The debiased Whittle likelihood

By ADAM M. SYKULSKI

*Department of Mathematics & Statistics, Lancaster University, Bailrigg,
Lancaster LA1 4YW, U.K.*

a.sykulski@lancaster.ac.uk

SOFIA C. OLHEDE, ARTHUR P. GUILLAUMIN

*Department of Statistical Science, University College London, Gower Street,
London WC1E 6BT, U.K.*

s.olhede@ucl.ac.uk   arthur.guillaumin.14@ucl.ac.uk

JONATHAN M. LILLY AND JEFFREY J. EARLY

*NorthWest Research Associates, 4118 148th Avenue NE, Redmond,
Washington 98052, U.S.A.*

lilly@nwra.com   jearly@nwra.com

Summary

The Whittle likelihood is a widely used and computationally efficient pseudolikelihood. However, it is known to produce biased parameter estimates with finite sample sizes for large classes of models. We propose a method for debiasing Whittle estimates for second-order stationary stochastic processes. The debiased Whittle likelihood can be computed in the same $O(n \log n)$ operations as the standard Whittle approach. We demonstrate the superior performance of our method in simulation studies and in application to a large-scale oceanographic dataset, where in both cases the debiased approach reduces bias by up to two orders of magnitude, achieving estimates that are close to those of the exact maximum likelihood, at a fraction of the computational cost. We prove that the method yields estimates that are consistent at an optimal convergence rate of $n^{-1/2}$ for Gaussian processes and for certain classes of non-Gaussian or nonlinear processes. This is established under weaker assumptions than in the standard theory, and in particular the power spectral density is not required to be continuous in frequency. We describe how the method can be readily combined with standard methods of bias reduction, such as tapering and differencing, to further reduce bias in parameter estimates.

Some key words: Computational efficiency; Fast Fourier transform; Frequency domain; Parameter estimation; Pseudolikelihood.

1. Introduction

This paper introduces an improved computationally efficient method of estimating time series model parameters for second-order stationary processes. The standard approach is to maximize the exact time-domain likelihood, which in general has computational efficiency of order n^2

for regularly spaced Gaussian observations, where n is the length of the observed time series, and produces estimates that are asymptotically efficient, converging at a rate of $n^{-1/2}$. A second approach is the method of moments, which in general has a computational efficiency of smaller order but with poorer statistical performance (Brockwell & Davis, 1991, p. 253), exhibiting both bias and, often, a higher variance. A third approach is to approximate the exact likelihood, often referred to as quasilikelihood, pseudolikelihood or composite likelihood, and has recently been receiving much attention across statistics; see, for example, Fan et al. (2014) and Guinness & Fuentes (2017). In time series analysis, such likelihood approximations offer the possibility of considerable improvements in computational performance, usually scaling as order $n \log n$, with only small changes in statistical behaviour; see, for example, Anitescu et al. (2016). Here we introduce a pseudolikelihood that is based on the Whittle likelihood (Whittle, 1953), which offers dramatic decreases in bias and mean squared error in applications, yet with no significant increase in computational cost and no loss in consistency or rate of convergence. We refer to our pseudolikelihood as the debiased Whittle likelihood.

The Whittle likelihood (Whittle, 1953) is a frequency-domain approximation to the exact likelihood and is considered a standard method in parametric spectral analysis on account of its $O(n \log n)$ computational efficiency (Choudhuri et al., 2004; Fuentes, 2007; Matsuda & Yajima, 2009; Krafty & Collinge, 2013; Jesus & Chandler, 2017). However, it has been observed that the Whittle likelihood, despite its desirable asymptotic properties, may exhibit poor behaviour when applied to real-world, finite-length time series, particularly in terms of estimation bias (Dahlhaus, 1988; Velasco & Robinson, 2000; Contreras-Cristan et al., 2006). Bias is caused by spectral blurring, sometimes referred to as spectral leakage (Percival & Walden, 1993). Furthermore, when the time series model is specified in continuous time but observed discretely, there is the added problem of aliasing, see also Eckley & Nason (2018), which if unaccounted for will further increase bias in Whittle estimates. The challenge is to account for such sampling effects and to debias Whittle estimates while retaining the computational efficiency of the method. We develop such a procedure here, which can be combined with tapering and appropriate differencing, as recommended by Dahlhaus (1988) and Velasco & Robinson (2000). This creates an automated procedure that incorporates all modifications simultaneously, without any hand-tuning or reliance on process-specific analytical derivations such as in Taniguchi (1983).

We compare pseudolikelihood approaches using simulated and real-world time series. In our example from oceanography, the debiased Whittle likelihood results in parameter estimates that are significantly closer to maximum likelihood than standard Whittle estimates, while reducing the computational time by a factor of 100 relative to maximum likelihood, thus demonstrating the practical utility and scalability of our method. Additionally, the theoretical properties of our new estimator are studied under relatively weak assumptions, in contrast to Taniguchi (1983), Dahlhaus (1988) and Velasco & Robinson (2000). Taniguchi (1983) considered autoregressive processes that depend on a scalar unknown parameter so that analytical calculations are possible. Dahlhaus (1988) examined processes whose spectral densities are the product of a known function with peaks that increase with sample size and a latent spectral density that is twice continuously differentiable in frequency. Velasco & Robinson (2000) investigated processes that exhibit power-law behaviour at low frequencies, and their results require continuous differentiability of the spectrum at all frequencies except zero. Our assumptions on the spectral density of the time series will be milder. In particular, we do not require that the spectral density be continuous in frequency. This is a useful generalization, as discontinuous spectra arise frequently, for example in oceanography, e.g., Polzin & Lvov (2011, p. 11). Despite these weaker assumptions, we are able to prove consistency of debiased Whittle estimates and establish a convergence rate matching the optimal $n^{-1/2}$, for large classes of Gaussian as well as non-Gaussian or nonlinear processes.

2. DEFINITIONS AND NOTATION

We shall assume that the stochastic process of interest is modelled in continuous time; however, the debiased Whittle likelihood can be readily applied to discrete-time models, as we will demonstrate later. We define $\{X_t\}$ to be the infinite sequence obtained from sampling a continuous-time real-valued process $X(t;\theta)$ with zero mean or a known nonzero mean such that it can be subtracted a priori, where θ is a length-p vector that specifies the process. That is, we let $X_t \equiv X(t\Delta;\theta)$, where t is a positive or negative integer ($t = \ldots, -2, -1, 0, 1, 2, \ldots$) and $\Delta > 0$ is the sampling interval. If the process is second-order stationary, we define the autocovariance sequence by $s(\tau;\theta) \equiv E\{X_t X_{t-\tau}\}$ for $\tau = \ldots, -2, -1, 0, 1, 2, \ldots$, where $E\{\cdot\}$ is the expectation operator. The power spectral density of $\{X_t\}$ forms a Fourier pair with the autocovariance sequence, and is almost everywhere given by

$$f(\omega;\theta) = \Delta \sum_{\tau=-\infty}^{\infty} s(\tau;\theta)\exp(-i\,\omega\tau\Delta), \quad s(\tau;\theta) = \frac{1}{2\pi}\int_{-\pi/\Delta}^{\pi/\Delta} f(\omega;\theta)\exp(i\,\omega\tau\Delta)\,d\omega. \quad (1)$$

As $\{X_t\}$ is a discrete sequence, its Fourier representation is only defined up to the Nyquist frequency $\pm\pi/\Delta$. Thus, there may be departures between $f(\omega;\theta)$ and the continuous-time process spectral density, denoted by $\tilde{f}(\omega;\theta)$, which for almost all $\omega \in \mathbb{R}$ is given by

$$\tilde{f}(\omega;\theta) = \int_{-\infty}^{\infty} \tilde{s}(\lambda;\theta)\exp(-i\,\omega\lambda)\,d\lambda, \quad \tilde{s}(\lambda;\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(\omega;\theta)\exp(i\,\omega\lambda)\,d\omega.$$

Here $\tilde{s}(\lambda;\theta) \equiv E\{X(t)X(t-\lambda)\}$ ($\lambda \in \mathbb{R}$) is the continuous-time process autocovariance, which is related to $s(\tau;\theta)$ via $\tilde{s}(\tau\Delta;\theta) = s(\tau;\theta)$ when τ is an integer. It follows that

$$f(\omega;\theta) = \sum_{k=-\infty}^{\infty} \tilde{f}\left(\omega + k\frac{2\pi}{\Delta};\theta\right), \quad \omega \in [-\pi/\Delta, \pi/\Delta]; \quad (2)$$

see Percival & Walden (1993). Thus, contributions to $\tilde{f}(\omega;\theta)$ outside the range of frequencies $\pm\pi/\Delta$ are said to be folded or wrapped into $f(\omega;\theta)$. We have defined both $f(\omega;\theta)$ and $\tilde{f}(\omega;\theta)$, as both quantities are important in separating aliasing from other artefacts in spectral estimation.

In addition to these theoretical quantities, we will also require certain quantities that are computed directly from a single length-n sample $\{X_t\}_{t=1}^{n}$. A widely used but inconsistent estimate of $f(\omega;\theta)$ is the periodogram, denoted by $I(\omega)$, which is the squared absolute value of the discrete Fourier transform:

$$I(\omega) \equiv |J(\omega)|^2, \quad J(\omega) \equiv \left(\frac{\Delta}{n}\right)^{1/2} \sum_{t=1}^{n} X_t \exp(-i\,\omega t\Delta), \quad \omega \in [-\pi/\Delta, \pi/\Delta]. \quad (3)$$

Note that $I(\omega)$ and $J(\omega)$ are taken to be properties of the observed realization and are formally not regarded as functions of θ.

3. MAXIMUM LIKELIHOOD AND THE WHITTLE LIKELIHOOD

Consider the discrete sample $X = \{X\}_{t=1}^{n}$, which is organized as a length-n column vector. Under the assumption that X is drawn from $X(t;\theta)$, the expected $n \times n$ autocovariance matrix is

$C(\theta) \equiv E\{XX^{\mathrm{T}}\}$, where $^{\mathrm{T}}$ denotes transpose, and the components of $C(\theta)$ are given by $C_{ij}(\theta) = s(i-j;\theta)$. Exact maximum likelihood inference can be performed for Gaussian data by evaluating the loglikelihood function (Brockwell & Davis, 1991, p. 254),

$$\ell(\theta) = -\log|C(\theta)| - X^{\mathrm{T}}C^{-1}(\theta)X, \qquad (4)$$

where $^{-1}$ denotes matrix inverse and $|C(\theta)|$ is the determinant of $C(\theta)$. In (4) we have omitted additive and multiplicative constants not affected by θ. The optimal choice of θ for our chosen model to characterize the sampled time series X is then found by maximizing the loglikelihood function in (4) so that

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ell(\theta),$$

where Θ defines the parameter space of θ. Because the time-domain maximum likelihood is known to have optimal properties, any other estimator will be compared with the properties of this quantity.

A standard technique for avoiding expensive matrix inversions is to approximate (4) in the frequency domain, following the seminal work of Whittle (1953). In this approach $C(\theta)$ is approximated using a Fourier representation, and special properties of Toeplitz matrices are utilized. Given the observed sampled time series X, the Whittle likelihood, denoted by $\ell_{\mathrm{W}}(\theta)$, is

$$\ell_{\mathrm{W}}(\theta) \equiv -\sum_{\omega \in \Omega} \left\{ \log \tilde{f}(\omega;\theta) + \frac{I(\omega)}{\tilde{f}(\omega;\theta)} \right\}, \qquad (5)$$

where Ω is the set of discrete Fourier frequencies,

$$\Omega \equiv (\omega_1, \omega_2, \ldots, \omega_n) = \frac{2\pi}{n\Delta}\left(-\lceil n/2 \rceil + 1, \ldots, -1, 0, 1, \ldots, \lfloor n/2 \rfloor\right). \qquad (6)$$

We have presented the Whittle likelihood in a discretized form here, as its usual integral representation must be approximated for finite-length time series. The Whittle likelihood approximates the time-domain likelihood, i.e., $\ell(\theta) \approx \ell_{\mathrm{W}}(\theta)$, and this statement can be made precise (Dzhaparidze & Yaglom, 1983). Its computational efficiency is $O(n \log n)$ as the periodogram can be computed using the fast Fourier transform, thus explaining its popularity in practice. Exact maximum likelihood, on the other hand, would require $O(n^2)$ computations for regularly sampled Gaussian processes, using for example the Trench algorithm (Trench, 1964) to compute (4), and often has higher complexity for non-Gaussian processes, as demonstrated by our non-Gaussian simulation example in the Supplementary Material.

The Whittle likelihood (5) is calculated using the periodogram, $I(\omega)$. This spectral estimate, however, is known to be a biased measure of the continuous-time process's spectral density for finite samples, due to blurring and aliasing effects (Percival & Walden, 1993). Aliasing results from the discrete sampling of the continuous-time process to generate an infinite sequence, whereas blurring is associated with the truncation of this infinite sequence over a finite-time interval. The desirable properties of the Whittle likelihood rely on the asymptotic behaviour of the periodogram for large sample sizes. The bias of the periodogram for finite samples, however, will translate into biased parameter estimates from the Whittle likelihood, as has been widely reported (e.g., Dahlhaus, 1988). In the next section we propose a procedure for debiasing Whittle estimates.

4. Modified pseudolikelihoods

4.1. *The debiased Whittle likelihood*

We introduce the pseudolikelihood function

$$\ell_{\mathrm{D}}(\theta) \equiv - \sum_{\omega \in \Omega} \left\{ \log \bar{f}_n(\omega; \theta) + \frac{I(\omega)}{\bar{f}_n(\omega; \theta)} \right\}, \tag{7}$$

$$\bar{f}_n(\omega; \theta) = \int_{-\pi/\Delta}^{\pi/\Delta} f(\nu; \theta) \mathcal{F}_{n,\Delta}(\omega - \nu) \, d\nu, \quad \mathcal{F}_{n,\Delta}(\omega) \equiv \frac{\Delta}{2\pi n} \frac{\sin^2(n\omega\Delta/2)}{\sin^2(\omega\Delta/2)}, \tag{8}$$

where a subscript D stands for debiased. Here $\tilde{f}(\omega; \theta)$ in (5) has been replaced by $\bar{f}_n(\omega; \theta) \equiv E\{I(\omega)\}$, which is the expected periodogram, and can be shown to be given by the convolution of the true modelled spectrum with the Fejér kernel $\mathcal{F}_{n,\Delta}(\omega)$. We call (7) the debiased Whittle likelihood, where the set Ω is defined as in (6).

Replacing the true spectrum $\tilde{f}(\omega; \theta)$ with the expected periodogram $\bar{f}_n(\omega; \theta)$ in (7) is a straightforward concept; however, our key innovation lies in formulating its efficient computation without losing $O(n \log n)$ efficiency. If we directly use (8), then this convolution would usually need to be approximated numerically and could be computationally expensive. Instead we utilize the convolution theorem to express the frequency-domain convolution as a time-domain multiplication (Percival & Walden, 1993, p. 198), such that

$$\bar{f}_n(\omega; \theta) = 2\Delta \times \mathrm{Re} \left\{ \sum_{\tau=0}^{n-1} \left(1 - \frac{\tau}{n} \right) s(\tau; \theta) \exp(-i\,\omega\tau\,\Delta) \right\} - \Delta \times s(0; \theta), \tag{9}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part. Therefore $\bar{f}_n(\omega; \theta)$ can be exactly computed at each Fourier frequency directly from $s(\tau; \theta)$, provided its functional form is known for $\tau = 0, \ldots, n-1$, by using a fast Fourier transform in $O(n \log n)$ operations. Care must be taken to subtract the variance term, $\Delta \times s(0; \theta)$, to avoid double counting contributions from $\tau = 0$. Both aliasing and blurring effects are automatically and conveniently accounted for in (9) in one operation; aliasing is accounted for by sampling the theoretical autocovariance function at discrete times, while the effect of blurring is accounted for by the truncation of this sequence to finite length and the inclusion of the triangle function $(1 - \tau/n)$ in the expression. The result is that $\bar{f}_n(\omega; \theta)$ is a blurred and aliased version of the true spectrum $\tilde{f}(\omega; \theta)$, which reflects the blurring and aliasing artefacts present in the periodogram.

The debiased Whittle likelihood can also be used with discrete-time processes, as (9) can be computed from the theoretical autocovariance sequence of the discrete process in exactly the same way. If the analytical form of $s(\tau; \theta)$ is unknown or expensive to evaluate, then it can be approximated from the spectral density using fast Fourier transforms, thus maintaining $O(n \log n)$ computational efficiency.

Computing the standard Whittle likelihood of (5) with the aliased spectrum $f(\omega; \theta)$ defined in (1), without accounting for spectral blurring, would in general be more complicated than using the expected periodogram $\bar{f}_n(\omega; \theta)$. This is because the aliased spectrum $f(\omega; \theta)$ seldom has an analytical form for continuous processes, and must instead be approximated either by explicitly wrapping in contributions from $\tilde{f}(\omega; \theta)$ from frequencies higher than the Nyquist frequency as in (2), or via an approximation to the Fourier transform in (1). This is in contrast to the debiased Whittle likelihood, where the effects of aliasing and blurring have been computed exactly in a single operation using (9). Thus, addressing aliasing and blurring together by means of the

debiased Whittle likelihood is simpler and computationally faster to implement than accounting for aliasing alone. This will become further apparent in the simulation studies of § 5.1.

4.2. *Combining with differencing or tapering*

A standard technique for reducing the effects of blurring on Whittle estimates is to apply the Whittle likelihood to the differenced process (Velasco & Robinson, 2000), as often differencing will decrease the dynamic range of the spectrum and hence decrease broadband blurring. The debiased Whittle likelihood can be readily implemented with differenced data. If we denote the differenced process by $U_t = X_{t+1} - X_t$, then the expected periodogram of (9) can be computed using the autocovariance of U_t, which is found from the autocovariance of X_t via $s_U(\tau) \equiv E\{U_t U_{t-\tau}\} = 2s_X(\tau) - s_X(\tau+1) - s_X(\tau-1)$, so that the procedure remains $O(n \log n)$.

Another standard approach to ameliorating the effects of blurring is to premultiply the data sequence by a weighting function known as a data taper (Thomson, 1982). The taper is chosen to have spectral properties such that broadband blurring will be minimized, and the variance of the spectral estimate at each frequency is reduced, although the trade-off is that tapering increases narrowband blurring as the correlation between neighbouring frequencies increases.

The tapered Whittle likelihood (Dahlhaus, 1988) corresponds to replacing the direct spectral estimator formed from $I(\omega)$ in (3) with one using the taper $h = \{h_t\}$:

$$J(\omega; h) \equiv \Delta^{1/2} \sum_{t=1}^{n} h_t X_t \exp(-i\,\omega t \Delta), \quad I(\omega; h) \equiv |J(\omega; h)|^2, \quad \sum_{t=1}^{n} h_t^2 = 1, \qquad (10)$$

where h_t is real-valued. Setting $h_t = 1/n^{1/2}$ for $t = 1, \ldots n$ recovers the periodogram estimate of (5). To estimate parameters, we then maximize

$$\ell_T(\theta) \equiv - \sum_{\omega \in \Omega} \left\{ \log \tilde{f}(\omega; \theta) + \frac{I(\omega; h)}{\tilde{f}(\omega; \theta)} \right\}, \qquad (11)$$

where a subscript T indicates that a taper has been used. Velasco & Robinson (2000) demonstrated that for certain discrete processes it is beneficial to use this estimator, rather than the standard Whittle likelihood, for parameter estimation, particularly when the spectrum exhibits a high dynamic range. Nevertheless, tapering will not remove all broadband blurring effects in the likelihood, because we are still comparing the tapered spectral estimate against the theoretical spectrum, and not against the expected tapered spectral estimate. Furthermore, there remain the issues of narrowband blurring and aliasing effects with continuous sampled processes.

Our debiasing procedure can be naturally combined with tapering. We define the pseudolikelihood

$$\ell_{TD}(\theta) \equiv - \sum_{\omega \in \Omega} \left\{ \log \bar{f}_n(\omega; h, \theta) + \frac{I(\omega; h)}{\bar{f}_n(\omega; h, \theta)} \right\}, \qquad (12)$$

$$\bar{f}_n(\omega; h, \theta) = \int_{-\pi/\Delta}^{\pi/\Delta} f(\nu; \theta) \mathcal{H}_{n,\Delta}(\omega - \nu) \, d\nu, \quad \mathcal{H}_{n,\Delta}(\omega) \equiv \Delta \left| \sum_{t=1}^{n} h_t \exp(-i\,\omega t \Delta) \right|^2,$$

with $I(\omega; h)$ as defined in (10) such that $\bar{f}_n(\omega; h, \theta) \equiv E\{I(\omega; h)\}$. We call $\ell_{TD}(\theta)$ the debiased tapered Whittle likelihood and $\bar{f}_n(\omega; h, \theta)$ the expected tapered spectral estimate. The function

$\bar{f}_n(\omega; h, \theta)$ can be computed exactly using a $O(n \log n)$ calculation similar to (9); that is,

$$\bar{f}_n(\omega; h, \theta) = 2\Delta \times \text{Re} \left\{ \sum_{\tau=0}^{n-1} s(\tau; \theta) \left(\sum_{t=1}^{n-\tau} h_t h_{t+\tau} \right) \exp(-i\,\omega\tau\Delta) \right\} - \Delta \times s(0; \theta).$$

Accounting in $\bar{f}_n(\omega; h, \theta)$ for the particular taper used accomplishes debiasing of the tapered Whittle likelihood, just as using the expected periodogram does for the standard Whittle likelihood. The time-domain kernel $\sum_{t=1}^{n-\tau} h_t h_{t+\tau}$ ($\tau = 0, \ldots, n-1$) can be pre-computed using fast Fourier transforms or using a known analytical form. Then, during optimization, a fast Fourier transform of this fixed kernel multiplied by the autocovariance sequence is taken at each iteration. Thus the debiased tapered Whittle likelihood is also an $O(n \log n)$ pseudolikelihood estimator.

Both the debiased Whittle and the debiased tapered Whittle likelihoods have their merits, but the trade-offs are different with nonparametric spectral density estimation than with parametric model estimation. Specifically, although tapering decreases the variance of nonparametric estimates at each frequency, it conversely may increase the variance of estimated parameters. This is because the taper reduces the degrees of freedom in the data, which increases correlations between local frequencies. On the other hand, the periodogram creates broadband correlations between frequencies, especially for processes with high dynamic range, which also contributes to variance in parameter estimates. We explore these trade-offs in greater detail in § 5.1.

5. SIMULATIONS AND APPLICATIONS

5.1. *The Matérn process*

In this section we investigate the performance of the debiased Whittle likelihood in a Monte Carlo study using observations from a Matérn process (Matérn, 1960), as motivated by the simulation studies of Anitescu et al. (2012), who investigated the same process. The Matérn process is a three-parameter continuous Gaussian process defined by its continuous-time unaliased spectrum

$$\tilde{f}(\omega) = \frac{A^2}{(\omega^2 + c^2)^\alpha}, \quad \omega \in \mathbb{R}. \tag{13}$$

The parameter $A \geqslant 0$ determines the magnitude of the variability, $1/c > 0$ is the damping time-scale, and $\alpha > 1/2$ controls the rate of spectral decay, or equivalently the smoothness or differentiability of the process. For large α the power spectrum exhibits a high dynamic range, and the periodogram will be a poor estimator of the spectral density due to blurring. Conversely, for small α there will be departures between the periodogram and the continuous-time spectral density because of aliasing. We therefore investigate the performance of estimators over a range of α values, and this is the reason why the Matérn process is a suitable process to study.

In Table 1 we display the average percentage bias, standard deviation and root mean square error, relative to the true parameter values, for six different pseudolikelihoods: the standard Whittle likelihood, (5), with the observed and the differenced processes; the tapered Whittle likelihood (11); and the debiased versions, (7) and (12). Our choice of data taper is the discrete prolate spheroidal sequence taper (Slepian & Pollak, 1961), with bandwidth parameter 4, for which performance was found to be broadly similar across different choices of bandwidth. We also include results for the exact maximum likelihood, (4). The results are averaged over estimates of the three parameters $\{A, \alpha, c\}$, all assumed to be unknown, where the true α varies from 0.6 to 2.5

Table 1. *Average percentage bias, standard deviation, and root mean square*
error relative to the true parameter values for a Matérn process across all
estimates of $\{A, \alpha, c\}$ *with* $n = 1000$

Inference method	Equation	Bias	SD	RMSE
Standard Whittle (periodogram)	(5)	23.69%	10.34%	26.66%
Debiased Whittle (periodogram)	(7)	3.96%	12.97%	13.75%
Standard Whittle (tapered)	(11)	18.11%	12.23%	23.12%
Debiased Whittle (tapered)	(12)	2.60%	14.15%	14.41%
Standard Whittle (differenced)	(5)	18.99%	9.33%	22.09%
Debiased Whittle (differenced)	(7)	1.19%	8.90%	8.99%
Maximum likelihood	(4)	1.10%	7.60%	7.68%

SD, standard deviation; RMSE, root mean square error.

in increments of 0.1 and we fix $A = 1$ and $c = 0.2$. This is to explore performance over spectra that have aliasing artefacts as well as high dynamic range. For each value of α, we simulate 10 000 time series, each of length $n = 1000$. The optimization is performed in Matlab using `fminsearch`, under identical settings for all likelihoods. Initialized guesses for the slope and amplitude are found using a least squares fit in the range $[\pi/4\Delta, 3\pi/4\Delta]$, and the initial guess for the damping parameter c is set at a mid-range value of 100 times the Rayleigh frequency, i.e., $c = 100\pi/n = \pi/10$.

All standard Whittle methods have performance significantly contaminated by bias. The debiased variants decrease this bias by an order of magnitude. The standard deviation is broadly similar across all Whittle methods, and tapering results in standard deviations that are approximately twice that of maximum likelihood, which is consistent with the loss of information from using a data taper. Of all the pseudolikelihood estimators considered, the debiased Whittle likelihood using the differenced process performs best, and yields results close to the exact maximum likelihood. Overall, of the three modifications to the standard Whittle likelihood, debiasing, tapering and differencing, the debiasing method proposed here is the procedure that yields the greatest overall improvement in parameter estimation.

In the Supplementary Material, we present a figure which separates out the bias and root mean square error improvements over different values of α, demonstrating that the debiased Whittle likelihood can effectively reduce bias from aliasing when α is low and bias from blurring when α is high. In the Supplementary Material we also include a comparison with a time-domain $O(n \log n)$ estimator from Anitescu et al. (2012), which is found to perform similarly to the debiased Whittle likelihood with differenced data in terms of bias and root mean square error, although the latter method requires only a fraction of the computational time.

In the next section we will prove that the debiased Whittle likelihood is a consistent estimator converging at the optimal $n^{-1/2}$ rate, under assumptions which are satisfied by the Matérn process. Motivated by this, we perform an additional experiment over different lengths of time series $n = 2^k$, with k taking integer values from 7 to 13, so that n ranges from 128 to 8192. To isolate the convergence of the parameter estimate, we fix $A = 1$ and $c = 0.2$ as before but this time assume that these are known, and now only estimate the slope parameter, which we set to $\alpha = 2$.

The average bias, standard deviation and root mean square error of each estimator are plotted in Fig. 1, together with average CPU times. We show results for the standard and debiased Whittle likelihoods, as well as the exact maximum likelihood. Motivated by Table 1, we include the debiased Whittle likelihood with differenced data. Finally, we also report results for the standard Whittle likelihood using an approximated aliased spectrum, which we find using (2) by
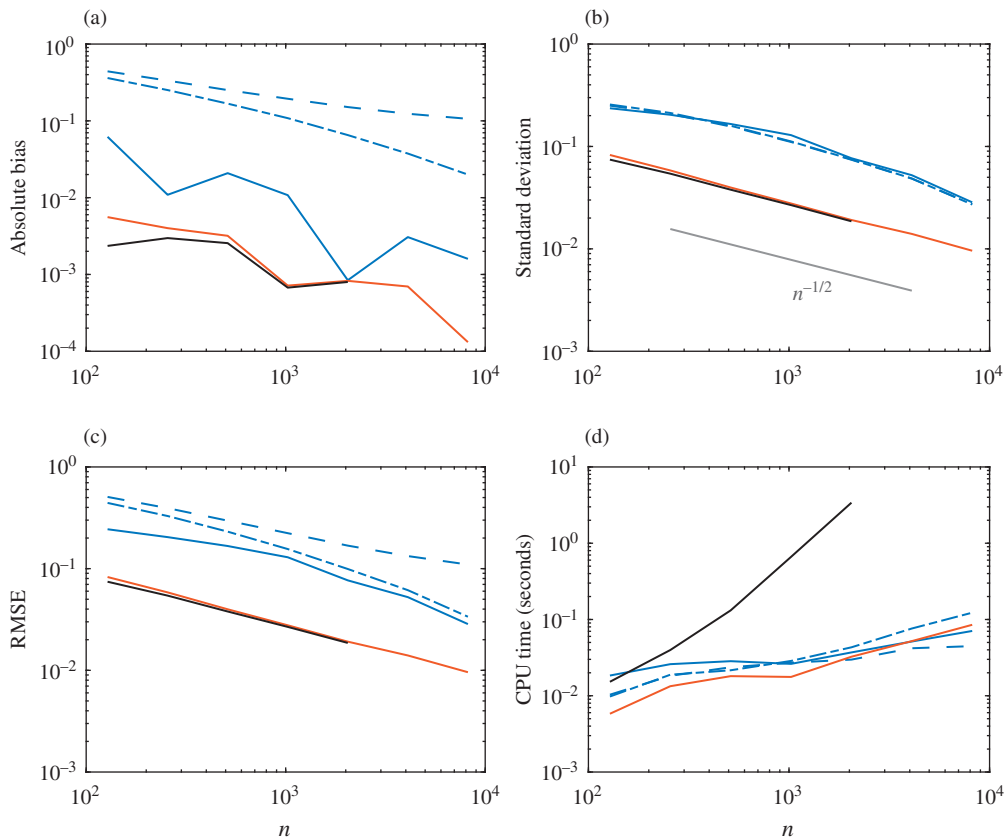
Fig. 1. Performance in terms of (a) absolute bias, (b) standard deviation, (c) root mean square error, and (d) CPU time of five different likelihoods: standard Whittle unaliased (blue dashed); standard Whittle aliased (blue dot-dashed); debiased Whittle periodogram (blue solid); debiased Whittle differenced (red solid); and exact maximum likelihood (black solid). Results were obtained over 1000 repetitions for various values of n when estimating the slope parameter ($\alpha = 2$) of a Matérn process. The axes are on a log-log scale, and CPU times are for computations performed on a 2.2 GHz Intel Core i7 processor.

truncating the summation limits to ± 5 to keep the computation efficient. The reason for including an approximate aliased version of the standard Whittle likelihood is to show that bias corrections are made by the debiased Whittle likelihood with regard to both blurring and aliasing. Here we see that the standard Whittle likelihood using the unaliased spectrum of (13) performs poorly with increasing n due to bias; this is because for growing domain asymptotics, bias due to aliasing does not decrease as n increases.

The standard deviations of all estimates converge at a rate consistent with $n^{-1/2}$, as Theorem 1 in § 6.1 will establish for debiased methods. Overall, the debiased approaches provide a good balance between statistical and computational efficiency over all sample sizes, whereas, in contrast, exact maximum likelihood is computed only up to $n = 2048$ due to rapidly increasing computational costs. For clarity of presentation the standard Whittle likelihood with differenced data is not included but was found to perform worse than the debiased Whittle likelihood with differenced data, consistent with the results in Table 1.

5.2. *Application to large-scale oceanographic data*

In this subsection we examine the performance of our method when applied to a real-world large-scale dataset, by analysing data obtained from the Global Drifter Program
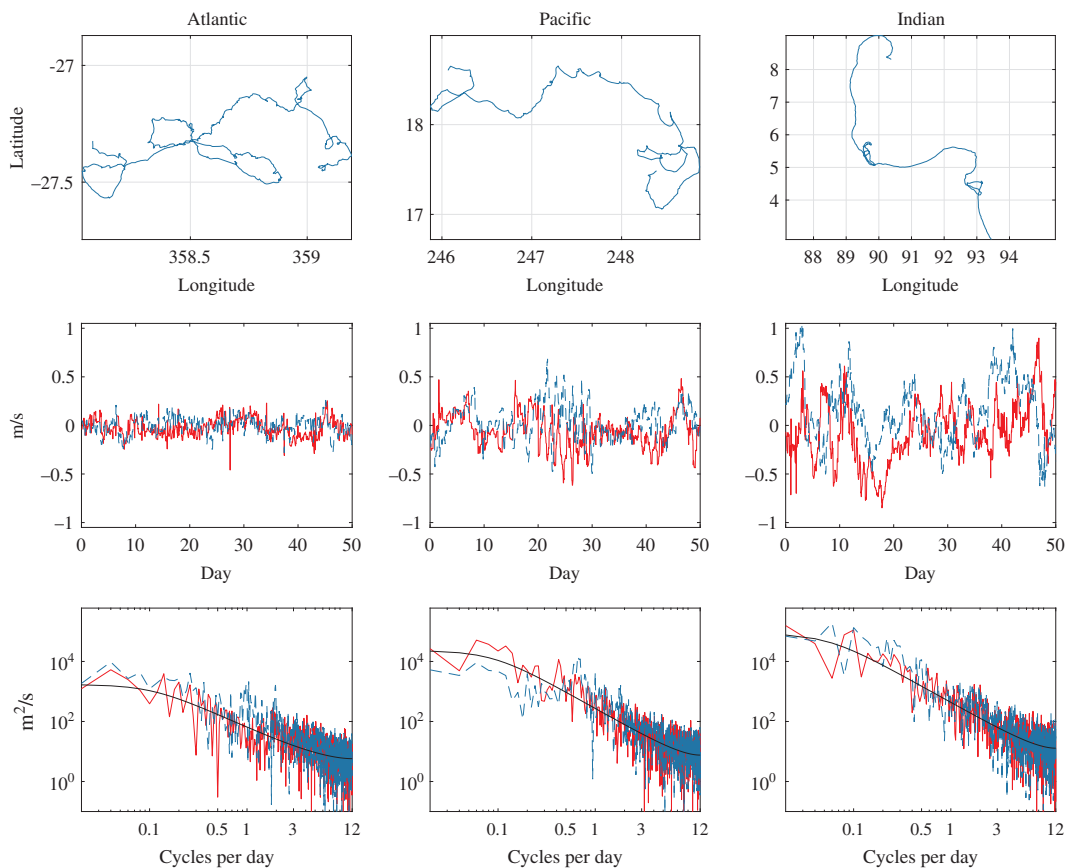
Fig. 2. The top row displays 50-day trajectories of drifters #2339255 (Atlantic Ocean), #49566 (Pacific Ocean) and #43577 (Indian Ocean). The middle row plots the corresponding east/west (red solid) and north/south (blue dashed) velocity time series for each trajectory. The bottom row shows the non-inertial (red solid) and inertial (blue dashed) sides of the periodogram of the complex-valued velocity series; the expected periodogram, $\bar{f}_n(\omega; \hat{\theta})$, from the debiased Whittle likelihood is overlaid in black, and the frequency axis is on a logarithmic scale.

`http://www.aoml.noaa.gov/phod/dac/index.php`, which maintains a publicly downloadable database of position measurements taken by freely drifting satellite-tracked oceanic instruments known as drifters. In total more than 23 000 drifters have been deployed, with interpolated six-hourly data available since 1979 and one-hourly data since 2005 (Elipot et al., 2016); over 100 million data points are available in total. Such data are pivotal to the understanding of ocean circulation and its impact on the global climate system (Griffa et al., 2007); it is therefore essential to have computationally efficient methods for their analysis.

In Fig. 2 we display 50-day position trajectories and corresponding velocity time series for three drifters from the one-hourly dataset, each from a different major ocean. These trajectories can be considered as complex-valued time series, with the real part corresponding to the east/west velocity component and the imaginary part corresponding to the north/south velocity component. We then plot the periodogram of the complex-valued series, which has different powers at positive and negative frequencies, distinguishing directions of rotation on the complex plane (Schreier & Scharf, 2010). The debiased Whittle likelihood for complex-valued proper processes is exactly the same as (7)–(9), see also Sykulski et al. (2016), where the autocovariance sequence of a complex-valued process Z_t is $s(\tau; \theta) = E(Z_t Z_{t-\tau}^*)$. For proper processes the complementary covariance is $r(\tau; \theta) = E(Z_t Z_{t-\tau}) = 0$ at all lags (Schreier & Scharf, 2010) and can

Table 2. *Estimated Matérn parameters (with corresponding estimated standard errors in parentheses) using the maximum, debiased Whittle, and standard Whittle likelihoods for the velocity time series of Fig. 2; the parameters are given in terms of the damping timescale $1/c$, the slope 2α, and the diffusivity κ; CPU times are for computations performed on a 2.8 GHz Intel Core i7 processor*

Drifter location	Inference method	Damping (days)	Slope (dimensionless)	Diffusivity (m²/s $\times 10^3$)	CPU (s)
	Maximum likelihood	10.65 (2.49)	1.460 (0.023)	0.49 (0.18)	7.42
Atlantic	Debiased Whittle	9.84 (5.51)	1.462 (0.062)	0.44 (0.28)	0.16
	Standard Whittle	30.19 (16.2)	1.097 (0.043)	0.65 (0.36)	0.04
	Maximum likelihood	10.62 (1.85)	1.829 (0.024)	5.09 (1.71)	7.47
Pacific	Debiased Whittle	11.82 (4.64)	1.827 (0.048)	6.00 (3.83)	0.10
	Standard Whittle	19.59 (6.51)	1.575 (0.036)	7.18 (3.60)	0.02
	Maximum likelihood	21.76 (4.83)	1.825 (0.025)	30.48 (12.9)	10.06
Indian	Debiased Whittle	19.90 (9.41)	1.802 (0.053)	22.70 (17.2)	0.10
	Standard Whittle	39.99 (16.9)	1.545 (0.038)	31.19 (19.7)	0.02

therefore be ignored in the likelihood, as $s(\tau; \theta)$ captures all the second-order structure in the zero-mean process.

We model the velocity time series as a complex-valued Matérn process, with power spectral density given in (13), as motivated by Sykulski et al. (2016) and Lilly et al. (2017). To account for a type of circular oscillations in each time series known as inertial oscillations, which create an off-zero spike on one side of the spectrum, we fit the Matérn process semiparametrically to the opposite, non-inertial side of the spectrum, represented by the solid red lines in the bottom row of Fig. 2; we overlay the fit of the debiased Whittle likelihood on the periodograms in Fig. 2. For a full parametric model of surface velocity time series, see Sykulski et al. (2016). We have selected drifters without noticeable tidal effects; for detiding procedures see Pawlowicz et al. (2002).

We estimate the Matérn parameters for each time series using the debiased and regular Whittle likelihoods, as well as the exact maximum likelihood. The last of these methods can be performed over only positive or negative frequencies by first decomposing the time series into analytic and anti-analytic components using the discrete Hilbert transform (see Marple, 1999) and then fitting the corresponding signal to an adjusted Matérn autocovariance that accounts for the effects of the Hilbert transform. The details of this procedure are provided in the Supplementary Material.

The parameter estimates from the three likelihoods are displayed in Table 2 along with the corresponding CPU times. We also report the estimated parameter standard errors using the method described in § 6.2, with more details in the online code. We reparameterize the Matérn process to output three important oceanographic quantities: the damping time-scale, the decay rate of the spectral slope, and the diffusivity, which is the rate of particle dispersion, $\kappa \equiv A^2/4c^{2\alpha}$ (Lilly et al., 2017, equation (43)). From Table 2 it can be seen that the debiased Whittle and maximum likelihoods yield similar values for the slope and damping time-scale; however, regular Whittle likelihood yields parameters that underestimate the slope by around 15% and overestimate the damping time-scale by a factor of two, which if used would incorrectly specify underdamped and rougher-than-expected trajectories. These biases are consistent with the significant biases discussed in § 5.1. Diffusivity estimates vary across all estimation procedures and have large standard errors; this variability is likely due to the fact that diffusivity is a measure of the spectrum at frequency zero, and hence estimation is performed over relatively few frequencies.

The maximum likelihood is two orders of magnitude slower to execute than the debiased Whittle likelihood. When this is scaled to fitting all time series in the Global Drifter Program database,

Table 3. *Average parameter estimates and root mean square errors when estimating all five* AR(4) *parameters using different estimation methods, for $n = 256$ and $n = 1024$; results are obtained over* 1000 *replicated time series for each time series length*

AR(4) Parameters:	$\phi_1 = 2.7607$	$\phi_2 = -3.8106$	$\phi_3 = 2.6535$	$\phi_4 = -0.9238$	$\sigma = 1$
			$n = 256$		
			Average parameter estimate		
Yule–Walker	1.7669	−1.6555	0.6081	−0.1685	4.3621
Standard Whittle	1.8989	−1.9485	0.8895	−0.2746	4.0231
Debiased Whittle	2.5309	−3.3065	2.1754	−0.7439	1.5341
Maximum likelihood	2.7478	−3.7490	2.5799	−0.8798	1.0525
			Root mean square error		
Yule–Walker	1.0591	2.2725	2.1473	0.7861	3.6499
Standard Whittle	0.9800	2.0775	1.9523	0.7076	3.5326
Debiased Whittle	0.5136	1.0539	0.9777	0.3456	1.7368
Maximum likelihood	0.0330	0.1618	0.2031	0.1318	0.2320
			$n = 1024$		
			Average parameter estimate		
Yule–Walker	2.1959	−2.5237	1.4004	−0.4328	2.9207
Standard Whittle	2.2642	−2.6878	1.5644	−0.5008	2.7092
Debiased Whittle	2.7030	−3.6704	2.5161	−0.8665	1.0370
Maximum likelihood	2.7574	−3.8006	2.6428	−0.9185	1.0028
			Root mean square error		
Yule–Walker	0.6409	1.4513	1.4094	0.5489	2.1214
Standard Whittle	0.6225	1.3923	1.3441	0.5164	2.0663
Debiased Whittle	0.2001	0.4346	0.4133	0.1550	0.6632
Maximum likelihood	0.0131	0.0438	0.0526	0.0326	0.0623

time-domain maximum likelihood becomes impractical. The debiased Whittle likelihood, on the other hand, retains the speed of the Whittle likelihood while returning estimates that are close to the maximum likelihood. The results of this subsection therefore serve as a proof of concept of how the debiased Whittle likelihood is a useful tool for efficient parameter estimation from large datasets.

5.3. *Autoregressive processes*

We now investigate the performance of the debiased Whittle likelihood in estimating parameters of a discrete-time autoregressive process, $X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \epsilon_t$, where ϵ_t are independent and identically distributed as $N(0, \sigma^2)$. Specifically, we generate time series from the AR(4) autoregressive process studied in Percival & Walden (1993), used throughout the book as a motivating example of a process that generates high spectral blurring in spectral density estimation. As the process is in discrete-time, there is no issue with aliasing, so this example assesses how well the debiased Whittle likelihood accounts for bias that is purely due to blurring.

In Table 3 we display the average parameter estimates and root mean square errors in estimating all five parameters of the AR(4) process $\{\phi_1, \phi_2, \phi_3, \phi_4, \sigma\}$. We compare four approaches: maximum likelihood, the standard Whittle likelihood, the debiased Whittle likelihood, and the standard Yule–Walker estimation procedure, which is used to initialize parameter estimates for the likelihood-based methods. We do not include results obtained using the differenced process, as it was not found to give improved parameter estimates for this particular example.

The Yule–Walker and standard Whittle estimates perform similarly and quite poorly for both sample sizes considered, which is consistent with the fact that the former uses the biased sample autocovariance to solve the Yule–Walker equations, while the latter uses the periodogram, which

is the Fourier pair of the biased sample autocovariance. The debiased Whittle likelihood accounts for this bias and yields average estimates that are close to the exact maximum likelihood and the true values; it eliminates around half the root mean square error when $n = 256$, and two thirds when $n = 1024$. The debiased Whittle likelihood is therefore an effective pseudolikelihood for discrete-time as well as continuous-time processes. In the Supplementary Material we report further simulation results, including a performance comparison for a non-Gaussian process, where again the debiased Whittle likelihood is found to provide a good trade-off between statistical and computational efficiency.

6. Properties of the debiased Whittle likelihood

6.1. *Consistency and optimal convergence rates*

We now establish consistency and optimal convergence rates for debiased Whittle estimates with Gaussian and certain classes of non-Gaussian or nonlinear processes. To show that debiased Whittle estimates converge at the optimal rate, the main challenge is that although our pseudo-likelihood accounts for the bias of the periodogram, correlation between different frequencies caused by the leakage associated with the Fejér kernel is still present. This is what prevents the debiased Whittle likelihood from being exactly equal to the time-domain maximum likelihood for Gaussian data. To establish optimal convergence rates, we bound the asymptotic behaviour of this correlation. The statement is provided in Theorem 1, with the proof given in the Supplementary Material. The proof is composed of several lemmas, which, for example, place useful bounds on the expected periodogram, the variance of linear combinations of the periodogram at different frequencies, and also the first and second derivatives of the debiased Whittle likelihood. Together these establish that the debiased Whittle likelihood is a consistent estimator with estimates that converge in probability at an optimal rate of $n^{-1/2}$, under relatively weak assumptions.

THEOREM 1. *Assume that $\{X_t\}$ is an infinite sequence obtained from sampling a zero-mean continuous-time real-valued process $X(t; \theta)$ which satisfies the following assumptions:*
  (i) *the parameter set $\Theta \subset \mathbb{R}^p$ is compact with a nonnull interior, and the true length-p parameter vector θ lies in the interior of Θ;*
 (ii) *for all $\theta \in \Theta$ and $\omega \in [-\pi, \pi]$, the spectral density of the sequence $\{X_t\}$ is bounded below by $f(\omega; \theta) \geqslant f_{\min} > 0$ and bounded above by $f(\omega; \theta) \leqslant f_{\max}$;*
(iii) *$\theta_1 \neq \theta_2$ implies $f(\cdot; \theta_1) \neq f(\cdot; \theta_2)$ on a set of positive Lebesgue measure;*
 (iv) *$f(\omega; \theta)$ is continuous in θ and Riemann-integrable in ω;*
  (v) *the expected periodogram $\bar{f}_n(\omega; \theta)$, as defined in (9), has two continuous derivatives in θ which are bounded above in magnitude uniformly for all n, and the first derivative in θ also has $\Theta(n)$ frequencies in Ω that are nonzero;*
 (vi) *$\{X_t\}$ is a fourth-order stationary process with finite fourth-order moments and absolutely summable fourth-order cumulants.*

Then the estimator $\hat{\theta} = \arg\max_{\theta \in \Theta} \ell_D(\theta)$ for a sample $\{X_t\}_{t=1}^n$, where $\ell_D(\theta)$ is the debiased Whittle likelihood of (7), satisfies $\hat{\theta} = \theta + O_p(n^{-1/2})$.

The fourth-order cumulant is formally defined in the Supplementary Material. All stationary Gaussian processes automatically satisfy assumption (vi) as the fourth-order cumulant is identically zero. In the Supplementary Material we consider a class of nonlinear processes and prove that it satisfies assumption (vi). Specifically, we study the process $Y_t = X_t^2$ where X_t is a Gaussian process with bounded spectral density and absolutely summable autocovariance.

6.2. *Standard error estimation*

Here we present a novel method of obtaining standard error estimates for debiased Whittle estimates. This method was used to calculate standard errors for our application example in Table 2. In the Supplementary Material we show that the $p \times p$ covariance matrix of the estimated vector $\hat{\theta}$ satisfies

$$\text{var}(\hat{\theta}) = E\{H(\theta)\}^{-1} \, \text{var}\{\nabla \ell_D(\theta)\} E\{H(\theta)\}^{-1} \{1 + o(1)\}, \tag{14}$$

where $\nabla = (\partial/\partial \theta_1 \; \partial/\partial \theta_2 \; \ldots \; \partial/\partial \theta_p)^T$ and $\nabla \ell_D(\theta)$ is known as the score. The $p \times p$ matrix $H(\theta)$, known as the Hessian, is defined entrywise by $H_{ij}(\theta) = \partial^2 \ell_D(\theta)/(\partial \theta_i \partial \theta_j)$, and its expectation can be approximated either analytically or numerically at $\hat{\theta}$. The remaining term in (14), $\text{var}\{\nabla \ell_D(\theta)\}$, is the $p \times p$ covariance matrix of the score. The diagonal elements of this matrix, which are variances of individual components of the score, can be expressed using (7) as

$$\text{var}\left\{ \frac{\partial}{\partial \theta_i} \ell_D(\theta) \right\} = \text{var}\left\{ \sum_{\omega \in \Omega} \frac{\partial \bar{f}_n(\omega; \theta)}{\partial \theta_i} \frac{I(\omega)}{\bar{f}_n^2(\omega; \theta)} \right\} = \text{var}\left\{ \sum_{j=1}^{n} a_{ij}(\theta) I(\omega_j) \right\}$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij}(\theta) a_{ik}(\theta) \, \text{cov}\{I(\omega_j), I(\omega_k)\},$$

where w_j are the elements of Ω defined in (6) and

$$a_{ij}(\theta) \equiv \frac{\partial \bar{f}_n(\omega_j; \theta)}{\partial \theta_i} \frac{1}{\bar{f}_n^2(\omega_j; \theta)}.$$

Here we have made use of the fact that the $\partial \log\{\bar{f}_n(\omega; \theta)\}/\partial \theta$ term is deterministic and therefore has no variance. As we have established asymptotic efficiency for $\hat{\theta}$, we can now use the invariance principle of maximum likelihood estimators (Casella & Berger, 2002, p. 320) to construct an estimator of the variance,

$$\hat{\text{var}}\left\{ \frac{\partial}{\partial \theta_i} \ell_D(\theta) \right\} = \sum_{j=1}^{n} \sum_{k=1}^{n} \hat{a}_{ij}(\theta) \hat{a}_{ik}(\theta) \, \hat{\text{cov}}\{I(\omega_j), I(\omega_k)\},$$

and by the same reasoning we can approximate $\hat{a}_{ij}(\theta)$ by $a_{ij}(\hat{\theta})$. Then, to estimate the covariance of the periodogram, we compute

$$\hat{\text{cov}}\{I(\omega_j), I(\omega_k)\} = \left| \frac{1}{2\pi n} \int_{-\pi}^{\pi} f(\omega'; \hat{\theta}) D_n(\omega_j - \omega') D_n^*(\omega_k - \omega') \, d\omega' \right|^2,$$

where the asterisk denotes the complex conjugate and $D_n(\omega)$ is the noncentred Dirichlet kernel defined by

$$D_n(\omega) \equiv \frac{\sin(n\omega/2)}{\sin(\omega/2)} \exp\{-i\,\omega(n+1)/2\},$$

so that we arrive at estimates of the diagonal elements of $\text{var}\{\nabla \ell_D(\theta)\}$. Estimates of $\text{cov}[\partial/\partial \theta_i\{\ell_D(\theta)\}, \, \partial/\partial \theta_j\{\ell_D(\theta)\}]$, which are the off-diagonal terms of $\text{var}\{\nabla \ell_D(\theta)\}$, can be found in the same way. Then, substituting all estimated entries of $\text{var}\{\nabla \ell_D(\theta)\}$ into (14), along with the estimate of the Hessian, gives estimates of the variance of the estimators.

7. Discussion

Standard theory shows that standard Whittle estimates are consistent with optimal convergence rates if the spectrum and its first and second partial derivatives in θ are continuous in ω and bounded from above and below (Dzhaparidze & Yaglom, 1983), as well as being twice continuously differentiable in θ. In contrast, we do not require that the spectrum or its derivatives be continuous in ω, so Theorem 1 will hold for discontinuous spectra, as long as the other assumptions, such as Riemann integrability, are satisfied. As detailed in the Supplementary Material, this is possible because the expectation of the score is zero after debiasing, which would not be the case for the standard Whittle likelihood; so we only need to consider the variance of the score and the Hessian. To control these variances, we make repeated use of a bound on the variance of linear combinations of the periodogram, a result previously established in Theorem 3.1 of Giraitis & Koul (2013) under a different set of assumptions.

It can easily be shown that the assumptions in Theorem 1 are weaker than standard Whittle assumptions, despite conditions on the behaviour of the expected periodogram $\bar{f}_n(\omega; \theta)$ in assumption (v). This is because if the spectral density $f(\omega; \theta)$ and its first and second partial derivatives in θ are continuous in both ω and θ, then it can be shown, by applying the Leibniz integration rule to the first and second derivatives of (8) with respect to θ, that $f(\omega; \theta)$ being twice continuously differentiable in θ implies that $\bar{f}_n(\omega; \theta)$ is twice continuously differentiable in θ. To show this, we make use of Proposition 3.1 in Stein & Shakarchi (2003), which states that the convolution of two integrable and periodic functions is itself continuous. This result can also be used to show that $\bar{f}_n(\omega; \theta)$ is always continuous in ω, even if $f(\omega; \theta)$ is not, as from (8) we see that $\bar{f}_n(\omega; \theta)$ is the convolution of $f(\omega; \theta)$ and the Fejér kernel, two functions which are integrable and 2π-periodic in ω. Therefore, not only does $\bar{f}_n(\omega; \theta)$ remove bias from blurring and aliasing and is computationally efficient to evaluate, but it also has desirable theoretical properties leading to consistency and optimal convergence rates of debiased Whittle estimates under weaker assumptions.

Supplementary material

Supplementary material available at *Biometrika* online includes additional simulation results, technical proofs, and details of a class of non-Gaussian processes satisfying the assumptions of our theory. All simulation results can be reproduced exactly in Matlab, and the data can be downloaded using the software available at `https://github.com/AdamSykulski/SPG`.

References

ANITESCU, M., CHEN, J. & STEIN, M. L. (2016). An inversion-free estimating equations approach for Gaussian process models. *J. Comp. Graph. Statist.* **26**, 98–107.

ANITESCU, M., CHEN, J. & WANG, L. (2012). A matrix-free approach for solving the parametric Gaussian process maximum likelihood problem. *SIAM J. Sci. Comp.* **34**, A240–62.

BROCKWELL, P. J. & DAVIS, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer.

CASELLA, G. & BERGER, R. L. (2002). *Statistical Inference*. Pacific Grove, California: Duxbury.

CHOUDHURI, N., GHOSAL, S. & ROY, A. (2004). Contiguity of the Whittle measure for a Gaussian time series. *Biometrika* **91**, 211–8.

Contreras-Cristan, A., Gutiérrez-Peña, E. & Walker, S. G. (2006). A note on Whittle's likelihood. *Commun. Statist.* B **35**, 857–75.

Dahlhaus, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.* **16**, 808–41.

Dzhaparidze, K. O. & Yaglom, A. M. (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics*, P. R. Krishnaiah, ed. New York: Academic Press, pp. 1–96.

Eckley, I. A. & Nason, G. P. (2018). A test for the absence of aliasing or white noise in locally stationary wavelet time series. *Biometrika* **105**, 833–48.

Elipot, S., Lumpkin, R., Perez, R. C., Lilly, J. M., Early, J. J. & Sykulski, A. M. (2016). A global surface drifter data set at hourly resolution. *J. Geophys. Res. Oceans* **121**, 2937–66.

Fan, J., Qi, L. & Xiu, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *J. Bus. Econ. Statist.* **32**, 178–91.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Am. Statist. Soc.* **102**, 321–31.

Giraitis, L. & Koul, H. L. (2013). On asymptotic distributions of weighted sums of periodograms. *Bernoulli* **19**, 2389–413.

Griffa, A., Kirwan, A. D., Mariano, A. J., Özgökmen, T. & Rossby, T. (2007). *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*. Cambridge: Cambridge University Press.

Guinness, J. & Fuentes, M. (2017). Circulant embedding of approximate covariances for inference from Gaussian data on large lattices. *J. Comp. Graph. Statist.* **26**, 88–97.

Jesus, J. & Chandler, R. E. (2017). Inference with the Whittle likelihood: A tractable approach using estimating functions. *J. Time Ser. Anal.* **38**, 204–24.

Krafty, R. T. & Collinge, W. O. (2013). Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika* **100**, 447–58.

Lilly, J. M., Sykulski, A. M., Early, J. J. & Olhede, S. C. (2017). Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. *Nonlin. Proc. Geophys.* **24**, 481–514.

Marple, S. L. (1999). Computing the discrete-time "analytic" signal via FFT. *IEEE Trans. Sig. Proces.* **47**, 2600–3.

Matérn, B. (1960). *Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Stockholm: Statens Skogsforskningsinstitut.

Matsuda, Y. & Yajima, Y. (2009). Fourier analysis of irregularly spaced data on R^d. *J. R. Statist. Soc.* B **71**, 191–217.

Pawlowicz, R., Beardsley, B. & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Comp. Geosci.* **28**, 929–37.

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

Polzin, K. L. & Lvov, Y. V. (2011). Toward regional characterizations of the oceanic internal wavefield. *Rev. Geophys.* **49**. DOI: 10.1029/2010RG000329.

Schreier, P. J. & Scharf, L. L. (2010). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge: Cambridge University Press.

Slepian, D. & Pollak, H. O. (1961). Prolate spheroidal wave functions, Fourier analysis and uncertainty — I. *Bell Syst. Tech. J.* **40**, 43–63.

Stein, E. M. & Shakarchi, R. (2003). *Fourier Analysis: An Introduction*. Princeton, New Jersey: Princeton University Press.

Sykulski, A. M., Olhede, S. C., Lilly, J. M. & Danioux, E. (2016). Lagrangian time series models for ocean surface drifter trajectories. *Appl. Statist.* **65**, 29–50.

Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11**, 157–69.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–96.

Trench, W. F. (1964). An algorithm for the inversion of finite Toeplitz matrices. *J. SIAM* **12**, 515–22.

Velasco, C. & Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J. Am. Statist. Soc.* **95**, 1229–43.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2**, 423–34.