of a spectral estimate will be different at different frequencies. This means that low density areas of the spectral density function (such as the high frequency tail) tend to be underweighted. For this reason, log transforms are often used in least squares objective functions, especially in the statistics literature (Bloomfield, 1973). Because standard least squares is widely used in the ocean waves literature, we present a comparison of standard least square in this chapter. However, in simulations not shown in this chapter, log least squares still does not perform as well as the debiased Whittle likelihood. Whilst log least squares does provide better estimates of the spectral tail decay index than standard least squares, some of the other parameter estimates have increased bias when compared to standard least squares. Plots of these log least squares simulations are available on GitHub (Grainger, 2021).

2.3.2 Maximum likelihood

Maximum likelihood inference treats the sea surface data as observations of a random variable with a given distribution. The parameters for this distribution are chosen by maximising the probability of observing the data given that the underlying distribution has certain parameters. For the moment, let the sea surface observations be multivariate Gaussian with expectation zero and an unknown covariance matrix $\Sigma_{\underline{\theta}}$. The log-likelihood function for observations of such a process is

$$\ell_{ML}(\underline{\theta} \mid X_{\Delta,n}) = \frac{1}{2} \left(-n \log(2\pi) - \log |\Sigma_{\underline{\theta}}| - X_{\Delta,n}^T \Sigma_{\underline{\theta}}^{-1} X_{\Delta,n} \right), \quad (2.3.2)$$

where $X_{\Delta,n}^T$ denotes the transpose of $X_{\Delta,n}$ and $|\Sigma_{\underline{\theta}}|$ denotes the determinant of $\Sigma_{\underline{\theta}}$. The maximum likelihood estimator is obtained by maximising the log-likelihood function. More formally, the maximum likelihood estimator of $\underline{\theta}$ is given by $\hat{\underline{\theta}}_{ML} = \operatorname{argmax}_{\underline{\theta} \in \Theta} \ell_{ML}(\underline{\theta} \mid X_{\Delta,n})$. Provided that the underlying random variable is actually multivariate Gaussian, this technique will provide asymptotically optimal estimates of $\underline{\theta}$, in the sense that they converge at an optimal rate, see Pawitan (2001), Chapter 8.5, for more details.

This approach can be computationally expensive because evaluating the objective function given by (2.3.2) requires the inversion of an $n \times n$ matrix. Also, if we want to model a distribution that is not Gaussian, then a different log-likelihood function must be used. This may take significantly longer to compute, or may not even be tractable. As previously discussed, wave processes will not typically be precisely Gaussian. However, in Section 2.5 we shall compare fitting techniques on simulated Gaussian processes in the first instance. In this case, full maximum likelihood provides a useful benchmark to compare the performance of other estimators to the optimal choice of estimator.

2.3.3 Spectral likelihood

To avoid some of the problems associated with maximum likelihood estimation we can use approximations to the likelihood, known as pseudo- or quasi-likelihoods, to gain some of the accuracy and precision of maximum likelihood, while keeping computational costs (and distributional assumptions) low. One such quasi-likelihood⁵ is known as the Whittle likelihood (Whittle, 1953b). The Whittle likelihood has been used in a wide range of applications due to its computational efficiency and fairly free distributional assumptions (in particular, we no longer need to assume that the underlying process is Gaussian). In its discretised form, the Whittle likelihood is

$$\ell_W(\underline{\theta} \mid X_{\Delta,n}) = - \sum_{\omega \in \Omega} \left\{ \log(f(\omega; \underline{\theta})) + \frac{I(\omega)}{f(\omega; \underline{\theta})} \right\}, \quad (2.3.3)$$

where $I(\omega)$ denotes the *periodogram* ordinate at angular frequency ω . The corresponding estimator is again obtained by maximising this spectral likelihood, which we shall denote by $\hat{\underline{\theta}}_W = \operatorname{argmax}_{\underline{\theta} \in \Theta} \ell_W(\underline{\theta} \mid X_{\Delta,n})$. This estimator also does not account for aliasing. However, by replacing $f(\omega; \underline{\theta})$ with $f_{\Delta}(\omega; \underline{\theta})$ in (2.3.3), we obtain an estimator that does account for aliasing. We shall refer to this as the aliased Whittle likelihood, though it should be noted that some authors refer to this as simply the Whittle likelihood.

Though this aliased approach accounts for some of the bias in the Whittle likelihood, other forms of bias introduced through phenomena such as blurring⁶ are still present (Percival and Walden, 1993, chapter 6). Sykulski et al. (2019) introduced the debiased Whittle likelihood to deal with both aliasing and blurring simultaneously.

⁵These likelihoods are usually referred to as quasi-likelihoods. However, we also use the term spectral likelihood as it integrates nicely with current terminology used in the literature, as well as giving an intuitive sense of what a spectral likelihood does.

⁶Blurring results from sampling for a finite duration in time, which corresponds to convolution with the Fejér kernel in the frequency domain.

The debiased Whittle likelihood is

$$\ell_{DW}(\underline{\theta} \mid X_{\Delta,n}) = - \sum_{\omega \in \Omega} \left\{ \log(\bar{f}_n(\omega; \underline{\theta})) + \frac{I(\omega)}{\bar{f}_n(\omega; \underline{\theta})} \right\},$$

where $\bar{f}_n(\omega; \underline{\theta}) = \mathbb{E}[I(\omega); \underline{\theta}]$ is the *expected periodogram*. As noted by Sykulski et al. (2019), the expected periodogram can be calculated in $O(n \log n)$ time by using the relation:

$$\mathbb{E}[I(\omega); \underline{\theta}] = \frac{\Delta}{2\pi} \sum_{\tau=-n+1}^{n-1} \left(1 - \frac{|\tau|}{n}\right) c(\tau; \underline{\theta}) e^{-i\tau\Delta\omega}. \quad (2.3.4)$$

The resulting estimator can then be expressed as $\hat{\underline{\theta}}_{DW} = \operatorname{argmax}_{\underline{\theta} \in \Theta} \ell_{DW}(\underline{\theta} \mid X_{\Delta,n})$. Despite being constructed from the periodogram, an inconsistent estimator of the spectral density function, the debiased Whittle likelihood is a consistent estimator of the parameters for the parametric model. The debiased Whittle likelihood is able to address the deficiencies in the periodogram without introducing bias, by accounting for the finite sample properties of the periodogram. Sykulski et al. (2019) also show that, under certain conditions, the debiased Whittle estimator converges optimally. These conditions are discussed further in Appendix A.3.

2.3.4 Comparison

In Section 2.5, we perform a simulation study to compare each of the estimators that we have discussed. However, we can also try to build some intuition as to why certain approaches are likely to be more effective than others. To achieve this we shall consider the conditions under which each technique would be equivalent to full maximum likelihood for a finite sample, then evaluate how likely it is that said assumptions are satisfied. Note that this is not (and nor is it intended to be) a formal proof; results related to the convergence of debiased Whittle estimators and their

proofs can be found in Sykulski et al. (2019). Rather, this is a sketch of what is going on under the hood that causes the debiased Whittle likelihood to outperform least squares based techniques.

Maximum likelihood inference works by making probabilistic statements about the distribution of data and then using this to work out the value of the parameters for which the data are most likely. The part of this process of interest to us here is making such distributional statements. For this comparison, we shall think of the non-parametric spectral density estimates as “the data”, and shall consider what their distribution would need to be for least squares or the Whittle likelihood to be the optimum likelihood function for this data.⁷

Then, for the least squares approach to yield the same parameter estimates as the optimum likelihood function, we would need the non-parametric spectral density estimator used in the fitting routine, $\hat{f}(\omega)$, to satisfy the following four assumptions. Firstly, at each frequency, the non-parametric estimator must be Gaussian. In general, this is not true for non-parametric spectral density estimators, though it is true asymptotically for some of them (e.g. Bartlett’s method). Secondly, the expectation of the non-parametric estimator must be equal to the spectral density function at a given frequency. This is not actually true for non-parametric spectral density estimators, as these are constructed to estimate the aliased spectral density function, not the spectral density function of the continuous time process. Though this aliasing could be accounted for by modifying the spectral form used in the fitting routine, such modification is not standard practice and many non-parametric spectral density estimators are still biased. Thirdly, the variance of the non-parametric spectral density estimator must be the same for each frequency. This is not the case for non-parametric spectral density estimators in general, as the variance at a given

⁷This differs from full maximum likelihood on the time series as we have lost the phase information in calculating a spectral density estimate.

frequency depends on the spectral density function at that frequency (Brockwell and Davis, 2006). Though weighted least squares approaches, such as the approach proposed by Chiu (1988), do begin to address the problem of assumption three, they are not widely used and still make the first and second assumptions. Fourthly, the non-parametric estimators at any two different frequencies must be uncorrelated. This assumption is discussed further in Section 2.3.5.

For the Whittle likelihood to yield the same parameter estimates as the optimum likelihood function, we would need the following three assumptions on the periodogram to hold. Firstly, we would require the periodogram to be exponentially distributed at each Fourier frequency.⁸ Secondly, we would require that the expectation of the periodogram is equal to the spectral density function at a given frequency (and consequently that the variance is the square of the spectral density function). Thirdly, the periodogram at any two different frequencies must be uncorrelated. At fixed frequencies, the first assumption is true asymptotically for linear processes (Brockwell and Davis, 2006) and for some classes of non-linear processes (Shao and Wu, 2007). At first glance, this may seem to be similar to the asymptotic normality of Bartlett modified periodograms that are often used in least squares. However, it should be noted that in the case of the periodogram, this asymptotic result is in terms of the number of observations; whereas for Bartlett modified periodograms, this result is in terms of the number of segments that are used, which is much smaller. When it comes to the second assumption, the periodogram is an asymptotically unbiased estimator of the aliased spectral density function. For this reason, the aliased version of the Whittle likelihood should be used over the standard version. Again it may seem that this is also true for Bartlett modified periodograms, as Bartlett's method averages periodograms and each periodogram is an asymptotically unbiased esti-

⁸A slightly different assumption is made about the zero and Nyquist frequency, though in practice they are often omitted.

mator of the aliased spectral density function. Therefore if we were to adjust for this aliasing, least squares would be justified. However, each of these component periodograms are calculated from small segments of the full record, so it is difficult to invoke asymptotic results. Indeed, this creates somewhat of a catch-22 for Bartlett least squares: to get asymptotic normality we must average many periodograms; but this results in using shorter segments for each periodogram, introducing bias (and vice versa). The debiased Whittle likelihood (Sykulski et al., 2019) bypasses the second assumption (made by the Whittle likelihood) altogether, as it uses the theoretical expectation of the periodogram in place of the spectral density function. This means that even for small sample sizes the debiased Whittle likelihood produces estimates with very small to no bias. The final assumption, the assumption of independence between frequencies, is required by both least squares and spectral likelihoods; however, the Whittle likelihood is also in a strong position when it comes to satisfying this assumption. This is because asymptotically the periodogram is uncorrelated at different frequencies, and we are using the longest periodogram possible, given the length of the data. Of course, least squares techniques could be used on the raw periodogram, meaning that the second and last assumptions are just as likely to be satisfied as when using spectral likelihoods, but in this case, the asymptotic normality required for least squares will not be satisfied (nor in general will the constant variance assumption).

When it comes to the final assumption for both least squares and spectral likelihood techniques, there are some practical concerns that should be considered. In particular, when the aliased spectral density has high dynamic range, the frequencies are often correlated. As we shall shortly show, in the case of wind-generated waves, this issue does not present itself for 1Hz data. Although, for higher sampling frequencies, such as 4Hz data, the periodogram is often highly correlated. To solve this problem we can turn to differencing, a technique that is well established for reducing

correlations in the periodogram (Velasco and Robinson, 2000).

2.3.5 Differencing

If the periodogram is highly correlated across frequencies, spectral likelihoods will perform poorly when compared to full maximum likelihood (Velasco and Robinson, 2000). Differencing can sometimes provide a convenient mechanism for removing such correlations. Define the differenced process as $W(t) = X(t + \Delta) - X(t)$. We briefly switch notation and let $c_X(\tau)$ and $f_X(\omega)$ denote the autocovariance and spectral density function of X at τ and ω respectively, and likewise $c_W(\tau)$ and $f_W(\omega)$ for the differenced process W . First notice that

$$\begin{aligned} c_W(\tau) &= \mathbb{E} [W(\tau)W(0)] \\ &= \mathbb{E} [X(\tau + \Delta)X(\Delta) - X(\tau + \Delta)X(0) - X(\tau)X(\Delta) + X(\tau)X(0)] \\ &= 2c_X(\tau) - c_X(\tau + \Delta) - c_X(\tau - \Delta), \end{aligned}$$

by stationarity. Then from (2.2.1) we can see that

$$\begin{aligned} f_W(\omega) &= \int_{-\infty}^{\infty} (2c_X(\tau) - c_X(\tau + \Delta) - c_X(\tau - \Delta)) e^{-i\tau\omega} d\tau \\ &= 2f_X(\omega) - e^{i\omega\Delta} f_X(\omega) - e^{-i\omega\Delta} f_X(\omega) \\ &= 2(1 - \cos(\omega\Delta)) f_X(\omega) \\ &= 4 \sin^2 \left(\frac{\omega\Delta}{2} \right) f_X(\omega). \end{aligned} \tag{2.3.5}$$

Therefore, differencing can be easily incorporated into the fitting techniques that have been discussed in this chapter, by simply replacing X with the differenced process W and f_X with f_W using the relation given by (2.3.5). Consider the correlation matrix of the periodogram: the matrix with i, j th element defined to be the correlation between the periodogram at the i th and j th Fourier frequencies. Figure 2.3.1

shows a plot of the correlation matrix for the periodogram of a wind-generated wave process, estimated from 1000 realisations of the process in question. We can see that for data recorded at a 1Hz sampling rate there is little correlation in the periodogram; however, this is not the case for 4Hz data. We can also see that the periodogram of the differenced process is almost completely uncorrelated, even for the 4Hz data.⁹ From the signal processing perspective, this has reduced the dynamic range of the spectrum as we are multiplying the spectral density function by something that is close to zero for angular frequencies that are small, but is close to one near the Nyquist, down-weighting the peak far more than the tail.

2.4 Practical concerns for implementation with the generalised JONSWAP

In Section 2.3, we described some techniques that can be used to estimate model parameters. When implementing these techniques for ocean wave models, there are some practical concerns that must be addressed. Firstly, we need not use all of the Fourier frequencies when fitting the model. Indeed, it may be preferable to remove some frequencies that are contaminated by some other process or by observational noise. Secondly, there is no known analytical form for the autocovariance corresponding to many of the spectral density functions used when modelling ocean waves. Therefore, numerical techniques for estimating the autocovariance play an important role in many of the fitting procedures discussed in Section 2.3. In particular, it is necessary for both the debiased Whittle likelihood and for full maximum likelihood.

⁹It should be noted that the region of high correlation in the bottom left corner of each of the correlation matrices is part of the reason for removing such frequencies from the objective function when performing fits, as discussed further in Section 2.4.1.

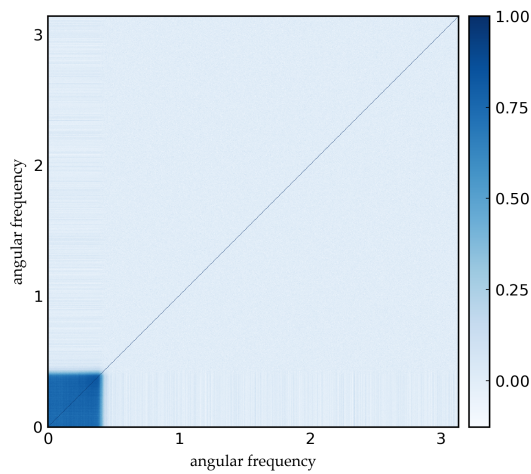
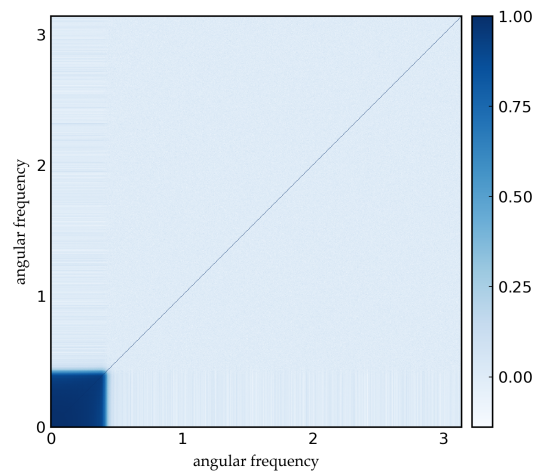
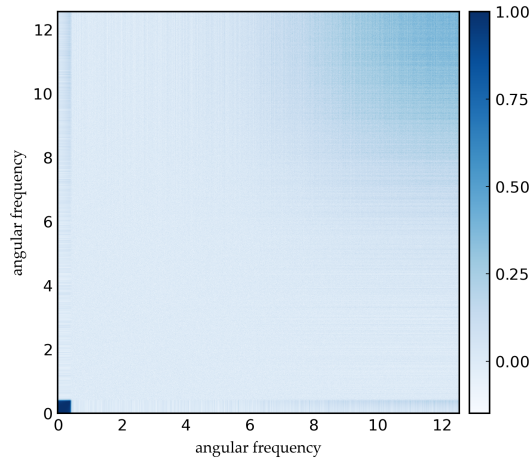
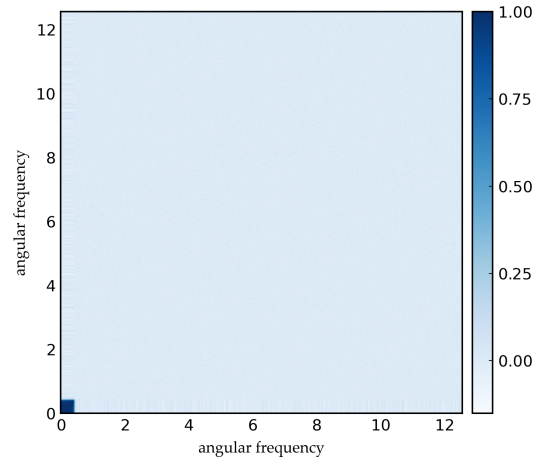
(a) No differencing and $\Delta = 1$.(b) Differencing and $\Delta = 1$.(c) No differencing and $\Delta = 1/4$.(d) Differencing and $\Delta = 1/4$.

Figure 2.3.1: Image plots of the correlation matrix of the periodogram of a wind-generated wave process for different values of Δ and the corresponding images for the differenced process, for a generalised JONSWAP with parameters $\alpha = 0.7$, $\omega_p = 0.7$, $\gamma = 3.3$ and $r = 5$.

2.4.1 Frequency selection

Many of the estimators defined in Section 2.3 involve minimising or maximising objective functions, which are expressed as the sum over some set of frequencies $\Omega \subseteq \Omega_{n,\Delta}$. The most simple choice for this set Ω is just the set of Fourier frequencies $\Omega_{n,\Delta}$. At first glance, this would seem like the most sensible choice (as omitting frequencies is essentially the same as throwing away data-points). However, there are many different circumstances in which it is preferable to remove some of the frequencies from the fit.

One practical reason for removing certain frequencies is that for very low frequencies, the generalised JONSWAP spectra is zero to machine precision. This often introduces numerical instabilities, especially for objective functions that involve dividing by the spectral density function (such as the Whittle likelihood). As can be seen in Figure 2.3.1, there is also a region of high correlation in the low frequencies, which provides an additional motivation for removing such frequencies. An alternative method to reducing correlations in the periodogram is to use tapered versions of the spectral density estimate in the Whittle likelihood (Dahlhaus, 1988), but in simulations (available on GitHub) we found omitting frequencies from the fit to be a better solution than tapering in terms of the resulting bias and variance of parameter estimates. Another reason for removing certain frequencies from the fit is that it can help to remove noise processes that are present in a record. For example, wave records often contain a low-frequency swell component, but we are interested in the parameters of the wind-sea component. By removing frequencies in which the swell is dominant, we are better able to model the wind-sea component of a sea state. On top of this, there is an added technical concern when using the Whittle and debiased Whittle likelihoods. The zero and Nyquist frequencies must be omitted (or a modified version of the summand must be used for those frequencies). This is because

these methods are based on the asymptotic distribution of the periodogram, which is different at the Nyquist and zero frequency than it is at other frequencies.

Fitting the model in this way can be thought of as fitting a semi-parametric model (as some of the frequencies are being modelled using a parametric model, and the remaining frequencies by some non-parametric model such as the periodogram). It is worth noting that this approach can actually be applied to full maximum likelihood as well. This can be achieved by transforming both the observations and autocovariance of the model into the frequency domain, applying a band pass filter, and then transforming back. While this is possible in theory, it is fiddly in practice and is no longer exact. This demonstrates another major advantage of spectral likelihoods: it is far easier to filter out undesired frequencies from the model fit. However, the choice of frequencies to be used in the fit should be made prior to the objective function being optimised. Otherwise the number of degrees of freedom could be changing throughout the optimisation routine, which would likely result in additional bias.

2.4.2 Numerical estimation of the autocovariance

To calculate both the multivariate Gaussian likelihood and debiased Whittle likelihood we require the autocovariance of the process given a certain parameter choice (in (2.3.2) for the multivariate Gaussian likelihood and (2.3.4) for the debiased Whittle likelihood). For the generalised JONSWAP spectra, there is no analytical form for this autocovariance. As such, the autocovariance must be approximated numerically. Firstly, recall that the autocovariance is the Fourier transform of the spectral density function, as defined in (2.2.3), and we wish to obtain the autocovariance at lags $0, \Delta, \dots, (n-1)\Delta$. The first problem we encounter is that this integral is over the entire real line. Clearly, we cannot approximate such an integral numerically

and must instead settle for integrating up to some finite frequency, such that the spectral density function beyond that frequency is sufficiently small. In particular, it is convenient to choose a multiple of the Nyquist frequency, as the integral will be approximated using a Fast Fourier Transform, so the desired lags can be extracted by sub-sampling if a multiple of the Nyquist is used in the integration. Therefore, based on equation (2.2.3), we can construct the approximate autocovariance

$$\hat{c}(\tau) = \int_{-L\pi/\Delta}^{L\pi/\Delta} f(\omega)e^{i\tau\omega} d\omega, \quad (2.4.1)$$

for $L \in \mathbb{N} = \{1, 2, 3, \dots\}$.

Alternatively, we could consider the relation given in equation (2.2.2), between the autocovariance and the discrete time spectral density function. In this case, we would first need to approximate the spectral density function for the discrete time process. To do this, we use a truncated version of the relation given by equation (2.2.4), between the spectral density of the continuous and discrete time processes. Consider the approximation of the aliased spectral density function given by

$$\tilde{f}_{\Delta}(\omega) = \sum_{k=-K}^K f\left(\omega + \frac{2\pi k}{\Delta}\right),$$

for $K \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Then we can construct the alternative approximation to the autocovariance given by

$$\tilde{c}(\tau) = \int_{-\pi/\Delta}^{\pi/\Delta} \tilde{f}_{\Delta}(\omega)e^{i\tau\omega} d\omega. \quad (2.4.2)$$

Notice that we may write

$$\begin{aligned}
 \tilde{c}(\tau) &= \int_{-\pi/\Delta}^{\pi/\Delta} \sum_{k=-K}^K f\left(\omega + \frac{2\pi k}{\Delta}\right) e^{i\tau\omega} d\omega \\
 &= \sum_{k=-K}^K \int_{-\pi/\Delta}^{\pi/\Delta} f\left(\omega + \frac{2\pi k}{\Delta}\right) e^{i\tau\omega} d\omega \\
 &= \sum_{k=-K}^K \int_{(2k-1)\pi/\Delta}^{(2k+1)\pi/\Delta} f(\omega) e^{i\tau\omega} d\omega \\
 &= \int_{(-2K-1)\pi/\Delta}^{(2K+1)\pi/\Delta} f(\omega) e^{i\tau\omega} d\omega. \tag{2.4.3}
 \end{aligned}$$

From (2.4.3) we can see that, if $L = 2K + 1$, then $\hat{c}(\tau)$ and $\tilde{c}(\tau)$ are equivalent. In practice, these integrals must be approximated numerically. To do this, we consider a Riemann approximation with bins of width $2\pi/m\Delta$. By choosing m to be some integer bigger than $2n$, we can obtain the desired lags by performing a Fast Fourier Transform and then sub-sampling appropriately.

We can now see that the approximation based on $\hat{c}(\tau)$ (in (2.4.1)) can be computed in $O(Lm \log Lm)$ time. However, computing the second approximation, based on $\tilde{c}(\tau)$ (in (2.4.2)), requires first computing $\tilde{f}(\omega)$ at m frequencies (taking $O(Lm)$ operations) and then performing a Fourier transform on m frequencies, requiring $O(m \log m)$ operations. In other words, the first approach requires $O(m(L \log L + L \log m))$ operations, whereas the second only requires $O(m(L + \log m))$ operations. For this reason, we use the latter approach when approximating the autocovariance: first approximating the aliased spectral density, then approximating the autocovariance.

The choice of K (or equivalently L) depends on the tail behaviour of the spectral density function in question. In practice, we choose K so that for frequencies beyond $(2K + 1)\pi/\Delta$, the spectral density is below some threshold (e.g. $10^{-6} \text{ m}^2 \text{ s rad}^{-1}$), though K should really be chosen so that it scales with n , for convergence results to still apply. The choice of m is based on the required accuracy of the integral approx-

imation and should be tuned accordingly. For the generalised JONSWAP, we have found that $m = \max\{8192, 2n\}$ is a good choice.

2.5 Simulation study

Though it is possible to make theoretical statements about the asymptotic behaviour of different estimators, from a practical perspective, their finite sample behaviour is of primary interest. To investigate this, we perform a simulation study to assess the performance of the estimators described in Section 2.3. In this simulation study, we compare six different fitting techniques based on these estimators. The first, which we call least squares, uses the curve fitting approach with the periodogram. The second approach is similar, but uses Bartlett's method to estimate the spectral density function, which we refer to as Bartlett least squares. The window size is chosen so that we have a spectral resolution of 0.2π , i.e. the window size is $100/\Delta$.¹⁰ For 1.28Hz data, this corresponds to a window size of 128. The third and fourth approaches are the Whittle and aliased Whittle likelihoods respectively. The final two approaches are the debiased Whittle likelihood and full time domain maximum likelihood.

2.5.1 Method

To investigate the effectiveness of different fitting approaches we simulate a linear wave record with a known parametric spectral density function and then re-estimate the parameters from the simulated record. By repeatedly performing this process, we can assess the bias and variance of each of the estimators discussed in Section 2.3. For the purposes of the simulation study we let $X_{\Delta,n}$ be a random variable with a multivariate normal distribution resulting from sub-sampling the continuous-time

¹⁰Clearly for some values of Δ this would not be an integer; however, for the values of Δ that we choose it is.

mean-zero stationary Gaussian process X , where X has spectral density function $f_G(\omega; \theta)$, defined by (2.2.6). We then simulate a realisation of $X_{\Delta, n}$ using the circulant embedding method described by Davies and Harte (1987) (and for complex valued processes by Percival (2006)). We choose to use circulant embedding over the typical approaches for simulating Gaussian processes often used in the ocean waves literature, such as the method due to Tucker et al. (1984), as these methods only approximately simulate a Gaussian process with the given spectral density function, whereas circulant embedding is exact (up to the quality of the approximation of the autocovariance that is used). Furthermore, many techniques, such as the method proposed by Tucker et al. (1984), or the more recent modification due to Méri­gaud and Ringwood (2018) do not account for aliasing when simulating the process. Since we are explicitly interested in the effect that aliasing has on recovered parameters, it is important that we simulate something that is as close as we can get to a Gaussian process with the desired aliased spectral density function. Circulant embedding generates time series with all of the sampling effects discussed by Tucker et al. (1984), but also includes additional finite sampling effects such as aliasing and correlations between spectral estimates at different frequencies. Such effects should be present in generated time series, but are not in time series generated using the method suggested by Tucker et al. (1984). More details can be found in Davies and Harte (1987); Dietrich and Newsam (1997); Wood and Chan (1994).

To perform the fitting we first choose one of the objective functions described in Section 2.3 and optimise this using the `fmincon` function in MATLAB (with maximisation done by minimising the negative of the objective function). An initial guess for the fitting procedure needs to be provided for each of the parameters. For ω_p , we use the frequency corresponding to the largest value of the periodogram. For r , we use a basic linear regression coefficient between the log spectral density and log periodogram over the tail frequencies (where the tail is chosen to be all frequencies that

are closer to the Nyquist than the peak). We choose to initialise γ by setting it equal to 3. This is because choosing γ heuristically is not easy, and $\gamma = 3$ is close to the value commonly assumed by many oceanographers. In practice, the initial choice of γ does not seem to have a huge impact on the final fitted values; however, the optimisation could also be run with multiple starting values of γ and the best estimate could then be selected. Once these parameters are initialised, α is initialised so that the area under the initial parametric spectral density function matches the area under the periodogram. In simulations, we find that the inference is not sensitive to the initial guess (provided it is sensible). In practice we are often fitting models to multiple consecutive sea states. In this case, it can be more efficient to use the parameter estimates for the previous sea state as initial values when optimising.

2.5.2 A canonical sea state

We shall begin by considering how each of the estimators perform for one choice of true parameters, before showing that the results are robust to the true parameters. In particular, we begin by considering a spectral density function of the form described in Section 2.2.2, with $\sigma_1 = 0.07$, $\sigma_2 = 0.09$ and $s = 4$ treated as known, and with $\alpha = 0.7$, $\omega_p = 0.7$, $\gamma = 3.3$ and $r = 4$ treated as unknown parameters to be estimated. The reason for choosing these parameters is that $\alpha = 0.7$ roughly corresponds to the scaling present when using Phillip's constant in a JONSWAP spectra, $\omega_p = 0.7$ is a reasonable choice for peak frequency, $\gamma = 3.3$ is commonly assumed to be the peak enhancement factor, and $r = 4$ is one of the suggested values for the tail decay index. Half hour records sampled at 1.28Hz (a standard time interval and sampling frequency for wave records) were simulated and the parameters were estimated using each of the six estimation methods described above. The resulting estimates across 1000 repeated simulations are summarised in Figure 2.5.1, alongside the time taken to perform the optimisation. For comparison, the true value of

each parameter is given by a horizontal red dashed line.

Perhaps the most striking feature of Figure 2.5.1 is the difference in the variability in estimates of the tail parameter, r , when comparing least squares type techniques to likelihood based techniques. Least squares techniques recover parameter estimates ranging from well beyond three to five, making it very difficult to make any statements about the true value of the tail decay. However, we can see that statistical techniques such as the debiased Whittle likelihood are able to recover the original tail parameter to within a few decimal places. Therefore, by using the debiased Whittle likelihood, practitioners would be able to distinguish between ω^{-4} and ω^{-5} spectral tails in observed records. Though it should be noted that this assumes the wind-sea had a spectral density that is well described by a generalised JONSWAP and so we cannot provide model free estimates of the tail decay. In particular, a different model would be required for a transition in tail decays over frequency, such as the effect discussed by Babanin (2010); however, the debiased Whittle likelihood is generic and could be applied to other models provided they satisfied certain conditions (see Appendix A.3). We can also see that the debiased Whittle likelihood offers an improvement in estimates of γ , performing almost as well as full maximum likelihood.

It is also interesting that bias can be seen in both the Whittle and aliased Whittle likelihood estimates, but that this bias is not present in the debiased Whittle likelihood estimates. This verifies that the debiased Whittle likelihood is indeed accounting for some of the bias present in standard Whittle likelihood, and demonstrates why debiased Whittle likelihood is necessary over the aliased Whittle likelihood, which can still be seen to be biased for some parameters. In the analysis of Figure 2.5.1 full maximum likelihood provides the best estimates, in terms of root mean square error. However, this comes at significant computational cost, whilst giving limited

improvement in bias and variance when compared to the debiased Whittle likelihood.

Often, during optimisation, parameters may trade off against one-another. Therefore it is also important to look at the joint behaviour of parameter estimates. Figure 2.5.2 shows a scatter plot of the debiased Whittle likelihood estimates from Figure 2.5.1. We can see that there is very strong correlation between the estimates of α and r , and some negative correlation between α and γ . This likely occurs because γ and r change the area under the spectral density function, so α is likely to be adjusted to compensate. Though it would be possible to reparameterise to try and avoid this, it does not seem to have a significant impact on the resulting estimates and is therefore not necessary.

In practice, longer sea states are often used to estimate model parameters. Therefore, we also compare some of the methods for 3 hour records. Figure 2.5.3 shows the comparison of least squares, Bartlett least squares and debiased Whittle likelihood estimates for these 3 hour records. The variance in the first two estimators has indeed decreased when compared to the estimates from half hour records shown in Figure 2.5.1. However, by comparing the debiased Whittle likelihood estimates in Figure 2.5.1 to the least squares estimates in Figure 2.5.3, we can see that the debiased Whittle likelihood used on a half hour record yields better estimates than the least squares based estimates performed on 3 hour records. The longer record reduces the variance of the least squares and Bartlett least squares techniques enough to allow us to see another interesting feature, namely that there is significant bias present in the Bartlett least squares estimates that is not present in the standard least squares estimates. This demonstrates that non-parametric smoothing can have unexpected consequences when used to fit a parametric spectral density function.

In essence, by using the debiased Whittle likelihood, we can obtain more accurate

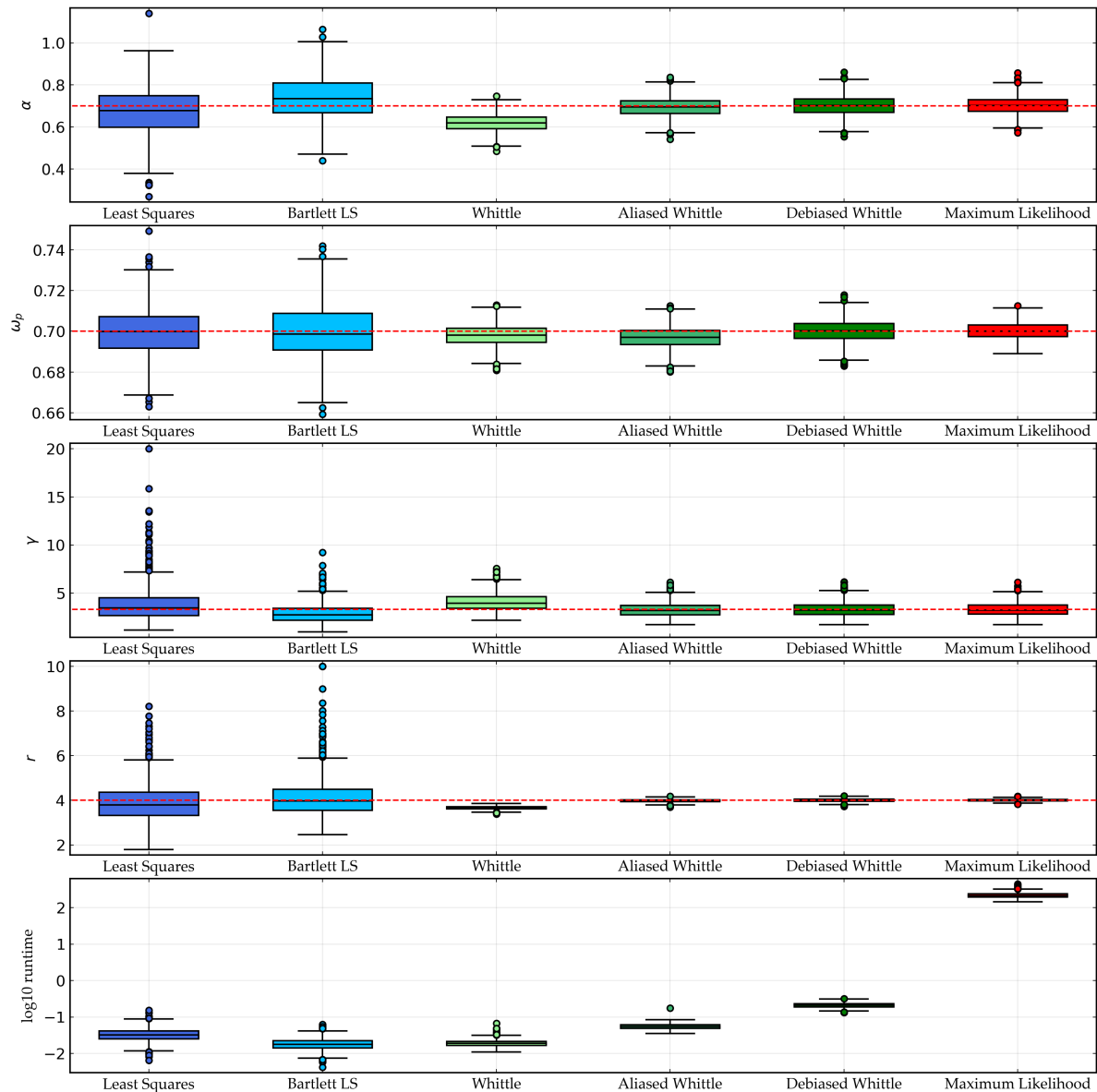


Figure 2.5.1: Boxplots of parameter estimates and time taken for six different fitting routines with true parameters denoted by the dashed red lines. Each row displays the results for a given fitting technique, as well as the log of the time taken to perform the optimisation, recorded in seconds. The fitting was performed on simulations of 1.28Hz observations recorded for half an hour (2304 observations). The process was repeated 1000 times.

and precise parameter estimates, whilst simultaneously reducing the length of the record required to obtain the estimates. This improvement is especially noticeable (and important) for the peak enhancement factor, γ , and tail decay index, r . In practice, this has two important consequences. Firstly, we can reduce the amount of time for which the surface is assumed to be stationary. This means that we can fit stationary models to weather systems that evolve very quickly, such as tropical cyclones. Secondly, we can obtain parameter estimates at more frequent time intervals. This higher parameter resolution means that we gain a more detailed insight into how certain parameters evolve throughout the course of a meteorological event.

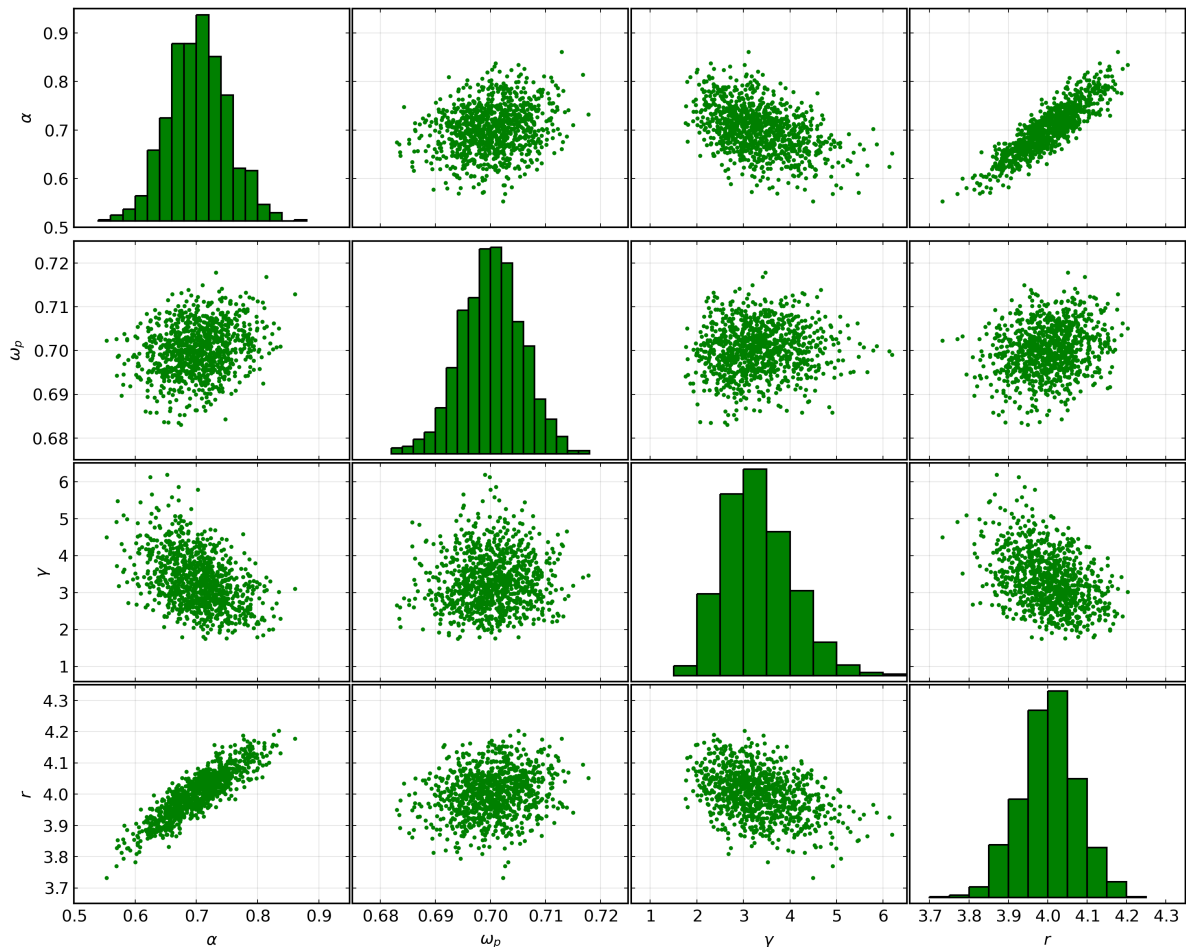


Figure 2.5.2: Scatter plot of the debiased Whittle estimates obtained from half hour records shown in Figure 2.5.1.

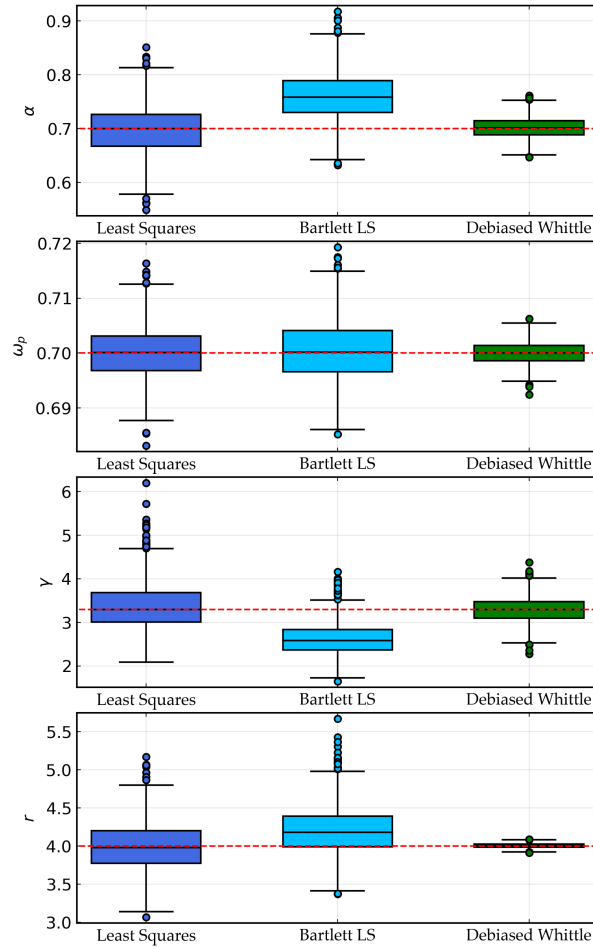


Figure 2.5.3: Boxplots of 1000 parameter estimates using three different fitting routines on the same simulations using the same true parameters as Figure 2.5.1 and same sampling interval but simulating 3 hour records.

2.5.3 Robustness of results

In Section 2.5.2 we have demonstrated that the debiased Whittle likelihood can produce parameter estimates that are both more accurate and more precise than those produced by least squares techniques, without making huge sacrifices in terms of computational time. However, it is also important to check that these results extend to different choices of the true parameter.

In Table 2.5.1, we present the results of a simulation study comparing least squares, Bartlett least squares and the debiased Whittle likelihood over 24 different choices of

true parameter. For each choice of true parameter, we calculated the percentage bias, standard deviation (SD) and root mean square error (RMSE) of the estimates (relative to the true parameter). We then averaged over all the choices of true parameter, yielding an average percent bias, SD and RMSE for each parameter. The parameter choices used to perform these simulations were all the combinations of $\alpha = 0.7$; $\omega_p = 0.7, 0.9, 1.2$; $\gamma = 1, 2, 3.3, 5$; and $r = 4, 5$. Boxplots for each set of parameter choices are available on GitHub (Grainger, 2021). We can see from Table 2.5.1 that there are substantial reductions in both the bias and standard deviation of all the estimated parameters.

	Bias (%)			SD (%)			RMSE (%)		
	LS	BLS	DW	LS	BLS	DW	LS	BLS	DW
α	4.70	10.88	0.80	34.64	30.69	9.13	35.46	33.49	9.19
ω_p	0.08	0.11	0.06	1.73	1.85	0.77	1.74	1.86	0.77
γ	15.88	10.20	2.98	43.29	29.27	18.00	46.39	31.75	18.71
r	3.89	5.60	0.18	17.57	15.76	2.09	18.35	17.44	2.11
average	6.14	6.70	1.01	24.31	19.39	7.50	25.49	21.13	7.69

Table 2.5.1: Average percentage bias, standard deviation and root mean squared error across 24 different parameter choices calculated from 1000 repetitions per parameter choice for least squares (LS), Bartlett least squares (BLS) and debiased Whittle likelihood (DW). Parameter choices were all the combinations of $\alpha = 0.7$; $\omega_p = 0.7, 0.9, 1.2$; $\gamma = 1, 2, 3.3, 5$ and $r = 4, 5$. Simulated records were half an hour long and sampled at 1.28Hz. The bottom row shows the average of these quantities over all the parameters in the model.

True parameters on the boundary of the parameter space can cause problems when performing parametric estimation. As such, we shall further discuss the results for the special case when $\gamma = 1$. The fitted parameters for half hour simulated records with true parameter values $\alpha = 0.7$, $\omega_p = 0.7$, $\gamma = 1$ and $r = 5$, using the least squares, Bartlett least squares and debiased Whittle likelihood techniques can be seen in Figure 2.5.4. This is an interesting case because not only does the true parameter lie on the boundary of the parameter space, but this value of γ corresponds to a Pierson-Moskowitz spectrum, for a fully developed sea. This means that such

a value of γ could occur in nature, and as such it is important that we can model this case. The problem is that the theoretical guarantees for an approach such as the debiased Whittle likelihood rely on the assumption that the true parameter does not lie on the boundary of the parameter space. Since this is not the case for $\gamma = 1$, therefore we must take care when dealing with records for which the true value of γ may indeed be 1. In particular, we can see from Figure 2.5.4 that the debiased Whittle likelihood estimates for γ is not normally distributed, a result that is assumed when constructing the confidence intervals that are described later in Section 2.6. One way to deal with this problem is to fit both a model with $\gamma = 1$ fixed and one with γ as a free parameter, and then test to see if there is evidence for $\gamma > 1$. Such a procedure allows us to circumvent potential issues caused by γ lying on the boundary of the parameter space. This could be performed by using the procedure developed by Sykulski et al. (2017), adapted to the 1D case.

2.5.4 Differencing for high sampling frequencies

As we discussed in Section 2.3.5, for wind generated waves observed at a 4Hz sampling rate, the periodogram is highly correlated. As noted in Section 2.3.4, in this case, we would expect spectral techniques to perform poorly compared to full maximum likelihood. Differencing can reduce the correlation in the periodogram, and therefore can be a powerful tool to remove bias from spectral methods. Figure 2.5.5 shows box plots of estimated parameters for least squares, Bartlett least squares and debiased Whittle likelihood techniques, both with and without differencing, for both 1Hz and 4Hz data (in this case, 2048 seconds worth of data was simulated per record, i.e. 2048 observations for the 1Hz simulation and 8192 for the 4Hz simulation). We can see little benefit from differencing in the 1Hz estimates (Figure 2.5.5a). However, when it comes to 4Hz data, there is a major benefit to differencing, especially for the debiased Whittle likelihood. Because there is little difference between the debiased

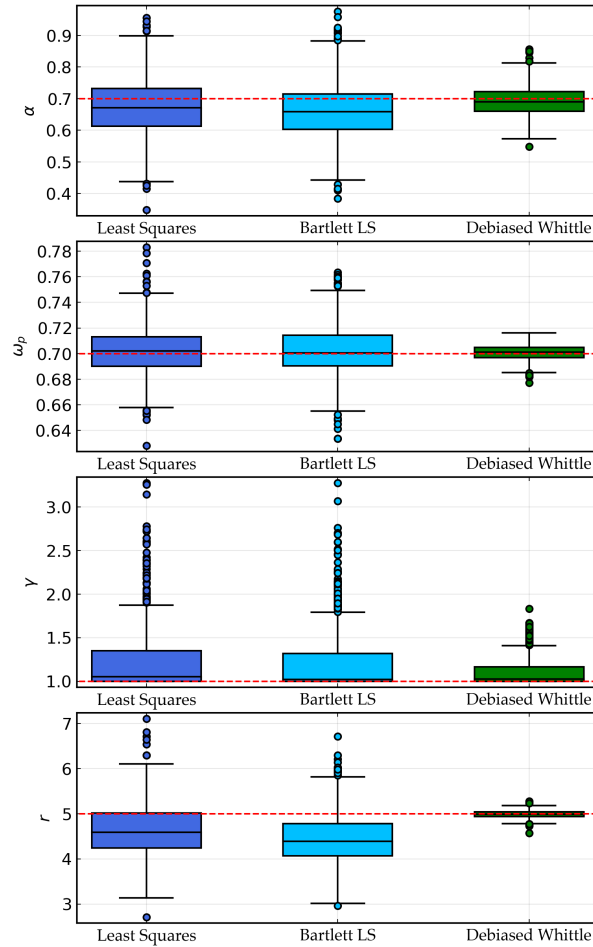
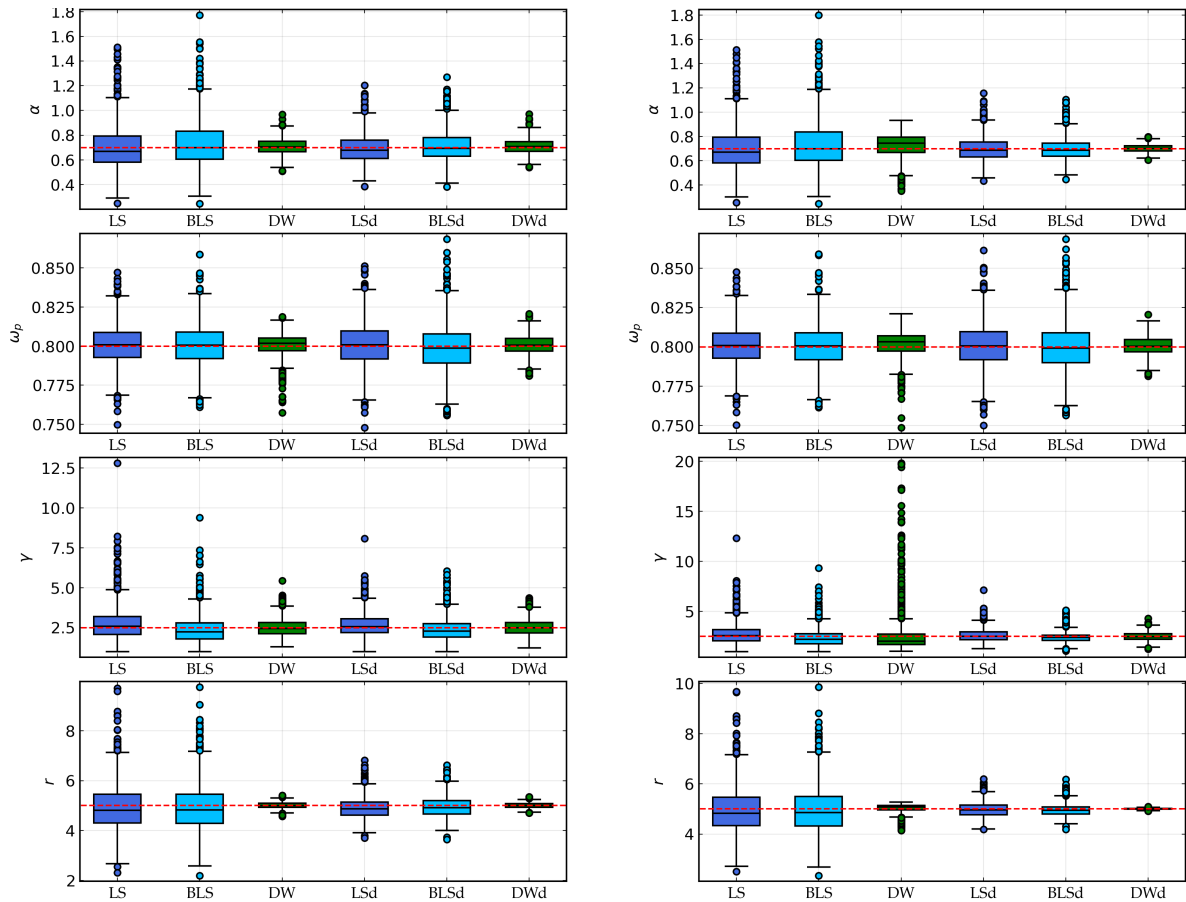


Figure 2.5.4: Boxplots of 1000 parameter estimates using three different fitting routines with true parameters $\alpha = 0.7$, $\omega_p = 0.7$, $\gamma = 1$ and $r = 5$, estimated from simulated half hour records sampled at 1.28Hz.

Whittle fits on the original and differenced 1Hz data, we would recommend that differencing is used as standard, to protect against the issues seen in Figure 2.5.5b. At the very least, investigating the correlation matrix of the periodogram should be an important diagnostic when fitting spectral models.

2.6 Quantifying estimation uncertainty

As with any statistical analysis, it is important to quantify the uncertainty in the parameter estimates that are obtained using the debiased Whittle likelihood. Sykulski et al. (2019) developed an approach for quantifying the asymptotic variance of the



(a) Parameter estimates for 1Hz data.

(b) Parameter estimates for 4Hz data.

Figure 2.5.5: Boxplots of 1000 parameter estimates for 1Hz and 4Hz data. LS, BLS, DW denote least squares, Bartlett least squares and debiased Whittle respectively, and LSd, BLSd and DWd denote the differenced versions.

estimator for a single time series. Say that we have p parameters that are written in a vector $\underline{\theta}$ and are interested in the debiased Whittle likelihood estimator $\hat{\underline{\theta}}_{DW}$ of the true parameter vector $\underline{\theta}$. Sykulski et al. (2019) decompose the variance as

$$\text{var}(\hat{\underline{\theta}}_{DW}) \approx \mathbb{E}[\mathbf{H}(\underline{\theta})]^{-1} \text{var}(\nabla \ell_{DW}(\underline{\theta})) \mathbb{E}[\mathbf{H}(\underline{\theta})]^{-1}, \quad (2.6.1)$$

where $\mathbf{H}(\underline{\theta})$ denotes the matrix of second derivatives of the log-likelihood function $\ell_{DW}(\underline{\theta})$, and ∇ denotes the vector of first derivatives of $\ell_{DW}(\underline{\theta})$.

Now for the debiased Whittle likelihood, we may write

$$\frac{\partial}{\partial \underline{\theta}_j} \ell_D(\underline{\theta}) = - \sum_{\omega \in \Omega} \left\{ \left(\frac{1}{\bar{f}_n(\omega; \underline{\theta})} - \frac{I(\omega)}{\bar{f}_n^2(\omega; \underline{\theta})} \right) \frac{\partial}{\partial \underline{\theta}_j} \bar{f}_n(\omega; \underline{\theta}) \right\},$$

for $j = 1, \dots, p$. As will be seen in Appendix A.2, this is required to calculate $\text{var}(\nabla \ell_{DW}(\underline{\theta}))$. Furthermore, we may write

$$\mathbb{E} \left[\frac{\partial^2}{\partial \underline{\theta}_j \partial \underline{\theta}_k} \ell_D(\underline{\theta}) \right] = - \sum_{\omega \in \Omega} \left\{ \frac{1}{\bar{f}_n^2(\omega; \underline{\theta})} \frac{\partial}{\partial \underline{\theta}_j} \bar{f}_n(\omega; \underline{\theta}) \frac{\partial}{\partial \underline{\theta}_k} \bar{f}_n(\omega; \underline{\theta}) \right\},$$

for $j, k = 1, \dots, p$. Therefore, we require only the first derivative of the expected periodogram in order to compute both parts of the variance decomposition given by (2.6.1).

Furthermore, the triangle function $(1 - \tau/n)$ and Fourier basis are constant with respect to $\underline{\theta}$, so from (2.3.4) we have that

$$\frac{\partial}{\partial \underline{\theta}_j} \bar{f}_n(\omega; \underline{\theta}) = \frac{\Delta}{2\pi} \sum_{\tau=-n+1}^{n-1} \left(1 - \frac{|\tau|}{n} \right) \frac{\partial c(\tau; \underline{\theta})}{\partial \underline{\theta}_j} e^{-i\tau\Delta\omega},$$

for $j = 1, \dots, p$. In other words, to calculate the derivative of the expected periodogram, we may first calculate the derivative of the autocovariance, then calculate

the expected periodogram of the process with that derivative as its autocovariance. In the case of the generalised JONSWAP spectral density function, when an analytical form for such autocovariance is unavailable, we may instead approximate the derivative of the autocovariance by first approximating the derivatives of the aliased spectral density function, and then of the autocovariance.

In Appendix A.1 we show that, for the generalised JONSWAP form, partial derivatives of $f_{\Delta}(\omega; \underline{\theta})$ can be calculated from the derivatives of the spectral density function, and that the resulting derivatives are continuous. Therefore, by Leibniz's rule

$$\frac{\partial}{\partial \underline{\theta}_j} c(\tau; \underline{\theta}) = \int_{-\pi/\Delta}^{\pi/\Delta} \frac{\partial}{\partial \underline{\theta}_j} f_{\Delta}(\omega; \underline{\theta}) e^{-i\tau\omega} d\omega.$$

As a result, we may approximate the derivative of the autocovariance from the derivative of the aliased spectral density function in the same way that we approximated the autocovariance from the aliased spectral density function in Section 2.4. Therefore, each partial derivative can be computed in the same time it would take to do one function evaluation, which is faster than using numerical approximations for the derivative (e.g. finite differencing methods).

In Appendix A.2, we discuss a novel procedure for estimating $\text{var}(\nabla \ell_{DW}(\underline{\theta}))$ in a computationally efficient manner. Combining this with the approach for obtaining derivatives, we can estimate the variance of the debiased Whittle likelihood estimator using (2.6.1). This enables the computation of approximate confidence intervals for our estimates, by using the asymptotic normality of the estimator and standard theory for confidence intervals.

2.7 Humboldt Bay data

In this section, we analyse data recorded as part of the Coastal Data Information Program (CDIP), to address some of the practical concerns when applying the techniques discussed in this chapter to observed wave records. Data were furnished by the Coastal Data Information Program, Integrative Oceanography Division, operated by the Scripps Institution of Oceanography, and is freely available on their website. The data we use were recorded at CDIP station 168, Humboldt Bay North Spit offshore California (40.8883, -124.3567), with a water depth of 113m. Data were recorded using a Datawell Waverider MkIII buoy (Datawell, 2006) with a sampling rate of 1.28Hz. Significant wave height and a spectrogram¹¹ on a decibel scale are shown in Figure 2.7.1. The vertical dashed lines on the spectrogram delimit the start and end of the period we analyse, and the dotted lines show the frequency region we used to fit the wind-sea model.

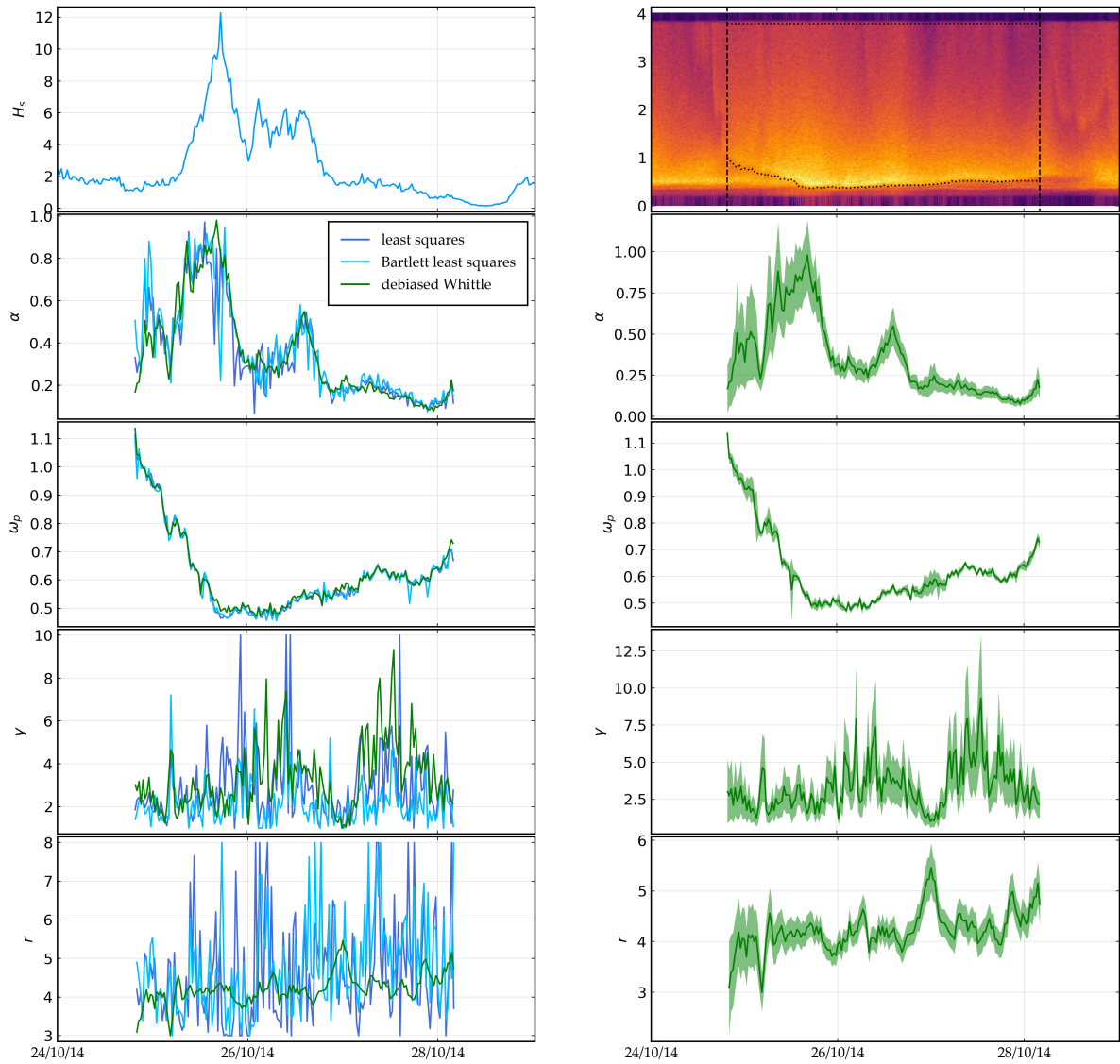
For the sake of illustration, we shall analyse the behaviour of spectral parameters throughout a single storm event, from the 24th to the 29th of October, 2014.¹² As can be seen from Figure 2.7.1, from the start of the analysis period a high frequency component begins to develop and then transitions to lower frequencies, increasing in magnitude. This is a classic example of a developing wind-sea. We also see the presence of persistent swell, which dies down as time goes on. Parameters of the wind-sea component of these bi-modal seas are estimated by first removing contaminated frequencies from the objective function, with optimisation proceeding as previously described, as seen in the top panel of Figure 2.7.1b. Ideally we would aim to fit a bi-modal model that was designed to also describe the swell component.

¹¹Computed using multitapering half-overlapped 30 minute records, we used $nW = 4/\Delta$ and $K = 7$ to compute the multitaper estimates.

¹²Note that this is not intended to make general statements about the behaviour of these parameters. Rather, we are illustrating some of the additional practical concerns faced when analysing recorded wave time series.

However, such an approach is beyond the scope of this chapter, though we note that preliminary results for such a procedure are encouraging and are discussed further in Section 2.8. It was also observed that frequencies beyond 3.8 rad s^{-1} exhibited a different behaviour from the rest of the spectral tail. We suspect this may be related to the physical response of the buoy to the waves, though we do not know exactly how such behaviour should be modelled. We have seen similar effects in buoy data from a variety of sources, including in the Maui data analysed by Grainger et al. (2021), and the North Sea data analysed in Chapter 3. Therefore, we also chose to omit those frequencies in the spectral fitting procedure when using each of the fitting procedures (see Chapter 3 for more details on this problem). After filtering out the aforementioned frequencies, a generalised JONSWAP spectral form was fitted to each of the sea states in turn, using least squares, Bartlett least squares, and the debiased Whittle likelihood. These estimates are shown in Figure 2.7.1a.

From Figure 2.7.1a, we can see that, especially for the tail decay index, the debiased Whittle estimates seem more stable than both the least squares and Bartlett least squares estimates. Figure 2.7.1b shows the debiased Whittle likelihood estimates and approximate 95% point-wise confidence intervals, calculated using the technique described in Section 2.6. It is worth noting that these confidence intervals are based on the asymptotic distribution of the estimator, in particular, we assume asymptotic normality of the estimator. This is especially problematic for values near to or on the boundary of the parameter space, as the confidence interval may include values outside the boundary, so care should be taken in interpreting the confidence intervals when γ is close to 1. Such confidence intervals enable us to better understand and communicate the uncertainty surrounding parameter estimates. This uncertainty, we argue, should be considered when using these parameters as inputs for other related models.



(a) Parameter estimates using least squares, Bartlett least squares and debiased Whittle likelihood.

(b) Parameter estimates using debiased Whittle likelihood with approximate 95% confidence intervals.

Figure 2.7.1: Plots of H_s , the spectrogram on a decibel scale and the estimated parameters over time. Note that some of the plots in (a) have not been truncated, but that upper bounds were set on the optimisation at 10 for γ and 8 for r .

From Figure 2.7.1b we can see the parameters behave broadly as expected through the storm. We see a steady evolution in peak frequency throughout the growth of the wind-sea component, alongside increasing values of γ as the storm develops. We see that γ falls off slightly before increasing again later in the record. Despite a few anomalous points, the estimated parameters seem physically reasonable.

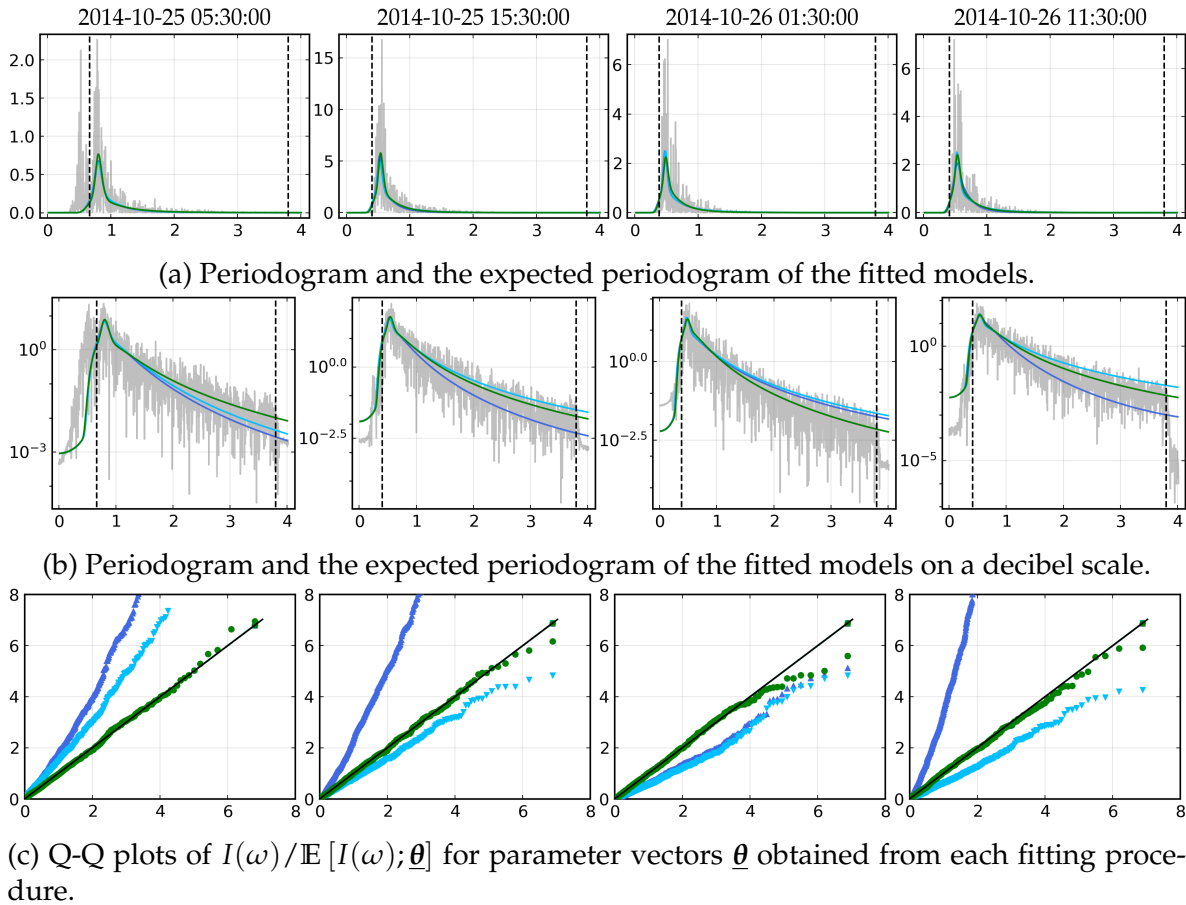


Figure 2.7.2: Plot of periodograms and fitted models on normal and decibel scale, alongside diagnostic Q-Q plots. The times of each record are given at the top of (a). Colours and line styles are used to denote different fitting techniques in the same manner as Figure 2.7.1a. The symbols “ \circ ”, “ \triangle ” and “ ∇ ” denote quantiles from the debiased Whittle likelihood, least squares estimates, and Bartlett least squares estimates, respectively. Note that the plots in (c) have been truncated in the y axis so that they have a 1:1 aspect ratio, so some quantiles are not shown.

A useful diagnostic tool for spectral models can be constructed by noticing that (under certain technical conditions) $I(\omega)/\mathbb{E}[I(\omega); \theta]$ should be approximately exponentially distributed with mean 1 (Brockwell and Davis, 2006), where $I(\omega)$ denotes the *periodogram* (this result is specific to the periodogram and is not true in general for non-parametric spectral density estimators). Therefore, if we calculate this ratio for a given model at each Fourier frequency, we can then compare the quantiles of these fractions to the quantiles of an exponential distribution with mean 1 using a Q-Q plot. Figure 2.7.2a and 2.7.2b show the periodograms and fitted models on both the

standard scale and the decibel scale respectively, for 4 time periods during the early development of the storm. Figure 2.7.2c shows the corresponding Q-Q plots, which demonstrate how well each of the models fits the data. The closer the points on the Q-Q plot are to the reference line $y = x$, the better. We can see from Figure 2.7.2c that the debiased Whittle likelihood consistently outperforms the other techniques.

2.8 Discussion and conclusion

The debiased Whittle likelihood has been shown to yield major improvements in both the bias and variance of estimated parameters for wind-generated waves. In particular, the tail decay index can be estimated to much greater levels of accuracy and precision than when using least squares techniques. Such an improvement will enable reliable tracking of the tail decay index's behaviour throughout the course of meteorological events, allowing oceanographers to gain fresh insights into the behaviour of wind-generated waves. We have also demonstrated some improvement in the estimation of the peak enhancement factor. The debiased Whittle estimator recovers estimates that are of similar quality to full maximum likelihood, which can be thought of as optimal. Since information about the peak enhancement factor is contained in a small region around the peak frequency, it is not surprising that it is so hard to estimate. Because there is significant variability in estimates of the peak enhancement factor, it is essential that we can describe uncertainty surrounding parameter estimates when performing an analysis. For this reason, the development of computationally efficient techniques for quantifying uncertainty surrounding the estimated parameters is important. We have shown that the method presented by Sykulski et al. (2019) for estimating the variance of the debiased Whittle estimator can be modified so that it can be computed using 2D Fast Fourier Transforms. Combining this with an analytical approach for calculating derivatives, we are able to

calculate the uncertainty in parameter estimates accurately and quickly. In addition, differencing can be used so that we can cope with high sampling frequency data, which tend to be correlated in the frequency domain.

As we discussed previously, when performing simulations we chose to use circulant embedding (Davies and Harte, 1987) to obtain realisations of Gaussian processes with the desired covariance matrix. This is different to the standard method used for simulating linear ocean waves, presented by Tucker et al. (1984) (or to the adapted version due to Mérigaud and Ringwood (2018), which is preferable). The first difference is that the standard waves method is only approximate, and does not exactly simulate a Gaussian process with the desired covariance (equivalently spectral density). The second is that these methods do not account for the aliasing that we would expect to be present in a record, essentially treating the spectral density as if it is zero beyond the Nyquist frequency. Though for many applications this will not matter, the methods discussed in this chapter will be sensitive to this difference. If the simulated record does not have the aliasing that should be present, then the parameters estimated using the debiased Whittle likelihood will seem biased (and standard Whittle will often seem better). This is important because if a method such as Tucker et al. (1984) was used to perform the simulation study described in this chapter, the results would be different, as likelihood based estimation will be sensitive to this problem (which could be thought of as model miss-specification, as we are trying to fit a model with a non-zero density beyond the Nyquist to a process that has been simulated with no density above the Nyquist). Least squares is somewhat invariant to this problem, as it mainly effects frequencies where the spectral density function is small, and least squares is not heavily influenced by such frequencies.

We have also developed a MATLAB toolbox, implementing the methods discussed in this chapter. This is available, alongside code to generate the figures in this chap-

ter and additional supplementary figures, on GitHub (Grainger, 2021). The toolbox contains code to perform each of the fitting techniques discussed in this chapter on arbitrary processes, as well as a function implementing the generalised JONSWAP (including first and second derivatives), which can then be used straight out the box. On top of this, the user may provide any spectral density function they wish, provided it satisfies the assumptions in Appendix A.3, and then use the toolbox to obtain parameter estimates from observations. Alongside this, an implementation of circulant embedding is provided, enabling exact simulation of a desired Gaussian process. Note that this has now be superseded by the Julia package WhittleLikelihoodInference.jl (Grainger, 2022c), which provides not only univariate Whittle methods, but also multivariate (which we will make use of in Chapter 3). The models for this paper and more are implemented OceanWaveSpectralFitting.jl (Grainger, 2022b).

Though the debiased Whittle likelihood has been seen to perform well in the estimation of wind-seas, we have yet to fully explore its potential when we wish to describe multi-modal seas (e.g. including one or more swell components). In this chapter, we were able to describe the wind-sea component of such sea states by first removing swell with a high pass filter. However, it would be preferable to develop and fit models that were bi-modal themselves, avoiding the need for partitioning schemes to determine which frequencies should be filtered. Indeed, such a procedure could be extended to describe seas with any number of components, using model selection to determine the number of components that are actually present. Such techniques could also allow for the development of model-based partitioning schemes, that could to separate overlapping wind-sea and swell components. In the development of such multi-modal models, it would be important to consider the interactions between different component weather systems. In particular, techniques such as higher-order spectral analysis could be used to determine if any non-linear

interactions were present and help characterise them. Such interactions could then also be parameterised and fitted using similar techniques to those discussed in this chapter.

Another important aspect of wind-generated waves is their directional characteristics. One approach to describing these is to assume some dispersion relation between wave-number and frequency, and then look at frequency direction spectra, which can be estimated, for example, from heave-pitch-roll buoys (Longuet-Higgins et al., 1963). We discuss this further in Chapter 3.

In summary, we have demonstrated that, by using the debiased Whittle likelihood, we are able to obtain more accurate and precise estimates of parameters for the wind-sea component of a wind-generated wave process in $O(n \log n)$ time. Using differencing, we are able to overcome correlations in the periodogram that are common in data sampled at a high frequency. Furthermore, we have described a procedure which can be used to obtain the variance of such estimates in $O(n^2 \log n)$ time. Together these developments will improve the tools available to practitioners, both in terms of fitting models to data and describing the associated uncertainty.

Chapter 3

A multivariate pseudo-likelihood approach to estimating directional ocean wave models

The content of this chapter has been submitted for publication as Grainger, J. P., Sykulski, A. M., Ewans, K., Hansen, H. F., and Jonathan, P. (2022). A multivariate pseudo-likelihood approach to estimating directional ocean wave models. With the exception of some minor notational changes, the content of the chapter is unchanged from that of the paper. As in Chapter 2, we use slightly simplified notation where appropriate. Section 3.1 motivates the chapter, ending with an overview of the chapter's structure.

3.1 Introduction

Applications of multivariate time series and spatiotemporal statistics are ubiquitous, for example using the affordable and widespread availability of GPS and accelerom-

eter technology to track individuals and objects in three spatial dimensions over time. Applications include clinical studies of human rest/activity cycles (actigraphy) (Geraci, 2019), player activity in sports (Tierney et al., 2016), motor vehicle tracking (telematics) (Verbelen et al., 2018), animal and wildlife tracking (Rivest et al., 2016), the tracking of large-scale currents and drifting objects in oceanography (Sykulski et al., 2016; O'Malley et al., 2021), as well as measuring ocean surface waves using buoys—the final of which is the focus of this chapter. The raw time series obtained from such devices are high frequency, but often noisy, and current practices throw away or over-smooth data without utilising their full potential. In this chapter we present a likelihood-based stochastic modelling approach that can meaningfully extract and estimate more spatiotemporal features from ocean wave observations than current methods—but we present our methodology in such a way that it can be applied more broadly to spatiotemporal data, including handling model misspecification, anisotropy, high- and low-frequency noise, aliasing, non-stationarity, and uncertainty quantification.

Wind-generated surface-gravity waves are an important feature of the ocean environment. Understanding the behaviour of such waves is of great scientific and engineering interest, with applications ranging from the design of ships and marine structures to modelling coastal flood risk. As such, large quantities of high frequency time series data are routinely recorded in order to help improve our understanding of the waves, generating a variety of statistical challenges. Characterisation of the wave environment needs to reflect the evolving nature of multiple weather systems, and the presence of measurement uncertainty.

From a modelling perspective, we typically seek to model the vertical displacement of the ocean surface over two-dimensional horizontal space and time. The second-order characteristics of this spatiotemporal process are usually summarised by the

frequency-direction spectrum, which “is the fundamental quantity of wave modelling and the quantity that allows us to calculate the consequences of interactions between waves and other matter”(Barstow et al., 2005). Heuristically, the frequency-direction spectrum quantifies the contribution to the variance of the wave process from multiple sinusoidal components with different frequencies travelling in different directions.¹ This description assumes that the wave process is stationary; however, in reality this is not generally the case. To address this, wave records are usually partitioned into shorter intervals of time series (referred to as sea states), each of which can be treated as stationary (Holthuijsen, 2007).

High resolution measurements of the ocean surface in space and time are rarely available. However, recordings of the vertical displacement of the ocean’s surface at a single location (e.g. using a wave staff or downward-facing radar) or of the motions of floating devices (e.g. buoys) are common (Jensen et al., 2011). In particular, modern buoy measurements provide time series of the buoy’s full three-dimensional displacement. Such measurements are then used to estimate the frequency-direction spectrum, though in general this estimation is not trivial to perform.

Existing techniques for parametric estimation of the frequency-direction spectrum typically use either a method-of-moments or least squares approach. In general, neither approach is optimal statistically. Furthermore, these techniques rely on non-parametric estimates of the frequency-direction spectrum, which exhibit substantial bias. As a result, these approaches perform poorly and can only reliably estimate simple location parameters such as the peak frequency or mean direction of the observed wave process. We propose using likelihood inference directly on the buoy data, avoiding both the poor performance of method-of-moments and least squares; and the issues generated by the non-parametric estimation. Ideally, parametric in-

¹See Appendix for a more formal definition.

ference would be made using maximum likelihood estimation with the full sample likelihood; however, the full likelihood is expensive to compute. Fortunately, adoption of the Whittle likelihood (Whittle, 1953a) provides a computationally efficient alternative to full maximum likelihood inference, which produces consistent parameter estimates and is optimally convergent. Furthermore, debiased Whittle likelihood inference (Sykulski et al., 2019) removes the finite sample bias present in Whittle likelihood inference, without compromising standard error or computational efficiency.

In Chapter 2, we demonstrated in a univariate setting that debiased Whittle likelihood inference yields significant improvements over competitors, when estimating parameters of the spectral density function of ocean waves recorded only over time. The chapter we present here seeks to generalise this methodology to incorporate directional characteristics of the wavefield. However, this extension is nontrivial, since the full spatiotemporal process, which constitutes the wavefield, is not recorded, and hence the spatial debiased Whittle likelihood of Guillaumin et al. (2022) cannot be applied directly. Instead we describe computationally efficient parametric estimation of a frequency-direction spectrum fitted directly to multivariate time series buoy data. Using a multivariate extension of the debiased Whittle likelihood we are able to obtain parameter estimates with lower bias and variance than competitor techniques. Our real-world data analysis reveals robust parameter estimates and captures their evolution over time during a storm; in contrast, such an analysis using existing techniques results in estimates that evolve erratically over time.

The structure of this chapter is as follows. Section 3.2 gives some background on ocean waves, introduces an example data set and then describes a model for the frequency-direction spectrum of wind-sea waves. Section 3.3 describes the debiased Whittle likelihood inference, and demonstrates its performance by simulation. In

Section 3.4, we then apply the debiased Whittle inference to the example data set introduced in Section 3.2.2, discussing a number of important practicalities of the analysis. Finally, Section 3.5 provides a discussion and conclusions.

3.2 Background

3.2.1 Ocean waves and frequency-direction spectra

Much of the interest in ocean waves relates to the surface displacement of the water over space and time, which is treated as a stochastic random field. Usually, the waves are assumed to be stationary and mean-zero within a given time window (often 30 minutes), referred to as a sea state. The covariance structure of the random field for this sea state is then described by the frequency-direction spectrum, $S(\omega, \phi)$, the frequency-domain equivalent of the spatio-temporal autocovariance (see Section 1.2 for more details).

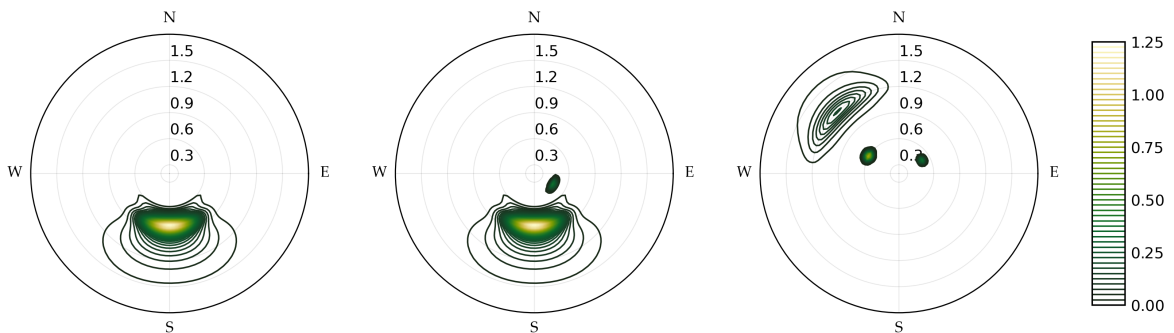


Figure 3.2.1: Example frequency-direction spectra. The left hand panel shows the frequency-direction spectrum corresponding to a single wind-sea, the middle shows a wind-sea and single swell and the right shows a wind-sea and two swells. Direction here is the direction the waves are travelling *from*. The polar plots have direction from north (rad) on the angular axis and angular frequency (rad Hz) on the radial axis.

Examples of frequency-direction spectra are shown in Figure 3.2.1, corresponding to: left, wind-sea only; middle, wind-sea and swell; and right, wind-sea with two swells. Heuristically, if we think about the spectral representation of a process as

decomposing that process into a “sum of sinusoids”, then $S(\omega, \phi)$ can be thought of as describing the contribution to the variance from a wave of a given angular frequency, ω (measured in rad Hz), travelling from a given direction, ϕ (measured in radians). For example, the left hand panel of Figure 3.2.1 describes a process where most of the variance is generated by sinusoids travelling from a southwards direction (π radians) with angular frequencies around 0.8 rad Hz. In contrast, the right hand example describes a process with major contributions to the variance from sinusoids with three different directions and frequencies. Notice that the direction is measured clockwise from North in radians and is the direction a wave is travelling *from* and not *towards*.²

Direct characterisation of the wavefield would require measurements of surface displacement over space and time. Outside of laboratory wave tanks (Forristall, 2015; Schubert et al., 2020), shallow lakes (Young et al., 1996) or coastal regions (Long and Oltman-Shay, 1991; Eastoe et al., 2013), this is very difficult to achieve with current technology. However it is relatively straightforward to measure some characteristics of the wavefield, and to use these measurements to infer properties of the latent spatiotemporal process. For example, we can use measurements of the motion of a floating buoy to approximate the Lagrangian motion of a particle on the water’s surface, recording time series $Z = \{Z(t)\}_{t \in \Delta Z}$, $X = \{X(t)\}_{t \in \Delta Z}$ and $Y = \{Y(t)\}_{t \in \Delta Z}$, of the vertical, northwards and eastwards displacements of the buoy respectively.

We may also describe the covariance structure of the vector-valued stochastic process $\underline{P} = \{\underline{P}(t) = [Z(t), X(t), Y(t)]^T\}_{t \in \mathbb{R}}$ (which is assumed to be stationary and mean-

²Both conventions are used in the literature (Barstow et al., 2005); however, we favour direction from as it means that the relation to the autocovariance is the same as the one used in the statistics literature and is the same as the convention for wind direction, making comparisons easier.

zero) by the spectral density matrix function

$$\mathbf{f}(\omega) = \begin{bmatrix} f_{ZZ}(\omega) & f_{ZX}(\omega) & f_{ZY}(\omega) \\ f_{XZ}(\omega) & f_{XX}(\omega) & f_{XY}(\omega) \\ f_{YZ}(\omega) & f_{YX}(\omega) & f_{YY}(\omega) \end{bmatrix} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{c}(\tau) e^{-i\tau\omega} d\tau,$$

provided certain technical conditions are satisfied (see Brockwell and Davis, 2006, for example), where $\mathbf{c}(\tau) = \mathbb{E} [\underline{\mathbf{P}}(\tau)\underline{\mathbf{P}}(0)^T]$. Under linear wave theory (see Holthuijsen, 2007, for example), there is a transfer function $\underline{\mathbf{G}}(\omega, \phi) = [1, i \cos \phi, i \sin \phi]^T$ for $\omega > 0$, which is conjugate symmetric and zero at $\omega = 0$ (Isobe et al., 1984), such that

$$\mathbf{f}(\omega) = \int_0^{2\pi} \underline{\mathbf{G}}(\omega, \phi) \underline{\mathbf{G}}(\omega, \phi)^H S(\omega, \phi) d\phi, \quad (3.2.1)$$

where A^H denotes the conjugate transpose of a matrix A . This directly relates the frequency-direction spectrum, $S(\omega, \phi)$, to the spectral density matrix function, $\mathbf{f}(\omega)$, which is a feature we shall exploit to perform inference. From (3.2.1) we can immediately see that, for all $\omega \in \mathbb{R}$, $\mathbf{f}(\omega)$ is non-negative definite for any non-negative choice of $S(\omega, \phi)$ (and indeed for any choice of $G(\omega, \phi)$, which may be required for other measurement systems, such as a heave-pitch-roll buoy). Therefore, if we specify a model for $S(\omega, \phi)$ then we can obtain a model for $\mathbf{f}(\omega)$. However, the relation in (3.2.1) is not invertible in general.

Figure 3.2.2 shows an example of the relationship between $S(\omega, \phi)$ and $\mathbf{f}(\omega)$ for four different sea states, differing only by mean direction (indicated by the four colours). The difference in mean direction is obvious in $S(\omega, \phi)$ in the left hand panel, and can still be identified from $\mathbf{f}(\omega)$ in the right hand panel, even though $\mathbf{f}(\omega)$ does not provide a complete description of $S(\omega, \phi)$.

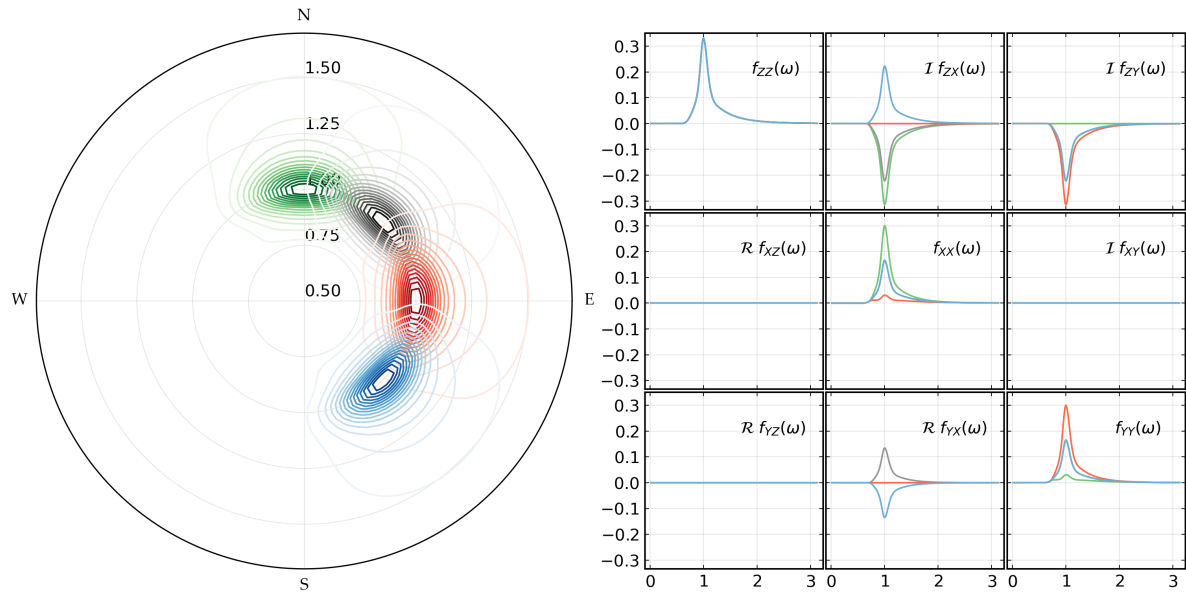


Figure 3.2.2: The effect of varying the mean direction of a wind-sea model on both the frequency-direction spectrum ($S(\omega, \phi)$, left) and the corresponding spectral density matrix function ($f(\omega)$, right). The elements of $f(\omega)$ are shown as a “matrix of functions”, plotting the real part in the lower triangle and imaginary part in the upper triangle. Due to the conjugate symmetry of $f(\omega)$, this representation contains all of the information present.

3.2.2 Example data

For the purpose of illustration, we consider a z, x, y time series recorded using a Datawell Waverider MkIII buoy (Datawell, 2006) in the southern North Sea, at a sampling rate of 1.28Hz. This particular five-day period is chosen to provide an illustration of various physical phenomena often present in the ocean. Within the period, 30-minute sea states (assumed stationary) range from being straightforward to being difficult to model, allowing us to explore the practical applicability of the technique we propose.

Figure 3.2.3 shows a summary of the five-day period in question. The first panel of Figure 3.2.3 shows significant wave height, $H_s = 4\sqrt{\text{var}(Z(t))}$, for each of the sea states, quantifying the roughness of the ocean’s surface. The second panel shows wind speed recorded at a nearby platform. The third panel shows a spectrogram

plotted on the decibel scale, computed using multitapering (Thomson, 1982) with half overlapped 30-minute windows (with $nW = 4/\Delta$ and $K = 7$ Slepian tapers), describing the time-frequency characteristics of the record. The fourth panel shows the mean direction of the waves at different frequencies over time, as defined by Kuik et al. (1988), again computed using half overlapped 30-minute windows. The final panel shows the wind direction over time at a nearby platform.

The record is made up of a variety of component weather systems, which are most easily identified from the mean wave direction (fourth panel). At the start of the record there are two components present, a high-frequency wind-sea and lower-frequency swell. These components fade out throughout day 0, as can be seen from H_s (first panel). A new high-frequency wind-sea develops from the start of day 1, with a clear change in mean wave direction (fourth panel). Throughout day 1, this new component increases in magnitude and transitions to lower frequencies (see third panel), peaking at the start of day 2. Half way through day 2, the wind drops off and then increases (second panel) and changes direction (fifth panel). In response, we see another wind-sea component develop with a different direction to the previous component (fourth panel). Towards the end of day 3, a similar event occurs (though to a lesser degree) and we again see a change in direction. Meanwhile, the swell persists in a low frequency band throughout (third and fourth panels), though it is small in magnitude and narrow-banded in frequency compared to the wind-sea (as can be seen from the spectrogram).

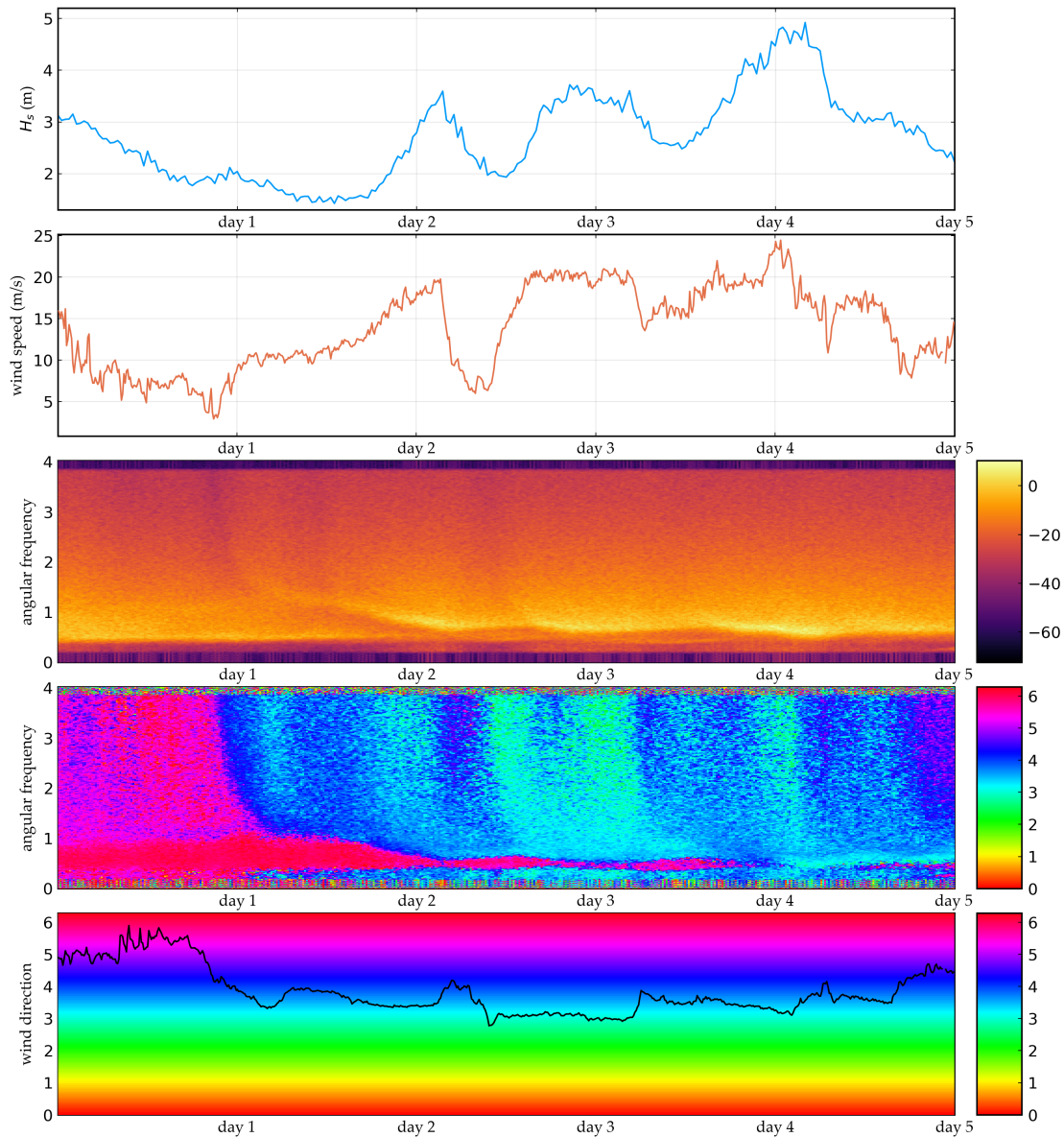


Figure 3.2.3: Summary of the storm data, showing significant wave height, wind speed, a spectrogram on the decibel scale, mean wave direction over time and frequency (direction the waves are travelling from, in radians clockwise from North) and the wind direction over time (direction the wind is travelling from, in radians clockwise from North) recorded at a nearby platform. The x -axis labels are at the start of the day, e.g. day 1 denotes the start of day 1.

3.2.3 Models for wind-sea

When modelling the frequency-direction spectrum, the spectrum is decomposed as

$$S(\omega, \phi) = f(\omega)D(\omega, \phi) \quad (3.2.2)$$

where $f(\omega)$ is the marginal spectral density function of the vertical displacement and $D(\omega, \phi)$ is the so called spreading function. The marginal spectral density function, $f(\omega)$, can be thought of as describing the contribution to the variance from waves of a given frequency regardless of direction, whereas the spreading function, $D(\omega, \phi)$, describes the distribution of wave variance for waves of a given frequency over direction. The spreading function satisfies $\int_0^{2\pi} D(\omega, \phi) d\phi = 1$ and $D(\omega, 0) = D(\omega, 2\pi)$ for all $\omega \in \mathbb{R}$. Figure 3.2.4 shows an example of the decomposition given in (3.2.2) for the model described in the remainder of this section.

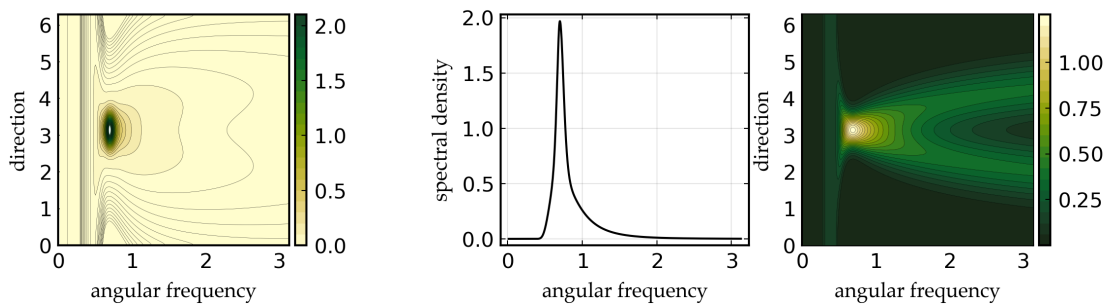


Figure 3.2.4: Example of the decomposition of a frequency-direction spectrum, showing the frequency-direction spectrum (left), marginal spectral density function (middle) and spreading function (right). Plots are given using Cartesian coordinates as this makes the arms of the spreading function easier to visualise.

For the purpose of this chapter, we use the JONSWAP spectral density function first described by Hasselmann et al. (1973), which we denote $f(\omega; \underline{\theta})$, where $\underline{\theta}$ is the vector of parameters. The JONSWAP spectral density function is widely used for modelling the univariate vertical surface displacement resulting from wind-sea waves. Based on physical observations, Hasselmann et al. (1973) developed the JONSWAP

spectral density function with an asymmetric peak and a polynomial decay in the high frequency tail. There is debate about the rate of this tail decay (e.g. Hasselmann et al., 1973; Toba, 1973; Phillips, 1985; Hwang et al., 2017), and so we treat the tail decay index as a free parameter in our analysis. The form of the JONSWAP spectral density function for ocean surface gravity waves can also be motivated from basic physical considerations. Wind waves are generated by the wind blowing over the ocean's surface, through a combination of three physical processes. Wind field turbulence disturbs the water's surface, creating high frequency surface water waves. Then, wind-wave interaction causes these surface waves to grow in amplitude. Thereafter, wave-wave interactions propagate wave energy from higher to lower frequencies. This produces a wave spectral density function with a single spectral peak and long high-frequency tail, with peak frequency evolving from higher to lower frequency during an ocean storm of limited duration. In particular,

$$f(\omega; \underline{\theta}) = \alpha \omega^{-r} \exp \left\{ -\frac{r}{4} \left(\frac{|\omega|}{\omega_p} \right)^{-4} \right\} \gamma^{\delta(\omega; \underline{\theta})}$$

where the specific details can be found in Section 1.3.

Various models have been proposed for the directional spreading of wind-sea waves. A large number of experimental studies (e.g. Young et al., 1995; Ewans, 1998; Wang and Hwang, 2001) indicate that the spreading function is bimodal with direction, for frequencies exceeding the peak frequency. This finding is supported by theoretical arguments involving directional energy transfer through wave-wave interactions (Banner and Young, 1994; Young et al., 1995; Toffoli et al., 2010). For this reason, we adopt the bimodal wrapped Gaussian model of Ewans (1998) in this work. At each frequency, the spreading function over direction is assumed to be a bimodal wrapped Gaussian with means $\phi_{m1}(\omega; \underline{\theta})$ and $\phi_{m2}(\omega; \underline{\theta})$, but the same standard de-

viation³ $\sigma(\omega; \underline{\theta})$. In other words,

$$D(\omega, \phi; \underline{\theta}) = \frac{1}{2\sigma(\omega; \underline{\theta})\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \sum_{i=1}^2 \exp \left\{ -\frac{1}{2} \left(\frac{\phi - \phi_{mi}(\omega; \underline{\theta}) - 2\pi k}{\sigma(\omega; \underline{\theta})} \right)^2 \right\}.$$

The mean directions are

$$\phi_{m1}(\omega; \underline{\theta}) = \phi_m + \beta \exp\{-\nu \cdot \min(\omega_p/|\omega|, 1)\}/2$$

$$\phi_{m2}(\omega; \underline{\theta}) = \phi_m - \beta \exp\{-\nu \cdot \min(\omega_p/|\omega|, 1)\}/2$$

and the spreading function is given by

$$\sigma(\omega; \underline{\theta}) = \sigma_l - \frac{\sigma_r}{3} \left\{ 4 \left(\frac{\omega_p}{|\omega|} \right)^2 - \left(\frac{\omega_p}{|\omega|} \right)^8 \right\}.$$

Table 3.2.1 gives a description of the parameters of the model, and their parameter spaces. A more complete description of the model is given in Section 1.3. Note that the inference approach described in this chapter is applicable for other models, but the model described here has been chosen for definiteness.

Parameter	Description	Parameter space
α	scaling parameter	$(0, \infty)$
ω_p	peak frequency	$(0, \infty)$
γ	peak enhancement factor	$[1, \infty)$
r	spectral tail decay index	$(1, \infty)$
ϕ_m	mean direction	$[0, 2\pi)$
β	limiting peak separation	$[0, 2\pi)$
ν	peak separation shape	$[0, \infty)$
σ_l	limiting angular width	$[0, \infty)$
σ_r	angular width shape	$[0, \infty)$

Table 3.2.1: Parameter descriptions for the JONSWAP spectrum and bimodal wrapped Gaussian spreading function.

³The standard deviation is referred to as angular width by Ewans (1998).

3.3 Modelling process

We aim to jointly estimate all the parameters of Table 3.2.1, both marginal parameters and spreading parameters, given a sample of three-dimensional displacement. In this section, we describe the proposed inference technique, and demonstrate in simulations that it yields significant improvements in performance over the existing least squares and moments-matching approaches, described in Appendix 1.3.3. For brevity, we shall refer to such techniques as competitor techniques for the remainder of this chapter. In contrast to competitor techniques, we convert the model for the frequency-direction spectrum to a model for the spectral density matrix function of the data we actually observe, and then fit the model directly to the data. This is statistically more appealing as we fully exploit the degrees of freedom in the observational data, rather than performing unnecessary smoothing transformations before model fitting, and is the key reason our method performs better.

3.3.1 Model fitting

Due to the quantity of available data, computationally efficient inference techniques are desirable. For a Gaussian process, full maximum likelihood would require the inversion of a $3n \times 3n$ matrix. This is expensive when $n = 2304$ as in our case, especially given that we have a different time series every half an hour. Furthermore, we may wish to only model a certain frequency range (see e.g. Section 3.4.1 for our application study), which is hard to achieve with full maximum likelihood. Frequency domain pseudo-likelihoods such as the debiased Whittle likelihood (Sykulski et al., 2019) provide a nice alternative to full maximum likelihood inference. Debiased Whittle likelihood inference has been shown to perform well in a variety of applications, including for planetary topography (Guillaumin et al., 2022), ocean drifters (Sykulski et al., 2016) and univariate recordings of ocean waves (Grainger et al.,

2021). For these reasons, we use a multivariate extension of the debiased Whittle likelihood due to Guillaumin et al. (2022).

Assume we have a sample of length n , the periodogram, $\mathbf{I}_n(\omega)$, is defined as

$$\mathbf{I}_n(\omega) = \mathbf{J}_n(\omega)\mathbf{J}_n(\omega)^H \quad \text{where} \quad \mathbf{J}_n(\omega) = \sqrt{\frac{\Delta}{2\pi n}} \sum_{t=0}^{n-1} \mathbf{P}(t\Delta)e^{-it\Delta\omega},$$

usually evaluated at $\Omega_n = \{2\pi j/n \mid j \in \{-\lceil n/2 \rceil + 1, \dots, \lfloor n/2 \rfloor\}\}$, the Fourier frequencies, using the Fast Fourier Transform (Cooley and Tukey, 1965). The multivariate Whittle likelihood (Whittle, 1953a), in its discrete form, is given by

$$\ell_W(\underline{\boldsymbol{\theta}}) = - \sum_{\omega \in \Omega} \log |\mathbf{f}(\omega; \underline{\boldsymbol{\theta}})| + \text{tr}\{\mathbf{I}_n(\omega)\mathbf{f}(\omega; \underline{\boldsymbol{\theta}})^{-1}\}, \quad (3.3.1)$$

where $\Omega \subseteq \Omega_n$ and $\mathbf{f}(\omega; \underline{\boldsymbol{\theta}})$ denotes a parametric spectral density matrix function with parameter vector $\underline{\boldsymbol{\theta}}$. The multivariate Whittle likelihood suffers from finite sample bias, especially as the dimension grows, so a debiased version may be used to improve estimates, accounting for sampling effects such as aliasing and blurring. Aliasing results from regular discrete sampling of a continuous time process. Blurring results from sampling a time series for finite duration. Specifically, the finite sample results in a convolution in the frequency domain which causes spectral density estimates to “blur” or “leak” across frequencies. In the case of 1.28Hz wave data recorded for 30 minutes, the blurring is minor; however, for shorter records this is not the case, and using the debiased Whittle likelihood is beneficial.

The multivariate debiased Whittle likelihood (Guillaumin et al., 2022) is

$$\ell_D(\underline{\boldsymbol{\theta}}) = - \sum_{\omega \in \Omega} \log |\mathbb{E}[\mathbf{I}_n(\omega); \underline{\boldsymbol{\theta}}]| + \text{tr}\{\mathbf{I}_n(\omega)\mathbb{E}[\mathbf{I}_n(\omega); \underline{\boldsymbol{\theta}}]^{-1}\}, \quad (3.3.2)$$

where the expected periodogram can be efficiently computed using the relation

$$\mathbb{E} [I(\omega); \underline{\theta}] = \frac{\Delta}{2\pi} \sum_{\tau=-n+1}^{n-1} (1 - |\tau|/n) c(\tau; \underline{\theta}) e^{-i\omega\Delta\tau}.$$

In our case, the autocovariance, $c(\tau; \underline{\theta})$, is not known analytically, and instead must be approximated numerically from the spectral density matrix function. Since models are specified for the continuous time process, the most efficient way to approximate the autocovariance is to first approximate the spectral density of the discrete time process, then approximate the autocovariance (as described in Chapter 2). The first step requires aliasing the spectral density of the continuous time process by wrapping in contributions from infinitely many frequencies above the Nyquist frequency, i.e. computing

$$f_{\Delta}(\omega) = \sum_{k=-\infty}^{\infty} f(\omega + 2\pi k/\Delta). \quad (3.3.3)$$

To do this numerically, we have to use a truncated version of the sum in (3.3.3). In practice, the instrument may not respond to waves with frequencies above a certain threshold, or the data may have been filtered in preprocessing (Datawell, 2006). Therefore, the recorded process may not be aliased to the same extent as the theoretical sampled process. In our case, we treat the buoys as if no aliasing has occurred (i.e. retaining only the $k = 0$ term in (3.3.3)) due to the observed drop-off in the spectral density at the highest frequencies, as can be seen from panel 3 of Figure 3.2.3. However, we note that this technique is able to account for aliasing, should it be felt that aliasing is present.

In both (3.3.1) and (3.3.2), summation is over a set Ω . Usually $\Omega = \Omega_n$; however, we may wish to remove some frequencies to avoid model misspecification (see Section 3.4.1) or because at some frequencies in the periodogram the ordinates can

be highly correlated for finite samples, which harms Whittle estimation. We then maximise this likelihood function using numerical methods, detailed further in Appendix B.3.

3.3.2 Simulation study

We now present a simulation study comparing the debiased Whittle likelihood inference proposed in Section 3.3.1 with the least squares and moments-matching techniques described in Appendix 1.3.3. We have chosen three different scenarios that represent possible conditions seen in practice, including cases where certain parameters are on the boundary of the parameter space (as this is likely to cause problems for estimation techniques). The parameters for each scenario are given in Table 3.3.1, and the corresponding frequency-direction spectra are given in Figure 3.3.1.

	α	ω_p	γ	r	ϕ_m	β	ν	σ_l	σ_r
Scenario 1	0.7	0.8	3.3	5	$\pi/2$	4	2.7	0.55	0.26
Scenario 2	0.7	1.1	3.3	5	$\pi/2$	4	2.7	0.55	0.00
Scenario 3	0.7	1.0	1.0	5	$\pi/2$	4	2.7	0.55	0.26

Table 3.3.1: Table showing the parameters for each scenario in the simulation study.

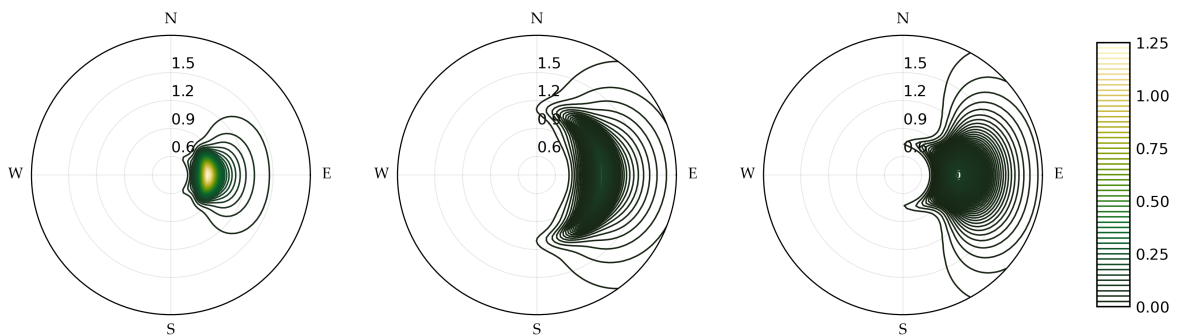


Figure 3.3.1: Frequency-direction spectra for Scenario 1 (left), Scenario 2 (middle) and Scenario 3 (right), as defined in in Table 3.3.1.

Scenario 1 is a classic example of a fetch-limited wind-sea, with directional shape parameters fixed to the standard values from Ewans (1998), and $\gamma = 3.3$ from Hasselmann et al. (1973). Scenario 2 is almost identical, except that $\sigma_r = 0$, meaning that

$\sigma(\omega; \underline{\theta})$ is constant over frequency. Heuristically, this corresponds to a frequency-direction spectrum where the width of each arm in the spreading function is constant over frequency (see Figure 3.2.4 for the notion of an arm). This scenario is included because we often see this parameter tending towards the boundary of the parameter space in practice (as in Section 3.4.2) and it is useful to explore the effect of this on other parameter estimates (though we cannot say anything about the impact of model misspecification from this). Finally, Scenario 3 is a Pierson-Moskowitz spectrum for a fully developed sea (Pierson and Moskowitz, 1964), also using the standard spreading parameters from Ewans (1998). This is a special case of the JON-SWAP spectrum with $\gamma = 1$, and so is of particular interest as it lies on the boundary of the parameter space.

We simulate 1000 time series from each of the scenarios and estimate the model parameters using each of the techniques from Appendix 1.3.3 alongside the debiased Whittle likelihood inference from Section 3.3.1.⁴ In particular, we use the least squares technique described in Appendix 1.3.3 with both MLM and MEM based estimation of frequency-direction spectrum and the moments-matching approach described in Appendix 1.3.3. Whilst there are three different methods from the existing literature in our comparison, they all use the same technique to estimate the parameters of the marginal spectral density function. As such, Figure 3.3.2 considers the marginal parameters estimated using least squares (the marginal technique for the competitor techniques), univariate debiased Whittle on only the vertical displacement, and multivariate debiased Whittle on all three time series. Figure 3.3.3 considers the spreading parameters where the least squares technique is now split into the three directional variants: least squares with MLM, least squares with MEM, and the moments-matching approach; and the univariate debiased Whittle is not in-

⁴In scenario 2, for nine of the replications, the least squares with MEM optimisation did not converge. For this reason, in the results for scenario 2 we include only the 991 replications for which the optimisation of all objective functions converged.

cluded (as it cannot be used to estimate the spreading parameters).

From Figure 3.3.2, we see a clear improvement in the debiased Whittle when comparing to least squares, especially in terms of variance, as has already been reported Chapter 2. Additionally to the results already seen in Chapter 2, there is also a benefit to estimating the parameters of the marginal spectral density function from all three series (as opposed to from the vertical displacement alone). Traditionally, estimating the marginal parameters has been treated as a separate problem from estimating the spreading parameters, with only the vertical displacements used to estimate the marginal parameters. However, this clearly throws away useful information about the marginal parameters which is present in the x and y time series. Furthermore, in Scenario 3, debiased Whittle likelihood recovers all of the parameters well, despite the true value of γ being on the boundary of the parameter space (though clearly the estimates of γ are not normally distributed).

Similarly, Figure 3.3.3 demonstrates a stark difference between the competitor techniques and debiased Whittle likelihood inference. Other than the mean direction ϕ_m , we see substantial bias in all the other parameter estimates from each of the three existing techniques which is not present in the debiased Whittle likelihood estimates. We also see that the debiased Whittle estimates exhibit significantly less variance across all parameters and scenarios. From Scenario 2, we see that debiased Whittle likelihood inference still performs well when σ_r is on the boundary of the parameter space (though again estimates are not normally distributed). Additionally, when estimating β , we see that least squares with MLM in scenario 1 and moments-matching in scenarios 2 and 3 the majority of the estimates are on the upper boundary of the parameter space, an issue which debiased Whittle likelihood inference does not have.

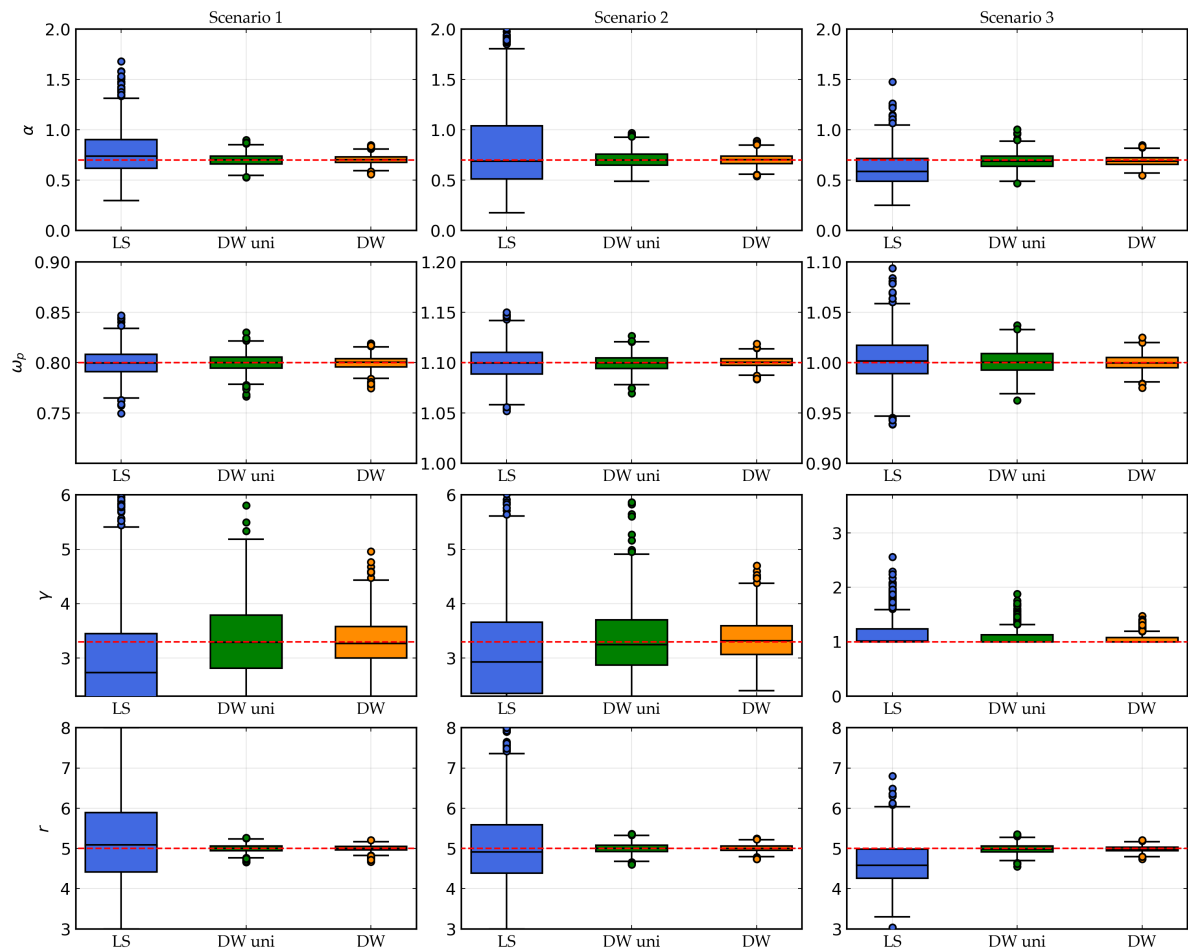


Figure 3.3.2: Boxplots of the parameter estimates from the simulation study for parameters of the marginal spectral density function, with the true values indicated by red dashed lines. Marginal parameters estimated using least squares (LS), univariate debiased Whittle (DW uni) and full multivariate debiased Whittle (DW) are shown.

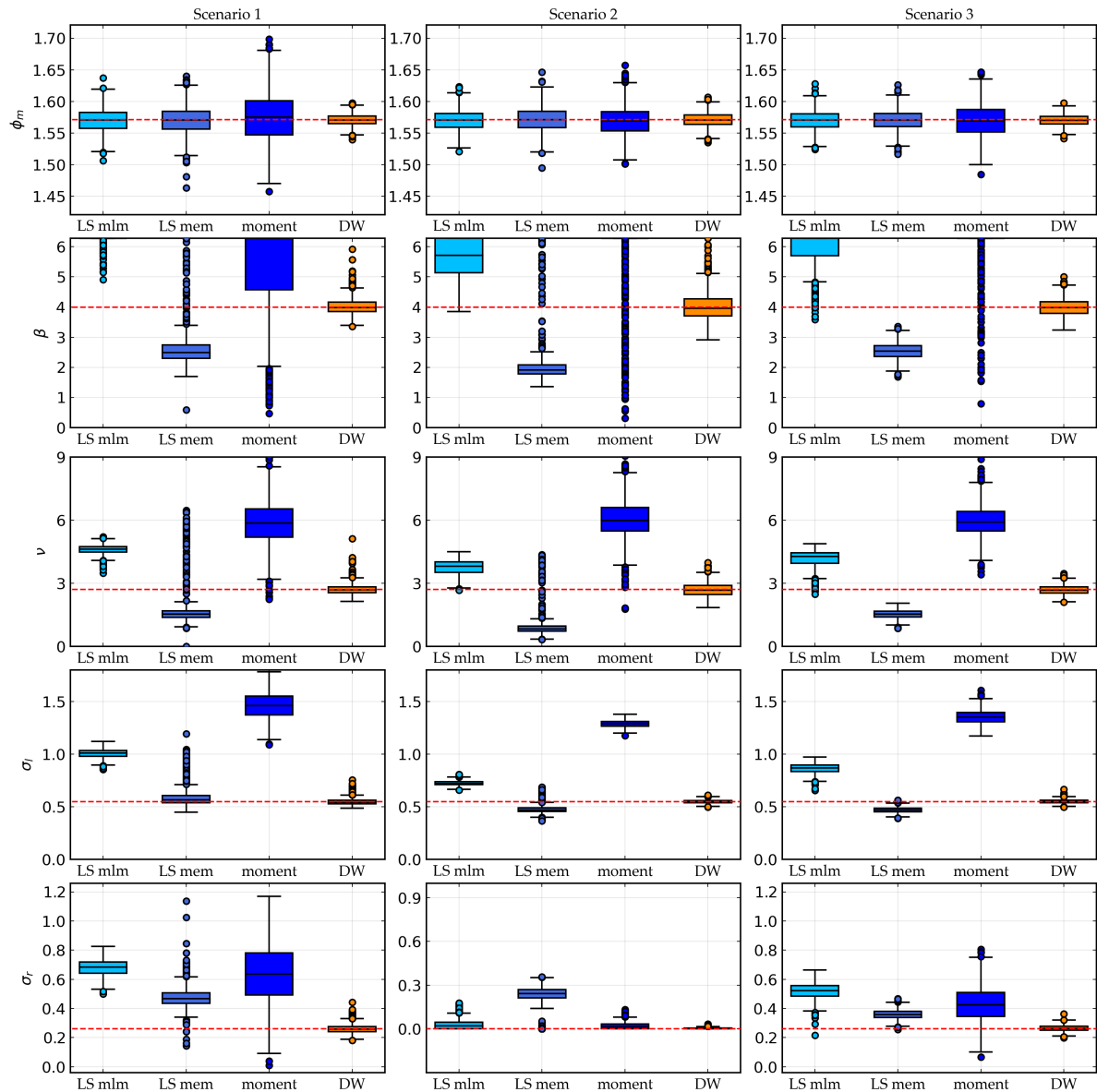


Figure 3.3.3: Boxplots of parameter estimates from the simulation study for parameters of the spreading function, with the true values indicated by red dashed lines. Spreading parameters estimated using least squares with MLM (LS mlm), least squares with MEM (LS mem), the moments-matching approach (moment) and multivariate debiased Whittle (DW) are shown.

3.4 Modelling the example data set

We now apply debiased Whittle inference for $S(\omega, \phi; \underline{\theta})$, Table 3.2.1, to the data set introduced in Section 3.2.2. Both wind-sea and swell are present in our example record. However, we have chosen to model only the wind-sea as the purpose of this chapter is to introduce a new inference technique, and this is easiest to illustrate and scrutinise with a simple wind-sea only model. The debiased Whittle procedure could naturally be used on a swell-only model (or indeed a joint wind-sea and swell model) should the swell characteristics be of further interest, but this is reserved for future work.

Due to issues with the measurement device and other contaminating processes, certain frequency regions do not reflect the process we are interested in modelling. Therefore, careful selection of the frequencies included in the objective function must be performed prior to inference. Selecting these frequencies is difficult, but there are principled ways to choose them. In particular, the buoy data does not accurately represent the data which we are interested in modelling at the lowest and highest frequencies (van Essen et al., 2018). As such, we select a low- and high-frequency threshold and use only the frequency interval between the thresholds in our analysis, as we shall now detail in Section 3.4.1.

3.4.1 Specification of low- and high-frequency thresholds for inference

Model misspecification presents a significant challenge for the fitting techniques discussed in this chapter. Such misspecification can be generated in a variety of ways. Firstly, other component weather systems that we do not want to (or cannot) model may be present. Secondly, there may be noise due to the buoy not following the true motion of a particle on the water's surface. Finally, the approximations made by lin-

ear wave theory that justify the transfer function in (3.2.1) may not be valid. All of the aforementioned problems affect some frequencies more than others. Therefore, we shall remove frequencies that are heavily contaminated before fitting models to the data. Because we are using a frequency domain pseudo-likelihood, this is easy to do, and essentially just involves omitting the appropriate Fourier frequencies from the likelihood (as discussed in Section 3.3.1).

However, choosing which frequencies to remove is not trivial. One useful guide comes from (3.2.1), which implies that $f_{ZZ}(\omega) = f_{XX}(\omega) + f_{YY}(\omega)$ under linear wave theory. Motivated by this, we define the error function $R(\omega) = \log(f_{XX}(\omega) + f_{YY}(\omega)) - \log(f_{ZZ}(\omega))$.⁵ An estimate, $\hat{R}(\omega)$, of the error function can be obtained by first estimating the spectral density functions, then plugging them into the above formula for $R(\omega)$. Clearly we would expect $\hat{R}(\omega) \approx 0$ for all $\omega \in [0, \pi/\Delta]$, so deviations from zero may indicate that there is a problem with a certain frequency range. Figure 3.4.1 shows a plot of $\hat{R}(\omega)$ for each half hour period from our example data set introduced in Section 3.2.2 using multitapering (again with $nW = 4/\Delta$ and $K = 7$ Slepian tapers).

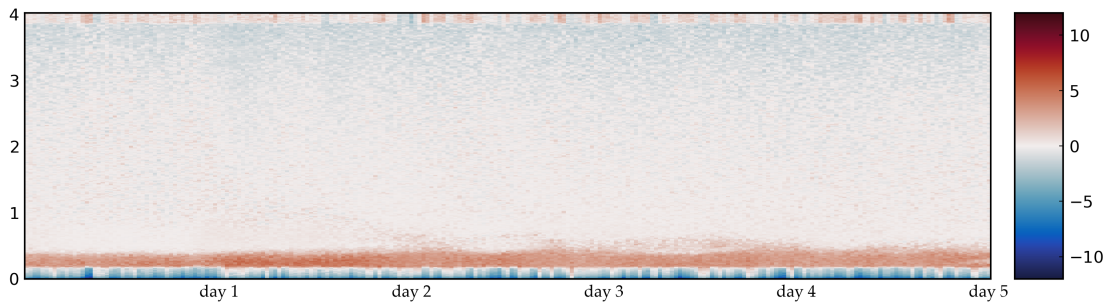


Figure 3.4.1: Heatmap of $\hat{R}(\omega)$ for each half hour period in the example data set, computed using multitapering.

From Figure 3.4.1, we see a blue band in the very lowest frequency range with a

⁵Note that this relation is for the deep water case. The finite water depth version is slightly different and given in Appendix B.2. The finite water depth version is used in Figure 3.4.1, though for simplicity we state the deep water version here. The quantity $R(\omega)$ is related to the check ratio often used in quality control for buoy data (Integrated Ocean Observing System (U.S.), 2019).

red band sitting in the frequency range just above this, where the absolute value of the error function is significantly larger than zero. Therefore, in low frequencies the transfer function mentioned above is not valid, and as a result these frequencies are removed when fitting the model. Additionally, $\hat{R}(\omega)$ is slightly negative in the highest frequencies. In other words, the spectral density of the X and Y processes decays more rapidly than that of the Z process in the high frequency tail. This is possibly because the accelerometers for measuring the horizontal displacement of the buoy are mounted in a different way to the accelerometer measuring the vertical displacement, though more investigation is needed to ascertain the source of this discrepancy. Regardless, it is the general consensus that these instruments are more reliable for the middle of the frequency range than they are at the highest and lowest frequencies, and standard quality control of buoy data include checks for excessive level of low and high frequency spectral density (Christou and Ewans, 2014, for example).

Additionally, an old wind-sea and a swell are present in the early sea states with the swell persisting, albeit with little energy, for most of the record. Since models for such conditions are beyond the scope of this chapter, we only begin modelling when the new wind-sea has become dominant, and remove frequencies in which the swell is large or \hat{R} is sufficiently far from zero. In particular, for each half hour period, we pick the cutoff frequency to be the largest frequency below the peak frequency which either has an average \hat{R} larger than some threshold⁶ or has a mean direction sufficiently far from the average of the wind-sea mean direction. For more details, see the code provided on GitHub (Grainger, 2022a). Additionally, frequencies beyond 3.8 rad Hz are also removed due to the observed drop-off in spectral density. Figure 3.4.2 shows this choice of modelling period and low frequency threshold with

⁶By average here we mean the average \hat{R} in some window centred on the frequency in question, and we use a threshold of 2.

the modelling period delimited by dashed vertical lines and the threshold shown by dotted lines, indicating that only frequencies between these lines are included.⁷

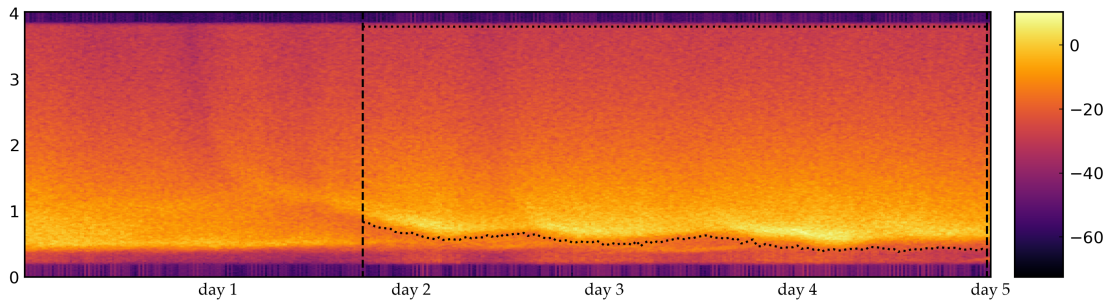


Figure 3.4.2: Spectrogram of the example data set on the decibel scale, with the period used in the fitting delimited by solid vertical lines, and the choice of frequency range over the period of interest shown by the dotted lines.

3.4.2 Parameter estimates

Here we estimate model parameters for $S(\omega, \phi; \theta)$ using debiased Whittle inference, for the frequency intervals specified in Section 3.4.1. Most of the parameters are initialised from standard values, with the only exceptions being ω_p and ϕ_m which are initialised by picking the frequency corresponding to the maximum of a non-parametric estimate of the marginal spectral density function, and the mean direction corresponding to this frequency respectively.

Figure 3.4.3 shows the parameter estimates, with 95% approximate confidence intervals, calculated using the expected Hessian matrix and assuming parameter estimators are Gaussian distributed. The location parameters ω_p (the peak frequency) and ϕ_m (the mean direction) behave as expected, following the spectral mode and reacting to changes in wind-direction respectively. They also evolve smoothly in time, despite fits being performed independently on non-overlapping sea states. The shape parameters for the marginal spectral density function (γ and r), clearly have time

⁷Some of the highest frequencies are also removed from the objective function. This is because the response of the buoy falls off rapidly at the highest frequency, which is likely a result of the instrument's inability to respond to the waves and the use of digital filters during post processing, details of which can be found in Datawell (2006).

varying behaviour. The peak enhancement factor, γ , increases as each component wind-sea evolves, then decreases as the component wind-sea dies out. Similarly, from the estimates for r , the tail decay becomes less steep between components. It is likely that this is due to model misspecification, as we really have two wind-seas present, but are only modelling one of them. Furthermore, the shape parameters for the directional spreading (β, ν, σ_l and σ_r) also show anomalous behaviour during these overlaps. In particular, we see large values of β (hitting the upper bound of the parameter space). Large values of β correspond to a wide spreading over direction, which likely occurs because there is another component present with different directional properties. However, outside these overlap periods we see stability in the parameters estimates, which is encouraging. Additionally, some of the estimates of σ_r drop off to zero, because the low frequency threshold can make σ_r unidentifiable as much of the information about σ_r resides in frequencies below the peak frequency. As a result, we have an identifiability-bias trade off as lowering the threshold frequency introduces more of the noise processes, which tends to result in biasing of β , but raising the threshold makes σ_r unidentifiable. This is a difficult problem, and is an important area of further research which we discuss more thoroughly in Section 3.5.

In summary, the parameter estimates converge to sensible values in the majority of sea states where a single wind-sea is present. Furthermore, looking at sea states where parameter estimates go to boundaries or unrealistic values helps to extract time periods of interest where the model fails and separate investigation is warranted.

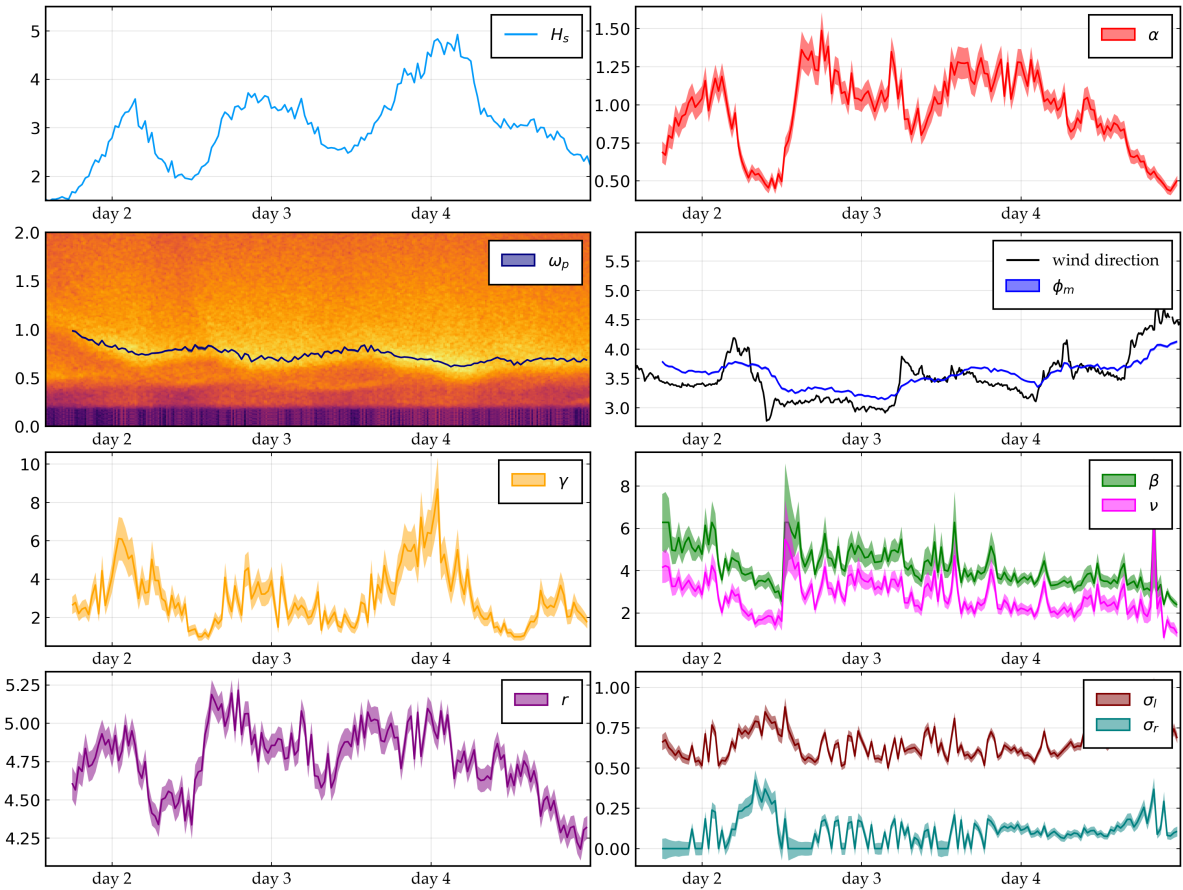


Figure 3.4.3: Parameter estimates using debiased Whittle likelihood inference over the period in question with approximate 95% confidence intervals. The two panels in the second row also include the spectrogram and wind direction for context. In order left to right then down, the panels show: \hat{H}_s ; $\hat{\alpha}$; $\hat{\omega}_p$ over the spectrogram on the decibel scale; $\hat{\phi}_m$ and wind direction; $\hat{\gamma}$; $\hat{\beta}$ and $\hat{\nu}$; \hat{r} ; and $\hat{\sigma}_l$ and $\hat{\sigma}_r$, with approximate 95% confidence intervals.

3.5 Discussion and conclusions

This chapter describes estimation of the parameters of frequency-direction spectra for ocean surface gravity waves from three-dimensional buoy displacement time series, using debiased Whittle likelihood inference. In simulation studies, debiased Whittle inference is shown to outperform inference using competitor techniques. Debiased Whittle inference for a sequence of sea states provides a means to characterise the joint evolution of spectral parameters in time, and allows uncertainties in parameter estimates to be quantified in a principled manner. The observed smooth nature of parameter evolution estimated from North Sea data, and the dependencies evident between parameters, are consistent with physical intuition.

Typically, the wave environment at a location is the product of different physical drivers, including swell and local wind forcing. In the current work, we focus on sea states corresponding to wind-sea conditions only, for clarity of description. More generally, debiased Whittle inference for mixed seas consisting of wind-sea and one or more swells is possible; in simulation studies of data for mixed seas (not shown), debiased Whittle inference again performs well. In simulation studies on samples of 30-minute records corresponding to wind-sea conditions, the debiasing procedure makes a small but marginal improvement over standard multivariate Whittle estimation. However, when fitting the joint wind-sea and swell model to mixed sea states, or when using shorter records, estimates using standard Whittle inference exhibit substantially greater bias than those from debiased Whittle inference.

In-situ measurement of the ocean environment is invariably problematic. In the present study, buoy displacement time series are contaminated by additional low-frequency processes, leading to spurious low-frequency spectral features not represented in the assumed parametric spectral form to be estimated. At very high frequencies, buoy displacement time series are further subject to on-board low-pass

filtering effects not represented in the assumed spectral form. We adjust the inference procedure for these sources of model misspecification by only considering a central band of frequencies in the likelihood, set using low-frequency and high-frequency thresholds. In general, the low-frequency threshold in particular should be chosen carefully, to achieve a good balance between model misspecification (when the threshold is too low) and identifiability (when the threshold is set so high that aspects of the spectral form cannot be resolved). We have explored extending the spectral form to accommodate an additional low-frequency noise feature, but found that achieving this reliably required a large number of extra parameters, and resulted in greater loss of efficiency in estimating the wind-sea (and swell) components of interest compared to frequency thresholding.

Spectral estimates in the current work are based on data for the ocean's surface displacement only. In general, it would be advantageous to incorporate the effects of covariates such as the evolving wind field on the spectral form, particularly for characterisation of mixed seas. For example, the direction associated with a wind-sea component at a location is dependent on local wind speed and direction, whereas the characteristics of a swell component do not vary substantially with the local wind field. These covariate dependencies are often exploited by physical oceanographers to partition the frequency-direction domain into sub-domains corresponding to wind-sea and swell components (Hanson and Phillips, 2001, for example).

The spectral characteristics of ocean waves evolve smoothly in time. In this chapter, as is common practice, we accommodate temporal non-stationarity by partitioning time series into consecutive 30-minute sea states which are considered stationary for purposes of spectral inference. Improved bias-variance properties of parameter estimates from debiased Whittle inference suggest that spectral estimation using sea states of shorter duration is feasible for more-rapidly evolving ocean environments;

initial simulation studies (not shown) support this finding. More generally, simultaneous spectral estimation for a sequence of consecutive sea states exploiting smooth time-varying basis representations for spectral parameters (e.g. using splines), or adaptive estimation of evolving spectral forms (e.g. using dynamic linear models) are obvious research avenues.

The methodology for spectral inference described in this chapter is generally applicable, provided that an appropriate model for the spectral density matrix function can be obtained by applying a suitable transfer function to the model for the frequency-direction spectrum. Thus, in addition to three-dimensional buoy displacement time series, debiased Whittle inference is applicable to heave-pitch-roll buoy data, for example. A collection of useful transfer functions for commonly used oceanographic devices is given by Benoit et al. (1997). The methodology can be modified for similar applications involving observations of a process viewed as a linear time-invariant filter of some latent process of interest. Practical issues encountered in the current work, relating to time series aliasing, unusual sources of measurement noise and complex likelihood functions are common across many applications (e.g. involving accelerometers and GPS tracking). Hence, we hope that the methodology presented and the ideas it incorporates will prove useful to the practising oceanographer, ocean engineer and applied statistician.

The environmental data used has been provided by TotalEnergies E&P Danmark A/S, and the data is the sole property of TotalEnergies E&P Danmark A/S, and it may not be used or reproduced without the written consent of TotalEnergies E&P Danmark A/S. TotalEnergies E&P Danmark A/S has not assisted with or had any influence on the use of the data or the subject or content of this chapter.

Chapter 4

Conclusions of Part I

In Part I of this thesis, we have considered parameter estimation for models of the frequency domain behaviour of ocean waves. In particular, Chapter 2 considered parameter estimation in the univariate case, whilst Chapter 3 described parameter estimation for the frequency direction spectrum.

In Chapter 2, we saw that parameter estimates can be substantially improved by utilising pseudo-likelihoods, without losing the computational efficiency of the methodology currently in use. We demonstrated such improvement in detailed simulation studies, and discussed computational and practical aspects of applying the method. In particular, many of the popular models for the spectral density function of ocean wave time series do not have analytically known Fourier transforms (autocovariance). As such, techniques such as the debiased Whittle likelihood (or even maximum likelihood) require additional approximation steps. However, this can be achieved with the use of an additional FFT, meaning that we do not increase the computational cost, which remains $O(n \log n)$. Similarly, gradients and Hessians can also be computed very efficiently with the same technique.

Chapter 3 then discussed similar estimation for the frequency-direction spectrum. This is not a straightforward generalisation of the problem discussed in Chapter 2 because we cannot directly record the ocean surface over space. Instead, we record the multivariate displacement of a buoy on the surface, and then try to estimate the parameters of the frequency-direction spectrum. At a high-level, existing methodology transforms the data to be closer to the model. In contrast, we transform the model to describe the data, avoiding many smoothing steps and extra assumptions. As such, we see dramatic improvements in both bias and variance over existing techniques for certain parameters. Consequently, we can estimate the parameters for significantly more complicated models than was feasible with existing techniques. Resulting in improved descriptions of the frequency-domain behaviour of ocean waves.

4.1 Further work

In Part I, we only discussed parameter estimation for models of a single wind-sea component. Whilst preliminary analysis suggests that the methodology developed in Chapters 2 and 3 could be used in more complicated scenarios, further work should be undertaken to investigate the performance of parameter estimates in this case. In particular, the issues with low-frequency noise discussed in Chapter 3 become more problematic in this setting, as the swell often substantially overlaps the noise in frequency. Models for swell processes are implemented in `OceanWaveSpectralFitting.jl` (Grainger, 2022b), and the infrastructure for handling additive models is provided by `WhittleLikelihoodInference.jl` (Grainger, 2022c), meaning that fitting additive models for joint wind-sea and swell conditions is fairly straightforward.¹

¹In the sense that they are already implemented (and gradients, Hessians and memory preallocation are handled in a sensible way).

Ideally, we would develop a model for the low-frequency noise, and estimate the resulting nuisance parameters alongside the parameters of interest. Unfortunately, this is not as easy as it first sounds. In particular, because we have three dimensions the number of parameters required for even simple models can be quite large. Combined with the wind-sea parameters already present in the model, and additional swell parameters, this can make optimisation much harder.

Two more concrete examples of further work are modelling the evolution of parameters over time, and applying the methodology we have developed to other kinds of data, such as data recorded by spatial arrays of wave staffs, in shallow water. We discuss the former problem in Section 4.1.1 and the latter in Section 4.1.2.

4.1.1 Modelling the behaviour of spectral parameters over time

We aim to model waves in the ocean. Typically, multiple component weather systems will be present in the ocean. These weather systems gradually appear at the recording location, persist for some time, and then fade away. To simplify the modelling, we shall have some start and end point between which the component is treated as present.

Model specification

Imagine that the spectral density function of a particular component weather system can be represented by some parametric family, such that, at any given time t , the true process had spectral density function $f(\cdot; \underline{\theta}(t))$. We aim to model the behaviour of $\underline{\theta}(\cdot)$ over time. Guillaumin et al. (2017) propose modelling such evolution for ocean drifters as modulated time series, and show how Whittle likelihood inference can be used in this context. In particular, such an approach models the time series as a stationary process multiplied (or modulated) by a deterministic function. However,

in our setting, it is not obvious that parameter evolution can be expressed in this manner, as the modulation function must be known a priori, at least its functional form, which is not usually the case in ocean wave records. Therefore, we suggest taking a similar approach to that of [Oscroft et al. \(2020\)](#), modelling the time varying parameters. Note that this is much more complicated in our setting, as component weather systems come and go. This results in the number of parameters essentially changing over time.

A rough outline of the proposed methodology is as follows. Say that we consider a stretch of time during which a number of different weather systems are present. Let $T_1^{(j)}$ be the start time and $T_2^{(j)}$ be the end time of component j . Then the time-varying parameter function is

$$\underline{\theta}^{(j)} : [T_1^{(j)}, T_2^{(j)}] \rightarrow \Theta^{(j)} \subseteq \mathbb{R}^{p^{(j)}},$$

where $p^{(j)}$ is the number of components of the model used for the component, and $\Theta^{(j)}$ is the parameter space of that model.

Let $\theta_i^{(j)} : [T_1^{(j)}, T_2^{(j)}] \rightarrow \Theta_i^{(j)}$, be the i^{th} parameter function (for $1 \leq i \leq p^{(j)}$), i.e. for $t \in [T_1^{(j)}, T_2^{(j)}]$ we have $\underline{\theta}^{(j)}(t) = [\theta_1^{(j)}(t), \dots, \theta_{p^{(j)}}^{(j)}(t)]^T$. To model $\theta_i^{(j)}$ we use the following form:

$$\theta_i^{(j)}(t) = g_i \left(\sum_{k=1}^{m_i^{(j)}} b_{i,k}^{(j)} \phi_{i,k}^{(j)}(t; \zeta_{i,k}^{(j)}) \right)$$

where $\phi_{i,k}^{(j)}$ are basis functions, $\zeta_{i,k}^{(j)} \in \Xi^{(j)}$ are parameters of the basis functions and g is a bijector (a differentiable bijection with a differentiable inverse). If we are using splines, $\zeta_{i,k}^{(j)}$ are the knots. Typically, g will be chosen such that $g_i : \mathbb{R}^{p^{(j)}} \rightarrow \Theta_i^{(j)}$, i.e. so that we do not need to worry about constraining $b_{i,k}^{(j)}$ (e.g. to ensure that the peak enhancement satisfies $\gamma > 1$).

Thus for a given component, we want to estimate:

1. $T_1^{(j)}, T_2^{(j)}$
2. $m_i^{(j)}$
3. $b_{i,k}^{(j)}, \forall k \in \{1, \dots, m_i^{(j)}\}, \forall i \in \{1, \dots, p^{(j)}\}$
4. $\zeta_{i,k}^{(j)}, \forall k \in \{1, \dots, m_i^{(j)}\}, \forall i \in \{1, \dots, p^{(j)}\}$

for all components j . Clearly we may also want to estimate the number of components, but this will be treated as known for now. Indeed, we will start by fixing $T_1^{(j)}$, $T_2^{(j)}$, $m_i^{(j)}$ and $\zeta_{i,k}^{(j)}$, which we assume can be determined during preprocessing.

Choice of basis function

One sensible choice of basis function are B-splines. B-splines of order k with knots ζ form a basis for all splines of order k with knots ζ , and so are a fairly natural choice (De Boor, 1978). We will assume that ζ is an increasing sequence of unique knots. The B-spline basis of order k with knots ζ is a basis comprised of $|\zeta| + k - 2$ B-splines, i.e. we have $m_i = |\zeta| + k - 2$ when using B-splines.

Inference and priors

At least initially, we propose using a composite likelihood (Varin and Vidoni, 2005, for example) to estimate the time evolving parameters. In particular, computing the debiased Whittle likelihood on contiguous windows (perhaps of length 30 minutes) and summing over all time windows. In simulation experiments, we generated hour long time series and computed parameter estimates for each half hour section. We found that the correlation between both the parameter estimates and the debiased Whittle likelihood at the parameter estimates across the two time periods was fairly

low, and so this does not seem overly unreasonable. However, a short gap could be left between time periods in order to reduce the dependence.

For notational simplicity, we drop the superscripts for now. To achieve smoothness of parameters, we suggest following Lang and Brezger (2004) and using a second-order random walk prior for the splines parameters (which corresponds to penalising), i.e. for $i \in \{1, \dots, p\}$

$$b_{i,k} - 2b_{i,k-1} + b_{i,k-2} \sim \mathcal{N}(0, \tau_i^2)$$

for $k > 2$ and uninformative priors for $b_{i,1}$ and $b_{i,2}$. Lang and Brezger (2004) assume a hierarchical model, so that

$$\tau_i^2 \sim \text{IG}(a_i, b_i).$$

In this case we have

$$\pi(\underline{\beta} | X) = L(X | \underline{\beta}) \prod_{i=1}^p \pi(\underline{\beta}_i | \tau_i) \pi(\tau_i).$$

Sampling from the posterior

Lang and Brezger (2004) use a Metropolis-Hastings algorithm to sample from the posterior distribution. Adaptive variations designed to be self tuning can be useful to improve performance, such as Roberts and Rosenthal (2009). However, we have many parameters, and so gradient based techniques such as the Metropolis adjusted Langevin algorithm (Roberts and Rosenthal, 1998, for example) are likely to be preferable. In preliminary simulations, they seem to perform much better.

Alternatively, random knot locations can be used as in DiMatteo et al. (2001), who

use reversible jump MCMC (Green, 1995) to perform parameter inference. This has the advantage that, in some sense the location of the first and last knot determine the start and end time of a component weather system, which would therefore not need to be fixed a priori.

However, it should be noted that the inference procedure described in this section is very hard, and even short time periods (such as a week) might require hundreds of knots, and so we have to estimate hundreds (or even thousands) of parameters, as each model parameter requires its own spline.

4.1.2 Modelling an array of wave staffs

An interesting, albeit less common, experimental setup for monitoring waves is that of an array of wave staffs. Such a setup was used by Donelan et al. (1985) in their experiments to develop the sech^2 spreading function, and Young et al. (1996) who took measurements at Lake George in Australia. Young et al. (1996) observed bimodal spreading, which, in part motivated the bimodal wrapped Gaussian spreading function developed by Ewans (1998). However, the relation to the frequency-direction spectrum in this context is somewhat harder to work with.

In particular, consider wave staffs at locations $\underline{x}_a = [x_a, y_a]^T$ and $\underline{x}_b = [x_b, y_b]^T$ and write $\underline{x} = \underline{x}_a - \underline{x}_b = [x, y]^T$ for convenience. Recall that $\eta = \{\eta(t, x, y)\}_{t, x, y \in \mathbb{R}}$ denotes the surface of the ocean over space and time. Let $\eta_a = \{\eta(t, x_a, y_a)\}_{t \in \mathbb{R}}$ and $\eta_b = \{\eta(t, x_b, y_b)\}_{t \in \mathbb{R}}$ for the time series recorded at locations a and b respectively. We have from linear wave theory that the cross-spectral density function between η_a and η_b is

$$f_{ab}(\omega) = f(\omega) \int_0^{2\pi} \exp\{ik(x \cos \phi + y \sin \phi)\} D(\omega, \phi) d\phi$$

where $\omega^2 = kg \tanh(kh)$ with h being water depth (Benoit et al., 1997). Integrals of this form are surprisingly unpleasant, and as far as we are aware are not solvable analytically.

Therefore, we are in an even more difficult situation than usual: not only must we approximate the autocovariance from the spectral density function, but we also need to approximate the spectral density function from the frequency-direction spectrum. We cannot use the usual trick of writing this as a Fourier transform and using an FFT. Instead when using quadrature methods, we have to approximate each integral in turn (for all required frequencies).

However, we can actually do slightly better. In particular, we show in Appendix B.5 that

$$\int_0^{2\pi} \exp\{ik(x \cos \phi + y \sin \phi)\} D(\omega, \phi) d\phi = \sum_{k=0}^{\infty} \sum_{n=0}^k a^n b^{k-n} J_1(k, n) \quad (4.1.1)$$

where

$$J_1(k, n) = \sum_{r=0}^{k-n} \frac{i^{-n} \cdot (-1)^{k+r}}{2^{k+1}} J_2(k, n, r),$$

such that

$$J_2(k, n, r) = \sum_{j=0}^n \frac{e^{-(2j+2r-k)^2 \sigma^2 / 2} \left(e^{i(2j+2r-k)\mu_1} + e^{i(2j+2r-k)\mu_2} \right)}{r!(k-n-r)!j!(n-j)!}.$$

Note that $J_1(n, k)$ does not depend on a, b so can be computed once and used for all pairs of sensors in the array. We found truncating the sum in (4.1.1) to be more efficient than quadrature methods. If we specify a model for S , then the resulting cross-spectral density functions could be used in the methodology developed in Chapter 3 to perform parameter inference. However, it is still slow, and often unstable. Be-

cause such data are uncommon, this is unlikely to be an issue for most practitioners, but approaches similar to this might be useful in some contexts.

Part II

Detecting departures from Gaussianity in records of ocean waves

" 2π is not one."

— DAVID R. BRILLINGER

Chapter 5

Higher-order spectra

Higher-order spectra have been used to detect and describe non-linearities in a wide variety of applications, from seismology (Haubrich, 1965) to plasma physics (Kim and Powers, 1979) and from neuroscience (He and Thomson, 2010) to oceanography (Hasselmann, 1962; Elgar and Guza, 1985; Chandran et al., 1994; Elgar et al., 1995; Cherneva and Soares, 2007; Ewans et al., 2021). In this chapter, we shall introduce higher-order spectra, and review some of the surrounding literature. We will often first give definitions in the third-order case and then give the general k^{th} -order case. We do this because, whilst Chapter 6 only requires the third-order case, there is growing interest in orders higher than three in the ocean waves literature (Ewans et al., 2021, for example).

The chapter is structured as follows. Some basic definitions are given in Section 5.1, and then the definitions of higher-order spectra are given in Section 5.2. Section 5.3 covers different estimates of higher-order spectra. Section 5.4 discusses different choices of normalisation in the literature. We then discuss existing tests for Gaussianity and linearity based on the third-order spectrum (the bispectrum) in Section 5.5

and give concluding remarks in Section 5.6.

5.1 Background

In this section, we recall definitions for stationarity, Gaussianity and *time series* linearity. Note that linearity means different things to different communities. In particular, from the perspective of ocean waves, a linear process often means that the governing equations of motion are linear (Holthuijsen, 2007, for example).

5.1.1 Stationarity, Gaussianity and linearity

A stochastic process $X = \{X(t)\}_{t \in T}$ is said to be k^{th} -order stationary if the joint distribution of $X(t_1), \dots, X(t_k)$ is equal to the joint distribution of $X(t_1 + \tau), \dots, X(t_k + \tau)$ for all $t_1, \dots, t_k \in T$ and $\tau \in T$. A stochastic process is said to be strictly stationary if for all $k \in \mathbb{N}$, the process is k^{th} -order stationary. Here the domain of the process T is \mathbb{R} in the case of a continuous-time process or $\Delta\mathbb{Z}$ in the case of a discrete-time process sampled evenly every Δ time points.

A stochastic process X is said to be a Gaussian process if, for any $n \in \mathbb{N}$ and $t_1, \dots, t_n \in T$, the multivariate random variable $X(t_1), \dots, X(t_n)$ is multivariate Gaussian.

For a discrete-time process, we will say that a process Y is the result of applying a linear filter to X if

$$Y(t) = \sum_{j \in \Delta\mathbb{Z}} a(j)X(t-j) = [a * X](t)$$

where $a : \Delta\mathbb{Z} \rightarrow \mathbb{R}$ is referred to as the impulse response of the filter and its Fourier

transform $h : [-\pi/\Delta, \pi/\Delta] \rightarrow \mathbb{C}$ such that

$$h(\omega) = \sum_{t \in \Delta\mathbb{Z}} a(t)e^{-it\omega}$$

is called the transfer function of the filter (Koopmans, 1995, for example).

A univariate stationary time series is said to be linear if, for all $t \in \Delta\mathbb{Z}$,

$$X(t) = \sum_{j \in \Delta\mathbb{Z}} a(j)\epsilon(t-j) \quad (5.1.1)$$

where $\epsilon(t)$ are i.i.d. (sometimes referred to as a pure white noise process). Such processes are stationary by construction, and so linear is a stronger condition than stationary under this definition (Brillinger, 1974). Note that, by Wold's Theorem (Wold, 1938), any stationary time series can be written in the form of (5.1.1) where the noise process $\{\epsilon(t)\}_{t \in \Delta\mathbb{Z}}$ is a white noise process, but not necessarily an i.i.d. process. Processes for which this noise is not i.i.d. are non-linear. Interestingly, this means that all of the non-linearity present in a stationary time series is actually present in the noise of the Wold decomposition.

5.1.2 Cumulants

Consider univariate random variables Y_1, \dots, Y_k , and assume that all of their moments exist. The cumulant generating function of Y_1, \dots, Y_k is defined as the log of the characteristic function, i.e.

$$\psi(\lambda_1, \dots, \lambda_k) = \log \mathbb{E} \left[e^{i \sum_{j=1}^k \lambda_j Y_j} \right].$$

The k^{th} order cumulant is then defined as

$$\mathcal{C}[Y_1, \dots, Y_k] = (-i)^k \frac{\partial^k \psi(\lambda_1, \dots, \lambda_k)}{\partial \lambda_1 \dots \partial \lambda_k} \Bigg|_{\lambda_j=0, \forall j \in \{1, \dots, k\}}.$$

Cumulants are related to moments as follows (Leonov and Shiryaev, 1959)

$$\mathcal{C}[Y_1, \dots, Y_k] = \sum_{P \in \mathcal{P}_k} (-1)^{|P|-1} (|P|-1)! \prod_{A \in P} \mathbb{E} \left[\prod_{j \in A} Y_j \right]$$

where \mathcal{P}_k is the set of all partitions of $\{1, \dots, k\}$. We also have the inverse relation

$$\mathbb{E} \left[\prod_{j=1}^k Y_j \right] = \sum_{P \in \mathcal{P}_k} \prod_{A \in P} \mathcal{C}[Y_j \mid j \in A]$$

where

$$\mathcal{C}[Y_j \mid j \in A] = \mathcal{C}[Y_{A_1}, \dots, Y_{A_{|A|}}].$$

For example, consider mean zero random variables Y_1, Y_2, Y_3, Y_4 , then

$$\begin{aligned} \mathcal{C}[Y_1, Y_2] &= \mathbb{E}[Y_1 Y_2], \\ \mathcal{C}[Y_1, Y_2, Y_3] &= \mathbb{E}[Y_1 Y_2 Y_3], \\ \mathcal{C}[Y_1, Y_2, Y_3, Y_4] &= \mathbb{E}[Y_1 Y_2 Y_3 Y_4] - \mathbb{E}[Y_1 Y_2] \mathbb{E}[Y_3 Y_4] \\ &\quad - \mathbb{E}[Y_1 Y_3] \mathbb{E}[Y_2 Y_4] \\ &\quad - \mathbb{E}[Y_1 Y_4] \mathbb{E}[Y_2 Y_3]. \end{aligned}$$

5.2 Higher-order spectra

In this section, we shall give some basic definitions and properties of higher-order spectra. Such definitions can be found in Brillinger (1965, 1974); Rosenblatt (1980); Nikias (1993) for example. We begin in Section 5.2.1 with the third-order case, and then discuss the general case in Section 5.2.2. Note that the general case is not required for Chapter 6, and could be skipped on first reading.

Unless specified otherwise, the quantities in this chapter will be defined for the processes $X_a = \{X_a(t)\}_{t \in \mathbb{R}}$ for $a \in \{1, \dots, d\}$, which are real-valued continuous-time mean-zero stochastic processes, and their discrete-time counterparts $X_a^{(\Delta)} = \{X_a(t)\}_{t \in \Delta\mathbb{Z}}$, arising from sampling X_a every Δ seconds. We assume further that all moments (equivalently cumulants) of the process exist.

5.2.1 The third-order case

The continuous-time case

Let $a_1, a_2, a_3 \in \{1, \dots, d\}$, then the third-order cumulant function between processes X_{a_1}, X_{a_2} and X_{a_3} is $\tilde{c}_{a_1 a_2 a_3} : \mathbb{R}^3 \rightarrow \mathbb{R}$ such that

$$\tilde{c}_{a_1 a_2 a_3}(t_1, t_2, t_3) = \mathcal{C}[X_{a_1}(t_1), X_{a_2}(t_2), X_{a_3}(t_3)].$$

Since the processes in question are mean-zero, the cumulant function is equal to the moment function. Under the additional assumption of stationarity, $\tilde{c}_{a_1 a_2 a_3}$ is invariant to time shifts, so define the function $c_{a_1 a_2 a_3} : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

$$c_{a_1 a_2 a_3}(\tau_1, \tau_2) = \tilde{c}_{a_1 a_2 a_3}(\tau_1, \tau_2, 0).$$

This contains all of the information present in $\tilde{c}_{a_1 a_2 a_3}$ as we have for any $t \in \mathbb{R}$

$$\begin{aligned} c_{a_1 a_2 a_3}(\tau_1, \tau_2) &= \tilde{c}_{a_1 a_2 a_3}(\tau_1, \tau_2, 0) \\ &= \tilde{c}_{a_1 a_2 a_3}(\tau_1 + t, \tau_2 + t, t), \end{aligned}$$

similarly to how the autocovariance can be reduced to be dependent on just one variable τ in the second-order stationary case.

Assuming that $c_{a_1 a_2 a_3}$ is absolutely integrable¹, the third-order cumulant spectral density function $f_{a_1 a_2 a_3} : \mathbb{R}^2 \rightarrow \mathbb{C}$, is defined as (Brillinger, 1965)

$$f_{a_1 a_2 a_3}(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} c_{a_1 a_2 a_3} e^{-i(\tau_1 \omega_1 + \tau_2 \omega_2)} d\tau_1 d\tau_2.$$

The third-order spectral density function is often referred to as the bispectrum, a name first introduced by J. W. Tukey (Brillinger, 1991).

The discrete-time case

Similar quantities can be defined in the discrete-time case. In particular, the third-order cumulant sequence² between processes X_{a_1} , X_{a_2} and X_{a_3} is $\tilde{c}_{a_1 a_2 a_3}^{(\Delta)} : \Delta\mathbb{Z}^3 \rightarrow \mathbb{R}$ such that

$$\tilde{c}_{a_1 a_2 a_3}^{(\Delta)}(t_1, t_2, t_3) = \mathcal{C}[X_{a_1}(t_1), X_{a_2}(t_2), X_{a_3}(t_3)].$$

¹The third-order cumulant spectrum can be defined without this assumption, but it is not a function in the standard sense, see Section 5.2.2 (or Brillinger, 1965) for more details.

²The slightly unusual notation for a sequence is used to avoid overwhelming subscripts, but the distinction is denoted by the presence of Δ in the superscript.

Note that this is equal to the third-order cumulant function on $\Delta\mathbb{Z}^3$. Under stationarity, again define the sequence $c_{a_1 a_2 a_3}^{(\Delta)} : \Delta\mathbb{Z}^2 \rightarrow \mathbb{R}$ such that

$$c_{a_1 a_2 a_3}^{(\Delta)}(\tau_1, \tau_2) = \tilde{c}_{a_1 a_2 a_3}^{(\Delta)}(\tau_1, \tau_2, 0).$$

Under the further assumption that $c_{a_1 a_2 a_3}^{(\Delta)}$ is absolutely summable, the aliased third-order spectral density function is defined to be $f_{a_1 a_2 a_3}^{(\Delta)} : [-\pi/\Delta, \pi/\Delta]^2 \rightarrow \mathbb{C}$, such that

$$f_{a_1 a_2 a_3}^{(\Delta)} = \frac{\Delta^2}{(2\pi)^2} \sum_{\tau_1 \in \Delta\mathbb{Z}} \sum_{\tau_2 \in \Delta\mathbb{Z}} c_{a_1 a_2 a_3}^{(\Delta)} e^{-i(\tau_1 \omega_1 + \tau_2 \omega_2)}.$$

Aliasing

For all $\omega_1, \omega_2 \in [-\pi/\Delta, \pi/\Delta]$,

$$f_{a_1 a_2 a_3}^{(\Delta)}(\omega_1, \omega_2) = \sum_{j_1 \in \Delta\mathbb{Z}} \sum_{j_2 \in \Delta\mathbb{Z}} f_{a_1 a_2 a_3}(\omega_1 + \frac{2\pi}{\Delta} j_1, \omega_2 + \frac{2\pi}{\Delta} j_2)$$

analogously to aliasing in the second order case, assuming certain regularity conditions on $f_{a_1 a_2 a_3}$ to allow for an interchange of limits, see Brillinger and Rosenblatt (1967b).

Symmetry

The bispectrum of a (real-valued) stochastic process has a number of symmetries (Brillinger and Rosenblatt, 1967b), in particular

$$\begin{aligned} f_{a_1 a_2 a_3}^{(\Delta)}(\omega_1, \omega_2) &= \overline{f_{a_1 a_2 a_3}^{(\Delta)}(-\omega_1, -\omega_2)}, \\ f_{a_1 a_2 a_3}^{(\Delta)}(\omega_1, \omega_2) &= f_{a_{\rho_1} a_{\rho_2} a_{\rho_3}}^{(\Delta)}(\omega_{\rho_1}, \omega_{\rho_2}) \end{aligned}$$

where \bar{z} denotes the complex conjugate of z ; ρ is any permutation of $\{1, 2, 3\}$; and $\omega_3 = -\omega_1 - \omega_2 \pmod{2\pi/\Delta}$ (usually so that $|\omega_3| < \pi/\Delta$).³ These are shown in Figure 5.2.1 (left), which we plot in the full three frequency space, in order to make the presence of the third frequency more clear. Therefore, we will often only consider the principle domain

$$\Omega_{PD} = \{(\omega_1, \omega_2) \mid 0 \leq \omega_1 \leq \omega_2 \text{ \& \ } \omega_1 + 2\omega_2 \leq 2\pi/\Delta\},$$

as discussed by Brillinger and Rosenblatt (1967b) in the univariate case and Wong (1997) in the multivariate case. The principle domain is shown in Figure 5.2.1 (right). Note that these symmetries are more than just a mathematical curiosity. In particular, in Chapter 6 we wish to account for multiple testing, and so we must first be careful to remove any symmetries before testing.

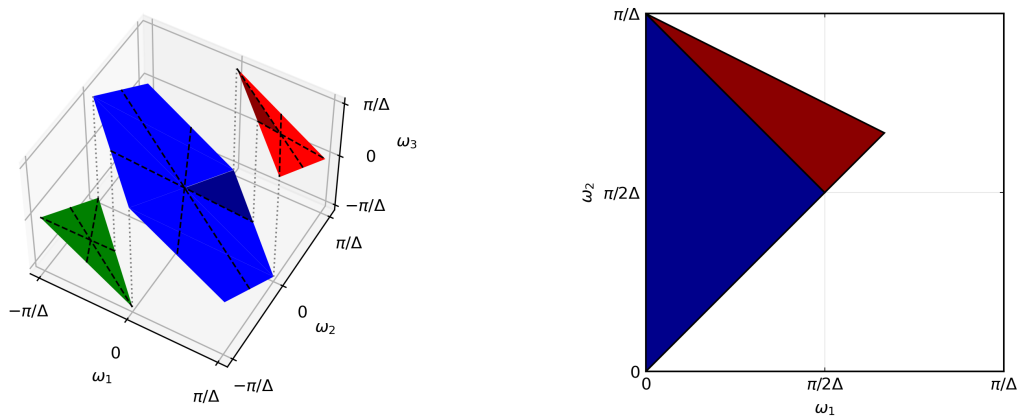


Figure 5.2.1: Symmetries of the bispectrum (left) and principle domain (right). The principle domain corresponds to the triangles highlighted in the darker blue and red in the plot of the symmetries (left).

³The symmetries are analogous in the continuous-time case.

5.2.2 The general case

The continuous-time case

The k^{th} -order cumulant function of X_{a_1}, \dots, X_{a_k} for $a_j \in \{1, \dots, d\}$ is defined as $\tilde{c}_{a_1 \dots a_k} : \mathbb{R}^k \rightarrow \mathbb{R}$ such that

$$\tilde{c}_{a_1 \dots a_k}(t_1, \dots, t_k) = \mathcal{C}[X_{a_1}(t_1), \dots, X_{a_k}(t_k)].$$

For a stationary process, write the function $c_{a_1 \dots a_k} : \mathbb{R}^{k-1} \rightarrow \mathbb{R}$, so that

$$c_{a_1 \dots a_k}(\tau_1, \dots, \tau_{k-1}) = \tilde{c}_{a_1 \dots a_k}(\tau_1, \dots, \tau_{k-1}, 0)$$

which by shift invariance of $\tilde{c}_{a_1 \dots a_k}$ contains all available information. Assuming that $c_{a_1 \dots a_k}$ is absolutely integrable, define the k^{th} -order (cumulant) spectral density function $f_{a_1 \dots a_k} : \mathbb{R}^{k-1} \rightarrow \mathbb{C}$ such that

$$f_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) = \frac{1}{(2\pi)^{k-1}} \int_{\mathbb{R}^{k-1}} c_{a_1 \dots a_k}(\tau_1, \dots, \tau_{k-1}) e^{-i \sum_{j=1}^{k-1} \tau_j \omega_j} d\tau_1 \dots d\tau_{k-1}.$$

It is convenient at this point to discuss the difference between moments and cumulants, as pointed out by Brillinger (1965).⁴ Analogously to cumulants, the k^{th} -order moment function is defined by Brillinger (1965) as $\tilde{m}_{a_1 \dots a_k} : \mathbb{R}^k \rightarrow \mathbb{R}$ such that

$$\tilde{m}_{a_1 \dots a_k}(t_1, \dots, t_k) = \mathbb{E} \left[\prod_{j=1}^k X_{a_j}(t_j) \right]$$

and for stationary processes again write $m_{a_1 \dots a_k} : \mathbb{R}^{k-1} \rightarrow \mathbb{R}$ such that

$$m_{a_1 \dots a_k}(\tau_1, \dots, \tau_{k-1}) = \tilde{m}_{a_1 \dots a_k}(\tau_1, \dots, \tau_{k-1}, 0).$$

⁴In the third-order case they are equal (when the process is mean-zero).

The k^{th} -order moment spectral density function $g_{a_1 \dots a_k}$ with domain \mathbb{R}^{k-1} is then

$$g_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) = \frac{1}{(2\pi)^{k-1}} \int_{\mathbb{R}^{k-1}} m_{a_1 \dots a_k}(\tau_1, \dots, \tau_{k-1}) e^{-i \sum_{j=1}^{k-1} \tau_j \omega_j} d\tau_1 \dots d\tau_{k-1}.$$

Note we do not specify the codomain, as this may not be a function in the usual sense, and often isn't (Brillinger, 1965). If the k^{th} -order moment function is absolutely integrable then, as in the case of cumulants, it will be a complex valued function (Brillinger and Rosenblatt, 1967b). However, for many classes of processes (e.g. Gaussian processes), this is not true. In particular, it will require a similar treatment to the Dirac delta function as a measure. Similarly, for now, we relax the previous assumption of $c_{a_1 \dots a_k}$ being absolutely integrable. The moment spectra are related to the cumulant spectra by the following:

$$f_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) = \sum_{P \in \mathcal{P}_k} (-1)^{|P|-1} (|P|-1)! \prod_{A \in P} \tilde{g}_A(\underline{\omega}_A) \delta\left(\sum_{j \in A} \omega_j\right) \quad (5.2.1)$$

$$g_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) = \sum_{P \in \mathcal{P}_k} \prod_{A \in P} \tilde{f}_A(\underline{\omega}_A) \delta\left(\sum_{j \in A} \omega_j\right) \quad (5.2.2)$$

where $\omega_k = -\sum_{j=1}^{k-1} \omega_j$ and

$$\tilde{f}_A(\underline{\omega}_A) = f_{a_{A_1} \dots a_{A_r}}(\omega_{A_1}, \dots, \omega_{A_{r-1}}),$$

$$\tilde{g}_A(\underline{\omega}_A) = g_{a_{A_1} \dots a_{A_r}}(\omega_{A_1}, \dots, \omega_{A_{r-1}}),$$

where $r = |A|$.⁵ Here we write A_j for the j^{th} element of A in order to distinguish them from a_j . Notice that the k^{th} -order cumulant and moment spectral density functions are equal except when $\sum_{j \in A} \omega_j = 0$ for any proper subset A of $\{1, \dots, k\}$. The

⁵The "tilde" notation here is deliberate, because we are using the r frequency version of the r^{th} -order spectral density function, analogously to the r -lag version of the cumulant function.

set of such frequencies

$$\left\{ [\omega_1, \dots, \omega_k]^T \in \mathbb{R}^k \mid \sum_{j \in A} \omega_j = 0 \right\}$$

is referred to as a *proper submanifold* by Brillinger and Rosenblatt (1967a). Also, the set

$$\left\{ [\omega_1, \dots, \omega_k]^T \in \mathbb{R}^k \mid \sum_{j=1}^k \omega_j = 0 \right\}$$

is referred to as the *principle manifold*, because the k^{th} -order spectral density function really describes the density of a measure supported on this manifold (Brillinger and Rosenblatt, 1967a). We saw an example of this pictorially in Figure 5.2.1 for the third-order case.

Brillinger (1965) points out that, from (5.2.1) and (5.2.2), at least one of the cumulant spectral density function or moment spectral density function are not functions in the usual sense (unless they are equal). Under the assumption that $c_{a_1 \dots a_k}$ is absolutely integrable, then $f_{a_1 \dots a_k}$ is a complex-valued function, and the corresponding moment spectral density function $g_{a_1 \dots a_k}$ is not. In particular, in the Gaussian case for $k > 2$ the cumulant spectral density function is zero everywhere. For odd k , the moment spectral density function is also zero; however, for even k this is not the case. For this reason Brillinger (1965) suggests using cumulants over moments in general, and so we follow this convention, and will refer to the k^{th} -order cumulant spectral density function as simply the k^{th} -order spectral density function.⁶

⁶This discussion holds similarly in the discrete-time case.

The discrete-time case

As in the continuous-time case the k^{th} -order cumulant sequence is $\tilde{c}_{a_1 \dots a_k} : \Delta\mathbb{Z}^k \rightarrow \mathbb{R}$ such that

$$\tilde{c}_{a_1 \dots a_k}^{(\Delta)}(t_1, \dots, t_k) = \mathcal{C}[X_{a_1}(t_1), \dots, X_{a_k}(t_k)].$$

Clearly where it is defined, the cumulant sequence is the same as the cumulant function. In other words, the cumulant sequence is just the domain restriction of the cumulant function, i.e. $\tilde{c}_{a_1 \dots a_k}^{(\Delta)} = \tilde{c}_{a_1 \dots a_k} |_{\Delta\mathbb{Z}^k}$.

As in the continuous-time case, for a stationary process we write $c_{a_1 \dots a_k}^{(\Delta)} : \mathbb{R}^{k-1} \rightarrow \mathbb{R}$ such that

$$c_{a_1 \dots a_k}^{(\Delta)}(\tau_1, \dots, \tau_{k-1}) = \tilde{c}_{a_1 \dots a_k}^{(\Delta)}(\tau_1, \dots, \tau_{k-1}, 0).$$

Assuming that $c_{a_1 \dots a_k}^{(\Delta)}$ is absolutely summable, the k^{th} -order (cumulant) spectral density function $f_{a_1 \dots a_k}^{(\Delta)} : \mathbb{R}^{k-1} \rightarrow \mathbb{C}$ is defined by

$$f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) = \frac{\Delta^{k-1}}{(2\pi)^{k-1}} \sum_{\tau_1 \in \Delta\mathbb{Z}} \dots \sum_{\tau_{k-1} \in \Delta\mathbb{Z}} c_{a_1 \dots a_k}^{(\Delta)}(\tau_1, \dots, \tau_{k-1}) e^{-i \sum_{j=1}^{k-1} \tau_j \omega_j}.$$

The relation to the amplitude process

By the spectral representation theorem (Cramer and Leadbetter, 1967), for a stationary processes X_a , for all $t \in \mathbb{R}$

$$X_a(t) = \int_{-\infty}^{\infty} e^{it\omega} d\zeta_a(\omega).$$

The process $\zeta_a = \{\zeta_a(\omega)\}_{\omega \in \mathbb{R}}$ is called the spectral process (or sometimes the amplitude process⁷) associated with X , and is conjugate symmetric about zero.

Similarly in discrete-time, for any $t \in \Delta\mathbb{Z}$

$$X_a(t) = \int_{-\pi/\Delta}^{\pi/\Delta} e^{it\omega} d\zeta_a^{(\Delta)}(\omega),$$

where $\zeta_a^{(\Delta)} = \{\zeta_a^{(\Delta)}(\omega)\}_{\omega \in \mathbb{R}}$ is the spectral process associated with $X_a^{(\Delta)}$. These are related to the k^{th} -order cumulant spectral density function by (Brillinger and Rosenblatt, 1967b)

$$\mathcal{C} [d\zeta_{a_1}(\omega_1), \dots, d\zeta_{a_k}(\omega_k)] = \delta \left(\sum_{j=1}^k \omega_j \right) f_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) d\omega_1 \dots d\omega_k,$$

in the continuous-time case⁸ and

$$\mathcal{C} [d\zeta_{a_1}^{(\Delta)}(\omega_1), \dots, d\zeta_{a_k}^{(\Delta)}(\omega_k)] = \delta_{\Delta} \left(\sum_{j=1}^k \omega_j \right) f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) d\omega_1 \dots d\omega_k,$$

in the discrete-time case, where δ is the Dirac delta function and δ_{Δ} is a periodic extension of the Dirac delta function satisfying

$$\delta_{\Delta}(\omega) = \sum_{j \in \mathbb{Z}} \delta \left(\omega + \frac{2\pi j}{\Delta} \right)$$

for $\omega_1, \dots, \omega_k \in \mathbb{R}^k$, which we shall refer to as the aliased Dirac delta function. This yields a helpful heuristic interpretation of the higher-order spectra as describing the higher-order cumulants of the amplitude/phase of random regular waves which constitute the time series of interest.

⁷Really ζ_a contains both the amplitude and phase.

⁸Note that this notation is shorthand, and only makes sense under an integral. See Brillinger (1974), Section 4.6, for example.

Aliasing

The aliasing seen in the case of $k = 2$ is also seen for $k > 2$. In particular, Brillinger and Rosenblatt (1967b) show that for all $\omega, \lambda \in [-\pi/\Delta, \pi/\Delta]$,

$$\zeta_a^{(\Delta)}(\omega) - \zeta_a^{(\Delta)}(\lambda) = \sum_{r \in \mathbb{Z}} \zeta_a(\omega + \frac{2\pi r}{\Delta}) - \zeta_a(\lambda + \frac{2\pi r}{\Delta})$$

and that for $\omega_1, \dots, \omega_{k-1} \in [-\pi/\Delta, \pi/\Delta]$

$$f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) = \sum_{r \in \mathbb{Z}^{k-1}} f_{a_1 \dots a_k}(\omega_1 + \frac{2\pi r_1}{\Delta}, \dots, \omega_{k-1} + \frac{2\pi r_{k-1}}{\Delta}),$$

under certain regularity conditions (see Brillinger and Rosenblatt, 1967b).

Symmetry

In the case of a real valued process Brillinger and Rosenblatt (1967a) show

$$f_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) = \overline{f_{a_1 \dots a_k}(-\omega_1, \dots, -\omega_{k-1})}.$$

Additionally, if ρ is any permutation of $\{1, \dots, k\}$, then Brillinger and Rosenblatt (1967b) show that for $\omega_1, \dots, \omega_{k-1} \in \mathbb{R}$

$$f_{a_1 \dots a_k}(\omega_1, \dots, \omega_{k-1}) = f_{a_{\rho_1} \dots a_{\rho_k}}(\omega_{\rho_1}, \dots, \omega_{\rho_{k-1}}),$$

where as usual, we take $\omega_k = -\sum_{j=1}^{k-1} \omega_j$. The discrete-time symmetries are equivalent.

5.3 Estimating higher-order spectra

We begin in Section 5.3.1 with a discussion of some of the more popular estimators for the bispectrum. We will state them in their simplest cases, with the more general cases covered by the discussion for general k in Section 5.3.2.

5.3.1 The third-order case

As in the second-order case, estimation of the third-order spectral density function tends to begin with the discrete Fourier transform of the observed series. For a time series $X_a(0), X_a(\Delta), \dots, X_a([n-1]\Delta)$ of length n , the discrete Fourier transform is given by

$$J_a^{(n)}(\omega) = \sum_{t=0}^{n-1} X_a(t\Delta) e^{-it\Delta\omega}.$$

The periodogram

The third-order periodogram is defined to be

$$I_{a_1 a_2 a_3}^{(n)}(\omega_1, \omega_2) = \frac{\Delta^{k-1}}{(2\pi)^{k-1}} \frac{1}{n} J_{a_1}^{(n)}(\omega_1) J_{a_2}^{(n)}(\omega_2) J_{a_3}^{(n)}(\omega_3),$$

where $\omega_3 = -\omega_1 - \omega_2 \pmod{2\pi/\Delta}$ as usual. The third-order periodogram is asymptotically unbiased but not consistent (Brillinger and Rosenblatt, 1967a). In fact, the variance of the third-order periodogram tends to infinity as $n \rightarrow \infty$. In particular, as shown by (Brillinger and Rosenblatt, 1967a),

$$\text{var} \left(I_{a_1 a_2 a_3}^{(n)}(\omega_1, \omega_2) \right) = n f_{a_1 a_1}^{(\Delta)}(\omega_1) f_{a_2 a_2}^{(\Delta)}(\omega_2) f_{a_3 a_3}^{(\Delta)}(\omega_3) + O(1).$$

Therefore, some kind of smoothing is required. Here we state some of the most commonly used estimators in practice. More detailed and general estimators are

covered in Section 5.3.2.

Segment averaging

The first approach is segment averaging (Hinich and Wolinsky, 2005, for example). The simplest version of segment averaging breaks up a record into B contiguous blocks of length n_l , and computes the third-order periodogram for each block and averages them, i.e.

$$\hat{f}_{a_1 a_2 a_3}^{(n_l, B)}(\omega_1, \omega_2) = \frac{1}{B} \sum_{b=0}^{B-1} I_{a_1 a_2 a_3}^{(bn_l : b+n_l-1)}(\omega_1, \omega_2)$$

where $I_{a_1 a_2 a_3}^{(r:s)}$ denotes the third-order periodogram computed from the data at time points $r\Delta, \dots, s\Delta$. A more general case with tapering and overlapping blocks is stated in Section 5.3.2.

Hinich and Wolinsky (2005) show that as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{E} \left[\hat{f}_{a_1 a_2 a_3}^{(n_l, B)}(\omega_1, \omega_2) \right] &= I_{a_1 a_2 a_3}^{(n_l)}(\omega_1, \omega_2) \\ &= f_{a_1 a_2 a_3}(\omega_1, \omega_2) + O(1/n_l) \end{aligned}$$

and

$$\begin{aligned} \text{var} \left(\hat{f}_{a_1 a_2 a_3}^{(n_l, B)}(\omega_1, \omega_2) \right) &= \frac{1}{B} \text{var} \left(I_{a_1 a_2 a_3}^{(n_l)}(\omega_1, \omega_2) \right) \\ &= \frac{n_l}{B} f_{a_1 a_1}^{(\Delta)}(\omega_1) f_{a_2 a_2}^{(\Delta)}(\omega_2) f_{a_3 a_3}^{(\Delta)}(\omega_3) + O(1). \end{aligned}$$

Therefore, we require $n_l/B \rightarrow 0$ as $n \rightarrow \infty$. Without loss of generality, consider the case that $n_l B = n$. Then taking $n_l = n^{1-c}$ we have $B = n^c$ and so $n_l/B = n^{1-2c}$ which tends to zero provided $0.5 < c < 1$. It will be convenient to write $C_n = B n_l / n \approx B^2$

so that

$$\text{var} \left(\hat{f}_{a_1 a_2 a_3}^{(n_l, B)}(\omega_1, \omega_2) \right) = \frac{n}{C_n} f_{a_1 a_1}^{(\Delta)}(\omega_1) f_{a_2 a_2}^{(\Delta)}(\omega_2) f_{a_3 a_3}^{(\Delta)}(\omega_3) + O(1),$$

as $n \rightarrow \infty$.

Frequency domain smoothing

Another popular technique is uniform smoothing over neighbouring Fourier frequencies (Hinich, 1982, for example).⁹ These estimates take the form

$$\hat{f}_{a_1 a_2 a_3}^{(W, b_n, n)}(\omega_1, \omega_2) = \sum_{j_1=-b_n}^{b_n} \sum_{j_2=-b_n}^{b_n} W_n \left(\omega_1 + \frac{2\pi j_1}{n\Delta}, \omega_2 + \frac{2\pi j_2}{n\Delta} \right)$$

where W_n is a uniform smoothing window and b_n is an integer determining the width of the smoothing (e.g. Birkelund and Hanssen, 2009).

Two popular windows are rectangular windows Hinich (1982); Wong (1997) and hexagonal windows Nikias and Raghuveer (1987); Birkelund and Hanssen (2009).

These take the form

$$W_n^{\square}(\omega_1, \omega_2) = \begin{cases} 1/C_n & \text{if } |\omega_1|, |\omega_2| \leq 2\pi b_n/n\Delta, \\ 0 & \text{otherwise,} \end{cases}$$

for rectangular windows and

$$W_n^{\diamond}(\omega_1, \omega_2) = \begin{cases} 1/C_n & \text{if } \max\{|\omega_1|, |\omega_2|, |\omega_1 + \omega_2|\} \leq 2\pi b_n/n\Delta, \\ 0 & \text{otherwise,} \end{cases}$$

for hexagonal windows, where in both cases C_n is the number of points in the region

⁹There are more general variants of frequency domain smoothing, which we state in Section 5.3.2 for the general case.

used in the average (equal to $(2b_n + 1)^2$ in the rectangular case).

In this case, the variance of the estimate is

$$\text{var} \left(\hat{f}_{a_1 a_2 a_3}^{(W, b_n, n)}(\omega_1, \omega_2) \right) = \frac{n}{C_n} f_{a_1 a_1}^{(\Delta)}(\omega_1) f_{a_2 a_2}^{(\Delta)}(\omega_2) f_{a_3 a_3}^{(\Delta)}(\omega_3) + O(1)$$

as $n \rightarrow \infty$ (Hinich, 1982). If we choose $b_n = \lfloor n^c/2 \rfloor$ with $0.5 < c < 1$ then $n/C_n \rightarrow 0$ as $n \rightarrow \infty$ as desired. We define the quantity C_n for all of these methods because it appears in the tests for Gaussianity and linearity described in Section 5.5, and in the bias correction in Section 6.3.3.

5.3.2 The general case

Brillinger and Rosenblatt (1967a) show that, assuming

$$\sum_{\boldsymbol{\tau} \in \mathbb{Z}^{k-1}} |\tau_j c_{a_1 \dots a_k}^{(\Delta)}(\tau_1, \dots, \tau_{k-1})| < \infty,$$

for $j \in \{1, \dots, k-1\}$, then for $\omega_1, \dots, \omega_k \in [-\pi/\Delta, \pi/\Delta]^k$ the joint cumulant of the discrete Fourier transform is

$$\mathcal{C} \left[J_{a_1}^{(n)}(\omega_1), \dots, J_{a_k}^{(n)}(\omega_k) \right] = \left(\frac{2\pi}{\Delta} \right)^{k-1} \mathcal{D}_n \left(\Delta \sum_{j=1}^k \omega_j \right) f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) + O(1)$$

as $n \rightarrow \infty$, where $\mathcal{D}_n : \mathbb{R} \rightarrow \mathbb{R}$ is the Dirichlet kernel defined to be

$$\mathcal{D}_n(\omega) = \frac{\sin(\omega n/2)}{\sin(\omega/2)} e^{i\omega(n-1)/2}.$$

Importantly, $\mathcal{D}_n(\omega) = n$ if $\omega \equiv 0 \pmod{2\pi}$ and $\mathcal{D}_n(\omega) = 0$ if $\omega = 2\pi k/n$ for $k \not\equiv 0 \pmod{n}$ (Brillinger and Rosenblatt, 1967a). This means that when $\omega_1, \dots, \omega_k$ are Fourier frequencies on the principle manifold but not on any of the proper sub-

manifolds, we have

$$\mathbb{E} \left[\frac{\Delta^{k-1}}{n(2\pi)^{k-1}} \prod_{j=1}^k J_{a_j}^{(n)}(\omega_j) \right] = f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) + O(1/n),$$

as $n \rightarrow \infty$. The behaviour on the proper submanifolds can be substantially different, and can cause issues in certain circumstances, see for example the discussion of [Hinich and Wolinsky \(2005\)](#), which we review in [Section 5.4](#).

Additionally, define the tapered discrete Fourier transform

$$J_{a,v}^{(n)}(\omega) = \sum_{t=0}^{n-1} v_n(t) X_a(t\Delta) e^{-it\Delta\omega}$$

where v is called a taper (and actually depends on n in general, but we suppress this for notational convenience). In this more general case, [Brillinger \(1974\)](#) shows that

$$\mathcal{C} \left[J_{a_1, v_1}^{(n)}(\omega_1), \dots, J_{a_k, v_k}^{(n)}(\omega_k) \right] = \left(\frac{2\pi}{\Delta} \right)^{k-1} \mathcal{V}_k^{(n)} \left(\Delta \sum_{j=1}^k \omega_j \right) f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) + O(1).$$

as $n \rightarrow \infty$, where

$$\mathcal{V}_k^{(n)}(\omega) = \sum_{t=1}^n \prod_{j=1}^k v_j(t) e^{-it\omega}.$$

Notice that we have more than one taper here. This is required for multitapering, which we describe in this section for the general case (which can of course be used in the third-order case, but is less commonly used and the description in the general case is fairly similar to the third-order case).

The periodogram

Most estimation techniques begin with the k^{th} -order periodogram, defined to be

$$I_{\underline{a}, \underline{v}}^{(n)}(\omega_1, \dots, \omega_{k-1}) = \frac{\Delta^{k-1}}{(2\pi)^{k-1} \mathcal{V}_k^{(n)}(0)} \prod_{j=1}^k J_{a_j, v_j}^{(n)}(\omega_j),$$

where $\underline{a} = [a_1, \dots, a_k]^T$ and $\underline{v} = [v_1, \dots, v_k]^T$. Notice that the tapers do not need to be the same for each series, though they often are chosen to be the same. In the case when no taper is used, we shall write $I_{\underline{a}}$.

For the untapered version, Brillinger and Rosenblatt (1967a) show that, off the proper submanifolds,

$$\mathbb{E} \left[I_{\underline{a}}^{(n)}(\omega_1, \dots, \omega_{k-1}) \right] = f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) + O(1/n),$$

and

$$n^{2-k} \text{var} \left(I_{\underline{a}}^{(n)}(\omega_1, \dots, \omega_{k-1}) \right) = \prod_{j=1}^k f_{a_1 \dots a_k}(\omega_j) + O(1/n),$$

as $n \rightarrow \infty$. Thus, for $k > 2$ the k^{th} -order periodogram is anti-consistent (Thomson, 1989). As such, some form of averaging is required. Typically this comes in one of three forms: segment averaging, smoothing over frequency, or multitapering.

Segment averaging

Originally proposed for $k = 2$ with non-overlapped segments, segment averaging was generalised by Welch (1967) to overlapping segments with tapers. If we use segments of length n_l with a shift of n_s , then we have an overlap of $n_l - n_s$. The number of blocks used is $B = \lfloor (n - n_l)/n_s \rfloor + 1$ and the segment averaged estimate

is

$$\hat{f}_{\underline{a}, \underline{v}}^{(n_l, n_s, B)}(\omega_1, \dots, \omega_{k-1}) = \frac{1}{B} \sum_{b=0}^{B-1} I_v^{(bn_s:bn_s+n_l-1)}(\omega_1, \dots, \omega_{k-1})$$

where

$$I_{\underline{a}, \underline{v}}^{(r:s)}(\omega_1, \dots, \omega_{k-1}) = \frac{\Delta^{k-1}}{(2\pi)^{k-1} \mathcal{V}_k^{(n_l)}(0)} \prod_{j=1}^k J_{a_j, v_j}^{(r:s)}(\omega_j)$$

is the k^{th} -order periodogram of a segment from observation r to observation s and

$$J_{a, v}^{(r:s)}(\omega) = \sum_{t=0}^{s-r} v(t) X_a([t+r]\Delta) e^{-it\Delta\omega}$$

is the discrete Fourier transform of the segment with taper v . Note that the taper used here is for a series of length n_l not n .

For segment averaging with no taper, (Hinich and Wolinsky, 2005) show that asymptotically, as $n \rightarrow \infty$,

$$\mathbb{E} \left[\hat{f}_{\underline{a}}^{(n_l, n_s, B)}(\omega_1, \dots, \omega_{k-1}) \right] = f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) + O(1/n_l)$$

and for segment averaging with no overlap

$$n_l^{2-k} B \text{var} \left(\hat{f}_{\underline{a}}^{(n_l, 0, B)}(\omega_1, \dots, \omega_{k-1}) \right) = \prod_{j=1}^k f_{a_j a_j}^{(\Delta)}(\omega_j) + O(1/n_l).$$

Frequency domain smoothing

Frequency domain smoothing works takes a weighted average of the periodogram over nearby frequencies (essentially as a convolution). This can either be discrete-frequency (Brillinger and Rosenblatt, 1967a) or continuous-frequency (Lii and Rosen-

blatt, 1990). In the continuous-frequency case,

$$\begin{aligned} f_{\underline{a}}^{(W,n)}(\underline{\omega}) &= \int_{[-\pi/\Delta, \pi/\Delta]^{k-1}} W_n(\underline{\omega} - \underline{\lambda}) I_{\underline{a}}(\underline{\lambda}) d\underline{\lambda} \\ &= \frac{\Delta}{2\pi} \sum_{\underline{\tau} \in \Delta \mathbb{Z}^{k-1}} w_n(\underline{\tau}) \hat{c}(\underline{\tau}) \end{aligned}$$

where $W_n(\underline{\omega}) = W(\underline{\omega}/b_n)/(b_n)^{k-1}$ and $w_n(\underline{\tau}) = w(b_n \underline{\tau})$ and $\hat{c}(\underline{\tau})$ is the sample k^{th} -order cumulant function at lag $\underline{\tau}$, which is the Fourier transform of the k^{th} -order periodogram.

When $b_n \rightarrow 0$ and $b_n^{k-1} n \rightarrow \infty$ as $n \rightarrow \infty$, such estimates have been shown to be consistent and asymptotically normal under certain assumptions (Brillinger and Rosenblatt, 1967a; Lii and Rosenblatt, 1990). Approaches of this type are often referred to as indirect (Nikias, 1993).

Multitapering

Multitapering was first introduced for estimating the second-order spectral density function by Thomson (1982). Thomson (1989) introduced multitapering for bispectra, which was further investigated by Birkelund and Hanssen (1999); He and Thomson (2009). Birkelund et al. (2003) developed multitapering in the general case for arbitrary order spectral density functions. Multitapering averages over a collection of tapered spectral estimates with different tapers. Say that we have a collection of K orthogonal tapers $\underline{v}_1, \dots, \underline{v}_K$ and let M denote the set of all k -tuples with elements in $\{1, \dots, K\}$, i.e.

$$M = \{1, \dots, K\}^k.$$

The multitaper estimate is

$$f_{a_1 \dots a_k}^{(M,n)}(\omega_1, \dots, \omega_{k-1}) = \frac{1}{U} \sum_{\underline{m} \in M} P(\underline{m}) \prod_{j=1}^k J_{v_{m_j}}^{(n)}(\omega_j)$$

where $P(\underline{m}) = A(\underline{m}) \sum_{t=1}^n \prod_{j=1}^k v_{m_j}(t)$ and $U = \sum_{\underline{m} \in M} P(\underline{m})^2 A(\underline{m})$. Additionally, $A(\underline{m}) = \prod_{j=1}^k a_{m_j}$ where the values of a_j are weightings, often chosen based on the eigenvalues of the respective tapers. In the case of $k = 2$, for orthogonal tapers, the sum collapses to just a sum of K tapered estimates as in Thomson (1982). Usually the dpss (discrete prolate spheroidal sequence) tapers are used, see Birkelund et al. (2003) for more details.

5.3.3 Practical considerations

From a practical standpoint it is worth noting some computational aspects of the estimators described in this section. Firstly, for a periodogram of a time series of length n , we must compute the discrete Fourier transform of the time series, an $O(n \log n)$ operation, and then compute the periodogram at each frequency. Ignoring symmetries, there are n^{k-1} Fourier frequencies, making this process $O(n^{k-1})$. Additionally, we have a memory cost of $O(n^{k-1})$. For $k > 3$, this very quickly becomes an issue, in fact, if we have $n = 1000$ and $k = 4$, since the periodogram is complex valued, if we used 64 bit floating point numbers, we would require 16GB of memory to store the periodogram at all of the Fourier frequencies. Of course, there are symmetries which reduce this cost, but we often have n much larger than 1000. If we want to perform bootstrap tests, as in Section 5.5 and Chapter 6, this could quickly become an issue.

Furthermore, note that for the multivariate case, we have d^k different cross higher-order spectra to compute. So a memory and computational cost of $O(d^k n^{k-1})$. One

way to avoid some of the memory problems is to use segment averaging. Alternatively, we can just evaluate the spectra at some subset of the Fourier frequencies, typically based on the bandwidth of the smoothing being used. Such an approach is preferable not only because it reduces the memory and computational costs, but also because estimates are then less correlated with each other, an approach used in tests for Gaussianity and linearity (Hinich, 1982; Wong, 1993, for example). This is described in more detail in the third-order case in Chapter 6.

5.4 Normalisation

The higher-order spectra alone can be hard to interpret. Even for Gaussian processes, the variance of estimators of the higher-order spectra depends on the second-order spectral density function. Additionally, the higher-order spectra of linear processes are not necessarily constant, but all of the structure comes from the linear filter. Therefore, it is often useful to normalise higher-order spectra. There are two competing normalisations in the literature, the older normalising due to Brillinger (1965) and the alternate normalisation proposed by Kim and Powers (1979), designed to be bounded between zero and one. Hinich et al. (2005) points out the differences, and describes why the former is preferable over the latter. Here we will describe these normalisations and detail the objection of Hinich et al. (2005) to the latter normalisation. We will also describe a general class of normalisations which are invariant to linear filtering.

In this section, we will use an augmented notation, writing

$$f_{X_{a_1} \dots X_{a_k}}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) = f_{a_1 \dots a_k}^{(\Delta)}(\omega_1, \dots, \omega_{k-1})$$

for all $\omega_1, \dots, \omega_{k-1} \in [-\pi/\Delta, \pi/\Delta]$. When all of the series are the same, we will use

the simpler notation $f_X^{(\Delta)}$, where the order will be clear from the number of arguments. From now on, we will focus solely on discrete-time processes.

5.4.1 The third-order case

We shall begin by considering the simple case of the bispectrum of a univariate discrete-time process X_Δ . If X_Δ is linear, i.e. a time-invariant linear filter of a pure white noise process ϵ (with transfer function h), from (Brillinger, 1965) we have that

$$\begin{aligned} f_X^{(\Delta)}(\omega) &= \frac{\kappa_2}{(2\pi/\Delta)} h(\omega)h(-\omega) \\ &= \frac{\sigma^2}{(2\pi/\Delta)} |h(\omega)|^2, \\ f_X^{(\Delta)}(\omega_1, \omega_2) &= \frac{\kappa_3}{(2\pi/\Delta)^2} h(\omega_1)h(\omega_2)h(-\omega_1 - \omega_2) \\ &= \frac{\kappa_3}{(2\pi/\Delta)^2} h(\omega_1)h(\omega_2)\overline{h(\omega_1 + \omega_2)}, \end{aligned}$$

where $\kappa_2 = \sigma^2 = \text{var}(\epsilon(t))$ and $\kappa_3 = \mathcal{C}_3[\epsilon] = \mathbb{E}[\epsilon^3]$.

Motivated by this, the squared skewness function is defined to be

$$\Gamma_X(\omega_1, \omega_2) = \frac{|f_X^{(\Delta)}(\omega_1, \omega_2)|^2}{f_X^{(\Delta)}(\omega_1)f_X^{(\Delta)}(\omega_2)f_X^{(\Delta)}(\omega_3)}.$$

It is often claimed (Hinich, 1982; Birkelund and Hanssen, 2009; Berg et al., 2010, for example), that if a process is linear, then

$$\begin{aligned} \Gamma_X(\omega_1, \omega_2) &= \frac{[\kappa_3^2/(2\pi/\Delta)^4] |h(\omega_1)h(\omega_2)h(-\omega_1 - \omega_2)|^2}{[\sigma^6/(2\pi/\Delta)^3] |h(\omega_1)|^2 |h(\omega_2)|^2 |h(-\omega_1 - \omega_2)|^2} \\ &= \frac{1}{(2\pi/\Delta)} \frac{|\kappa_3|^2}{\sigma^6}. \end{aligned}$$

Meaning that the squared skewness function is constant for a linear process. How-

ever, this ignores the possibility that the spectral density function is zero at some frequencies. In this case, the squared skewness function is not defined at those frequencies. This happens precisely when the transfer function is zero. If the transfer function of a linear filter is never zero, then it is said to be *invertible*. Therefore, we shall refer to processes obtained by an invertible linear filter of an i.i.d. process as invertible linear processes.

It is fairly easy to construct filters which are not invertible. In particular, consider the filter where $a(0) = 1$ and $a(\Delta) = 1$ but $a(t) = 0$ for all other $t \in \Delta\mathbb{Z}$. The transfer function of this filter is given by

$$h(\omega) = e^{-i\omega\Delta} + 1$$

for all $\omega \in [-\pi/\Delta, \pi/\Delta]$. Clearly this is zero when $\omega = \pm\pi/\Delta$. In fact, any filter of the form

$$a(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq m, \\ 0 & \text{otherwise} \end{cases}$$

where $m \geq 1$ will have a transfer function which is zero at certain frequencies. This result has an interesting interpretation. By the central limit theorem, such averaging should produce a process which is “more Gaussian” than the original noise. Recall that the bispectrum of a Gaussian process is zero, and in the case of these filters, the resulting process has a bispectrum which is zeroed at certain frequencies. The more points we include in the average, the more often the transfer function hits zero, and the more of the bispectrum is zero.¹⁰

The square root of Γ_X is often referred to as the skewness function¹¹, and so we will

¹⁰This holds for spectra of all orders.

¹¹Some authors (Kim and Powers, 1978) refer to this as bicoherence, though, as pointed out by

refer to Γ_X as the squared skewness function. Note that this is not the skewness of a univariate random variable, but rather, can be heuristically interpreted as the cross skewness of the associated spectral process.

Kim and Powers (1979) suggest an alternate normalisation, usually referred to as the bicoherence. In particular, they define

$$\bar{\Gamma}_X(\omega_1, \omega_2) = \frac{|\mathbb{E} [\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)\mathcal{X}(\omega_3)]|^2}{\mathbb{E} [|\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)|^2] \mathbb{E} [|\mathcal{X}(\omega_3)|^2]}$$

where \mathcal{X} is the discrete Fourier transform of the stochastic process X for some long, but finite time (He and Thomson, 2009).¹² The Cauchy-Schwartz inequality gives that $\bar{\Gamma}(\omega_1, \omega_2) \leq 1$. However, as pointed out by Hinich et al. (2005), this is misleading. The bicoherence (and its generalisation to arbitrary k) is still widely used, and therefore, we will briefly reiterate the argument of Hinich and Wolinsky (2005).

As we mentioned, \mathcal{X} is defined to be the discrete Fourier transform of X for some long period of time, say N . In the notation of the previous section,

$$\mathcal{X}(\omega) = J^{(N)}(\omega).$$

The argument of Hinich and Wolinsky (2005) is as follows. From the results of Brillinger and Rosenblatt (1967a) that if we avoid the proper submanifolds, as $N \rightarrow \infty$ we have

$$\begin{aligned} \mathbb{E} [|\mathcal{X}(\omega_3)|^2] &= Nf_X^{(\Delta)}(\omega_3) + O(1) \\ \mathbb{E} [\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)\mathcal{X}(\omega_3)] &= Nf_X^{(\Delta)}(\omega_1, \omega_2) + O(1). \end{aligned}$$

Hinich and Wolinsky (2005), this is misleading as the squared skewness function is not bounded above by one. Note we use bicoherence to refer to the quantity defined by Kim and Powers (1979).

¹²Again, this is not always defined, though it is for invertible linear processes.

The final term in the denominator is

$$\begin{aligned}
\mathbb{E} \left[|\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)|^2 \right] &= \mathbb{E} [\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)\mathcal{X}(-\omega_1)\mathcal{X}(-\omega_2)] \\
&= \mathcal{C} [\mathcal{X}(\omega_1), \mathcal{X}(\omega_2), \mathcal{X}(-\omega_1), \mathcal{X}(-\omega_2)] \\
&\quad + \mathbb{E} \left[|\mathcal{X}(\omega_1)|^2 \right] \mathbb{E} \left[|\mathcal{X}(\omega_2)|^2 \right] + O(1) \\
&= Nf_X^{(\Delta)}(\omega_1, \omega_2, -\omega_1) \\
&\quad + [Nf_X^{(\Delta)}(\omega_1) + O(1)][Nf_X^{(\Delta)}(\omega_2) + O(1)] + O(1) \\
&= Nf_X^{(\Delta)}(\omega_1, \omega_2, -\omega_1) + N^2 f_X^{(\Delta)}(\omega_1)f_X^{(\Delta)}(\omega_2) + O(N).
\end{aligned}$$

Putting this together we have

$$\begin{aligned}
\bar{\Gamma}(\omega_1, \omega_2) &= \frac{|\mathbb{E} [\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)\mathcal{X}(\omega_3)]|^2}{\mathbb{E} [|\mathcal{X}(\omega_1)\mathcal{X}(\omega_2)|^2] \mathbb{E} [|\mathcal{X}(\omega_3)|^2]} \\
&= \frac{f_X^{(\Delta)}(\omega_1, \omega_2) + O(1/N)}{f_X^{(\Delta)}(\omega_1, \omega_2, -\omega_1)f_X^{(\Delta)}(\omega_3) + Nf_X^{(\Delta)}(\omega_1)f_X^{(\Delta)}(\omega_2)f_X^{(\Delta)}(\omega_3) + O(1)}.
\end{aligned}$$

So as $N \rightarrow \infty$, this tends to zero. As a result, this statistic is artificially bounded above by one, and will likely be misleading (Hinich and Wolinsky, 2005).

5.4.2 The general case

In the third-order case, one of the motivations for considering the squared skewness function is that it is constant for an invertible linear processes. In fact, the squared skewness function is invariant to invertible linear filters. In general, it is possible to construct a broader class of normalisations for which this is true. Consider

$$\Gamma_X^{(P)}(\omega_1, \dots, \omega_k) = \frac{|f_X^{(\Delta)}(\omega_1, \dots, \omega_{k-1})|^2}{\prod_{A \in P} \tilde{f}_X^{(\Delta)}(\underline{\omega}_A, -\underline{\omega}_A)} \quad (5.4.1)$$

where P is some partition of $\{1, \dots, k\}$ and

$$\tilde{f}_X^{(\Delta)}(\underline{\omega}_A, -\underline{\omega}_A) = f_X^{(\Delta)}(\omega_{A_1}, \dots, \omega_{A_r}, -\omega_1, \dots, -\omega_{A_{r-1}}),$$

with $r = |A|$. Essentially, the standard normalisation corresponds to the largest partition i.e. with $|P| = k$. A normalisation similar to, but not the same as, the Kim and Powers normalisation corresponds to partitions with $|P| = 2$ (of which there are k choices).

Brillinger (1965) show that if Y arises from a linear filter of X then

$$f_Y^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) = f_X^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) \prod_{j=1}^k h(\omega_j).$$

Therefore, assuming that the linear filter is invertible, and that $f_X^{(\Delta)}(\omega) > 0$ for all ω , we have

$$\begin{aligned} \Gamma_Y^{(P)}(\omega_1, \dots, \omega_k) &= \frac{|f_Y^{(\Delta)}(\omega_1, \dots, \omega_{k-1})|^2}{\prod_{A \in P} \tilde{f}_Y^{(\Delta)}(\underline{\omega}_A, -\underline{\omega}_A)} \\ &= \frac{|f_X^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) \prod_{j=1}^k h(\omega_j)|^2}{\prod_{A \in P} \tilde{f}_X^{(\Delta)}(\underline{\omega}_A, -\underline{\omega}_A) \prod_{j \in A} h(\omega_j) h(-\omega_j)} \\ &= \frac{|f_X^{(\Delta)}(\omega_1, \dots, \omega_{k-1})|^2}{\prod_{A \in P} \tilde{f}_X^{(\Delta)}(\underline{\omega}_A, -\underline{\omega}_A)} \\ &= \Gamma_X^{(P)}(\omega_1, \dots, \omega_k) \end{aligned}$$

where the cancellation of the transfer functions follows from P being a partition of $\{1, \dots, k\}$ and conjugate symmetry of the transfer function. Thus these normalisations are all invariant to invertible linear filters.

Since an invertible linear process is the result of applying an invertible linear filter to a pure white noise process, and a pure white noise process has constant spectra for

all k , linearly-invariant normalisations of the form (5.4.1) are constant for invertible linear processes. In particular, for a pure white noise process $\epsilon = \{\epsilon(t)\}_{t \in \Delta\mathbb{Z}}$, letting $\kappa_k = \mathcal{C}_k[\epsilon(t)]$ we have

$$f_\epsilon^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) = \frac{\kappa_k}{(2\pi/\Delta)^{k-1}}$$

and therefore

$$\begin{aligned} \Gamma_\epsilon(\omega_1, \dots, \omega_{k-1}) &= \frac{\kappa_k}{(2\pi/\Delta)^{k-1}} \prod_{A \in \mathcal{P}} \frac{(2\pi/\Delta)^{2|A|-1}}{\kappa_{|A|}} \\ &= \left| \frac{\kappa_k}{\prod_{A \in \mathcal{P}} \kappa_{|A|}} \right|^2 \left(\frac{2\pi}{\Delta} \right)^{2-|P|}. \end{aligned}$$

Note that the only form of (5.4.1) which avoids normalising by a higher-order spectral density on a proper submanifold is when $|P| = k$. Additionally, by normalising by the second order spectral density function ($|P| = k$), we standardise the variance of most non-parametric estimators. For these reasons, we prefer the statistical normalisation

$$\Gamma_X(\omega_1, \dots, \omega_{k-1}) = \frac{|f_X^{(\Delta)}(\omega_1, \dots, \omega_{k-1})|^2}{\prod_{j=1}^k f_X^{(\Delta)}(\omega_j)}.$$

In particular, in the third-order case, we use the squared skewness function.

5.4.3 Multivariate squared skewness

Consider a d -dimensional real-valued mean-zero stationary process denoted by $\underline{\mathbf{X}} = \{\underline{\mathbf{X}}(t)\}_{t \in \Delta\mathbb{Z}}$. Hannan (2009) define the process to be linear if, for all $t \in \Delta\mathbb{Z}$,

$$\underline{\mathbf{X}}(t) = \sum_{j \in \Delta\mathbb{Z}} \mathbf{A}(j) \epsilon(t-j)$$

where $\{\underline{\epsilon}(t)\}_{t \in \Delta\mathbb{Z}}$ is i.i.d. and A_j are matrices satisfying $\sum_{j \in \mathbb{Z}} \|A_j\|^2 < \infty$ (where $\|A\|$ is the largest singular value of the matrix A). As in the univariate case, the process is a linear filter of an i.i.d. process, with impulse response sequence $A : \Delta\mathbb{Z} \rightarrow \mathbb{R}^{d \times d}$ and transfer Function $H : [-\pi/\Delta, \pi/\Delta] \rightarrow \mathbb{C}^{d \times d}$ such that

$$H(\omega) = \frac{\Delta}{2\pi} \sum_{t \in \Delta\mathbb{Z}} A(t) e^{-it\omega}.$$

Whilst we could consider the cross squared skewness, Wong (1997) point out that this would not be constant for processes that are linear in the multivariate sense. For this reason, they propose a multivariate version of the squared skewness function, which is constant for linear processes. This is then used in a test for non-Gaussianity and non-linearity (see Section 5.5). As in the univariate case, this is only true for invertible linear filters. In the multivariate case, a linear filter is said to be invertible if the the transfer function, which is matrix valued, is invertible at every frequency (Hannan, 2009).

We now need the concept of the cumulant vector sequence. For a full definition, see Jammalamadaka et al. (2006); however, for our purposes we use the definition of Wong (1997). In particular, the k^{th} -order vector cumulant sequence is denoted by $\underline{c}_{\underline{X}}^{(\Delta)} : \Delta\mathbb{Z}^{k-1} \rightarrow \mathbb{R}^{d^k}$ such that

$$\underline{c}_{\underline{X}}^{(\Delta)}(\tau_1, \dots, \tau_{k-1}) = [c_{X_{a_1} \dots X_{a_k}}^{(\Delta)}(\tau_1, \dots, \tau_{k-1})]_{a_1, \dots, a_k},$$

where the indices are ordered in the same way as a Kronecker product of k vectors, each of length d . The k^{th} -order spectral density vector function $\underline{f}_{\underline{X}}^{(\Delta)} : \mathbb{R}^{k-1} \rightarrow \mathbb{C}^{d^k}$, is the Fourier transform of $\underline{c}_{\underline{X}}^{(\Delta)}$.

Note that in the second-order case, this is different to the spectral density matrix

function, $\underline{f}_{\underline{X}}^{(\Delta)}$, encountered in Chapter 3. We make the distinction by underlining vector valued functions (and vectors). The spectral density vector function and spectral density matrix function are related by

$$\underline{f}_{\underline{X}}^{(\Delta)}(\omega) = \text{vec}(\underline{f}_{\underline{X}}^{(\Delta)}(\omega)^T)$$

where vec denotes the operator that stacks a matrix by its columns. The transpose is required because the Kronecker product stacks things differently to the vec operator (see Graham, 1981).

If the process is multivariate linear, then Wong (1997) show that

$$\underline{f}_{\underline{X}}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}) = \frac{\Delta^{k-1}}{(2\pi)^{k-1}} \underline{\kappa}_k \prod_{j=1}^{\otimes k} \mathbf{H}(\omega_j)$$

where $\prod_{j=1}^{\otimes k} \mathbf{H}(\omega_j) = \mathbf{H}(\omega_1) \otimes \dots \otimes \mathbf{H}(\omega_k)$ and

$$\underline{f}_{\underline{X}}(\omega) = \frac{\Delta}{(2\pi)} \underline{\Sigma} \mathbf{H}(\omega_j)^{\otimes 2}$$

with $\mathbf{H}(\omega)^{\otimes 2} = \mathbf{H}(\omega) \otimes \mathbf{H}(\omega)$.

Wong (1997) defines the normalisation $\Gamma_{\underline{X}} : \Delta \mathbb{Z}^{k-1} \rightarrow \mathbb{R}$ such that

$$\Gamma_{\underline{X}}(\omega_1, \dots, \omega_{k-1}) = \underline{f}_{\underline{X}}^{(\Delta)}(\omega_1, \dots, \omega_{k-1})^H \left(\prod_{k=1}^{\otimes k} \underline{f}_{\underline{X}}^{(\Delta)}(\omega_j) \right)^{-1} \underline{f}_{\underline{X}}^{(\Delta)}(\omega_1, \dots, \omega_{k-1}).$$

They show that, if the process is linear we have

$$\Gamma_{\underline{X}}(\omega_1, \dots, \omega_{k-1}) = \frac{1}{(2\pi)^{k-2}} \underline{\kappa}_k^H (\underline{\Sigma}^{\otimes k})^{-1} \underline{\kappa}_k,$$

where

$$\begin{aligned}\underline{\kappa}_k &= \mathcal{C}[\underbrace{\epsilon(t), \dots, \epsilon(t)}_{k \text{ times}}] \\ \Sigma &= \mathbb{E} [\underline{\epsilon}(t)\underline{\epsilon}(t)^H] - \mathbb{E} [\underline{\epsilon}(t)] \mathbb{E} [\underline{\epsilon}(t)]^H\end{aligned}$$

so that $\text{vec}(\Sigma^T) = \underline{\kappa}_2$. However, as in the univariate case, their proof only holds if the process is also invertible (the squared skewness function is not defined at some frequencies when it isn't). Again, this result can be seen in the slightly more general light as an invariance to invertible linear filtering (assuming of course that the spectral density matrix function of the process being filtered is itself invertible at all frequencies).

5.5 Global tests for linearity and Gaussianity

In what follows, when we refer to a linear process, we will mean that the process is invertible, i.e. it has a positive spectral density function (positive definite in the multivariate case). We also drop the dependence on X in many quantities, as the difference between univariate and multivariate will either not matter, or be clear from context. [Hinich \(1982\)](#) suggested tests for Gaussianity and linearity based on the asymptotic distribution of the skewness function under the respective nulls (an approach which has been used by [Cherneva and Soares \(2007\)](#) to test for non-linearity in ocean wave records). [Wong \(1997\)](#) then extended this to the multivariate third-order case, and [Dalle Molle and Hinich \(1995\)](#) to the fourth-order univariate case. Testing is based on the asymptotic distribution, under the null, of

$$\hat{R}_n(\omega_1, \omega_2) = 2\pi 2C_n n^{-1} \hat{\Gamma}_n(\omega_1, \omega_2),$$

where $\hat{\Gamma}_n(\omega_1, \omega_2)$ is an estimator of $\Gamma(\omega_1, \omega_2)$ (typically a plug in estimator) and C_n is a normalising constant depending on the method used to estimate the various spectra required for computing $\hat{\Gamma}_n$. [Hinich \(1982\)](#) in the univariate case and [Wong \(1997\)](#) in the multivariate case show that for large n , $\hat{R}_n(\omega_1, \omega_2)$ is approximately $\chi_{2d^3}^2(\lambda_n(\omega_1, \omega_2))$, where $\lambda_n(\omega_1, \omega_2) = 2\pi 2C_n n^{-1} \Gamma(\omega_1, \omega_2)$, provided $n/C_n \rightarrow 0$ as $n \rightarrow \infty$.

Gaussianity

Under the null hypothesis of Gaussianity, $\lambda(\omega_1, \omega_2) = 0$ for all $\omega_1, \omega_2 \in \Omega_{PD}$. Therefore [Hinich \(1982\)](#) suggest the test statistic

$$D_G = \sum_{(\omega_1, \omega_2) \in \Omega} \hat{R}_n(\omega_1, \omega_2),$$

where $\Omega \subset \Omega_{PD}$. In the univariate case, under the null assumption of Gaussianity, D_G is approximately $\chi_{2|\Omega|}^2$ ([Hinich, 1982](#)). Or more generally, in the multivariate case, D_G is approximately $\chi_{2d^3|\Omega|}^2$ as $n \rightarrow \infty$ ([Wong, 1997](#)), assuming the process is Gaussian. Therefore, the null hypothesis of Gaussianity is rejected at the q significance level if $P(D_G > \chi_{2d^3|\Omega|}^2) \leq q$.

Linearity

Under the null of a linear, but not necessarily Gaussian process, ([Hinich, 1982](#)) and [Wong \(1997\)](#) show that for large n , $\hat{R}_n(\omega_1, \omega_2)$ is approximately $\chi_{2d^3|\Omega|}^2(\lambda_0)$, where $\lambda_0 = 2C_n n^{-1} \underline{\kappa}_3^H (\underline{\Sigma}^{\otimes 3})^{-1} \underline{\kappa}_3$, in particular, the noncentrality parameter is constant. Since λ_0 is unknown, [Hinich \(1982\)](#) suggest estimating it with the method of moments estimator

$$\tilde{\lambda}_0 = \frac{1}{|\Omega|} \sum_{(\omega_1, \omega_2) \in \Omega} \hat{R}_n(\omega_1, \omega_2) - d^3.$$

This could be replaced by the maximum likelihood estimator

$$\hat{\lambda}_0 = \operatorname{argmax}_{\lambda \in \mathbb{R}} \sum_{(\omega_1, \omega_2) \in \Omega} \log f_R(\hat{R}_n(\omega_1, \omega_2); 2d^3, \lambda)$$

where

$$f_R(x; \nu, \lambda) = \frac{1}{2} e^{-(x+\lambda)/2} \left(\frac{x}{\lambda}\right)^{\nu/4-1/2} I_{\nu/2-1}(\sqrt{\lambda x})$$

is the pdf of a noncentral chi-squared distribution with degree of freedom ν and noncentrality parameter λ (and I_k is a modified Bessel function of the first kind), but we have found little difference in simulations. If the process is not linear, then the noncentrality parameter may not be constant. In this case, there should be more density in the tails of $\hat{R}_n(\omega_1, \omega_2)$ than in a $\chi_{2d^3}^2(\hat{\lambda}_0)$ distribution. To test this, [Hinich \(1982\)](#) suggest using the inter quartile range of $\hat{\mathbf{R}} = [\hat{R}_n(\omega_1, \omega_2)]_{(\omega_1, \omega_2) \in \Omega}$, whilst [Wong \(1997\)](#) suggest the 80% inter quantile range due to the results from [Ashley et al. \(1986\)](#). We shall refer to these test statistics as $D_L^{(IQR)}$ and $D_L^{(80)}$ for the interquartile range and 80% interquantile range respectively.

More formally, let $\hat{R}_{(j)}$ denote the j^{th} order statistic of $\hat{\mathbf{R}}$. Then the y^{th} sample quantile of $\hat{\mathbf{R}}$ is $\hat{R}_{(\lfloor y|\Omega| \rfloor + 1)}$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x . Assume that the null of linearity is true (so that each element of $\hat{\mathbf{R}}$ is approximately $\chi_{2d^3}^2(\hat{\lambda}_0)$ distributed) and write $\tilde{\zeta}_y = F_R^{-1}(y)$ where F_R^{-1} is the inverse CDF of a $\chi_{2d^3}^2(\hat{\lambda}_0)$ distribution. Then by Theorem 10.3 of [David and Nagaraja \(2004\)](#), we have for arbitrary quantiles $0 < y_1 < y_2 < 1$ the quantile range $D_L(y_1, y_2) = \hat{R}_{(\lfloor y_2|\Omega| \rfloor + 1)} - \hat{R}_{(\lfloor y_1|\Omega| \rfloor + 1)}$ satisfies

$$|\Omega|^{1/2} [D_L(y_1, y_2) - (\tilde{\zeta}_{y_2} - \tilde{\zeta}_{y_1})] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(y_1, y_2))$$

as $|\Omega| \rightarrow \infty$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution and

$$\sigma^2(y_1, y_2) = \frac{y_2(1-y_2)}{f_R(\xi_{y_2})^2} + \frac{y_1(1-y_1)}{f_R(\xi_{y_1})^2} - 2 \frac{y_1(1-y_2)}{f_R(\xi_{y_2})f_R(\xi_{y_1})}$$

and f_R is the pdf of a $\chi_{2d^3}^2(\hat{\lambda}_0)$. In the case of the inter quartile range, $y_1 = 0.25$ and $y_2 = 0.75$, so we have (Hinich, 1982)

$$\sigma^2(0.25, 0.75) = \frac{1}{16} \left(\frac{3}{f_R(\xi_{0.75})^2} + \frac{3}{f_R(\xi_{0.25})^2} - \frac{2}{f_R(\xi_{0.75})f_R(\xi_{0.25})} \right).$$

For the 80% quantile range, we have $y_1 = 0.1$ and $y_2 = 0.9$ and thus

$$\sigma^2(0.1, 0.9) = \frac{1}{100} \left(\frac{9}{f_R(\xi_{0.9})^2} + \frac{9}{f_R(\xi_{0.1})^2} - \frac{2}{f_R(\xi_{0.9})f_R(\xi_{0.1})} \right).$$

Note that this is different to the result stated by Wong (1997), who claim

$$\sigma^2(0.1, 0.9) = \frac{1}{16} \left(\frac{3}{f_R(\xi_{0.9})^2} + \frac{3}{f_R(\xi_{0.1})^2} - \frac{2}{f_R(\xi_{0.9})f_R(\xi_{0.1})} \right)$$

which is not correct. In any case, given a quantile range with lower quantile y_1 and upper quantile y_2 we reject the null hypothesis at the q significance level if

$$P \left(D_L(y_1, y_2) > \mathcal{N}(\xi_{y_2} - \xi_{y_1}, \sigma^2(y_1, y_2)/|\Omega|) \right) \leq q.$$

The test statistics used by Hinich (1982) and Wong (1997) are $D_L^{(IQR)} = D_L(0.25, 0.75)$ and $D_L^{(80)} = D_L(0.1, 0.9)$ respectively.

5.5.1 Univariate bootstrap tests

There are many cases in which the finite sample distribution of the aforementioned test statistics deviates greatly from the asymptotic distribution (e.g. Birkelund and Hanssen, 2009; Berg et al., 2010). In the univariate case, Birkelund and Hanssen

(2009) address this problem by using the method of surrogate data to generate series under the null hypothesis, in order to better approximate the distribution of the test statistics. Additionally, they suggest two new test statistics

$$\begin{aligned}\tilde{D}_G &= \frac{1}{|\Omega|} \sum_{(\omega_1, \omega_2) \in \Omega} \hat{R}_n(\omega_1, \omega_2) \\ \tilde{D}_L &= \frac{1}{|\Omega|} \sum_{(\omega_1, \omega_2) \in \Omega} (\hat{R}_n(\omega_1, \omega_2) - \tilde{D}_G)^2\end{aligned}$$

which are sample mean and (biased) sample variance of \hat{R} respectively.

Under the Gaussian null hypothesis, Birkelund and Hanssen (2009) suggest using the randomised phase approach of Theiler et al. (1992) to generate surrogate time series, namely

$$X^*(t) = \frac{1}{n} \sum_{j=1}^n d_X^{(n)}(\omega_j) e^{i\theta_j} e^{it\omega_j}$$

where ω_j is the j^{th} Fourier frequency and $\theta_j \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, 2\pi)$. In other words, compute the discrete Fourier transform of the series, scramble the phases and use the inverse discrete Fourier transform to return to the time domain.

To simulate under the null of linearity, Birkelund and Hanssen (2009) propose the linearly filtered non-Gaussian (LFNG) surrogate data generator. In particular, they propose first fitting an $\text{AR}(p)$ model to the time series, with parameter estimate $\hat{a}_1, \dots, \hat{a}_p$, and then bootstrapping the residuals. Birkelund and Hanssen (2009) claim that, because an AR model averages, by the central limit theorem, the simulated time series are ‘‘Gaussianified’’. To circumvent this, Birkelund and Hanssen (2009) propose adjusting the bootstrapped residuals as follows. Start with initial noise $\hat{\epsilon}_0(t)$

generated by bootstrapping the residuals of the AR model, i.e. bootstrapping

$$\hat{\epsilon}(t) = X(t) - \sum_{j=1}^p a_j X(t-j)$$

for $t \in \{p+1, \dots, n\}$. Adjust the noise iteratively by computing

$$\hat{\epsilon}_{j+1}(t) = \hat{\epsilon}_j^*(t) + X(t) - X_{(t)}^{(j)}$$

where $\hat{\epsilon}_j^*(t)$ is a bootstrap sample from $\{\hat{\epsilon}_j(t)\}_{t=p+1}^n$ and $X_{(t)}^{(j)}$ is ordered output of the AR filter of $\{\hat{\epsilon}_j(t)\}_{t=p+1}^n$. This procedure is terminated when $|\hat{\kappa}_3/\hat{\kappa}_3^{(j)} - 1| \leq \varepsilon$, where ε is some tolerance and $\hat{\kappa}_3$ and $\hat{\kappa}_3^{(j)}$ are estimates of the skewness of $\{X(t)\}$ and $\{X^{(j)}(t)\}$ respectively (which is not the squared skewness function of those processes).

Berg et al. (2010) suggest using the AR(∞) “seive” bootstrap (first introduced by Kreiss, 1988). This proceeds as follows:

1. Fit an AR(p) to the time series, for some large p .
2. Generate AR realisations from residuals drawn from a distribution which depends on the null in question:
 - Gaussian null: $\mathcal{N}(0, \hat{\sigma}^2)$ where $\hat{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^n (\hat{\epsilon}(t) - \bar{\epsilon})$ and $\bar{\epsilon} = \frac{1}{n-p} \sum_{t=p+1}^n \hat{\epsilon}(t)$.
 - Linear null: empirical distribution of $\epsilon(t) - \bar{\epsilon}$.
 - Linear null with symmetric errors (zero skewness): symmetrised empirical distribution of $\epsilon(t) - \bar{\epsilon}$ (achieved by multiplying a realisation from the empirical distribution by -1 or 1 based on a Bernoulli(0.5) random variable).

This is referred to as an $AR(\infty)$ bootstrap as the order of the $AR(p)$ is taken to infinity in theory. This is in contrast to the method proposed by Birkelund and Hanssen (2009), who choose the order of the AR model based on standard model comparison techniques (not taking p to infinity).

At this point, the reader may be wondering why the algorithm due to Birkelund and Hanssen (2009) involved a correction for “Gaussianification”, but that the $AR(\infty)$ approach of Berg et al. (2010) did not. The claim of Birkelund and Hanssen (2009) is that, due to the central limit theorem, a linear filter will make the filtered process more Gaussian than the original. Note however, that the Wold decomposition yields a plethora of counter examples, since it says that *any* stationary process is the result of applying a linear filter to a white noise process, which includes all non-Gaussian (and indeed non-linear) stationary discrete-time processes.

However, clearly some linear filters will make the resultant process more Gaussian, e.g. a filter which computes an unweighted average of a finite number of time points from the input process. At this point it helps to make the concept of “Gaussianification” more concrete. The property that concerns us here is that the result of applying the AR filter to the bootstrapped residuals (which we hope are white noise) would not change the squared skewness function. Provided the filter is invertible, then this will not happen. For our purpose, note that AR processes are stationary if and only if they are produced by an invertible filter (Grenander and Rosenblatt, 2008), and so when the AR filter is applied to noise the squared skewness function will not change. Thus the adjustment suggest by Birkelund and Hanssen (2009) is not necessary (as is confirmed by the results of Berg et al. (2010), where AR filtering is shown to work without such an adjustment).

5.6 Conclusion

In this chapter, we have reviewed higher-order versions of the spectral density function. We gave a detailed discussion of different normalisation choices, and recommendations for their use. Additionally, we reviewed a variety of estimation techniques. We finished with a discussion of global testing for Gaussianity and linearity of time series. However, in the case of ocean waves, the interest is not in global tests. Rather, the interest is in detecting frequency regions in which the higher-order spectra are non-zero. It is these local tests that we focus on in Chapter 6.

Chapter 6

Testing for non-Gaussianity in ocean wave records

In Chapter 5, we reviewed higher-order spectra in the general case, and discussed some global tests for Gaussianity and linearity (in the time series sense). From the perspective of ocean waves, the interest in detecting non-linearities in wave records usually corresponds to detecting regions of higher-order spectra which are non-zero. In the language of the tests seen in the previous section, the interest is in tests of non-Gaussianity. In contrast to the global tests discussed in Section 5.5, there is significant interest in the frequencies at which these (oceanographic) non-linearities occur.

A technique for detecting local non-linearities, based on generating surrogate background data, has been proposed by Schulte (2016). They propose using a Gaussian AR(1) process to generate a background level used for detecting non-linearities. However, the processes we are interested in are very different from AR(1) processes. To some extent, this is mitigated against because the test is for departures in the squared skewness function, which is invariant to the second-order spectral density

function of the process. Since this is all that distinguishes two (mean-zero) Gaussian processes, the choice should not matter. However, whilst this is true of the process, the finite sample properties of estimates of the squared skewness function will in general not be the same for two different Gaussian processes. As such, we propose generating the surrogate data using circulant embedding from non-parametric estimates of the spectral density function, a technique developed by Percival and Constantine (2006) for univariate time series and Chandna and Walden (2013) for multivariate time series.

At this point we should note that the squared skewness function is different from the marginal skewness of the process at a given point in time. In particular, if the squared skewness function is zero, then so is the skewness of the process at a given time point. However, the converse is not true (Raghuveer, 1994).

This chapter is structured as follows. Firstly, we review some definitions of third order spectra in Section 6.1. Then in Section 6.2 we introduce a family of simple processes with interesting higher-order spectra based on polynomials of two different Gaussian processes. Section 6.3 presents simulation studies exploring the performance of the proposed technique. An application to wave data recorded at Humboldt Bay in California is then presented in Section 6.4. A discussion and concluding remarks are then made in Section 6.5.

6.1 Background

In this section, we briefly review some necessary background literature. Chapter 5 provided a more thorough overview of these concepts, but we review some of them here for convenience. Note that we restrict our attention to the discrete-time case.

6.1.1 The bispectrum

Let $X = \{X(t)\}_{t \in \Delta\mathbb{Z}}$ be a discrete-time strictly-stationary real-valued stochastic process. The autocovariance (second-order cumulant function) and third-order cumulant function are defined as

$$c_X(\tau) = \text{cov}(X(\tau), X(0)),$$

$$c_X(\tau_1, \tau_2) = \mathcal{C}[X(\tau_1), X(\tau_2), X(0)],$$

respectively, where \mathcal{C} denotes the joint cumulant.¹ If the process is mean zero, then the third-order cumulant function is equal to the third-order moment function. Assuming that the second- and third-order cumulant functions are absolutely summable, the second-order cumulant spectral density function (or power spectrum) is the Fourier transform of the autocovariance, and the third-order cumulant spectral density function (often referred to as the bispectrum) is the Fourier transform of the third-order cumulant function. In other words, the power spectrum and bispectrum are

$$f_X^{(\Delta)}(\omega) = \sum_{\tau \in \Delta\mathbb{Z}} c_X(\tau) e^{-i\omega\tau},$$

$$f_X^{(\Delta)}(\omega_1, \omega_2) = \sum_{\tau_1 \in \Delta\mathbb{Z}} \sum_{\tau_2 \in \Delta\mathbb{Z}} c_X(\tau_1, \tau_2) e^{-i(\omega_1\tau_1 + \omega_2\tau_2)},$$

respectively. A stationary process has a spectral representation (Cramer and Leadbetter, 1967) so that

$$X(t) = \int_{-\pi/\Delta}^{\pi/\Delta} e^{it\omega} d\zeta_X^{(\Delta)}(\omega).$$

¹Note that for real valued processes, $\text{cov}(X(\tau), X(0)) = \mathcal{C}[X(\tau), X(0)]$, but for complex processes this is not the case. In particular, $\text{cov}(X(\tau), X(0)) = \mathcal{C}[X(\tau), \bar{X}(0)]$.

The associated spectral process $\zeta_X^{(\Delta)}$ can be related to the power spectrum and bispectrum by

$$\begin{aligned}\mathbb{E} \left[d\zeta_X^{(\Delta)}(\omega_1) d\zeta_X^{(\Delta)}(\omega_2) \right] &= \delta_\Delta(\omega_1 + \omega_2) f_X^{(\Delta)}(\omega_1) d\omega_1 d\omega_2, \\ \mathbb{E} \left[d\zeta_X^{(\Delta)}(\omega_1) d\zeta_X^{(\Delta)}(\omega_2) d\zeta_X^{(\Delta)}(\omega_3) \right] &= \delta_\Delta(\omega_1 + \omega_2 + \omega_3) f_X^{(\Delta)}(\omega_1, \omega_2) d\omega_1 d\omega_2 d\omega_3\end{aligned}$$

where $\delta_\Delta(x) = \sum_{k \in \mathbb{Z}} \delta(x + 2\pi k/\Delta)$ denotes the aliased Dirac delta function. Heuristically, this suggests that we can think about power spectra as describing the variance of the spectral process (which is uncorrelated for different frequencies) and the bispectra as describing the third-order joint cumulant between the spectral process at different frequencies.

6.1.2 Normalisation

The bispectrum can be difficult to interpret alone, as linear aspects of the process are often present in the bispectrum. Therefore, it is common to first normalise the bispectrum. Many authors (including Schulte, 2016) favour the use of bicoherence, introduced by Kim and Powers (1979). However, as we discussed in Section 5.4, we will use the squared skewness function. In the univariate case this is defined as

$$\Gamma_X(\omega_1, \omega_2) = \frac{|f_X^{(\Delta)}(\omega_1, \omega_2)|^2}{f_X^{(\Delta)}(\omega_1) f_X^{(\Delta)}(\omega_2) f_X^{(\Delta)}(\omega_3)}$$

where $\omega_3 = -\omega_1 - \omega_2 \pmod{2\pi/\Delta}$ such that $|\omega_3| < \pi/\Delta$ (see Chapter 5 for the multivariate definition). To summarise Section 5.4, the use of the squared skewness function over bicoherence, is for two main reasons. Firstly, estimates of the bicoherence tend to zero as $n \rightarrow \infty$ (Hinich and Wolinsky, 2005). Secondly, the bicoherence down-weights non-linearities in the bispectrum by non-linearities of higher-order, which can be misleading. Therefore, we prefer the squared skewness function.

6.1.3 Estimation

In Section 5.3, we described a variety of different estimation techniques for higher-order spectra, in particular bispectra. Whilst in theory these could be evaluated at any frequency pair, in practice we only evaluate them at some finite number of frequency pairs, which we shall call *evaluation frequencies*. In the case of the periodogram, a fairly natural choice is the Fourier frequencies. However, for smoothed estimates there are other considerations. In particular, because such smoothing correlates over frequency, we will often only estimate the bispectra at a subset of the Fourier frequencies, with a spacing determined by the bandwidth of the estimation technique in question. This is similar to the approach used in global testing (e.g. Hinich, 1982; Wong, 1997). For example, consider a kernel smoothed estimate of the bispectra using uniform rectangular windows (see Section 5.3.1). Then we choose evaluation frequencies by placing the appropriate sized squares over the grid of Fourier frequencies such that they are separated by a gap of one Fourier frequency, in other words, so that no periodogram ordinate is used in more than one estimate (avoiding additional correlation between estimates at different frequencies). We then only include frequencies for which the entire square is in the interior² of the principle domain

$$\Omega_{PD} = \{(\omega_1, \omega_2) \mid 0 \leq \omega_1 \leq \omega_2 \ \& \ \omega_1 + 2\omega_2 \leq 2\pi/\Delta\}.$$

The procedure is analogous for other kinds of estimates. We choose a bandwidth of $b_n = \lfloor n^{0.55}/2 \rfloor$, as this satisfies the consistency requirements, but does not lose too much resolution.

Figure 6.1.1 shows the evaluation frequencies for both rectangular and hexagonal smoothing when $n = 200$ and $b_n = 9$. The potential evaluation frequencies are shown as dots at the centre of the smoothing region, and are black if the frequency

²We follow Wong (1997) in not including the boundary as estimates of the squared skewness function on the boundary have a different distribution to the interior.

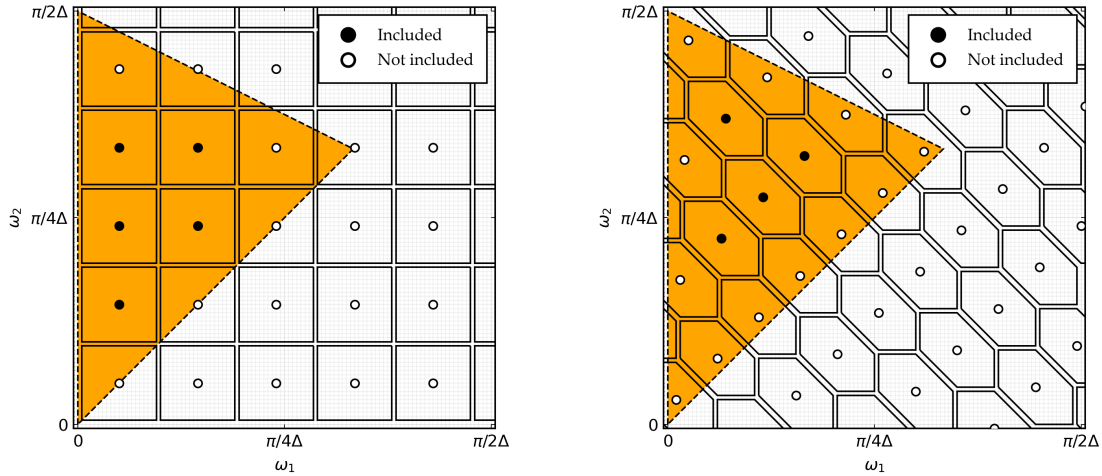


Figure 6.1.1: Evaluation frequencies for uniform smoothing over rectangles (left) and hexagons (right). Minor ticks are at each of the Fourier frequencies and the filled orange triangle is the principle domain.

pair is included and white if not. Solid lines indicate that the boundary is included in a region, whilst dashed lines indicate that the boundary is not included.

In the remainder of this chapter, we will use the uniform hexagonal smoothing kernel to estimate the bispectrum, as this is fast and has a very obvious definition of smoothing region, whilst improving over the rectangular kernel in that it smooths over a region which is more centred on the frequency pair of interest.

6.1.4 Local testing for non-Gaussianity

Schulte (2016) proposed an approach for determining significant deviations from zero in wavelet bicoherence (which is analogous to the Kim and Powers normalisation, but for wavelets). We will use the squared skewness function in this work, but the procedure is analogous. Their approach was to generate background red-noise processes (AR(1) processes) with the same covariance structure as the observed processes, i.e. the same lag-1 correlation³. Let N_r denote the number of replications used in the procedure. Estimates of the squared skewness function of these back-

³The same variance is not required as the squared skewness function is invariant to this (as is the bicoherence).

ground processes are used to compute p -values for the squared skewness function of the true process at each frequency pair. This can then be used to approximate the true distribution of such p -values under the null hypothesis that the process is Gaussian.

In general, for recorded time series $X(0), X(\Delta), \dots, X([n-1]\Delta)$ and some frequency pair ω_1, ω_2 ,

1. Estimate the squared skewness function from the original record, $\hat{\Gamma}_X(\omega_1, \omega_2)$.
2. For $j \in \{1, \dots, N_r\}$:
 - (a) Generate $\tilde{X}_j(0), \tilde{X}_j(\Delta), \dots, \tilde{X}_j([n-1]\Delta)$ from some Gaussian process \tilde{X} .
 - (b) Compute $\hat{\Gamma}_{\tilde{X}_j}(\omega_1, \omega_2)$.
3. Estimate the p -value by $\frac{1}{N_r} \sum_{j=1}^{N_r} \mathbb{1}_{\hat{\Gamma}_{\tilde{X}_j}(\omega_1, \omega_2) \leq \hat{\Gamma}_X(\omega_1, \omega_2)}$.

This procedure is then repeated for each desired frequency pair. In the version proposed by Schulte (2016), \tilde{X} is an AR(1) with lag-1 coefficient equal to the sample lag-1 correlation of the recorded time series. However, the processes we are interested in are not well represented by an AR(1) process. A sensible first choice would be to consider the AR(∞) bootstrap technique (Kreiss, 1988), which has already been used by Berg et al. (2010) in the case of global testing (see Section 5.5). However, whilst the AR model here is infinite in theory, clearly in practice a finite AR model must be used. To investigate the effect of this, we generate 10,000 time series with JONSWAP spectral density functions (with parameters $\alpha = 0.7$, $\omega_p = 0.8$, $\gamma = 3.3$, $r = 5$). For each time series we fit AR models of order 1, 10, 30 and 50. Figure 6.1.2 shows the average spectral density function of the residuals of such model fits. If the residuals were white noise, then the spectral density function would be flat. We can see that this is clearly not the case, even for the AR(50). The AR(∞) bootstrap assumes that

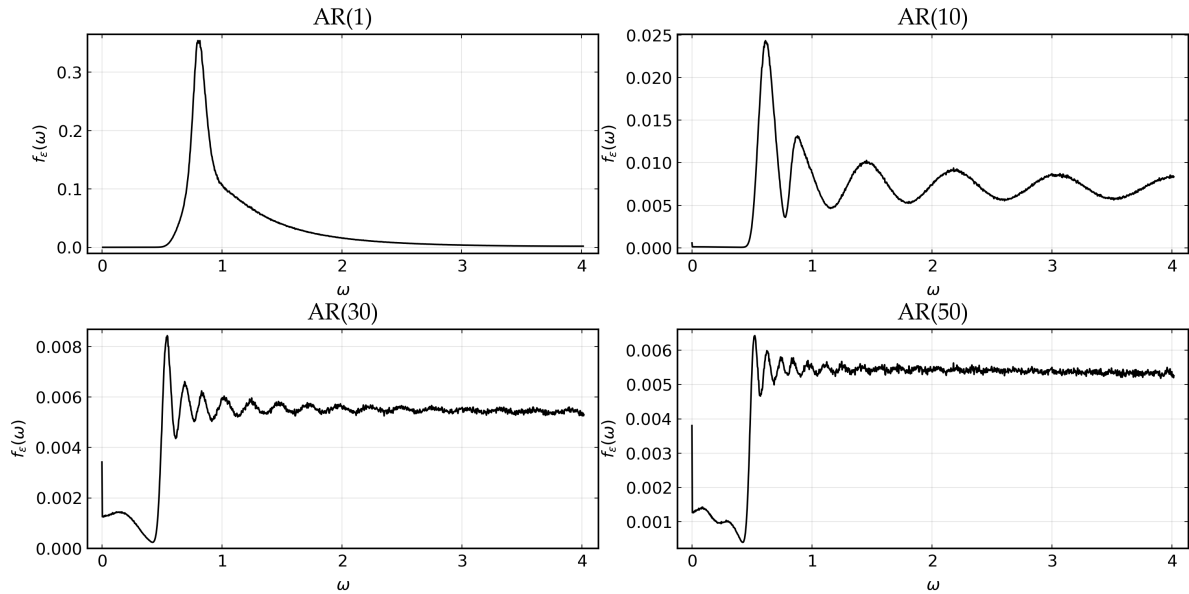


Figure 6.1.2: The spectral density function of the residuals of AR(p) processes fitted to Gaussian processes with a JONSWAP spectral density function ($\alpha = 0.7$, $\omega_p = 0.8$, $\gamma = 3.3$, $r = 5$).

the residuals are white noise, and so is unlikely to work well in this case.

Therefore, we propose replacing this with \tilde{X} , a Gaussian process where $f_{\tilde{X}}(\omega)$ is equal to an estimate of the spectral density function (typically a multitaper estimate). Such a process can be simulated efficiently with circulant embedding (Percival and Constantine, 2006; Chandna and Walden, 2013). We review circulant embedding in Appendix C.1. Note that this is slightly restrictive, as we can now only test the null hypothesis that the process is Gaussian (and not that it is linear). However, this is still useful in our context as the default assumption is that recordings of ocean waves are Gaussian. Furthermore, testing for linearity locally would be much more challenging, as we would need to test for deviations from some constant but non-zero level. The exception being processes with zero squared skewness functions that are not Gaussian. The methodology developed here will not be able to simulate under such a null hypothesis, but in our context this is not a major problem, though this is an avenue of possible further work.

Accounting for multiple testing

Schulte (2016) suggest accounting for multiple testing by controlling the false discovery rate (FDR). The false discovery rate is defined as the expected proportion of the rejected null hypotheses which were rejected in error, and was originally proposed by Benjamini and Hochberg (1995). In other words, the FDR is $\mathbb{E}[V/R]$ where V is the number of true null hypotheses which were rejected and R is the total number of rejected null hypotheses.⁴ Because the test statistics exhibit some dependence, Schulte (2016) suggest using the correction due to Benjamini and Yekutieli (2001) for data with arbitrary dependence structure.

Say that we are testing m hypothesis (so we have p -values p_1, \dots, p_m) and test at a significance level of q . The procedure proposed by Benjamini and Hochberg (1995) first computes ordered p -values $p_{(1)}, \dots, p_{(n)}$. Then compute

$$r = \max \left\{ 1 \leq j \leq m : p_{(j)} \leq \frac{j}{m}q \right\} \quad (6.1.1)$$

and reject the null hypotheses corresponding to $p_{(1)}, \dots, p_{(r)}$. For convenience, we will refer to this as the *first FDR method*. In contrast, the correction for arbitrary correlation due to Benjamini and Yekutieli (2001) rejects the hypotheses corresponding to the first

$$\tilde{r} = \max \left\{ 1 \leq j \leq m : p_{(j)} \leq \frac{j}{m}q / \left(\sum_{j=1}^m 1/j \right) \right\} \quad (6.1.2)$$

p -values, i.e. $p_{(1)}, \dots, p_{(\tilde{r})}$. Such a corrected test is less powerful, but decreases false positives and controls the FDR in the more general case (Benjamini and Yekutieli, 2001). We will refer to this as the *second FDR method*. We compare each of these and the standard Bonferroni correction (i.e. using q/m in place of q as a significance

⁴The ratio is taken to be zero when $R = 0$.

level) in Section 6.3.1. Benjamini and Yekutieli (2001) show that under positive dependence, the first FDR method controls the FDR to be less than or equal to $\frac{m_0}{m}q$ (where m_0 is the number of true null hypotheses), and the second FDR method always controls the FDR at this level.

6.1.5 Choosing the number of surrogate realisations

Schulte (2016) suggest generating “a large ensemble” of surrogate realisations. In this section, we give a lower bound on the number of required surrogate realisations, N_r , motivated by the FDR. Notice that, at a given frequency pair ω_1, ω_2 , if the observed squared skewness function $\hat{\Gamma}_X(\omega_1, \omega_2)$ is below the squared skewness function of all of the surrogate data, i.e.

$$\hat{\Gamma}_X(\omega_1, \omega_2) < \hat{\Gamma}_{X_j}(\omega_1, \omega_2) \quad \text{for all } j \in \{1, \dots, N_r\}$$

then the p -value will be zero, and will be rejected by any method for controlling the FDR, regardless of how conservative the method is.⁵ If we do not generate enough surrogate realisations, then this may happen often, and can drastically increase the FDR.

Say that X is a Gaussian process and that we have a perfect surrogate data generator (in the sense that we can simulate realisations of the same Gaussian process perfectly). If we generate N_r surrogate realisations, then the probability that we obtain a p -value of zero at a given frequency pair is $1/(N_r + 1)$. This occurs because we have $N_r + 1$ total realisations of the process (including the original), and thus at a given frequency pair have $N_r + 1$ i.i.d realisations of the squared skewness function, and the p -value is zero if the smallest of these realisations corresponds to the original process X .

⁵Assuming the method is sensible.

Say that we test at a total of N_T frequency pairs. Because we assumed the process was Gaussian, we have a false rejection if *any* of these tests are rejected. The probability that at least one test is rejected is bounded below by the probability that at least one of the p -values is zero. Let M be the number of p -values that are zero, then

$$\begin{aligned}\mathbb{P}(M \geq 1) &= 1 - \mathbb{P}(M = 0) \\ &= 1 - (N_r / (N_r + 1))^{N_T}.\end{aligned}$$

Say that we wish to have an FDR of q . Then we have

$$\begin{aligned}q &\geq \mathbb{P}(M \geq 1) \\ &= 1 - (N_r / (N_r + 1))^{N_T}\end{aligned}$$

and so,

$$\begin{aligned}q &\geq 1 - (N_r / (N_r + 1))^{N_T} \\ \Leftrightarrow \frac{N_r}{N_r + 1} &\geq (1 - q)^{1/N_T} \\ \Leftrightarrow N_r &\geq \frac{(1 - q)^{1/N_T}}{1 - (1 - q)^{1/N_T}}.\end{aligned}$$

In our experiments in Section 6.3, we have 90 evaluation frequencies and $q = 0.05$. Therefore, we require at the very least $N_r \geq 1755$. Of course this is only a lower bound, and so we often use a larger N_r . In particular, in our simulation studies in Section 6.3, we used $N_r = 4000$.

6.2 Toy example

In many practical studies of higher-order spectra (Schulte, 2016; Ewans et al., 2021, for example) the toy examples examined are discrete frequency, often determinis-

tic processes⁶ embedded in noise. However, the processes we are interested in are not discrete frequency. Models for non-linear time series are of course available, see Tjøstheim (1994), but the processes described below are more similar to the processes we are interested in. As such, we shall study toy examples constructed as polynomials of Gaussian processes. Such models can induce interesting behaviours, whilst being simple enough to study analytically. In particular, we shall consider polynomial interactions between two processes, loosely designed to mimic an interaction between a swell and a wind-sea process.

6.2.1 Univariate

In the univariate case, consider two independent Gaussian processes U and V (we might think of them as a swell and a wind-sea component). Define a new process W by

$$W(t) = U(t) + V(t) + aU(t)V(t). \quad (6.2.1)$$

The spectral density function is

$$f_W(\omega) = f_U(\omega) + f_V(\omega) + a^2[f_U * f_V](\omega).$$

Similarly, the bispectrum is

$$\begin{aligned} f_W(\omega_1, \omega_2) = & a[f_U(\omega_1)f_V(\omega_2) + f_U(\omega_2)f_V(\omega_1) + \\ & f_U(-\omega_2)f_V(\omega_1 + \omega_2) + f_U(\omega_1 + \omega_2)f_V(-\omega_2) + \\ & f_U(\omega_1)f_V(\omega_1 + \omega_2) + f_U(\omega_1 + \omega_2)f_V(\omega_1)]. \end{aligned}$$

⁶Such processes are not stationary. In this case, time averaged spectra and bispectra can be defined, but we do not consider that here. See Nikias (1993) for more details.

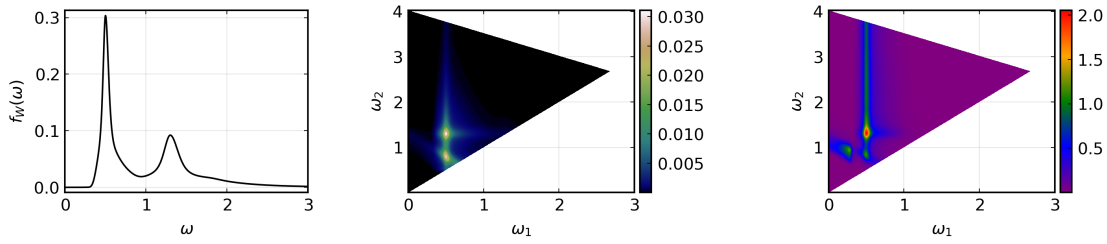


Figure 6.2.1: Example model spectral density function (left), bispectrum (middle) and squared skewness function (right).

The derivation of this is given in Appendix C.2. For example, consider U and V with JONSWAP spectral density functions with parameters given in Table 6.3.1 and $a = 0.8$. Then the spectral density functions of W are shown in Figure 6.2.1 (left) and the bispectrum and squared skewness function are shown in the middle and right panes of Figure 6.2.1.

6.2.2 Multivariate

In the multivariate case, let U_j and V_j for $j \in \{1, \dots, d\}$ be Gaussian processes such that for any $i, j \in \{1, \dots, d\}$, U_i is independent of V_j . We shall consider models of the form

$$\begin{aligned} W_j(t) &= U_j(t) + V_j(t) + \sum_{r=1}^d \sum_{s=1}^d a_{rs}^{(j)} U_r(t) V_s(t) \\ &= U_j(t) + V_j(t) + \text{vec}(\mathbf{a}_j^T)^T [\underline{\mathbf{U}}(t) \otimes \underline{\mathbf{V}}(t)] \end{aligned}$$

where $\mathbf{a}_j = [a_{rs}^{(j)}]_{1 \leq r, s \leq d} \in \mathbb{R}^{d \times d}$ are matrices of coefficients. Essentially these models are additive models with additional interaction terms which can be cross series.

The cross-spectral density functions are given by

$$f_{Z_i Z_j}(\omega) = f_{U_i U_j}(\omega) + f_{V_i V_j}(\omega) + \sum_{r_1=1}^d \sum_{s_1=1}^d \sum_{r_2=1}^d \sum_{s_2=1}^d a_{r_1 s_1}^{(i)} a_{r_2 s_2}^{(j)} [f_{U_{r_1} U_{r_2}} * f_{V_{s_1} V_{s_2}}](\omega)$$

and the cross-bispectra are

$$\begin{aligned}
 f_{Z_i Z_j Z_k}(\omega_1, \omega_2) = & \sum_{r=1}^d \sum_{s=1}^d a_{rs}^{(i)} \left[f_{U_r U_k}(\omega_1 + \omega_2) \overline{f_{V_s V_j}(\omega_2)} + \overline{f_{U_r U_j}(\omega_2)} f_{V_s V_k}(\omega_1 + \omega_2) \right] \\
 & + a_{rs}^{(j)} \left[f_{U_r U_k}(\omega_1 + \omega_2) f_{V_i V_s}(\omega_1) + f_{U_i U_r}(\omega_1) f_{V_s V_k}(\omega_1 + \omega_2) \right] \\
 & + a_{rs}^{(k)} \left[f_{U_i U_r}(\omega_1) f_{V_j V_s}(\omega_2) + f_{U_j U_r}(\omega_2) f_{V_i V_s}(\omega_1) \right].
 \end{aligned}$$

The derivation of this is analogous to the univariate case.

6.3 Monte-Carlo simulation studies

In this section, we evaluate the performance of the proposed technique with a Monte Carlo simulation. These simulation studies all take on the following form (where N_m is the number of Monte Carlo replications).

1. For $j \in \{1 \dots, N_m\}$
 - (a) Generate a realisation of the process of length n .
 - (b) Compute the p -values with the desired technique (e.g. using surrogate processes generated by circulant embedding).
 - (c) Test hypotheses using the desired method for correcting for multiple testing, setting $r_j = 1$ if rejected and $r_j = 0$ otherwise.
2. Compute the proportion of rejected hypotheses, i.e. $\frac{1}{N_m} \sum_{j=1}^{N_m} r_j$.

Note that if we use surrogate data we must compute $N_m N_r$ estimates of the squared skewness function for each simulation study, at each pair of frequencies. This can be quite expensive, especially in the multivariate case.

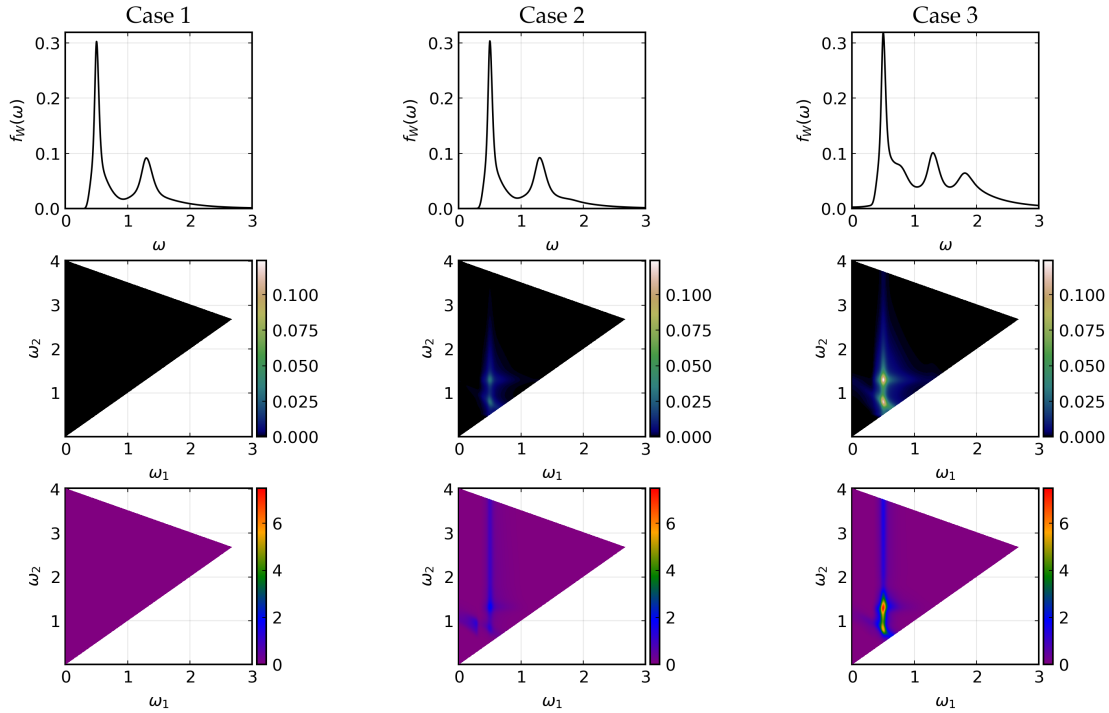


Figure 6.3.1: The spectral density functions (top), bispectra (middle) and squared skewness functions (bottom) of Cases 1, 2 and 3, the models used in the univariate simulation study. Here a is equal to zero (left), one (middle) and four (right).

6.3.1 Univariate simulation study

To evaluate the performance of circulant embedding in the univariate case, consider two independent Gaussian processes U and V with JONSWAP spectral density functions, the parameters of which are given in Table 6.3.1. We use three different cases from the polynomial model given in (6.2.1), $a = 0, 1, 4$. The spectral density functions, bispectra and squared skewness functions are shown in Figure 6.3.1.

	α	ω_p	γ	r
θ_U	0.02	0.5	3.3	5.0
θ_V	0.7	1.3	3.3	5.0

Table 6.3.1: Parameters used to generate the models for the simulation study.

To evaluate the performance of the AR(1) and circulant embedding techniques, we perform a Monte Carlo simulation as described above, with $N_m = 1000$ Monte Carlo replications, for $n = 2304$, $\Delta = 1/1.28$ with $N_r = 4000$ surrogate data replications.

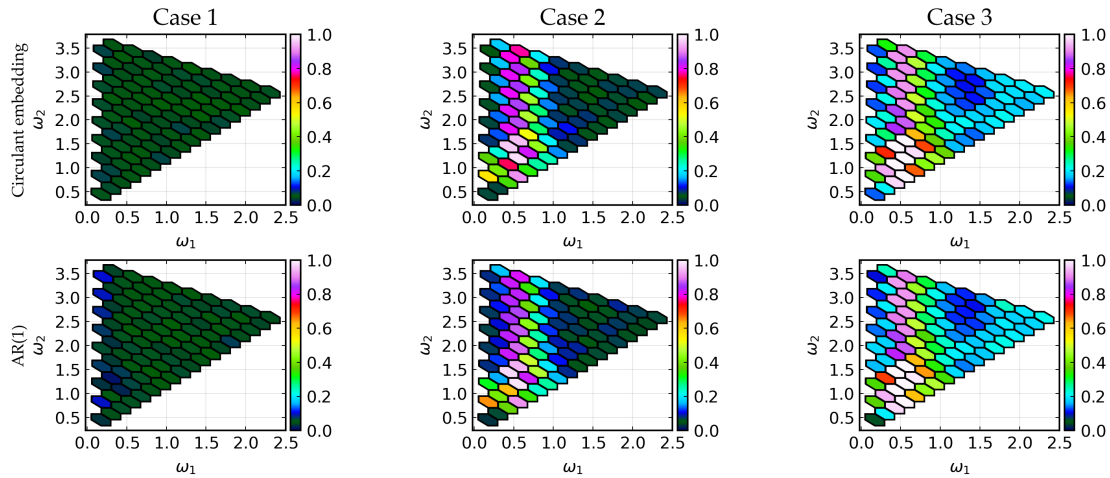


Figure 6.3.2: Proportion of rejections for different models using both circulant embedding and an AR(1) process to generate surrogate data for Cases 1, 2 and 3 (the univariate cases).

Figure 6.3.2 shows the proportion of rejections of the null at the $q = 0.05$ significance level for both circulant embedding and the AR(1) approach.

The first thing to note is that with circulant embedding in Case 1 (the Gaussian case) we have a rejection rate very close to 0.05 for all frequencies, as expected. This is in contrast to the AR(1), which has an inflated rejection rate near the boundaries. Looking at Case 2, we see that both techniques have high rejection rates in the region where the squared skewness function is (substantially) non-zero, but preserve the rejection rate of 0.05 away from this region. Finally, in Case 3, we see that both techniques detect the region in which the squared skewness function is large but, although the rejection rates away from this region are low, they are above 0.05. However, the process has a fairly extreme degree of non-linearity, so in practice this is unlikely to be a problem. In fact, the squared skewness function in both Case 2 and 3 is greater than zero everywhere, so in some sense we should detect something in these other regions.

At this point, we have not yet corrected for multiple testing. In Figure 6.3.3, we show the proportion of replications for which the null hypothesis is rejected using

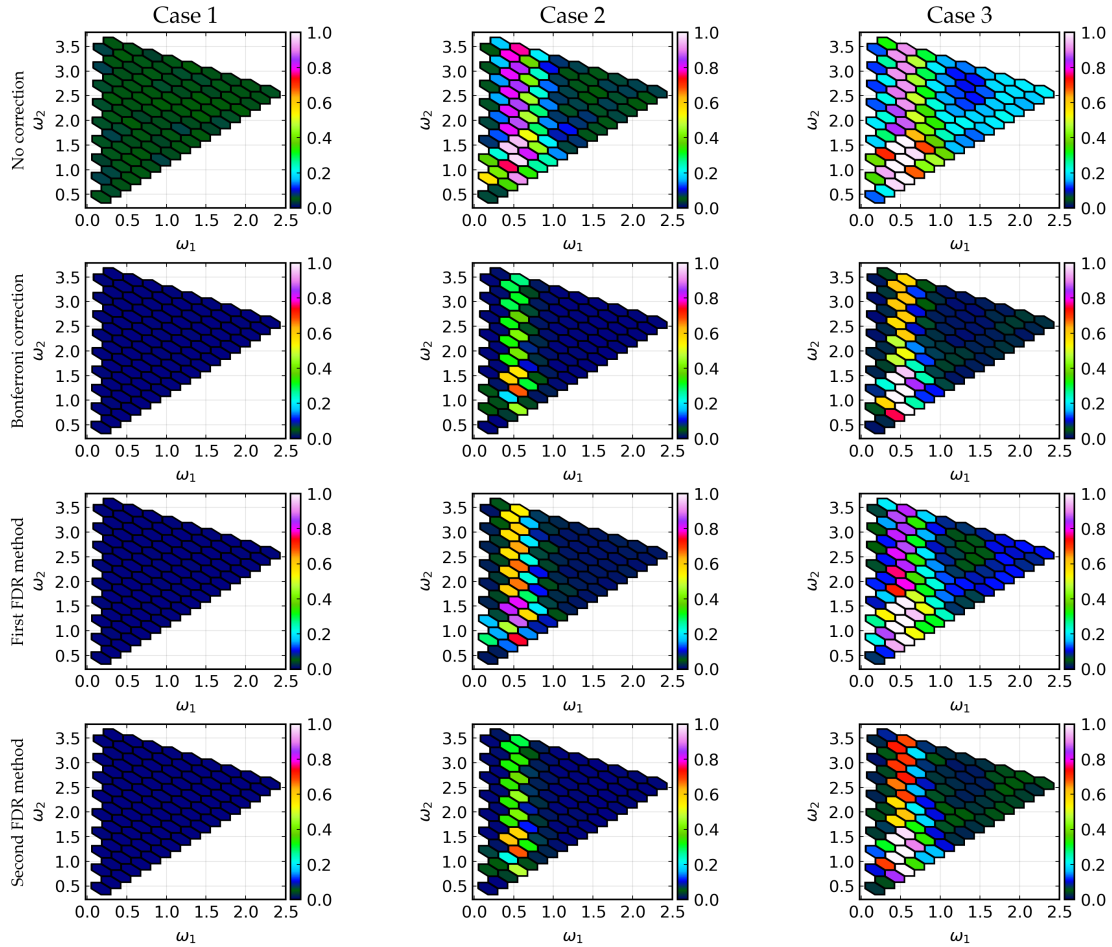


Figure 6.3.3: Proportion of rejections with different correction levels using circulant embedding to generate under the null for Cases 1, 2 and 3.

each correction method. We notice that all correction methods do a good job in the null case of reducing the detection rate. Similarly, we see as expected that, of all of the correction methods, the first FDR method has the highest power of the three, but this comes at the cost of higher detection rates away from the peak.

To evaluate the performance of the different corrections discussed in Section 6.1.4 more formally, we first must determine for which frequency pairs the null hypothesis (squared skewness function equal to zero) is true, and for which pairs the alternate is true. This is not trivial (except in Case 1), as in Cases 2 and 3, true squared skewness function is often very close to zero, but is never exactly zero (since the aliased

spectral density functions of U and V are never zero). We choose to define the null hypothesis to be true whenever the squared skewness function of the underlying process satisfies $\Gamma_W(\omega_1, \omega_2) < 0.1$.⁷ A plot of the regions considered to have true null and true alternate hypotheses is shown in Figure 6.3.4. Based on this definition, we estimate the FDR and power (number of detected false nulls over the total number of false nulls) of the corrected tests using, no correction, Bonferroni correction, and the two FDR methods, defined in (6.1.1) and (6.1.2). Again we are testing at the $q = 0.05$ significance level. In Cases 1, 2 and 3, the total number of true null hypotheses, m_0 , equals 90, 65 and 62 respectively and the total number of tested hypotheses is $m = 90$. Therefore, we expect to control the FDR at less than or equal to 0.05 in Case 1 and approximately 0.0361 and 0.0344 in Cases 2 and 3 respectively.

Table 6.3.2 shows the FDR and power for the different correction methods for Cases 1, 2 and 3. In Case 1, we see that without correction the FDR is very high, as expected. However, we also see that the FDR is slightly elevated in both the Bonferroni correction and first FDR method. In Cases 2 and 3, we see that, ignoring no correction, Bonferroni has the lowest FDR and lowest power as expected, followed by the second FDR method then the first FDR method. The first FDR method seems to not be conservative enough, so we will use the second FDR method, though in practice we will often also examine the results of the first FDR method. The results of additional simulation studies varying the γ parameters in the two JONSWAP processes are shown in Appendix C.3.

6.3.2 Multivariate

In the multivariate case, simulation studies yield similar results to the univariate case. In particular, we generate non-linear interactions between a swell process \mathbf{U}

⁷Note that this is entirely for evaluating the performance of our method, and not a part of the methodology itself.

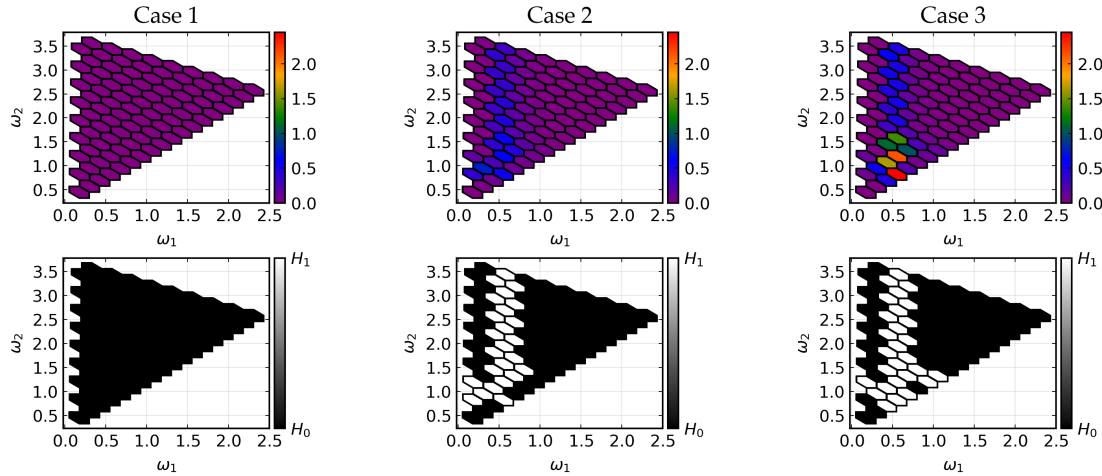


Figure 6.3.4: Definition of a true null, top row showing the true bispectra for each process on the estimation grid and bottom row showing which hypothesis is considered true for each frequency pair (black for H_0 and white for H_1). Here we use a definition of $\Gamma(\omega_1, \omega_2) < 0.1$ for the null.

and a wind-sea process \underline{V} . For the swell process, we use a JONSWAP with a single wrapped Gaussian spreading function⁸, and for the wind-sea we use a JONSWAP with bimodal wrapped Gaussian. The parameters of which are given in Table 6.3.3.

In Case 4, we do not include any interactions, so the resulting process is also Gaussian (i.e. $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3 = \mathbf{0}$). In Cases 5 and 6, we add non-linearities to the second time series, so that

$$\mathbf{a}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{Case 5}), \quad \mathbf{a}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 4 & 0 \\ 0 & 2 & 0 \end{bmatrix} \quad (\text{Case 6}),$$

but in both cases $\mathbf{a}_1 = \mathbf{a}_3 = \mathbf{0}$. Figure 6.3.5 shows the squared skewness functions of these models.

We compare the performance of the multivariate circulant embedding described in

⁸This is a wrapped Gaussian with fixed mean ϕ_m and fixed standard deviation σ at all frequencies.

technique	statistic	Case 1	Case 2	Case 3
No correction	FDR	0.987 (± 0.007)	0.284 (± 0.028)	0.363 (± 0.030)
	power	—	0.703 (± 0.028)	0.750 (± 0.026)
Bonferroni	FDR	0.065 (± 0.015)	0.048 (± 0.013)	0.104 (± 0.019)
	power	—	0.265 (± 0.027)	0.510 (± 0.031)
First FDR	FDR	0.065 (± 0.015)	0.133 (± 0.021)	0.258 (± 0.027)
	power	—	0.487 (± 0.031)	0.699 (± 0.028)
Second FDR	FDR	0.019 (± 0.009)	0.050 (± 0.014)	0.145 (± 0.022)
	power	—	0.282 (± 0.028)	0.578 (± 0.031)

Table 6.3.2: FDR and power for various corrections for multiple testing. These results are for the univariate simulation study using circulant embedding to generate surrogate data. We give 95% confidence intervals in brackets, so that the significance of any differences can be ascertained.

	α	ω_p	γ	r	ϕ_m	β	ν	σ_l	σ_r	σ
θ_U	0.02	0.5	3.3	5.0	1.5	—	—	—	—	0.5
θ_V	0.7	1.3	3.3	5.0	0.5	4.0	2.7	0.55	0.26	—

Table 6.3.3: Parameters used to generate the models for the multivariate simulation study (Cases 4, 5 and 6).

Section C.1 with a marginal circulant embedding approach, which simulates each individual component process independently. Figure 6.3.6 shows the result of this comparison. As expected, we see that simulating with multivariate circulant embedding yields the correct false positive rate, but if each series is simulated independently, we do not get the correct rate. This highlights the need for the multivariate surrogate data generator, as opposed to using the univariate version independently for each series. Results for the different multiple testing corrections behave similarly to univariate case, see Figure C.3.3 in Appendix C.3.

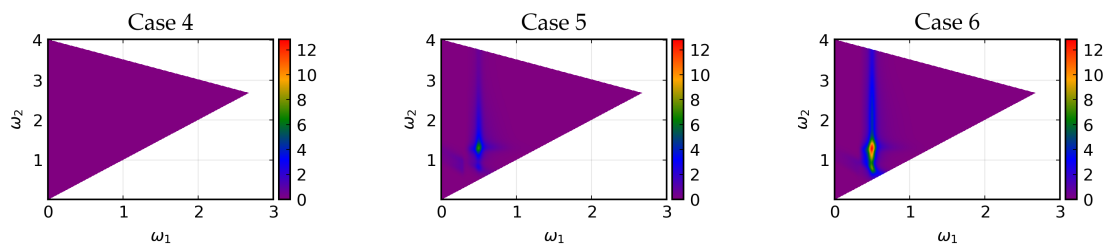


Figure 6.3.5: Squared skewness function of Cases 4, 5 and 6, the models used in the multivariate simulation study.

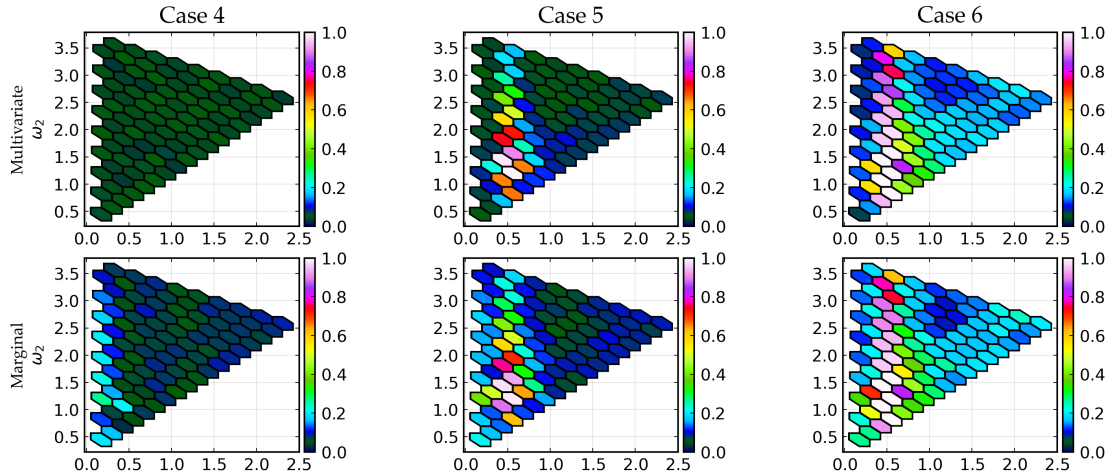


Figure 6.3.6: Proportion of rejections using multivariate circulant embedding and marginal univariate circulant embedding to generate under the null for Cases 4, 5 and 6 (the multivariate cases).

6.3.3 Bias correction

In Section 6.1.3 (and Section 5.3), we discussed estimation of the spectral density function and bispectrum. Of course, we also want to estimate the squared skewness function. Taking absolute values and squaring shifts variance into the mean, introducing bias into the estimate. However, as we shall show in this section, we can correct for this to some extent.

Recall from Section 5.5 (Hinich, 1982; Wong, 1997) that $\hat{R}_n(\omega_1, \omega_2)$ is approximately $\chi_{2d^3}^2(\lambda_n(\omega_1, \omega_2))$ where

$$\hat{R}_n(\omega_1, \omega_2) = 2\pi 2C_n n^{-1} \hat{\Gamma}_n(\omega_1, \omega_2),$$

and $\lambda_n(\omega_1, \omega_2) = 2\pi 2C_n n^{-1} \Gamma_{\underline{X}}(\omega_1, \omega_2)$. Therefore

$$\mathbb{E} [\hat{R}_n(\omega_1, \omega_2)] \approx 2\pi 2C_n n^{-1} \Gamma(\omega_1, \omega_2) + 2d^3,$$

which implies that

$$\mathbb{E} [\hat{\Gamma}_n(\omega_1, \omega_2)] \approx \Gamma(\omega_1, \omega_2) + \frac{d^3 n}{2\pi C_n}.$$

Furthermore, $d^3 n / (2\pi C_n)$ is known a priori, so we can reduce the bias in the estimate by simply subtracting this from $\hat{\Gamma}$, i.e. by using the

$$\check{\Gamma}_n(\omega_1, \omega_2) = \hat{\Gamma}_n(\omega_1, \omega_2) - \frac{d^3 n}{2\pi C_n},$$

as the estimate for squared skewness function.⁹ Since for consistency we already require $C_n/n \rightarrow \infty$ as $n \rightarrow \infty$, we have $d^3 n / (2\pi C_n) \rightarrow 0$ as $n \rightarrow \infty$. As a result, both $\hat{\Gamma}_n$ and $\check{\Gamma}_n$ are asymptotically unbiased, but the latter will have better finite sample properties. This is particularly relevant for the multivariate case, as the bias is proportional to d^3 . Figure 6.3.7 shows the bias of the estimated squared skewness function for the multivariate experiments, both before and after the correction. We see that the correction gives a major improvement in bias. However, estimates can now be negative. Whilst this may seem undesirable, we argue that it actually serves as a reminder of the substantial uncertainties present in such estimation, and is preferable to the bias present in the original estimate.¹⁰

6.4 Application to ocean wave data

We will now apply this methodology to the buoy data used in Chapter 2, from CDIP station 168, Humboldt Bay North Spit in California. Data were furnished by the Coastal Data Information Program, Integrative Oceanography Division, operated by the Scripps Institution of Oceanography, and is freely available on their website.

⁹Note that this bias adjustment involves an additive constant, and thus does not change the variance.

¹⁰The p -values are invariant to this correction, so it is not necessary for the testing, only for the estimate itself.

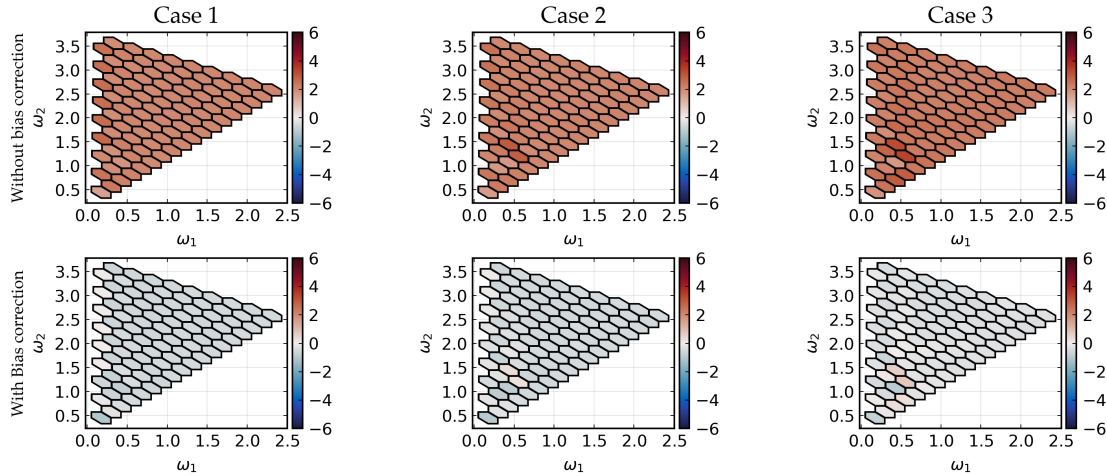


Figure 6.3.7: Average bias of estimated skewness function for Cases 4, 5 and 6 of the multivariate simulation, from 5000 Monte Carlo replications. Top row shows the bias of the standard estimates, and bottom row shows the bias of the corrected estimates.

Specifically, we shall demonstrate the technique on the half hour record with the largest value of H_s from the same storm analysed in Chapter 2. We choose the record with the largest H_s as this is most likely to contain non-linearities.

Recall that we have three-dimensional displacement data for the buoy. Therefore, we have three time series to analyse. We call them Z , X and Y , as in Chapter 3. We shall begin by considering each series as univariate. Figure 6.4.1 shows the resulting estimates of the squared skewness function. In order to show the results of significance testing, we highlight the borders of hexagons for which the null is rejected with colours depending on the correction level. In particular, we use purple if the p -value is less than 0.05, red if we reject the null hypothesis using the first FDR method, and orange if the null hypothesis is rejected using the second (more conservative) FDR method. We see that, provided we correct for multiple testing, the null is not rejected for any frequency pair in the case of Z and Y , but that for X , the null is rejected at many frequency pairs. However, we suspect this is spurious.

As previously mentioned, the squared skewness function is invariant to invertible linear filters, but estimates of the squared skewness function will not be. In particu-

lar, filtering the series to flatten the power spectrum and bispectrum will reduce bias from smoothing. The idea of first filtering the time series in this manner and then estimating spectra is referred to as prewhitening (Press and Tukey, 1956).¹¹ When prewhitening is used for estimating the spectral density function, the resulting estimate must be divided by some product of the transfer function of the filter to obtain an estimate of the spectrum of interest. However, in our case we do not need to do this, as the transfer function cancels out when we compute the squared skewness function.

In particular, we fit moving average, $MA(q)$, models with $q \in \{5, 10, 15, 20\}$, and then estimate the squared skewness function of the residuals (see Brockwell and Davis, 2006, for example). The results of this are shown in Figure 6.4.2. We now see that far fewer frequencies are detected as significant when compared to the unfiltered version in Figure 6.4.1. However, interestingly, we do see that we still detect an interaction near the anomalous low frequency peak (which we encountered earlier, in Section 3.4.1). In particular, across all of the prewhitening techniques, we detect non-linearity at approximately the frequencies $\omega_1 = 0.3$, $\omega_2 = 0.6$ and $\omega_3 = -0.9$. This is an interaction between the low frequency peak and the wind-sea component. In Figure 6.4.3 we show the estimated multivariate squared skewness function and test results, both with no filtering and after fitting marginal $MA(q)$ models. Again we see that such prewhitening seems necessary to stabilise results in practice. However, the low frequency peak is no longer detected as significant when using the more conservative FDR correction.

In addition to the results shown here for the sea state with the largest significant wave height, we also ran the analysis for the previous 20 sea states and the subsequent 20 sea states. The results of this analysis were similar to those shown above,

¹¹Differencing can be an example of such a prewhitening filter (which we used in Chapter 2).

in that often the low frequency peak is detected as a significant non-linearity, but other frequencies are not. Whilst this may seem to be somewhat of a non-result, this actually means that the second-order modelling we developed in Part I is sufficient to describe these processes (with the exception of the low-frequency peak, and the caveat that we have not examined spectra of higher order).

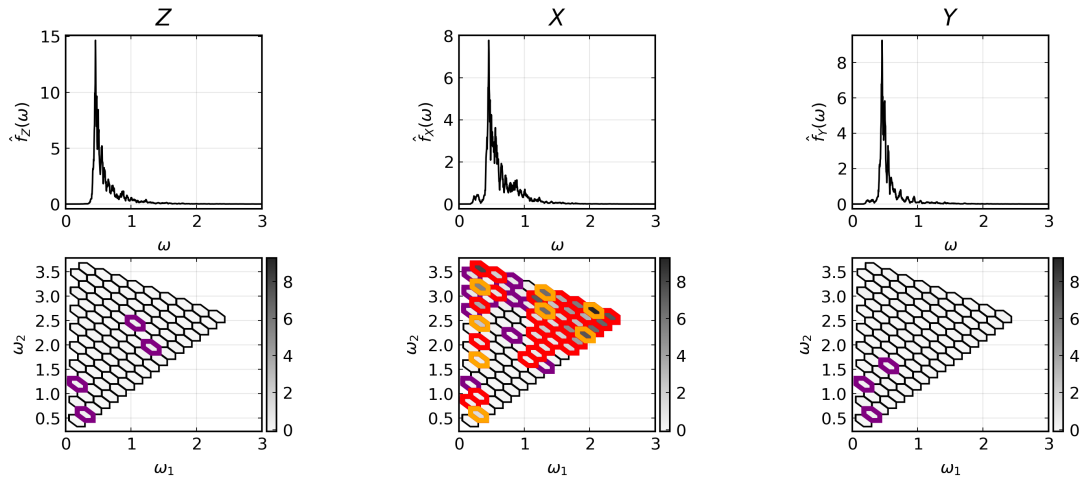


Figure 6.4.1: The squared skewness function of the X , Y and Z processes. If a p -value was classed as significant the border of the hexagon is highlighted with a colour depending on the correction method. In particular, in purple (no correction), red (first FDR method) and orange (second FDR method).

6.5 Conclusion

In this chapter, we have developed methodology for detecting non-linearities locally in frequency space. In particular, we demonstrated that existing surrogate data techniques based on simulating AR(1) processes have an elevated false detection rate for certain processes. In contrast, by simulating more representative processes, we are able to correct for this, without losing power. Furthermore, we explore the effect of different corrections for multiple testing and show how to remove the bias generated by taking absolute values. We also give recommendations for selecting the number of surrogate realisations, motivated by controlling the FDR.

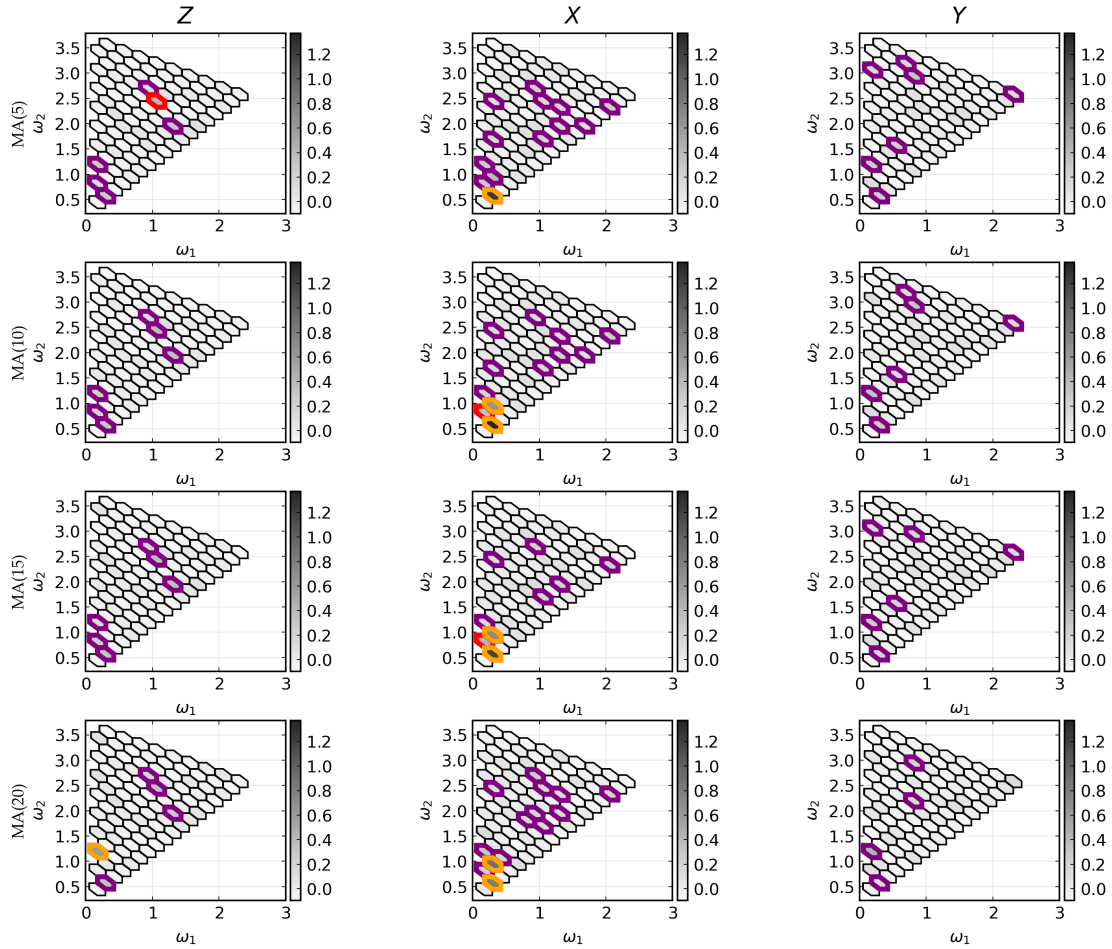


Figure 6.4.2: The squared skewness function of the X, Y and Z processes after applying various filters. The filters are fitted MA models with $q = 5, 10, 15$ and 20 for rows from top to bottom. Test results are highlighted analogously to Figure 6.4.1.

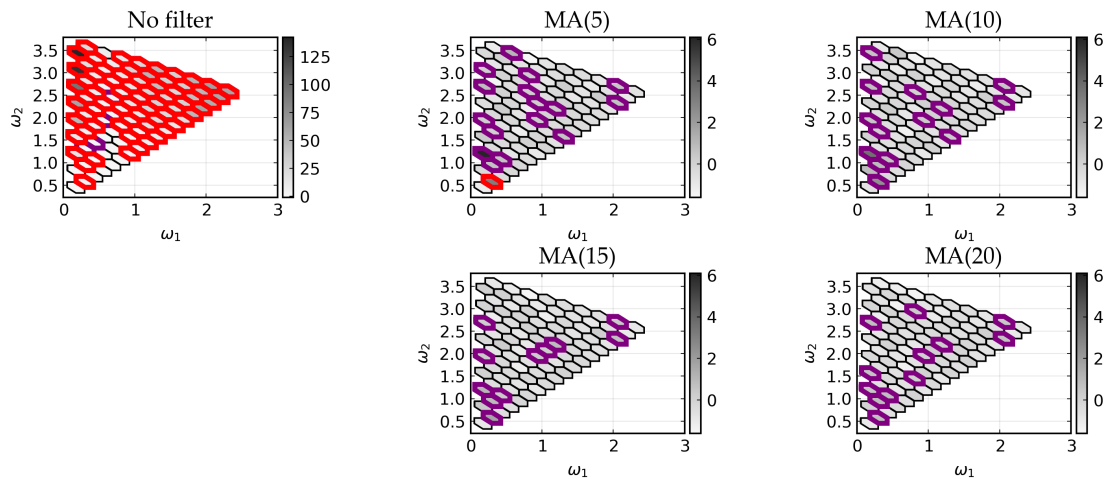


Figure 6.4.3: Estimate of the multivariate squared skewness function of the CDIP data, both with no filter and marginal MA models as applied in Figure 6.4.2.

Additionally, we demonstrated the technique on a time series recorded at Humboldt Bay. We found that the low frequency peak was detected as significant when applying the technique to the X time series in isolation, but it was not detected in the other series. However, we should note the detection in the X and not the Y series is likely related to the direction of the phenomena at this particular time, as it is unlikely that there is something special about the X displacement of the buoy when compared to the Y displacement in general. Furthermore, we note that we found similar results for other time series taken from the Humboldt Bay data, often detecting non-linearity in the low-frequency peak, but not in the frequencies associated with the processes of interest.

This is encouraging, as it suggests that the modelling we perform in Part I is sufficient to describe the behaviour of the observed processes. To end on a note of caution, whilst we have made some progress in making the analysis of the bispectrum of ocean wave records more conservative, results can be unstable without careful preprocessing of the data, e.g. by prewhitening.

Chapter 7

Conclusions of Part II

In Part II of this thesis, we have discussed the detection of non-linearities in records of ocean waves. We began, in Chapter 5, with a detailed background on higher-order spectra. This included a discussion of different normalisations and global tests for Gaussianity and linearity.

In Chapter 6, we discussed local testing for Gaussianity. Existing surrogate data generation techniques do not have the correct marginal false positive rate under the null. We demonstrated that circulant embedding can be utilised to resolve this issue. Additionally, we extended the methodology to the multivariate case, again using circulant embedding to generate surrogate realisations. Furthermore, we investigated both power and FDR for simulated processes with varying degrees of non-linearity, as well as various corrections for multiple testing.

In order to use such surrogate data techniques, we need to specify the number of surrogate realisations to generate. If we do not generate enough surrogate realisations, we will often obtain p-values of zero, which cannot be adjusted correctly for multiple testing. Motivated by this we give a lower bound on the number of required

surrogate replications so that the FDR *could* be at the required level, though this does not guarantee that it is at that level. Additionally, we demonstrated how the bias due to squaring the estimate of the bispectrum when computing the squared skewness function can easily be removed without increasing the variance.

We then applied this methodology to buoy data from Humboldt Bay, and discussed practical issues this raises. In particular, the impact of prewhitening on the resulting estimates and tests. It seems that the anomalous low-frequency peak seen in the spectral density function of the horizontal motion of the buoy may have been generated by non-linear interactions, but further investigation would be required to ascertain the exact nature of this phenomenon. For the most part, we do not detect significant non-linearities for most of the spectral domain, which is encouraging as it suggests that the linear modelling developed in Part I was sufficient to describe such processes.

7.1 Further work

In Chapter 6, we focused on the bispectrum. However, there is growing interest in spectra of higher order, i.e. the fourth-order spectral density function (Ewans et al., 2021, for example), which is usually called the trispectrum. The testing procedures described in Chapter 6 could be extended to such a setting, though it is likely that there will be substantial difficulties. Firstly, such estimation is much more costly from a computational perspective, even in the univariate case. Furthermore, we are somewhat doubtful that such quantities can be reliably estimated from ocean wave records without having to assume stationarity for an unpalatable length of time. In order to gain consistency, the loss of resolution is even worse than in the third-order case.

One drawback of the proposed methodology is that we can only test the null hypothesis of Gaussianity. In part, this is due to the method used for generating the surrogate data, however, it is also not obvious how we would test for a non-constant squared skewness function. One possibility is to note that under the null of linearity, estimates of the squared skewness function would have the same distribution (asymptotically). In particular, if we suspect that the squared skewness function is constant except in a certain region, we might expect a change in mean in this region. However, such an approach would likely be unstable.

This technique could also be applied to physical models, where, although the true squared skewness function may not be known, non-linear physical terms can be turned on and off, giving at least some ground truth. For example, the HOS model (Ducrozet et al., 2016), which was used by Ewans et al. (2021) to investigate the trispectrum.

We focused on the stationary case in this thesis, but ocean waves are not actually stationary. Approaches such as dual frequency spectra (Olhede and Ombao, 2013, for example) in the second-order case, or wavelets (both for second-order, e.g. Percival and Walden, 2000, and higher-order, e.g. Schulte, 2016) could be explored further for this kind of data. However, surrogate data generation will be substantially more challenging in this setting.

Epilogue

In this thesis, we developed methodology to solve two broad problems. Part I focused on parametric estimation for the frequency and frequency-direction spectrum. Part II discussed quantifying the significance of peaks in the bispectrum. Whilst different in theme, Parts I and II are joined by a strong practical motivation around the analysis of buoy data.

In Part I, we saw that substantial improvements can be made in the quality of estimated parameters, without the need to use computationally expensive methodology. The methodology developed throughout Chapters 2 and 3 can be used to model ever more complicated situations, and to make inferences about more intricate aspects of the frequency domain behaviour of ocean wave time series.

In Part II, methodology was developed to improve the detection of non-linearities present in the bispectrum. In this case, we saw that by simulating surrogate processes with similar linear structure to the original process, we are able to reduce the number of false detections. The methodology can then be used to determine whether an observed peak in the normalised bispectrum should be considered as significant.

In general, we have demonstrated that careful application of statistical methodology to buoy data yields improved insights about the behaviour of ocean waves.

Appendices to Parts I and II

“There he stood far into the night, hearing only the sigh and murmur of the waves on the shores of Middle-Earth, and the sound of them sank deep into his heart.”

— J. R. R. TOLKIEN, *THE RETURN OF THE KING*

Appendix A

Appendix to Chapter 2

A.1 Derivatives of ocean wave spectra

When calculating the derivatives of the generalised JONSWAP form, special care must be taken around the peak frequency and zero frequency, to check that the generalised JONSWAP form is actually differentiable. We do not show this here, but for the first derivatives there is no issue (though this is not the case for the second derivatives, which do not all exist at the peak). Consider the generalised JONSWAP

spectral form, $f_G(\omega; \underline{\theta})$. We may write, for $\omega > 0$,

$$\frac{\partial}{\partial \omega} f_G(\omega; \underline{\theta}) = f_G(\omega; \underline{\theta}) \left[\frac{r}{\omega_p} \left(\frac{\omega}{\omega_p} \right)^{-s-1} - \frac{r}{\omega} - \frac{\delta(\omega; \underline{\theta}) \cdot (\omega - \omega_p) \log \gamma}{\omega_p^2 \sigma(\omega; \underline{\theta})^2} \right]$$

$$\rightarrow 0 \text{ as } \omega \rightarrow 0^+,$$

$$\frac{\partial}{\partial \alpha} f_G(\omega; \underline{\theta}) = f_G(\omega; \underline{\theta}) / \alpha$$

$$\rightarrow 0 \text{ as } \omega \rightarrow 0^+,$$

$$\frac{\partial}{\partial \omega_p} f_G(\omega; \underline{\theta}) = f_G(\omega; \underline{\theta}) \left[\delta(\omega; \underline{\theta}) \log(\gamma) \frac{\omega}{\sigma(\omega; \underline{\theta})^2} \left(\frac{\omega - \omega_p}{\omega_p^3} \right) - \frac{\omega r}{\omega_p^2} \left(\frac{\omega}{\omega_p} \right)^{-s-1} \right]$$

$$\rightarrow 0 \text{ as } \omega \rightarrow 0^+,$$

$$\frac{\partial}{\partial \gamma} f_G(\omega; \underline{\theta}) = \alpha \omega^{-r} \exp \left\{ -\frac{r}{s} \left(\frac{\omega}{\omega_p} \right)^{-s} \right\} \delta(\omega; \underline{\theta}) \gamma^{\delta(\omega; \underline{\theta})-1} / 2$$

$$= f_G(\omega; \underline{\theta}) \delta(\omega; \underline{\theta}) / \gamma$$

$$\rightarrow 0 \text{ as } \omega \rightarrow 0^+,$$

$$\frac{\partial}{\partial r} f_G(\omega; \underline{\theta}) = f_G(\omega; \underline{\theta}) \left[-\log(\omega) - \frac{1}{s} \left(\frac{\omega}{\omega_p} \right)^{-s} \right]$$

$$\rightarrow 0 \text{ as } \omega \rightarrow 0^+.$$

By the chain rule, for $\omega < 0$

$$\frac{\partial}{\partial \omega} f_G(\omega; \underline{\theta}) = -\frac{\partial}{\partial \omega} f_G(-\omega; \underline{\theta}),$$

$$\frac{\partial}{\partial \alpha} f_G(\omega; \underline{\theta}) = \frac{\partial}{\partial \alpha} f_G(-\omega; \underline{\theta}),$$

$$\frac{\partial}{\partial \omega_p} f_G(\omega; \underline{\theta}) = \frac{\partial}{\partial \omega_p} f_G(-\omega; \underline{\theta}),$$

$$\frac{\partial}{\partial \gamma} f_G(\omega; \underline{\theta}) = \frac{\partial}{\partial \gamma} f_G(-\omega; \underline{\theta}),$$

$$\frac{\partial}{\partial r} f_G(\omega; \underline{\theta}) = \frac{\partial}{\partial r} f_G(-\omega; \underline{\theta}).$$

Finally, for $\omega = 0$, since $f_G(0, \alpha, \omega_p, \gamma, r) = 0$, we may write

$$\begin{aligned}\frac{\partial}{\partial \omega} f_G(\omega; \underline{\theta})|_{\omega=0} &= 0, \\ \frac{\partial}{\partial \alpha} f_G(0; \underline{\theta}) &= 0, \\ \frac{\partial}{\partial \omega_p} f_G(0; \underline{\theta}) &= 0, \\ \frac{\partial}{\partial \gamma} f_G(0; \underline{\theta}) &= 0, \\ \frac{\partial}{\partial r} f_G(0; \underline{\theta}) &= 0.\end{aligned}$$

Therefore, we see that f_G has continuous first derivatives.

Proposition A.1.1. *For all $x \in \{\omega, \alpha, \omega_p, \gamma, r\}$ the series*

$$\sum_{k=-\infty}^{\infty} \frac{\partial}{\partial x} f_G \left(\omega + \frac{2\pi k}{\Delta}; \underline{\theta} \right)$$

converges uniformly.

Proof. Firstly, due to symmetry, we may consider the series

$$\sum_{k=L}^{\infty} \frac{\partial}{\partial x} f_G \left(\omega + \frac{2\pi k}{\Delta}; \underline{\theta} \right)$$

where L is such that $\omega_p < \frac{\pi}{\Delta} + \frac{2\pi L}{\Delta}$. Now write

$$\begin{aligned}g_k(\omega; \underline{\theta}) &= f_G \left(\omega + \frac{2\pi k}{\Delta}; \underline{\theta} \right) \quad \text{and} \\ g_{k,x}(\omega; \underline{\theta}) &= \frac{\partial}{\partial x} f_G \left(\omega + \frac{2\pi k}{\Delta}; \underline{\theta} \right).\end{aligned}$$

Next notice that

$$\begin{aligned} |g_k(\omega; \underline{\theta})| &\leq \alpha\gamma \left(\omega + \frac{2\pi k}{\Delta} \right)^{-r} \\ &\leq \alpha\gamma \left(\frac{-\pi}{\Delta} + \frac{2\pi k}{\Delta} \right)^{-r} \\ &= \alpha\gamma \left(\frac{\pi}{\Delta} \right)^{-r} (2k-1)^{-r}. \end{aligned}$$

Let $M_k = \alpha\gamma (\pi/\Delta)^{-r} (2k-1)^{-r}$. It now suffices to notice that each of the partial derivatives can be written as $f_G(\omega; \underline{\theta})$ multiplied by some other function. Furthermore, each of these functions can be bounded. Therefore we may write

$$|g_{k,x}(\omega; \underline{\theta})| \leq C_x M_k,$$

for some constants $C_x > 0$. Finally

$$\sum_{k=L}^{\infty} M_k < \infty,$$

so by the Weierstrass M-test, we have uniform convergence. By extension we can see that

$$\sum_{k=-\infty}^{\infty} \frac{\partial}{\partial x} f_G \left(\omega + \frac{2\pi k}{\Delta}; \underline{\theta} \right)$$

is also uniformly convergent. □

Proposition A.1.2. For all $x \in \{\omega, \alpha, \omega_p, \gamma, r\}$

$$\frac{\partial}{\partial x} f_{\Delta}(\omega; \underline{\theta}) = \sum_{k=-\infty}^{\infty} \frac{\partial}{\partial x} f_G \left(\omega + \frac{2\pi k}{\Delta}; \underline{\theta} \right).$$

Furthermore, $\frac{\partial}{\partial x} f_{\Delta}(\omega; \underline{\theta})$ is continuous.

Proof. This first part follows from Proposition A.1.1, the convergence of

$$\sum_{k=-\infty}^{\infty} f_G(\omega + 2\pi k/\Delta; \underline{\theta}),$$

the continuous differentiability of f_G , and Theorem 4.4.20 of Trench (2013). The continuity follows from the continuity of the derivatives of f_G and the uniform limit theorem. \square

A.2 Computing the variance of the first derivative

Sykulski et al. (2019) decompose the variance of the first derivative of the debiased Whittle likelihood as follows:

$$\text{var} \left(\frac{\partial}{\partial \underline{\theta}_i} \ell_{DW}(\underline{\theta}) \right) = \sum_{j=1}^n \sum_{k=1}^n a_{ij}(\underline{\theta}) a_{ik}(\underline{\theta}) \text{cov}(I(\omega_j), I(\omega_k)),$$

where ω_j, ω_k denote Fourier frequencies and

$$a_{ij}(\underline{\theta}) = \frac{\partial \bar{f}_n(\omega_j; \underline{\theta})}{\partial \underline{\theta}_i} \frac{1}{\bar{f}_n^2(\omega_j; \underline{\theta})}.$$

To estimate the variance, they propose using

$$\hat{\text{var}} \left(\frac{\partial}{\partial \underline{\theta}_i} \ell_{DW}(\underline{\theta}) \right) = \sum_{j=1}^n \sum_{k=1}^n a_{ij}(\hat{\underline{\theta}}_{DW}) a_{ik}(\hat{\underline{\theta}}_{DW}) \hat{\text{cov}}(I(\omega_j), I(\omega_k)).$$

For ocean wave models, $a_{ij}(\hat{\underline{\theta}}_{DW})$ can be easily computed by using the results from Section A.1. So, from Sykulski et al. (2019), we are interested in computing

$$\hat{\text{cov}}(I(\omega_j), I(\omega_k)) = \left| \frac{\Delta}{2\pi n} \int_{-\pi/\Delta}^{\pi/\Delta} f_{\Delta}(\omega'; \underline{\theta}) \mathcal{D}_n(\Delta(\omega_j - \omega')) \bar{\mathcal{D}}_n(\Delta(\omega_k - \omega')) d\omega' \right|^2,$$

where

$$\mathcal{D}_n(v) = \frac{\sin(nv/2)}{\sin(v/2)} e^{-iv(n+1)/2}.$$

To do this efficiently, first note that

$$\begin{aligned} \mathcal{D}_n(v) &= \frac{\sin(nv/2)}{\sin(v/2)} e^{-iv(n+1)/2} = \frac{e^{inv/2} - e^{-inv/2}}{e^{iv/2} - e^{-iv/2}} e^{-iv(n+1)/2} \\ &= \frac{e^{-inv/2} e^{inv} - 1}{e^{-iv/2} e^{iv} - 1} e^{-iv(n+1)/2} = e^{-i(n-1)v/2} e^{-iv(n+1)/2} \sum_{s=0}^{n-1} e^{isv} \\ &= e^{-inv} \sum_{s=0}^{n-1} e^{isv}. \end{aligned}$$

It is also convenient to write

$$\bar{\mathcal{D}}_n(v) = \frac{\sin(nv/2)}{\sin(v/2)} e^{iv(n+1)/2} = e^{iv} \sum_{s=0}^{n-1} e^{isv}.$$

Now, consider the function

$$\begin{aligned} h_{jk}(\omega) &= f_{\Delta}(\omega; \underline{\theta}) \mathcal{D}_n(\Delta(\omega_j - \omega)) \bar{\mathcal{D}}_n(\Delta(\omega_k - \omega)) \\ &= f_{\Delta}(\omega; \underline{\theta}) \left(e^{-in(\Delta(\omega_j - \omega))} \sum_{s=0}^{n-1} e^{is(\Delta(\omega_j - \omega))} \right) \left(e^{i(\Delta(\omega_k - \omega))} \sum_{r=0}^{n-1} e^{ir(\Delta(\omega_k - \omega))} \right) \\ &= e^{-i\Delta(n\omega_j - \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k} \left[f_{\Delta}(\omega; \underline{\theta}) e^{i\Delta(n-1)\omega} e^{-i\Delta(s+r)\omega} \right], \end{aligned}$$

where we have rearranged for later convenience. We can now see, by linearity of integration, that $\int_{-\pi/\Delta}^{\pi/\Delta} h_{jk}(\omega) d\omega$ equals

$$e^{-i\Delta(n\omega_j - \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k} \int_{-\pi/\Delta}^{\pi/\Delta} f_{\Delta}(\omega; \underline{\theta}) e^{i\Delta(n-1)\omega} e^{-i\Delta(s+r)\omega} d\omega.$$

Thus for $r = 0, \dots, n-1$ and $s = 0, \dots, n-1$ we must calculate

$$\int_{-\pi/\Delta}^{\pi/\Delta} f_{\Delta}(\omega; \underline{\theta}) e^{i\Delta(n-1)\omega} e^{-i\Delta(r+s)\omega} d\omega.$$

Notice that by letting $t = r + s$ we need to calculate the following integral for $t = 0, \dots, (2n-1) - 1$

$$\int_{-\pi/\Delta}^{\pi/\Delta} f_{\Delta}(\omega; \underline{\theta}) e^{i\Delta(n-1)\omega} e^{-i\Delta t\omega} d\omega.$$

For clarity let

$$q(\omega) = f_{\Delta}(\omega; \underline{\theta}) e^{i\Delta(n-1)\omega}$$

then we need to compute

$$Q(t) = \int_{-\pi/\Delta}^{\pi/\Delta} q(\omega) e^{-i\Delta t\omega} d\omega, \tag{A.2.1}$$

for $t = 0, \dots, 2n-2$. We notice that this is a Fourier transform and we can obtain an approximation of this integral at each of the required t by doing an FFT on the relevant length $2n-1$ sequence (the details of this are discussed in Section A.2.1). This means that we only need to do one Fourier transform, and can then substitute this into the previous sums. Now we require

$$\begin{aligned} \int_{-\pi/\Delta}^{\pi/\Delta} h_{jk}(\omega) d\omega &= e^{-i\Delta(n\omega_j - \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k} Q(r+s) \\ &= e^{-i\Delta(n\omega_j - \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} \tilde{Q}(s, r) e^{i\Delta 2\pi(sj+rk)/n\Delta} \\ &= e^{-i\Delta(n\omega_j - \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} \tilde{Q}(s, r) e^{i2\pi(sj+rk)/n} \end{aligned}$$

where $\tilde{Q}(s, r) = Q(r + s)$. This is a 2D Fourier transform, and can be computed efficiently for $j, k = 0, \dots, n - 1$. This means we take $O(n^2 \log n)$ time. Importantly, libraries exist to compute this very quickly in most programming languages. Finally

$$\begin{aligned} \text{cov}(I(\omega_j), I(\omega_k)) &= \left| \frac{\Delta}{2\pi n} \int_{-\pi/\Delta}^{\pi/\Delta} h_{jk}(\omega) d\omega \right|^2 \\ &= \left| \frac{\Delta}{2\pi n} e^{-i\Delta(n\omega_j - \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} \tilde{Q}(s, r) e^{i2\pi(sj+rk)/n} \right|^2 \\ &= \left| \frac{\Delta}{2\pi n} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} \tilde{Q}(s, r) e^{i2\pi(sj+rk)/n} \right|^2. \end{aligned}$$

A.2.1 Approximating $Q(t)$

Note that we must still approximate $Q(t)$ at $t = 0, \dots, 2n - 2$. We aim to approximate the integral in (A.2.1). To achieve this, we use the Riemann sum given by

$$\bar{Q}(t) = \frac{2\pi}{m\Delta} \sum_{j=0}^{m-1} q(2\pi j/m\Delta) e^{-i(t\Delta)(2\pi j/m\Delta)}.$$

An FFT will produce values of $\bar{Q}(t)$ for $t = 0, 1, \dots, (m - 1)$. Therefore, provided $M \geq 2n - 1$ we can obtain approximations of the desired integrals.

A.3 Assumptions for debiased Whittle

Theorem 1 of Sykulski et al. (2019) gives assumptions for the optimal convergence of the debiased Whittle likelihood estimator. The first of these assumptions is that the parameter space is compact with non-null interior and that the true value of the parameter lies in the interior of the parameter space. This is not strictly satisfied by the generalised JONSWAP spectral density; however, for physical reasons the parame-

ter space can be restricted so that it is compact. The greater problem is that a value of $\gamma = 1$, which corresponds to the fully developed sea described by Pierson and Moskowitz (1964), may occur in nature and would violate this assumption. In Section 2.5.3 we demonstrate that, in this case, the debiased Whittle likelihood estimator still performs well.

The second assumption is that the spectral density of the aliased process is bounded above and is bounded below by some positive real number. This is satisfied by the generalised JONSWAP spectral density. The aliased spectrum is bounded below by the non-aliased spectrum, and is bounded above by the variance of $X(t)$ (which is finite). Therefore, the only frequency remaining to consider is zero, as the spectral density is zero when $\omega = 0$. However, contributions from above the Nyquist frequency are positive, and as such the aliased spectral density at zero will also be positive.

Assumption three relates to parameter identifiability. Informally, this requires the aliased spectral density function to be different for different choices of $\underline{\theta}$. Intuitively, provided the sampling interval is sufficiently small (so that the peak frequency is smaller than the Nyquist), then each of the parameters is changing the shape of the generalised JONSWAP spectrum in a different way, such that the parameters will in general be identifiable for a sufficient sample size.

The fourth assumption states that the aliased spectral density function must be continuous in $\underline{\theta}$ and Riemann integrable in ω . This is satisfied for the generalised JONSWAP spectral form as it is continuous in $\underline{\theta}$ and ω .

Assumption five states that the expected periodogram has two continuous derivatives in $\underline{\theta}$, and that these derivatives are bounded uniformly for all n . Furthermore the first derivatives are required to have $\Theta(n)$ non-zero frequencies. Strictly speak-

ing, the generalised JONSWAP is not twice differentiable (the second derivative does not exist at the peak, due to the step function $\sigma(\omega; \underline{\theta})$). However, a simple adaptation can be made, by replacing $\sigma(\omega; \underline{\theta})$ with

$$\bar{\sigma}(\omega; \underline{\theta}) = \sigma_1 + (\sigma_2 - \sigma_1) \left(\frac{1}{2} + \frac{1}{\pi} \arctan (C (\omega - \omega_p)) \right),$$

where $C \in (0, \infty)$ is chosen to be large. This has the advantage of having continuous second derivatives, and is essentially equivalent to the generalised JONSWAP, because C can be chosen such that it would be impossible to distinguish between the two models from observed data. Since this part of the generalised JONSWAP was developed empirically, there is no practical difference in using the step function over this reformulation. Indeed, in keeping with the general philosophy of statistical modelling — “all models are wrong, but some models are useful” — we suggest that this alternative formulation is just as appropriate as the generalised JONSWAP, but more useful here as it allows us to show this assumption is satisfied. By similar arguments to those presented in Appendix A.1 and Appendix A.2, we can see that the aliased spectral density has continuous second derivatives. The autocovariance then has continuous second derivatives by a similar argument to that in Section 2.6, and noting that the second derivative of the aliased spectral density function is integrable, and so therefore the second derivative of the autocovariance must be continuous (as they can be shown to be Fourier pairs). Therefore, the expected periodogram also has continuous second derivatives by linearity of derivatives and the fact that linear combinations of continuous functions are continuous. These derivatives are also bounded uniformly for all n due to the compactness of the set of frequencies Ω and the parameter space $\underline{\theta}$.

The final assumption states that the process in question is fourth-order stationary with finite fourth order moments and absolutely summable fourth order cumulants.

Clearly this is true for a Gaussian process (as second order stationarity implies strict stationarity for Gaussian) and it is also true for some non-linear processes, such as the class of non-linear processes discussed by Sykulski et al. (2019).

Finally, in our simulations the estimator based on the debiased Whittle likelihood performs in broad agreement with the theory, for example we observed desirable properties such as root n convergence when exploring different values of n . We also did not find any problems with local minima during optimisation for any of the record lengths considered, suggesting that the parameters of the generalised JON-SWAP form are indeed identifiable in practice for sufficiently long records.

Appendix B

Appendix to Chapter 3

B.1 Deriving the spectral density matrix function from the frequency-direction spectrum

We wish to obtain an expression for the spectral density matrix function of a buoy assuming that the surface has a bimodal wrapped Gaussian spreading function. Firstly, observe that many of these integrals can be rearranged, in ways which will be demonstrated shortly, such that they can be split into integrals that are of the form

$$\int_{-\pi}^{\pi} \cos(n\phi)g_X(\phi)d\phi = \mathbb{E}[\cos(nX)] = \mathcal{R}\left(\mathbb{E}\left[e^{inX}\right]\right),$$

or

$$\int_{-\pi}^{\pi} \sin(n\phi)g_X(\phi)d\phi = \mathbb{E}[\sin(nX)] = \mathcal{I}\left(\mathbb{E}\left[e^{inX}\right]\right),$$

where $g_X(\phi)$ is the pdf of some wrapped random variable X and $n \in \mathbb{Z}$. In other words, we need only to compute the characteristic function of the random variable X , and then can just take real and imaginary parts. Fortunately, many random variables (including Gaussian random variables) have a known characteristic function with a nice close form. Furthermore, the characteristic function of a random variable and its wrapped form are related.¹ Denote the unwrapped random variable by Y , so that $g_X(x) = \sum_{k=-\infty}^{\infty} g_Y(x + 2\pi k)$. Then, for integer values of n ,

$$\begin{aligned}
 \mathbb{E} \left[e^{inX} \right] &= \int_{-\pi}^{\pi} e^{inx} g_X(x) dx \\
 &= \int_{-\pi}^{\pi} e^{inx} \sum_{k=-\infty}^{\infty} g_Y(x + 2\pi k) dx \\
 &= \sum_{k=-\infty}^{\infty} \int_{-\pi}^{\pi} e^{inx} g_Y(x + 2\pi k) dx \\
 &= \sum_{k=-\infty}^{\infty} \int_{\pi(2k-1)}^{\pi(2k+1)} e^{inx} e^{-ink2\pi} g_Y(x) dx \\
 &= \sum_{k=-\infty}^{\infty} \int_{\pi(2k-1)}^{\pi(2k+1)} e^{inx} g_Y(x) dx \\
 &= \int_{-\infty}^{\infty} e^{inx} g_Y(x) dx \\
 &= \mathbb{E} \left[e^{inY} \right],
 \end{aligned}$$

where the interchange of sum and integral is justified because $g_X(\cdot)$ is a probability density so has an absolute integral which is finite (hence we may apply the Fubini-Tonelli theorem). For a wrapped Gaussian with mean μ and variance σ^2 , we see that

$$\mathbb{E} \left[e^{inX} \right] = e^{in\mu - \frac{1}{2}\sigma^2 n^2},$$

¹Interestingly, this is actually an analogous relation to the one between the spectral density functions of a continuous time process and the corresponding sampled discrete time process.

for $n \in \mathbb{Z}$. Therefore, by extension

$$\begin{aligned}\mathbb{E} [\cos(nX)] &= \cos(n\mu)e^{-\frac{1}{2}\sigma^2n^2}, \\ \mathbb{E} [\sin(nX)] &= \sin(n\mu)e^{-\frac{1}{2}\sigma^2n^2}.\end{aligned}$$

We shall give results for $\omega > 0$, noting that the remainder of the domain can be obtained by conjugate symmetry (and that all the cross spectra are 0 for $\omega = 0$). We also only give some of the cross spectra, noting that the remainder can also be obtained by the Hermitian symmetry of the spectral density matrix. We introduce the simplified notation $g_{X_j}(\cdot) = g_X(\cdot; \phi_{mj}(\omega), \sigma(\omega)^2)$ for $j = 1, 2$.

Rewriting equation (3.2.1) for the cross spectra (Longuet-Higgins et al., 1963) we see that

$$\begin{aligned}f_{ZZ}(\omega) &= f(\omega), \\ f_{XX}(\omega) &= f(\omega) \int_{-\pi}^{\pi} \cos^2(\phi) D(\omega, \phi) d\phi, \\ f_{YY}(\omega) &= f(\omega) \int_{-\pi}^{\pi} \sin^2(\phi) D(\omega, \phi) d\phi, \\ f_{XZ}(\omega) &= if(\omega) \int_{-\pi}^{\pi} \cos(\phi) D(\omega, \phi) d\phi, \\ f_{YZ}(\omega) &= if(\omega) \int_{-\pi}^{\pi} \sin(\phi) D(\omega, \phi) d\phi, \\ f_{YX}(\omega) &= f(\omega) \int_{-\pi}^{\pi} \cos(\phi) \sin(\phi) D(\omega, \phi) d\phi.\end{aligned}$$

Then we see that

$$f_{ZZ}(\omega) = f(\omega),$$

also

$$\begin{aligned}
f_{XX}(\omega) &= f(\omega) \int_{-\pi}^{\pi} \cos^2(\phi) D(\omega, \phi) d\phi \\
&= f(\omega) \int_{-\pi}^{\pi} \cos^2(\phi) \frac{1}{2} [g_{X_1}(\phi) + g_{X_2}(\phi)] d\phi \\
&= f(\omega) \int_{-\pi}^{\pi} \frac{1}{2} (\cos(2\phi) + 1) \frac{1}{2} [g_{X_1}(\phi) + g_{X_2}(\phi)] d\phi \\
&= f(\omega) \frac{1}{4} \left(\mathcal{R} \left\{ \mathbb{E} \left[e^{i2X_1} \right] \right\} + \mathcal{R} \left\{ \mathbb{E} \left[e^{i2X_2} \right] \right\} + 2 \right) \\
&= f(\omega) \frac{1}{4} \left(\cos(2\phi_{m1}(\omega)) e^{-2\sigma(\omega)^2} + \cos(2\phi_{m2}(\omega)) e^{-2\sigma(\omega)^2} + 2 \right) \\
&= f(\omega) \frac{1}{2} \left(\cos^2(\phi_{m1}(\omega)) e^{-2\sigma(\omega)^2} + \cos^2(\phi_{m2}(\omega)) e^{-2\sigma(\omega)^2} - e^{-2\sigma(\omega)^2} + 1 \right),
\end{aligned}$$

similarly

$$\begin{aligned}
f_{YY}(\omega) &= f(\omega) \int_{-\pi}^{\pi} \sin^2(\phi) D(\omega, \phi) d\phi \\
&= f(\omega) \int_{-\pi}^{\pi} \frac{1}{2} (1 - \cos(2\phi)) \frac{1}{2} [g_{X_1}(\phi) + g_{X_2}(\phi)] d\phi, \\
&= f(\omega) \frac{1}{4} \left(2 - \mathcal{R} \left\{ \mathbb{E} \left[e^{i2X_1} \right] \right\} - \mathcal{R} \left\{ \mathbb{E} \left[e^{i2X_2} \right] \right\} \right) \\
&= f(\omega) \frac{1}{4} \left(2 - \cos(2\phi_{m1}(\omega)) e^{-2\sigma(\omega)^2} - \cos(2\phi_{m2}(\omega)) e^{-2\sigma(\omega)^2} \right) \\
&= f(\omega) \frac{1}{2} \left(\sin^2(\phi_{m1}(\omega)) e^{-2\sigma(\omega)^2} + \sin^2(\phi_{m2}(\omega)) e^{-2\sigma(\omega)^2} - e^{-2\sigma(\omega)^2} + 1 \right), \\
f_{XZ}(\omega) &= if(\omega) \int_{-\pi}^{\pi} \cos(\phi) D(\omega, \phi) d\phi, \\
&= if(\omega) \frac{1}{2} \left(\cos(\phi_{m1}(\omega)) e^{-\sigma(\omega)^2/2} + \cos(\phi_{m2}(\omega)) e^{-\sigma(\omega)^2/2} \right), \\
f_{YZ}(\omega) &= if(\omega) \int_{-\pi}^{\pi} \sin(\phi) D(\omega, \phi) d\phi, \\
&= if(\omega) \frac{1}{2} \left(\sin(\phi_{m1}(\omega)) e^{-\sigma(\omega)^2/2} + \sin(\phi_{m2}(\omega)) e^{-\sigma(\omega)^2/2} \right),
\end{aligned}$$

and finally

$$\begin{aligned}
 f_{YX}(\omega) &= f(\omega) \int_{-\pi}^{\pi} \cos(\phi) \sin(\phi) D(\omega, \phi) d\phi \\
 &= f(\omega) \int_{-\pi}^{\pi} \frac{1}{2} \sin(2\phi) D(\omega, \phi) d\phi \\
 &= f(\omega) \frac{1}{4} \left(\mathcal{I} \left\{ \mathbb{E} \left[e^{i2X_1} \right] \right\} + \mathcal{I} \left\{ \mathbb{E} \left[e^{i2X_2} \right] \right\} \right) \\
 &= f(\omega) \frac{1}{4} \left(\sin(2\phi_{m1}(\omega)) e^{-2\sigma(\omega)^2} + \sin(2\phi_{m2}(\omega)) e^{-2\sigma(\omega)^2} \right).
 \end{aligned}$$

B.2 Finite water depth correction

The relation given in (3.2.1) is for waves in deep water. For finite water depths a slightly different relation is required. In particular, we have

$$G(\omega, \phi) = \begin{bmatrix} 1 & i \cos \phi / \tanh(kh) & i \sin \phi / \tanh(kh) \end{bmatrix}^T$$

where h is the water depth (in metres)² and $\omega^2 = k \tanh(kh)$. Consequently we have

$$f_{ZZ}(\omega) = (f_{XX}(\omega) + f_{YY}(\omega)) \tanh(kh)^2$$

meaning that the finite water depth definition for $R(\omega)$ is

$$R(\omega) = \log(f_{XX}(\omega) + f_{YY}(\omega)) + 2 \log(\tanh(kh)) - \log(f_{ZZ}(\omega)).$$

It is this definition for $R(\omega)$ we use in Figure 3.4.1.

²In our case, $h \approx 40$.

B.3 Optimisation and gradient calculation

Parameters are estimated jointly by optimising the debiased Whittle likelihood using the interior point Newton method as implemented in `Optim.jl` (Mogensen and Riseth, 2018). We use Fisher scoring as the expected Hessian of the debiased Whittle likelihood can be computed analytically from the first derivatives of the autocovariance (whilst the Hessian would require the second derivatives as well). This results in very fast optimisation compared to other approaches. For further details, see the Julia package `WhittleLikelihoodInference.jl` (Grainger, 2022c).

B.4 Confidence intervals of the multivariate debiased Whittle likelihood

In Chapter 3, we used the inverse expected Hessian to compute confidence intervals. It is possible to obtain improved confidence intervals, analogously to those used in Chapter 2. However, they are much harder to compute from a computational perspective, so we did not use them. However, for completeness, we present the derivation here.

From Proposition 4 of Guillaumin et al. (2022), we have

$$\text{var}(\hat{\underline{\theta}}_{DW}) \approx \mathbb{E}[\mathbf{H}(\underline{\theta})]^{-1} \text{var}(\nabla \ell_{DW}(\underline{\theta})) \mathbb{E}[\mathbf{H}(\underline{\theta})]^{-1}.$$

Sykulski et al. (2019); Guillaumin et al. (2022) give the form for $\text{var}(\nabla \ell_{DW}(\underline{\theta}))$ in the univariate case, for a Gaussian process and Gaussian random field respectively. Here we extend this to the multivariate case.

Firstly, we have that

$$\begin{aligned} \frac{\partial \ell_D(\underline{\theta})}{\partial \theta_j} &= \sum_{\omega \in \Omega_N} \text{tr} \left(\frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j} \mathbf{E}(\omega; \underline{\theta})^{-1} - \mathbf{I}(\omega) \mathbf{E}(\omega; \underline{\theta})^{-1} \frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j} \mathbf{E}(\omega; \underline{\theta})^{-1} \right), \\ &= \sum_{\omega \in \Omega_N} \text{tr} \left([\mathbf{E}(\omega; \underline{\theta}) - \mathbf{I}(\omega)] \mathbf{E}(\omega; \underline{\theta})^{-1} \frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j} \mathbf{E}(\omega; \underline{\theta})^{-1} \right), \end{aligned}$$

where $\mathbf{E}(\omega; \underline{\theta}) = \mathbb{E}[\mathbf{I}(\omega); \underline{\theta}]$ and $\frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j}$ is the matrix of element-wise derivatives of $\mathbf{E}(\omega; \underline{\theta})$ with respect to θ_j . For convenience, write

$$\begin{aligned} \mathbf{B}_j(\omega; \underline{\theta}) &= \mathbf{E}(\omega; \underline{\theta})^{-1} \frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j} \mathbf{E}(\omega; \underline{\theta})^{-1} \\ \mathbf{A}_j(\omega; \underline{\theta}) &= [\mathbf{E}(\omega; \underline{\theta}) - \mathbf{I}(\omega)] \mathbf{B}_j(\omega; \underline{\theta}) \end{aligned}$$

so that

$$\frac{\partial \ell_D(\underline{\theta})}{\partial \theta_j} = \sum_{\omega \in \Omega_N} \text{tr}(\mathbf{A}_j(\omega; \underline{\theta})).$$

Therefore we have

$$\begin{aligned} \text{cov} \left(\frac{\partial \ell_D(\underline{\theta})}{\partial \theta_j}, \frac{\partial \ell_D(\underline{\theta})}{\partial \theta_k} \right) &= \text{cov} \left(\sum_{\omega \in \Omega_N} \text{tr}(\mathbf{A}_j(\omega; \underline{\theta})), \sum_{\lambda \in \Omega_N} \text{tr}(\mathbf{A}_k(\lambda; \underline{\theta})) \right) \\ &= \sum_{\omega \in \Omega_N} \sum_{\lambda \in \Omega_N} \text{cov}(\text{tr}(\mathbf{A}_j(\omega; \underline{\theta})), \text{tr}(\mathbf{A}_k(\lambda; \underline{\theta}))) \\ &= \sum_{\omega \in \Omega_N} \sum_{\lambda \in \Omega_N} \text{tr}(\text{cov}(\mathbf{A}_j(\omega; \underline{\theta}), \mathbf{A}_k(\lambda; \underline{\theta}))) \end{aligned}$$

where the covariance here is matrix covariance (see [Wong, 1997](#), for example). The matrix covariance is essentially comprised from stacking marginal covariances in the same way as a Kronecker product, i.e. for random matrices $\mathbf{A}_1 \in \mathbb{C}^{m_1 \times n_1}$ and

$A_2 \in \mathbb{C}^{m_2 \times n_2}$ we have

$$\text{cov}(A_1, A_2) = \left[\text{cov} \left(a_{rs}^{(1)}, a_{kl}^{(2)} \right) \right]_{r,s,k,l}$$

where the indices are as in $A_1 \otimes A_2$. From Lemma 2.1 of Wong (1997), we have that if we have constant matrices of appropriate dimensions B_1, B_2, C_1, C_2 , that

$$\begin{aligned} \text{cov}(B_1 A_1 C_1, B_2 A_2 C_2) &= \mathcal{C} \left[B_1 A_1 C_1, \overline{B_2 A_2 C_2} \right] \\ &= (B_1 \otimes \overline{B_2}) \text{cov}(A_1, A_2) (C_1 \otimes \overline{C_2}). \end{aligned}$$

Note that Lemma 2.1 of Wong (1997) is for general cumulants, so we need to conjugate the $k = 2$ case for covariance. Furthermore, covariance is shift invariant. Therefore, we see that

$$\begin{aligned} \text{cov}(A_j(\omega; \underline{\theta}), A_k(\lambda; \underline{\theta})) &= \text{cov}([E(\omega; \underline{\theta}) - I(\omega)] B_j(\omega; \underline{\theta}), [E(\lambda; \underline{\theta}) - I(\lambda)] B_k(\lambda; \underline{\theta})) \\ &= \text{cov}(I(\omega), I(\lambda)) (B_j(\omega; \underline{\theta}) \otimes \overline{B_k(\lambda; \underline{\theta})}). \end{aligned} \quad (\text{B.4.1})$$

We now consider each of these parts in turn. Firstly, $\text{cov}(I(\omega), I(\lambda))$ is entirely determined by

$$\text{cov}(I_{rs}(\omega), I_{kl}(\lambda)) = \frac{\Delta}{2\pi n} \text{cov}(J_r(\omega) \overline{J_s(\omega)}, J_k(\lambda) \overline{J_l(\lambda)})$$

where $I_{rs}(\omega) = [I(\omega)]_{rs}$ and $J_r(\omega) = [J(\omega)]_r$. As in Percival (1992), we make use of Israelis Theorem (Isserlis, 1918) which states that if we have four mean-zero complex Gaussian random variables Z_1, Z_2, Z_3, Z_4 then

$$\text{cov}(Z_1 Z_2, Z_3 Z_4) = \text{cov}(Z_1, Z_3) \text{cov}(Z_2, Z_4) + \text{cov}(Z_1, Z_4) \text{cov}(Z_2, Z_3).$$

Now $J_r(\omega)$ is approximately mean-zero and complex Gaussian and so assuming it is we can write

$$\begin{aligned} \text{cov}(J_r(\omega)\bar{J}_s(\omega), J_k(\lambda)\bar{J}_l(\lambda)) &= \text{cov}(J_r(\omega), J_k(\lambda)) \text{cov}(\bar{J}_s(\omega), \bar{J}_l(\lambda)) \\ &\quad + \text{cov}(J_r(\omega), \bar{J}_l(\lambda)) \text{cov}(\bar{J}_s(\omega), J_k(\lambda)). \end{aligned}$$

Again in parallel to the argument presented by Percival (1992) in the univariate case, we note that

$$\begin{aligned} J_k(\omega) &= \int_{-\pi/\Delta}^{\pi/\Delta} \mathcal{D}_n(\Delta[\omega - u]) d\zeta_k(u) \\ &= - \int_{-\pi/\Delta}^{\pi/\Delta} \mathcal{D}_n(\Delta[\omega + u]) \overline{d\zeta_k(u)}. \end{aligned}$$

and so writing

$$\begin{aligned} \mathfrak{C}_{rl}(\omega, \lambda) &= \text{cov}(J_r(\omega), \bar{J}_l(\lambda)) \\ &= \int_{-\pi/\Delta}^{\pi/\Delta} \mathcal{D}_n(\Delta[\omega - u]) \overline{\mathcal{D}_n(\Delta[\lambda - u])} f_{rl}^{(\Delta)}(u) du \\ \mathfrak{D}_{rk}(\omega, \lambda) &= \text{cov}(J_r(\omega), J_k(\lambda)) \\ &= - \int_{-\pi/\Delta}^{\pi/\Delta} \mathcal{D}_n(\Delta[\omega - u]) \mathcal{D}_n(\Delta[\lambda + u]) f_{rk}^{(\Delta)}(u) du \end{aligned}$$

where $f_{rl}^{(\Delta)}(u) = [f_{\Delta}(u; \theta)]_{rl}$.

Putting this together we have

$$\begin{aligned}
\text{cov}(J_r(\omega)\bar{J}_s(\omega), J_k(\lambda)\bar{J}_l(\lambda)) &= \text{cov}(J_r(\omega), J_k(\lambda)) \text{cov}(\bar{J}_s(\omega), \bar{J}_l(\lambda)) \\
&\quad + \text{cov}(J_r(\omega), \bar{J}_l(\lambda)) \text{cov}(\bar{J}_s(\omega), J_k(\lambda)) \\
&= \text{cov}(J_r(\omega), J_k(\lambda)) \overline{\text{cov}(J_s(\omega), J_l(\lambda))} \\
&\quad + \text{cov}(J_r(\omega), \bar{J}_l(\lambda)) \overline{\text{cov}(J_s(\omega), \bar{J}_k(\lambda))} \\
&= \mathfrak{D}_{rk}(\omega, \lambda) \overline{\mathfrak{D}_{sl}(\omega, \lambda)} + \mathfrak{C}_{rl}(\omega, \lambda) \overline{\mathfrak{C}_{sk}(\omega, \lambda)}.
\end{aligned}$$

Integrals of the type \mathfrak{C} can easily be computed efficiently using the approach described in Appendix A.2. Integrals of the type \mathfrak{D} , can be computed analogously by noting that, under the notation of Appendix A.2, if

$$\begin{aligned}
h_{jk}(\omega) &= f_\Delta(\omega) \mathcal{D}_n(\Delta[\omega_j - \omega]) \mathcal{D}_n(\Delta[\omega_k + \omega]) \\
&= f_\Delta(\omega; \underline{\theta}) \left(e^{-in(\Delta(\omega_j - \omega))} \sum_{s=0}^{n-1} e^{is(\Delta(\omega_j - \omega))} \right) \left(e^{-in(\Delta(\omega_k - \omega))} \sum_{r=0}^{n-1} e^{ir(\Delta(\omega_k - \omega))} \right) \\
&= e^{-i\Delta n(\omega_j + \omega_k)} \sum_{s=0}^{n-1} \sum_{r=0}^{n-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k} \left[f_\Delta(\omega; \underline{\theta}) e^{i\Delta(1-n)\omega} e^{-i\Delta(s-r+n-1)\omega} \right],
\end{aligned}$$

using $q(\omega) = f_\Delta(\omega; \underline{\theta}) e^{i\Delta(1-n)\omega}$ and $\tilde{Q}(s, r) = Q(s - r + n - 1)$ in the algorithm in Appendix A.2, we can approximate the desired integral. Clearly we also need to replace $f_\Delta(\omega; \underline{\theta})$ with the appropriate cross spectra, but the algorithm works the same regardless. Note that the term $e^{-i\Delta n(\omega_j + \omega_k)}$ will cancel when we compute $\mathfrak{D}_{rk}(\omega, \lambda) \overline{\mathfrak{D}_{sl}(\omega, \lambda)}$.

Now moving to the second part of (B.4.1), i.e.

$$\begin{aligned} \mathbf{B}_j(\omega; \underline{\theta}) \otimes \bar{\mathbf{B}}_k(\lambda; \underline{\theta}) &= \left(\mathbf{E}(\omega; \underline{\theta})^{-1} \frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j} \mathbf{E}(\omega; \underline{\theta})^{-1} \right) \\ &\quad \otimes \left(\mathbf{E}(\lambda; \underline{\theta})^{-1} \frac{\partial \mathbf{E}(\lambda; \underline{\theta})}{\partial \theta_k} \mathbf{E}(\lambda; \underline{\theta})^{-1} \right) \\ &= \mathfrak{E}(\omega, \lambda; \underline{\theta}) \left(\frac{\partial \mathbf{E}(\omega; \underline{\theta})}{\partial \theta_j} \otimes \frac{\partial \mathbf{E}(\lambda; \underline{\theta})}{\partial \theta_k} \right) \mathfrak{E}(\omega, \lambda; \underline{\theta}) \end{aligned}$$

where

$$\mathfrak{E}(\omega, \lambda; \underline{\theta}) = \left(\mathbf{E}(\omega; \underline{\theta})^{-1} \otimes \bar{\mathbf{E}}(\lambda; \underline{\theta})^{-1} \right).$$

Note that the covariance matrix of interest $\text{cov}(\mathbf{I}(\omega), \mathbf{I}(\lambda))$ is a $d^2 \times d^2$ matrix.

B.5 Proof of the approximation of cross-spectral density function for two wave staffs

Recall that we need to compute

$$I_1 = \int_0^{2\pi} \exp\{ik(x \cos \phi + y \sin \phi)\} D(\omega, \phi) d\phi$$

Noting that ω is fixed and so $D(\omega, \cdot)$ is a pdf for a random variable supported on $[0, 2\pi]$, call it X . Thus we want to compute $I_1 = \mathbb{E}[\exp\{i(a \cos X + b \sin X)\}]$, where $a = kx$ and $b = ky$. Firstly

$$\forall x \in \mathbb{R}, \quad |a \cos x + b \sin x|^k \leq |a + b|^k$$

therefore

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{E} \left[\left| \frac{i^k (a \cos X + b \sin X)^k}{k!} \right| \right] &= \sum_{k=0}^{\infty} \mathbb{E} \left[\frac{|a \cos X + b \sin X|^k}{k!} \right] \\ &\leq \sum_{k=0}^{\infty} \mathbb{E} \left[\frac{|a + b|^k}{k!} \right] \\ &= \sum_{k=0}^{\infty} \frac{|a + b|^k}{k!} = e^{|a+b|} < \infty. \end{aligned}$$

So we may apply the Fubini-Tonelli theorem and get

$$\begin{aligned} I_1 &= \mathbb{E} \left[\sum_{k=0}^{\infty} i^k \frac{(a \cos X + b \sin X)^k}{k!} \right] \\ &= \sum_{k=0}^{\infty} \frac{i^k}{k!} \mathbb{E} \left[(a \cos X + b \sin X)^k \right]. \end{aligned}$$

Thus we must compute, for all $k \in \mathbb{N}_0$,

$$\begin{aligned} I_2(k) &= \mathbb{E} \left[(a \cos X + b \sin X)^k \right] \\ &= \mathbb{E} \left[\sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \cos^n X \sin^{k-n} X \right] \\ &= \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \mathbb{E} \left[\cos^n X \sin^{k-n} X \right]. \end{aligned}$$

In other words, we must compute, for all $n, m \in \mathbb{N}_0$,

$$\begin{aligned} I_3(n, m) &= \mathbb{E} [\cos^n X \sin^m X] \\ &= \mathbb{E} \left[\left(\frac{1}{2^n} \sum_{j=0}^n \binom{n}{j} e^{i(2j-n)X} \right) \left(\frac{1}{(2i)^m} \sum_{r=0}^m \binom{m}{r} (-1)^{m-r} e^{i(2r-m)X} \right) \right] \\ &= \frac{1}{2^{n+m} i^m} \sum_{r=0}^m (-1)^{m-r} \binom{m}{r} \sum_{j=0}^n \binom{n}{j} \mathbb{E} \left[e^{i(2j+2r-n-m)X} \right] \\ &= \frac{1}{2^{n+m} i^m} \sum_{r=0}^m (-1)^{m-r} \binom{m}{r} \sum_{j=0}^n \binom{n}{j} \varphi_X(2j + 2r - n - m), \end{aligned}$$

where $\varphi_X(t)$ is the characteristic function of X . Note at this point X could be any distribution.

Putting this together we have

$$\begin{aligned} I_2(k) &= \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} I_3(n, k-n) \\ &= \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \frac{1}{2^k i^{k-n}} \sum_{r=0}^{k-n} \binom{k-n}{r} (-1)^{k-n-r} \sum_{j=0}^n \binom{n}{j} \varphi_X(2j+2r-k) \\ &= \frac{1}{2^k} \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \sum_{r=0}^{k-n} \binom{k-n}{r} i^{k-n-2r} \sum_{j=0}^n \binom{n}{j} \varphi_X(2j+2r-k), \end{aligned}$$

and finally

$$I_1 = \sum_{k=0}^{\infty} \frac{i^k}{k!} \frac{1}{2^k} \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \sum_{r=0}^{k-n} \binom{k-n}{r} i^{k-n-2r} \sum_{j=0}^n \binom{n}{j} \varphi_X(2j+2r-k).$$

In our case, we wish to compute this for $X \sim \mathcal{N}(\mu, \sigma^2)$. So we have that $\varphi_X(t) = e^{it\mu - t^2\sigma^2/2}$. Therefore

$$I_1 = \sum_{k=0}^{\infty} \frac{i^k}{k!} \frac{1}{2^k} \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \sum_{r=0}^{k-n} \binom{k-n}{r} i^{k-n-2r} \sum_{j=0}^n \binom{n}{j} e^{i(2j+2r-k)\mu - (2j+2r-k)^2\sigma^2/2}.$$

In fact, we are interested in $Y \sim \text{BN}(\mu_1, \mu_2, \sigma^2)$ where BN is the bimodal Normal distribution, with pdf given by $[g(y; \mu_1, \sigma^2) + g(y; \mu_2, \sigma^2)]/2$ where g is the pdf of a Gaussian random variable. Therefore, from linearity of Fourier transforms we

have

$$\begin{aligned}\varphi_Y(t) &= \frac{\varphi_{X_1}(t) + \varphi_{X_2}(t)}{2} \\ &= \frac{1}{2} \left(e^{it\mu_1 - t^2\sigma^2/2} + e^{it\mu_2 - t^2\sigma^2/2} \right) \\ &= \frac{e^{-t^2\sigma^2/2}}{2} \left(e^{it\mu_1} + e^{it\mu_2} \right).\end{aligned}$$

Now note that

$$\begin{aligned}I_1 &= \sum_{k=0}^{\infty} \frac{i^k}{k!} \frac{1}{2^k} \sum_{n=0}^k \binom{k}{n} a^n b^{k-n} \sum_{r=0}^{k-n} \binom{k-n}{r} i^{k-n-2r} \sum_{j=0}^n \binom{n}{j} \varphi_Y(2j+2r-k) \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^k a^n b^{k-n} \sum_{r=0}^{k-n} \frac{i^k}{k!} \frac{1}{2^k} i^{k-n-2r} \sum_{j=0}^n \binom{k}{n} \binom{k-n}{r} \binom{n}{j} \varphi_Y(2j+2r-k) \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^k a^n b^{k-n} \sum_{r=0}^{k-n} \frac{i^{-n} \cdot (-1)^{k-r}}{2^k} \sum_{j=0}^n \frac{1}{r!(k-n-r)!j!(n-j)!} \varphi_Y(2j+2r-k) \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^k a^n b^{k-n} J_1(k, n),\end{aligned}$$

where

$$\begin{aligned}J_1(n, k) &= \sum_{r=0}^{k-n} \frac{i^{-n} \cdot (-1)^{k-r}}{2^k} \sum_{j=0}^n \frac{1}{r!(k-n-r)!j!(n-j)!} \varphi_Y(2j+2r-k) \\ &= \sum_{r=0}^{k-n} \frac{i^{-n} \cdot (-1)^{k-r}}{2^{k+1}} \sum_{j=0}^n \frac{e^{-(2j+2r-k)^2\sigma^2/2} \left(e^{i(2j+2r-k)\mu_1} + e^{i(2j+2r-k)\mu_2} \right)}{r!(k-n-r)!j!(n-j)!} \\ &= \sum_{r=0}^{k-n} \frac{i^{-n} \cdot (-1)^{k+r}}{2^{k+1}} J_2(k, n, r),\end{aligned}$$

with

$$J_2(k, n, r) = \sum_{j=0}^n \frac{e^{-(2j+2r-k)^2\sigma^2/2} \left(e^{i(2j+2r-k)\mu_1} + e^{i(2j+2r-k)\mu_2} \right)}{r!(k-n-r)!j!(n-j)!}.$$

Appendix C

Appendix to Chapter 6

C.1 Circulant embedding

In this appendix, we review circulant embedding for univariate and multivariate time series, in particular from a non-parametric estimate of the spectral density function (Percival and Constantine, 2006; Chandna and Walden, 2013). It is this approach which we propose using to generate surrogate realisations.

C.1.1 Univariate circulant embedding

A matrix $A \in \mathbb{R}^{m \times m}$ is said to be circulant if each row is a cycle one to the right of the previous row. In other words, if

$$A = \begin{bmatrix} a(0) & a(1) & a(2) & \dots & a(m-1) \\ a(m-1) & a(0) & a(1) & \dots & a(m-2) \\ a(m-2) & a(m-1) & a(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a(1) \\ a(1) & a(2) & \dots & a(m-1) & a(0) \end{bmatrix}. \quad (\text{C.1.1})$$

If A is such a matrix, we can find a matrix square root of A in $O(m \log m)$ time (Brockwell and Davis, 2006).

Circulant embedding (Davies and Harte, 1987; Wood and Chan, 1994) makes use of this by noting the following: if \underline{Y} is multivariate Normal with zero mean and covariance matrix

$$A = \begin{bmatrix} \Sigma & \dots \\ \vdots & \ddots \end{bmatrix}$$

then the first n observations of \underline{Y} are multivariate Normal with mean zero and covariance matrix Σ . Furthermore, if A were circulant, we could obtain such a \underline{Y} using Fourier transforms. Remember that we are interested in second-order stationary processes, and Σ is symmetric and Toeplitz. As such,

$$\Sigma = \begin{bmatrix} c(0) & c(\Delta) & \dots & c([n-1]\Delta) \\ c(\Delta) & c(0) & \dots & c([n-2]\Delta) \\ \vdots & \vdots & \ddots & \vdots \\ c([n-1]\Delta) & c([n-2]\Delta) & \dots & c(0) \end{bmatrix}$$

is the covariance matrix of multivariate Gaussian from which we wish to simulate our univariate time series. If A has the representation given in (C.1.1), then we can achieve such an embedding by setting

$$a(j) = \begin{cases} c(j\Delta) & 0 \leq j \leq n, \\ c([m-j]\Delta) & m - (n-1) \leq j \leq m-1, \\ b(j) & \text{otherwise,} \end{cases}$$

where the $b(j)$ is chosen out of convenience so that the matrix is circulant (often set

equal to 0). If $m = 2n$ then these $b(j)$ s need not be specified, but this is no guarantee that this matrix A will be non-negative definite, so a larger m may have to be used. Because A is circulant, its eigenvalues are the Fourier transform of $a(0), \dots, a(m-1)$ (Brockwell and Davis, 2006), so A being non-negative definite is equivalent to the Fourier transform of the sequence $a(j)$ being non-negative and therefore is easy to check (Percival and Constantine, 2006). In our setting, we can choose estimators of the spectral density function which guarantee this, avoiding the problem.

Given a non-negative definite circulant embedding A , Dietrich and Newsam (1997) suggest performing the following procedure:

1. For $k \in \{0, \dots, m-1\}$
 - (a) Compute $S(k) = \sum_{j=0}^{m-1} a(j)e^{-i2\pi jk/m}$.
 - (b) Compute $\sqrt{\frac{1}{m}S(k)}Z(k)$ where $Z(k) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_c(0, 1)$.
2. Form $Y(t) = \sum_{k=0}^{m-1} Z(k)e^{-i2\pi tk/m}$ for $t \in \{0, \dots, m-1\}$.
3. Set $X^{(1)}(t\Delta) = \mathcal{R}(Y(t))$ and $X^{(2)}(t\Delta) = \mathcal{I}(Y(t))$ for $t \in \{0, \dots, n-1\}$.

Here $\mathcal{N}_c(0, 1)$ denotes a standard complex Gaussian random variable, i.e. $Z \sim \mathcal{N}_c(0, 1)$ if and only if $Z = X + iY$ where $X, Y \sim \mathcal{N}(0, 1/2)$. The obtained series $X^{(1)}$ and $X^{(2)}$ will both be realisations of a multivariate Gaussian distribution with covariance matrix A (Dietrich and Newsam, 1997). Furthermore, if Σ has been embedded as described before, the first N observations of $X^{(1)}$ and $X^{(2)}$ form two independent realisations from a multivariate Gaussian with covariance matrix Σ (and mean $\mathbf{0}$).

C.1.2 Multivariate circulant embedding

Circulant embedding for simulating multivariate Gaussian processes was first introduced by Chan and Wood (1999) in general case of multivariate Gaussian random fields, but (Helgason et al., 2011) give a more detailed account for the case of multivariate Gaussian processes, so we follow their approach here. The approach provides a fast and exact simulation of regularly sampled stationary Gaussian processes given some specified covariance structure. Let \underline{X} be the d -dimensional process from which we wish to simulate, and write

$$\Sigma_{rs} = \begin{bmatrix} c_{X_r X_s}(0) & c_{X_r X_s}(-\Delta) & \dots & c_{X_r X_s}(-[n-1]\Delta) \\ c_{X_r X_s}(\Delta) & c_{X_r X_s}(0) & \dots & c_{X_r X_s}(-[n-2]\Delta) \\ \vdots & \vdots & \ddots & \vdots \\ c_{X_r X_s}([n-1]\Delta) & c_{X_r X_s}([n-2]\Delta) & \dots & c_{X_r X_s}(0) \end{bmatrix}$$

for $1 \leq r \leq s \leq d$. These are embedded in larger circulant matrices analogously to the univariate case by setting

$$A_{rs} = \begin{bmatrix} a_{rs}(0) & a_{rs}(1) & a_{rs}(2) & \dots & a_{rs}(m-1) \\ a_{rs}(m-1) & a_{rs}(0) & a_{rs}(1) & \dots & a_{rs}(m-2) \\ a_{rs}(m-2) & a_{rs}(m-1) & a_{rs}(0) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a_{rs}(1) \\ a_{rs}(1) & a_{rs}(2) & \dots & a_{rs}(m-1) & a_{rs}(0) \end{bmatrix}$$

where for $1 \leq r \leq s \leq d$

$$a_{rs}(j) = \begin{cases} c_{X_r X_s}(-j\Delta) & 0 \leq j \leq n, \\ c_{X_r X_s}([m-j]\Delta) & m - (n-1) \leq j \leq m-1, \\ b_{rs}(j) & \text{otherwise.} \end{cases}$$

For $r > s$, set $\mathbf{A}_{rs} = \mathbf{A}_{sr}^T$. Note that this is different from the expression in Helgason et al. (2011) because we defined autocovariance as lagged covariance rather than led covariance, i.e. we define $c_{rs}(\tau) = \mathbb{E}[X_r(\tau)X_s(0)]$ whereas Helgason et al. (2011) use $\mathbb{E}[X_r(0)X_s(\tau)]$. The difference is essentially a matter of preference. (In the frequency domain, this difference corresponds to taking conjugates.)

Helgason et al. (2011) show that the following procedure simulates from a process with covariance and cross covariance matrices given by \mathbf{A}_{rs} .

1. For $k \in \{0 \dots m-1\}$:
 - (a) Compute $S_{rs}(k) = \sum_{j=0}^{m-1} a_{rs}(j) \exp\{-i2\pi jk/m\}$.
 - (b) Compute $\underline{\mathbf{C}}(k) = [2S_{rs}(k)/(S_{rr}(k)S_{ss}(k))^{1/2}]_{1 \leq r,s \leq d}$, setting terms equal to zero if $S_{rr}(k)S_{ss}(k) = 0$.
 - (c) Compute the Cholesky decomposition $\underline{\mathbf{L}}(k)\underline{\mathbf{L}}(k)^H = \underline{\mathbf{C}}(k)$.
 - (d) Set $\underline{\mathbf{Z}}(k) = \underline{\mathbf{L}}(k)\underline{\mathbf{W}}(k)$ where $\underline{\mathbf{W}}(k) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_c(0, 1)$.
2. Form $\underline{\mathbf{Y}}(t) = \sum_{k=0}^{m-1} \underline{\mathbf{Z}}(k) \exp\{-i2\pi tk/m\}$ for $t \in \{0, \dots, m-1\}$.
3. Set $\underline{\mathbf{X}}^{(1)}(t\Delta) = \mathcal{R}(\underline{\mathbf{Y}}(t))$ and $\underline{\mathbf{X}}^{(2)}(t\Delta) = \mathcal{I}(\underline{\mathbf{Y}}(t))$ for $t \in \{0, \dots, n-1\}$.

As in the univariate case, $\underline{\mathbf{X}}^{(1)}$ and $\underline{\mathbf{X}}^{(2)}$ are independent Gaussian processes with mean zero and covariance/cross-covariance matrices $\underline{\Sigma}_{rs}$.

C.1.3 Circulant embedding from non-parametric estimates

We describe the simulation from non-parametric estimates in the multivariate case, as proposed by Chandna and Walden (2013), because the univariate version (Percival and Constantine, 2006) is analogous. We wish to simulate Gaussian processes with

the same covariance structure as some observed process. Let $\underline{\hat{f}}(\omega)$ denote some consistent estimate of the spectral density matrix function. We can write

$$\underline{\hat{f}}(\omega) = \frac{\Delta}{2\pi} \sum_{\tau=-n+1}^{n-1} \underline{\hat{c}}(\tau) e^{-i\omega\tau}.$$

where $\underline{\hat{c}}(\tau)$ is an estimate of the autocovariance. We aim to simulate a Gaussian process with autocovariance given by $\underline{\hat{c}}(\tau)$ for $\tau \in \{-(n-1)\Delta, \dots, (n-1)\Delta\}$ and 0 otherwise. Taking the algorithm due to Helgason et al. (2011) above and setting $m = 2n$, we have

$$\begin{aligned} S_{rs}(k) &= \sum_{j=0}^{m-1} a_{rs}(j) e^{-i2\pi jk/m} \\ &= \sum_{j=0}^n \hat{c}_{rs}(-j\Delta) e^{-i2\pi jk/m} + \sum_{j=n+1}^{m-1} \hat{c}_{rs}((m-j)\Delta) e^{-i2\pi jk/m} \\ &= \sum_{j=0}^{n-1} \hat{c}_{rs}(-j\Delta) e^{-i2\pi jk/m} + \sum_{j=-(n-1)}^{-1} \hat{c}_{rs}(-j\Delta) e^{-i2\pi jk/m} e^{-i2\pi k} \\ &= \sum_{j=-(n-1)}^{n-1} \hat{c}_{rs}(-j\Delta) e^{-i2\pi jk/m} \\ &= \frac{2\pi}{\Delta} \overline{\hat{f}_{rs}\left(\frac{2\pi k}{m\Delta}\right)}. \end{aligned}$$

So setting

$$S_{rs}(k) = \frac{2\pi}{\Delta} \overline{\hat{f}_{rs}\left(\frac{2\pi k}{m\Delta}\right)}$$

and running the circulant embedding algorithm will produce simulations from the desired Gaussian process. Additionally, from Theorem 2.1 of Helgason et al. (2011), the embedding will be valid (will produce a process with non-negative definite covariance matrices) if and only if the matrices $\mathbf{S}(k) = [S_{rs}(k)]_{1 \leq r, s \leq d}$ are non-negative

definite for all $k \in \{0, \dots, m-1\}$. Note that

$$S(k) = \frac{2\pi}{\Delta} \overline{\hat{f}\left(\frac{2\pi k}{m\Delta}\right)} = \frac{2\pi}{\Delta} \hat{f}\left(\frac{2\pi k}{m\Delta}\right)^T$$

and so this is equivalent to the non-parametric spectral density estimate being non-negative definite for all of the Fourier frequencies. Multitaper estimates satisfy this property, and will be used in our simulations, as in [Chandna and Walden \(2013\)](#).

C.2 Derivation of univariate model

Recall that we have the model

$$W(t) = U(t) + V(t) + aU(t)V(t),$$

where U and V are univariate mean-zero stationary Gaussian processes which are independent of each other. We have that the autocovariance is given by

$$\begin{aligned} c_W(\tau) &= \mathbb{E}[W(\tau)W(0)] \\ &= \mathbb{E}[\{U(\tau) + V(\tau) + aU(\tau)V(\tau)\}\{U(0) + V(0) + aU(0)V(0)\}] \\ &= c_U(\tau) + c_V(\tau) + a^2c_U(\tau)c_V(\tau) \end{aligned}$$

because we have by independence that $\mathbb{E}[U(t)^m V(t)^n] = \mathbb{E}[U(t)^m] \mathbb{E}[V(t)^n]$ and $\mathbb{E}[U(t)] = \mathbb{E}[V(t)] = \mathbb{E}[U(t)^3] = \mathbb{E}[V(t)^3] = 0$. Therefore the spectral density function is

$$f_W(\omega) = f_U(\omega) + f_V(\omega) + a^2[f_U * f_V](\omega).$$

Similarly, the third-order cumulant function is

$$\begin{aligned} c_W(\tau_1, \tau_2) = & a[c_U(\tau_1)c_V(\tau_2) + c_U(\tau_2)c_V(\tau_1) + \\ & c_U(\tau_1 - \tau_2)c_V(\tau_1) + c_U(\tau_1)c_V(\tau_1 - \tau_2) + \\ & c_U(\tau_1 - \tau_2)c_V(\tau_2) + c_U(\tau_2)c_V(\tau_1 - \tau_2)] \end{aligned}$$

and so

$$\begin{aligned} f_W(\omega_1, \omega_2) = & a[f_U(\omega_1)f_V(\omega_2) + f_U(\omega_2)f_V(\omega_1) + \\ & f_U(-\omega_2)f_V(\omega_1 + \omega_2) + f_U(\omega_1 + \omega_2)f_V(-\omega_2) + \\ & f_U(\omega_1)f_V(\omega_1 + \omega_2) + f_U(\omega_1 + \omega_2)f_V(\omega_1)]. \end{aligned}$$

C.3 Additional simulation studies

In this appendix, we provide the results of additional simulation studies to accompany those in Chapter 6. Firstly, Figure C.3.1 shows models of the type Case 2, but with varying γ parameters taking values 1, 3.5 and 5, in both the swell and wind-sea. Secondly, Figure C.3.3 shows proportion of rejections using different levels of multiple testing correction, accompanying Figure 6.3.6.

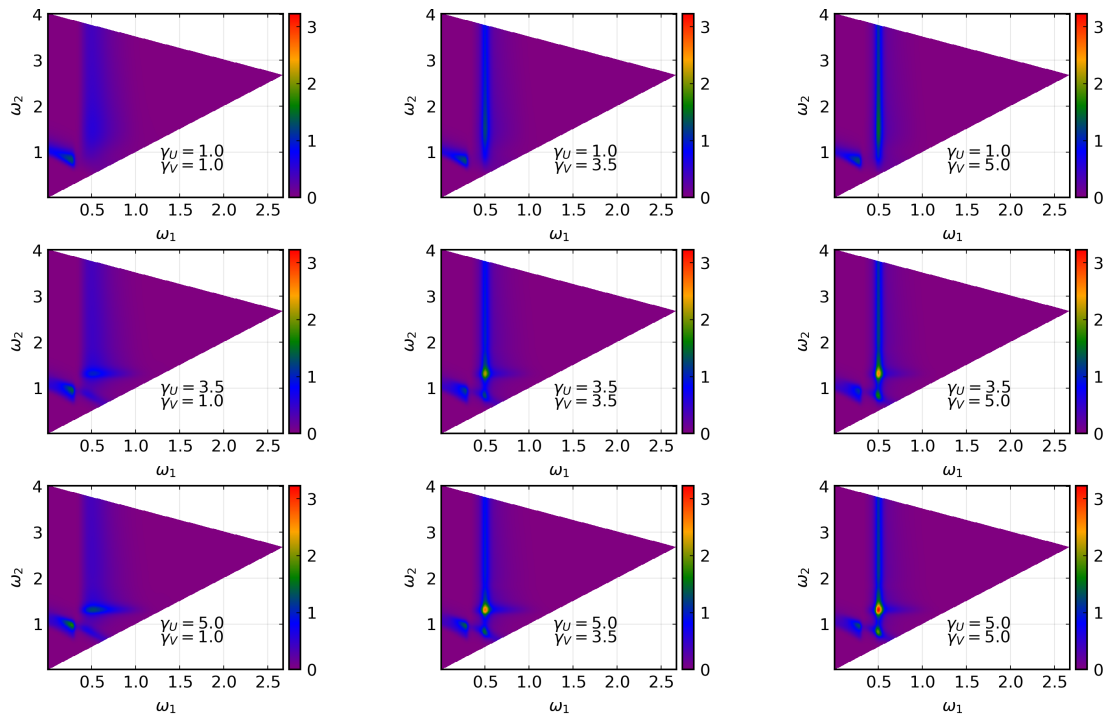


Figure C.3.1: Squared skewness function for univariate processes with $a = 1$ but with varying γ parameters.

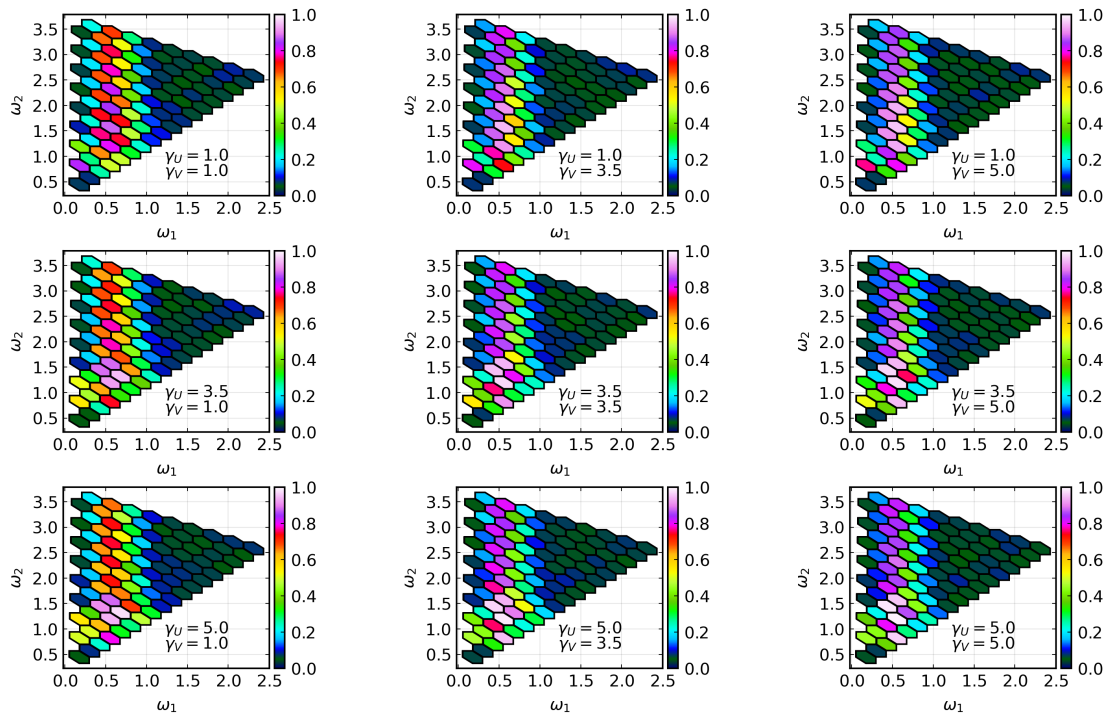


Figure C.3.2: Results of simulation study with the models given in Figure C.3.1.

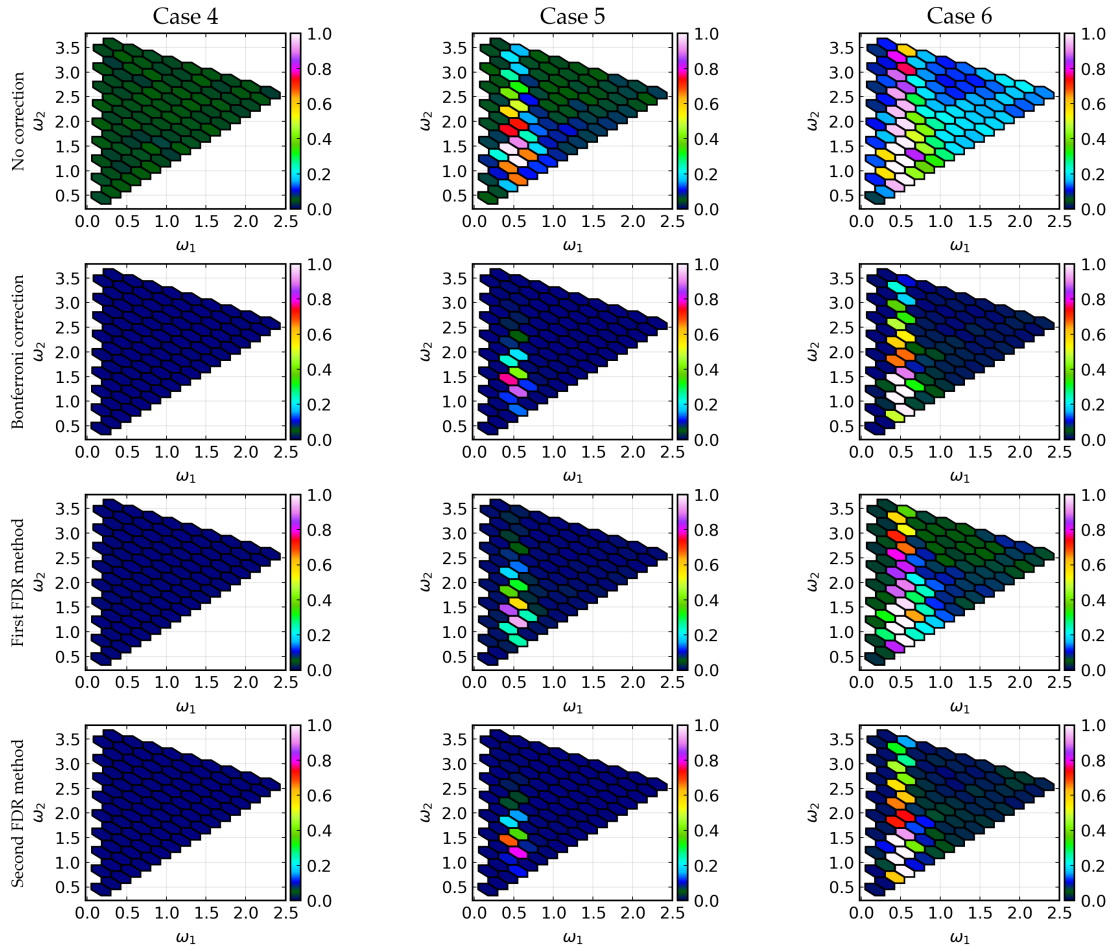


Figure C.3.3: Proportion of rejections with different correction levels using circulant embedding to generate under the null for multivariate processes.

Bibliography

- Airy, G. (1845). Tides and waves. In *Encyclopædia Metropolitana*, pages 241–396. B. Fellowes, London.
- Ashley, R. A., Patterson, D. M., and Hinich, M. J. (1986). A diagnostic test for non-linear serial dependence in time series fitting errors. *Journal of Time Series Analysis*, 7(3):165–178.
- Babanin, A. (2010). Wind input, nonlinear interactions and wave breaking at the spectrum tail of wind-generated waves, transition from f^{-4} to f^{-5} behaviour. *Ecological Safety of Coastal and Shelf Zones and Comprehensive Use of Shelf Resources*, pages 173–187.
- Banner, M. L. (1990). Equilibrium spectra of wind waves. *Journal of Physical Oceanography*, 20:966–984.
- Banner, M. L. and Young, I. R. (1994). Modeling spectral dissipation in the evolution of wind waves. Part I: assessment of existing model performance. *Journal of Physical Oceanography*, 24:1550–1571.
- Barstow, S. F., Bidlot, J.-R., Caires, S., Donelan, M. A., Drennan, W. M., Dupuis, H., Graber, H. C., Green, J. J., Gronlie, O., Guérin, C., et al. (2005). *Measuring and analysing the directional spectrum of ocean waves*. COST Office.

-
- Bartlett, M. S. (1948). Smoothing periodograms from time-series with continuous spectra. *Nature*, 161(4096):686–687.
- Battjes, J. A., Zitman, T. J., and Holthuisen, L. H. (1987). A reanalysis of the spectra observed in JONSWAP. *Journal of Physical Oceanography*, 17:1288–1295.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*, 57(1):289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, pages 1165–1188.
- Benoit, M., Frigaard, P., and Schäffer, H. (1997). Analyzing multidirectional wave spectra: a tentative classification of available methods. In *Proceedings of the IAHR Seminar on Multidirectional Waves and their Interactions with Structures*, pages 131–158.
- Berg, A., Paparoditis, E., and Politis, D. N. (2010). A bootstrap test for time series linearity. *Journal of Statistical Planning and Inference*, 140(12):3841–3857.
- Birkelund, Y. and Hanssen, A. (1999). Multitaper estimators for bispectra. In *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics. SPW-HOS'99*, pages 207–211. IEEE.
- Birkelund, Y. and Hanssen, A. (2009). Improved bispectrum based tests for Gaussianity and linearity. *Signal Processing*, 89(12):2537–2546.
- Birkelund, Y., Hanssen, A., and Powers, E. J. (2003). Multitaper estimators of polyspectra. *Signal Processing*, 83(3):545–559.
- Blackman, R. B. and Tukey, J. W. (1958). The measurement of power spectra from the

-
- point of view of communications engineering—Part I. *Bell System Technical Journal*, 37(1):185–282.
- Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, 60(2):217–226.
- Brillinger, D. R. (1965). An introduction to polyspectra. *The Annals of Mathematical Statistics*, pages 1351–1374.
- Brillinger, D. R. (1974). *Time Series: Data Analysis and Theory*. International series in decision processes. Holt, Rinehart, and Winston, New York.
- Brillinger, D. R. (1991). Some history of the study of higher-order moments and spectra. *Statistica Sinica*, pages 465–476.
- Brillinger, D. R. and Rosenblatt, M. (1967a). Asymptotic theory of estimates of k -th order spectra. In Harris, B., editor, *Spectral Analysis of Time Series*, pages 153–188. Wiley.
- Brillinger, D. R. and Rosenblatt, M. (1967b). Computation and interpretation of k -th order spectra. In Harris, B., editor, *Spectral Analysis of Time Series*, pages 189–232. Wiley.
- Brockwell, P. J. and Davis, R. A. (2006). *Time Series : Theory and Methods*. Springer, second edition.
- Capon, J., Greenfield, R., and Kolker, R. (1967). Multidimensional maximum-likelihood processing of a large aperture seismic array. *Proceedings of the IEEE*, 55:192–211.
- Chan, G. and Wood, A. T. (1999). Simulation of stationary Gaussian vector fields. *Statistics and computing*, 9(4):265–268.

-
- Chandna, S. and Walden, A. T. (2013). Simulation methodology for inference on physical parameters of complex vector-valued signals. *IEEE Transactions on Signal Processing*, 61(21):5260–5269.
- Chandran, V., Elgar, S., and Vanhoff, B. (1994). Statistics of tricoherence. *IEEE Transactions on Signal Processing*, 42(12):3430–3440.
- Cherneva, Z. and Soares, G. (2007). Estimation of the bispectra and phase distribution of storm sea states with abnormal waves. *Ocean Engineering*, 34:2009–2020.
- Chiu, S.-T. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *The Annals of Statistics*, 16(3):1315–1326.
- Christou, M. and Ewans, K. (2014). Field measurements of rogue water waves. *Journal of Physical Oceanography*, 44(9):2317–2335.
- Committee on Climate Change (2018). Managing the coast in a changing climate. *Committee on Climate Change*.
- Constable, C. G. (1988). Parameter estimation in non-Gaussian noise. *Geophysical Journal International*, 94(1):131–142.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301.
- Cramer, H. and Leadbetter, M. (1967). *Stationary and Related Stochastic Processes-Sample Function Properties and their Applications*. John Wiley and Sons, Inc.,.
- Dahlhaus, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *The Annals of Statistics*, 16(2):808–841.
- Dahlhaus, R. (2012). Locally Stationary Processes. In *Handbook of Statistics*, volume 30, pages 351–413. Elsevier B.V.

-
- Dalle Molle, J. W. and Hinich, M. J. (1995). Trispectral analysis of stationary random time series. *The Journal of the Acoustical Society of America*, 97(5):2963–2978.
- Datawell, B. (2006). Datawell waverider reference manual. *Datawell BV, Zomerluststraat 4, 2012 LM Haarlem, The Netherlands*.
- David, H. A. and Nagaraja, H. N. (2004). *Order statistics*. John Wiley & Sons.
- Davies, R. B. and Harte, D. S. (1987). Tests for Hurst effect. *Biometrika*, 74(1):95–101.
- De Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York.
- Dietrich, C. R. and Newsam, G. N. (1997). Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal of Scientific Computing*, 18(4):1088–1107.
- DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4):1055–1071.
- Donelan, M. A., Hamilton, J., and Hui, W. H. (1985). Directional spectra of wind-generated waves. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 315:509–562.
- Douglas, B. C. and Cheney, R. E. (1990). Geosat: Beginning a new era in satellite oceanography. *Journal of Geophysical Research: Oceans*, 95(C3):2833–2836.
- Ducrozet, G., Bonnefoy, F., Le Touzé, D., and Ferrant, P. (2016). Hos-ocean: Open-source solver for nonlinear waves in open ocean based on high-order spectral method. *Computer Physics Communications*, 203:245–254.
- Dzharidze, K. and Yaglom, A. (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics*, volume 4, pages 1–96. Elsevier.

-
- Eastoe, E., Koukoulas, S., and Jonathan, P. (2013). Statistical measures of extremal dependence illustrated using measured sea surface elevations from a neighbourhood of coastal locations. *Ocean Engineering*, 62:68–77.
- Elgar, S. and Guza, R. T. (1985). Observations of bispectra of shoaling surface gravity waves. *Journal of Fluid Mechanics*, 161:425–448.
- Elgar, S., Herbers, T. H. C., Chandran, V., and Guza, R. T. (1995). Higher-order spectral analysis of nonlinear ocean surface gravity waves. *Journal of Geophysical Research*, 100:4977–4983.
- Ewans, K., Christou, M., Ilic, S., and Jonathan, P. (2021). Identifying higher-order interactions in wave time-series. *Journal of Offshore Mechanics and Arctic Engineering*, 143(2).
- Ewans, K. and McConochie, J. (2018). OMAE2018-78386: On the uncertainties of estimating JONSWAP spectrum peak parameters. In *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering*, volume 3, Madrid, Spain. American Society of Mechanical Engineers (ASME).
- Ewans, K. C. (1998). Observations of the directional spectrum of fetch-limited waves. *Journal of Physical Oceanography*, 28(3):495–512.
- Ewans, K. C. (2002). Directional spreading in ocean swell. In *Ocean Wave Measurement and Analysis (2001)*, pages 517–529.
- Ewans, K. C. and Kibblewhite, A. C. (1986). Spectral characteristics of ocean waves undergoing generation in New Zealand waters. In *9th Australasian Fluid Mechanics Conference*, Auckland, New Zealand. Auckland Univ.
- Forristall, G. Z. (1981). Measurements of a saturated range in ocean wave spectra. *Journal of Geophysical Research*, 86(C9):8075–8084.

-
- Forristall, G. Z. (2004). On the use of directional wave criteria. *Journal of Waterway, Port, Coastal and Ocean Engineering*, 130(5):272–275.
- Forristall, G. Z. (2015). Maximum crest heights over an area: laboratory measurements compared to theory. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 56499, page V003T02A044. American Society of Mechanical Engineers.
- Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102(477):321–331.
- Geraci, M. (2019). Additive quantile regression for clustered data with an application to children’s physical activity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(4):1071–1089.
- Graham, A. (1981). *Kronecker Products and Matrix Calculus : with Applications*. Ellis Horwood series in mathematics and its applications. Horwood ; Halsted Press, Chichester : New York.
- Grainger, J. P. (2021). Software for spectral fitting. <https://github.com/JakeGrainger/SpectralFitting>.
- Grainger, J. P. (2022a). JakeGrainger/NorthSeaBuoyDisplacementAnalysis: v1.0.0. 10.5281/zenodo.6922892.
- Grainger, J. P. (2022b). JakeGrainger/OceanWaveSpectralFitting.jl: v0.3.0. 10.5281/zenodo.7042974.
- Grainger, J. P. (2022c). JakeGrainger/WhittleLikelihoodInference.jl: v0.3.0. 10.5281/zenodo.6922101.
- Grainger, J. P., Sykulski, A. M., Jonathan, P., and Ewans, K. (2021). Estimating the parameters of ocean wave spectra. *Ocean Engineering*, 229:108934.

-
- Greb, U. and Rusbridge, M. (1988). The interpretation of the bispectrum and bi-coherence for non-linear interactions of continuous spectra. *Plasma Physics and Controlled Fusion*, 30(5):537.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.
- Grenander, U. and Rosenblatt, M. (2008). *Statistical Analysis of Stationary Time Series*, volume 320. American Mathematical Soc.
- Guillaumin, A. P., Sykulski, A. M., Olhede, S. C., Early, J. J., and Lilly, J. M. (2017). Analysis of non-stationary modulated time series with applications to oceanographic surface flow measurements. *Journal of Time Series Analysis*, 38(5):668–710.
- Guillaumin, A. P., Sykulski, A. M., Olhede, S. C., and Simons, F. J. (2022). The de-biased spatial Whittle likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(4):1526–1557.
- Hannan, E. J. (2009). *Multiple Time Series*. John Wiley & Sons.
- Hanson, J. L. and Phillips, O. M. (2001). Automated analysis of ocean surface directional wave spectra. *Journal of Atmospheric and Oceanic Technology*, 18:277–293.
- Hasselmann, D. E., Dunckel, M., and Ewing, J. A. (1980). Directional wave spectra observed during JONSWAP 1973. *Journal of Physical Oceanography*, 10:1264–1280.
- Hasselmann, K. (1962). On the non-linear energy transfer in a gravity-wave spectrum Part 1. General theory. *Journal of Fluid Mechanics*, 12:481–500.
- Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Müller, P., Olbers, D., Richter, K., Sell, W., and Walden, H. (1973). Measurements of wind-wave

- growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Ergänzungsheft* 8-12.
- Haubrich, R. A. (1965). Earth noise, 5 to 500 millicycles per second: 1. Spectral stationarity, normality, and nonlinearity. *Journal of Geophysical Research*, 70:1415–1427.
- He, H. and Thomson, D. J. (2009). The canonical bicoherence—part I: Definition, multitaper estimation, and statistics. *IEEE Transactions on Signal Processing*, 57(4):1273–1284.
- He, H. and Thomson, D. J. (2010). Canonical bicoherence analysis of dynamic EEG data. *Journal of Computational Neuroscience*, 29(1):23–34.
- Helgason, H., Pipiras, V., and Abry, P. (2011). Fast and exact synthesis of stationary multivariate gaussian time series using circulant embedding. *Signal Processing*, 91(5):1123–1133.
- Hinich, M. J. (1982). Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3(3):169–176.
- Hinich, M. J., Mendes, E. M., and Stone, L. (2005). Detecting nonlinearity in time series: Surrogate and bootstrap approaches. *Studies in Nonlinear Dynamics & Econometrics*, 9(4).
- Hinich, M. J. and Wolinsky, M. (2005). Normalizing bispectra. *Journal of Statistical Planning and Inference*, 130(1-2):405–411.
- Holthuijsen, L. H. (2007). *Waves in Oceanic and Coastal Waters*. Cambridge University Press.
- Hwang, P. A., Fan, Y., Ocampo-Torres, F. J., and García-Nava, H. (2017). Ocean surface wave spectra inside tropical cyclones. *Journal of Physical Oceanography*, 47(10):2393–2417.

-
- Hwang, P. A., Walsh, E. J., Krabill, W. B., Swift, R. N., Manizade, S. S., Scott, J. F., and Earle, M. D. (1998). Airborne remote sensing applications to coastal wave research. *Journal of Geophysical Research: Oceans*, 103:18791–18800.
- Integrated Ocean Observing System (U.S.) (2019). Manual for real-time quality control of in-situ surface wave data version 2.1 : A guide to quality control and quality assurance of in-situ surface wave observations. Manuals & Handbooks.
- Isobe, M., Kondo, K., and Horikawa, K. (1984). Extension of MLM for estimating directional wave spectrum. In *Proc. Symp. on Description and Modeling of Directional Seas*, pages 1–15.
- Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139.
- Jammalamadaka, S. R., Rao, T. S., and Terdik, G. (2006). Higher order cumulants of random vectors and applications to statistical inference and time series. *Sankhyā: The Indian Journal of Statistics*, pages 326–356.
- Jensen, R., Swail, V., Lee, B., and O’Reilly, W. (2011). Wave measurement evaluation and testing. In *Proceedings of the 12th International Workshop on Wave Hindcasting and Forecasting (Kohala Coast, Hawaii)*.
- Kim, Y. and Powers, E. (1978). Digital bispectral analysis of self-excited fluctuation spectra. *The Physics of Fluids*, 21(8):1452–1453.
- Kim, Y. C. and Powers, E. J. (1979). Digital bispectral analysis and its applications to nonlinear wave interactions. *IEEE Transactions on Plasma Science*, 7(2):120–131.
- Koopmans, L. (1995). *The Spectral Analysis of Time Series*. Elsevier.

-
- Kreiss, J.-P. (1988). *Asymptotic Statistical Inference for a Class of Stochastic Processes*. Habilitationsschrift, Faculty of Mathematics, Univ. Hamburg, Germany.
- Kuik, A. J., van Vledder, G. P., and Holthuijsen, L. H. (1988). A method for the routine analysis of pitch-and-roll buoy wave data. *Journal of Physical Oceanography*, 18:1020–1034.
- Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.
- Leonov, V. P. and Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory of Probability & Its Applications*, 4(3):319–329.
- Lii, K. and Rosenblatt, M. (1990). Asymptotic normality of cumulant spectral estimates. *Journal of Theoretical Probability*, 3(2):367–385.
- Long, C. E. and Oltman-Shay, J. M. (1991). Directional characteristics of waves in shallow water. Technical report, Coastal Engineering Research Center, Vicksburg, MS.
- Longuet-Higgins, M. S. (1957). The statistical analysis of a random, moving surface. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 249(966):321–387.
- Longuet-Higgins, M. S., Cartwright, D. E., and Smith, N. D. (1963). Observations of the directional spectrum of sea waves using the motions of a floating buoy. In *Ocean Wave Spectra*, pages 111–136.
- Lygre, A. and Krogstad, H. E. (1986). Maximum entropy estimation of the directional distribution in ocean wave spectra. *Journal of Physical Oceanography*, 16:2052–2060.
- Mackay, E. B. L. (2016). A unified model for unimodal and bimodal ocean wave spectra. *International Journal of Marine Energy*, 15:17–40.

-
- Mérigaud, A. and Ringwood, J. V. (2018). Free-surface time-series generation for wave energy applications. *IEEE Journal of Oceanic Engineering*, 43(1):19–35.
- Mitsuyasu, H., Tasai, F., Suhara, T., Mizuno, S., Ohkusu, M., Honda, T., and Rikishi, K. (1975). Observations of the directional spectrum of ocean waves using a cloverleaf buoy. *Journal of Physical Oceanography*, 5:750–760.
- Mogensen, P. K. and Riseth, A. N. (2018). Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615.
- Nikias, C. L. (1993). *Higher-order Spectra Analysis : a Nonlinear Signal Processing Framework*. Prentice Hall signal processing series. PTR Prentice Hall, Englewood Cliffs, N.J.
- Nikias, C. L. and Raghuveer, M. R. (1987). Bispectrum estimation: A digital signal processing framework. *Proceedings of the IEEE*, 75(7):869–891.
- Nordman, D. J. and Lahiri, S. N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Annals of Statistics*, 34(6):3019–3050.
- Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):93–120.
- Olagnon, M., Ewans, K., Forristall, G., and Prevosto, M. (2013). West Africa swell spectral shapes. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 55331, page V02BT02A029. American Society of Mechanical Engineers.
- Olhede, S. C. and Ombao, H. (2013). Modeling and estimation of covariance of replicated modulated cyclical time series. *IEEE Transactions on Signal Processing*, 61(8):1944–1957.

-
- O'Malley, M., Sykulski, A. M., Laso-Jadart, R., and Madoui, M.-A. (2021). Estimating the travel time and the most likely path from Lagrangian drifters. *Journal of Atmospheric and Oceanic Technology*, 38(5):1059 – 1073.
- Oscroft, S., Sykulski, A. M., and Early, J. J. (2020). Separating mesoscale and submesoscale flows from clustered drifter trajectories. *Fluids*, 6(1):14.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling And Inference Using Likelihood*. Oxford University Press.
- Percival, D. B. (1992). Simulating Gaussian random processes with specified spectra. *Computing Science and Statistics*, 24:534–538.
- Percival, D. B. (2006). Exact simulation of complex-valued Gaussian stationary processes via circulant embedding. *Signal Processing*, 86:1470–1476.
- Percival, D. B. and Constantine, W. L. (2006). Exact simulation of Gaussian time series from nonparametric spectral estimates with application to bootstrapping. *Statistics and Computing*, 16(1):25–35.
- Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications : Multitaper and Conventional Univariate Techniques*. Cambridge University Press, Cambridge.
- Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*, volume 4. Cambridge University Press.
- Phillips, O. M. (1958). The equilibrium range in the spectrum of wind-generated waves. *Journal of Fluid Mechanics*, 4:426–434.
- Phillips, O. M. (1985). Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *Journal of Fluid Mechanics*, 156:505–531.

-
- Pierson, W. J. and Moskowitz, L. (1964). A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *Journal of Geophysical Research*, 69(24):5181–5190.
- Press, H. and Tukey, J. W. (1956). *Power Spectral Methods Of Analysis And Their Application To Problems In Airplane Dynamics*. Bell Telephone System Monograph 2606.
- Raghuveer, M. R. (1994). Higher-order statistics: laying a myth to rest. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 5–8. IEEE.
- Riggs, G. A., Nogami, S., Koepke, M., Melnikov, A., Eliseev, L., Lysenko, S., Khabanov, P., Drabinskij, M., Kharchev, N., Kozachek, A., and et al. (2021). Bispectral analysis of broadband turbulence and geodesic acoustic modes in the t-10 tokamak. *Journal of Plasma Physics*, 87(3):885870301.
- Rivest, L.-P., Duchesne, T., Nicosia, A., and Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):445–463.
- Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268.
- Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.
- Rodríguez, G. and Soares, C. (1999). Uncertainty in the estimation of the slope of the high frequency tail of wave spectra. *Applied Ocean Research*, 21(4):207–213.
- Rosenblatt, M. (1980). Linear processes and bispectra. *Journal of Applied Probability*, 17:265–270.

-
- Schubert, M., Wu, Y., Tychsen, J., Dixen, M., Faber, M. H., Sørensen, J. D., and Jonathan, P. (2020). On the distribution of maximum crest and wave height at intermediate water depths. *Ocean Engineering*, 217:107485.
- Schulte, J. A. (2016). Wavelet analysis for non-stationary, nonlinear time series. *Non-linear Processes in Geophysics*, 23(4):257–267.
- Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35(4):1773–1801.
- Shooter, R., Ross, E., Ribal, A., Young, I. R., and Jonathan, P. (2022). Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Engineering*, 247:110647.
- Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430.
- Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., and Early, J. J. (2019). The debiased Whittle likelihood. *Biometrika*, 106:251–266.
- Sykulski, A. M., Olhede, S. C., Lilly, J. M., and Danioux, E. (2016). Lagrangian time series models for ocean surface drifter trajectories. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 65:29–50.
- Sykulski, A. M., Olhede, S. C., Lilly, J. M., and Early, J. J. (2017). Frequency-domain stochastic modeling of stationary bivariate or complex-valued signals. *IEEE Transactions on Signal Processing*, 65(12):3136–3151.
- Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28:39–58.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992). Testing

-
- for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94.
- Thompson, R. (1973). Generation of stochastic processes with given spectrum. *Util. Math.*
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.
- Thomson, D. J. (1989). Multi-window bispectrum estimates. In *Proc. IEEE Workshop on Higher-order Spectral Analysis*, pages 19–23.
- Tierney, P. J., Young, A., Clarke, N. D., and Duncan, M. J. (2016). Match play demands of 11 versus 11 professional football using global positioning system tracking: Variations across common playing formations. *Human Movement Science*, 49:1–8.
- Tjøstheim, D. (1994). Non-linear time series: a selective review. *Scandinavian Journal of Statistics*, pages 97–130.
- Toba, Y. (1973). Local balance in the air-sea boundary processes. *Journal of the Oceanographical Society of Japan*, 29:209–220.
- Toffoli, A., Onorato, M., Bitner-Gregersen, E., and Monbaliu, J. (2010). Development of a bimodal structure in ocean wave spectra. *Journal of Geophysical Research: Oceans*, 115(C3).
- Trench, W. F. (2013). *Introduction to Real Analysis*. Faculty Authored and Edited Books & CDs.
- Tucker, M. J., Challenor, P. G., and Carter, D. J. (1984). Numerical simulation of a random sea: a common error and its effect upon wave group statistics. *Applied Ocean Research*, 6(2):118–122.

-
- Tucker, M. J. and Pitt, E. G. (2001). *Waves in Ocean Engineering*. Elsevier.
- van Essen, S., Ewans, K., and McConochie, J. (2018). Wave buoy performance in short and long waves, evaluated using tests on a hexapod. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 51272, page V07BT06A001. American Society of Mechanical Engineers.
- van Zutphen, H. J., Jonathan, P., and Ewans, K. C. (2008). A generic method to model frequency-direction wave spectra for FPSO motions. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume Volume 1: Offshore Technology, pages 463–473.
- Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.
- Velasco, C. and Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, 95(452):1229–1243.
- Verbelen, R., Antonio, K., and Claeskens, G. (2018). Unravelling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(5):1275–1304.
- Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.
- Wang, D. W. and Hwang, P. A. (2001). Evolution of the bimodal directional distribution of ocean waves. *Journal of Physical Oceanography*, 31(5):1200–1221.
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73.

-
- Whittle, P. (1953a). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 15:125–139.
- Whittle, P. (1953b). Estimation and information in stationary time series. *Arkiv för Matematik*, 2(5):423–434.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksells boktryckeri-a.-b.
- Wong, W. (1997). Frequency domain tests of multivariate Gaussianity and linearity. *Journal of Time Series Analysis*, 18(2):181–194.
- Wong, W. K. (1993). *Some Contributions to Multivariate Stationary Non-linear Time Series*. The University of Manchester (United Kingdom).
- Wood, A. T. A. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3(4):409–432.
- Young, I., Verhagen, L., and Khatri, S. (1996). The growth of fetch limited waves in water of finite depth. Part 3. Directional spectra. *Coastal Engineering*, 29(1-2):101–121.
- Young, I. R., Verhagen, L. A., and Banner, M. L. (1995). A note on the bimodal directional spreading of fetch-limited wind waves. *Journal of Geophysical Research*, 100:773–778.