Estimating the parameters of ocean wave spectra

Jake P. Grainger [a],[*], Adam M. Sykulski [b], Philip Jonathan [b],[c], Kevin Ewans [d]

[a] *STOR-i Centre for Doctoral Training, Department of Mathematics and Statistics, Lancaster University, Lancaster, UK*
[b] *Department of Mathematics and Statistics, Lancaster University, Lancaster, UK*
[c] *Shell Research Ltd., London, UK*
[d] *MetOcean Research Ltd., New Plymouth, New Zealand*

ABSTRACT

Wind-generated waves are often treated as stochastic processes. There is particular interest in their spectral density functions, which are often expressed in some parametric form. Such spectral density functions are used as inputs when modelling structural response or other engineering concerns. Therefore, accurate and precise recovery of the parameters of such a form, from observed wave records, is important. Current techniques are known to struggle with recovering certain parameters, especially the peak enhancement factor and spectral tail decay. We introduce an approach from the statistical literature, known as the de-biased Whittle likelihood, and address some practical concerns regarding its implementation in the context of wind-generated waves. We demonstrate, through numerical simulation, that the de-biased Whittle likelihood outperforms current techniques, such as least squares fitting, both in terms of accuracy and precision of the recovered parameters. We also provide a method for estimating the uncertainty of parameter estimates. We perform an example analysis on a data-set recorded off the coast of New Zealand, to illustrate some of the extra practical concerns that arise when estimating the parameters of spectra from observed data.

1. Introduction

Due to the random nature of wind-generated 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, such as that of Bartlett (1948), have been developed. 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 JONSWAP to refer to the original formulation given by Hasselmann et al. (1973), with a tail decay of ω^{-5} (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 5 we demonstrate with simulated half hour records that estimates for the tail decay (using current estimation techniques) range from ω^{-3} to ω^{-6}, when the true tail decay is known to be ω^{-4}. Because there is too much variability in the estimates, it is impossible to determine from data of such lengths if the true tail decay is ω^{-4} or ω^{-5}. 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 ω^{-4} and ω^{-5} 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). 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, 1953). 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 de-biased version of the Whittle likelihood that corrects for this bias, without sacrificing the computational speed or making extra distributional assumptions. In Section 3.4, we will provide some intuition as to why we would expect the de-biased Whittle likelihood to perform better than least squares, both in terms of accuracy (bias) and precision (variance). Then in Section 5, we evidence this claim using numerical simulations.

The contributions of this paper are as follows. Firstly, we introduce the de-biased Whittle likelihood estimator for use on wind-generated wave processes (Section 3). Secondly, we detail practical concerns regarding the implementation of the estimation procedure for wind-generated ocean wave processes (Section 4), with accompanying MATLAB code provided on GitHub (Grainger, 2021). 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 de-biased Whittle likelihood estimator which can be used to quantify the uncertainty of parameter estimates (Section 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 (Sections 5 and 7).

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 paper 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 $X_\Delta = \{X_{t\Delta}\}_{t \in \mathbb{Z}}$.

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 $c(\tau) = \mathbb{E}[X_\tau X_0] - \mathbb{E}[X_\tau]\mathbb{E}[X_0]$. 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 removing the mean of the data we can obtain a process with the desired property. Therefore,

---

[1] To 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).

the autocovariance simplifies to $c(\tau) = \mathbb{E}\left[X_\tau X_0\right]$. 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. A formal construction of the spectral density function can be found in, for example, Brockwell and Davis (2006, page 118) (for the discrete time process), or Doob (1953, page 522) (for the continuous time process). However, for our purposes it suffices to note the following relations. Firstly for the discrete time process,

$$f_\Delta(\omega) = \frac{\Delta}{2\pi} \sum_{\tau=-\infty}^{\infty} c(\tau\Delta) \exp\{-i\omega\tau\Delta\}, \tag{1}$$

for $\omega \in [-\pi/\Delta, \pi/\Delta]$, where π/Δ is the *Nyquist frequency* and is the highest observable frequency of the sampled process. Secondly for the continuous time process,

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(\tau) \exp\{-i\omega\tau\}\mathrm{d}\tau, \tag{2}$$

for almost every $\omega \in \mathbb{R}$ (i.e. equal except on a set of measure zero).[2] Similarly, the inverse relations are

$$c(\tau\Delta) = \int_{-\pi/\Delta}^{\pi/\Delta} f_\Delta(\omega) \exp\{i\omega\tau\Delta\}\mathrm{d}\omega, \tag{3}$$

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

$$c(\tau) = \int_{-\infty}^{\infty} f(\omega) \exp\{i\omega\tau\}\mathrm{d}\omega, \tag{4}$$

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:

$$f_\Delta(\omega) = \sum_{k=-\infty}^{\infty} f\left(\omega + \frac{2\pi k}{\Delta}\right), \tag{5}$$

for $\omega \in [-\pi/\Delta, \pi/\Delta]$ (Percival and Walden, 1993, chapter 4). In Section 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.

2.1. 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 are used throughout this paper. 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 $\epsilon > 0$, $\mathbb{P}\left(|\hat{\theta}_N - \theta_0| > \epsilon\right) \to 0$ as $N \to \infty$.

The most basic non-parametric estimator for the spectral density function of a discrete time process is the *periodogram*. Let $X_{\Delta,N} = \{X_0, X_\Delta, \ldots, X_{\Delta(N-1)}\}$ be a series of N consecutive random variables from X_Δ, then the periodogram is defined as

$$I(\omega) = \frac{\Delta}{2\pi N} \left| \sum_{t=0}^{N-1} X_{\Delta t} \exp\{-it\Delta\omega\} \right|^2, \tag{}$$

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 $j = -\lceil N/2 \rceil + 1, \ldots, \lfloor N/2 \rfloor$. 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 window, 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

$$I_B(\omega) = \frac{\Delta}{2\pi PL} \sum_{p=1}^{P} \left| \sum_{t=(p-1)L+1}^{pL} X_{\Delta t} \exp\{-it\Delta\omega\} \right|^2, \tag{6}$$

where P is the number of segments and L is the number of observations in each segment (with $PL \leq N$).

2.2. Models for the sea surface

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 spectra, 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 parameterisation:

$$S_G(\omega \mid \theta) = \alpha\omega^{-r} \exp\left\{-\frac{r}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right\} \gamma^{\delta(\omega|\theta)}, \tag{7}$$

where

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

and

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

for $\omega > 0$; where $\alpha, \omega_p, s > 0$, $\gamma \geq 1$, $r > 1$[3] and θ denotes the vector of parameters. Typically, and for the remainder of this paper, 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 $\theta = (\alpha, \omega_p, \gamma, r)$. Also let Θ denote the *parameter space* — the set of possible values that θ 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 (7) 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 $f_G : \mathbb{R} \times \Theta \to [0, \infty)$,

$$f_G(\omega \mid \theta) = \begin{cases} S_G(\omega \mid \theta)/2 & \text{for } \omega > 0, \\ 0 & \text{for } \omega = 0, \\ S_G(-\omega \mid \theta)/2 & \text{for } \omega < 0. \end{cases} \tag{8}$$

---

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

[3] Mackay (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 $r > 1$.

We shall refer to the function defined by (8) as the generalised JON-SWAP spectral form. In this formulation, α is measured in units of $\mathrm{m^2\ s^{1-r}\ rad^{r-1}}$, ω_p in rad s^{-1} and γ and r are dimensionless. For convenience, we omit the units in future references.

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 θ. 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 $f(\omega \mid \theta)$ for the spectral density function of the continuous time process, given some choice of parameters θ. We shall also write $f_\Delta(\omega \mid \theta)$ and $c(\tau \mid \theta)$ 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 Σ_θ 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 paper.

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. As such, we shall refer to such approaches as least squares fitting techniques.

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

$$\ell_{LS}\left(\theta \mid X_{\Delta,N}\right) = \sum_{\omega \in \Omega} \left(f(\omega \mid \theta) - \bar{I}(\omega)\right)^2, \qquad (9)$$

where $\Omega \subseteq \Omega_{N,\Delta}$ (the choice of Ω is discussed in Section 4.1). In other words, the least squares estimator for θ is defined as $\hat{\theta}_{LS} = \mathrm{argmin}_{\theta \in \Theta}\,\ell_{LS}\left(\theta \mid X_{\Delta,N}\right)$. This approach could be adapted to account for aliasing by replacing $f(\omega \mid \theta)$ with $f_\Delta(\omega \mid \theta)$, 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 (9) 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 under-weighted. 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 paper. However, in simulations not shown in this paper, log least squares still does not perform as well as the de-biased 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).

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 Σ_θ. The log-likelihood function for observations of such a process is

$$\ell_{ML}\left(\theta \mid X_{\Delta,N}\right) = \frac{1}{2}\left(-N\log(2\pi) - \log|\Sigma_\theta| - X_{\Delta,N}^T \Sigma_\theta^{-1} X_{\Delta,N}\right), \qquad (10)$$

where $X_{\Delta,N}^T$ denotes the transpose of $X_{\Delta,N}$. The maximum likelihood estimator is then obtained by maximising the log-likelihood function. More formally, the maximum likelihood estimator of θ is $\hat{\theta}_{ML} = \mathrm{argmax}_{\theta \in \Theta}\,\ell_{ML}\left(\theta \mid X_{\Delta,N}\right)$. Provided that the underlying random variable is actually multivariate Gaussian, this technique will provide asymptotically optimal estimates of θ, 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 (10) 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 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.

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-likelihood[4] is known as the Whittle likelihood (Whittle, 1953). 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

$$\ell_W\left(\theta \mid X_{\Delta,N}\right) = -\sum_{\omega \in \Omega}\left\{\log\left(f(\omega \mid \theta)\right) + \frac{I(\omega)}{f(\omega \mid \theta)}\right\}, \qquad (11)$$

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 $\hat{\theta}_W = \mathrm{argmax}_{\theta \in \Theta}\,\ell_W\left(\theta \mid X_{\Delta,N}\right)$. This estimator also does not account for aliasing. However, by replacing $f(\omega \mid \theta)$ with $f_\Delta(\omega \mid \theta)$ in (11), 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 blurring are still present Percival and Walden (1993, chapter 6). Sykulski et al. (2019) introduced the de-biased Whittle likelihood to deal with both aliasing and blurring simultaneously. The de-biased Whittle likelihood is

$$\ell_{DW}\left(\theta \mid X_{\Delta,N}\right) = -\sum_{\omega \in \Omega}\left\{\log\left(\bar{f}_N(\omega \mid \theta)\right) + \frac{I(\omega)}{\bar{f}_N(\omega \mid \theta)}\right\}, \qquad (12)$$

---

[4] These 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.

where $\bar{f}_N(\omega \mid \theta) = \mathbb{E}[I(\omega) \mid \theta]$ 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:

$$\mathbb{E}[I(\omega) \mid \theta] = \frac{\Delta}{2\pi} \text{Re}\left(2 \sum_{\tau=0}^{N-1} \left(1 - \frac{\tau}{N} \right) c(\tau \mid \theta) \exp\{-i\omega\tau\Delta\} - c(0 \mid \theta) \right). \quad (13)$$

The resulting estimator can then be expressed as $\hat{\theta}_{DW} = \text{argmax}_{\theta \in \Theta} \ell_{DW}(\theta \mid X_{\Delta,N})$. Despite being constructed from the periodogram, an inconsistent estimator of the spectral density function, the de-biased Whittle likelihood is a consistent estimator of the parameters for the parametric model. The de-biased 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 de-biased Whittle estimator converges optimally. These condition are discussed further in Appendix C.

3.4. Comparison

In Section 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 in 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 de-biased 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 de-biased 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 parameter choice that would have been most likely to have given rise to the data in question. 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.[5]

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, $\bar{I}(\omega)$, 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 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 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.[6] 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 estimator 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 de-biased 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 de-biased 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 1 Hz data. Although, for higher sampling frequencies, such as 4 Hz 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).

3.5. Differencing

If the periodogram is highly correlated across frequencies, spectral-likelihoods will perform poorly when compared to full maximum

---

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

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

likelihood Velasco and Robinson (2000). Differencing can sometimes provide a convenient mechanism for removing such correlations. Define the differenced process as $Y_t = X_{t+\Delta} - X_t$. 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 $c_Y(\tau)$ and $f_Y(\omega)$ for the differenced process Y. First notice that

$$
\begin{aligned}
c_Y(\tau) &= \mathbb{E}\left[Y_0 Y_\tau\right] \\
&= \mathbb{E}\left[X_\Delta X_{\tau+\Delta} - X_0 X_{\tau+\Delta} - X_\Delta X_\tau + X_0 X_\tau\right] \\
&= 2c_X(\tau) - c(\tau + \Delta) - c(\tau - \Delta),
\end{aligned}
\tag{14}
$$

by stationarity. Then from (2) we can see that

$$
\begin{aligned}
f_Y(\omega) &= \int_{-\infty}^{\infty} \left(2c_X(\tau) - c(\tau + \Delta) - c(\tau - \Delta)\right) \exp\{-i\tau\omega\}\mathrm{d}\tau \\
&= 2f_X(\omega) - e^{i\omega\Delta} f_X(\omega) - e^{-i\omega\Delta} f_X(\omega) \\
&= 2(1 - \cos(\omega\Delta)) f_X(\omega) \\
&= 4\sin^2\left(\frac{\omega\Delta}{2}\right) f_X(\omega).
\end{aligned}
\tag{15}
$$

Therefore, differencing can be easily incorporated into the fitting techniques that have been discussed in this paper, by simply replacing X with the differenced process Y and f_X with f_Y using the relation given by (15). Consider the correlation matrix of the periodogram: the matrix with i,jth element defined to be the correlation between the periodogram at the ith and jth Fourier frequencies. Fig. 1 shows a plot of the correlation matrix for the periodogram of a wind-generated wave process, estimated using the technique described in Appendix B. We can see that for data recorded at a 1 Hz sampling rate there is little correlation in the periodogram; however, this is not the case for 4 Hz data. We can also see that the periodogram of the differenced process is almost completely uncorrelated, even for the 4 Hz data.[7] 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.

4. Practical concerns for implementation with the generalised JONSWAP

In Section 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 3. In particular, it is necessary for both the de-biased Whittle likelihood and for full maximum likelihood.

4.1. Frequency selection

Many of the estimators defined in Section 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 Fig. 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 de-biased Whittle likelihoods. The zero and Nyquist frequencies must be omitted (or a modified version of the summand must 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.

4.2. Numerical estimation of the autocovariance

To calculate both the multivariate Gaussian likelihood and de-biased Whittle likelihood we require the autocovariance of the process given a certain parameter choice (in (10) for the multivariate Gaussian likelihood and (13) for the de-biased 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 (4), and we wish to obtain the autocovariance at lags $0, \Delta, \ldots, (N-1)\Delta$. 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 Eq. (4), we can construct the approximate autocovariance

$$
\hat{c}(\tau) = \int_{-L\pi/\Delta}^{L\pi/\Delta} f(\omega) \exp\{i\tau\omega\}\mathrm{d}\omega,
\tag{16}
$$

---

[7] It should be noted that the region of high correlation in the top 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 4.1.
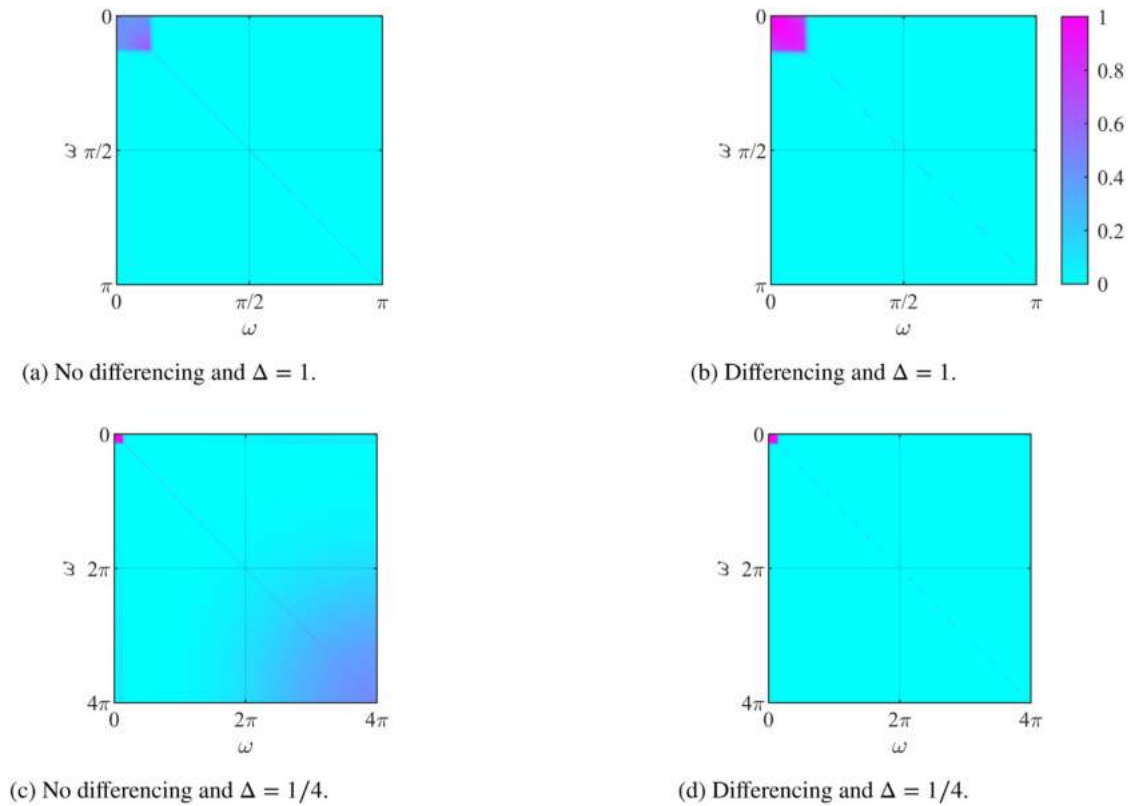
(a) No differencing and $\Delta = 1$.

(b) Differencing and $\Delta = 1$.

(c) No differencing and $\Delta = 1/4$.

(d) Differencing and $\Delta = 1/4$.

Fig. 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 $r = 5$.

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

Alternatively, we could consider the relation given in Eq. (3), 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 Eq. (5), between the spectral density of the continuous and discrete time processes. Consider the approximation of the aliased spectral density function given by

$$\tilde{f}_\Delta(\omega) = \sum_{k=-K}^{K} f\left(\omega + \frac{2\pi k}{\Delta}\right), \tag{17}$$

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

$$\tilde{c}(\tau) = \int_{-\pi/\Delta}^{\pi/\Delta} \tilde{f}_\Delta(\omega) \exp\{i\tau\omega\} d\omega. \tag{18}$$

Notice that we may write

$$
\begin{aligned}
\tilde{c}(\tau) &= \int_{-\pi/\Delta}^{\pi/\Delta} \sum_{k=-K}^{K} f\left(\omega + \frac{2\pi k}{\Delta}\right) \exp\{i\tau\omega\} d\omega \\
&= \sum_{k=-K}^{K} \int_{-\pi/\Delta}^{\pi/\Delta} f\left(\omega + \frac{2\pi k}{\Delta}\right) \exp\{i\tau\omega\} d\omega \\
&= \sum_{k=-K}^{K} \int_{(2k-1)\pi/\Delta}^{(2k+1)\pi/\Delta} f(\omega) \exp\{i\tau\omega\} d\omega \\
&= \int_{(-2K-1)\pi/\Delta}^{(2K+1)\pi/\Delta} f(\omega) \exp\{i\tau\omega\} d\omega.
\end{aligned}
\tag{19}
$$

From (19) 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 (16)) can be computed in $O(LM \log LM)$ time. However, computing the second approximation, based on $\tilde{c}(\tau)$ (in (18)), 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 (typically 10^{-6} m^2 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 approximation and should be tuned accordingly. For the generalised JONSWAP, we have found that $M = \max\{8192, 2N\}$ is a good choice.

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 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 $100/\Delta$.[8] For 1.28 Hz 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 de-biased Whittle likelihood and full time domain maximum likelihood.

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 3. For the purposes of the simulation study we let $\boldsymbol{X}_{\Delta,N}$ be a random variable with a multivariate normal distribution resulting from sub-sampling the continuous-time mean-zero stationary Gaussian process $\boldsymbol{X}$, where $\boldsymbol{X}$ has spectral density function $f_G(\omega \mid \theta)$, defined by (8). We then simulate a realisation of $\boldsymbol{X}_{\Delta,N}$ 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) and Wood and Chan (1994).

To perform the fitting we first choose one of the objective functions described in Section 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.

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, 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 1.28 Hz (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 Fig. 2, alongside the time taken to perform the optimisation. For comparison, the true value of each parameter is given by a horizontal black dashed line.
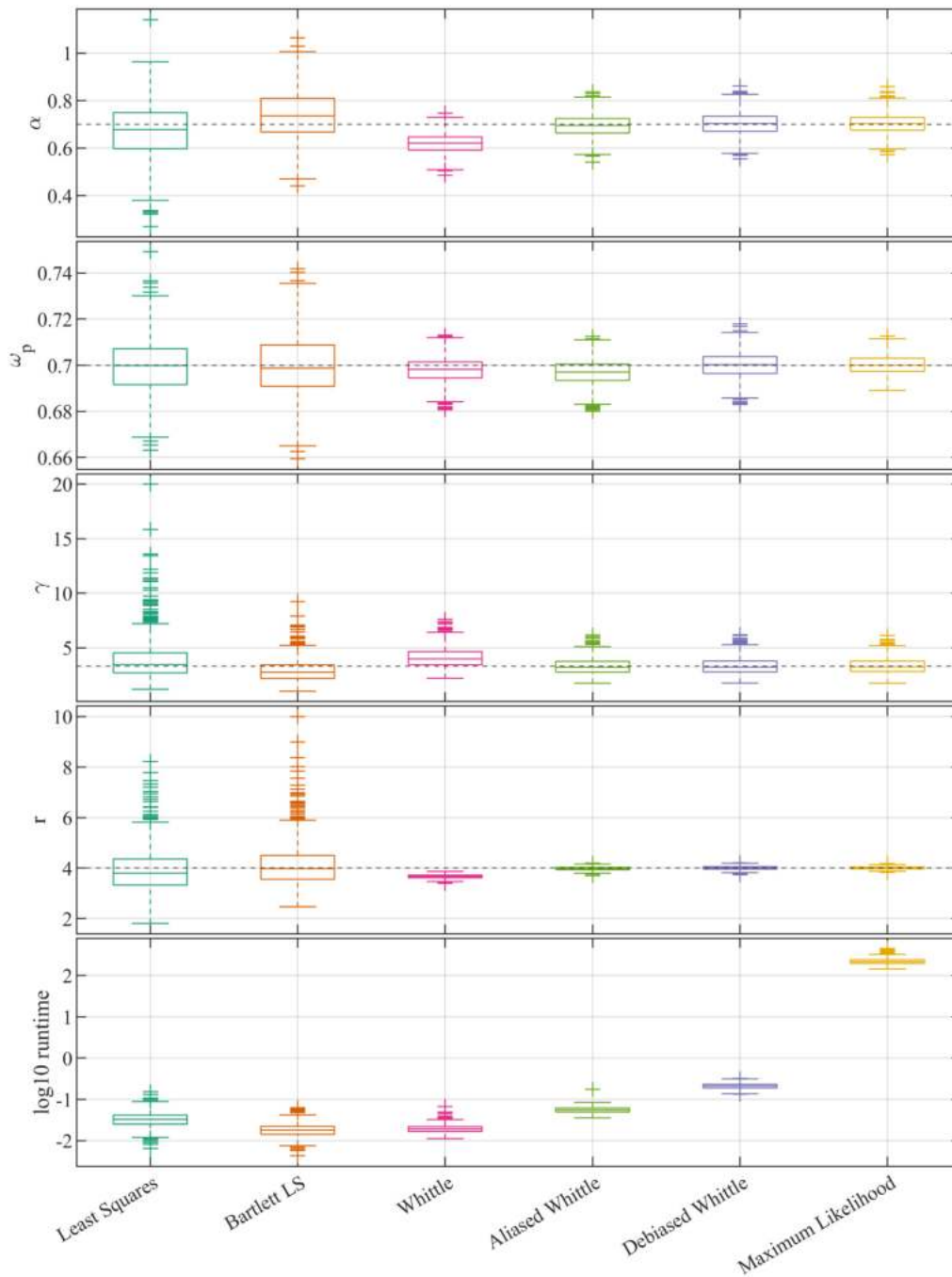
Perhaps the most striking feature of Fig. 2 is the difference in the variability in estimates of the tail parameter, r, when comparing least squares type techniques to likelihood based techniques. Least squares techniques recover parameter estimates ranging from well beyond three to five, making it very difficult to make any statements about the true value of the tail decay. However, we can see that statistical techniques such as the de-biased Whittle likelihood are able to recover the original tail parameter to within a few decimal places. Therefore, by using the de-biased 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 de-biased Whittle likelihood is generic and could be applied to other models provided they satisfied certain conditions (see Appendix C). We can also see that the de-biased 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 de-biased Whittle likelihood estimates. This verifies that the de-biased Whittle likelihood is indeed accounting for some of the bias present in standard Whittle likelihood, and demonstrates why de-biased Whittle likelihood is necessary over the aliased Whittle likelihood, which can still be seen to be biased for some parameters. In the analysis of Fig. 2 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 de-biased 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. Fig. 3 shows a scatter plot of the de-biased Whittle likelihood estimates from Fig. 2. 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 h records. Fig. 4 shows the comparison of least squares, Bartlett least squares and de-biased Whittle likelihood estimates for these 3 h records. The variance in the first two estimators has indeed decreased when compared to the estimates from half hour records shown in Fig. 2.

---

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

Fig. 2. Boxplots of parameter estimates and time taken for six different fitting routines with true parameters denoted by the dashed black 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 1.28 Hz observations recorded for half an hour (2304 observations). The process was repeated 1000 times.

However, by comparing the de-biased Whittle likelihood estimates in Fig. 2 to the least squares estimates in Fig. 4, we can see that the de-biased Whittle likelihood used on a half hour record yields better estimates than the least squares based estimates performed on 3 h 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 de-biased Whittle likelihood, we can obtain more accurate 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.
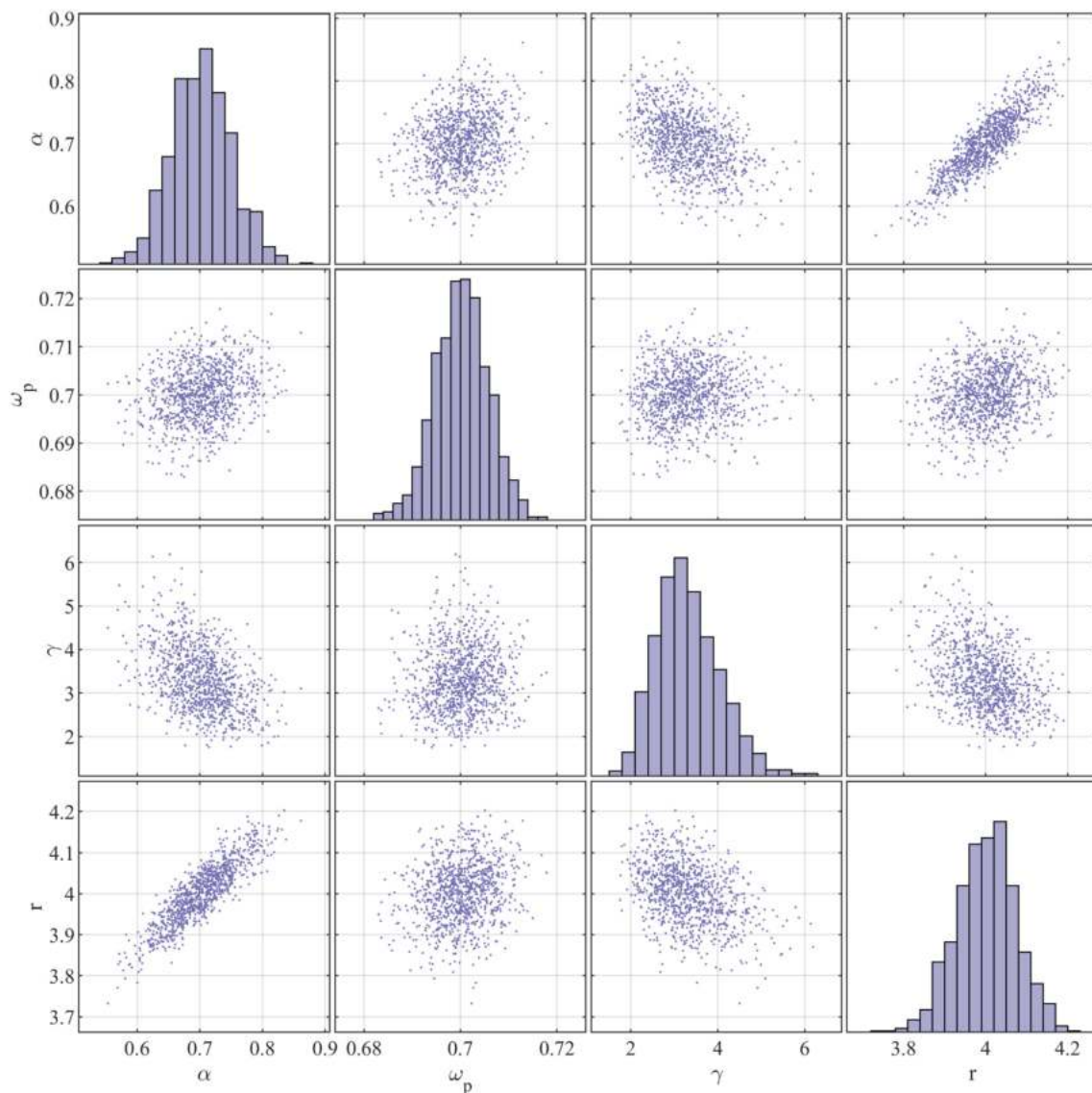
Fig. 3. Scatter plot of the de-biased Whittle estimates obtained from half hour records shown in Fig. 2.

5.3. Robustness of results

In Section 5.2 we have demonstrated that the de-biased 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 1, we present the results of a simulation study comparing least squares, Bartlett least squares and the de-biased 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 1 that there are substantial reductions in both the bias and standard deviation of all the estimated parameters.

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 de-biased Whittle likelihood techniques can be seen in Fig. 5. 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 de-biased 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 Fig. 5 that the de-biased 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 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
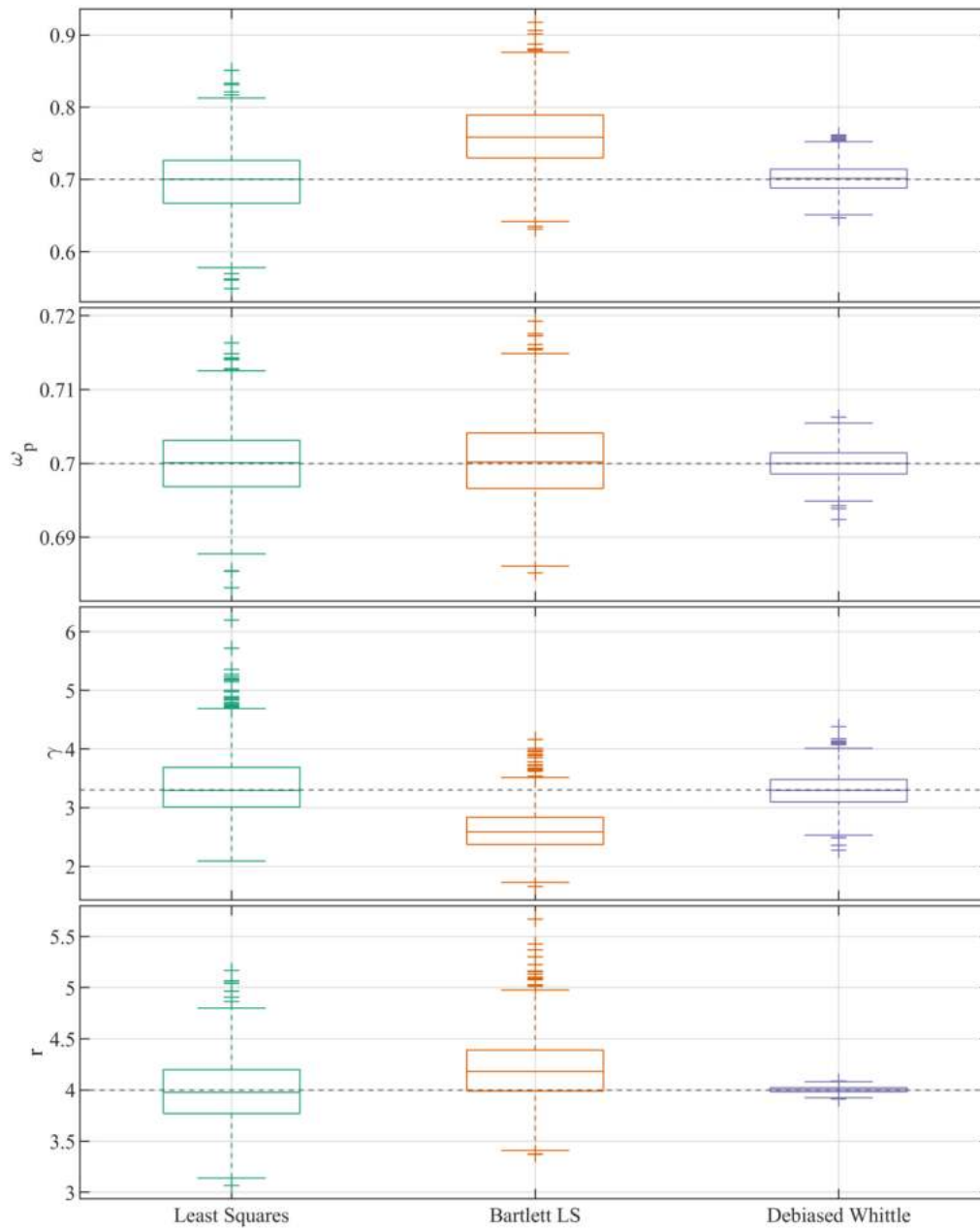
Fig. 4. Boxplots of 1000 parameter estimates using three different fitting routines on the same simulations using the same true parameters as Fig. 2 and same sampling interval but simulating 3 h records.

Table 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 de-biased 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 1.28 Hz. The bottom row shows the average of these quantities over all the parameters in the model.

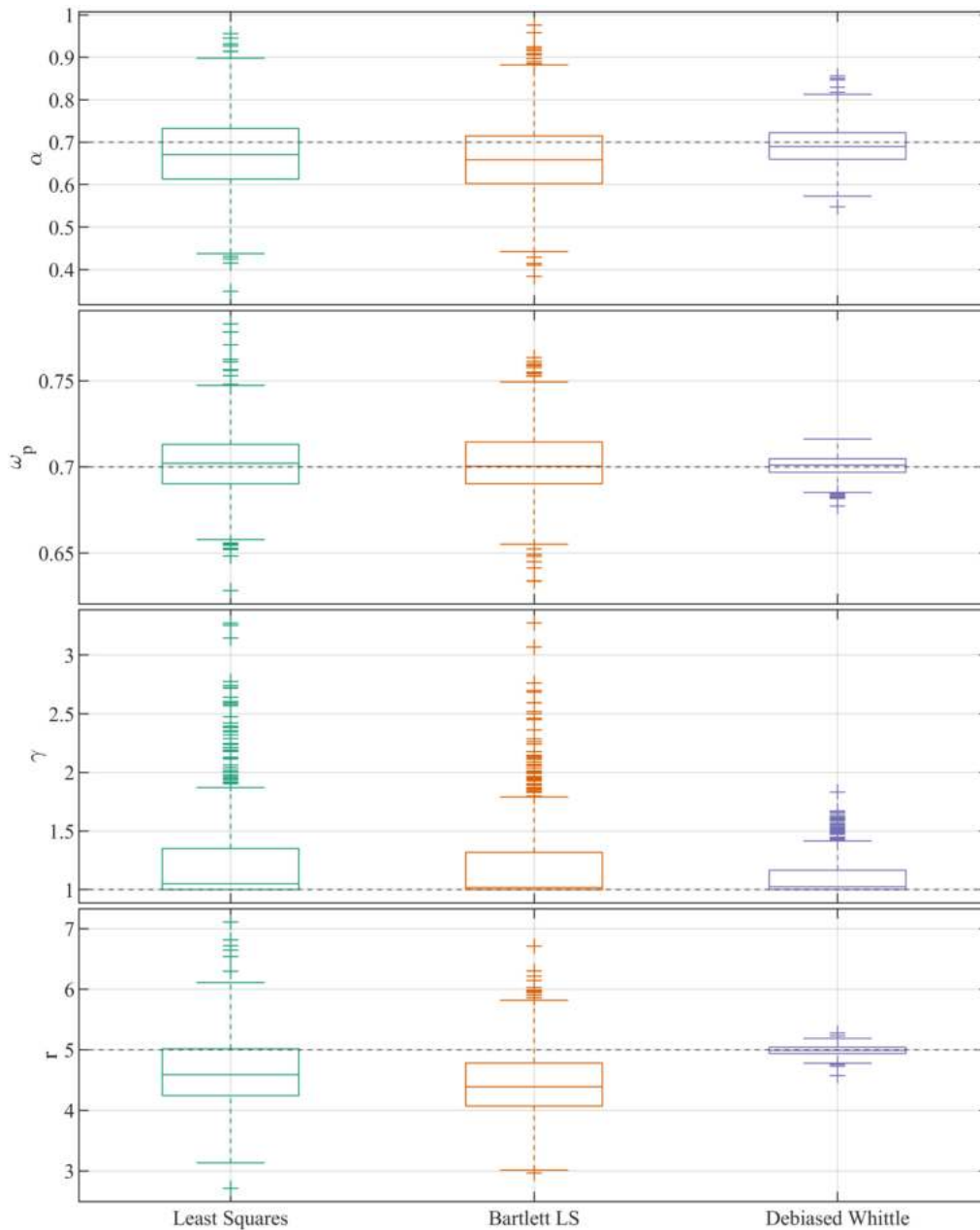|	Bias			SD			RMSE		
	LS	BLS	DW	LS	BLS	DW	LS	BLS	DW
α	4.70%	10.88%	0.80%	34.64%	30.69%	9.13%	35.46%	33.49%	9.19%
ω_p	0.08%	0.11%	0.06%	1.73%	1.85%	0.77%	1.74%	1.86%	0.77%
γ	15.88%	10.20%	2.98%	43.29%	29.27%	18.00%	46.39%	31.75%	18.71%
r	3.89%	5.60%	0.18%	17.57%	15.76%	2.09%	18.35%	17.44%	2.11%
Average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

Fig. 5. 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 1.28 Hz.

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.

5.4. Differencing for high sampling frequencies

As we discussed in Section 3.5, for wind generated waves observed at a 4 Hz sampling rate, the periodogram is highly correlated. As noted in Section 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. Fig. 6 shows box plots of estimated parameters for least squares, Bartlett least squares and de-biased Whittle likelihood techniques, both with and without differencing, for both 1 Hz and 4 Hz data (in this case, 2048 s worth of data was simulated per record, i.e. 2048 observations for the 1 Hz

simulation and 8192 for the 4 Hz simulation). We can see little benefit from differencing in the 1 Hz estimates (Fig. 6(a)). However, when it comes to 4 Hz data, there is a major benefit to differencing, especially for the de-biased Whittle likelihood. Because there is little difference between the de-biased Whittle fits on the original and differenced 1 Hz data, we would recommend that differencing is used as standard, to protect against the issues seen in Fig. 6(b). At the very least, investigating the correlation matrix of the periodogram should be an important diagnostic when fitting spectral models.

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 de-biased Whittle likelihood. Sykulski et al. (2019) developed an approach for quantifying the asymptotic variance of the estimator for a single

(a) Parameter estimates for 1Hz data.

(b) Parameter estimates for 4Hz data.
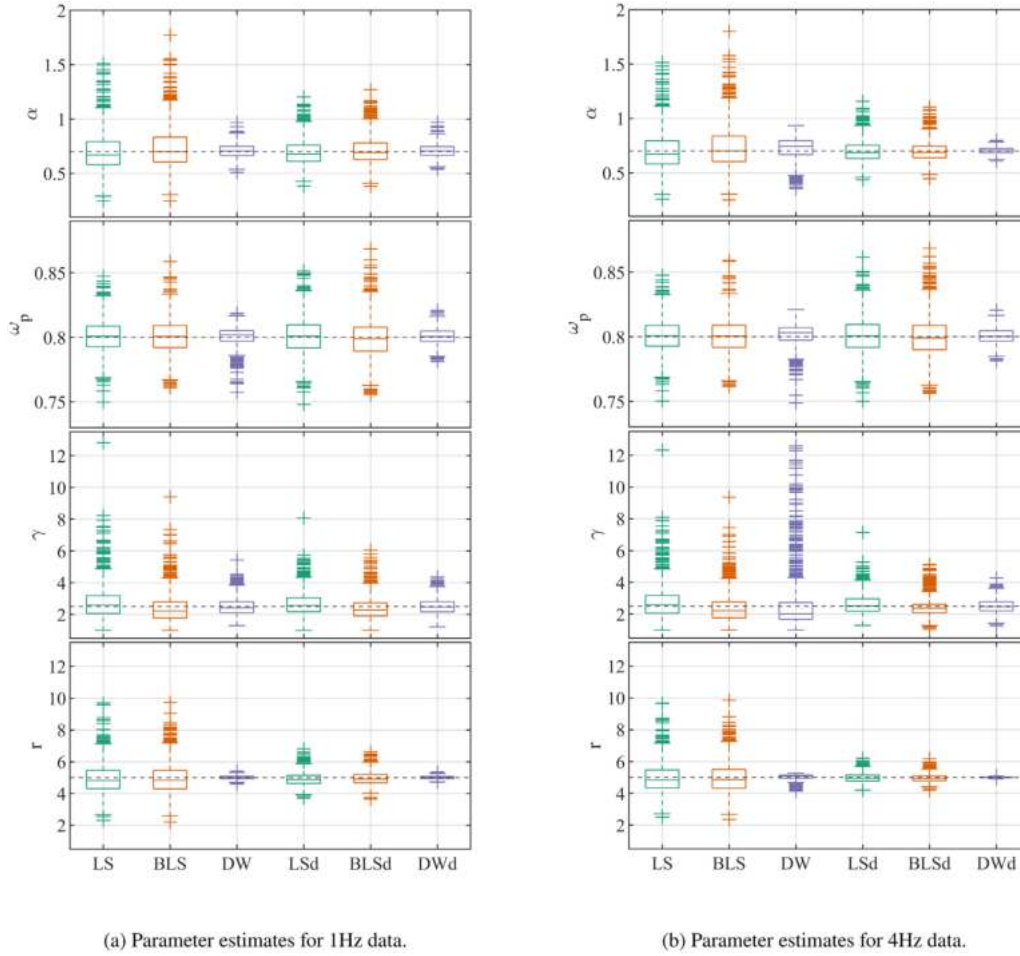
Fig. 6. Boxplots of 1000 parameter estimates for 1 Hz and 4 Hz data. LS, BLS, DW denote least squares, Bartlett least squares and de-biased Whittle respectively, and LSd, BLSd and DWd denote the differenced versions. Note that the values of γ for the 4 Hz DW have been truncated, the maximum was 19.81.

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

$$\text{var}\left(\hat{\theta}_{DW}\right) \approx \mathbb{E}\left[H(\theta)\right]^{-1} \text{var}\left(\nabla\ell_{DW}(\theta)\right) \mathbb{E}\left[H(\theta)\right]^{-1}, \quad (20)$$

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

Now for the de-biased Whittle likelihood, we may write

$$\frac{\partial}{\partial\theta_j}\ell_D(\theta) = -\sum_{\omega\in\Omega}\left\{\left(\frac{1}{\bar{f}_N(\omega\mid\theta)} - \frac{I(\omega)}{\bar{f}_N^2(\omega\mid\theta)}\right)\frac{\partial\bar{f}_N(\omega\mid\theta)}{\partial\theta_j}\right\}, \quad (21)$$

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

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell_D(\theta)\right] = -\sum_{\omega\in\Omega}\left\{\frac{1}{\bar{f}_N^2(\omega\mid\theta)}\frac{\partial\bar{f}_N(\omega\mid\theta)}{\partial\theta_j}\frac{\partial\bar{f}_N(\omega\mid\theta)}{\partial\theta_k}\right\}, \quad (22)$$

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 (20).

Furthermore, the triangle function and Fourier basis are constant with respect to θ, so from (13) we have that

$$\frac{\partial}{\partial\theta_j}\bar{f}_N(\omega\mid\theta) = \frac{\Delta}{2\pi}\text{Re}\left(2\sum_{\tau=0}^{N-1}\left(1-\frac{\tau}{N}\right)\frac{\partial c(\tau\mid\theta)}{\partial\theta_j}\exp\{-i\omega\tau\Delta\} - \frac{\partial c(0\mid\theta)}{\partial\theta_j}\right), \quad (23)$$

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

$$\frac{\partial}{\partial\theta_j}c(\tau\mid\theta) = \int_{-\pi/\Delta}^{\pi/\Delta}\frac{\partial f_\Delta(\omega\mid\theta)}{\partial\theta_j}\exp\{-i\omega\tau\}d\omega. \quad (24)$$

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 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 B, we discuss a novel procedure for estimating $\text{var}\left(\nabla\ell_{DW}(\theta)\right)$ in a computationally efficient manner. Combining this with the approach for obtaining derivatives, we can estimate the variance of the de-biased Whittle likelihood estimator using (20). 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.
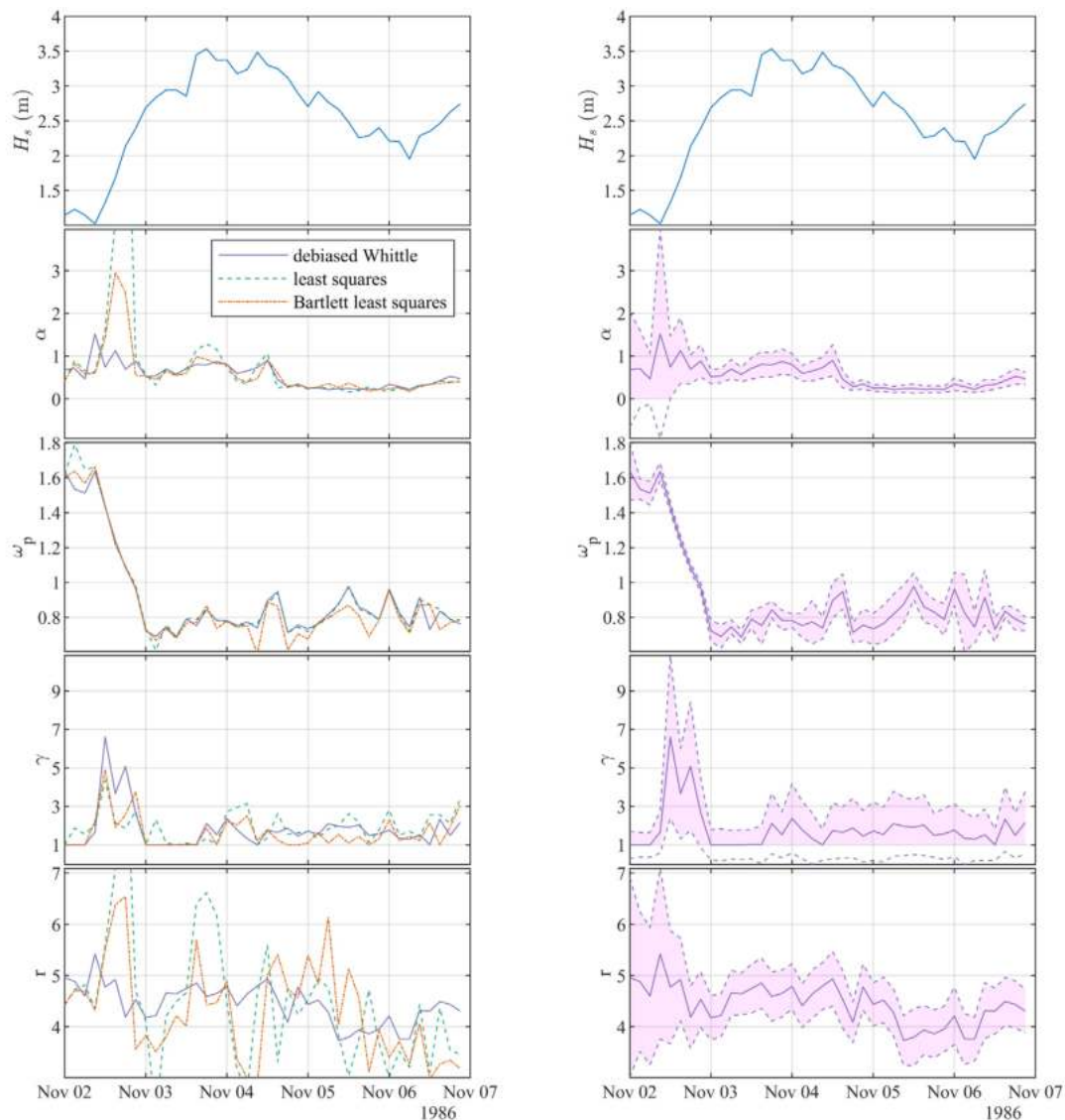
(a) Parameter estimates using least squares, Bartlett least squares and de-biased Whittle likelihood.

(b) Parameter estimates using de-biased Whittle likelihood with approximate 95% confidence intervals.

Fig. 7. Plots of H_s and the estimated parameters over time, from the second to the seventh of November, 1986. Estimates are from 20 min records starting every 3 h and are plotted at the start time of each record. Note that some of the plots in (a) have been truncated. The maximum of the least squares estimate for α was 9.37 and maximum and minimum for r were 9.66 and 2.28.

7. Maui data

Observations of fetch-limited seas made near the Maui-A platform off the coast of New Zealand between November 1986 and November 1987 have been studied previously by Ewans and Kibblewhite (1986) and Ewans (1998). We revisit these observations to demonstrate additional practical concerns when applying the techniques discussed in this paper to observed wave records. Observations were taken using a Datawell WAVEC buoy at a location with a water depth of 110 m. Data were recorded for 20 min at a sampling rate of 1.28 Hz, then there is a gap of 10 min while data is processed. Non-parametric spectral density estimates were recorded for each of these intervals, while the raw time series was only retained for the first 20 min of a 3 h segment. In other words, a 20 min record is available starting at 00:00, 03:00, 06:00 etc. for each day.

For the sake of illustration, we shall analyse the behaviour of spectral parameters throughout a single storm event, from the 2nd to

the 7th of November, 1986.[9] As can be seen from Fig. 7, throughout the first and second days, the conditions become steadily more severe, before calming down over the proceeding days. Transforming to the frequency domain, we see the presence of persistent swell, which remains fairly constant throughout all of the records. 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.[10] Ideally we would aim to fit a bi-modal model that was designed to also describe the swell component. However, such an approach is beyond the scope of this paper, though we note that preliminary results for such a procedure are

---

[9] Note 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.

[10] This can be thought of as first applying a high pass filter, then fitting a uni-modal model to the resulting filtered record.

encouraging and are discussed further in Section 8. It was also observed that frequencies beyond 3 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. Therefore, we also chose to omit those frequencies in the spectral fitting procedure when using each of the fitting procedures. 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 de-biased Whittle likelihood. These estimates are shown in Fig. 7(a).

From Fig. 7(a), we can see that, especially for the tail index, the de-biased Whittle estimates seem more stable than both the least squares and Bartlett least squares estimates. Fig. 7(b) shows the de-biased Whittle likelihood estimates and approximate 95% point-wise confidence intervals, calculated using the technique described in Section 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. To represent this, in Fig. 7(b), we only shade regions of the confidence interval that are within the parameter space. 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.

From Fig. 7(b) we can see that the parameters behave broadly as expected when moving through the storm. We see a steady evolution in peak frequency throughout the growth of the wind sea component, alongside large values of γ as the storm is developing. After this point, the values of γ reside close to 1. We can also see that there is a greater level of uncertainty when H_s was low. This is because there is very little energy in the wind-sea component. However, as the storm evolves, this uncertainty quickly reduces. The uncertainty surrounding the peak enhancement factor remains fairly high through the second half of the first day, even when the uncertainty around other parameters has reduced. This is because when the true value of γ is large, at frequencies around the peak the periodogram has more variance, and it is these frequencies that contain the information about γ.

A useful diagnostic tool for spectral models can be constructed by noticing that (under certain technical conditions) $I(\omega)/\mathbb{E}\left[I(\omega) \mid \theta\right]$ 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. Figs. 8(a) and 8(b) 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. Fig. 8(c) 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 Fig. 8(c) that the de-biased Whittle likelihood consistently outperforms the other techniques. Plots for the remaining sea states are available on GitHub (Grainger, 2021).
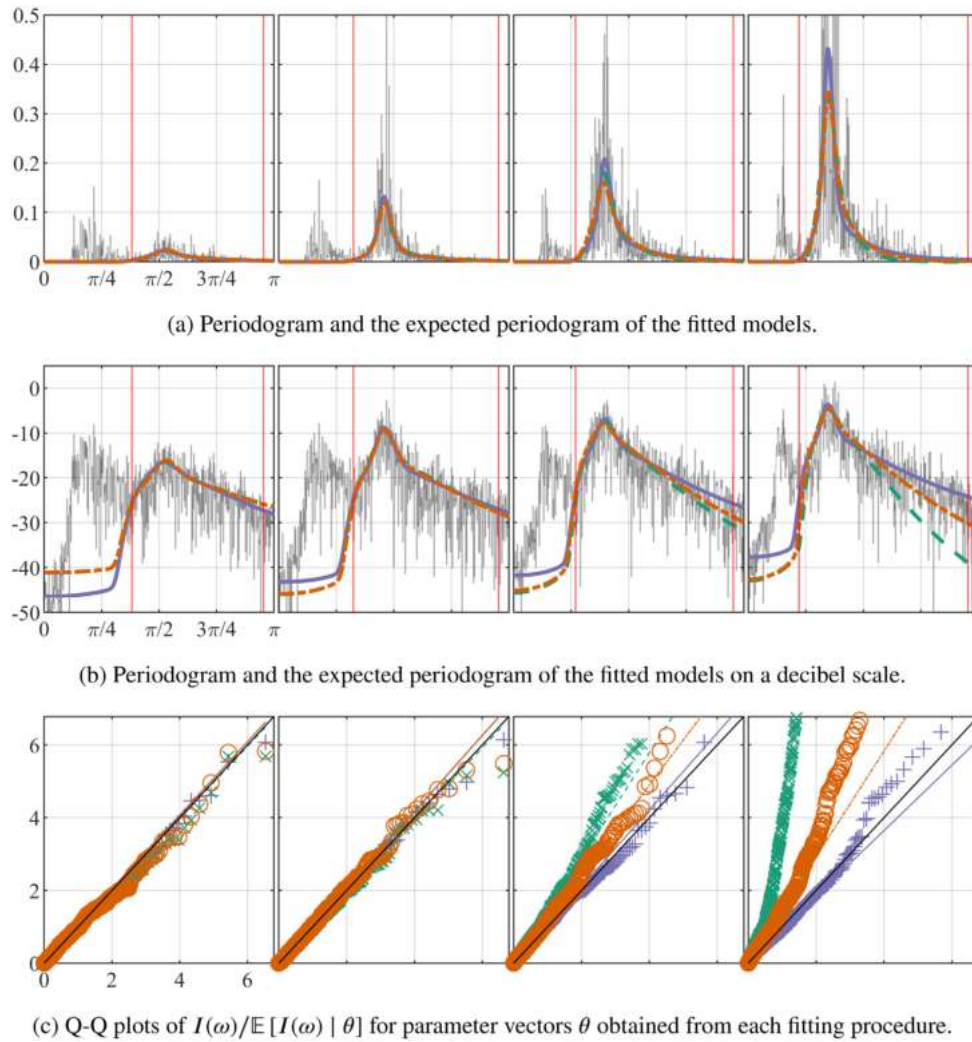
8. Discussion and conclusion

The de-biased 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 de-biased 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 de-biased 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 paper 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 de-biased 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 paper, 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 paper. This is available, alongside code to generate the figures in this paper and additional supplementary figures, on GitHub (Grainger, 2021). The toolbox contains code to perform each of the fitting techniques discussed in this paper 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 C, 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.

Though the de-biased 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 paper, 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.

(a) Periodogram and the expected periodogram of the fitted models.

(b) Periodogram and the expected periodogram of the fitted models on a decibel scale.

(c) Q-Q plots of $I(\omega)/\mathbb{E}[I(\omega) \mid \theta]$ for parameter vectors θ obtained from each fitting procedure.

Fig. 8. Plot of periodograms and fitted models on normal and decibel scale, alongside diagnostic Q-Q plots. Each column is from records recorded on the fourth of November, 1986, starting at 09:00, 12:00, 15:00 and 18:00 respectively. Colours and line styles are used to denote different fitting techniques in the same manner as Fig. 7(a). In (c), reference lines are added to each of the fitting methods to aid comparison to the black line denoting $x = y$. The symbols "+", "×" and "∘" denote quantiles from the de-biased 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. The untruncated figures can be found on GitHub (Grainger, 2021). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 would 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 paper.

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). However, it may preferable to model the three recorded series instead, requiring the use of a multivariate extension to the de-biased Whittle likelihood. These are all important avenues of further investigation.

In summary, we have demonstrated that, by using the de-biased 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.

CRediT authorship contribution statement

Jake P. Grainger: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Writing, Visualisation. **Adam M. Sykulski:** Conceptualisation, Methodology, Formal analysis, Writing. **Philip Jonathan:** Conceptualisation, Methodology, Writing, Visualisation. **Kevin Ewans:** Conceptualisation, Methodology, Resources, Data curation, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. 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, $f_G(\omega \mid \theta)$. We may write, for $\omega > 0$,

$$\frac{\partial}{\partial \omega} f_G(\omega \mid \theta) = \frac{\alpha}{2} \omega^{-r} \exp\left\{-\frac{r}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right\} \gamma^{\delta(\omega\mid\theta)}$$
$$\times \left[\frac{r}{\omega_p}\left(\frac{\omega}{\omega_p}\right)^{-s-1} - \frac{r}{\omega} - \frac{\delta(\omega \mid \theta) \cdot (\omega - \omega_p) \log \gamma}{\omega_p^2 \sigma(\omega \mid \theta)^2}\right]$$
$$= f_G(\omega \mid \theta)\left[\frac{r}{\omega_p}\left(\frac{\omega}{\omega_p}\right)^{-s-1} - \frac{r}{\omega} - \frac{\delta(\omega \mid \theta) \cdot (\omega - \omega_p) \log \gamma}{\omega_p^2 \sigma(\omega \mid \theta)^2}\right]$$
$$\to 0 \text{ as } \omega \to 0^+,$$

$$\frac{\partial}{\partial \alpha} f_G(\omega \mid \theta) = \omega^{-r} \exp\left\{-\frac{r}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right\} \gamma^{\delta(\omega\mid\theta)}/2$$
$$= f_G(\omega \mid \theta)/\alpha$$
$$\to 0 \text{ as } \omega \to 0^+,$$

$$\frac{\partial}{\partial \omega_p} f_G(\omega \mid \theta) = \frac{\alpha}{2} \omega^{-r} \exp\left\{-\frac{r}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right\} \gamma^{\delta(\omega\mid\theta)}$$
$$\times \left[\delta(\omega \mid \theta) \log(\gamma) \frac{\omega}{\sigma(\omega \mid \theta)^2}\left(\frac{\omega - \omega_p}{\omega_p^3}\right) - \frac{\omega r}{\omega_p^2}\left(\frac{\omega}{\omega_p}\right)^{-s-1}\right]$$
$$= f_G(\omega \mid \theta)\left[\delta(\omega \mid \theta) \log(\gamma) \frac{\omega}{\sigma(\omega \mid \theta)^2}\left(\frac{\omega - \omega_p}{\omega_p^3}\right) - \frac{\omega r}{\omega_p^2}\left(\frac{\omega}{\omega_p}\right)^{-s-1}\right]$$
$$\to 0 \text{ as } \omega \to 0^+,$$

$$\frac{\partial}{\partial \gamma} f_G(\omega \mid \theta) = \alpha \omega^{-r} \exp\left\{-\frac{r}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right\} \delta(\omega \mid \theta)\gamma^{\delta(\omega\mid\theta)-1}/2$$
$$= f_G(\omega \mid \theta)\delta(\omega \mid \theta)/\gamma$$
$$\to 0 \text{ as } \omega \to 0^+,$$

$$\frac{\partial}{\partial r} f_G(\omega \mid \theta) = \alpha \omega^{-r} \exp\left\{-\frac{r}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right\} \gamma^{\delta(\omega\mid\theta)}\left[-\log(\omega) - \frac{1}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right]/2$$
$$= f_G(\omega \mid \theta)\left[-\log(\omega) - \frac{1}{s}\left(\frac{\omega}{\omega_p}\right)^{-s}\right]$$
$$\to 0 \text{ as } \omega \to 0^+.$$

By the chain rule, for $\omega < 0$

$$\frac{\partial}{\partial \omega} f_G(\omega \mid \theta) = -\frac{\partial}{\partial \omega} f_G(-\omega \mid \theta),$$
$$\frac{\partial}{\partial \alpha} f_G(\omega \mid \theta) = \frac{\partial}{\partial \alpha} f_G(-\omega \mid \theta),$$
$$\frac{\partial}{\partial \omega_p} f_G(\omega \mid \theta) = \frac{\partial}{\partial \omega_p} f_G(-\omega \mid \theta),$$
$$\frac{\partial}{\partial \gamma} f_G(\omega \mid \theta) = \frac{\partial}{\partial \gamma} f_G(-\omega \mid \theta),$$
$$\frac{\partial}{\partial r} f_G(\omega \mid \theta) = \frac{\partial}{\partial r} f_G(-\omega \mid \theta).$$

Finally, for $\omega = 0$, since $f_G(0, \alpha, \omega_p, \gamma, r) = 0$, we may write

$$\frac{\partial}{\partial \omega} f_G(\omega, \mid \theta)\mid_{\omega=0} = 0,$$
$$\frac{\partial}{\partial \alpha} f_G(0, \mid \theta) = 0,$$
$$\frac{\partial}{\partial \omega_p} f_G(0, \mid \theta) = 0,$$
$$\frac{\partial}{\partial \gamma} f_G(0, \mid \theta) = 0,$$
$$\frac{\partial}{\partial r} f_G(0, \mid \theta) = 0.$$

Therefore, we see that f_G has continuous first derivatives.

Proposition 1. *For all $x \in \{\omega, \alpha, \omega_p, \gamma, r\}$ the series*

$$\sum_{k=-\infty}^{\infty} \frac{\partial}{\partial x} f_G\left(\omega + \frac{2\pi k}{\Delta} \mid \theta\right)$$

converges uniformly.

Proof. Firstly, due to symmetry, we may consider the series

$$\sum_{k=L}^{\infty} \frac{\partial}{\partial x} f_G\left(\omega + \frac{2\pi k}{\Delta} \mid \theta\right)$$

where L is such that $\omega_p < \frac{\pi}{\Delta} + \frac{2\pi L}{\Delta}$. Now write

$$g_k(\omega \mid \theta) = f_G\left(\omega + \frac{2\pi k}{\Delta} \mid \theta\right) \quad \text{and}$$
$$g_{k,x}(\omega \mid \theta) = \frac{\partial}{\partial x} f_G\left(\omega + \frac{2\pi k}{\Delta} \mid \theta\right).$$

Next notice that

$$|g_k(\omega \mid \theta)| \le \alpha\gamma\left(\omega + \frac{2\pi k}{\Delta}\right)^{-r}$$
$$\le \alpha\gamma\left(\frac{-\pi}{\Delta} + \frac{2\pi k}{\Delta}\right)^{-r}$$
$$= \alpha\gamma\left(\frac{\pi}{\Delta}\right)^{-r}(2k-1)^{-r}.$$

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 $f_G(\omega \mid \theta)$ multiplied by some other function. Furthermore, each of these functions can be bounded. Therefore we may write

$$|g_{k,x}(\omega \mid \theta)| \le C_x M_k,$$

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

$$\sum_{k=-\infty}^{\infty} \frac{\partial}{\partial x} f_G\left(\omega + \frac{2\pi k}{\Delta} \mid \theta\right)$$

is also uniformly convergent.

Proposition 2. *For all $x \in \{\omega, \alpha, \omega_p, \gamma, r\}$*

$$\frac{\partial}{\partial x} f_\Delta(\omega \mid \theta) = \sum_{k=-\infty}^{\infty} \frac{\partial}{\partial x} f_G\left(\omega + \frac{2\pi k}{\Delta} \mid \theta\right).$$

Furthermore, $\frac{\partial}{\partial x} f_\Delta(\omega \mid \theta)$ is continuous.

Proof. This first part follows from Proposition 1, the convergence of $\sum_{k=-\infty}^{\infty} f_G(\omega + 2\pi k/\Delta \mid \theta)$, 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.

Appendix B. Computing the variance of the first derivative

Sykulski et al. (2019) decompose the variance of the first derivative of the de-biased Whittle likelihood as follows:

$$\text{var}\left(\frac{\partial}{\partial \theta_i} \ell_{DW}(\theta) \right) = \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij}(\theta) a_{ik}(\theta) \, \text{cov}\left(I(\omega_j), I(\omega_k) \right),$$

where ω_j, ω_k denote Fourier frequencies and

$$a_{ij}(\theta) = \frac{\partial \bar{f}_N(\omega \mid \theta)}{\partial \theta_i} \frac{1}{\bar{f}_N^2(\omega_j \mid \theta)}.$$

To estimate the variance, they propose using

$$\widehat{\text{var}}\left(\frac{\partial}{\partial \theta_i} \ell_{DW}(\theta) \right) = \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij}(\hat{\theta}_{DW}) a_{ik}(\hat{\theta}_{DW}) \, \widehat{\text{cov}}\left(I(\omega_j), I(\omega_k) \right).$$

For ocean wave models, $a_{ij}(\hat{\theta}_{DW})$ can be easily computed by using the results from Appendix A. So, from Sykulski et al. (2019), we are interested in computing

$$\widehat{\text{cov}}\left(I(\omega_j), I(\omega_k) \right) = \left| \frac{\Delta}{2\pi N} \int_{-\pi/\Delta}^{\pi/\Delta} f_\Delta(\omega' \mid \theta) D_N(\Delta(\omega_j - \omega)) D_N^*(\Delta(\omega_k - \omega)) d\omega' \right|^2,$$

where

$$D_N(v) = \frac{\sin(Nv/2)}{\sin(v/2)} e^{-iv(N+1)/2}.$$

To do this efficiently, first note that

$$D_N(v) = \frac{\sin(Nv/2)}{\sin(v/2)} e^{-iv(N+1)/2} = \frac{e^{iNv/2} - e^{-iNv/2}}{e^{iv/2} - e^{-iv/2}} e^{-iv(N+1)/2}$$

$$= \frac{e^{-iNv/2}}{e^{-iv/2}} \frac{e^{iNv} - 1}{e^{iv} - 1} e^{-iv(N+1)/2} = e^{-i(N-1)v/2} e^{-iv(N+1)/2} \sum_{s=0}^{N-1} e^{isv}$$

by geometric series

$$= e^{-iNv} \sum_{s=0}^{N-1} e^{isv}.$$

It is also convenient to write

$$D_N^*(v) = \frac{\sin(Nv/2)}{\sin(v/2)} e^{iv(N+1)/2} = e^{iv} \sum_{s=0}^{N-1} e^{isv}.$$

Now, consider the function

$$h_{jk}(\omega) = f_\Delta(\omega \mid \theta) D_N(\Delta(\omega_j - \omega)) D_N^*(\Delta(\omega_k - \omega))$$

$$= f_\Delta(\omega \mid \theta) \left(e^{-iN(\Delta(\omega_j - \omega))} \sum_{s=0}^{N-1} e^{is(\Delta(\omega_j - \omega))} \right)$$

$$\times \left(e^{i(\Delta(\omega_k - \omega))} \sum_{r=0}^{N-1} e^{ir(\Delta(\omega_k - \omega))} \right)$$

$$= e^{-i\Delta(N\omega_j - \omega_k)} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k} \left[f_\Delta(\omega \mid \theta) e^{i\Delta(N-1)\omega} e^{-i\Delta(s+r)\omega} \right],$$

where we have rearranged for later convenience. We can now see, by linearity of integration, that

$$\int_{-\pi/\Delta}^{\pi/\Delta} h_{jk}(\omega) d\omega = e^{-i\Delta(N\omega_j - \omega_k)} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k}$$

$$\times \int_{-\pi/\Delta}^{\pi/\Delta} f_\Delta(\omega \mid \theta) e^{i\Delta(N-1)\omega} e^{-i\Delta(s+r)\omega} d\omega.$$

Thus for $r = 0, \ldots, N-1$ and $s = 0 \ldots, N-1$ we must calculate

$$\int_{-\pi/\Delta}^{\pi/\Delta} f_\Delta(\omega \mid \theta) e^{i\Delta(N-1)\omega} e^{-i\Delta(r+s)\omega} d\omega.$$

Notice that by letting $t = r+s$ we need to calculate the following integral for $t = 0, \ldots, (2N-1) - 1$

$$\int_{-\pi/\Delta}^{\pi/\Delta} f_\Delta(\omega \mid \theta) e^{i\Delta(N-1)\omega} e^{-i\Delta t\omega} d\omega.$$

For clarity let

$$q(\omega) = f_\Delta(\omega \mid \theta) e^{i\Delta(N-1)\omega}$$

then we need to compute

$$Q(t) = \int_{-\pi/\Delta}^{\pi/\Delta} q(\omega) e^{-i\Delta t\omega} d\omega, \tag{25}$$

for $t = 0, \ldots, 2N - 2$. 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 Appendix B.1). This means that we only need to do one Fourier transform, and can then sub this into the previous sums. Now we require

$$\int_{-\pi/\Delta}^{\pi/\Delta} h_{jk}(\omega) d\omega = e^{-i\Delta(N\omega_j - \omega_k)} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} e^{is\Delta\omega_j} e^{ir\Delta\omega_k} Q(r+s)$$

$$= e^{-i\Delta(N\omega_j - \omega_k)} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} \tilde{Q}(r,s) e^{i\Delta 2\pi(rk+sj)/N\Delta}$$

$$= e^{-i\Delta(N\omega_j - \omega_k)} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} \tilde{Q}(r,s) e^{i2\pi(rk+sj)/N}$$

where $\tilde{Q}(r,s) = Q(r+s)$. This is a 2D Fourier transform, and can be computed efficiently for $j, k = 0, \ldots, N-1$. This means we take $O(N^2 \log N)$ time. Importantly, libraries exist to compute this very quickly in most programming languages. Finally

$$\widehat{\text{cov}}\left(I(\omega_j), I(\omega_k) \right) = \left| \frac{\Delta}{2\pi N} \int_{-\pi/\Delta}^{\pi/\Delta} h_{jk}(\omega) d\omega \right|^2$$

$$= \left| \frac{\Delta}{2\pi N} e^{-i\Delta(N\omega_j - \omega_k)} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} \tilde{Q}(r,s) e^{i2\pi(rk+sj)/N} \right|^2$$

$$= \left| \frac{\Delta}{2\pi N} \sum_{s=0}^{N-1} \sum_{r=0}^{N-1} \tilde{Q}(r,s) e^{i2\pi(rk+sj)/N} \right|^2.$$

B.1. Approximating $Q(t)$

Note that we must still approximate $Q(t)$ at $t = 0, \ldots, 2N - 2$. We aim to approximate the integral in (25). To achieve this, we use the Riemann sum given by

$$\bar{Q}(t) = \frac{2\pi}{M\Delta} \sum_{j=0}^{M-1} q(2\pi j/M\Delta) e^{-i(t\Delta)(2\pi j/M\Delta)}. \tag{26}$$

An FFT will produce values of $\bar{Q}(t)$ for $t = 0, 1, \ldots, (M-1)$. Therefore, provided $M \geq 2N - 1$ we can obtain approximations of the desired integrals.

Appendix C. Assumptions for de-biased Whittle

Theorem 1 of Sykulski et al. (2019) gives assumptions for the optimal convergence of the de-biased 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 parameter 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 5.3 we demonstrate that, in this case, the de-biased 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 θ. 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 θ and Riemann integrable in ω. This is satisfied for the generalised JONSWAP spectral form as it is continuous in θ and ω.

Assumption five states that the expected periodogram has two continuous derivatives in θ, and that these derivatives are bounded uniformly for all N. Furthermore the first derivatives are required to have $\Theta(N)$ non-zero frequencies. Strictly speaking, the generalised JONSWAP is not twice differentiable (the second derivative does not exist at the peak, due to the step function $\sigma(\omega \mid \theta)$). However, a simple adaptation can be made, by replacing $\sigma(\omega \mid \theta)$ with

$$\bar{\sigma}(\omega \mid \theta) = \sigma_1 + (\sigma_2 - \sigma_1)\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(C\left(\omega - \omega_p\right)\right)\right), \qquad (27)$$

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 Appendices A and B, 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 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 Θ.

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 de-biased 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 JONSWAP form are indeed identifiable in practice for sufficiently long records.

References

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. Ecol. Saf. Coast. Shelf Zones Compr. Use Shelf Resour. 173–187.

Bartlett, M.S., 1948. Smoothing periodograms from time-series with continuous spectra. Nature 161 (4096), 686–687.

Battjes, J.A., Zitman, T.J., Holthuusen, L.H., 1987. A reanalysis of the spectra observed in JONSWAP. J. Phys. Oceanogr. 17, 1288–1295.

Bloomfield, P., 1973. An exponential model for the spectrum of a scalar time series. Biometrika 60 (2), 217–226.

Brockwell, P.J., Davis, R.A., 2006. Time Series : Theory and Methods, second ed. Springer, p. 577.

Chiu, S.-T., 1988. Weighted least squares estimators on the frequency domain for the parameters of a time series. Ann. Statist. 16 (3), 1315–1326.

Constable, C.G., 1988. Parameter estimation in non-Gaussian noise. Geophys. J. Int. 94 (1), 131–142.

Dahlhaus, R., 1988. Small sample effects in time series analysis: A new asymptotic theory and a new estimate. Ann. Statist. 16 (2), 808–841.

Dahlhaus, R., 2012. Locally stationary processes. In: Handbook of Statistics, Vol. 30. Elsevier B.V., pp. 351–413.

Davies, R.B., Harte, D.S., 1987. Tests for Hurst effect. Biometrika 74 (1), 95–101.

Dietrich, C.R., Newsam, G.N., 1997. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. SIAM J. Sci. Comput. 18 (4), 1088–1107.

Doob, J.L., 1953. Stochastic Processes. In: Wiley Publications in Statistics, Wiley, New York.

Dzhaparidze, K., Yaglom, A., 1983. Spectrum parameter estimation in time series analysis. In: Developments in Statistics, Vol. 4. Elsevier, pp. 1–96.

Ewans, K.C., 1998. Observations of the directional spectrum of fetch-limited waves. J. Phys. Oceanogr. 28 (3), 495–512.

Ewans, K.C., Kibblewhite, A.C., 1986. Spectral characteristics of ocean waves undergoing generation in New Zealand waters. In: 9th Australasian Fluid Mechanics Conference. Auckland Univ, Auckland, New Zealand.

Ewans, K., 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, Vol. 3. American Society of Mechanical Engineers (ASME), Madrid, Spain.

Forristall, G.Z., 1981. Measurements of a saturated range in ocean wave spectra. J. Geophys. Res. 86 (C9), 8075–8084.

Fuentes, M., 2007. Approximate likelihood for large irregularly spaced spatial data. J. Amer. Statist. Assoc. 102 (477), 321–331.

Grainger, J.P., 2021. Software for spectral fitting. In: GitHub Repository. GitHub, https://github.com/JakeGrainger/SpectralFitting.

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., Walden, H., 1973. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Ergänzungsheft 8-12.

Hwang, P.A., Fan, Y., Ocampo-Torres, F.J., García-Nava, H., 2017. Ocean surface wave spectra inside tropical cyclones. J. Phys. Oceanogr. 47 (10), 2393–2417.

Longuet-Higgins, M.S., Cartwright, D.E., Smith, N.D., 1963. Observations of the directional spectrum of sea waves using the motions of a floating buoy. In: Ocean Wave Spectra. pp. 111–136.

Mackay, E.B.L., 2016. A unified model for unimodal and bimodal ocean wave spectra. Int. J. Mar. Energy 15, 17–40.

Mérigaud, A., Ringwood, J.V., 2018. Free-surface time-series generation for wave energy applications. IEEE J. Ocean. Eng. 43 (1), 19–35.

Nordman, D.J., Lahiri, S.N., 2006. A frequency domain empirical likelihood for short- and long-range dependence. Ann. Statist. 34 (6), 3019–3050.

Pawitan, Y., 2001. In all Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press.

Percival, D.B., 2006. Exact simulation of complex-valued Gaussian stationary processes via circulant embedding. Signal Process. 86, 1470–1476.

Percival, D.B., Walden, A.T., 1993. Spectral Analysis for Physical Applications : Multitaper and Conventional Univariate Techniques. Cambridge University Press, Cambridge.

Phillips, O.M., 1958. The equilibrium range in the spectrum of wind-generated waves. J. Fluid Mech. 4, 426–434.

Phillips, O.M., 1985. Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. J. Fluid Mech. 156, 505–531.

Pierson, W.J., Moskowitz, L., 1964. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. J. Geophys. Res. 69 (24), 5181–5190.

Rodríguez, G., Soares, C.G., 1999. Uncertainty in the estimation of the slope of the high frequency tail of wave spectra. Appl. Ocean Res. 21 (4), 207–213.

Shao, X., Wu, W.B., 2007. Asymptotic spectral theory for nonlinear time series. Ann. Statist. 35 (4), 1773–1801.

Sykulski, A.M., Olhede, S.C., Guillaumin, A.P., Lilly, J.M., Early, J.J., 2019. The de-biased Whittle likelihood. Biometrika 106 (2), 251–266.

Sykulski, A.M., Olhede, S.C., Lilly, J.M., Early, J.J., 2017. Frequency-domain stochastic modeling of stationary bivariate or complex-valued signals. IEEE Trans. Signal Process. 65 (12), 3136–3151.

Toba, Y., 1973. Local balance in the air-sea boundary processes. J. Oceanogr. Soc. Japan 29, 209–220.

Trench, W.F., 2013. Introduction to Real Analysis. Faculty Authored and Edited Books & CDs, URL: https://digitalcommons.trinity.edu/mono/7.

Tucker, M.J., Challenor, P.G., Carter, D.J., 1984. Numerical simulation of a random sea: a common error and its effect upon wave group statistics. Appl. Ocean Res. 6 (2), 118–122.

Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. J. Amer. Statist. Assoc. 95 (452), 1229–1243.

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 Trans. Audio Electroacoust. 15 (2), 70–73.

Whittle, P., 1953. Estimation and information in stationary time series. Arkiv Mat. 2 (5), 423–434.

Wood, A.T.A., Chan, G., 1994. Simulation of stationary Gaussian processes in $[0, 1]^d$. J. Comput. Graph. Statist. 3 (4), 409–432.