Wavelet Spectra for Multivariate Point Processes

Edward A. K. Cohen[1] and Alexander J. Gibberd[2]

[1]Department of Mathematics, Imperial College London, South Kensington Campus,
London SW7 2AZ, U.K.
[2]Department of Mathematics and Statistics, Lancaster University, Bailrigg,
Lancaster LA1 4YF, U.K.

Abstract

Wavelets provide the flexibility to analyse stochastic processes at different scales. Here, we apply them to multivariate point processes as a means of detecting and analysing unknown non-stationarity, both within and across data streams. To provide statistical tractability, a temporally smoothed wavelet periodogram is developed and shown to be equivalent to a multi-wavelet periodogram. Under a stationary assumption, the distribution of the temporally smoothed wavelet periodogram is demonstrated to be asymptotically Wishart, with the centrality matrix and degrees of freedom readily computable from the multi-wavelet formulation. Distributional results extend to wavelet coherence; a time-scale measure of inter-process correlation. This statistical framework is used to construct a test for stationarity in multivariate point-processes. The methodology is applied to neural spike train data, where it is shown to detect and characterise time-varying dependency patterns.

1   Introduction

We adopt the construction of Hawkes (1971) which presents a p-dimensional multivariate point process $(p \geq 1)$ as a counting vector $N(t) \equiv \{N_1(t), \ldots, N_p(t)\}^{\mathrm{T}}$ where the random element $N_i(t)$ $(i = 1, \ldots, p)$ states the number of events of type i over the interval $(0, t]$. Its

first order properties are characterized by its rate $\lambda(t) \in \mathbb{R}^p$, defined as $\lambda(t) \equiv E\{dN(t)\}/dt$ where $dN(t) = N(t + dt) - N(t)$, and its second order properties at times s and t characterized by its covariance density matrix

$$\Gamma(s,t) = E\{dN(s)dN^{\mathrm{T}}(t)\}/(dt\ ds) - \lambda(s)\lambda^{\mathrm{T}}(t)\ .$$

Process $N(t)$ is second-order stationary, henceforth referred to simply as "stationary", if $\lambda(t)$ is constant for all t and $\Gamma(t,s)$ depends only on $\tau = s - t$. In this setting we will denote the covariance density matrix $\Gamma(\tau)$.

The spectral domain provides a rich environment for representing this second order structure and is based on the fact that stationary stochastic processes can be considered as a composite of subprocesses operating at different frequencies. The spectral density matrix of a stationary point process is the Fourier transform of its covariance density matrix (Bartlett, 1963), namely

$$S(f) = \mathrm{diag}(\lambda) + \int_{-\infty}^{\infty} \Gamma(\tau)e^{-i2\pi f\tau}d\tau, \quad -\infty < f < \infty.$$

A fundamental summary of the second order relationship between a pair of component processes, $N_i(t)$ and $N_j(t)$ say, is their coherence defined as

$$\rho_{ij}^2(f) = \frac{|S_{ij}(f)|^2}{S_{ii}(f)S_{jj}(f)}. \tag{1}$$

This provides a normalized measure on $[0, 1]$ of the correlation structure between the processes in the frequency domain. For time series data, it has been used extensively in several disciplines, including climatology, oceanography and medicine. For event data, it has been an important tool in neuroscience for the analysis of neuron spike train data.

Estimation of the coherence can be achieved by substituting smoothed spectral estimators into (1). Failure to smooth, i.e. simply using the periodogram, will result in a coherence estimate of one for all frequencies, irrespective of whether correlation exists between the pair of processes or not. Tractability of the coherence estimator's distribution is crucial for principled statistical testing and dependent on the smoothing procedure used (Walden, 2000).

Often, stochastic processes do not conform to the assumptions of stationarity. This might occur through simple first-order trends in the underlying data generating process, or more

typically, complex changes in its second, or higher, order structure. For point processes, the key objective now is to analyse how correlations within and across event streams change in time. One approach is to model the process and its associated time-varying spectrum, with the locally stationary Hawkes process (Roueff et al., 2016; Roueff & Von Sachs, 2019) representing an important recent advancement. The present paper complements this by developing a versatile non-parametric approach for exploring second-order structure in a time-localised way. Wavelets form a natural basis with which to do this due to their inherent ability to trade-off time and frequency resolution. Wavelet methods designed for time series analysis (e.g. Nason et al., 2000) could be deployed on discrete-time count sequences. However, should precise timestamps for the events be available, the necessary binning procedure required to implement them would discard information. This motivates the development of continuous time methodologies that preserves all temporal information.

For a wavelet $\psi(t)$, the continuous wavelet transform at scale $a > 0$ and translation (or time) $b \in \mathbb{R}$ of $N(t)$, observed on the interval $(0, T]$, is defined by Brillinger (1996) as

$$w(a, b) = a^{-1/2} \int_0^T \psi^*\{(t-b)/a\}\mathrm{d}N(t) , \tag{2}$$

where * denotes the complex conjugate. The ith element of this stochastic integral is computed as $w_i(a, b) = \sum_{k=1}^{N_i(T)} \psi_{a,b}^*(s_{i,k})$, where $s_{i,1}, ..., s_{i,N_i(T)}$ are the ordered event times of $N_i(t)$ and $\psi_{a,b}(t) \equiv a^{-1/2}\psi\{(t-b)/a\}$. Thus, working with the continuous time process is possible if the finite set of event times are known. The wavelet periodogram is subsequently defined as $W(a, b) = w(a, b)w^{\mathrm{H}}(a, b)$, where $^{\mathrm{H}}$ denotes the complex conjugate transpose.

As is the case with the Fourier periodogram, smoothing is required for two reasons. Firstly to control variance, and secondly to give meaningful values of the wavelet coherence estimator. Wavelet coherence is an analogue of coherence which provides a normalized measure on $[0, 1]$ of the correlation between a pair of processes in time-scale space. It is defined as

$$\gamma_{ij}^2(a, b) = \frac{|\Omega_{ij}(a, b)|^2}{\Omega_{ii}(a, b)\Omega_{jj}(a, b)},$$

where Ω is a smoothed version of W. In the time series setting, wavelet coherence has been extensively applied in a wide range of disciplines (e.g. Torrence & Webster, 1999; Grinsted et al., 2004). Understanding the distributional properties of these smoothed coherence

estimators is vital for rigorous statistical analysis and testing. In the Gaussian discrete-time setting the asymptotic distribution of coherence is widely studied (Cohen & Walden, 2010a,b), however, the point-process case has received little attention.

There are a wide range of ways in which non-stationarity can occur. Hence, rather than assume a specific model of non-stationarity, we propose to study the properties of the temporally smoothed wavelet periodogram and coherence for stationary point-processes. This provides a statistical framework in which we deploy methods for exploratory data analysis and construct a test for second order stationarity, complementing stationarity tests in time series analysis (e.g. Von Sachs & Neumann, 2000; Nason, 2013; Preuss et al., 2013).

2   Temporally smoothed wavelet periodogram

2.1   Formulation

Assumption 1. *Wavelet $\psi(t)$ is a real or complex valued continuous function that satisfies (i) $\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t = 0$, (ii) $\|\psi\|_2 = 1$, and (iii) the admissibility condition $\int_{-\infty}^{\infty} f^{-1}|\Psi(f)|^2\mathrm{d}f < \infty$, where Ψ is the Fourier transform of $\psi(t)$.*

Assumption 2. *Smoothing function $h(t)$ is a non-negative, symmetric function supported and continuous on $(-1/2, 1/2)$, and normalized such that $\int_{-\infty}^{\infty} h(t)\mathrm{d}t = 1$.*

Let $\psi(t)$ and $h(t)$ satisfy Assumptions 1 and 2, respectively. We define the *temporally smoothed wavelet periodogram* as

$$\Omega(a, b) = \int_{-\infty}^{\infty} h_\xi(u - b)W(a, u)\mathrm{d}u, \tag{3}$$

where $h_\xi \equiv \xi^{-1}h(t/\xi)$ with $\xi > 0$ controlling the level of smoothing. This is a wavelet analogue to Welch's *weighted overlapping sample averaging* spectral estimator for stationary time series (Welch, 1967; Carter, 1987). It will prove convenient for the level of smoothing to scale with a, and we therefore let $\xi = \kappa a$, with $\kappa > 0$.

For a particular choice of κ, and defining the Hermitian kernel function (at scale $a = 1$) as

$$K(s,t) = \int_{-\infty}^{\infty} h_\kappa(u)\psi(s-u)\psi^*(t-u)\mathrm{d}u \,, \tag{4}$$

the temporally smoothed wavelet periodogram in (3) can be expressed as

$$\Omega(a,b) \equiv \int_0^T \int_0^T K_{a,b}(s,t)\mathrm{d}N(t)\mathrm{d}N^{\mathrm{T}}(s) \,,$$

where $K_{a,b}(s,t) = a^{-1}K\{(s-b)/a, (t-b)/a\}$. The (i,j)th element of $\Omega(a,b)$ is computed as

$$\Omega_{ij}(a,b) = \sum_{k=1}^{N_i(T)} \sum_{k'=1}^{N_j(T)} K_{a,b}(s_{i,k}, s_{j,k'}) \,.$$

Given a choice for $h(t)$ and κ, the form of $K(s,t)$ will depend on $\psi(t)$. Throughout this paper, we use the examples of the complex valued Morlet wavelet and the real valued Mexican hat wavelet. These are examples of wavelets for which $K(s,t)$ is analytically tractable.

2.2 Practical implementation

For continuous time wavelet analysis, the wavelets themselves are often non-compactly supported. However, the region of significant support is typically well localized and a close approximation to $w(a,b)$ can be obtained through utilising the approximating wavelet

$$\bar{\psi}(t) = \begin{cases} \psi(t) & |t| < \alpha/2 \\ 0 & \text{otherwise.} \end{cases}$$

For example, the Morlet wavelet $\psi(t) = \pi^{-1/4}\mathrm{e}^{-t^2/2}\mathrm{e}^{\mathrm{i}2\pi t}$ shown in Fig. 1 has infinite support but can be well approximated by $\bar{\psi}(t)$ for $\alpha = 8$. In practice, to speed up computation, it can make sense to use the approximating wavelet as only a subset of the data is required to compute the wavelet transform. From herein, to simplify notation, will we use $\psi(t)$ to represent both the original and approximating wavelet, assuming that α is chosen appropriately.

In a finite data setting we are restricted to regions of the time-scale space in which we can fairly evaluate (2) without the consequences of edge effects at either ends of the data.
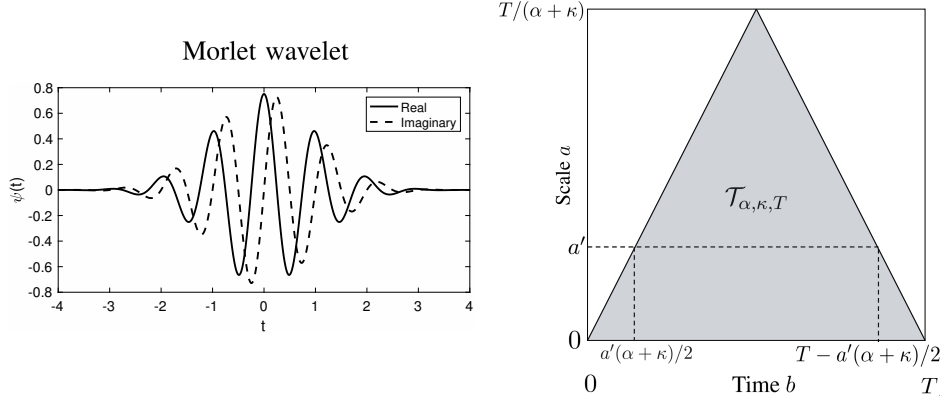
5

Figure 1: The Morlet wavelet and the valid region for analysis $\mathcal{T}_{\alpha,\kappa,T}$. Note this has been plotted with time b on the horizontal axis and scale a on the vertical axis, as is convention

These issues are compounded when smoothing across time, for a smoothing window $h_\kappa(t)$ with $\text{supp}(h) = (-\kappa/2, \kappa/2)$, the effective size of support for $K(s,t)$ is $\alpha + \kappa$, therefore we restrict ourselves to values of a and b for which $\text{supp}(K_{a,b}) = (b - a(\alpha + \kappa)/2, b + a(\alpha + \kappa)/2) \times (b - a(\alpha + \kappa)/2, b + a(\alpha + \kappa)/2) \subseteq (0, T] \times (0, T]$. This defines an isosceles triangle $\mathcal{T}_{\alpha,\kappa,T} \subset \mathbb{R}^2$ with vertices $(0,0)$, $(0,T)$ and $(a_{\max}(T), T/2)$, where $a_{\max}(T) = T/(\alpha + \kappa)$. This is an adaptation to the *cone of influence* (Mallat & Peyré, 2008, p. 215) that also mitigates for smoothing distances. In practice, a positive minimum value of a should be imposed to ensure a reasonable amount of event data exists in the smoothing range.

3 Multi-wavelet representation

3.1 Formulation

Given $K(s,t)$ is continuous and non-negative definite by construction, associated with kernel $K(s,t)$ is the Hermitian linear operator T_K defined as $[T_K f](s) = \int_{-\infty}^{\infty} K(s,t) f(t) \mathrm{d}t$. It follows from Mercer's Theorem (Mercer, 1909) that $K(s,t) = \sum_{l=0}^{\infty} \eta_l \varphi_l(s) \varphi_l^*(t)$ where $\{\varphi_l(t);\ l = 0, 1, ...\}$ are the orthonormal eigenfunctions of T_K with non-zero eigenvalues $\{\eta_l;\ l = 0, 1, ...\}$ ordered in decreasing size. Noting that $\text{tr}(T_K) := \int_{-\infty}^{\infty} K(t,t) \mathrm{d}t = 1$, it

follows that $\sum_{l=0}^{\infty} \eta_l = 1$. From here on, we refer to $\{\varphi_l(t); \ l = 0, 1, ...\}$ as the eigenfunctions of $K(s, t)$. The following proposition shows that these orthonormal eigenfunctions are themselves wavelets.

Proposition 1. *Let $\psi(t)$ satisfy Assumption 1, $h(t)$ satisfy Assumption 2, and for $\kappa > 0$ the corresponding non-negative definite kernel $K(s, t)$ have eigenfunctions $\{\varphi_l(t); \ l = 0, 1, ...\}$. Every eigenfunction $\varphi_l(t)$ with a non-zero eigenvalue is a wavelet that satisfies the conditions of Assumption 1.*

We adopt the term *eigen-wavelets* for the functions $\{\varphi_l(t); \ l = 0, 1, ...\}$.

Turning our attention back to the temporally smoothed wavelet periodogram, it is straightforward to show

$$\int_{-\infty}^{\infty} K_{a,b}(s, t)\varphi_l\{(t - b)/a\}\mathrm{d}t = \eta_l\varphi_l\{(s - b)/a\}.$$

Thus, the scaled and shifted versions $\varphi_{l,a,b}(t) = a^{-1/2}\varphi_l\{(t - b)/a\}, \ l = 0, 1, \ldots$ of the eigen-wavelets are themselves the eigenfunctions of $K_{a,b}$, and again from Mercer's theorem $K_{a,b}(s, t) = \sum_{l=0}^{\infty} \eta_l\varphi_{l,a,b}(s)\varphi_{l,a,b}^*(t)$. The temporally smoothed wavelet periodogram can thus be represented as

$$\Omega(a, b) = \sum_{l=0}^{\infty} \eta_l v_l(a, b)v_l^{\mathrm{H}}(a, b), \tag{5}$$

where $v_l(a, b) = \int_0^T \varphi_{l,a,b}(t)\mathrm{d}N(t)$ is the continuous wavelet transform of $N(t)$ at scale a and translation b with respect to eigen-wavelet $\varphi_l(t)$. Therefore the temporally smoothed wavelet periodogram is equivalent to the weighted sum of wavelet spectra arising from the orthonormal eigen-wavelet system. This is analogous to multitapering (Thomson, 1982) and comparisons can also be drawn with the multi-wavelet spectrum of Cohen & Walden (2010b). In that setting, multiple orthogonal wavelets were derived in Olhede & Walden (2002) from a time-frequency concentration problem, whereas here we have shown they can be generated by any arbitrary wavelet $\psi(t)$ and smoothing window $h(t)$.

The representation in (5) will be crucial for deriving the distributional results in Section 4, as well as offering computational speed-up. In particular, we will make use of the following proposition which shows the effective frequency response of the eigen-wavelet system is equal to the frequency response of the generating wavelet $\psi(t)$.

Proposition 2. *Let $\psi(t)$ satisfy Assumption 1, $h(t)$ satisfy Assumption 2, and for $\kappa > 0$ the corresponding non-negative definite kernel $K(s,t)$ have eigenfunctions $\{\varphi_l(t);\ l = 0,1,...\}$ and eigenvalues $\{\eta_l;\ l = 0,1,...\}$. It holds that $\sum_l \eta_l |\Phi_l(f)|^2 = |\Psi(f)|^2$ where $\Phi_l(f)$ and $\Psi(f)$ are the Fourier transforms of $\varphi_l(t)$ and $\psi(t)$, respectively.*

In general, closed form expressions for the eigen-wavelets $\{\varphi_l(t);\ l = 0,1,\ldots\}$ will be unobtainable and numerical procedures need to be used to find the solutions of $\int_{-\infty}^{\infty} K(s,t)\varphi(t)\mathrm{d}t = \eta\varphi(s)$. Details for an implementation of the Nystrom method for doing just this can be found in Appendix 1.

3.2   Worked example

The Morlet wavelet can be seen as a complex sinusoid enveloped with a Gaussian window, and therefore the wavelet transform at scale $a > 0$ and translation b is the Fourier transform of the tapered process, localized at b and evaluated at frequency $1/a$. The temporally smoothed wavelet periodogram using a rectangular smoothing function

$$h(t) = \begin{cases} 1 & -1/2 < t < 1/2 \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

emits kernel $K(s,t) = k(s,t)\mathrm{e}^{-\mathrm{i}2\pi(t-s)}$, where

$$k(s,t) = (2\kappa)^{-1}\mathrm{e}^{-(t-s)^2}[\mathrm{erf}\{\kappa - (t+s)\} + \mathrm{erf}\{\kappa + (t+s)\}]$$

and $\mathrm{erf}(x) = \pi^{-1/2}\int_{-x}^{x}\exp(-t^2)\mathrm{d}t$ is the Gauss error function. The real part of this kernel is shown in Fig. 2a.

The function $k(s,t)$ is itself a real valued non-negative kernel with its own set of real valued orthonormal eigenfunctions $\{\phi_l(t);\ l = 0,1,...\}$ and associated eigenvalues $\{\eta_l;\ l = 0,1,...\}$. It follows that $\varphi_l(t) = \mathrm{e}^{\mathrm{i}2\pi t}\phi_l(t)$ is an eigenfunction of $K(s,t)$ with corresponding eigenvalue η_l and hence $\{\varphi_l(t) = \mathrm{e}^{\mathrm{i}2\pi t}\phi_l(t); l = 0,1,...\}$ is the eigen-wavelet system emitted by the Morlet wavelet with a rectangular smoothing function. The first five of these eigen-wavelets for $\kappa = 10$ are shown in Fig. 2b. This eigen-wavelet system follows the same spirit of the generating Morlet wavelet, with themselves being complex sinusoids enveloped by a
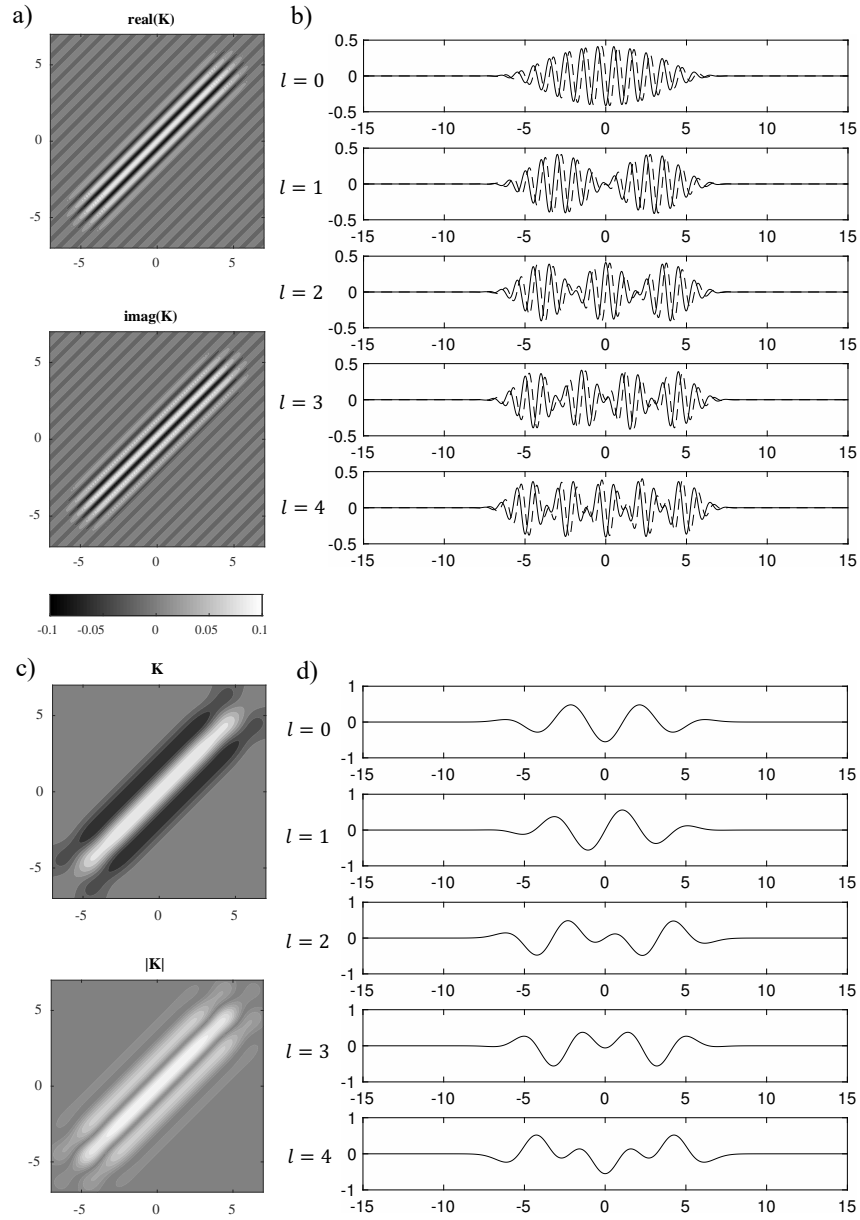
8

Figure 2: The kernel and first five eigen-wavelets for the Morlet wavelet (panels a,b) and Mexican hat wavelet (panels c,d) subject to a rectangular smoothing window of width $\kappa = 10$. In panel b) the solid and dashed line represent the real and imaginary components respectively.

taper. Thus, performing a continuous wavelet transform with one of the eigen-wavelets is equivalent to a time localized tapered Fourier transform evaluated at frequency $1/a$, and the temporally smoothed wavelet periodogram as represented in (5) is equivalent to a time localized multitaper spectral estimator. As means of a comparison, the kernel and associated eigen-wavelets of the Mexican hat wavelet using a rectangular smoothing function are shown in Fig. 2c and Fig. 2d, respectively.

4 Statistical Properties under Stationarity

4.1 Preliminaries

Let us define the kth order cumulant q of the differential process as

$$q_{i_1,...,i_k}(t_1, ..., t_k)\mathrm{d}t_1 \cdots \mathrm{d}t_k \equiv \mathrm{cum}\{\mathrm{d}N_{i_1}(u_1), ..., \mathrm{d}N_{i_k}(u_k)\} \,.$$

The following mixing condition (Assumption 2.2 in Brillinger, 1972) is sufficient for the asymptotic results that follow. It ensures that dependency structure in the point process decays at a sufficient rate for central limit arguments to be invoked.

Assumption 3. *The p-dimensional point process $N(t)$ is strictly stationary, i.e. $q_{i_1,...,i_k}(t_1 + t, ..., t_k + t) = q_{i_1,...,i_k}(t_1, ..., t_k)$, and we set $r_{i_1,...,i_k}(u_1, ..., u_{k-1}) = q_{i_1,...,i_k}(u_1, ..., u_{k-1}, 0)$. Furthermore, all moments exist, the cumulant function satisfies*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |r_{i_1,...,i_k}(u_1, ..., u_{k-1})|\mathrm{d}u_1 \cdots \mathrm{d}u_{k-1} < \infty,$$

for $i_1, ..., i_k = 1, ..., p$ and $k = 2, 3, ...,$ and

$$\int_{-\infty}^{\infty} |u||r_{i_1,i_2}(u)|\mathrm{d}u < \infty,$$

for $i_1, i_2 = 1, ..., p$.

The distributional results that follow differ slightly depending on whether a real valued wavelet (e.g. Mexican hat) or complex valued wavelet (e.g. Morlet) is chosen. We present

the results for a complex valued wavelet and relegate the derivation for a real valued wavelet to the Supplementary Material. For wavelet $\psi(t)$ with Fourier transform $\Psi(f)$, its central frequency is defined as the first moment of $|\Psi(f)|^2$, namely $f_0 := \int_0^\infty f|\Psi(f)|^2\mathrm{d}f$ (Cohen & Walden, 2010a). The central frequency of $\psi_{a,b}$ is therefore f_0/a and can be interpreted as the central analysing frequency of the wavelet at scale a. For example, the Morlet wavelet has a central frequency of $f_0 = 1$ and the Mexican hat wavelet has a central frequency of (approx.) $f_0 = 0.21$. We further define the frequency concentration of $\psi(t)$ to be the second central moment of $|\Psi(f)|^2$, namely $\sigma_\psi = \|(f - f_0)\Psi(f)\|_2$. It immediately follows from Proposition 2 that the central frequency and frequency concentration of the eigen-wavelet system is f_0 and σ_ψ, respectively.

Assumption 4. *Wavelet $\psi(t)$ is complex valued, satisfies Assumption 1, has approximating support $(-\alpha/2, \alpha/2)$ for some finite $\alpha > 0$, and $\sigma_\psi < \infty$. Furthermore, there exists a finite C such that $\int |\psi(t + u) - \psi(t)|\mathrm{d}t < C|u|$ for all real u, and it is orthogonal to its complex conjugate, i.e. $\int_{-\infty}^\infty \psi(t)\psi(t)\mathrm{d}t = 0$.*

The Morlet wavelet is an example of a complex valued wavelet that satisfies Assumption 4.

Assumption 5. *Smoothing function $h(t)$ satisfies Assumption 2 and furthermore there exists a finite C' such that $\int |h(t + u) - h(t)|\mathrm{d}t < C'|u|$.*

4.2   Asymptotic distributional results

We allow the wavelet to scale with T by defining $\psi^{(T)}(t) = \{(\alpha + \kappa)/T\}^{-1/2}\psi\{t(\alpha + \kappa)/T\}$, and appropriately normalize the scale and translation parameters as $\tilde{a} = a(\alpha + \kappa)/T$ and $\tilde{b} = b/T$, respectively. Under this rescaling (2) becomes

$$w(a, b) = w^{(T)}(\tilde{a}, \tilde{b}) = \tilde{a}^{-1/2} \int_0^T \psi^{*(T)}\{(t - \tilde{b}T)/\tilde{a}\}\mathrm{d}N(t),$$

and the normalized temporally smoothed wavelet periodogram is defined as

$$\Omega^{(T)}(\tilde{a}, \tilde{b}) = \int_{-\infty}^\infty h_{\kappa\tilde{a}}^{(T)}(u)W^{(T)}(\tilde{a}, u)\mathrm{d}u,$$

11

where $h^{(T)}(t) = T^{-1}h(t/T)$. For any T, the valid region of analysis is normalized to $\tilde{\mathcal{T}}_{\alpha,\kappa}$, an isosceles triangle with vertices $(0,0)$, $(0,1)$ and $(1,1/2)$ whose interior contains all valid pairs of $(\tilde{a}, \tilde{b})$. Asymptotic results are presented for any fixed point $(\tilde{a}, \tilde{b}) \in \tilde{\mathcal{T}}_{\alpha,\kappa}$ as $T \to \infty$. In doing so we define the frequency $f_{\tilde{a}} = f_0^{(T)}/\tilde{a}$, where $f_0^{(T)} := \int_0^\infty f|\Psi^{(T)}(f)|^2 \mathrm{d}f$ with $\Psi^{(T)}(f)$ denoting the Fourier transform of $\psi^{(T)}(t)$.

Proposition 3. *Let $N(t)$ be a p-dimensional stationary process with spectral density matrix $S(f)$. Let $\psi(t)$ be a wavelet satisfying Assumption 1 and let $h(t)$ be a smoothing function satisfying Assumption 2. For all $\kappa > 0$ and for all $(\tilde{a}, \tilde{b}) \in \tilde{T}_{\alpha,\kappa}$,*

$$E\{\Omega^{(T)}(\tilde{a}, \tilde{b})\} = E\{W^{(T)}(\tilde{a}, \tilde{b})\} = \int_{-\infty}^{\infty} \tilde{a}|\Psi^{(T)}(\tilde{a}f)|^2 S(f) \mathrm{d}f$$

and $E\{\Omega^{(T)}(\tilde{a}, \tilde{b})\} = S(f_{\tilde{a}}) + O(T^{-2})$ as $T \to \infty$.

In the following theorem, $\mathcal{N}_p^{\mathcal{C}}(\mu, \Sigma)$ denotes the circular p-dimensional complex normal distribution with mean μ and covariance matrix Σ.

Theorem 1. *Let $N(t)$ be a p-dimensional stationary process satisfying Assumption 3 with spectral density matrix $S(f)$, and let $\psi(t)$ be a wavelet with central frequency f_0 satisfying Assumption 4. The continuous wavelet transform $w^{(T)}(\tilde{a}, \tilde{b})$ is asymptotically $\mathcal{N}_p^{\mathcal{C}}\{0, S(f_{\tilde{a}})\}$ as $T \to \infty$, for all $(\tilde{a}, \tilde{b}) \in \tilde{\mathcal{T}}_{\alpha,\kappa}$.*

Let $\mathcal{W}_p^{\mathcal{C}}(n, \Sigma)$ denote the p-dimensional complex Wishart distribution with n degrees of freedom and centrality matrix Σ.

Theorem 2. *Let $N(t)$ be a p-dimensional stationary process satisfying Assumption 3 with spectral density matrix $S(f)$. Let $\psi(t)$ be a wavelet with central frequency f_0 satisfying Assumption 4, let $h(t)$ be a smoothing function satisfying Assumption 5, and for $\kappa > 0$ let $\{\eta_l; \ l = 0,1,...\}$ be the eigenvalues of the kernel $K(s,t)$ defined in (4). The temporally smoothed wavelet periodogram $\Omega^{(T)}(\tilde{a}, \tilde{b})$ is asymptotically $(1/n)\mathcal{W}_p^{\mathcal{C}}\{n, S(f_{\tilde{a}})\}$ as $T \to \infty$ for all $(\tilde{a}, \tilde{b}) \in \tilde{\mathcal{T}}_{\alpha,\kappa}$, where $n = 1/(\sum_{l=1}^{\infty} \eta_l^2)$.*

The following distributional result for the wavelet coherence is now immediate from Theorem 2 and Goodman (1963). We let $_2F_1(\alpha_1, \alpha_2; \beta_1; z)$ denote the hypergeometric function with 2 and 1 parameters α_1, α_2 and β_1 and scalar argument z.

Corollary 1. *Under the conditions of Theorem 2, the temporally smoothed wavelet coherence* $\gamma_{ij}^2(\tilde{a}, \tilde{b}) = |\Omega_{ij}^{(T)}(\tilde{a}, \tilde{b})|^2 / \{\Omega_{ii}^{(T)}(\tilde{a}, \tilde{b}) \Omega_{jj}^{(T)}(\tilde{a}, \tilde{b})\}$ *between component processes* $N_i(t)$ *and* $N_j(t)$ *($i \neq j$) asymptotically has density function*

$$g_{\gamma^2}(x) = (n-1)(1-\rho^2)^n (1-x)^{n-2} \, {}_2F_1(n, n; 1; \rho^2 x),$$

where ρ^2 *is shorthand for* $\rho_{ij}^2(f_{\tilde{a}})$*, the spectral coherence between* $N_i(t)$ *and* $N_j(t)$ *at frequency* $f_{\tilde{a}}$.

In the case of the rectangular smoothing function given in (6), the effective degrees of freedom n scale linearly with κ according to the following proposition.

Proposition 4. *Let* $\psi(t)$ *satisfy Assumption 4, let* $h(t)$ *be the rectangular smoothing function given in (6), and for* $\kappa > 0$ *let corresponding kernel* $K(s,t)$ *have ordered eigenvalues* $\{\eta_l; \; l = 0, 1, ...\}$. *Provided* $\kappa > \alpha$*, then* $n = (\sum_{l=0}^{\infty} \eta_l^2)^{-1} = \kappa \{\int_{-\infty}^{\infty} |\mathcal{P}(x)|^2 dx\}^{-1}$*, where* $\mathcal{P}(x) \equiv \int_{-\infty}^{\infty} \psi(t) \psi^*(t-x) dt$.

5 Test for stationarity

5.1 Formulation

Consider testing the null hypothesis H_0 that states $N(t)$ is a stationary process, against the alternative hypothesis H_A that states $N(t)$ is non-stationary. Under H_0 and from Proposition 3 it is true that $E\{\Omega(a,b)\}$ is constant in b. We therefore consider testing for stationarity at different scales. For convenience, we perform a dyadic partition of the time-scale space, performing a test at each scale in the set $\{\tilde{a}_j = 2^{-j}; j = 1, ..., J\}$. At scale $\tilde{a}_j$, we partition time into 2^j non-overlapping equal size segments, each centred at time points $\{\tilde{b}_{j,k} = (2k-1)/(2^{j+1}); k = 1, ..., 2^j\}$ and each the width of the approximate support of the wavelet at that scale. Our test at scale $\tilde{a}_j$ therefore becomes a test of the null hypothesis

$$H_j : E\{\Omega^{(T)}(\tilde{a}_j, \tilde{b}_1)\} = ... = E\{\Omega^{(T)}(\tilde{a}_j, \tilde{b}_{2^j})\} = \Omega_j,$$

13

where Ω_j is unspecified.

We construct a likelihood ratio test based on the asymptotic distribution of $\Omega^{(T)}(\tilde{a}, \tilde{b})$ stated in Theorem 1. To attain asymptotic results we require smoothing parameter κ and consequently degrees of freedom n, see Proposition 4, to grow with T. Furthermore, we permit the maximum resolution of analysis J to also grow.

Assumption 6. *Let $\kappa = O(T^c)$ and $2^J = o(T^{1/2-c})$, for some $c \in (0, 1/2)$.*

Proposition 5. *Let $B_1, ..., B_K$ be independent samples where $B_i \sim (1/n)\mathcal{W}_p^{\mathcal{C}}(n, \Sigma_i)$ ($i = 1, ..., K$). The likelihood ratio test statistic for the null hypothesis $H : \Sigma_1 = ... = \Sigma_K = \Sigma$, with unspecified Σ, is*

$$U(K) = K^{pKn} \frac{\prod_{i=1}^{K} \det(B_i)^n}{\det \left(\sum_{i=1}^{K} B_i \right)^{Kn}}.$$

Furthermore, when H is true, $-2\log(U(K))$ is asymptotically χ_f^2 as $n \to \infty$, where $f = (K-1)p^2$.

In the following proposition, we let $\tilde{U}_j$ be a random variable that is equal in distribution to $U(2^j)$ under the null hypothesis.

Proposition 6. *Let Assumptions 3, 4, 5 and 6 be true, and for a fixed $j \in \{1, ..., J\}$ define the test statistic for H_j as*

$$V_j = K^{pKn} \frac{\prod_{i=1}^{K} \det\{\Omega^{(T)}(\tilde{a}_j, \tilde{b}_i)\}^n}{\det \left\{ \sum_{i=1}^{K} \Omega^{(T)}(\tilde{a}_j, \tilde{b}_i) \right\}^{Kn}},$$

where $K = 2^j$ and n is as given in Theorem 2. Under H_j, $V_j \stackrel{d}{=} \tilde{U}_j + o(1)$ as $T \to \infty$.

Theorem 3. *Let Assumptions 3, 4, 5 and 6 be true. For a fixed $j \in \{1, ..., J\}$, under H_j, $\mathrm{pr}\{-2\log(V_j) \leq x\} = \mathrm{pr}(\chi_{\nu_j}^2 \leq x) + o(1)$ as $T \to \infty$, where $\nu_j = (2^j - 1)p^2$.*

With the following assumption, we can now combine the test statistics for $H_1, ..., H_J$, to form $V \equiv \prod_{j=1}^{J} V_j$ with which to test H_0.

Assumption 7. *Wavelet $\psi(t)$ satisfies $\int_{-\infty}^{\infty} \psi_{j,k}^{(T)}(t)\psi_{l,m}^{*(T)}(t)\mathrm{d}t = 0$ for all $(j,k) \neq (l,m)$ where $\psi_{j,k}^{(T)}(t)$ denotes the wavelet at the jth scale and kth translation ($j = 1, ..., J$; $k = 1, ..., K$).*

Note, this is only an approximation for the Morlet wavelet.

Corollary 2. *Let Assumptions 3, 4, 5, 6 and 7 be true. Under H_0, $\mathrm{pr}\{-2\log(V) \leq x\} = \mathrm{pr}(\chi^2_\nu \leq x) + o(1)$ as $T \to \infty$, where $\nu = \sum_{j=1}^J \nu_j = p^2(2^{J+1} - 2 - J)$.*

5.2 Simulations

The following simulations all use a Morlet wavelet and tests are conducted on H_1, H_2, H_3 and combined hypothesis H_0 at the $\alpha = 0.05$ level.

To demonstrate the test asymptotically attains the nominal level, we consider a pair of independent homogeneous Poisson processes each with intensity $\lambda = 1$ and increase T. We let smoothing parameter κ grow as $\kappa = 10\ T^{1/4}$. The type I error rate of the tests as a function of T is shown in Fig. 3(a). The effective support of a smoothed wavelet periodogram at scale j is $2^{-j}T$, hence, convergence is fastest at $j = 1$.

To study power with respect to first-order non-stationarity, we consider a pair of independent inhomogeneous Poisson processes. The first has a linearly increasing intensity function $\lambda_1(t) = \tan(\theta)t + 1 - (T/2)\tan(\theta)$, the second is linearly decreasing with intensity function $\lambda_2(t) = \tan(-\theta)t + 1 - (T/2)\tan(-\theta)$. We vary θ, the slope of the intensity functions, between $-\arctan(2/T)$ and $\arctan(2/T)$; the process is stationary if and only if $\theta = 0$. This model ensures that the expected number of events in $(0, T]$ is equal for all θ. We set $T = 1024$ and again have $\kappa = 10\ T^{1/4}$, resulting in 22 degrees of freedom. Power as a function of θ is shown in Fig. 3(b). As expected, power increases with $|\theta|$, noting the nominal level is attained (for small j) at $\theta = 0$.

To study power with respect to second-order non-stationarity, we consider a bivariate piecewise stationary Hawkes process, see Appendix 2, made up of three segments. The first and third segments are the time intervals $(0, 256]$ and $(768, 1024]$, respectively. On these exist a bivariate process with baseline intensities $\nu_1 = \nu_2 = 1$ and an exponential excitation kernel with excitation parameter $\alpha_{11} = \alpha_{22} = 1$ and decay parameter $\beta_{11} = \beta_{22} = 2$; there is no mutual (cross) excitation between the processes, i.e. $\alpha_{12} = \alpha_{21} = 0$. On the second segment at time interval $(256, 768]$ exists a mutually exciting Hawkes process with identical

15

self-excitation parameters to segments one and three, and now with mutual excitation between the individual processes. The decay parameters are set at $\beta_{12} = \beta_{21} = 2$, and power is explored as α_{12} and α_{21} are increased together to introduce increasing levels of mutual excitation. This in turn gives an increasing deviation away from the distribution of the process in the first and third segments. The baseline intensity on $(256, 768]$ is decreased accordingly to ensure the expected number of events is constant across the entirety of the time interval $(0, 1024]$. Fig. 3(c) shows power increases with the value of α_{12} in the second segment, i.e. as the distribution of the process in the second segment deviates further from that of the first and third segments.

The final study we consider is a pair of independent locally-stationary Hawkes processes as developed in Roueff et al. (2016). We let each process have a baseline intensity of $\nu_1 = \nu_2 = 0.5$ and a Gamma excitation kernel of the form

$$g\{u; \delta(t/T)\} = \{u - \delta(t/T)\} \frac{\mathrm{e}^{-\{u - \delta(t/T)\}}}{2} \mathbb{1}_{\{u > \delta(t/T)\}}, \quad 0 \le u < \infty . \tag{7}$$

This excitation kernel only allows an offspring event to be generated after a delay of $\delta(\cdot)$, which varies as a function of time to introduce non-stationary second order structure. In this example we set $\delta(s) = 1 + \delta_0|1 - 2s|$; when $\delta_0 = 0$ the process is stationary and we can introduce time-varying second order behaviour of increasing magnitude by increasing δ_0. Power as a function of δ_0 is shown in Fig. 3(d). As expected, power increases with δ_0 and the nominal level is attained at $\delta_0 = 0$.

5.3   Practical implementation

In practice, one must make a choice of κ and J that is both suitable for the dataset under analysis and the time-scales on which one wishes to detect deviations from stationarity, while accounting for traditional trade-offs. Assumption 6 dictates the growth rates of κ and J must be balanced such that $2^J \kappa = o(T^{1/2})$. If analysing variation on small time-scales, one benefits from lowering κ and increasing J, at the consequence of asymptotic limits being less appropriate. Conversely, asymptotic limits can be practically attained at the cost of loosing temporal resolution. Therefore, there does not exist a universal
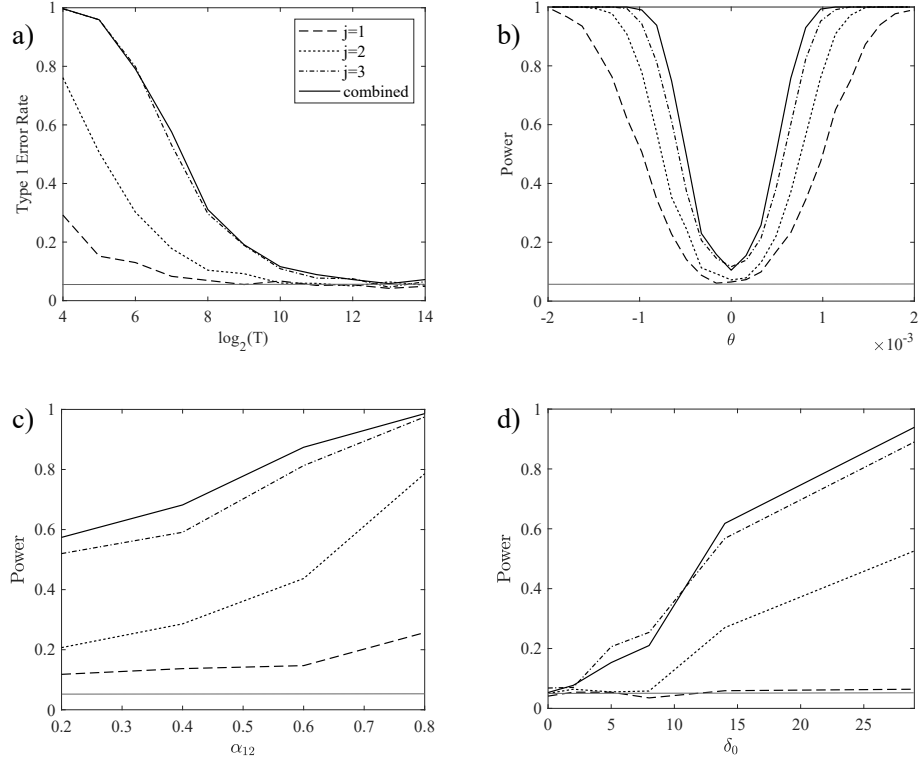
16

Figure 3: Results for simulation studies detailed in Section 5.2. Tests are performed at the $\alpha = 0.05$ level (indicated by the light grey line) on hypotheses H_j for $j = 1, 2, 3$, and combined hypothesis H_0. In all cases, 1000 realisations of the processes were used. (a) The type I error rate for a pair of independent homogeneous Poisson processes. (b) Power as a function of θ, the slope of the intensity function for the pair of independent inhomogeneous Poisson processes. (c) Power as a function of α_{12} (and α_{21}), the mutual excitation parameter in the second segment of the piecewise stationary bivariate Hawkes process. (d) Power as a function of δ_0, the magnitude of the excitation delay effect in the locally stationary Hawkes process.

heuristic for making this choice. If attaining the nominal level of the test is of importance, experiments with independent Poisson processes over a time length and intensity that is indicative of the data can be used select values of κ and J. Example results for experiments of this type are presented in Supplementary Fig. 2.

6 Application to neural signalling

To give an example of the methods in practice, we analyse signalling regions within the *lateral geniculate nucleus* of a mouse. Specifically, we consider a set of neurons examined in Tang et al. (2015), where the authors are primarily concerned with analysing firing properties in order to understand how visual signals are encoded and transferred throughout the brain. To demonstrate the ability of our smoothed coherence estimator and stationarity test to operate with a single trial we consider only a single firing sequence from the paper. In this case, the mouse is shown a visual stimulus in the form of an liquid crystal display screen showing a sinusoidal monochromatic drifting grating with spatial stimulus at a frequency of 0.04 cycles per second and temporal flicker of 1Hz. The firing pattern is 7 seconds in length and represents data for cells 108 and 117; these cells were picked for the example as they demonstrate relatively high firing rates. We use the Morlet wavelet with temporal smoothing parameter $\kappa = 10$ and approximating support parameter $\alpha = 8$. For completeness, this example was performed using exact kernel sampling, however, an approximate computation based on the Nystrom method, see Appendix 1, provides visually indistinguishable results and p-values.

The analysis of the experimental data is provided in Fig. 4. Tests for stationarity were performed at scale levels $j = 1$ and 2, informed by the analysis of Supplementary Material Section 3, with dyadic sampling points as marked by the crosses. With a p-value of 0.032, there is strong evidence that the process demonstrates non-stationary behaviour at the coarsest scale $j = 1$, corresponding to a scale of 0.25s and frequency of 4Hz through the $f = 1/a$ relationship for a Morlet wavelet. However, there is not enough evidence to reject the wavelet spectrum stays constant at the finer scale $j = 2$. With the parameters specified,
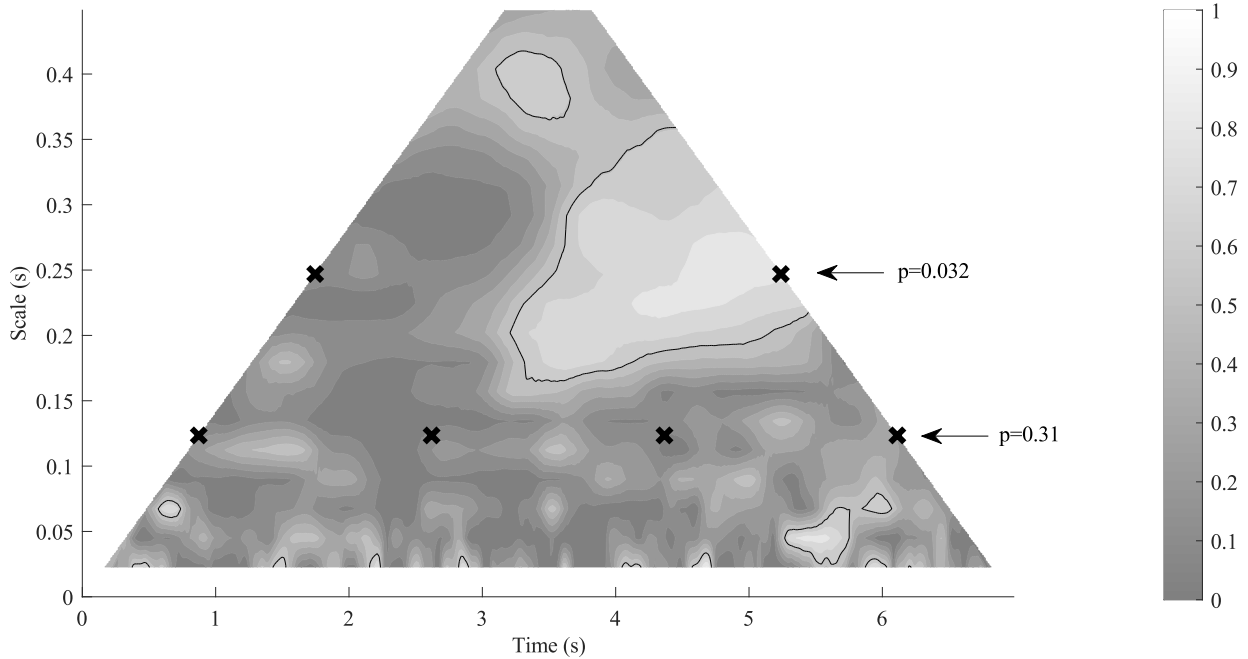
Figure 4: Wavelet coherence estimated from the mouse neuron firing data with the Morlet wavelet. Black lines represent contours drawn at the 95th percentile of the distribution for zero coherence (see Corollary 1). Crosses indicate dyadic sampling points for the stationarity test, with arrows depicting the scale specific p-values.

the 95th percentile of the distribution for zero coherence is 0.593 and is represented by the black contour line. This indicates that the non-stationarity at $j = 1$ involves a change in the correlation between the two data streams half way through the experiment, with significant coherent signalling becoming present in the latter half.

Acknowledgement

Supplementary material

Supplementary Material Section 1 contains the proofs to propositions, theorems and corollary presented here. Supplementary Material Section 2 provides the results for real valued wavelets. Supplementary Material Section 3 provides a guide to practical implementation. Supplementary Material Section 4 contains a link to a MATLAB package for implementing the presented methods.

Appendix 1

Computing eigen-wavelets and eigenvalues

The Nystrom method (Chapter 1, Kythe & Puri, 2001) is an efficient method for computing the eigenfunctions of kernel $K(s,t)$ for the multiwavelet representation described in Section 3. We can approximate the integral using the quadrature rule to solve the approximate eigen-problem $\sum_{j=1}^{n} w_j K(s, t_j) \tilde{\varphi}_l(t_j) = \tilde{\eta} \tilde{\varphi}_l(s)$ for a discrete set of values for s. The quadrature points $\{t_1, ..., t_n\}$ (n large) are regularly spaced across $(-(\alpha+\kappa)/2, (\alpha+\kappa)/2)$ and the weights are set to be $w_j = (\alpha + \kappa)/n$. For simplicity, the Nystrom points $\{s_1, ..., s_n\}$ are set to equal $\{t_1, ..., t_n\}$. In matrix form, the eigen-problem now becomes

$$KW\tilde{\varphi} = \tilde{\eta}\tilde{\varphi} ,$$

where K is the $\mathbb{R}^{n \times n}$ matrix $(K(s_i, t_j))$, $\tilde{\varphi} \equiv [\tilde{\varphi}(t_1), \ldots, \tilde{\varphi}(t_n)]^{\mathrm{T}}$, and $W \equiv \mathrm{diag}(w_1, \ldots, w_n)$. Solving the above gives approximations to the first n eigenvalues and eigen-wavelets of kernel $K(s,t)$.

Should it be required, the Nystrom extension of the sampled vector $\tilde{\varphi}_l = [\tilde{\varphi}(s_1), \ldots, \tilde{\varphi}(s_n)]$ is the function

$$\tilde{\varphi}_l(x) = \tilde{\lambda}_l \sum_{j=1}^{n} w_j K(x, s_j) \tilde{\varphi}_l(s_j).$$

The sum in (5) is over an infinite set of (eigen-)wavelet periodograms. However, in practice,

the size of the eigenvalues drop away rapidly indicating that the kernel can be accurately reconstructed using only a small number of its eigen-wavelets, hence (5) can be approximated with only a small number of terms. For example, in the case of the $\kappa = 10$, the first nine eigenvalues contain 99.9% (3.s.f.) of the overall energy.

Appendix 2

Hawkes process

A stationary bivariate Hawkes process (Hawkes, 1971) contains both inter and cross dependencies, and is defined through its stochastic intensity function

$$\Lambda = \nu + \int_{-\infty}^{t} G(t - s)\mathrm{d}N(s),$$

where $\nu \equiv (\nu_1, \nu_2)^T$ is the baseline intensity vector and $G(\cdot)$ is the matrix valued excitation kernel. The diagonal elements of $G(\cdot)$ characterise the self-exciting behaviour of each individual process, and the off-diagonal elements characterise the mutually exciting behaviour. The exponential excitation kernel used in Section 5.2 is

$$G(s) = \left\{ \begin{array}{cc} \alpha_{11} \exp(-\beta_{11}s) & \alpha_{12} \exp(-\beta_{12}s) \\ \alpha_{12} \exp(-\beta_{21}s) & \alpha_{22} \exp(-\beta_{22}s) \end{array} \right\}.$$

For simulation we use the thinning algorithm of Ogata (1981).

References

BARTLETT, M. S. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B* **25**, 264–296.

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics* , 483–513.

BRILLINGER, D. R. (1996). Some uses of cumulants in wavelet analysis. *Journal of Nonparametric Statistics* **6**, 93–114.

CARTER, G. (1987). Coherence and time delay estimation. *Proceedings of the IEEE* **75**, 236–255.

COHEN, E. A. K. & WALDEN, A. T. (2010a). A statistical analysis of Morse wavelet coherence. *IEEE Transactions on Signal Processing* **58**, 980–989.

COHEN, E. A. K. & WALDEN, A. T. (2010b). A statistical study of temporally smoothed wavelet coherence. *IEEE Transactions on Signal Processing* **58**, 2964–2973.

GOODMAN, N. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Annals of Mathematical Statistics* **34**, 152–177.

GRINSTED, A, J., MOORE, C. & JEVREJEVA, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics* **11**, 561–566.

HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.

KYTHE, P. K. & PURI, P. (2001). *Computational Methods for Linear Integral Equations.* Springer.

MALLAT, S. & PEYRÉ, G. (2008). *A Wavelet Tour of Signal Processing.* Elsevier Science and Technology, 3rd ed.

MERCER, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **209**, 415–446.

NASON, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **75**, 879–904.

NASON, G. P., VON SACHS, R. & KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society, Series B* **62**, 1–28.

OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* **27**, 23–31.

OLHEDE, S. C. & WALDEN, A. T. (2002). Generalized Morse wavelets. *IEEE Transactions on Signal Processing* **50**, 2661–2670.

PREUSS, P., VETTER, M. & DETTE, H. (2013). Testing semiparametric hypotheses in locally stationary processes. *Scandinavian Journal of Statistics* **25**, 417–437.

ROUEFF, F. & VON SACHS, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli* **25**, 1355–1385.

ROUEFF, F., VON SACHS, R. & SANSONNET, L. (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications* **126**, 1710–1743.

TANG, S., ARDILA JIMENEZ, S. C., CHAKRABORTY, S. & SCHULTZ, S. R. (2015). Visual receptive field properties of neurons in the mouse lateral geniculate nucleus. *Plos One* **11**, 1–34.

THOMSON, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* **70**.

TORRENCE, C. & WEBSTER, P. (1999). Interdecadal changes in the ESNO-monsoon system. *Journal of Climate* **12**, 2679–2690.

VON SACHS, R. & NEUMANN, M. H. (2000). A wavelet-based test for stationarity. *Journal of Time Series Analysis* **21**, 597–613.

WALDEN, A. T. (2000). A unified view of multitaper multivariate spectral estimation. *Biometrika* **87**, 767–788.

WELCH, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio Electroacoustics* **15**, 70–73.

Correspondence

Correspondence should be addressed to

Edward Cohen
Department of Mathematics
Imperial College London
London SW7 2AZ.

Email: e.cohen@imperial.ac.uk