

SPECIAL ISSUE

ANALYSIS OF NON-STATIONARY MODULATED TIME SERIES WITH APPLICATIONS TO OCEANOGRAPHIC SURFACE FLOW MEASUREMENTS

ARTHUR P. GUILLAUMIN,^{a*} ADAM M. SYKULSKI,^b SOFIA C. OLHEDE,^a JEFFREY J. EARLY^c AND JONATHAN M. LILLY^c

^a *Department of Statistical Science, University College London, London, UK*

^b *Data Science Institute and Department of Mathematics & Statistics, Lancaster University, Lancaster, UK*

^c *NorthWest Research Associates, Seattle, WA, USA*

We propose a new class of univariate non-stationary time series models, using the framework of modulated time series, which is appropriate for the analysis of rapidly evolving time series as well as time series observations with missing data. We extend our techniques to a class of bivariate time series that are isotropic. Exact inference is often not computationally viable for time series analysis, and so we propose an estimation method based on the Whittle likelihood, a commonly adopted pseudo-likelihood. Our inference procedure is shown to be consistent under standard assumptions, as well as having considerably lower computational cost than exact likelihood in general. We show the utility of this framework for the analysis of drifting instruments, an analysis that is key to characterizing global ocean circulation and therefore also for decadal to century-scale climate understanding.

Received 8 May 2016; Accepted 22 March 2017

Keywords: Modulation; non-stationary; periodogram; Whittle likelihood; missing data; surface drifters.

MOS subject classification: 62M15

1. INTRODUCTION

This article introduces a new family of rapidly evolving time series models, inspired by real-data applications, and then develops the appropriate analysis tools for their computationally efficient and consistent inference. Statistical models for time series observations are usually described by their expectations and covariance structure. Classic families of covariance structure correspond to stationary covariances, governed only by the temporal lags between observed values of the process. The assumption of stationarity greatly simplifies analysis, as it renders the covariance structure homogeneous across time, and this motivates averaging for estimation. Unfortunately, most often, this homogeneous time structure is inadequate as a model for real-world applications and does not reflect the variability of the observed time series.

To analyse non-stationary time series, the framework of locally stationary time series is standard (Priestley, 1988; Dahlhaus, 1997). The idea is to allow for a time-varying spectral density. Parametric models for the time-varying spectral density can be fitted via the use of local Fourier transforms, usually requiring a spectral smoothness assumption. The concept of infill asymptotics developed by Dahlhaus (1997) is based on the idea that a growing amount of data are obtained locally in time. Normally, for non-stationary time series analysis, there is a bias–variance trade-off that occurs when selecting the length of an analysis window. Longer windows will decrease variance but will simultaneously increase bias owing to the variation of the covariance function over

* Correspondence to: Arthur P. Guillaumin, Department of Statistical Science, University College London, London, UK. E-mail: arthur.guillaumin.14@ucl.ac.uk

the analysis window (Adak, 1998). In our case, we shall eliminate the bias, and this will enable us to use longer time window lengths. For this purpose, we exploit the notion of a modulated process (Parzen, 1963; Priestley, 1965). A modulated process is a latent stationary process multiplied pointwise by a modulating function. If we observe the modulating function, this framework allows us to define an expected sample autocovariance function, despite the clear non-stationarity of the modulated process. This in turn allows us to introduce the Fourier transform of the expected sample autocovariance function of the modulated process, which is equal to its expected periodogram. Through examining the expected periodogram, the properties controlling the latent random process may be inferred even when the modulating function changes very rapidly.

The standard class of modulated processes is that of asymptotically stationary modulated processes (Parzen, 1963; Toloi and Morettin, 1989; Jiang and Hui, 2004). There, the autocovariance of the modulating function converges to a fixed function, which is too restrictive for our real-world application. We introduce a more general class, which we call *modulated processes with a significant correlation contribution*. This more flexible model will still allow us to infer the parameters of the driving process using likelihood-based methodologies. An alternative approach might be to simply divide the observed process by the known modulating sequence to recover the latent process and then perform inference directly on the recovered latent process. However, this is not possible in general, as the modulating function may contain zeros, or the observed process may in fact be an aggregation of different processes, as in the case in the real-world application that motivated us to develop this model class.

Anticipating our application to oceanographic surface flow measurements, we present a novel generalization of modulated processes for isotropic bivariate processes, or equivalently proper complex-valued processes (Schreier and Scharf, 2010). The wealth of possible structure in multi-variate processes is considerable in general. Inherent documented challenges in modelling include producing valid joint representations (Tong 1973, 1974; Priestley and Tong, 1973). This problem does not apply here, as we shall modulate both stationary processes under consideration simultaneously, thus automatically removing such problems.

Having set up our model of modulated processes with a significant correlation contribution, we show how a modified version of a frequency-domain likelihood allows us to consistently estimate parameters with a high degree of computational efficiency. More specifically, the Whittle likelihood for stationary Gaussian processes is an approximation of the exact likelihood that is asymptotically equivalent and can be computed in $\mathcal{O}(N \log N)$ elementary operations. We adapt this pseudo-likelihood to our class of models, making use of the expected periodogram, and conserve the $\mathcal{O}_P(N^{-1/2})$ convergence rate. We also conserve the $\mathcal{O}(N \log N)$ computational cost in the minimization procedure. Exact likelihood for non-stationary time series, on the other hand, will in general require more than $\mathcal{O}(N^2)$ operations, owing to the need to manipulate large covariance matrices.

We apply this method to an important dataset measuring ocean currents. There are only a handful of observational platforms capable of providing continuous global coverage of the Earth's oceans, and so it is critical that we fully utilize these datasets to advance our understanding of the oceans and their impact on climate. One of these studies is the Global Drifter Program (www.aoml.noaa.gov/phod/dac), consisting of freely drifting instruments, or 'drifters' (Lumpkin and Pazos, 2007). Figure 1(a) shows positions from multiple trajectories obtained from drifters at or near the equator. From the positions of the trajectories, we may also calculate the velocities of the instruments, and these velocity time series are useful measurements for understanding ocean dynamics. Depending on the instrument, it may not be reasonable to model the velocity time series as locally stationary, as is assumed in Sykulski *et al.* (2016c). In particular, for reasons to be discussed, regions near the equator are likelier to yield drifter trajectories with highly non-stationary velocities where locally stationary modelling breaks down. Instead, to capture such rapid time-variability, we use a modulated stochastic process from our class of non-stationary models. This model allows us to capture the rapid frequency modulation of oscillations known to geophysicists as 'inertial oscillations'. An example of a time series with such rapid frequency modulation can be seen in Figure 1(c).

We organize the article into the following sections. Section 2 reviews the model family of modulated processes, the standard assumption of asymptotic stationarity associated with such processes, and introduces our generalized class called modulated processes with a significant correlation contribution. This section also includes extensions to bivariate processes. Section 3 describes a computationally efficient pseudo-likelihood estimation procedure. In Section 4, we apply our methods to real-world oceanographic data and various numerical experiments; we also

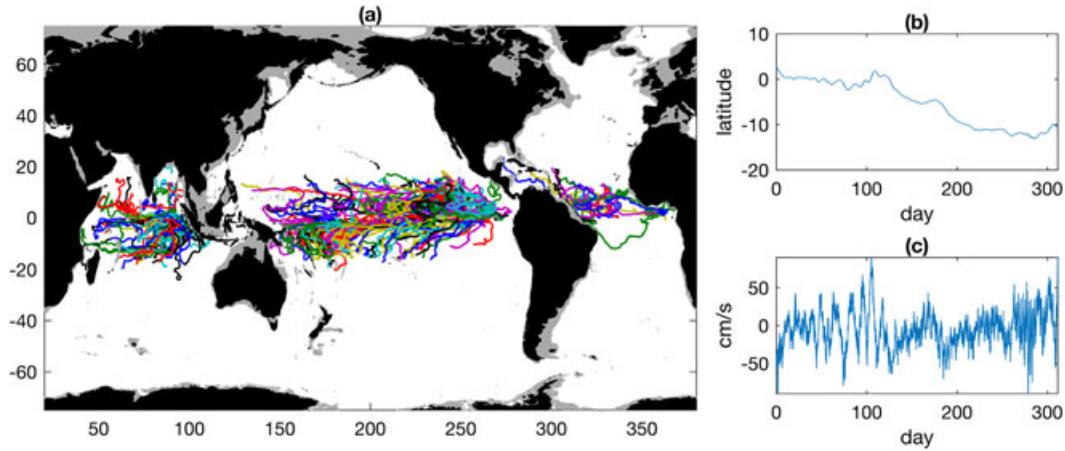


Figure 1. (a) The trajectories of the 200 drifters from the Global Drifter Program, analysed in Section 4.1.1, that exhibit the greatest change in Coriolis frequency (f) across 60 inertial cycles, as described in that section. (b) A segment of data of the meridional (latitudinal) positions over time from drifter ID 43594. (c) A segment of data of the latitudinal velocities from this drifter in centimetre per second. This figure is produced using the jLab toolbox (Lilly, 2016)

apply our methods to a simulated missing data problem. We establish consistency of our proposed procedure in Section 5, under the assumption of significant correlation contribution, as well as standard assumptions on the stationary process that is modulated. Finally, concluding remarks can be found in Section 6.

2. MODULATED TIME SERIES

In this article, we review and study modelling and inference methods for univariate and bivariate non-stationary Gaussian processes. We have that the first moment of a univariate stochastic discrete Gaussian process $\{X_t : t \in \mathbb{N}\}$, with index set $\mathbb{N} = \{0, 1, 2, \dots\}$, is provided pointwise by

$$\mu_X(t) = E\{X_t\},$$

and the second-order structure is given by

$$c_X(t_1, t_2) = \text{cov}\{X_{t_1}, X_{t_2}\},$$

where moments are finite as a direct consequence of the Gaussianity of $\{X_t\}$. We shall assume throughout this article that $\mu_X(t) = 0$. In practice, this may require us to subtract the sample mean from the observed series, or more generally remove trends and seasonal components (Brockwell and Davis, 1991, chap. 1).

Second-order stationarity implies that the function $c_X(t, t + \tau)$ does not depend on the index t and takes the simplified form $c_X(\tau)$. An alternative way to represent $c_X(\tau)$, assuming it is absolutely summable, is via its Fourier transform $S_X(\cdot)$, also known as the spectral density of $\{X_t\}$,

$$S_X(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} c_X(\tau) e^{-i\omega\tau}, \quad \omega \in [-\pi, \pi].$$

The spectral density $S_X(\omega)$ is then a continuous function of ω . The corresponding inversion formula is given for all integer value τ by

$$c_X(\tau) = \int_{-\pi}^{\pi} S_X(\omega) e^{i\omega\tau} d\omega.$$

A consequence of stationarity is that the quantities in question can be stably estimated by averaging in time (Brockwell and Davis, 1991). If $c_X(t, t + \tau)$ is not stationary, but is slowly varying in time, then it can be estimated by dividing the observed data into multiple segments and performing inference on each segment (Adak, 1998). This does not hold in settings where the time variation is too rapid. Our goal is to estimate $c_X(t, t + \tau)$ in such settings, in particular when a parametric specification is made for the function.

2.1. Classes of Modulated Processes

Modulation is a natural and simple method of producing a non-stationary process (Parzen, 1963). A univariate modulated process is defined as follows.

Definition 1 (Modulated process). Let $\{X_t : t \in \mathbb{N}\}$ be a Gaussian, real-valued, zero-mean stationary process. Let $\{g_t : t \in \mathbb{N}\}$ be a given bounded real-valued deterministic sequence. Then a modulated process is defined as one taking the form

$$\tilde{X}_t = g_t X_t \quad (1)$$

at all time points $t \in \mathbb{N}$.

Herein we treat $\{g_t\}$ as a known deterministic signal. In our setting, the process $\{X_t\}$, which is referred to as the *latent* process, is modelled through a finite set of parameters $\theta \in \Theta \subset \mathbb{R}^d$, where d is a positive integer and Θ is the parameter space. Usually, our object of interest is θ , the particular values of parameters that generated the observed realization. For example, if the latent process is an autoregressive process of order p , we then have $d = p + 1$ if the mean is known (p autoregressive parameters and the variance of the innovations). We denote the autocovariance function of the stationary zero-mean process $\{X_t\}$ by $c_X(\tau)$, or $c_X(\tau; \theta)$ when we want to make the dependence on θ explicit. Its Fourier transform, the spectral density, is denoted as $S_X(\omega)$ or $S_X(\omega; \theta)$ respectively.

The modulation of the latent process X_t is a convenient mechanism to account for a wide range of non-stationary processes. In particular, this mechanism has been widely used as a modelling tool for missing data problems in time series, where g_t is assigned values 0 or 1 when respectively missing or observing a data point in time (Jones, 1962).

To understand when we can recover the parameters controlling the latent process X_t from observing $\tilde{X}_t$, we need to put further conditions in place on g_t . The time series $\tilde{X}_t/g_t$ cannot always be formed as g_t may be zero for some time indices, corresponding to missing observations. Another reason is that we may not directly observe $\tilde{X}_t$, but instead we may observe an aggregated process $\tilde{X}_t + Z_t$, where Z_t is a stationary process independent from $\tilde{X}_t$, this preventing us from recovering the stationary latent process X_t by division.

We assume that $\tilde{X}_t$ satisfies (1) for a Gaussian, real-valued, zero-mean stationary X_t with absolutely summable autocovariance sequence. Then $E\{\tilde{X}_t\} = g_t E\{X_t\} = 0$, and the time-varying autocovariance sequence is defined by $c_{\tilde{X}}(t, t + \tau; \theta) = E\{\tilde{X}_t \tilde{X}_{t+\tau}\}$. Given a single length- N realization $\tilde{X}_0, \dots, \tilde{X}_{N-1}$, we start by computing the usual method of moments estimator according to

$$\hat{c}_{\tilde{X}}^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} \tilde{X}_t \tilde{X}_{t+\tau}, \quad (2)$$

for $\tau = 0, 1, \dots, N - 1$, such that τ is within the range of time offsets that is permissible given the length- N sample. Equation (2) is the biased sample autocovariance sequence of the modulated time series, which we define even though the process is non-stationary, as this object will become pivotal in our estimation procedure. The

expectation of this object, which we denote as $\bar{c}_{\tilde{X}}^{(N)}(\tau; \theta)$ or simply $\bar{c}_{\tilde{X}}^{(N)}(\tau)$, takes the following form:

$$\bar{c}_{\tilde{X}}^{(N)}(\tau) = E\{\hat{c}_{\tilde{X}}^{(N)}(\tau)\} = E\left\{\frac{1}{N} \sum_{t=0}^{N-\tau-1} \tilde{X}_t \tilde{X}_{t+\tau}\right\} = c_X(\tau) \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t g_{t+\tau} = c_g^{(N)}(\tau) \cdot c_X(\tau), \quad (3)$$

where we have introduced the (deterministic) sample autocovariance of the modulating sequence,

$$c_g^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t g_{t+\tau}. \quad (4)$$

In the specific case where the modulating sequence $\{g_t\}$ is constant and equal to unity everywhere, which would correspond to observing the latent stationary process directly, we recover the expectation of the biased sample autocovariance for stationary time series, $(1 - \tau/N) c_X(\tau)$, for $\tau = 0, \dots, N - 1$. More generally, a standard assumption is to say that the modulated process $\tilde{X}_t$ is an asymptotically stationary process (Parzen, 1961, 1963), which arises if for all lags τ , the quantity $c_g^{(N)}(\tau)$ in (3) converges as N tends to infinity. We define this formally as follows.

Definition 2 (Asymptotically stationary process). Let $\{\tilde{X}_t\}$ be a discrete-time random process. We say that $\{\tilde{X}_t\}$ is an asymptotically stationary process if there exists a fixed function $\{\gamma(\tau) : \tau \in \mathbb{N}\}$ such that for all $\tau \in \mathbb{N}$,

$$\lim_{N \rightarrow \infty} E\left\{\frac{1}{N} \sum_{t=0}^{N-\tau-1} \tilde{X}_t \tilde{X}_{t+\tau}\right\} = \gamma(\tau), \quad (5)$$

or specifically if $\tilde{X}_t$ is a modulated process as defined in Definition 1, $\tilde{X}_t$ is asymptotically stationary if,

$$\lim_{N \rightarrow \infty} \bar{c}_{\tilde{X}}^{(N)}(\tau) = \gamma(\tau), \quad (6)$$

where $\bar{c}_{\tilde{X}}^{(N)}(\tau)$ is defined in (3).

An example of a non-stationary but asymptotically stationary process is given by Parzen (1963), where a stationary process is observed according to a periodically missing data pattern, such that the first k values are observed, the next l values are missed, the next k values are observed, and so on, where k and l are two strictly positive integers.

The class of asymptotically stationary modulated processes (Parzen, 1963; Dunsmuir and Robinson, 1981b; Toloi and Morettin, 1989; Jiang and Hui, 2004) corresponds to that for which there exists a sequence $\{R_g(\tau) : \tau \in \mathbb{N}\}$ such that

$$\lim_{N \rightarrow \infty} c_g^{(N)}(\tau) = R_g(\tau), \quad \forall \tau \in \mathbb{N}. \quad (7)$$

Indeed we then note that $\bar{c}_{\tilde{X}}^{(N)}(\tau) \rightarrow R_g(\tau) c_X(\tau)$ as $N \rightarrow \infty$, so we could estimate $c_X(\tau)$ by defining

$$\hat{c}_X^{(N)}(\tau) = \frac{\hat{c}_{\tilde{X}}^{(N)}(\tau)}{R_g(\tau)}, \quad (8)$$

assuming $R_g(\tau) \neq 0$ for all $\tau \in \mathbb{N}$ and is known. It is shown in Parzen (1963) that $\hat{c}_X^{(N)}(\tau)$ is a consistent estimator of the autocovariance sequence $c_X(\tau)$ of the latent stationary process, under some rather mild conditions. Further

results are found in Dunsmuir and Robinson (1981b). Consistent spectral density estimates can be obtained by a Fourier transformation of the sequence $\{k^{(N)}(\tau)\hat{c}_X^{(N)}(\tau) : \tau = 0, \dots, N-1\}$, where $k^{(N)}(\tau)$ is chosen suitably for $\tau = 0, \dots, N-1$.

The key feature in Definition 2 is that in (6), we average the time-varying autocovariance sequence $c_{\tilde{X}}(t, t+\tau) = E\{\tilde{X}_t \tilde{X}_{t+\tau}\}$ across a time period N to produce an average autocovariance across the time period, written as $\tilde{c}_{\tilde{X}}^{(N)}(\tau)$. If this converges (in N) to a function of τ , then by observing the modulated process over a suitably long time interval, we can recover the second-order properties of the stationary latent process.

We now wish to explore a more general assumption than that of asymptotic stationarity for modulated processes. Specifically, we seek a larger class of models where consistent inference is still achievable. This will be smaller than the full class of models for g_t , as using a trivial example, if $g_t \equiv 0$ always then we would not be able to infer properties of the generating mechanism of X_t . For consistent inference, we propose the following class of modulated processes.

Definition 3 (Modulated process with a significant correlation contribution). Assume that $\tilde{X}_t$ is specified by (1). We say that $\tilde{X}_t$ is a *modulated process with a significant correlation contribution* if there exists a finite subset of non-negative lags $\Gamma \subset \mathbb{N}$ such that,

1. The mapping $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$ is one-to-one (injective).
2. For all lags $\tau \in \Gamma$,

$$\liminf_{N \rightarrow \infty} |c_g^{(N)}(\tau)| > 0, \quad (9)$$

where $\liminf_{N \rightarrow \infty}$ is the limit inferior.

Because of the symmetry of autocovariance sequences, we do not need to consider $\tau < 0$ in this definition. Point 1 of Definition 3 means that for any two distinct parameter vectors $\theta, \theta' \in \Theta$, there exists at least one lag τ in the finite set Γ such that $c_X(\tau; \theta) \neq c_X(\tau; \theta')$. It is therefore an assumption about the latent process model. The sequence $|c_g^{(N)}(\tau)|$ is bounded above since the modulating sequence is assumed to be bounded above. Therefore, the limit inferior in (9) is always finite. We observe that, for $\tau \in \Gamma$, (9) is equivalent to

$$\exists \alpha_\tau > 0, \exists N_\tau \in \mathbb{N}, \forall N \in \mathbb{N}, N \geq N_\tau \Rightarrow |c_g^{(N)}(\tau)| \geq \alpha_\tau, \quad (10)$$

which we interpret as the fact that the sequence $|c_g^{(N)}(\tau)|$ is bounded below for N large enough. For further understanding of point 1 in Definition 3, we provide the following two simple examples.

1. Let the latent process $\{X_t\}$ be an autoregressive process of order p , denoted as $AR(p)$, with known mean zero and unknown innovation variance, and with the parameter set Θ that is a subset of $\mathbb{R}^{p+1}$. If the parameter set Θ is chosen appropriately, that is, such that the roots of the characteristic equation all lie outside the unit circle, the Yule–Walker equations (Brockwell and Davis, 1991) show that $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$, where $\Gamma = \{0, \dots, p\}$, is a one-to-one mapping. Similarly if $\{X_t\}$ is a moving average process of order q , denoted as $MA(q)$, with known mean an unknown innovation variance and if the parameter set Θ is chosen appropriately (Dzhaparidze and Yaglom, 1983), then the mapping $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$, where $\Gamma = \{0, \dots, q\}$, is one-to-one.
2. Let the latent process $\{X_t\}$ be the $MA(2)$ process defined by

$$X_t = \sigma(\epsilon_t + \theta_2 \epsilon_{t-2}), \quad (11)$$

where the innovations ϵ_t are i.i.d. and have a standard normal distribution and $\sigma > 0$. The parameters of the model are (θ_2, σ) , and the parameter set $\Theta = \mathbb{R} \times \mathbb{R} \setminus \{0\}$ ensures that the mapping $\theta \mapsto \{c_X(\tau; \theta) : \tau \in \Gamma\}$, where $\Gamma = \{0, 2\}$, is one-to-one. Note that observing lag -1 is not required here as we have assumed $\theta_1 = 0$ in the model.

The definition of a *significant correlation contribution* constrains how much energy adds up for any fixed lag $\tau \in \Gamma$. We see directly from (3) that if we assume a significant correlation contribution, the expectation of the sample autocovariance sequence of $\tilde{X}_t$ does not vanish with the length of the observation N , at least for lags in Γ . This allows for consistent estimation of the parameter θ as we will see in Section 5. As a trivial counterexample, assume for instance that $c_g^{(N)}(\tau)$ goes to zero when N goes to infinity. Then $\hat{c}_{\tilde{X}}^{(N)}(\tau)$ in (2) goes to zero as well, independently of the parameter vector θ , either resulting in infeasible estimation or requiring a change of estimation approach.

Asymptotically stationary modulated processes are a subclass of modulated processes with a significant correlation contribution. Specifically, for the class of asymptotically stationary modulated processes, (7) requires that $c_g^{(N)}(\tau)$ converges to the non-zero quantity $R_g(\tau)$, which is a stronger requirement than (9), where we only require an asymptotic positive lower bound rather than convergence.

2.2. Missing Observations

A particularly enticing use of modulated processes is to account for missing observations in stationary time series. Let $\{X_t : t \in \mathbb{N}\}$ be a stationary process. For each time point $t \in \mathbb{N}$, we set (Parzen, 1963)

$$g_t = \begin{cases} 0 & \text{if } X_t \text{ is missing} \\ 1 & \text{if } X_t \text{ is observed} \end{cases} \quad (12)$$

The process $\tilde{X}_t = g_t X_t$ is formed at all time points $t \in \mathbb{N}$, forming a modulated process in the sense of Definition 1.

An example where the missing observation pattern is deterministic and leads to an asymptotically stationary modulated process is the case of (k, l) -periodically missing data treated by Jones (1962) and Parzen (1963). This corresponds to observing the k first values, missing the l next values, observing the k next values and so on. Note that Parzen (1963) requires $k > l$ for non-parametric estimation of the spectral density of X_t based on (8). Our model of modulated processes with significant correlation contribution allows for $k \leq l$, as long as we observe the lags in Γ . A generalization of this missing data scheme was introduced by Clinger and Ness (1976) with an application to oceanography.

Missing observations can also occur according to a random mechanism. This can be modelled by a random modulation sequence taking values zero and one (Scheinok, 1965; Bloomfield, 1970), when the random mechanism according to which missing points occur is independent from the observed process, which we shall assume. Conditioning on the observed modulation function, we then return to the deterministic modulating sequence described in this article. Most works, to our knowledge, have assumed some sort of stationarity for the random modulation sequence, that is, that the sample autocovariance of the modulation sequence converges almost surely to a non-zero value at all lags (Dunsmuir and Robinson, 1981a, c). Some authors do not require such an assumption but have treated only specific models, usually autoregressive models (Jones, 1980; Broersen *et al.*, 2004). The definition of a modulated process with a significant correlation contribution in such a situation needs to be understood in a probabilistic fashion, that is, we require that property 2 of Definition 3 be satisfied with probability one. Indeed, if one sees the general random experiment as a two-step experiment, where first the random modulating sequence $\{g_t\}$ is generated and observed and then a stationary process $\{X_t\}$ is modulated by this modulating sequence to produce $\{\tilde{X}_t\}$, then with probability one, the modulating sequence $\{g_t\}$ in the first step makes $\{\tilde{X}_t\}$ a modulated process with significant correlation contribution. Such a situation will be described by saying that $\{\tilde{X}_t\}$ is a modulated process with an almost surely significant correlation contribution. We shall now give a few examples of cases satisfying the stated conditions.

1. Let X_t be an $\text{AR}(p)$ Gaussian process with mean zero. If we set $\Gamma = \{0, \dots, p\}$, and if the missing data occur deterministically according to a (k, l) -periodic pattern, $k \geq p$ is a sufficient condition for the resulting modulated process to have a significant correlation contribution. This is because we are able to observe an infinite number of time the lags in Γ . We do not require any additional condition on l .
2. Let X_t be an $\text{AR}(p)$ process and consider the missing data scheme treated by Scheinok (1965), where the random mechanism is a sequence of Bernoulli i.i.d. trials with identical probability of success (to be understood as observation here) $0 < p \leq 1$. According to the strong central limit theorem, for all lag $\tau \in \mathbb{N}$, $c_g^{(N)}(\tau)$ converges a.s. to $p^2 > 0$ and therefore $\liminf_{N \rightarrow \infty} |c_g^{(N)}(\tau)| > 0$ a.s. Therefore, the observed process is a modulated process with an almost surely significant correlation contribution.
3. Consider the random mechanism where the sequence $\{g_t\}$ is generated according to

$$g_t \sim \mathcal{B}(p_t), \quad (13)$$

where $\mathcal{B}(p)$ represents the Bernoulli distribution with parameter p , and where we set

$$p_t = \mathcal{P} + A_p \cos(\omega_p t), \quad (14)$$

with $0 < \mathcal{P} < 1$, $0 \leq A_p < \min(\mathcal{P}, 1 - \mathcal{P})$ (which ensures $0 < \mathcal{P} - A_p \leq p_t \leq 1, \forall t \in \mathbb{N}$), and $\omega_p \in [-\pi, \pi]$. The Bernoulli parameters p_t as given by (14) will oscillate periodically around their mean value $\mathcal{P}$. This also leads to $\liminf_{N \rightarrow \infty} |c_g^{(N)}(\tau)| > 0$ a.s., using the fact that p_t is bounded below by $\mathcal{P} - A_p > 0$.

In Section 4.2, we will provide a simulation study based on example 3. This is novel in comparison with previously studied missing observation schemes as we do not make an assumption of stationarity for the process g_t (Dunsmuir and Robinson, 1981b).

2.3. Sampling Properties of Modulated Processes

We shall review and study some distributional properties of the periodogram of a modulated time series. Dunsmuir and Robinson (1981b) used the periodogram as the basis for designing pseudo-likelihood methods for asymptotically stationary modulated time series, with an emphasis on treating the problem of missing data. Similarly, in Section 3, we will use the results of this section to formulate a pseudo-likelihood using the periodogram, for our class of modulated processes with a significant correlation contribution. Herein we shall denote Ω_N as the set of Fourier frequencies $2\pi/N \cdot (-\lceil N/2 \rceil + 1, \dots, -1, 0, 1, \dots, \lfloor N/2 \rfloor)$.

We denote $\tilde{\mathbf{X}} = \{\tilde{X}_t : t = 0, \dots, N-1\}$ as a single realization of a length- N sample of a modulated process $\{\tilde{X}_t\}$ defined in Definition 1. The unobserved sample of the latent stationary process is denoted as $\mathbf{X} = \{X_t : t = 0, \dots, N-1\}$ accordingly. The squared modulus of the Fourier transform of the time series $\mathbf{X}$, known as the *periodogram*, is a common statistic in stationary time series analysis (Percival and Walden, 1993) and is given by

$$\hat{S}_X^{(N)}(\omega) = \frac{1}{N} \left| \sum_{t=0}^{N-1} X_t e^{-i\omega t} \right|^2, \quad \omega \in \mathbb{R}. \quad (15)$$

Note that this quantity is 2π -periodic, that is, $\hat{S}_X^{(N)}(\omega + 2\pi) = \hat{S}_X^{(N)}(\omega)$, $\omega \in \mathbb{R}$. The periodogram of the sample $\mathbf{X}$ is an asymptotically unbiased estimator of the spectral density of the stationary process $\{X_t\}$, that is, $\lim_{N \rightarrow \infty} \mathbb{E}\{\hat{S}_X^{(N)}(\omega; \boldsymbol{\theta})\} = 2\pi S_X(\omega; \boldsymbol{\theta})$ for all $\omega \in [-\pi, \pi]$ (Brockwell and Davis, 1991). However, the variance of the periodogram does not decrease to zero as the sample size increases. A consistent non-parametric estimator of a smooth spectral density $S_X(\omega; \boldsymbol{\theta})$ of the latent process $\{X_t\}$, were it to be directly observed, could be obtained by smoothing the periodogram across frequencies (Percival and Walden, 1993, pp. 235–253), as long as $S_X(\omega; \boldsymbol{\theta})$ is continuous.

For the modulated process $\{\tilde{X}_t\}$, the latent time series $\{X_t\}$ is not observed, so we instead compute the periodogram of the modulated (and observed) process itself, $\hat{S}_{\tilde{X}}^{(N)}(\omega)$, and we define the expected periodogram to be

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = \mathbb{E} \left\{ \hat{S}_{\tilde{X}}^{(N)}(\omega); \theta \right\}, \quad \omega \in \mathbb{R}.$$

Note that this quantity is also 2π -periodic. It is necessary to understand how modulation in the time domain will affect the expected periodogram. Proposition 1 gives more insight on how $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)$ relates to the modulating sequence g_t and the spectral density $S_X(\omega; \theta)$ of the latent stationary process $\{X_t\}$.

Proposition 1 (Expectation of the periodogram of a modulated time series). The expectation of the periodogram of the modulated time series takes the form

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = 2\pi \int_{-\pi}^{\pi} S_X(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda, \quad \forall \omega \in \mathbb{R}, \quad (16)$$

which is a periodic convolution. Here $S_g^{(N)}(\lambda)$ is the squared value of the Fourier transform of the finite sequence $\{g_t\}_{t=0, \dots, N-1}$, that is,

$$S_g^{(N)}(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=0}^{N-1} g_t e^{-i\lambda t} \right|^2,$$

which is defined for $\lambda \in \mathbb{R}$ and is 2π -periodic.

Proof

The proof for this proposition, which is a well-known result, can be found in Dunsmuir and Robinson (1981b, p. 562). $\square$

When $g_t = 1$ everywhere, which corresponds to observing the stationary latent process directly, the quantity $S_g^{(N)}(\lambda)$ is the usual Féjer kernel (Bloomfield, 2000) defined by

$$\mathcal{F}^{(N)}(\lambda) = \frac{\sin^2\left(\frac{N\lambda}{2}\right)}{2\pi N \sin^2\left(\frac{\lambda}{2}\right)}, \quad \forall \lambda \in \mathbb{R} \setminus \Omega_N, \quad \mathcal{F}^{(N)}(0) = \frac{N}{2\pi}, \quad \mathcal{F}^{(N)}(\lambda) = 0, \quad \forall \lambda \in \Omega_N \setminus \{0\}, \quad (17)$$

which behaves asymptotically (as N tends to infinity) as a Dirac delta-function centred at zero. This explains why the periodogram is, asymptotically, an unbiased estimator of the spectral density of a stationary process up to a multiplicative factor of 2π (Brockwell and Davis, 1991).

When g_t is such that the modulated process is asymptotically stationary, Dunsmuir and Robinson (1981c) approximate $\frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{i\omega\tau}$, where $\gamma(\tau) = R_g(\tau) c_X(\tau)$ using the notation of (7), for ω at Fourier frequencies by

$$\tilde{S}_{\tilde{X}}^{(D)}(\omega; \theta) = \frac{(2\pi)^2}{N} \sum_{\lambda \in \Omega_N} S_X(\omega - \lambda; \theta) S_g^{(N)}(\lambda). \quad (18)$$

When g_t is such that the modulated process $\tilde{X}_t$ has a significant correlation contribution, we derive the exact value of $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)$ by using the theoretical autocovariances of the latent model, in a similar fashion as in Sykulski *et al.* (2016b) for stationary processes. This is the result of Proposition 2, which follows.

Proposition 2 (Computation of the expected periodogram). Let $\omega \in \mathbb{R}$. We have

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = 2\mathcal{R} \left\{ \sum_{\tau=0}^{N-1} \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) e^{-i\omega\tau} \right\} - \bar{c}_{\tilde{X}}^{(N)}(0; \theta), \quad (19)$$

where $\bar{c}_{\tilde{X}}^{(N)}(\tau; \theta)$ is defined in (3). By defining $\bar{c}_{\tilde{X}}^{(N)}(-\tau; \theta) = \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta)$ for $\tau = 1, \dots, N-1$, we can (equivalently) express this relationship as

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = \sum_{\tau=-(N-1)}^{N-1} \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) e^{-i\omega\tau}.$$

Proof

The proof, which is standard (Brockwell and Davis, 1991, p. 334), follows directly from (15) and (3) in a few lines of algebra after aggregating along the diagonal of the covariance matrix. $\square$

Therefore, the expectation of the periodogram of $\tilde{X}$ is the discrete Fourier transform of the expected sample autocovariance sequence. This is true even though we have not assumed stationarity; it is simply a consequence of the relation between the formal definitions of (3) and (15). Note that calculating the Fourier transform of the sequence $\bar{c}_{\tilde{X}}^{(N)}(\tau; \theta)$ will always give a real-valued positive $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)$ for $\theta \in \Theta$, as the latter is defined as the expectation of the squared modulus of the Fourier transform of the process.

Proposition 2 can be used to compute the expected periodogram of an asymptotically stationary modulated process. In such cases, the difference between (18) and Proposition 2 is that (18) is a finite approximation of (16), whereas Proposition 2 is exact. The difference occurs because (18) does not account for the bias of the periodogram that results from leakage (Sykulski *et al.*, 2016b), whereas these effects are naturally accounted for in Proposition 2.

To justify the use of the expected periodogram in the setting of modulated processes with a significant correlation contribution, we now consider what conditions are required for the expected periodogram to *carry sufficient information* so that the parameter vector is identifiable within the parameter set Θ .

Proposition 3 (Identifiability of the expected periodogram). If the modulated process has a significant correlation contribution, the expected periodogram is a one-to-one (i.e. injective) mapping from the parameter set Θ to the set of non-negative continuous functions on $[-\pi, \pi]$, for a large enough sample size. More specifically, for two distinct parameter vectors θ and θ' , the expected periodograms $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)$ and $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta')$ cannot be equal for all Fourier frequencies $\frac{2\pi}{N} (-\lceil \frac{N}{2} \rceil + 1, \dots, -1, 0, 1, \dots, \lfloor \frac{N}{2} \rfloor)$ for N large enough.

Proof

Let $\theta, \tilde{\theta} \in \Theta$ be distinct parameter vectors, and let N be a positive integer. Let Γ be as given by Definition 3. By the assumption of significant correlation contribution, the finite sequences $\{c_X(\tau; \theta) : \tau \in \Gamma\}$ and $\{c_X(\tau; \tilde{\theta}) : \tau \in \Gamma\}$ are not equal. Since $\bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) = c_g^{(N)}(\tau) c_X(\tau; \theta)$ for $\tau \in \Gamma$, and according to (10), for N large enough, the sequences $\{\bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) : \tau \in \Gamma\}$ and $\{\bar{c}_{\tilde{X}}^{(N)}(\tau; \tilde{\theta}) : \tau \in \Gamma\}$ are not equal. Hence for N large enough, the sequences $\{\bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) : \tau = -(N-1), \dots, N-1\}$ and $\{\bar{c}_{\tilde{X}}^{(N)}(\tau; \tilde{\theta}) : \tau = -(N-1), \dots, N-1\}$ are not equal. Their finite Fourier transforms $\{\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) : \omega \in \Omega_N\}$ and $\{\bar{S}_{\tilde{X}}^{(N)}(\omega; \tilde{\theta}) : \omega \in \Omega_N\}$ are, by the bijective nature of the Fourier transform, not equal either. $\square$

This means that for two distinct parameters vectors $\theta, \tilde{\theta} \in \Theta$, we will have two distinct expected periodograms. This is a necessary condition for an estimation procedure based on the expected periodogram. We will propose such an estimation procedure in Section 3 and derive its consistency and convergence rate in Section 5.

2.4. Bivariate Modulated Processes

It is common in practical applications to observe more than one time series at any time and to analyse a set together. Often, the series in the set are related via phase shifts and other small temporal inhomogeneities; see, for example, Allen and Robertson (1996), Rünstler (2004), Allefeld *et al.* (2009), & Lilly and Olhede (2012). Bivariate non-stationary processes can be challenging to model, as both components may not be representable in the same non-stationary oscillatory family (Tong, 1973, 1974). To explore the nature of multi-variate modulation, we shall investigate the representation of bivariate processes. For ease of exposition, we shall represent such series using complex-valued time series (Walker, 1993). We shall continue to assume that the latent process, now denoted as Z_t for complex-valued processes, is Gaussian and zero-mean, leaving only the second-order structure to be modelled. For complex-valued processes, both the autocovariance $c_Z(\tau) = E\{Z_t^* Z_{t+\tau}\}$ (the star denotes conjugation) and the relation $r_Z(\tau) = E\{Z_t Z_{t+\tau}\}$ sequences need to be modelled (Walden, 2013). Complex-valued processes, unlike real-valued, no longer have a spectrum that needs to satisfy Hermitian symmetry, and if the series represents motion in the plane, the positive and negative frequencies represent anti-clockwise and clockwise rotations respectively. Following the classical modelling framework (Miller, 1969) for complex-valued processes, we shall assume that the relation sequence takes the value zero for all lags. The complex-valued process is then said to be *proper*, which is equivalent to the isotropy of the corresponding bivariate real-valued process. The assumption of propriety has the consequence of directly extending equation (1) to the complex-valued case from the real-valued case. Specifically, let Z_t be a complex-valued Gaussian proper zero-mean process; a complex-valued modulated process is defined as one taking the form

$$\tilde{Z}_t = g_t Z_t, \quad (20)$$

at all times $t \in \mathbb{N}$, where $g_t = \rho_t e^{i\phi_t}$ is a bounded modulation sequence. We note that for complex-valued time series, the modulation sequence is complex-valued. With this definition, the modulation series g_t accomplishes a time-dependent rescaling or expansion/dilation, from ρ_t , together with a time-dependent rotation, from $e^{i\phi_t}$.

The autocovariance of the complex-valued modulated process $\tilde{Z}_t$ at times t_1 and t_2 is given by the conveniently simple form,

$$c_{\tilde{Z}}(t_1, t_2; \boldsymbol{\theta}) = E\{\tilde{Z}_{t_1}^* \tilde{Z}_{t_2}; \boldsymbol{\theta}\} = g_{t_1}^* g_{t_2} c_Z(t_2 - t_1; \boldsymbol{\theta}) = \rho_{t_1} \rho_{t_2} e^{i(\phi_{t_2} - \phi_{t_1})} c_Z(t_2 - t_1; \boldsymbol{\theta}),$$

and $c_{\tilde{Z}}(t_1, t_2; \boldsymbol{\theta})$ fully characterizes the process. Note that this quantity is not only a function of the lag $t_2 - t_1$ as the process is no longer stationary. Similarly to the univariate case (cf. (4)), let N be any positive integer, and we define for $\tau = 0, \dots, N-1$,

$$c_g^{(N)}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} g_t^* g_{t+\tau}. \quad (21)$$

Note that when g_t is real-valued, (21) and (4) are the same. We also extend the notion of a significant correlation contribution for complex-valued modulated processes, which naturally mimics Definition 3. We define the expected periodogram of a complex-valued modulated time series as $\bar{S}_{\tilde{Z}}^{(N)}(\omega; \boldsymbol{\theta}) = E\{\hat{S}_{\tilde{Z}}^{(N)}(\omega; \boldsymbol{\theta})\}$, which can be computed efficiently similarly to Proposition 2 for the univariate case, by replacing $\tilde{X}_t$ by $\tilde{Z}_t$ in (3) and (19).

A univariate real-valued modulated process is stationary if and only if the modulating sequence is a constant. A necessary and sufficient condition on the modulating sequence for the complex-valued modulated process (20) to be stationary is more complicated to obtain and is determined in the following proposition.

Proposition 4 (Stationary bivariate modulated processes). Let $\tilde{Z}_t$ be the complex-valued modulated process defined in (20). First, assume the latent process $\{Z_t\}$ is a white noise process. Then the modulated process $\{\tilde{Z}_t\}$ is

stationary if and only if the modulating sequence $g_t = \rho_t e^{i\phi_t}$ is of constant modulus, that is, $\rho_t = a \geq 0$. In such case, the modulated process is a white noise process with variance $a^2 \mathbb{E}\{|Z_0|^2\}$.

More generally, assume the stationary latent process $\{Z_t\}$ is not a white noise process, and let $\mu = \gcd\{\tau \neq 0 \in \mathbb{N} : |c_Z(\tau; \theta)| > 0\}$, where $\gcd$ denotes the greatest common divisor. Then the modulated process is stationary if and only if $\{g_t\}$ is zero everywhere or if there exists two constants $a > 0$ and $\gamma \in [-\pi, \pi)$ such that for all $t \in \mathbb{N}$, letting $r = t \bmod \mu$ be the remainder of t divided by μ ,

$$\begin{aligned}\rho_t &= a \\ \phi_t &= \phi_r + \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \bmod 2\pi,\end{aligned}$$

where $\left\lfloor \frac{t}{\mu} \right\rfloor$ denotes the floor of $\frac{t}{\mu}$ and $\bmod 2\pi$ indicates that the equality is true up to an additive multiple of 2π . In this case, the spectral density of the modulated process $\{\tilde{Z}_t\}$ is

$$S_{\tilde{Z}}(\omega) = a^2 S_Z\left(\omega - \frac{\gamma}{\mu}\right).$$

Proof

See Appendix A.1. □

The value of μ in Proposition 4 depends on the location of zeros in the covariance sequence of the latent process. In particular, if $|c_Z(1; \theta)| > 0$, then $\mu = 1$ and $\tilde{Z}_t$ is stationary only if there exists a constant $\gamma \in \mathbb{R}$ such that for all $t \in \mathbb{N}$, $\phi_t = \phi_0 + \gamma t \bmod 2\pi$. If $|c_Z(2; \theta)| > 0$ but $|c_Z(\tau; \theta)| = 0$ for all $\tau \in \mathbb{N}$, $\tau \neq 0, 2$, then $\mu = 2$ (this can occur with a second-order moving average process for instance). In that case, the modulated process $\tilde{Z}_t$ is stationary if and only if there exists a constant $\gamma \in [-\pi, \pi)$ such that for all $t \in \mathbb{N}$, $\phi_t = \phi_0 + \gamma \frac{t}{2} \bmod 2\pi$ if t is even, or $\phi_t = \phi_1 + \gamma \frac{t-1}{2} \bmod 2\pi$ if t is odd.

2.4.1. A Time-varying Bivariate Autoregressive Process

We now introduce the specific non-stationary bivariate autoregressive model that will be used in our real-world data application. We consider the discrete-time complex-valued process $\{\tilde{Z}_t : t \in \mathbb{N}\}$, defined by

$$\begin{aligned}\tilde{Z}_t &= r e^{i\beta_t} \tilde{Z}_{t-1} + \epsilon_t, \quad t \geq 1, \quad 0 \leq r < 1, \quad \beta_t \in \mathbb{R}, \\ \tilde{Z}_0 &\sim \mathcal{N}_C\left(0, \frac{\sigma^2}{1-r^2}\right), \quad \sigma > 0, \\ \epsilon_t &\sim \mathcal{N}_C(0, \sigma^2),\end{aligned}\tag{22}$$

where $\mathcal{N}_C(0, \sigma^2)$ denotes the complex-valued normal distribution with mean 0 and variance σ^2 , and with i.i.d. real and imaginary parts. Note that the real and imaginary parts of ϵ_t then each have variance $\sigma^2/2$. Here $0 \leq r < 1$ is commonly known as either the autoregressive or damping parameter, ensuring the mean-reversion of the process. By mean-reversion, we mean that, given any time $t \in \mathbb{N}$, we have $\lim_{\tau \rightarrow \infty} \mathbb{E}\{\tilde{Z}_{t+\tau} | \tilde{Z}_t\} = 0$, that is, irrespective of the size of the perturbation ϵ_t at time t , the process is expected to return to its mean. This is seen from the following inductive relationship:

$$\tilde{Z}_{t+\tau} = r^\tau e^{i \sum_{j=1}^{\tau} \beta_{t+j}} \tilde{Z}_t + \sum_{j=1}^{\tau} r^{\tau-j} e^{i \sum_{k=j+1}^{\tau} \beta_k} \epsilon_{t+j}, \quad \tau \geq 0,$$

which leads to

$$\mathbb{E}\{\tilde{Z}_{t+\tau} | \tilde{Z}_t\} = r^\tau e^{i \sum_{j=1}^{\tau} \beta_{t+j}} \tilde{Z}_t,$$

which goes to zero exponentially as τ goes to infinity, since $0 \leq r < 1$. A damping parameter r close to 1 will lead to a slowly decaying autocorrelation sequence. A value of r close to 0 will lead to a process with very short memory, with the limiting behaviour of a white noise process as $r \rightarrow 0$. The parameter β_t is a known, dimensionless time-varying frequency, which we shall take within the interval $[-\pi, \pi)$ without loss of generality.

Process (22) is a non-stationary version of the complex-valued first-order autoregressive process (Sykulski *et al.*, 2016a) introduced by Le Breton (1988), and also a discrete-time analogue of the complex-valued Ornstein–Uhlenbeck (OU) process (Arató *et al.*, 1962) with time-varying oscillation frequency. We now prove in Proposition 5 that the model defined in (22) belongs to our class of bivariate modulated processes.

Proposition 5 (Modulated process representation). Let $\{\tilde{Z}_t\}$ be the process defined in (22). There exists a unit-magnitude complex-valued modulating sequence g_t , and a stationary complex-valued proper process $\{Z_t\}$ such that $\{\tilde{Z}_t\}$ is the modulation of $\{Z_t\}$ by the non-random sequence $\{g_t\}$. More explicitly, we have $\tilde{Z}_t = g_t Z_t$, for all $t \in \mathbb{N}$, where

$$\begin{aligned} g_t &= e^{i \sum_{u=1}^t \beta_u}, \\ Z_t &= r Z_{t-1} + \epsilon'_t, \quad t \geq 1, \end{aligned} \quad (23)$$

and $Z_0 \sim \mathcal{N}_C(0, \sigma^2/(1-r^2))$. The process ϵ'_t is a Gaussian white noise process with the same properties (zero-mean, variance σ^2 and independence of real and imaginary parts) as those of ϵ_t . Defined as such, the latent complex-valued process Z_t is stationary and proper.

Proof

See Appendix A.2. □

The stationary latent process Z_t defined in (23) is a stationary complex-valued first-order autoregressive process and is Gaussian. Its autocovariance sequence is given by

$$c_Z(\tau) = \frac{\sigma^2}{1-r^2} r^\tau, \quad \tau \in \mathbb{Z}.$$

It is easy to verify that the mapping $(r, \sigma) \mapsto (c_Z(0), c_Z(1))$ is a one-to-one mapping. In the following proposition, we stipulate a sufficient condition on the frequencies β_t so that the process defined in (22) satisfies our assumption of significant correlation contribution, when represented as a modulated process as in Proposition 5.

Proposition 6. Let $\tilde{Z}_t$ be the process defined by (22). Assume that there exists $\Xi \in [-\pi, \pi)$ and $0 \leq \kappa < \frac{\pi}{2}$ such that for all $t \in \mathbb{N}$, $|\beta_t - \Xi| \leq \kappa$. Then $\tilde{Z}_t$ is a modulated process with significant correlation contribution.

Proof

See Appendix A.3. □

Hence with some conditions on β_t , the complex-valued autoregressive process defined by (22) belongs to the class of processes with a significant correlation contribution, and the expected periodogram is a one-to-one mapping from the parameter set $[0, 1) \times [0, \infty)$ to the set of non-negative continuous functions on $[-\pi, \pi]$, according to Proposition 3 – a proposition that is readily extended to the complex-valued processes of this section.

3. PARAMETRIC ESTIMATION OF MODULATED PROCESSES

We have explored a class of univariate and bivariate modulated processes. The next stage is to describe their efficient inference. In this section, we describe how the parameters of the latent model for $\{X_t\}$ can be inferred from observing a single realization of the modulated process $\{\tilde{X}_t\}$. Most authors have focused on the problem of estimating modulated processes under the assumption of asymptotic stationarity as defined in Definition 2 (Parzen,

1963; Dunsmuir and Robinson, 1981a, b; Toloï and Moretton, 1989). Although non-parametric estimates have been the key concern in most of the relevant literature, there have been instances where parametric estimation has been considered; see, for instance, Dunsmuir and Robinson (1981a). Parametric estimation ensures that the estimated autocovariance sequence is non-negative definite, as opposed to using non-parametric estimates of the form given in (8). Parametric estimation is also preferable when the true model is known, as it uses the observed degrees of freedom more efficiently. Herein we consider the problem of parametric estimation for our class of modulated processes with a significant correlation contribution, which, we recall, is a generalization of asymptotically stationary modulated processes. We propose an adaptation of the Whittle likelihood (Whittle, 1953), based on the expected periodogram.

We wish to infer the parameter vector θ of the latent univariate stationary process $\{X_t\}$ within the parameter set Θ , based on the sample $\tilde{\mathbf{X}} = \tilde{X}_0, \dots, \tilde{X}_{N-1}$ and the known modulating sequence $\{g_t : t = 0, \dots, N-1\}$. Because it has been assumed that the latent process is a zero-mean Gaussian process, the same is true for the modulated process. The vector $\tilde{\mathbf{X}}$ is multi-variate Gaussian with an expected $N \times N$ autocovariance matrix $C_{\tilde{X}}(\theta) = \{c_{\tilde{X}}(t_1, t_2; \theta)\}$ for $t_1, t_2 = 0, \dots, N-1$, where the components of this matrix are given by $c_{\tilde{X}}(t_1, t_2; \theta) = g_{t_1} g_{t_2} c_X(t_2 - t_1; \theta)$. However, the parameter vector θ of the latent process $\{X_t\}$ can be uniquely determined from the modulated process $\{\tilde{X}_t\}$ only if $\theta \rightarrow \{c_{\tilde{X}}(t_1, t_2; \theta) : t_1, t_2 \in \mathbb{N}\}$ is injective, that is, there is no $\theta' \in \Theta$ such that $\theta \neq \theta'$ and $c_{\tilde{X}}(t_1, t_2; \theta) = c_{\tilde{X}}(t_1, t_2; \theta') \forall t_1, t_2 \in \mathbb{N}$. This necessary condition is clearly achieved under the assumption of a modulated process with significant correlation contribution. The negative of the exact time-domain Gaussian log-likelihood is proportional to

$$\ell_G(\theta) = \frac{1}{N} \log |C_{\tilde{X}}(\theta)| + \frac{1}{N} \tilde{\mathbf{X}}^T C_{\tilde{X}}(\theta)^{-1} \tilde{\mathbf{X}}, \quad (24)$$

where $|C_{\tilde{X}}(\theta)|$ denotes the determinant of $C_{\tilde{X}}(\theta)$. Note that one may need to remove from $\tilde{\mathbf{X}}$ points where g_t is zero, to ensure that the determinant of the covariance matrix is non-zero, and since those observations carry no information about θ . We minimize ℓ_G to obtain the time-domain maximum likelihood estimator, that is,

$$\hat{\theta}_G^{(N)} = \arg \min_{\theta \in \Theta} \ell_G(\theta).$$

Parameter estimation based on time-domain likelihood has several drawbacks in the context of modulated processes. For a large sample size N , computing the determinant of the covariance matrix is expensive, requiring $\mathcal{O}(N^3)$ elementary operations in general (although in specific cases such as for Markovian processes, the likelihood is obtained in only $\mathcal{O}(N)$ computations). Moreover, each computation of the parametric covariance matrix $C_{\tilde{X}}(\theta)$ within the exact likelihood requires $\mathcal{O}(N^2)$ operations, compared with $\mathcal{O}(N)$ operations in the case of a stationary process.

We propose a computationally efficient estimation method for the parameters of the latent model based on the periodogram of the modulated time series. First, recall that for the stationary time series $\{X_t\}$, making use of the Toeplitz property of the autocovariance matrix, one can approximate the negative log-likelihood using the Whittle likelihood (Whittle, 1953), which once discretized is evaluated by

$$\ell_W(\theta) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log S_X(\omega; \theta) + \frac{\hat{S}_X^{(N)}(\omega)}{S_X(\omega; \theta)} \right\}, \quad (25)$$

where again Ω_N is the set of Fourier frequencies $\frac{2\pi}{N} \cdot (-\lceil \frac{N}{2} \rceil + 1, \dots, -1, 0, 1, \dots, \lfloor \frac{N}{2} \rfloor)$. This pseudo-likelihood has the benefit of $\mathcal{O}(N \log N)$ computational complexity using the discrete Fourier transform, with the resulting maximum pseudo-likelihood estimator being asymptotically equivalent to the time-domain maximum likelihood estimator. We adapt this pseudo-likelihood procedure to modulated processes with significant correlation contribution in the following definition.

Definition 4 (Spectral maximum pseudo-likelihood estimator for univariate modulated processes). Let $\{\tilde{X}_t\}$ be a modulated process with significant correlation contribution, and let $\tilde{\mathbf{X}}$ be its length- N sample. We define the following pseudo-likelihood:

$$\ell_M(\boldsymbol{\theta}) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) + \frac{\hat{S}_{\tilde{X}}^{(N)}(\omega)}{\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta})} \right\}, \quad (26)$$

where $\bar{S}_{\tilde{X}}^{(N)}(\omega)$ is defined in Section 2.3 as the expectation of the periodogram of the modulated time series and is computed using Proposition 2. The corresponding estimator of the parameter vector $\boldsymbol{\theta}$ is obtained by a minimization procedure over the parameter set,

$$\hat{\boldsymbol{\theta}}_M^{(N)} = \arg \min_{\boldsymbol{\theta} \in \Theta} \ell_M(\boldsymbol{\theta}).$$

The sequence $\{c_g^{(N)}(\tau) : \tau = 0, \dots, N-1\}$ defined in (4) and necessary to the computation of the expected periodogram requires $\mathcal{O}(N \log N)$ computations, as it can be computed as the biased sample autocovariance sequence of $g_0, \dots, g_{N-1}$ via a fast Fourier transform. This initial step is carried out independently of inferring the parameter of interest $\boldsymbol{\theta}$. Then any computation of $\{\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) : \omega \in \Omega_N\}$ for any value of the parameter vector $\boldsymbol{\theta}$ will require $\mathcal{O}(N \log N)$ computations, since we can compute $\{\bar{c}_{\tilde{X}}^{(N)}(\tau; \boldsymbol{\theta}) : \tau = 0, \dots, N-1\}$ in $\mathcal{O}(N)$ computations using (3) and the precomputed $\{c_g^{(N)}(\tau) : \tau = 0, \dots, N-1\}$, and the quantity $\{\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) : \omega \in \Omega_N\}$ is then computed via a fast Fourier transform (Proposition 2). The reason for separating the initial step of computing $\{c_g^{(N)}(\tau) : \tau = 0, \dots, N-1\}$ from the rest of the computation is that it is carried out independently of the parameter value, and therefore outside any call to a minimization procedure over the parameter set Θ involving the expected periodogram.

In the trivial case of a modulation sequence equal to 1 everywhere, then the pseudo-likelihood of Definition 4 does not exactly equal the Whittle likelihood of (25). This is because the spectral density $S_X(\omega)$ would be replaced by the expected periodogram $\bar{S}_{\tilde{X}}^{(N)}(\omega)$, which is the convolution of the true spectral density with the Fejér kernel (equation (18)). For stationary time series, this type of estimator was investigated in Sykulski *et al.* (2016b) and was found to significantly reduce bias and error in parameter estimation as compared with standard Whittle estimation. For modulated processes that are asymptotically stationary, this signifies the difference between using (18) and the quantity defined by Proposition 2 to fit the periodogram.

The same estimator to Definition 4 can be used for the complex-valued proper time series $\tilde{Z}_t$ considered in Section 2.4, that is, we define our estimator

$$\hat{\boldsymbol{\theta}}_M^{(N)} = \arg \min_{\boldsymbol{\theta} \in \Theta} \ell_M(\boldsymbol{\theta}), \quad (27)$$

with the objective function given by

$$\ell_M(\boldsymbol{\theta}) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\tilde{Z}}^{(N)}(\omega; \boldsymbol{\theta}) + \frac{\hat{S}_{\tilde{Z}}^{(N)}(\omega)}{\bar{S}_{\tilde{Z}}^{(N)}(\omega; \boldsymbol{\theta})} \right\}. \quad (28)$$

The comments on computational aspects hold for the complex-valued case as well. In Section 5, we will prove consistency of the frequency-domain estimator and its $\mathcal{O}(N^{-1/2})$ convergence rate.

4. APPLICATIONS

4.1. Application to Oceanographic Data

We analyse real-world data from the Global Drifter Program. Specifically, we model jointly the latitudinal and longitudinal velocities obtained from instruments known as drifters, which drift freely according to ocean surface flows (Sykulski *et al.*, 2016c). Those velocities are modelled as the aggregation of two independent complex-valued processes, one of which is non-stationary and which we model as the complex-valued AR(1) process described in Section 2.4.1. We use our estimator (27) to infer physical quantities that describe the ocean surface currents. To scrutinize our results and compare with alternative approaches, we also present two simulation studies, the first one being based on a dynamical model of the ocean surface currents and the second one being a simulated version of the model of Section 2.4.1. All data and code used in this article are available for download at <http://www.ucl.ac.uk/statistics/research/spg/software>, and all results in this section and Section 4.2 are exactly reproducible.

4.1.1. The Global Drifter Program

The Global Drifter Program database (www.aoml.noaa.gov/phod/dac) is a collection of measurements obtained from buoys known as surface drifters, which drift freely with ocean currents and regularly communicate measurements to passing satellites at unequally spaced time intervals averaging 1.4 hours since 2005. The data are then interpolated onto a regular temporal grid using the approach of Elipot *et al.* (2016). The measurements include position, and sea surface temperature. In total, over 11,000 drifters have been deployed since 2005, with approximately 70 million position recordings obtained. The analysis of this data is crucial to our understanding of ocean circulation (Lumpkin and Pazos, 2007), which is known to play a primary role in determining the global climate system; see, for example, Andrews *et al.* (2012). Furthermore, Global Drifter Program data are used to understand the dispersion characteristics of the ocean, which are critical in correctly modelling oil spills (Abascal *et al.*, 2010) and more generally assist in developing theoretical understanding of ocean fluid dynamics (Griffa *et al.*, 2007), which is necessary for global climate modelling.

In Figure 1(a), we display in the left panel the trajectories of 200 drifters, which either traverse or are near the equator, interpolated for this application onto a 2-hour grid from raw position fixes available at the Global Drifter Program web site. We focus on a single drifter trajectory, drifter ID 43594, in Figure 1(b, c), displaying both its latitudinal position and velocity respectively, the latter of which is obtained by differencing the positions. This velocity time series is non-stationary, as it has oscillations that appear to be modulated and change in frequency over time. The oscillations are known as *inertial oscillations* – one of the most ubiquitous and readily observable features of the ocean currents accounting for approximately half of the kinetic energy in the upper ocean (Ferrari and Wunsch, 2009). Inertial oscillations arise owing to the deviation of the rotating Earth from a purely spherical geometry, together with the appearance of the Coriolis force in the rotating reference frame of an Earth-based observer (Early, 2012). The modulation of these oscillations occurs because the drifters are changing latitude – and the *Coriolis frequency*, denoted as f , is equal to twice the rotation rate of the Earth Ω , multiplied by the sine of the latitude ζ , that is, $f = 2\Omega \sin \zeta$ radians per second. The rotation rate of the Earth Ω is computed as $2\pi/T$, where T is one sidereal day in seconds. Note that the Coriolis frequency f is a signed quantity, implying that oscillations occur in opposite rotational clockwise/anti-clockwise directions from one hemisphere to the other. The Coriolis frequency is positive in the Northern hemisphere, whereas the oscillations occur in the mathematically negative sense. Therefore, we define the inertial frequency $\omega^{\{f\}} = -f/2\pi K$ as the negative of the Coriolis frequency divided by $2\pi K$, where K is one solar day in seconds, so that $\omega^{\{f\}}$ is in cycles per day. The entire drifter dataset is split into segments of 60 inertial periods in length, accounting for the variation of the inertial period along drifter trajectories, and with 50% overlap between segments. The standard deviation of the inertial frequency along each data segment is taken, and the 200 segments exhibiting the largest ratio of the standard deviation of the inertial frequency, to the magnitude of its mean value along the segment, are identified for use in this study. These exhibit the largest fractional changes in the inertial frequency and, as shown in Figure 1(a), are located in the vicinity of the equatorial region where inertial frequency vanishes.

4.1.2. Stochastic Modelling

The stochastic modelling of Lagrangian trajectories was investigated in Sykulski *et al.* (2016c), where the term ‘Lagrangian’ is used because the moving object making the observations (i.e. the drifter) is the frame of reference, as opposed to fixed-point measurements known as *Eulerian* observations. In that article, the Lagrangian velocity time series was modelled as a stationary Gaussian complex-valued time series, with the following six-parameter power spectral density

$$S(\omega) = \frac{A^2}{(\omega - \omega^{\{f\}})^2 + \lambda^2} + \frac{B^2}{(\omega^2 + h^2)^\alpha}, \quad \omega \in \mathbb{R}, \quad (29)$$

$$A > 0, \quad \lambda > 0, \quad h > 0, \quad B > 0, \quad \alpha > \frac{1}{2},$$

where ω is given in cycles per day.

The first component of (29) is the spectral density of a complex-valued OU process (Arató *et al.*, 1962) and is used to describe the effect of inertial oscillations at frequency $\omega^{\{f\}}$. Denoting $\tilde{Z}_{OU}(t)$ the OU component, where $\tilde{Z}_{OU}(t)$ is complex-valued, these oscillations are described by the following stochastic differential equation:

$$d\tilde{Z}_{OU}(t) = (-\lambda + i2\pi\omega^{\{f\}})\tilde{Z}_{OU}(t)dt + AdW(t), \quad (30)$$

where t is expressed in days and $W(t)$ is a complex-valued Brownian process with independent real and imaginary parts. The damping parameter $\lambda > 0$ ensures that the OU process is mean-reverting. The corresponding continuous complex-valued autocovariance is given by

$$s(\tau) = \frac{A^2}{2\lambda} \exp\{-\lambda|\tau| + i2\pi\omega^{\{f\}}\tau\},$$

and the sampled process (Arató *et al.*, 1999) $\tilde{Z}_{OU,t} = \tilde{Z}_{OU}(t\Delta)$, where $\Delta = 1/12$ day is the sampling rate corresponding to the 2-hour grid, is a complex-valued AR(1),

$$\tilde{Z}_{OU,t} = r e^{i2\pi\omega^{\{f\}}\Delta} \tilde{Z}_{OU,t-1} + \epsilon_t. \quad (31)$$

Here ϵ_t is a Gaussian complex-valued white noise process with variance σ^2 and independent real and imaginary parts. The autocovariance sequence of the stationary sampled process is given by

$$c_{\tilde{Z}_{OU}}(\tau) = \frac{\sigma^2}{1-r^2} r^\tau e^{i2\pi\tau\Delta\omega^{\{f\}}}. \quad (32)$$

The transformation between the parameters of the complex-valued OU and the complex-valued AR(1) are given by

$$\sigma^2 = \frac{A^2(1-e^{-\lambda\Delta})}{2\lambda\Delta}, \quad r = e^{-\lambda\Delta}. \quad (33)$$

The second component of (29) is the spectral density of a stationary proper Matérn process (Gneiting *et al.*, 2010), denoted as $Z_{M,t}$, and is used to describe two-dimensional background turbulence (Lilly *et al.*, 2016). Although the parameter $\omega^{\{f\}}$ is varying as the drifter changes latitude, this parameter is fixed to its mean value in each trajectory segment in Sykulski *et al.* (2016c). This leaves five remaining parameters to estimate, $\{A, \lambda, B, h, \alpha\}$, in different regions of the ocean.

The model of (29) is stationary—slowly varying non-stationarity in the data is accounted for by windowing the data into chunks of approximately 60 inertial periods and treating the process as locally stationary within

each window. The estimated parameters can then be aggregated spatially to quantify the heterogeneity of ocean dynamics. This method works well on relatively quiescent and stationary regions of the ocean; however, this method cannot account for the rapidly varying non-stationarity evident in Figure 1 and leads to model misfit and biased parameter estimates, as we shall now investigate in detail.

4.1.3. Modulated Time Series Modelling and Estimation

We now apply the methodological contributions of this article to improve the model of (29) for highly non-stationary time series, such as those observed in Figure 1(a). We do this by accounting for changes in the inertial frequency, $\omega^{\{f\}}$, within each window of observation. We denote $\omega^{\{f\}}(t)$ as the continuous time-varying inertial frequency and $\omega_t^{\{f\}} = \omega^{\{f\}}(t\Delta)$ the inertial frequency value at each observed time step, $t = 0, \dots, N - 1$. The adapted version of the stochastic differential equation (30) is then given by

$$d\tilde{Z}_{OU}(t) = \left(-\lambda + i2\pi\omega^{\{f\}}(t)\right)\tilde{Z}_{OU}(t)dt + AdW(t). \quad (34)$$

In analogue to the proof in Appendix A.2, it is shown that the sampled process $\tilde{Z}_{OU,t} = \tilde{Z}_{OU}(t\Delta)$ satisfies

$$\tilde{Z}_{OU,t} = re^{i2\pi\int_{\Delta(t-1)}^{\Delta t}\omega^{\{f\}}(u)du}\tilde{Z}_{OU,t-1} + \epsilon_t.$$

As the inertial frequency is only observed at sampled points, we approximate the term $\int_{\Delta(t-1)}^{\Delta t}\omega^{\{f\}}(u)du$ by $\Delta\omega_t^{\{f\}}$. Specifically, we use the model of (22) for complex-valued time series, that is,

$$\tilde{Z}_{OU,t} = re^{i2\pi\Delta\omega_t^{\{f\}}}\tilde{Z}_{OU,t-1} + \epsilon_t, \quad t \geq 1, \quad (35)$$

where ϵ_t has the same properties as in (31) and the transformation between the parameters $\{A, \lambda, \omega_t^{\{f\}}\}$ of the non-stationary complex-valued OU process (34) and the parameters $\{r, \sigma, \omega_t^{\{f\}}\}$ of the non-stationary complex-valued AR(1) process (35) are given by (33).

The required methodology has been developed in Section 2.4 for bivariate (or complex-valued) time series. We only perform the modulation on the complex OU component in (29); the Matérn component for the turbulent background is unchanged and is considered to be stationary in the window, as it is not in general affected by changes in $\omega^{\{f\}}$. The two components are however observed in aggregation, and for this reason, we cannot simply demodulate the observed non-stationary signal to recover a stationary signal. Instead, to jointly estimate the parameters $\{A, \lambda, B, h, \alpha\}$, we first compute the modulating sequence, g_t , using (23) in Proposition 5 and accounting for the temporal sample rate Δ :

$$g_t = e^{i\sum_{u=1}^t 2\pi\Delta\omega_u^{\{f\}}}, \quad (36)$$

for $t = 0, \dots, N - 1$. Then we obtain the expected periodogram of the OU component, by computing $c_g(\tau)$ according to (21); then $\bar{c}_{\tilde{Z}_{OU}}^{(N)}(\tau)$, where we use the autocovariance of a stationary complex-valued AR(1) process with

$$c_{Z_{OU}}(\tau; r, \sigma) = \frac{\sigma^2}{1 - r^2} r^\tau,$$

and Fourier transform according to (19). Next, we compute the expected periodogram of the stationary Matérn as outlined in Sykulski *et al.* (2016c). Note that this can also be computed from the autocovariance of a Matérn,

which we denote as $c_{Z_M}(\tau)$, using (19), and by setting $g_t = 1$ for all t , from which the modulation kernel (4) is simply the triangle kernel $1 - |\tau|/N$. Finally, we additively combine the expected periodograms, that is,

$$\bar{S}^{(N)}(\omega; \theta) = \sum_{\tau=-(N-1)}^{N-1} \left[c_g^{(N)}(\tau) c_{Z_{OU}}(\tau) + \left(1 - \frac{|\tau|}{N}\right) c_{Z_M}(\tau) \right] e^{-i2\pi\omega\tau\Delta}, \quad \omega \in \left[-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right]$$

and then minimize the objective function, given in (28), to obtain parameter estimates for $\{A, \lambda, B, h, \alpha\}$.

Note that the modulation of a complex-valued AR(1) process by (36) will not lead to an asymptotically stationary process, as in general we cannot expect the quantities $c_g^{(N)}(\tau)$ to converge. However, we can see from Figure 1(a) that the drifters of our dataset have latitudes comprised between $\pm 20^\circ$. Therefore, the terms $2\pi\Delta\omega_t^{\{f\}}$ in (35) are comprised between ± 0.3591 radians (remembering the formula for the Coriolis frequency as a function of the latitude), so that the conditions of Proposition 6 are verified. Hence the sampled inertial component is a modulated process with a significant correlation contribution, which justifies the use of our estimator (27). Note that this results from both the latitudes of the drifters and the sampling rate used.

The assumption of Gaussianity is reasonable for modelling the velocity of instruments from the GDP as is discussed in Section 2.4 of LaCasce (2008) and references therein. To further inspect this, we tested the hypothesis of Gaussianity for the specific equatorial drifter velocity dataset that is the subject of our study. The code for this test is available online. Because the time series values are correlated in time, it would be incorrect to apply a Gaussianity test to all the velocities, as such tests typically require the samples to be independent (Paparoditis and Politis, 2012). Therefore, we perform our tests on the differenced process (i.e. the accelerations, noting that if the accelerations are Gaussian, then the velocities are Gaussian too) where correlation will decay more quickly in time. Furthermore, to account for any remaining correlations, we subsample the differenced time series. To choose a relevant sampling step, we selected a decorrelation length common to all our velocity time series based on an analysis inspired by Paparoditis and Politis (2012). We averaged the biased sample autocorrelation sequences and selected the first lag from which the averaged autocorrelation sequence is within the 95% confidence interval obtained under the hypothesis of zero correlation. The subsamples are tested via a Kolmogorov–Smirnov test with size 5%. We obtain a rejection rate of 6.5% for the overall dataset (including both longitudinal and latitudinal velocities), which is broadly consistent with a type I error rate of 5%. This shows that the assumption of Gaussianity is reasonable.

4.1.4. Parameter Estimation with Equatorial Drifters

We now compare the likelihood estimates and parametric fits for the stationary model (29) with those for the non-stationary version of this model described in the previous section. In particular, the damping timescale $1/\lambda$ is of primary interest in oceanography (Elipot *et al.*, 2010). In Figure 2, we display the frequency likelihood fits of each model to segments of data from drifter IDs 79243, 54656 and 71845, all of which are among the trajectory segments displayed in the left-hand panel of Figure 1. We also include model fits to a 60-inertial period window of drifter ID 44312, which is investigated in detail in Sykulski *et al.* (2016c), as this South Pacific drifter is from a more quiescent region of the ocean and does not exhibit significant changes in $\omega_t^{\{f\}}$. For the South Pacific drifter in Figure 2(d), both fits are almost equivalent (and hence are overlaid), capturing the sharp peak in inertial oscillations at approx 1.2 cycles per day. For the three equatorial drifters, the stationary model (29) has been fit with the inertial frequency set to the average of $\omega_t^{\{f\}}$ across the window. The stationary model is a relatively poor fit to the observed time series spectra. The non-stationary modulated model, which incorporates changes in $\omega_t^{\{f\}}$, is a better fit, capturing the spreading of inertial energy between the maximum and minimum values of $\omega_t^{\{f\}}$.

In this analysis, we have excluded frequencies higher than 0.8 cycles per day from all the likelihood fits to the equatorial drifters (the Nyquist is 6 cycles per day for this 2-hourly data) to ignore contamination from tidal energy occurring at 1 cycle per day or higher, which is not part of our stochastic model. Furthermore, we also only fit to the side of the spectrum dominated by inertial oscillations, as the model is not always seen to be a good fit on the other side of the spectrum. The modelling and inference approach is therefore semi-parametric (Robinson, 1995).

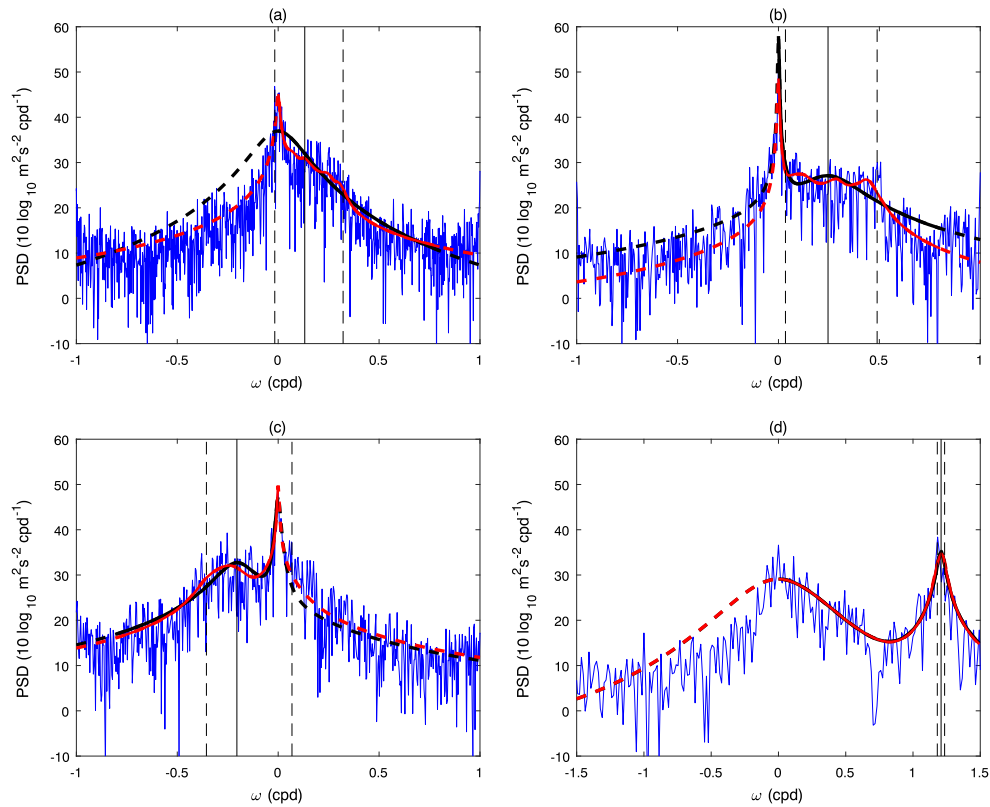


Figure 2. Fitted variance of the discrete Fourier transform using either the stationary model (in black) or the non-stationary model (in red) to the periodogram (in blue) for segments of data from drifter IDs (a) 79243, (b) 54656, (c) 71845 and (d) 44312. The solid black vertical line is the average inertial frequency, and the dashed vertical black lines are the minimum and maximum observed inertial frequency over the observed time window. The models are fit in the frequency range of 0 to 0.8 cycles per day (a)–(c), and from 0 to 1.5 cycles per day (d) as this drifter is at a higher latitude of 37°S where inertial oscillations occur at a frequency of about 1.2 cycles per day. The fitted models are shown in solid lines within the frequency range, and in dashed lines outside the frequency range

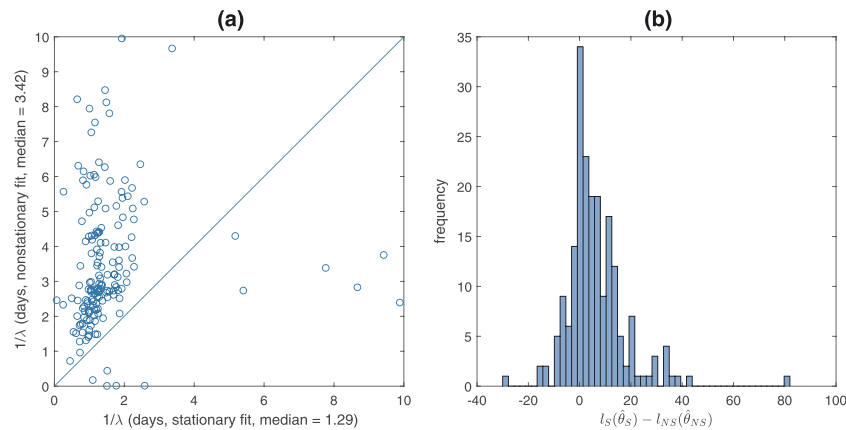


Figure 3. (a) A scatter plot of the damping timescale $1/\lambda$ as estimated by the stationary and non-stationary models, for each of the 200 trajectories displayed in Figure 1. (b) A histogram of the difference between the log-likelihoods of the non-stationary and stationary models for the same 200 trajectories

The significance of the misfit of the stationary model is that parameters of the model may be underestimated or overestimated as the model attempts to compensate for the misfit. For example, the damping parameter of the inertial oscillations, λ , will likely be overestimated in the stationary model, as it is used to try to capture the spread of energy around $\omega^{\{f\}}$, which is in fact mostly caused by the changing value of $\omega^{\{f\}}$, rather than a true high value of λ .

To investigate this further, we perform the analysis with all 200 drifters shown in Figure 1. In Figure 3(a), we show a scatter plot of the estimates of $1/\lambda$, known as the damping timescale, as estimated by both models. In general, the damping timescales are larger with the non-stationary model (consistent with a smaller λ), where the median value is 3.42 days, rather than 1.3 days with the stationary model. Previous estimates of the damping timescale in the literature have not included data from the equatorial region, so while direct comparisons are not possible, the former estimates are found to be more consistent with previous estimates at higher latitudes where values of around 3 days are reported in Elipot *et al.* (2010), and values ranging from 2 to 10 days are reported in Watanabe and Hibiya (2002).

The non-stationary model does not require more parameters to be fitted than the stationary model; both have five unknown parameters. Therefore, there is no need to penalize the non-stationary model using model choice or likelihood ratio tests. Even though the models are not nested, comparing the likelihood of the two approaches can be informative. We can directly compare the negative pseudo-likelihood value of each model using (26) with the stationary model (denoted $\ell_S(\hat{\theta}_S)$) and the nonstationary model (denoted $\ell_{NS}(\hat{\theta}_{NS})$). A histogram of the difference between the likelihoods for the 200 drifters is shown in Figure 3(b), where positive values indicate that the likelihood of the non-stationary model is higher. Overall, the non-stationary model has a smaller negative likelihood in 146 out of the 200 trajectories and is therefore seen to be the better model in general.

There are other regions of the global oceans, in addition to the equator, where the non-stationary methods of this article may significantly improve parameter estimates of drifter time series. These include drifters that follow currents that traverse across different latitudes, such as the Gulf Stream or the Kuroshio. Analysis of such data is an important avenue of future investigation.

4.1.5. Testing with Numerical Model Output

We test the accuracy of the non-stationary modelling and parameter estimation for drifters by analysing output in a controlled setting using a dynamical model for inertial oscillations. The model propagates particles on an ocean surface forced by winds – simulated white noise in our simulations – with a fixed damping parameter, similarly to the damped-slab model of Pollard and Millard (1970), but uses the correct spherical dynamics for the Earth from Early (2012) so that the oscillations occur at the correct Coriolis frequency given the particle's latitude and the model remains valid at the equator. The damping timescale parameter is fixed globally *a priori* in the model, and the goal is to see if it can be accurately estimated using parametric time series models.

The numerical model is constructed such that the particle can also be given a linear mean flow, $U + iV$. If this mean flow has a significant vertical component V , then the particle will cross different latitudes and the frequency of inertial oscillations will significantly change over a single analysis window. We display particle trajectories from the dynamical model in Figure 4, with various realistic mean flow values, where the spherical dynamics can clearly be seen for larger latitudinal mean flow values. We observe that the particles subject to small mean flows display stationary oscillation patterns, whereas for the particles with a large latitudinal mean flow, the oscillation frequency appears to diminish as the particle approaches latitude zero. A more complete description of the numerical model is available in the online code.

To explore the performance of the estimation of damping timescales, we assess the performance of the parameter estimates of our non-stationary model, by performing a Monte Carlo study based on the dynamical model described in the previous paragraph. We generate 100 trajectories, each of length 60 days and sampled every 2 hours, for a given damping timescale ($1/\lambda$) and latitudinal mean flow (V). We estimate the damping parameter using the stationary and non-stationary methods, in exactly the same way as with the real-world drifter data, and average the estimated damping timescales $1/\lambda$ over the 100 time series. We note that as this model has no background turbulence, we set $B = 0$ in (29) such that there is no Matérn component present. We then repeat this

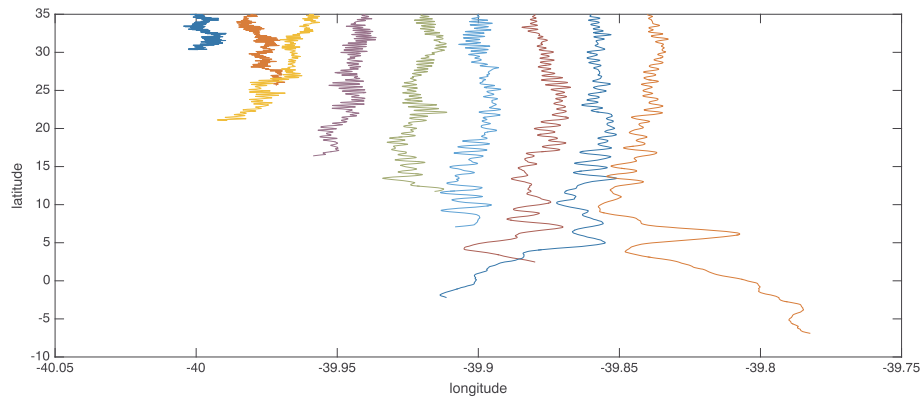


Figure 4. Trajectories of nine particles from the dynamical model, with the damping timescale set to 4 days. All particle trajectories are started at 35°N and 40°W with increasing meridional mean flow from $V = 0.1$ to $V = 0.9$ m/s going from left to right (U is set to zero for this example). The drifters are offset in longitude by 0.02° for representation

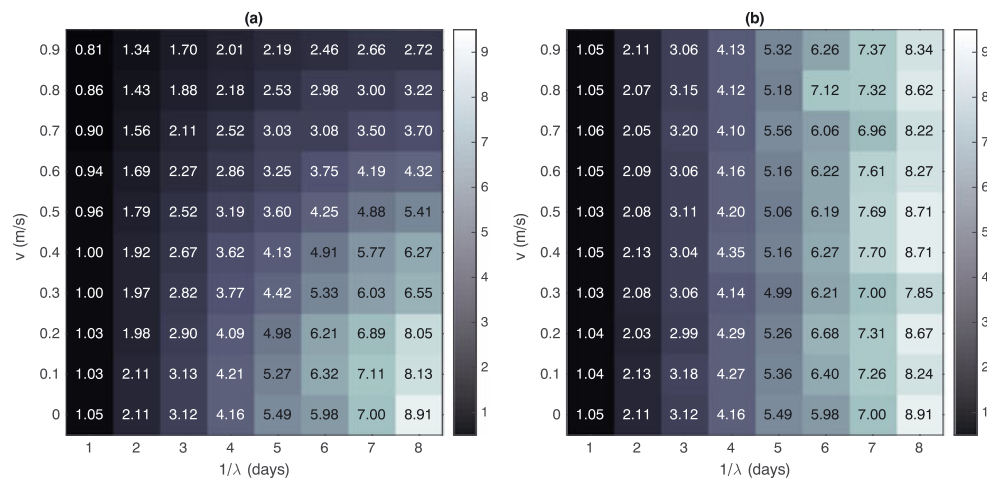


Figure 5. Mean estimates of the damping timescale $1/\lambda$ with (a) the stationary model of (29) and (b) the non-stationary model of Section 4.1.3, applied to 100 realizations of the dynamical model described in Section 4.1.5. The experiment is performed over a grid of latitudinal mean flow values V from 0 to 0.9 cm/s, and over a range of true damping timescales $1/\lambda$ from 1 to 8 days. The estimated damping timescale values, averaged over 100 repeat experiments, is written in each cell and shaded according to the color bar

analysis over a range of realistic values for $1/\lambda$ and V . The average estimates of $1/\lambda$ are reported in Figure 5. The stationary method breaks down for large mean flows and long damping timescales, with large overestimates of λ , confirming our observations from the real-data analysis. The non-stationary method performs well across the entire range of values. We note that long damping timescales are generally harder to estimate, as λ becomes close to zero and is estimated over relatively fewer frequencies. We have not reported mean-squared errors (MSEs) here for space considerations, but we found the parameter biases to be the main contribution to the errors, so it follows that the non-stationary method remains strongly preferable.

4.1.6. Testing with Stochastic Model Output

We test with purely stochastic output, which allows us to extensively compare biases, errors and computational times of the stationary and non-stationary methods in a much larger Monte Carlo study. We continue using the

bivariate model of (22), which is suitable for inertial oscillations, except this time we change β_t according to a stochastic process. Specifically, we set as our generative mechanism for the frequencies β_t ,

$$\beta_0 = \mathcal{D}(\gamma + A\epsilon_t), \quad (37)$$

$$\beta_t = \mathcal{D}(\beta_{t-1} + A\epsilon_t), \quad (38)$$

where $\gamma \in [-\pi, \pi)$, $A > 0$, ϵ_t is a standard normal white noise, and $\mathcal{D}(\cdot)$ is the bounding function defined by

$$\mathcal{D}(x) = \max\{\min(x, \gamma + \Delta), \gamma - \Delta\}, \quad (39)$$

where $\Delta > 0$, and this choice of $\mathcal{D}(x)$ constrains β_t in the interval $[\gamma - \Delta, \gamma + \Delta]$. This way the frequencies β_t are generated according to a bounded random walk, that is, a random walk, which is constrained to stay within a fixed bounded interval. According to Proposition 6, if Δ is smaller than $\pi/2$, then this ensures that the modulated process belongs to the class of modulated processes with a significant correlation contribution, and our estimator (28) is consistent.

In our simulations, we have set $\gamma = \pi/2$, $\Delta = 1$, $A = 1/20$. We simulate for a range of sample sizes ranging from $N = 128$ to $N = 4096$. For each sample size N , we independently simulate 2000 time series and estimate $\{r, \sigma\}$ for each series to report ensemble-averaged biases, errors and computational times. The results are reported in Table I. The bias and MSE of the estimated parameters with the stationary method are seen to increase with increasing sample size. This is because the random walk of β_t increases the range of β_t with larger N , such that the non-stationarity of the time series is increasing. Conversely, the non-stationary method accounts for these rapidly changing modulating frequencies, and the bias and MSE of parameter estimates rapidly decrease with increasing N . The average CPU time is only around 5% slower using the non-stationary method, as the method is still $\mathcal{O}(N \log N)$ in computational efficiency.

Finally, we consider the case in which the modulating sequence is only unknown up to a functional form, and we must also estimate its parameters, along with the parameters of the latent process. We consider the following parametric form for β_t

$$\beta_t = \gamma + \Delta \frac{2t - (N - 1)}{2(N - 1)}, \quad (40)$$

with parameters $\gamma \in [-\pi, \pi)$ and $0 < \Delta < \pi$. The upper bound for Δ is chosen so that the resulting modulated process satisfies the assumptions of Proposition 6. The modulated process then has a significant correlation contribution. Therefore, β_t varies linearly from $\gamma - \frac{\Delta}{2}$ to $\gamma + \frac{\Delta}{2}$. We can then show that for all integer value τ ,

$$c_g^{(N)}(\tau) = \frac{\sin\left[\frac{\Delta\tau}{2(N-1)}(N-\tau)\right]}{N \sin\left[\frac{\Delta\tau}{2(N-1)}\right]} e^{i\left\{\gamma\tau + \frac{\Delta\tau}{2(N-1)}\right\}}. \quad (41)$$

This allows the kernel in (3) to be precomputed in $\mathcal{O}(N)$ elementary operations for all values of $\tau = 0, \dots, N-1$. This helps to speed up the computation of the expected periodogram in the likelihood for the special case of a linearly varying β_t . In this problem, we have to estimate $\{\gamma, \Delta\}$ from β_t as well as $\{r, \sigma\}$ from Z_t . We perform a Monte Carlo simulation with a fixed sample size of $N = 512$, where we simulate 5000 independent time series each with parameters set to $r = 0.9$, $\sigma = 10$, $\gamma = 0.8$ and $\Delta = 1$. We report the biases, variances and MSEs with the stationary and non-stationary methods in Table II. As the stochastic process is Markovian, it is also possible to implement exact maximum likelihood in $\mathcal{O}(N)$ elementary operations for this specific problem, and we report these values in the table also. Our non-stationary inference method performs relatively close to that of exact

Table I. Performance of estimators with the stationary and non-stationary methods for the model of (22) with β_t evolving according to the bounded random walk described by (37)–(39)

	Sample size (N)					
	128	256	512	1024	2048	4096
Stationary frequency-domain likelihood						
Bias (r)	−2.35e−02	−3.24e−02	−4.81e−02	−6.98e−02	−9.33e−02	−1.12e−01
Variance (r)	1.82e−03	1.08e−03	1.14e−03	1.56e−03	1.40e−03	8.29e−04
MSE (r)	2.37e−03	2.13e−03	3.46e−03	6.42e−03	1.01e−02	1.33e−02
Bias (σ)	2.56e−02	5.50e−02	8.95e−02	1.32e−01	1.74e−01	2.07e−01
Variance (σ)	3.39e−03	2.82e−03	3.35e−03	4.47e−03	3.99e−03	2.16e−03
MSE (σ)	4.04e−03	5.84e−03	1.14e−02	2.20e−02	3.44e−02	4.48e−02
Average CPU time (seconds)	.65e−03	1.27e−02	1.83e−02	3.08e−02	2.86e−02	4.81e−02
Non-stationary frequency-domain likelihood						
Bias (r)	−4.62e−03	−2.01e−03	−1.42e−03	−2.90e−04	−2.70e−04	8.83e−05
Variance (r)	1.65e−03	7.54e−04	3.98e−04	2.07e−04	1.07e−04	5.32e−05
MSE (r)	1.67e−03	7.58e−04	4.00e−04	2.07e−04	1.07e−04	5.32e−05
Bias (σ)	−1.50e−02	−8.86e−03	−4.43e−03	−2.53e−03	−1.41e−03	−9.17e−04
Variance (σ)	2.25e−03	1.20e−03	6.42e−04	3.48e−04	2.01e−04	1.08e−04
MSE (σ)	2.48e−03	1.28e−03	6.62e−04	3.54e−04	2.03e−04	1.08e−04
Average CPU time (seconds)	1.11e−02	1.35e−02	1.90e−02	3.02e−02	2.82e−02	4.75e−02

The parameters are set as $r = 0.8$, $\sigma = 1$, $\gamma = \pi/2$, $\Delta = 1$ and $A = 1/20$. The results are averaged over 2000 independently generated time series for each sample size N . The average CPU times for the optimization are given in seconds, as performed on a 3.60-GHz Intel i7-4790MQ processor (four cores).
MSE, mean-squared error.

Table II. Performance of estimators with the stationary and non-stationary methods for the model of (22) with β_t evolving according to (40)

	Estimated parameter			
	r	σ	γ	Δ
Exact likelihood				
Bias	−1.32e−03	2.11e−02	2.12e−03	−3.61e−03
Variance	1.87e−04	5.21e−02	2.37e−04	2.83e−03
MSE	1.88e−04	5.26e−02	2.42e−04	2.85e−03
Stationary frequency-domain likelihood				
Bias	−1.54e−01	5.21e+00	2.59e−03	N/A
Variance	5.51e−04	8.69e−01	9.46e−03	N/A
MSE	2.42e−02	2.80e+01	9.47e−03	N/A
Non-stationary frequency-domain likelihood				
Bias	−1.71e−03	6.82e−03	1.14e−03	−3.71e−02
Variance	2.40e−04	1.53e−01	2.08e−03	1.64e−02
MSE	2.43e−04	1.53e−01	2.08e−03	1.78e−02

The parameters are set as $r = 0.9$, $\sigma = 10$, $\gamma = 0.8$ and $\Delta = 1$. The results are averaged over 5000 independently generated time series.
N/A, not applicable; MSE, mean-squared error.

maximum likelihood, despite the challenge of having to estimate parameters of the modulating sequence, as well as the latent process. The stationary method performs poorly, as with previous examples, as stationary modelling is not appropriate for such rapidly varying oscillatory structure.

Table III. Performance of our estimator for the missing data problem defined in (42)–(44)

	Sample size						
	128	512	1024	2048	4096	8192	16,384
Estimate of parameter a							
Bias	−2.08e−02	−4.81e−03	−3.09e−04	−4.17e−04	−2.73e−04	−4.94e−04	−1.52e−04
Variance	1.07e−02	2.71e−03	1.28e−03	6.31e−04	3.09e−04	1.44e−04	7.10e−05
MSE	1.12e−02	2.73e−03	1.28e−03	6.31e−04	3.09e−04	1.44e−04	7.10e−05
Estimate of parameter σ							
Bias	−1.71e−02	−5.49e−03	−6.89e−03	−2.79e−03	−1.31e−03	2.29e−04	−1.74e−04
Variance	3.37e−02	8.76e−03	4.17e−03	1.94e−03	9.80e−04	4.15e−04	2.27e−04
MSE	3.40e−02	8.79e−03	4.22e−03	1.95e−03	9.81e−04	4.15e−04	2.27e−04
Computational time							
Average CPU time (s)	7.82e−03	1.63e−02	2.77e−02	2.40e−02	3.99e−02	6.81e−02	1.22e−01

The unknown parameters are set as $a = 0.8$ and $\sigma = 1$. The results are averaged over 2000 independently generated time series for each sample size N . The average CPU times for the optimization are given in seconds, as performed on a 3.60-GHz Intel i7-4790 processor (four cores). MSE, mean-squared error.

4.2. Missing Data Simulation

We show that the estimator defined in Definition 4 can be used for the random missing data scheme 3 of Section 2.2. Therefore, we simulate a real-valued first-order autoregressive process with parameters $0 \leq a < 1$ and σ according to

$$X_t = aX_{t-1} + \epsilon_t, \quad t \geq 1, \quad (42)$$

where $X_0 \sim \mathcal{N}[0, \sigma^2/(1 - a^2)]$, and ϵ_t is a Gaussian white noise process with mean zero and variance σ^2 . The process $\{X_t\}$ is the latent process of interest. To account for the missing data, we generate a modulated time series $\tilde{X}_t = g_t X_t$ and assume we only observe the time series $\{\tilde{X}_t\}$, from which we estimate the parameters of the process $\{X_t\}$. The sequence $\{g_t\}$ takes its values in the set $\{0, 1\}$ and is generated according to

$$g_t \sim \mathcal{B}(p_t), \quad (43)$$

where $\mathcal{B}(p)$ represents the Bernoulli distribution with parameter p , and where we set

$$p_t = \frac{1}{2} + \frac{1}{4} \cos\left(\frac{2\pi}{10}t\right). \quad (44)$$

The observed modulating sequence $\{g_t\}$, made of zeros and ones, is clearly non-stationary as it does not admit a constant expectation. Therefore, a spectral representation of the second-order structure of the random modulating sequence $\{g_t\}$, as required in Dunsmuir and Robinson (1981c), does not exist. We simulate and estimate such a model for different sample sizes ranging from $N = 128$ to $N = 16,384$. For each value of N , we independently simulate 2000 time series, and for each time series, we estimate $\{a, \sigma\}$. The outcomes of our simulation study are reported in Table III. The bias, variance and MSE rapidly decrease with increasing N , while the computational time only increases gradually with N such that the methods are still computationally efficient for long time series. Comparing our technique with other methods from the literature is the subject of ongoing work.

5. CONSISTENCY

We show in Theorem 1 that the frequency-domain estimator $\hat{\theta}_M^{(N)}$ (which for simplicity we denote as $\hat{\theta}^{(N)}$ in this section) is consistent in the univariate real-valued case (extension to our class of bivariate processes follows

directly). In Theorem 2, we show that this estimator converges with a $\mathcal{O}(N^{-\frac{1}{2}})$ rate. To guarantee consistency, we require the following assumptions to be satisfied:

1. The parameter set $\Theta \subset \mathbb{R}^d$ is compact with a non-null interior, and the true parameter θ lies in the interior of Θ .
2. Assume that for all $\theta \in \Theta$, we have $\sum_{\tau \in \mathbb{N}} |c_X(\tau; \theta)| < \infty$ (short memory) and that the functions $\theta \rightarrow c_X(\tau; \theta)$ are continuous with respect to θ . It follows that the spectral densities are also continuous with respect to θ . We also assume that for all $\theta \in \Theta$ and $\omega \in [-\pi, \pi]$, $S_X(\omega; \theta) > 0$. By continuity on a compact set, the spectral densities $S_X(\omega; \theta)$ are therefore bounded below in both variables by a non-zero value. For the same reason, they are bounded above.
3. We assume that the spectral densities are continuously differentiable with respect to both variables ω and θ . By continuity on a compact set, the derivatives with respect to ω are bounded above, independently of θ .
4. The process $\tilde{X}_t$ is a modulated process with significant correlation contribution. We recall that this implies the existence of a finite subset $\Gamma \subset \mathbb{N}$ such that the mapping $\theta \mapsto \{c_X(\tau) : \tau \in \Gamma\}$ is one-to-one. We also assume that the modulating sequence $\{g_t\}$ is bounded above in absolute value by some finite constant $g_{\max} > 0$.

We start with the following two lemmas, which yield uniform bounds of the expected periodogram and its derivative.

Lemma 1 (Boundedness of the expected periodogram). For all $\theta \in \Theta$ and $N \in \mathbb{N}$, the expected periodogram $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)$ is bounded below (by a positive real number) and above independently of N and θ . We denote these bounds as $\bar{S}_{\tilde{X}, \min}$ and $\bar{S}_{\tilde{X}, \max}$ respectively.

Proof

See Appendix A.4. □

Lemma 2 (Boundedness of the derivative of the expected periodogram). The derivative of the expected periodogram with respect to ω exists is bounded in absolute value independently of θ and N .

Proof

See Appendix A.5. □

In analogue to Taniguchi (1979) for stationary processes, we introduce the following quantity:

$$D^{(N)}(\boldsymbol{\gamma}, f) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\gamma}) + \frac{f(\omega)}{\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\gamma})} \right\},$$

for all positive integer N , $\boldsymbol{\gamma} \in \Theta$ and non-negative real-valued function f defined on $[-\pi, \pi]$. We also define

$$T^{(N)}(f) = \arg \min_{\boldsymbol{\gamma} \in \Theta} D^{(N)}(\boldsymbol{\gamma}, f).$$

This minimum for fixed f is well defined since the set Θ is compact and since the function $\boldsymbol{\gamma} \mapsto D^{(N)}(\boldsymbol{\gamma}, f)$ is continuous. However, in cases where the minimum is reached not uniquely but at multiple parameter values, $T^{(N)}(f)$ will denote any of these values, chosen arbitrarily. Note that by the definition of our frequency-domain estimator, we have $\hat{\theta}^{(N)} = T^{(N)}(\hat{S}_{\tilde{X}}^{(N)}(\cdot))$. We derive three lemmas that will be required in proving Theorem 1, which establishes consistency.

Lemma 3. We have, for N large enough, $T^{(N)}(\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)) = \theta$, uniquely.

Proof

See Appendix A.6. □

This shows that for all N large enough, the function $\boldsymbol{\gamma} \rightarrow D\left(\boldsymbol{\gamma}, \bar{S}_{\bar{X}}^{(N)}(\cdot; \boldsymbol{\theta})\right)$ reaches a global minimum at the true parameter vector $\boldsymbol{\theta}$. However, because $\bar{S}_{\bar{X}}^{(N)}(\cdot; \boldsymbol{\theta})$ is changing with N and is not expected to converge to a given function, we need the following stronger result.

Lemma 4. If $\{\boldsymbol{\gamma}_N\}_{N \in \mathbb{N}} \in \Theta^{\mathbb{N}}$ is a sequence of parameter vectors such that $D\left(\boldsymbol{\gamma}_N, \bar{S}_{\bar{X}}^{(N)}(\cdot; \boldsymbol{\theta})\right) - D\left(\boldsymbol{\theta}, \bar{S}_{\bar{X}}^{(N)}(\cdot; \boldsymbol{\theta})\right)$ converges to zero when N goes to infinity, then $\boldsymbol{\gamma}_N$ converges to $\boldsymbol{\theta}$.

Proof

See Appendix A.7. □

We now show that the functions $D\left(\boldsymbol{\gamma}, \bar{S}_{\bar{X}}^{(N)}(\cdot; \boldsymbol{\theta})\right)$ and $D\left(\boldsymbol{\gamma}, \hat{S}_{\bar{X}}^{(N)}(\cdot)\right)$, defined on Θ , behave asymptotically in the same way. For this, we first need the following lemma where we bound the asymptotic variance of some linear functionals of the periodogram.

Lemma 5. Let $\{a^{(N)}(\omega) : \omega \in [-\pi, \pi)\}_{N \in \mathbb{N}}$ be a family of real-valued functions, uniformly bounded by a positive real number. We have

$$\text{var} \left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a^{(N)}(\omega) \hat{S}_{\bar{X}}^{(N)}(\omega) \right\} = \mathcal{O} \left(\frac{1}{N} \right).$$

Proof

See Appendix A.8. □

Remembering that $\bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\theta}) = \mathbb{E} \left\{ \hat{S}_{\bar{X}}^{(N)}(\omega); \boldsymbol{\theta} \right\}$, we thus have

$$\sum_{\omega \in \Omega_N} a^{(N)}(\omega) \hat{S}_{\bar{X}}^{(N)}(\omega) = \sum_{\omega \in \Omega_N} a^{(N)}(\omega) \bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\theta}) + \mathcal{O}_P \left(\frac{1}{\sqrt{N}} \right).$$

We are now able to state a consistency theorem for our estimator $\hat{\boldsymbol{\theta}}^{(N)}$.

Theorem 1 (Consistency of the frequency-domain estimator). We have $\hat{\boldsymbol{\theta}}^{(N)} \xrightarrow{P} \boldsymbol{\theta}$ in probability.

Proof

The proof is based on Taniguchi (1979). Denote $\bar{h}^{(N)}(\boldsymbol{\gamma}; \boldsymbol{\theta}) = D\left(\boldsymbol{\gamma}, \bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)$ and $\hat{h}^{(N)}(\boldsymbol{\gamma}) = D\left(\boldsymbol{\gamma}, \hat{S}_{\bar{X}}^{(N)}(\omega)\right)$ defined for any $\boldsymbol{\gamma} \in \Theta$. We have

$$\begin{aligned} \bar{h}^{(N)}(\boldsymbol{\gamma}; \boldsymbol{\theta}) - \hat{h}^{(N)}(\boldsymbol{\gamma}) &= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\gamma}) + \frac{\bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\theta})}{\bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\gamma})} - \log \bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\gamma}) - \frac{\hat{S}_{\bar{X}}^{(N)}(\omega)}{\bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\gamma})} \right\} \\ &= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{\bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\theta}) - \hat{S}_{\bar{X}}^{(N)}(\omega)}{\bar{S}_{\bar{X}}^{(N)}(\omega; \boldsymbol{\gamma})}. \end{aligned}$$

We have shown in Lemma 1 that $\bar{S}_{\bar{X}}^{(N)}(\omega; \gamma)$ is bounded below in both variables ω and γ by a positive real number, independently of N and γ . Therefore, making use of Lemma 5, we have

$$\sup_{\gamma \in \Omega} \left| \bar{h}^{(N)}(\gamma; \theta) - \hat{h}^{(N)}(\gamma) \right| \xrightarrow{P} 0, \quad (N \rightarrow \infty), \quad (45)$$

where the letter P indicates that the convergence is in probability, as the difference is of stochastic order $N^{-\frac{1}{2}}$. In particular, (45) implies that

$$\left| \min_{\gamma} \bar{h}^{(N)}(\gamma; \theta) - \min_{\gamma} \hat{h}^{(N)}(\gamma) \right| \leq \sup_{\gamma \in \Omega} \left| \bar{h}^{(N)}(\gamma; \theta) - \hat{h}^{(N)}(\gamma) \right| \xrightarrow{P} 0,$$

that is,

$$\left| \bar{h}^{(N)}\left(T^{(N)}(\bar{S}_{\bar{X}}^{(N)}(\omega; \theta)); \theta\right) - \hat{h}^{(N)}\left(T^{(N)}(\hat{S}_{\bar{X}}^{(N)}(\omega)); \theta\right) \right| \xrightarrow{P} 0. \quad (46)$$

Relation (45) also implies that

$$\left| \bar{h}^{(N)}\left(T^{(N)}(\hat{S}_{\bar{X}}^{(N)}(\omega)); \theta\right) - \hat{h}^{(N)}\left(T^{(N)}(\hat{S}_{\bar{X}}^{(N)}(\omega)); \theta\right) \right| \xrightarrow{P} 0, \quad (47)$$

so that using the triangle inequality, (46) and (47), we obtain

$$\left| \bar{h}^{(N)}\left(T^{(N)}(\hat{S}_{\bar{X}}^{(N)}(\omega)); \theta\right) - \bar{h}^{(N)}\left(T^{(N)}(\bar{S}_{\bar{X}}^{(N)}(\omega; \theta)); \theta\right) \right| \xrightarrow{P} 0.$$

We then obtain the stated theorem making use of Lemma 4. $\square$

We now study the convergence rate of our frequency-domain estimator. For this, we first need the following two lemmas. Although the Hessian matrix of the likelihood is not expected to converge for modulated processes with a significant correlation contribution, we can show that its norm is bounded below by a positive real number. For this, we need to strengthen the assumption of significant correlation contribution. Assuming that the spectral densities of the latent process are twice continuously differentiable with respect to θ , we assume that the Jacobian determinant of the mapping $\theta \mapsto [c_X(\tau; \theta) : \tau \in \Gamma]^T$ taken at the true parameter value θ , that is, the determinant of the matrix with elements $\frac{\partial c_X(\tau_i; \theta)}{\partial \theta_j}$ (with $\Gamma = \{\tau_1, \tau_2, \dots, \tau_d\}$ here), is non-zero.

Lemma 6. Let $\mathbf{U}_1, \dots, \mathbf{U}_d$ a family of vectors of $\mathbb{R}^d$ with rank d . Let $\alpha_1, \dots, \alpha_d$ be positive real numbers. There exists a positive constant $C > 0$ such that for all $\mathbf{V} \in \mathbb{R}^d$,

$$\sum_{i=1}^d \alpha_i^2 (\mathbf{U}_i^T \mathbf{V})^2 \geq C \|\mathbf{V}\|_2^2, \quad (48)$$

where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^N$.

Proof

See Appendix A.9. $\square$

Lemma 7. We have

$$\frac{\partial l_M^{(N)}}{\partial \theta_i}(\theta) = \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right). \quad (49)$$

The Hessian matrix of the function $l_M(\boldsymbol{\theta})$ satisfies

$$H(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta}) + \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right), \quad (50)$$

where the matrix norm of $\mathcal{I}(\boldsymbol{\theta})$ is bounded below by a positive value, independently of N .

Proof

See Appendix A.10. □

Theorem 2 (Convergence rate). We have $\hat{\boldsymbol{\theta}}^{(N)} = \boldsymbol{\theta} + \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right)$.

Proof

We have, by Taylor expansion with Lagrange form of the remainder term,

$$\nabla l_M(\hat{\boldsymbol{\theta}}^{(N)}) = \mathbf{0} = \nabla l_M(\boldsymbol{\theta}) + H(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}),$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}^{(N)}$ and $\boldsymbol{\theta}$. Therefore,

$$\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta} = -H(\tilde{\boldsymbol{\theta}})^{-1} \nabla l_M(\boldsymbol{\theta}). \quad (51)$$

We have shown that $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}$. By continuity of the Hessian, and using the results of Lemma 7, we obtain

$$\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta} = -\left[\mathcal{I}(\boldsymbol{\theta}) + \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right) + o_P(1)\right]^{-1} \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right) = \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right). \quad (52)$$

This concludes the proof. □

6. CONCLUSION

The well-established theory for the analysis of stationary time series is often in contradiction with real-world data applications. This is because most real time series are non-stationary. Non-stationary observations have required statisticians to develop new models and more broadly new generating mechanisms. Among the large class of non-stationary models, uniform modulation of time series is an easy way to create non-stationarity and presents all the advantages of a simple mechanism for the time-varying second-order structure of a process. Modulation has already been used to account for missing data when analysing stationary time series, as well as gentle time variation. In fact, if the modulation is slow, regular theory for locally stationary time series applies. Despite its popularity as a modelling tool, the concept of modulation, when variation can be moderate to rapid, is very poorly understood. We have in this article shown how modulation of time series can account for much more rapid changes of a time series model. Guinness and Stein (2013) already abandoned the assumption of smoothness in time of the time-varying spectral density (Dahlhaus, 1997). However, the class of modulated processes with a significant correlation contribution is one of few instances of non-stationary models where more data in time results in more accurate estimates, and asymptotic consistency under standard assumptions for the latent stationary process.

As we have generalized modulation beyond the assumptions where it is known that models can be estimated, the question naturally arises, as to what types of modulation still permit parameter estimation. Key to our understanding of modulated processes is the definition of modulated processes with a significant correlation contribution

(Definition 3), which generalizes the classical concept of asymptotically stationary modulated processes and corresponds to our main modelling innovation. We require that the sample autocorrelations of the modulating sequences be asymptotically bounded below, so that the information in the autocorrelation of the process does not fade in the observed process. The interpretation of this requirement is that there must be sufficient support in the autocorrelation to retain the information in the modulation.

With this new model class, we can implement estimation directly in the Fourier domain, directly after transforming the data from the temporal domain. Estimation is still possible in $\mathcal{O}(N \log N)$ computational effort, and the further required conditions to ensure consistency were studied. Most real-world data sets are aggregations of heterogeneous components. To fully show the promise of our newly proposed procedure, we show how estimation is still possible in the setting of unobserved components models, where different types of processes are superimposed. Real-world data from the GDP show its relevance for understanding surface flow measurements at the equator – a challenging region for studying inertial oscillations where the power of the new method shows that despite rapid modulation we can still uncover the generating mechanism of the process.

There are a number of questions still remaining in our understanding of modulation. We have extended the regimes when estimation is possible but do not know when an estimable process tips into one from which no information can be recovered. By introducing a new class of models, many new questions can both be posited and answered, especially as most sources of real-world data show aggregations of components, all obeying different generation mechanisms.

ACKNOWLEDGEMENTS

The work of Arthur Guillaumin was supported by the Statistical Department of University College London. The work of A. M. Sykulski and S. C. Olhede was supported by the UK Engineering and Physical Sciences Research Council via EP/I005250/1 and EP/L025744/1. A. M. Sykulski and S. C. Olhede also acknowledge support from the 7th European Community Framework Programme via a Marie Curie International Outgoing Fellowship (Sykulski), and via a European Research Council Fellowship via grant CoG 2015- 682172NETS (Olhede). The work of J. M. Lilly and J. J. Early was supported by the Physical Oceanography Program of the United States National Science Foundation under award no. 1235310. The authors thank the anonymous reviewers for their careful reading and their suggested improvements that significantly enhanced the quality of the article.

REFERENCES

- Abascal AJ, Castaneda S, Medina R, Liste M. 2010. Analysis of the reliability of a statistical oil spill response model. *Marine Pollution Bulletin* **60**(11): 2099–2110.
- Adak S. 1998. Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association* **93**(444): 1488–1501.
- Allefeld C, Atmanspacher H, Wackermann J. 2009. Mental states as macrostates emerging from brain electrical dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **19**: 015102.
- Allen MR, Robertson AW. 1996. Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dynamics* **12**(11): 775–784.
- Andrews T, Gregory JM, Webb MJ, Taylor KE. 2012. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters* **39**(9): L09712.
- Arató M, Baran S, Ispány M. 1999. Functionals of complex Ornstein–Uhlenbeck processes. *Computers & Mathematics with Applications* **37**(1): 1–13.
- Arató M, Kolmogorov AN, Sinai YG. 1962. Evaluation of the parameters of a complex stationary Gauss–Markov process. *Doklady Akademii Nauk SSSR* **146**: 747–750.
- Bloomfield P. 1970. Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society. Series B (Methodological)* **32**(3): 369–380.
- Bloomfield P. 2000. *Fourier Analysis of Time Series: An Introduction*. New York, USA: Wiley Series in Probability and Statistics.
- Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd edn. Berlin, Germany: Springer.
- Broersen PMT, De Waele S, Bos R. 2004. Autoregressive spectral analysis when observations are missing. *Automatica* **40**(9): 1495–1504.

- Clinger W, Ness JWV. 1976. On unequally spaced time points in time series. *The Annals of Statistics* **4**(4): 736–745.
- Dahlhaus R. 1997. Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**(1): 1–37.
- Dunsmuir W, Robinson PM. 1981. Asymptotic theory for time series containing missing and amplitude modulated observations. *Sankhyā: The Indian Journal of Statistics, Series A* **43**(3): 260–281.
- Dunsmuir W, Robinson PM. 1981. Estimation of time series models in the presence of missing data. *Journal of the American Statistical Association* **76**(375): 560–568.
- Dunsmuir W, Robinson PM. 1981. Parametric estimators for stationary time series with missing observations. *Advances in Applied Probability* **13**(1): 129–146.
- Dzhaparidze KO, Yaglom AM. 1983. *Spectrum parameter estimation in time series analysis*. In *Developments in Statistics*, Krishnaiah PR (ed.), chapter 1, vol. 4 New York: Academic Press, pp. 1–96.
- Early JJ. 2012. The forces of inertial oscillations. *Quarterly Journal of the Royal Meteorological Society* **138**(668): 1914–1922.
- Elipot S, Lumpkin R, Perez RC, Lilly JM, Early JJ, Sykulski AM. 2016. A global surface drifter data set at hourly resolution. *Journal of Geophysical Research: Oceans* **121**(5): 2937–2966.
- Elipot S, Lumpkin R, Prieto G. 2010. Modification of inertial oscillations by the mesoscale eddy field. *Journal of Geophysical Research: Oceans* **115**: C09010. <https://doi.org/10.1029/2009JC005679>.
- Ferrari R, Wunsch C. 2009. Ocean circulation kinetic energy: reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics* **41**(1): 253–282.
- Gneiting T, Kleiber W, Schlather M. 2010. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**(491): 1167–1177.
- Griffa A, Kirwan AD, Mariano AJ, Özgökmen T, Rossby T. 2007. *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*. Cambridge, UK: Cambridge University Press.
- Guinness J, Stein ML. 2013. Transformation to approximate independence for locally stationary Gaussian processes. *Journal of Time Series Analysis* **34**(5): 574–590.
- Horn RA, Johnson CR. 1985. *Matrix Analysis*. Cambridge, UK: Cambridge University Press.
- Jiang J, Hui YV. 2004. Spectral density estimation with amplitude modulation and outlier detection. *Annals of the Institute of Statistical Mathematics* **56**(4): 611–630.
- Jones RH. 1962. Spectral analysis with regularly missed observations. *The Annals of Mathematical Statistics* **33**(2): 455–461.
- Jones RH. 1980. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**(3): 389–395.
- LaCasce JH. 2008. Statistics from Lagrangian observations. *Progress in Oceanography* **77**(1): 1–29.
- Le Breton A. 1988. A note on maximum likelihood estimation for the complex-valued first-order autoregressive process. *Statistics & Probability Letters* **7**(2): 171–173.
- Lilly JM. 2016. *jLab: a data analysis package for Matlab*, v. 1.6.2. <http://www.jmlilly.net/jmlsoft.html>.
- Lilly JM, Olhede SC. 2012. Analysis of modulated multivariate oscillations. *IEEE Transactions on Signal Processing* **60**(2): 600–612.
- Lilly JM, Sykulski AM, Early JJ, Olhede SC. 2016. Fractional Brownian motion, the Matérn process, and stochastic modelling of turbulent dispersion. *arXiv preprint: 1605.01684*.
- Lumpkin R, Pazos M. 2007. *Measuring surface currents with Surface Velocity Program drifters: the instrument, its data, and some results*. In *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, Griffa A, Kirwan AD, Mariano AJ, Özgökmen T, Rossby T (eds.) Cambridge University Press: New York, USA, pp. 39–67.
- Miller KS. 1969. Complex Gaussian processes. *Siam Review* **11**(4): 544–567.
- Papadimitis E, Politis DN. 2012. Nonlinear spectral density estimation: thresholding the correlogram. *Journal of Time Series Analysis* **33**(March): 386–397.
- Parzen E. 1961. Spectral analysis of asymptotically stationary time series, *DTIC Document*, Stanford, CA.
- Parzen E. 1963. On spectral analysis with missing observations and amplitude modulation. *Sankhyā: The Indian Journal of Statistics, Series A* **25**(4): 383–392.
- Percival DB, Walden AT. 1993. *Spectral Analysis for Physical Applications*. Cambridge, UK: Cambridge University Press.
- Pollard RT, Millard RC. 1970. Comparison between observed and simulated wind-generated inertial oscillations. *Deep Sea Research and Oceanographic Abstracts* **17**(4): 813–821.
- Priestley MB. 1965. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Series B (Methodological)* **27**(2): 204–237.
- Priestley MB. 1988. *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.
- Priestley MB, Tong H. 1973. On the analysis of bivariate non-stationary processes. *Journal of the Royal Statistical Society, Series B (Methodological)* **35**(2): 153–166.
- Robinson PM. 1995. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23**(5): 1630–1661.
- Rüinstler G. 2004. Modelling phase shifts among stochastic cycles. *The Econometrics Journal* **7**: 232–248.
- Scheinok PA. 1965. Spectral analysis with randomly missed observations: the binomial case. *The Annals of Mathematical Statistics* **36**(3): 971–977.

- Schreiber PJ, Scharf LL. 2010. *Statistical Signal Processing of Complex-valued Data*. Cambridge, UK: Cambridge University Press.
- Sykulski AM, Olhede SC, Lilly JM. 2016a. A widely linear complex autoregressive process of order one. *IEEE Transactions on Signal Processing* **64**(23): 6200–6210.
- Sykulski AM, Olhede SC, Lilly JM. 2016b. The de-biased Whittle likelihood for second-order stationary stochastic processes. *arXiv preprint: 1605.06718v1*.
- Sykulski AM, Olhede SC, Lilly JM, Danioux E. 2016c. Lagrangian time series models for ocean surface drifter trajectories. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(1): 29–50.
- Taniguchi Masanobu. 1979. On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability* **16**(3): 575–591.
- Toloi CMC, Morettin PA. 1989. Spectral analysis for amplitude-modulated time series. *Journal of Time Series Analysis* **14**(4): 409–432.
- Tong H. 1973. Some comments on spectral representations of non-stationary stochastic processes. *Journal of Applied Probability* **10**(4): 881–885.
- Tong H. 1974. On time-dependent linear transformations of non-stationary stochastic processes. *Journal of Applied Probability* **11**(1): 53–62.
- Walden AT. 2013. Rotary components, random ellipses and polarization: a statistical perspective. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **371**: 20110554. <https://doi.org/10.1098/rsta.2011.0554>.
- Walker AM. 1993. *Periodogram analysis for complex-valued time series*. In *Developments in Time Series Analysis*, Rao TS (ed.) Chapman and Hall: London, UK, pp. 149–163.
- Watanabe M, Hibiya T. 2002. Global estimates of the wind-induced energy flux to inertial motions in the surface mixed layer. *Geophysical Research Letters* **29**(8): 64–1–64–3.
- Whittle P. 1953. Estimation and information in stationary time series. *Arkiv för Matematik* **2**(5): 423–434.

APPENDIX A

A.1. Proof of Proposition 4

Proof

We distinguish the case where $\{Z_t\}$ is a white noise process (the covariance is zero everywhere except for lag zero) from the case where $\{Z_t\}$ is not a white noise process.

1. Assume $\{Z_t\}$ is a white noise process.

→ Assume $\{\tilde{Z}_t\}$ is a stationary process. Being stationary, it has a constant variance, and therefore, the modulating sequence must have a constant modulus.

← Conversely, if $\{g_t\}$ has a constant modulus, $\{\tilde{Z}_t\}$ is stationary and is a white noise process.

2. Assume $\{Z_t\}$ is not a white noise process. The set $\{\tau \in \mathbb{N}^* : |c_Z(\tau; \theta)| > 0\}$ is therefore not empty, so $\mu = \gcd\{\tau \in \mathbb{N}^* : |c_Z(\tau; \theta)| > 0\}$ is well defined.

→ Assume $\{\tilde{Z}_t\}$ is stationary. Then it must have a constant variance, so there must exist a real number $a \geq 0$ such that $\rho_t = a \forall t \in \mathbb{N}$. Leaving aside the trivial case in which $\{g_t\}$ is zero everywhere, let t_1, t_2 be two natural integers. We have

$$c_{\tilde{Z}}(t_1, t_2; \theta) = g_{t_1}^* g_{t_2} c_Z(t_2 - t_1; \theta) = a^2 e^{i(\phi_{t_2} - \phi_{t_1})} c_Z(t_2 - t_1; \theta).$$

If $c_Z(t_2 - t_1; \theta) \neq 0$, then

$$e^{i(\phi_{t_2} - \phi_{t_1})} = \frac{c_{\tilde{Z}}(t_1, t_2; \theta)}{a^2 c_Z(t_2 - t_1; \theta)},$$

which leads to

$$\phi_{t_2} - \phi_{t_1} = \arg \left\{ \frac{c_{\tilde{Z}}(t_1, t_2; \theta)}{a^2 c_Z(t_2 - t_1; \theta)} \right\} \mod 2\pi,$$

where the equality is true up to a multiple of 2π , which we indicate by the use of the notation $\mod 2\pi$. Since $\{\tilde{Z}_t\}$ is assumed stationary, there exists a function ζ , defined on $\{\tau \in \mathbb{N} : c_Z(\tau; \boldsymbol{\theta}) \neq 0\}$, such that

$$\arg \left\{ \frac{c_{\tilde{Z}}(t_1, t_2; \boldsymbol{\theta})}{a^2 c_Z(t_2 - t_1; \boldsymbol{\theta})} \right\} = \zeta(t_2 - t_1) \mod 2\pi, \quad \forall t_1, t_2 \in \mathbb{N}.$$

Therefore,

$$\phi_{t_2} - \phi_{t_1} = \zeta(t_2 - t_1) \mod 2\pi.$$

Now let $t \in \mathbb{N}$ be any natural integer and write $t = \mu q + r$ where $0 \leq r < \mu$ and $q \in \mathbb{N}$ are uniquely defined as the remainder and quotient of the Euclidean division of t by μ .

$$\phi_t = \sum_{k=0}^{q-1} (\phi_{r+(k+1)\mu} - \phi_{r+k\mu}) + \phi_r = \sum_{k=0}^{q-1} \zeta(\mu) + \phi_r \mod 2\pi = q\zeta(\mu) + \phi_r \mod 2\pi.$$

Letting $\gamma = \zeta(\mu)$, we obtain

$$\phi_t = \gamma \left\lfloor \frac{t}{\mu} \right\rfloor + \phi_{t \mod \mu} \mod 2\pi.$$

← Conversely assume there exists two constants $\gamma \in \mathbb{R}$ and $a \geq 0$ such that for all $t \in \mathbb{N}$,

$$\begin{aligned} \rho_t &= a, \\ \phi_t &= \phi_{t \mod \mu} + \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \mod 2\pi. \end{aligned}$$

Let t, τ be two natural integers. We have

$$c_{\tilde{Z}}(t, t + \tau; \boldsymbol{\theta}) = g_t^* g_{t+\tau} c_Z(\tau; \boldsymbol{\theta}) = a^2 e^{i(\phi_{t+\tau} - \phi_t)} c_Z(\tau; \boldsymbol{\theta}).$$

If $c_Z(\tau; \boldsymbol{\theta}) = 0$, then $c_{\tilde{Z}}(t, t + \tau; \boldsymbol{\theta}) = 0$, which does not depend on t . Otherwise, τ is a multiple of μ by definition of μ . Therefore, there exists an integer q such that $\tau = q\mu$, and $(t + \tau) \mod \mu = t \mod \mu$. Finally,

$$\begin{aligned} \phi_{t+\tau} - \phi_t &= \phi_{(t+\tau) \mod \mu} + \gamma \left\lfloor \frac{t+\tau}{\mu} \right\rfloor - \phi_{t \mod \mu} - \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \mod 2\pi \\ &= \gamma \left\lfloor \frac{t}{\mu} + q \right\rfloor - \gamma \left\lfloor \frac{t}{\mu} \right\rfloor \mod 2\pi = \gamma \left(\left\lfloor \frac{t}{\mu} \right\rfloor + q - \left\lfloor \frac{t}{\mu} \right\rfloor \right) \mod 2\pi \\ &= \gamma q \mod 2\pi, \end{aligned}$$

where we have used the fact that $\left\lfloor \frac{t}{\mu} + q \right\rfloor = \left\lfloor \frac{t}{\mu} \right\rfloor + q$ as q is an integer. Again the obtained quantity does not depend on t . Therefore, $c_{\tilde{Z}}(t, t + \tau; \boldsymbol{\theta})$ does not depend on t but depend only on the lag τ . This proves that $\{\tilde{Z}_t\}$ is stationary with autocovariance sequence $c_{\tilde{Z}}(\tau; \boldsymbol{\theta}) = a^2 e^{i\gamma \frac{\tau}{\mu}} c_Z(\tau)$. As for the

spectral density of the resulting stationary modulated process $\{\tilde{Z}_t\}$, we have

$$\begin{aligned} S_{\tilde{Z}}(\omega; \theta) &= \sum_{\tau=-\infty}^{\infty} c_{\tilde{Z}}(\tau; \theta) e^{-i\omega\tau} = \sum_{\tau=-\infty}^{\infty} a^2 c_Z(\tau; \theta) e^{-i(\omega\tau - \gamma \frac{\tau}{\mu})} = \sum_{\tau=-\infty}^{\infty} a^2 c_Z(\tau; \theta) e^{-i(\omega - \frac{\gamma}{\mu})\tau} \\ &= a^2 \sum_{\tau=-\infty}^{\infty} c_Z(\tau; \theta) e^{-i(\omega - \frac{\gamma}{\mu})\tau} = a^2 S_Z\left(\omega - \frac{\gamma}{\mu}\right). \end{aligned}$$

This concludes the proof. Note that for a real-valued process, this shift would be impossible as the spectral density has to retain symmetry. As both Z_t and $\tilde{Z}_t$ are complex-valued, this is not a concern. $\square$

A.2. Proof of Proposition 5

Proof

Let us define the complex-valued stochastic process $\{Z_t\}$ according to

$$Z_t = e^{-i \sum_{u=1}^t \beta_u} \tilde{Z}_t, \quad t = 0, 1, 2, \dots$$

By applying the definition of the process $\{\tilde{Z}_t\}$, one can determine the following relationship, for all $t \geq 1$,

$$Z_t = e^{-i \sum_{u=1}^t \beta_u} \tilde{Z}_t = e^{-i\beta_t} e^{-i \sum_{u=1}^{t-1} \beta_u} \tilde{Z}_t = e^{-i\beta_t} e^{-i \sum_{u=1}^{t-1} \beta_u} (r e^{i\beta_t} \tilde{Z}_{t-1} + \epsilon_t) = r e^{-i \sum_{u=1}^{t-1} \beta_u} \tilde{Z}_{t-1} + \epsilon'_t,$$

and finally $Z_t = r Z_{t-1} + \epsilon'_t$, $t \geq 1$, where $\epsilon'_t = e^{-i \sum_{u=1}^t \beta_u} \epsilon_t$, $\forall t \in \mathbb{N}$ has the same distribution as ϵ_t , as we have assumed that the complex-valued white noise process ϵ_t has variance σ^2 and independent real and imaginary parts. Therefore, the process Z_t is a first-order complex-valued autoregressive process with constant stationary parameters. It is stationary if and only if $\text{var}\{Z_0\} = \frac{\sigma^2}{1-r^2}$. Since $Z_0 = \tilde{Z}_0$, it follows that $\text{var}\{Z_0\} = \text{var}\{\tilde{Z}_0\}$. Thus if $\text{var}\{\tilde{Z}_0\} = \frac{\sigma^2}{1-r^2}$, the process $\{Z_t\}$ is stationary. The fact that the process $\{Z_t\}$ is proper stems from the fact that the innovations $\{\epsilon_t\}$ as well as the random variable $\tilde{Z}_0$ are proper, as using the following relation, $Z_t = r^t Z_0 + \sum_{j=1}^t r^{t-j} \epsilon'_j$, we obtain for all $t, \tau \in \mathbb{N}$, $\mathbb{E}\{Z_t Z_{t+\tau}\} = 0$. This shows how the proposed process is generated by the stated mechanism of modulation as claimed in the proposition. $\square$

A.3. Proof of Proposition 6

Proof

Let $\Gamma = \{0, 1\}$. We show that conditions 1 and 2 given in Definition 3 are verified.

According to Proposition 5, there exists a stationary proper complex-valued process Z_t such that $\tilde{Z}_t = g_t Z_t$, where $g_t = e^{i \sum_{u=1}^t \beta_u}$, that is, $\tilde{Z}_t$ is a modulated process. The autocovariance sequence of the process Z_t is given by

$$c_Z(\tau) = \frac{\sigma^2}{1-r^2} r^\tau, \quad \tau \in \mathbb{Z},$$

and we observe that the function $(r, \sigma) \mapsto (c_Z(0), c_Z(1))$ is one-to-one. Let L be the largest positive (i.e. greater than or equal to 1) integer such that $\kappa \leq \frac{\pi}{2L}$. This is well defined as we have assumed $0 \leq \kappa < \frac{\pi}{2}$. Fix an integer

lag value $0 \leq \tau \leq L$. We have

$$\begin{aligned} \left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} g_t^* g_{t+\tau} \right| &= \left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t}^{t+\tau-1} \beta_u} \right| = \left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t}^{t+\tau-1} (\Xi + \beta_u - \Xi)} \right| \\ &= \left| \frac{1}{N} e^{i \tau \Xi} \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t}^{t+\tau-1} (\beta_u - \Xi)} \right| \geq \frac{1}{N} \left| \Re \left\{ \sum_{t=0}^{N-1-\tau} e^{i \sum_{u=t}^{t+\tau-1} (\beta_u - \Xi)} \right\} \right| \\ &= \frac{1}{N} \left| \sum_{t=0}^{N-1-\tau} \cos \left\{ \sum_{u=t}^{t+\tau-1} (\beta_u - \Xi) \right\} \right|. \end{aligned}$$

Using the triangle inequality, it follows $\left| \sum_{u=t}^{t+\tau-1} \beta_u - \Xi \right| \leq \sum_{u=t}^{t+\tau-1} |\beta_u - \Xi| \leq \tau \kappa$. With the fact that $\tau \kappa < \frac{\pi}{2}$ by assumption, and that the cosine function is decreasing on the interval $[0, \frac{\pi}{2}]$, we obtain

$$\left| \frac{1}{N} \sum_{t=0}^{N-1-\tau} g_t^* g_{t+\tau} \right| \geq \frac{1}{N} \sum_{t=0}^{N-1-\tau} \cos(\tau \kappa) = \left(1 - \frac{\tau}{N}\right) \cos(\tau \kappa) \xrightarrow{N \rightarrow \infty} \cos(\tau \kappa) > 0.$$

The preceding equation converges to a non-zero value as N goes to infinity, so that $\liminf_{N \rightarrow \infty} |c_g^{(N)}(\tau)| > 0$. It is true in particular for $\tau \in \Gamma$.

This shows that the process $\tilde{Z}_t$ is a modulated process with a significant correlation contribution. $\square$

A.4. Proof of Lemma 1

Proof

We denote $S_{X,\max} = \max_{\theta, \omega} S_X(\omega; \theta)$ and $S_{X,\min} = \min_{\theta, \omega} S_X(\omega; \theta)$.

1. We first show the existence of the upper bound. According to Proposition 1, the expected periodogram can be expressed, for $\omega \in [-\pi, \pi]$, $\theta \in \Theta$ and $N \in \mathbb{N}$, by

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = 2\pi \int_{-\pi}^{\pi} S_X(\omega - \lambda) S_g^{(N)}(\lambda) d\lambda.$$

Therefore,

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \leq 2\pi S_{X,\max} \int_{-\pi}^{\pi} S_g^{(N)}(\lambda) d\lambda = S_{X,\max} \frac{1}{N} \sum_{t=0}^{N-1} |g_t|^2,$$

by Parseval equality, and finally,

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \leq g_{\max}^2 S_{X,\max},$$

and this by assumption is finite.

2. Similarly, we show the existence of a lower bound. According to the assumption of a modulated process with significant correlation contribution, there exists a non-negative integer $\tau \in \Gamma$ and a positive real number α_τ

such that for N large enough, $c_g^{(N)}(\tau) \geq \alpha_\tau$. Then,

$$\begin{aligned}\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) &\geq 2\pi S_{X,\min} \int_{-\pi}^{\pi} S_g^{(N)}(\lambda) d\lambda = S_{X,\min} \frac{1}{N} \sum_{t=0}^{N-1} |g_t|^2 \\ &\geq S_{X,\min} \frac{1}{N} \sum_{t=0}^{N-\tau-1} |g_t|^2 \geq S_{X,\min} \frac{1}{N} \sqrt{\sum_{t=0}^{N-\tau-1} |g_t|^2} \sqrt{\sum_{t=\tau}^{N-1} |g_t|^2} \\ &\geq S_{X,\min} \frac{1}{N} \left| \sum_{t=0}^{N-\tau-1} g_t^* g_{t+\tau} \right|,\end{aligned}$$

by Cauchy–Schwartz inequality. Hence we obtain for N large enough $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) \geq \alpha_\tau S_{X,\min} > 0$. This proves the stated result.

A.5. Proof of Lemma 2

□

Proof

We have for all $\omega \in [-\pi, \pi]$, $\theta \in \Theta$ and $N \in \mathbb{N}$ that the form of $\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)$ is given by

$$\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) = 2\pi \int_{-\pi}^{\pi} S_X(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda.$$

We obtain (where the inversion of differentiation and integration is a consequence of the differentiability of the functions $\omega \rightarrow S_X(\omega; \theta)$ and the fact that the spectral densities are bounded above),

$$\begin{aligned}\left| \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \omega}(\omega; \theta) \right| &= 2\pi \left| \int_{-\pi}^{\pi} \frac{\partial S_X}{\partial \omega}(\omega - \lambda; \theta) S_g^{(N)}(\lambda) d\lambda \right| \leq 2\pi \max_{\lambda, \theta} \left\{ \left| \frac{\partial S_X}{\partial \omega}(\lambda; \theta) \right| \right\} \\ &\quad \times \int_{-\pi}^{\pi} S_g^{(N)}(\lambda) d\lambda \leq g_{\max}^2 \max_{\omega, \theta} \left\{ \left| \frac{\partial S_X}{\partial \omega}(\omega; \theta) \right| \right\},\end{aligned}$$

which concludes the proof.

□

A.6. Proof of Lemma 3

Proof

We will use repeatedly the fact that the function $x \rightarrow x - \log x$, defined on the set of positive real numbers, admits a global minimum for $x = 1$ where it takes value 1. It is an increasing function on $(1, \infty)$ and decreasing on $(0, 1)$. This is easily seen by studying the derivative. Now let N be a natural integer. We have for all $\gamma \in \Theta$

$$\begin{aligned}D(\gamma, \bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)) &= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\tilde{X}}^{(N)}(\omega; \gamma) + \frac{\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)}{\bar{S}_{\tilde{X}}^{(N)}(\omega; \gamma)} \right\} \\ &= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) + \overbrace{\frac{\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)}{\bar{S}_{\tilde{X}}^{(N)}(\omega; \gamma)} - \log \left[\frac{\bar{S}_{\tilde{X}}^{(N)}(\omega; \theta)}{\bar{S}_{\tilde{X}}^{(N)}(\omega; \gamma)} \right]}^{\geq 1} \right\} \\ &\geq \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \log \bar{S}_{\tilde{X}}^{(N)}(\omega; \theta) + 1 \right\},\end{aligned}$$

where we have an equality if and only if $\bar{S}_{\bar{X}}^{(N)}(\omega; \theta) = \bar{S}_{\bar{X}}^{(N)}(\omega; \gamma)$ for all $\omega \in \Omega_N$, which for N large enough is equivalent to $\gamma = \theta$ according to Proposition 3. $\square$

A.7. Proof of Lemma 4

Proof

We prove this in three steps.

1. We have for a positive integer N ,

$$D(\gamma_N, \bar{S}_{\bar{X}}^{(N)}(\cdot; \theta)) - D(\theta, \bar{S}_{\bar{X}}^{(N)}(\cdot; \theta)) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \frac{\bar{S}_{\bar{X}}^{(N)}(\omega; \theta)}{\bar{S}_{\bar{X}}^{(N)}(\omega; \gamma_N)} - \log \frac{\bar{S}_{\bar{X}}^{(N)}(\omega; \theta)}{\bar{S}_{\bar{X}}^{(N)}(\omega; \gamma_N)} - 1 \right\}. \quad (\text{A1})$$

By assumption, this converges to zero as N goes to infinity. For any non-negative integer τ smaller than N , we can write

$$\bar{c}_{\bar{X}}^{(N)}(\tau; \gamma_N) - \bar{c}_{\bar{X}}^{(N)}(\tau; \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\bar{S}_{\bar{X}}^{(N)}(\omega; \gamma_N) - \bar{S}_{\bar{X}}^{(N)}(\omega; \theta)) e^{i\omega\tau} d\omega, \quad (\text{A2})$$

so we have the following bound:

$$\left| \bar{c}_{\bar{X}}^{(N)}(\tau; \gamma_N) - \bar{c}_{\bar{X}}^{(N)}(\tau; \theta) \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \bar{S}_{\bar{X}}^{(N)}(\omega; \gamma_N) - \bar{S}_{\bar{X}}^{(N)}(\omega; \theta) \right| d\omega. \quad (\text{A3})$$

2. Now we assume, with the intent to arrive at a contradiction, that this quantity does not converge to zero. Then there exists an increasing function $\phi(N)$, defined on the set of non-negative integers and taking values in the set of non-negative integers and $\epsilon > 0$ such that

$$\left| \bar{c}_{\bar{X}}^{(\phi(N))}(\tau; \gamma_{\phi(N)}) - \bar{c}_{\bar{X}}^{(\phi(N))}(\tau; \theta) \right| \geq \epsilon, \quad \forall N \in \mathbb{N}. \quad (\text{A4})$$

Fix $N \in \mathbb{N}$. Denote M as the upper bound (independent of N) of the integrand in (A3) using Lemma 1. Let $B_{\phi(N)} \subset [-\pi, \pi]$ be the inverse image of $[\epsilon/2, \infty)$ by the function $\omega \mapsto \left| \bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - \bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \theta) \right|$. Let $\lambda_{\phi(N)}$ be the Lebesgue measure of the Borel set $B_{\phi(N)}$. We have

$$\epsilon \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - \bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \theta) \right| d\omega \leq \frac{\lambda_{\phi(N)}}{2\pi} M + \frac{2\pi - \lambda_{\phi(N)}}{2\pi} \frac{\epsilon}{2}, \quad (\text{A5})$$

and therefore,

$$\lambda_{\phi(N)} \geq \frac{\pi\epsilon}{M - \frac{\epsilon}{2}}. \quad (\text{A6})$$

Since $B_{\phi(N)}$ is defined such that

$$\left| \bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - \bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \theta) \right| \geq \frac{\epsilon}{2}, \quad \forall \omega \in B_{\phi(N)}, \quad (\text{A7})$$

it follows that, dividing each side of (A7) by $\bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})$,

$$\left| \frac{\bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \theta)}{\bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} - 1 \right| \geq \frac{\epsilon}{2\bar{S}_{\bar{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} \geq \frac{\epsilon}{2\bar{S}_{\bar{X}, \max}^{(\phi(N))}}, \quad \forall \omega \in B_{\phi(N)}. \quad (\text{A8})$$

We therefore have that for all $\omega \in B_{\phi(N)}$, $\left| \bar{S}_{\tilde{X}}^{(\phi(N))}(\omega; \theta) / \bar{S}_{\tilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)}) - 1 \right|$ is bounded below by $c\epsilon$, where $c = 1/(2\bar{S}_{\tilde{X},\max})$ is a positive constant independent of N . Denote

$$b : x \rightarrow x - \log x - 1, \quad x > 0, \quad \bar{b}^{(\phi(N))} : \omega \mapsto b \left(\frac{\bar{S}_{\tilde{X}}^{(\phi(N))}(\omega; \theta)}{\bar{S}_{\tilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} \right), \quad \omega \in [-\pi, \pi].$$

For all $\omega \in B_{\phi(N)}$, $\bar{b}^{(\phi(N))}(\omega)$ is bounded below by $d = \min(b(1 + c\epsilon), 1 - c\epsilon)$ (where $d > 0$ is a constant that depends on ϵ but not on N) because of the properties of the function $b(x)$, which we recalled at the beginning of the proof of Lemma 3. The function $b(x)$ has a bounded derivative on any interval of the form $[a_1, a_2]$, where $0 < a_1 < a_2 < \infty$. Since

$$\frac{\bar{S}_{\tilde{X},\max}}{\bar{S}_{\tilde{X},\min}} \geq \frac{\bar{S}_{\tilde{X}}^{(\phi(N))}(\omega; \theta)}{\bar{S}_{\tilde{X}}^{(\phi(N))}(\omega; \gamma_{\phi(N)})} \geq \frac{\bar{S}_{\tilde{X},\min}}{\bar{S}_{\tilde{X},\max}} > 0,$$

and using Lemmas 1 and 2, the function $\bar{b}^{(\phi(N))}(\omega)$ has a bounded derivative. We denote the corresponding bound $l_{\max}$, which is independent of N .

Recalling that $\lambda_{\phi(N)}$ is the measure of $B_{\phi(N)}$, there exist $T = \lfloor \frac{N\lambda_{\phi(N)}}{4\pi} \rfloor$ increasing elements $v_1, \dots, v_T \in B_{\phi(N)}$ such that $v_{i+1} - v_i \geq \frac{4\pi}{N}$, $i = 1, \dots, T-1$. Then there exist $T-1$ Fourier frequencies $v'_1, \dots, v'_{T-1}$, such that $v_i < v'_i < v_{i+1}$. Then we have

$$\left| \sum_{i=1}^{T-1} \bar{b}^{(\phi(N))}(v'_i) - \bar{b}^{(\phi(N))}(v_i) \right| \leq \sum_{i=1}^{T-1} \left| \bar{b}^{(\phi(N))}(v'_i) - \bar{b}^{(\phi(N))}(v_i) \right| \leq \sum_{i=1}^{T-1} (v'_i - v_i) l_{\max} \leq 2\pi l_{\max},$$

which implies

$$\begin{aligned} \sum_{i=1}^{T-1} \bar{b}^{(\phi(N))}(v'_i) &\geq \sum_{i=1}^{T-1} \bar{b}^{(\phi(N))}(v_i) - 2\pi l_{\max} \\ &\geq (T-1)d - 2\pi l_{\max}. \end{aligned}$$

Because T is of order N , we conclude that (A1) cannot converge to zero. We arrive at a contradiction.

So we obtain that for all integer τ , $\bar{c}_{\tilde{X}}^{(N)}(\tau; \gamma_N) - \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta)$ converges to zero when N goes to infinity.

3. In particular for $\tau \in \Gamma$, if we denote $\alpha_\tau = \liminf_{N \rightarrow \infty} |c_g^{(N)}(\tau)| > 0$, we have for N large enough,

$$\left| \bar{c}_{\tilde{X}}^{(N)}(\tau; \gamma_N) - \bar{c}_{\tilde{X}}^{(N)}(\tau; \theta) \right| = \left| c_g^{(N)}(\tau) (c_X(\tau; \gamma_N) - c_X(\tau; \theta)) \right| \geq \alpha_\tau |c_X(\tau; \gamma_N) - c_X(\tau; \theta)|,$$

so that (with the assumption of high correlation contribution) $|c_X(\tau; \gamma_N) - c_X(\tau; \theta)|$ converges to zero as N tends to infinity. Because of the compactness of Θ , and using the fact that the function $\theta \mapsto \{c_X(\tau) : \tau \in \Gamma\}$ is one-to-one and continuous, this yields the stated lemma. □

A.8. Proof of Lemma 5

Proof

Let $a_{\max}$ be a finite positive constant such that $|a^{(N)}(\omega)| \leq a_{\max}$, $\forall \omega \in [-\pi, \pi]$, $\forall N \in \mathbb{N}$. We start by looking at the covariance matrix of the Fourier transform. We shall denote the Fourier transform, for a fixed positive integer

N ,

$$J_{\tilde{X}}^{(N)}(\omega) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \tilde{X}_t e^{-i\omega t} = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} g_t X_t e^{-i\omega t}, \quad \omega \in \Omega_N.$$

Since the expectation of the latent process is assumed to be zero, the same holds for the Fourier transform by the linearity of the Fourier transform. Hence from the preceding linear equation, we see that the covariance matrix elements can be expressed in the following way:

$$\text{cov} \left\{ J_{\tilde{X}}^{(N)}(\omega), J_{\tilde{X}}^{(N)}(\omega') \right\} = \frac{1}{N} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) G_{\omega'}^{(N)}, \quad \omega, \omega' \in \Omega_N, \quad (\text{A9})$$

where superscript H denotes the Hermitian transpose, $C_X^{(N)}(\theta)$ denotes the finite theoretical autocovariance matrix of the latent process (i.e. with elements $c_X(i-j; \theta)$, $i, j = 0, \dots, N-1$) and $G_{\omega}^{(N)}$ is the vector $[g_t e^{i\omega t} : t = 0, \dots, N-1]^T$. Using Isserlis's theorem and the assumption of Gaussianity of the latent process (which in turns implies the Gaussianity of the Fourier transform of the modulated process), the covariances of the periodogram are related to the covariances of the Fourier transform according to the following simple relation:

$$\text{cov} \left\{ \hat{S}_{\tilde{X}}^{(N)}(\omega), \hat{S}_{\tilde{X}}^{(N)}(\omega') \right\} = \left| \text{cov} \left\{ J_{\tilde{X}}^{(N)}(\omega), J_{\tilde{X}}^{(N)}(\omega') \right\} \right|^2. \quad (\text{A10})$$

This can be written as

$$\begin{aligned} \text{cov} \left\{ \hat{S}_{\tilde{X}}^{(N)}(\omega), \hat{S}_{\tilde{X}}^{(N)}(\omega') \right\} &= \frac{1}{N^2} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) G_{\omega'}^{(N)} \left((G_{\omega}^{(N)})^H C_X^{(N)}(\theta) G_{\omega'}^{(N)} \right)^H \\ &= \frac{1}{N^2} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) G_{\omega'}^{(N)} G_{\omega'}^{(N)H} C_X^{(N)}(\theta)^H G_{\omega}^{(N)}. \end{aligned}$$

We then have

$$\begin{aligned} \text{var} \left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a^{(N)}(\omega) \hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} &= \\ \frac{1}{N^2} \sum_{\omega \in \Omega_N} \sum_{\omega' \in \Omega_N} a^{(N)}(\omega) a^{(N)}(\omega') \frac{1}{N^2} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) G_{\omega'}^{(N)} G_{\omega'}^{(N)H} C_X^{(N)}(\theta)^H G_{\omega}^{(N)} \\ &\leq \frac{a_{\max}^2}{N^4} \sum_{\omega \in \Omega_N} \sum_{\omega' \in \Omega_N} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) G_{\omega'}^{(N)} G_{\omega'}^{(N)H} C_X^{(N)}(\theta)^H G_{\omega}^{(N)} \\ &= \frac{a_{\max}^2}{N^4} \sum_{\omega \in \Omega_N} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) \sum_{\omega' \in \Omega_N} \left\{ G_{\omega'}^{(N)} G_{\omega'}^{(N)H} \right\} C_X^{(N)}(\theta)^H G_{\omega}^{(N)}, \end{aligned}$$

where the first inequality is legitimate as the covariances of the periodogram are positive real-valued numbers (equation (A10)), and where the last equality is obtained after factorizing. Now we use the fact that

$$\sum_{\omega' \in \Omega_N} \left\{ G_{\omega'}^{(N)} G_{\omega'}^{(N)H} \right\} = N \text{diag}(g_0^2, \dots, g_{N-1}^2). \quad (\text{A11})$$

Indeed, the (t_1, t_2) th term of the left-hand side of (A11) is given by

$$\sum_{\omega' \in \Omega_N} g_{t_1} g_{t_2} e^{i\omega'(t_1-t_2)} = \sum_{k=0}^{N-1} g_{t_1} g_{t_2} e^{\frac{i2k\pi(t_1-t_2)}{N}} = g_{t_1} g_{t_2} \sum_{k=0}^{N-1} e^{\frac{i2k\pi(t_1-t_2)}{N}}, \quad (\text{A12})$$

where we recognize the finite sum of the geometric sequence of term $e^{\frac{i2\pi(t_1-t_2)}{N}}$, which is N if $t_1 = t_2$, and otherwise,

$$\sum_{k=0}^{N-1} e^{\frac{i2k\pi(t_1-t_2)}{N}} = \sum_{k=0}^{N-1} \left(e^{\frac{i2\pi(t_1-t_2)}{N}} \right)^k = \frac{1 - \left(e^{\frac{i2\pi(t_1-t_2)}{N}} \right)^N}{1 - e^{\frac{i2\pi(t_1-t_2)}{N}}} = 0. \quad (\text{A13})$$

Therefore,

$$\text{var} \left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a(\omega) \hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} \leq \frac{a_{\max}^2}{N^3} \sum_{\omega \in \Omega_N} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) \text{diag}(g_0^2, \dots, g_{N-1}^2) C_X^{(N)}(\theta)^H G_{\omega}^{(N)} \quad (\text{A14})$$

$$\leq \frac{a_{\max}^2 g_{\max}^2}{N^3} \sum_{\omega \in \Omega_N} (G_{\omega}^{(N)})^H C_X^{(N)}(\theta) C_X^{(N)}(\theta)^H G_{\omega}^{(N)}. \quad (\text{A15})$$

Therefore, we now have

$$\text{var} \left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a(\omega) \hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} \leq \frac{a_{\max}^2 g_{\max}^2}{N^3} \sum_{\omega \in \Omega_N} \left\| C_X^{(N)}(\theta)^H G_{\omega}^{(N)} \right\|_2^2, \quad (\text{A16})$$

where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{C}^N$. For all $\mathbf{U} \in \mathbb{C}^N$, the matrix $C_X^{(N)}(\theta)$ is Hermitian, so it can be written PDP^H , where D is a diagonal matrix and P is unitary, so that

$$\left\| C_X^{(N)}(\theta) \mathbf{U} \right\|_2 \leq \|\mathbf{U}\|_2 \max_{\eta \in \text{sp}(C_X^{(N)}(\theta))} |\eta|, \quad (\text{A17})$$

where $\text{sp}(C_X^{(N)}(\theta))$ is the set of eigenvalues of $C_X^{(N)}(\theta)$. Furthermore, we have from Horn and Johnson (1985, p. 394) that, recalling that the spectral density $S_X(\omega; \theta)$ is assumed to be continuous in ω ,

$$\max_{\eta \in \text{sp}(C_X^{(N)}(\theta))} |\eta| = \max_{\mathbf{U} \in \mathbb{C}^n} \left\{ \frac{\mathbf{U}^H C_X^{(N)}(\theta) \mathbf{U}}{\mathbf{U}^H \mathbf{U}} \right\} \leq S_{X, \max}. \quad (\text{A18})$$

Combining (A16)–(A18) and replacing $\mathbf{U}$ by $G_{\omega}^{(N)}$,

$$\text{var} \left\{ \frac{1}{N} \sum_{\omega \in \Omega_N} a(\omega) \hat{S}_{\tilde{X}}^{(N)}(\omega) \right\} \leq \frac{(S_{X, \max} a_{\max} g_{\max}^2)^2}{N}, \quad (\text{A19})$$

as $\|G_{\omega}^{(N)}\|_2 \leq g_{\max} \sqrt{N}$. □

A.9. Proof of Lemma 6

Proof

We first show that the proposition is true for all vectors in $\mathcal{C} = \{\mathbf{V} \in \mathbb{R}^d : \|\mathbf{V}\|_2 = 1\}$, which is compact. The function $\mathcal{S} : \mathbf{V} \mapsto \sum_{i=1}^d \alpha_i^2 (\mathbf{U}_i^T \mathbf{V})^2$ is continuous, so the image of $\mathcal{C}$ by $\mathcal{S}$ is compact. Since $\mathcal{S}$ takes non-negative values, the image of $\mathcal{C}$ by $\mathcal{S}$ either contains zero or there exists a constant $C > 0$ such that it is bounded below by C . The image of $\mathcal{C}$ by $\mathcal{S}$ cannot contain zero, as otherwise there would be a vector of $\mathbb{R}^d$ with norm 1 whose scalar product with vectors $U_i, i = 1, \dots, d$ is zero, which is impossible as we have assumed that the family $U_1, \dots, U_d$ has rank d . Therefore, there exists a constant $C > 0$ such that

$$\sum_{i=1}^d \alpha_i^2 (\mathbf{U}_i^T \mathbf{V})^2 \geq C, \quad \forall \mathbf{V} \in \mathcal{C}.$$

Now in general, if $\mathbf{V}$ is any non-zero vector in $\mathbb{R}^d$, we have, using the result we have derived for vectors of $\mathcal{C}$,

$$\begin{aligned} \sum_{i=1}^d \alpha_i^2 (\mathbf{U}_i^T \mathbf{V})^2 &= \|\mathbf{V}\|_2^2 \sum_{i=1}^d \alpha_i^2 \left(\mathbf{U}_i^T \frac{\mathbf{V}}{\|\mathbf{V}\|_2} \right)^2 \\ &\geq \|\mathbf{V}\|_2^2 C. \end{aligned}$$

If $\mathbf{V} = 0$ the result is obvious. This concludes the proof in the general case. $\square$

A.10. Proof of Lemma 7

Proof

1. Direct calculations give that the score function takes the form, for $i = 1, \dots, d$,

$$\frac{\partial l_M^{(N)}}{\partial \theta_i}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \frac{\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta})}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) - \hat{S}_{\tilde{X}}^{(N)}(\omega) \right) \right\}.$$

Since by definition $\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) = E \left\{ \hat{S}_{\tilde{X}}^{(N)}(\omega); \boldsymbol{\theta} \right\}$, the null expectation of the score function is zero. Applying

Lemma 5, and the fact that $\frac{\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta})}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2}$ is bounded above in absolute value independently of N (as a direct consequence of Lemmas 1 and 2), we obtain the first result.

2. Again by direct calculation, we obtain the following Hessian matrix:

$$\begin{aligned} \frac{\partial^2 l_M^{(N)}}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}) &= \frac{1}{N} \sum_{\omega \in \Omega_N} \left\{ \frac{\frac{\partial^2 \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i \partial \theta_j}(\omega; \boldsymbol{\theta}) \left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2 - 2 \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta}) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_j}(\omega; \boldsymbol{\theta})}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^4} \right. \\ &\quad \times \left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) - \hat{S}_{\tilde{X}}^{(N)}(\omega) \right) + \frac{1}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta}) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_j}(\omega; \boldsymbol{\theta}) \left. \right\}. \end{aligned}$$

The expectation of the Hessian matrix is therefore

$$\mathcal{I}^{(N)}(\boldsymbol{\theta}) = \mathbb{E} \left\{ \frac{\partial^2 l_M^{(N)}(\omega; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\} = \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \frac{\partial \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T},$$

where we use the notation $\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}^T}(\omega; \boldsymbol{\theta})$ to denote the transpose of the gradient vector $\frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}}(\omega; \boldsymbol{\theta})$. For any of the $\omega \in \Omega_N$ (to which corresponds a term in the above sum), and for any vector $\mathbf{U} \in \mathbb{R}^d$,

$$\mathbf{U}^T \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}}(\omega; \boldsymbol{\theta}) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}^T}(\omega; \boldsymbol{\theta}) \mathbf{U} = \left| \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}^T}(\omega; \boldsymbol{\theta}) \mathbf{U} \right|^2 \geq 0,$$

so that the matrix $\mathcal{I}^{(N)}(\boldsymbol{\theta})$ is non-negative definite as a sum of non-negative definite matrices. Now to show that the matrix $\mathcal{I}(\boldsymbol{\theta})$ is positive definite, let $\mathbf{U} = [u_1, \dots, u_d]^T \in \mathbb{R}^d$ be non-zero. We have

$$\begin{aligned} \mathbf{U}^T \mathcal{I}^{(N)}(\boldsymbol{\theta}) \mathbf{U} &= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \mathbf{U}^T \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}}(\omega; \boldsymbol{\theta}) \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}^T}(\omega; \boldsymbol{\theta}) \mathbf{U} \\ &= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \left| \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \boldsymbol{\theta}^T}(\omega; \boldsymbol{\theta}) \mathbf{U} \right|^2 \\ &= \frac{1}{N} \sum_{\omega \in \Omega_N} \frac{1}{\left(\bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \left(\sum_{i=1}^d \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta}) u_i \right)^2 \\ &\geq \frac{1}{N \left(\sup_{\omega \in \Omega_N} \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \sum_{\omega \in \Omega_N} \left(\sum_{i=1}^d \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta}) u_i \right)^2. \end{aligned}$$

Seeing $\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^d u_i \frac{\partial \bar{S}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\omega; \boldsymbol{\theta}) \right\}_{\omega \in \Omega_N} \Leftrightarrow \left\{ \sum_{i=1}^d u_i \frac{\partial \bar{c}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\tau; \boldsymbol{\theta}) \right\}_{\tau=-(N-1), \dots, N-1}$ as a finite Fourier pair and applying Parseval's equality, we obtain that

$$\begin{aligned} \mathbf{U}^T \mathcal{I}^{(N)}(\boldsymbol{\theta}) \mathbf{U} &\geq \frac{1}{\left(\sup_{\omega \in \Omega_N} \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \sum_{\tau=-(N-1)}^{N-1} \left(\sum_{i=1}^d u_i \frac{\partial \bar{c}_{\tilde{X}}^{(N)}}{\partial \theta_i}(\tau; \boldsymbol{\theta}) \right)^2 \\ &= \frac{1}{\left(\sup_{\omega \in \Omega_N} \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta}) \right)^2} \sum_{\tau=-(N-1)}^{N-1} \left(\sum_{i=1}^d u_i c_g^{(N)}(\tau) \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta}) \right)^2, \end{aligned}$$

by definition of $\bar{c}_{\tilde{X}}^{(N)}(\tau; \boldsymbol{\theta})$ and noting that $c_g^{(N)}(\tau)$ does not depend on $\boldsymbol{\theta}$. Therefore,

$$\mathbf{U}^T \mathcal{I}^{(N)}(\boldsymbol{\theta}) \mathbf{U} \geq \frac{1}{\left(\sup_{\omega \in \Omega_N} \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)^2} \sum_{\tau \in \Gamma} \left(\sum_{i=1}^d u_i c_g^{(N)}(\tau) \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta}) \right)^2,$$

as long as N is larger than the greater integer value in Γ . Denote $\alpha_\tau = \liminf_{N \rightarrow \infty} |c_g^{(N)}(\tau)| > 0$ for $\tau \in \Gamma$, and we obtain that for N large enough (equation (10)),

$$\begin{aligned} \mathbf{U}^T \mathcal{I}^{(N)}(\boldsymbol{\theta}) \mathbf{U} &\geq \frac{1}{\left(\sup_{\omega \in \Omega_N} \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)^2} \sum_{\tau \in \Gamma} c_g^{(N)}(\tau)^2 \left(\sum_{i=1}^d u_i \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta}) \right)^2 \\ &\geq \frac{1}{\left(\sup_{\omega \in \Omega_N} \bar{S}_{\tilde{X}}^{(N)}(\omega; \boldsymbol{\theta})\right)^2} \sum_{\tau \in \Gamma} \alpha_\tau^2 \left(\sum_{i=1}^d u_i \frac{\partial c_X}{\partial \theta_i}(\tau; \boldsymbol{\theta}) \right)^2. \end{aligned}$$

Now according to the assumption of significant correlation contribution, the mapping $\boldsymbol{\theta} \mapsto [c_X(\tau) : \tau \in \Gamma]^T$ is one-to-one, so its Jacobian taken at the true parameter vector $\boldsymbol{\theta}$ is non-zero. Therefore, the family $\frac{\partial c_X}{\partial \boldsymbol{\theta}}(\tau; \boldsymbol{\theta}) : \tau \in \Gamma$ has rank d , and we can apply Lemma 6, that is, we can conclude that there exists a positive constant $C > 0$ such that

$$\mathbf{U}^T \mathcal{I}^{(N)}(\boldsymbol{\theta}) \mathbf{U} \geq C \|\mathbf{U}\|_2^2.$$

This implies that the norm of the expected Hessian matrix is bounded below by a positive real number. Similarly to the gradient, using Lemma 5, we obtain the stated result for the Hessian.

□