Article:
Palasciano, Henry A., Knight, Marina Iuliana and Nason, Guy P. (2025) Continuous Time Locally Stationary Wavelet Processes. Biometrika. asaf015. ISSN: 0006-3444

https://doi.org/10.1093/biomet/asaf015

Biometrika (2025), **112**, 2, asaf015 https://doi.org/10.1093/biomet/asaf015
Advance Access publication 13 March 2025

Continuous-time locally stationary wavelet processes

By H. A. PALASCIANO

Department of Mathematics, Imperial College London,
180 Queen's Gate, South Kensington, London SW7 2AZ, U.K.
henry.palasciano17@imperial.ac.uk

M. I. KNIGHT

Department of Mathematics, University of York,
Heslington, York YO10 5DD, U.K.
marina.knight@york.ac.uk

and G. P. NASON

Department of Mathematics, Imperial College London,
180 Queen's Gate, South Kensington, London SW7 2AZ, U.K.
g.nason@imperial.ac.uk

Summary

This article introduces the class of continuous-time locally stationary wavelet processes. Continuous-time models enable us to properly provide scale-based time series models for irregularly spaced observations for the first time, while also permitting a spectral representation of the process over a continuous range of scales. We derive results for both the theoretical setting, where we assume access to the entire process sample path, and a more practical one, which develops methods for estimating the quantities of interest from sampled time series. The latter estimates are accurately computable in reasonable time by solving the relevant linear integral equation using the iterative soft-thresholding algorithm of Daubechies et al. (2004). Appropriate smoothing techniques are also developed and applied in this new setting. Comparisons to previous methods are conducted on the heart rate time series of a sleeping infant. Additionally, we exemplify our new methods by computing spectral and autocovariance estimates on irregularly spaced heart rate data obtained from a recent sleep-state study.

Some key words: Continuous-time model; Irregular spacing; Local autocovariance; Nonstationary time series; Wavelet; Wavelet spectrum.

1. Introduction

It is often customary to assume that a time series is second-order stationary, since plenty of theoretically sound and well-tested methods are readily available for its analysis; see, for example, Brockwell & Davis (1991) or Percival & Walden (2020). It is well known that a

stationary stochastic process $X(t)$, $t \in \mathbb{R}$, has a spectral representation of the form

$$X(t) = \int_{-\infty}^{\infty} \exp(i\omega t)\, \mathrm{d}Z(\omega), \tag{1}$$

where $Z(\omega)$ is a square-integrable orthogonal increment process with $\mathbb{E}[|\mathrm{d}Z(\omega)|^2] = \mathrm{d}S^{(I)}(\omega)$ and $S^{(I)}(\omega)$ is the integrated spectrum of $X(t)$. The autocovariance function of $X(t)$, denoted $c(\tau)$, has a similar representation in the frequency domain,

$$c(\tau) = \int_{-\infty}^{\infty} \exp(i\omega\tau)\, \mathrm{d}S^{(I)}(\omega); \tag{2}$$

see, for example, Priestley (1983). However, in practice, many time series are unlikely to be stationary, such as is often the case in finance for instance (Mikosch & Stărică, 2004; Fryzlewicz, 2005). To extend these quantities to the nonstationary case, Priestley (1983) suggested introducing a time-varying amplitude function $A(t, \omega)$ in (1), giving

$$X(t) = \int_{-\infty}^{\infty} A(t, \omega) \exp(i\omega t)\, \mathrm{d}Z(\omega). \tag{3}$$

Models such as (3) form the basis for locally stationary time series, later developed by Dahlhaus (1997) in discrete time.

Nason et al. (2000) introduced the locally stationary wavelet processes, which replaced the set of functions $\{\exp(i\omega t)\}$ by a set of discrete nondecimated wavelets. As with locally stationary Fourier processes (Dahlhaus, 1997), the local stationarity is enforced via constraints on the amplitude functions. From locally stationary wavelet processes, one can then obtain a time-scale decomposition of the spectral density function, known as the evolutionary wavelet spectrum, and a wavelet representation of the local autocovariance similar to (2).

However, discrete-time locally stationary wavelet processes are constructed under the assumption that the time series data are obtained at a constant sampling rate, which is not always the case as, in practice, observations can either be missing or be irregularly spaced. Attempts to provide a suitable alternative for missing data were made by Knight et al. (2012) via the use of the nondecimated lifting transform from Knight & Nason (2009). Unfortunately, (i) the Knight et al. (2012) estimates only resemble the true spectrum defined in Nason et al. (2000) in that they can only estimate an underlying spectrum at the n time-points in the irregular sample and nowhere in-between; (ii) the lifting methods of Knight et al. (2012) bestow no complete analytical formula connection, nor asymptotic theory, between the expected periodogram and the original spectrum, at each time, such as can be found in equation (20) in Proposition 4 of Nason et al. (2000) or Theorem 3 below; (iii) the scales in Knight et al. (2012) are artificial and depend on the inter-point spacing of the samples and, hence, do not have a one-to-one correspondence to the actual wavelet scales underlying both the discrete and continuous locally stationary wavelet processes in Nason et al. (2000) and below, respectively. Ideally, spectral estimates should not depend on the sampling scheme. (iv) Knight et al. (2012) stated that their methods are 'highly computationally intensive' in that the periodogram evaluated at each time-point requires computation of a full lifting (second-generation wavelet) transform for multiple, P, runs corresponding to different lifting trajectories. For good estimation, typically P is large and was set to $P = 1000$

and $P = 5000$ in Knight et al. (2012). Hence, periodogram estimation in Knight et al. (2012) requires nP lifting wavelet transforms compared to our method that uses one, or certainly at least $O(1)$. Knight et al. (2012) did not specify spectral computational effort, but they remarked that it is considerably more computationally intensive than their periodogram computation.

Instead, we propose the use of genuine continuous-time locally stationary wavelet processes, which are natural counterparts to the discrete-time versions from Nason et al. (2000). To the best of our knowledge, our new methods are the only ones that can accurately estimate wavelet spectral properties of an irregularly sampled process within a reasonable computational time frame. Our method estimates the spectrum itself, not a simulacrum as in Knight et al. (2012), and is among the few guaranteeing the nonnegativity of the spectrum (the other we know of was for regularly spaced data in chapter 4 of P. Fryzlewicz's 2003 PhD thesis from the University of Bristol).

Aside from the practical advantages that a continuous-time model provides, our results are also interesting from a more theoretical perspective. In fact, many discrete-time models and methods also possess a continuous-time counterpart; see, for example, Brockwell et al. (1991, 2007), Brockwell (2001), Lucchese et al. (2023) and Calvello et al. (2024). Recently, Bitter et al. (2023) adapted locally stationary Fourier processes from the discrete- (Dahlhaus, 1997, 2012) to the continuous-time setting. Since, so far, the theory of locally stationary wavelet processes exists in discrete time only, our objective is to introduce parallel continuous-time models, study their theoretical properties and develop a practical means for their estimation and analysis. In doing so, we obtain a representation of the evolutionary wavelet spectrum across a continuous range of scales, providing a complete decomposition of the underlying frequencies of the process.

2. Background and notation

We briefly review the relevant background material for our continuous-time locally stationary wavelet processes. The Supplementary Material contains a more detailed discussion.

Discrete-time locally stationary wavelet processes (Nason et al., 2000; Nason, 2008) are time series models that have found applications in various domains, including finance (Fryzlewicz, 2005), economics (Winkelmann, 2016), aliasing detection (Eckley & Nason, 2018; Palasciano & Nason, 2023), biology (Hargreaves et al., 2019), gravitational-wave detection (Romano & Cornish, 2017) and energy (Nowotarski et al., 2013; de Menezes et al., 2016), to name but a few. They are defined as a sequence of doubly indexed stochastic processes having the representation

$$X_{t,T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k;T} \psi_{j,k-t} \xi_{j,k}$$

in the mean square sense, where $\{w_{j,k;T}\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ is a set of amplitudes, $\{\psi_{j,k-t}\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ are discrete wavelets and $\{\xi_{j,k}\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ is a collection of uncorrelated random variables with zero mean and unit variance. Conditions are imposed on the amplitude functions to prevent them from varying too quickly, thus preserving the local stationarity of the process. More importantly, one assumes that, as $T \to \infty$, $w_{j,k;T}$ converges to $W_j(k/T)$, a rescaled-time Lipschitz continuous amplitude function defined on $z = k/T \in (0, 1)$. The idea of rescaled time was first introduced by Dahlhaus (1997) and is frequently used when working with

locally stationary processes. The corresponding evolutionary wavelet spectrum, autocorrelation wavelet and local autocovariance of the locally stationary wavelet process, for $j \in \mathbb{N}$ and $\tau \in \mathbb{Z}$, are

$$S_j(z) = |W_j(z)|^2, \qquad \Psi_j(\tau) = \sum_k \psi_{j,k}\psi_{j,k-\tau} \quad \text{and} \quad c(z,\tau) = \sum_{j=1}^{\infty} S_j(z)\Psi_j(\tau),$$

respectively. To estimate the evolutionary wavelet spectrum, Nason et al. (2000) first estimated the raw wavelet periodogram using the nondecimated wavelet transform and then applied a correction with the inner product operator $A_{j,\ell} = \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau)$, $j, \ell \in \mathbb{N}$. See Nason et al. (2000) for further details. Section 3 below defines the continuous-time counterparts of all these quantities and §4 develops methods to estimate these in practice.

Lévy processes and bases play a crucial role in the definition of locally stationary wavelet processes in continuous time. The following overview draws heavily from Walsh (1986), Rajput & Rosinski (1989), Applebaum (2009) and Barndorff-Nielsen et al. (2018). We take $(\Omega, \mathcal{F}, \mathbb{P})$ to be the underlying probability space, $\{S, \mathcal{B}(S)\}$ to be a Borel space and denote the bounded Borel sets of $\mathcal{B}(S)$ by $\mathcal{B}_b(S)$. A real-valued Lévy basis L on S is a collection $\{L(A)\colon A \in \mathcal{B}_b(S)\}$ of random variables such that the law of $L(A)$ is infinitely divisible for all $A \in \mathcal{B}_b(S)$ and, if A and B are disjoint subsets in $\mathcal{B}_b(S)$, $L(A)$ and $L(B)$ are independent and $L(A \cup B) = L(A) + L(B)$ almost surely. We restrict our attention to homogeneous Lévy bases as defined by Barndorff-Nielsen et al. (2018), since this is synonymous with stationarity. We say that a homogeneous Lévy basis G is Gaussian if $G(A) \sim \mathcal{N}\{\mu\,\mathrm{Leb}(A), \sigma^2\mathrm{Leb}(A)\}$ for all $A \in \mathcal{B}(\mathbb{R}^n)$ and some $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$, where Leb denotes the Lebesgue measure. Homogeneous Gaussian Lévy bases are equivalent to the concept of white noise defined by Walsh (1986). Furthermore, if L is a homogeneous Lévy basis on $\{\mathbb{R}, \mathcal{B}(\mathbb{R})\}$ then the process $(L_t)_{t\geqslant 0}$ with $L_t = L([0,t])$ is a Lévy process. Conversely, given a Lévy process $(L_t)_{t\geqslant 0}$, we can obtain a Lévy basis L by setting $L((s,t]) = L_t - L_s$.

The iterative soft-thresholding algorithm of Daubechies (1992) is a powerful regularization method that we find useful in §4 below for estimating the continuous-time evolutionary wavelet spectrum from an estimate of the raw wavelet periodogram. This algorithm was developed to obtain regularized solutions to linear inverse problems $Th = f$, where $f = g + \varepsilon$ is some noisy observation of the true signal g. In our setting, T is a linear integral operator, f, g and h are some functions in a Hilbert space $\mathcal{H}$ and ε is the noise. In a purely theoretical setting, Mercer's theorem (Mercer, 1909) or methods from Kress (1999) can be used to obtain a solution. Instead, Daubechies (1992) suggested minimizing the functional

$$\Phi_{\mu,p}(h) = ||Th - f||^2 + \sum_{\gamma\in\Gamma} \mu_\gamma\,|\langle h, e_\gamma\rangle|^p \tag{4}$$

for $1 \leqslant p \leqslant 2$, where $\{e_\gamma\}_{\gamma\in\Gamma}$ is an orthogonal basis of $\mathcal{H}$ and $\mu = (\mu_\gamma)_{\gamma\in\Gamma}$ is a vector of weights, which act as regularization parameters. In our setting, we are interested in penalizing positive and negative coefficients asymmetrically for the case $p = 1$ and so replace each $\mu_\gamma \in \mu$ with the positive weights $(\mu_\gamma^+, \mu_\gamma^-)$. The solution is then given by the iterative soft-thresholding algorithm

$$h^n = \mathcal{S}_\mu\{h^{n-1} + T^\dagger(f - Th^{n-1})\},$$

where

$$\mathcal{S}_{\mu}(h) = \sum_{\gamma \in \Gamma} \mathcal{S}_{\mu_\gamma^+, \mu_\gamma^-}(\langle h, e_\gamma \rangle) e_\gamma, \quad \mathcal{S}_{\mu_\gamma^+, \mu_\gamma^-}(x) = \begin{cases} x - \mu_\gamma^+/2 & \text{if } x \geqslant \mu_\gamma^+/2, \\ 0 & \text{if } -\mu_\gamma^-/2 < x < \mu_\gamma^+/2, \\ x + \mu_\gamma^-/2 & \text{if } x \leqslant -\mu_\gamma^-/2, \end{cases} \tag{5}$$

and $T^\dagger$ denotes the adjoint of T. As shown by Daubechies et al. (2004), this iterative scheme converges to the unique minimizer of (4) for arbitrarily chosen $h^0 \in \mathcal{H}$. More details can be found in Daubechies et al. (2004, 2016) or Bertero et al. (2021).

3. Locally stationary wavelet processes in continuous time: theory

3.1. *Process definition*

Our continuous-time locally stationary wavelet processes rely on continuous wavelets $\psi \in L^2(\mathbb{R})$ for which we assume that

$$\int_{-\infty}^{\infty} \psi(t)\, \mathrm{d}t = 0, \qquad ||\psi||^2 = \int_{-\infty}^{\infty} |\psi(t)|^2\, \mathrm{d}t = 1. \tag{6}$$

From $\psi \in L^2(\mathbb{R})$, an entire family of wavelets $\{\psi_{u,v}(t),\ u \in \mathbb{R}^+,\ v \in \mathbb{R}\}$ can be generated via

$$\psi_{u,v}(t) = u^{-1/2} \psi\{u^{-1}(t - v)\} = \psi(u, t - v), \tag{7}$$

preserving the properties in (6), for all $u \in \mathbb{R}^+$ and $v \in \mathbb{R}$. In (7) we have introduced the notation $\psi(\cdot, \cdot)$, making the dependence on scale more explicit, with $\psi(1, t) = \psi(t)$. For more information on the continuous wavelet transform, see Daubechies (1992).

In continuous time, we assume knowledge of the entire path rather than a finite set of observations. Consequently, it becomes irrelevant to define a sequence of stochastic processes corresponding to increasing amounts of data becoming available. Hence, continuous-time locally stationary wavelet processes are defined as follows.

Definition 1. *A continuous-time locally stationary wavelet process* $\{X(t)\}_{t \in \mathbb{R}}$ *is a stochastic process defined in the mean square sense as*

$$X(t) = \int_0^{\infty} \int_{-\infty}^{\infty} W(u, v) \psi(u, v - t) L(\mathrm{d}u, \mathrm{d}v), \tag{8}$$

where L *is a homogeneous square-integrable Lévy basis with* $E\{L(A)\} = 0$ *and* $E\{L(A)^2\} = \mathrm{Leb}(A)$ *for all* $A \in \mathcal{B}_b(\mathbb{R}^+ \times \mathbb{R})$, $\{\psi_{u,v}(x),\ u \in \mathbb{R}^+,\ v \in \mathbb{R}\}$ *is a continuous wavelet family and* $W \in L^2(\mathbb{R}^+, \mathbb{R})$ *is an amplitude function. Furthermore, the following conditions on* W *hold.*

(i) *For each fixed scale* $u \in \mathbb{R}^+$, $W(u, \cdot)$ *is a Lipschitz continuous function with Lipschitz constant* $K(u)$. *The Lipschitz constant function* $K(\cdot)$ *is bounded and satisfies*

$$\int_0^{\infty} uK(u)\, \mathrm{d}u < \infty. \tag{9}$$

(ii) *For each fixed* $v \in \mathbb{R}$, $W(\cdot, v)$ *satisfies*

$$\int_0^\infty |W(u, v)|^2 \, \mathrm{d}u < \infty.$$

Locally stationary time series are stochastic processes whose statistical properties change gradually over time. The use of wavelets enables an implicit local representation of the process, while the smoothness assumptions imposed on W ensure that locally the process behaves in an approximately stationary fashion. The condition on the Lipschitz constants in (9) guarantees that the rate at which the statistical properties are allowed to vary decreases as the scales widen. This ensures that the process behaves in a realistic fashion and is required for estimation.

For representation (8) to converge in the mean square sense, the amplitude functions cannot be allowed to grow arbitrarily with u. Hence, assumption (ii) in Definition 1 ensures that $|W(u, v)|^2 \to 0$ as $u \to \infty$.

The requirement that the Lévy basis is homogeneous is essential, as this implies that the basis is stationary, and vice versa; see the Supplementary Material or Barndorff-Nielsen et al. (2018). This ensures that the nonstationarity is controlled entirely by the amplitude functions W. Furthermore, since L is, by definition, independent when acting on disjoint sets, any correlation structure in the process is captured entirely by the wavelets themselves.

Remark 1. The amplitude function $W(u, v)$ is defined for all $v \in \mathbb{R}$ as opposed to its discrete counterpart, which is only defined on $(0, 1)$. Later, in § 4, when adapting this definition for practical applications, we again return to the concept of rescaled time.

Example 1. A Haar moving average process of order α, or Haar MA(α), is a continuous-time locally stationary wavelet process $X(t)$ with amplitude function $W(u, v) = 1$ for $u = \alpha \in \mathbb{R}^+$ and zero otherwise. The Lévy basis is $L(\mathrm{d}u, \mathrm{d}v) = \delta_\alpha(\mathrm{d}u)G(\mathrm{d}v)$, where δ_α is the Dirac measure centred at α and G is a homogeneous Gaussian Lévy basis on $\mathcal{B}_b(\mathbb{R})$. The underlying wavelet family $\{\psi_{u,v}\}$ is derived from the Haar wavelet (see the Supplementary Material) through $\psi(x) = \psi_{\mathrm{H}}(-x)$, with the reflection about the y axis chosen to maintain a backward-looking process. The Haar MA(α) can be written as

$$X(t) = \int_0^\infty \int_{-\infty}^\infty \psi_{\mathrm{H}}(u, v - t)\delta_\alpha(\mathrm{d}u)G(\mathrm{d}v) = \alpha^{-1/2}\{(B_t - B_{t-\alpha/2}) - (B_{t-\alpha/2} - B_{t-\alpha})\},$$

where B_t is a standard Brownian motion. Essentially, Haar moving average processes capture the difference between consecutive Brownian increments, each of width $\alpha/2$. See Nason et al. (2000) for the discrete-time analogue. There, for example, the Haar MA(1) and MA(2) processes are

$$X_t^1 = 2^{-1/2}(\varepsilon_t - \varepsilon_{t-1}), \qquad X_t^2 = 2^{-1}(\varepsilon_t + \varepsilon_{t-1} - \varepsilon_{t-2} - \varepsilon_{t-3}),$$

respectively, where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance.

3.2. *Continuous-time evolutionary wavelet spectrum*

The evolutionary wavelet spectrum gives a location-scale decomposition of the spectral power of a continuous-time locally stationary wavelet process, and is defined as follows.

DEFINITION 2. *The continuous-time evolutionary wavelet spectrum S of a locally stationary wavelet process $X(t)$, with respect to the wavelet family ψ, is defined by*

$$S(u, v) = |W(u, v)|^2$$

for all scales $u \in \mathbb{R}^+$ and times $v \in \mathbb{R}$.

Assumption (ii) in Definition 1 implies that, for each fixed $v \in \mathbb{R}$,

$$\int_0^\infty S(u, v)\, \mathrm{d}u < \infty.$$

Remark 2. More generally and technically, we can define

$$S^{(I)}(\mathrm{d}u, \mathrm{d}v) = E\{|W(u, v)L(\mathrm{d}u, \mathrm{d}v)|^2\},$$

where $S^{(I)}$ is the continuous-time integrated evolutionary wavelet spectrum. We can write this as

$$S^{(I)}(\mathrm{d}u, \mathrm{d}v) = S(u, v)\, \mathrm{d}u\, \mathrm{d}v,$$

since L is a homogeneous Lévy basis. We can then define

$$S_v^{(I)}(u) = \int_0^u S(x, v)\, \mathrm{d}x = \frac{\partial S^{(I)}}{\partial v}(u, v), \tag{10}$$

representing the contribution to the total power of the process in a small neighbourhood of v for scales less than or equal to u, mirroring the evolutionary power spectrum of Priestley (1983).

Example 1 *(continued)*. The evolutionary wavelet spectrum of a Haar MA(α) process is

$$S(u, v) = \begin{cases} 1 & \text{for } u = \alpha, \\ 0 & \text{for } u \neq \alpha. \end{cases}$$

3.3. *Continuous-time local autocovariance*

Nason et al. (2000) showed that the autocovariance of a locally stationary wavelet process has a wavelet representation, known as the local autocovariance. We investigate continuous-time versions and begin with continuous autocorrelation wavelets; see Beylkin & Saito (1992).

DEFINITION 3. *For scale $u \in \mathbb{R}^+$ and lag $\tau \in \mathbb{R}$, the continuous autocorrelation wavelets are*

$$\Psi(u, \tau) = \int_{-\infty}^\infty \psi(u, v)\psi(u, v - \tau)\, \mathrm{d}v.$$

DEFINITION 4. *For time $t \in \mathbb{R}$ and lag $\tau \in \mathbb{R}$, the continuous-time local autocovariance is*

$$c(t, \tau) = \int_0^\infty S(u, t)\Psi(u, \tau)\, \mathrm{d}u, \tag{11}$$

where $S(u,t)$ *and* $\Psi(u,\tau)$, $u \in \mathbb{R}^+$, *are the continuous-time evolutionary wavelet spectrum and autocorrelation wavelets, respectively.*

The usual autocovariance function of a continuous-time locally stationary wavelet process is

$$c_X(t,\tau) = \operatorname{cov}\{X(t), X(t+\tau)\} = E\{X(t)X(t+\tau)\}, \tag{12}$$

since $X(t)$ has zero mean. This can also be written in terms of the evolutionary wavelet spectrum and the underlying wavelets of the process as

$$c_X(t,\tau) = \int_0^\infty \int_{-\infty}^\infty S(u,t+v)\psi(u,v)\psi(u,v-\tau)\,\mathrm{d}v\,\mathrm{d}u; \tag{13}$$

see the proof of Proposition 1 in the Supplementary Material. Comparing (11) and (13), we see that, at each time t, the local autocovariance effectively replaces $S(u,t+v)$ with the constant $S(u,t)$, as a function of v, over a finite interval centred around t of width matching that of the wavelet. Given the slow evolution of the statistical properties of the process, $S(u,t+v)$ is similar to $S(u,t)$ in this region and, if the process is stationary, the two expressions are equivalent. Therefore, the local autocovariance assumes that the spectrum is locally constant, or, equivalently, that the process is stationary at a local level, a concept that underpins the framework of locally stationary processes. This is similar to the discrete-time local autocovariance defined in Nason et al. (2000). The next proposition establishes a bound on the maximum difference between (11) and (13).

Proposition 1. *Let* $\{X(t)\}_{t\in\mathbb{R}}$ *be a continuous-time locally stationary wavelet process, and let* $c_X(t,\tau)$ *be its autocovariance function and* $c(t,\tau)$ *its local autocovariance. Then,*

$$|c_X(t,\tau) - c(t,\tau)| \leqslant \gamma \int_0^\infty uK(u)\,\mathrm{d}u$$

for all scales $u \in \mathbb{R}^+$ *and times* $v \in \mathbb{R}$, *where* γ *is a constant that depends on the decay structure of wavelet* ψ *and* $K(\cdot)$ *is the Lipschitz constant function from Definition 1.*

In essence, the more stationary process $X(t)$ is, the closer $c_X(t,\tau)$ and $c(t,\tau)$ will be, since $\int_0^\infty uK(u)\,\mathrm{d}u$ can be interpreted as a measure of the nonstationarity of the process. For a more local bound, if spectrum S admits a Taylor series representation, we can write

$$S(u,t+v) = S(u,t) + v\frac{\partial S}{\partial v}(u,t) + o(v^2)$$

for v close to t. Using this expansion, we can write

$$|c_X(t,\tau) - c(t,\tau)| \leqslant \gamma \int_0^\infty u\left|\frac{\partial S}{\partial v}(u,t)\right|\mathrm{d}u + M,$$

where $M \geqslant 0$ is a constant that bounds the discrepancy due to the higher-order terms. From the above expression, it becomes clearer that the difference between the $c_X(t,\tau)$ and $c(t,\tau)$ depends on the local structure of the process. A comparison of the local and standard autocovariance functions on a simple example can be found in the Supplementary Material.

DEFINITION 5. *The continuous-time local autocorrelation is defined as*

$$\rho(t,\tau) = \sigma^{-2}(t)c(t,\tau)$$

for all $t, \tau \in \mathbb{R}$, *where* $\sigma^2(t) = c(t,0) = \int_0^\infty S(u,t)\,\mathrm{d}u$ *is the local variance of process* $X(t)$ *and* $c(t,\tau)$ *is the local autocovariance defined above.*

Example 1 *(continued)*. For the Haar MA(α) processes, we have

$$c(t,\tau) = \Psi_{\mathrm{H}}(\alpha,\tau) = c_X(t,\tau), \tag{14}$$

where Ψ_{H} is the Haar autocorrelation wavelet; see the Supplementary Material. Since Haar MA processes are stationary, the local and standard autocovariance functions are equivalent. Furthermore, $\rho(t,\tau) = c(t,\tau)$, since $\Psi_{\mathrm{H}}(\alpha,0) = 1$, as shown in Proposition 2 below.

Remark 3. Strictly speaking, since the Dirac measure is not continuous with respect to the Lebesgue measure, (14) is calculated using $S_t^{(I)}$ from (10) as

$$c(t,\tau) = \int_0^\infty \Psi(u,\tau)\,S_t^{(I)}(\mathrm{d}u) = \int_0^\infty \Psi(u,\tau)\delta_\alpha(\mathrm{d}u) = \Psi(\alpha,\tau).$$

To conclude this section, we list some useful properties of autocorrelation wavelets.

PROPOSITION 2. *For all* $u \in \mathbb{R}^+$, *with* $\Psi(u,\tau)$ *denoting the autocorrelation wavelets and* $\Psi(\tau) = \Psi(1,\tau)$, *we have*

(i) $\Psi(u,\tau)$ *is symmetric as a function of* τ, $\Psi(u,\tau) = \Psi(u,-\tau)$;
(ii) $\Psi(u,0) = 1$;
(iii) $\Psi(u,\tau) = \Psi(\tau/u)$;
(iv) $\hat{\Psi}(u,\omega) = u|\hat{\psi}(u\omega)|^2$, *where* $\hat{\Psi}(u,\omega)$ *is the Fourier transform of* $\Psi(u,t)$;
(v) $\int_{-\infty}^{\infty} \Psi(u,\tau)\,\mathrm{d}\tau = 0$.

3.4. *Relationship between the wavelet periodogram and spectrum*

We now define the raw wavelet periodogram and its relationship to the evolutionary wavelet spectrum. First, we need to define the wavelet coefficients.

DEFINITION 6. *The wavelet coefficients of process* $X(t)$ *in terms of the wavelet basis* ψ *are*

$$d(u,v) = \int_{-\infty}^{\infty} X(t)\psi(u,t-v)\,\mathrm{d}t \tag{15}$$

for all $u \in \mathbb{R}^+$ *and* $v \in \mathbb{R}$.

DEFINITION 7. *The raw wavelet periodogram of process* $X(t)$ *is*

$$I(u,v) = |d(u,v)|^2$$

for all scales $u \in \mathbb{R}^+$ *and times* $v \in \mathbb{R}$, *where the* $d(u,v)$ *are wavelet coefficients. We define the notation* $\beta(u,v) = E\{I(u,v)\}$, *which turns out to be useful.*

Before going further, we introduce the continuous counterpart of the inner product matrix defined by Nason et al. (2000), the inner product kernel.

Definition 8. *For scales $u, x \in \mathbb{R}^+$, the inner product kernel is*

$$A(u,x) = \int_{-\infty}^{\infty} \Psi(u,\tau)\Psi(x,\tau)\,\mathrm{d}\tau.$$

Proposition 3. *The inner product kernel A is continuous, symmetric and nonnegative definite. Furthermore, for all $b > 0$ and $x, y \in \mathbb{R}^+$, $A(bx, by) = bA(x, y)$.*

The following theorem establishes a relationship between the raw wavelet periodogram and the evolutionary wavelet spectrum, paralleling and extending that of Nason et al. (2000).

Theorem 1. *Let $\beta(u, v)$ be the expectation of the raw wavelet periodogram of the locally stationary wavelet process $X(t)$, $S(u, v)$ be the corresponding evolutionary wavelet spectrum and $A(u, x)$ be the inner product kernel. Then*

$$\left|\beta(u,v) - \int_0^{\infty} A(u,x)S(x,v)\,\mathrm{d}x\right| \leqslant \gamma \int_0^{\infty} xK(x)\,\mathrm{d}x \tag{16}$$

for all scales $u \in \mathbb{R}^+$ and times $v \in \mathbb{R}$, where γ is a constant that depends on the decay structure of wavelet ψ and $K(\cdot)$ is the Lipschitz constant function from Definition 1.

As for Proposition 1, if $X(t)$ is stationary then the two terms on the left-hand side of (16) are equivalent. The expectation of the raw wavelet periodogram can be written as

$$\beta(u,v) = \int_0^{\infty}\int_{-\infty}^{\infty} S(x, y+v)\left\{\int_{-\infty}^{\infty} \psi(x, y-t)\psi(u,t)\,\mathrm{d}t\right\}^2 \mathrm{d}y\,\mathrm{d}x;$$

see the proof of Theorem 1 in the Supplementary Material, containing terms resembling the cross-scale autocorrelation wavelets from the discrete setting (Killick et al., 2020; Embleton et al., 2022). Hence, $\int_0^{\infty} S(x,v)A(u,x)\,\mathrm{d}x$ can be viewed as approximating the expectation of the raw wavelet periodogram under the assumption that $S(u, t+v) = S(u, t)$ at a local level, in the same way that the local autocovariance approximates the autocovariance. Because a similar discussion follows Proposition 1, we omit a more detailed analysis here for brevity.

3.5. *Solving for the spectrum and related quantities*

Theorem 1 motivates the following definition.

Definition 9. *The inner product operator $T_A\colon L^2(\mathbb{R}^+) \mapsto L^2(\mathbb{R}^+)$ is a linear integral operator defined by*

$$(T_A f)(x) = \int_0^{\infty} A(x,y)f(y)\,\mathrm{d}y \tag{17}$$

for $f \in L^2(\mathbb{R}^+)$, where $A(x, y)$ is the inner product kernel.

Theorem 1 and Definition 9 suggest that one could approximate $\beta(u, v)$ by

$$\beta(u, v) \approx \int_0^\infty A(u, x)S(x, v)\,\mathrm{d}x = \{T_A S(\cdot, v)\}(u). \tag{18}$$

Therefore, given $\beta(u, v)$, we could invert the above linear integral equation to obtain the evolutionary wavelet spectrum as a solution $\tilde{S}$. We now show how this could be done using the properties of the inner product kernel and Mercer's theorem (Mercer, 1909): this is a key difference with the earlier discrete-time version of Nason et al. (2000).

Since T_A has a continuous symmetric kernel, it is compact, bounded and self-adjoint. Furthermore, we have the following result.

Proposition 4. *Let T_A be the linear integral operator from Definition 9. Then T_A has a trivial null space, i.e., $N(T_A) = \{f \in L^2(\mathbb{R}^+)\colon T_A f = 0\} = \{0\}$.*

Since the inner product kernel is continuous, symmetric and nonnegative definite, we can apply Mercer's theorem (Mercer, 1909), see the Supplementary Material, to obtain a decomposition of A of the form

$$A(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x)\varphi_n(y),$$

where λ_n and φ_n are the eigenvalues and eigenfunctions of T_A. Therefore,

$$(T_A f)(x) = \sum_{n=1}^{\infty} \lambda_n \langle \varphi_n, f\rangle \varphi_n(x), \qquad \langle f, g\rangle = \int_0^\infty f(x)g(x)\,\mathrm{d}x,$$

for $f, g \in L^2(\mathbb{R}^+)$. Since T_A is self-adjoint, this allows us to obtain the solution

$$\tilde{S}(u, t) = \sum_{n=1}^{\infty} \lambda_n^{-1} \langle \beta(\cdot, t), \varphi_n(\cdot)\rangle \varphi_n(u) \tag{19}$$

to the linear integral (18) for all $u \in \mathbb{R}^+$ and $t \in \mathbb{R}$, which is unique, since $N(T_A) = \{0\}$.

The evolutionary wavelet spectral solution (19) can be used to obtain an estimate of the local autocovariance using (11). Furthermore, the ability to invert T_A also permits us to invert (11) to obtain a representation of the spectrum in terms of the local autocovariance as follows.

Theorem 2. *Let $c(t, \tau)$ be the local autocovariance of a continuous-time locally stationary wavelet process. Then the evolutionary wavelet spectrum can be written as*

$$S(u, t) = \int_{-\infty}^{\infty} B(u, \tau)c(t, \tau)\,\mathrm{d}\tau, \qquad B(u, \tau) = \sum_{n=1}^{\infty} \lambda_n^{-1} \langle \Psi(\cdot, \tau), \varphi_n(\cdot)\rangle \varphi_n(u),$$

where Ψ *is an autocovariance wavelet and* λ_n *and* φ_n *are the eigenvalues and eigenvectors of the inner product operator* T_A *from Definition 9.*

Example 2. The expectation of the Haar raw wavelet periodogram of a Gaussian white noise process $G((s,t]) \sim \mathcal{N}\{0,(t-s)\sigma^2\}$ *is* $\beta(u,v) = \sigma^2$ for all $u \in \mathbb{R}^+$ and $v \in \mathbb{R}$. From

$$\sigma^2 = \int_0^\infty A_{\mathrm{H}}(u,x)S(x,v)\,\mathrm{d}x,$$

where A_{H} is the Haar inner product kernel, we obtain

$$S(u,v) = \frac{\sigma^2}{\log(2)u^2}$$

for all $u \in \mathbb{R}^+$ and $v \in \mathbb{R}$. Further details can be found in the Supplementary Material.

So far, we have ignored the fact that generally one does not work directly with $\beta(u,v)$, but rather with a realization, or in some cases a set of samples, of the locally stationary wavelet process $X(t)$, which can then be used to obtain an estimate of $\beta(u,v)$. This, of course, introduces noise into the linear integral equation. To address this problem, regularization methods such as the spectral cut-off method discussed in the Supplementary Material could be employed.

4. Implementation of spectral and autocovariance estimators

4.1. *Sampled process definition*

In practice, one typically has access to a discrete and finite realization of the process, rather than the entire sample path. Therefore, estimation of spectral and autocovariance quantities requires adaptations of the previous section's theoretical framework. To begin, we provide a new definition of locally stationary wavelet processes in continuous time, reintroducing a sequence of stochastic processes and the concept of rescaled time.

Definition 10. *Let* $n(T)$ *be a nondecreasing integer-valued function of* $T \in \mathbb{R}^+$, *such that* $n(T) \to \infty$ *as* $T \to \infty$, *and* $\mathcal{T} = \{t_0, t_1, \ldots, t_{n(T)-1}\}$ *be a set of times, such that* $0 = t_0 < t_1 < \cdots < t_{n(T)-1} = T$. *A sequence of stochastic processes* $\{X_T(t)\}_{t\in\mathcal{T}}$ *indexed by* T *is said to be a sampled continuous-time locally stationary wavelet process if it admits the representation*

$$X_T(t) = \int_0^{J_\psi(T)} \int_{-\infty}^{\infty} w_T(u,v)\psi(u,v-t)L(\mathrm{d}u,\mathrm{d}v),$$

in the mean square sense, where L *is a square-integrable homogeneous Lévy basis with* $E\{L(A)\} = 0$ *and* $E\{L(A)^2\} = \mathrm{Leb}(A)$ *for all* $A \in \mathcal{B}_b(\mathbb{R}^+ \times \mathbb{R})$, $\{\psi_{u,v}(x),\ u \in \mathbb{R}^+,\ v \in \mathbb{R}\}$ *is a continuous wavelet family,* $w_T \in L^2(\mathbb{R}^+,\mathbb{R})$ *is a sequence of location-scale amplitude functions and* $J_\psi(T)$ *is an increasing function that depends on the wavelet* ψ. *Furthermore, there exists a function* $W\colon \mathbb{R}^+ \times (0,1) \mapsto \mathbb{R}$ *such that the following conditions hold.*

(i) *For each fixed scale* $u \in \mathbb{R}^+$, $W(u, \cdot)$ *is a Lipschitz continuous function with Lipschitz constant* $K(u)$. *The Lipschitz constant function* K *is bounded and satisfies*

$$\int_0^\infty uK(u)\,\mathrm{d}u < \infty.$$

(ii) *For each fixed* $z \in (0, 1)$, $W(\cdot, z)$ *satisfies*

$$\int_0^\infty |W(u, z)|^2\,\mathrm{d}u < \infty.$$

(iii) *For each fixed scale* $u \in \mathbb{R}^+$, *there exists a constant* $C(u)$ *such that, for each* $T \in \mathbb{R}^+$,

$$\sup_{v \in \mathcal{T}} \left| w_T(u, v) - W\left(u, \frac{v}{T}\right) \right| \leqslant \frac{C(u)}{T}, \qquad \int_0^\infty C(u)\,\mathrm{d}u < \infty.$$

We have now reintroduced the concept of rescaled time, $z = v/T \in (0, 1)$, as is standard in the literature on locally stationary processes; see Dahlhaus (1997) for example. In essence, this means that, as $T \to \infty$, we are collecting increasing amounts of data and so are learning more and more about the local structure of the process. Assumption (iii) in Definition 10 formalizes this notion, ensuring that $w_T(u, v)$ converges to its rescaled-time limit $W(u, v/T)$, while also becoming smoother in the process. In our setting, the growing number of observations is $n(T)$, whereas T now denotes the total length of the data. Thus, we still rescale the process by T, with the increased flexibility of being able to handle irregularly spaced observations.

Function $J_\psi(T)$ determines the maximum scale that contributes to the process. As more data become available, the model is able to include increasingly coarser scales and so $J_\psi(T) \to \infty$ as $T \to \infty$. We make the dependence of $J_\psi(T)$ on the structure of the underlying wavelet explicit by including ψ in the subscript.

Continuous wavelets can be defined for any scale in $\mathbb{R}^+$ and can be shifted to any location in $\mathbb{R}$. This makes the continuous wavelet transform particularly well suited for handling missing observations or irregularly spaced data.

Definition 11. *The evolutionary wavelet spectrum of the sampled continuous-time locally stationary wavelet process* $X_T(t)$ *with respect to* ψ *is*

$$S(u, z) = |W(u, z)|^2$$

for all $u \in \mathbb{R}^+$ *and* $z = v/T \in (0, 1)$.

Assumptions (ii) and (iii) of Definition 10, imply that, for each fixed $z \in (0, 1)$,

$$\int_0^\infty S(u, z)\,\mathrm{d}u < \infty, \qquad S(u, z) = \lim_{T \to \infty} |w_T(u, v)|^2,$$

for all $u \in \mathbb{R}^+$ and $v \in \mathbb{R}$.

The local autocovariance and autocorrelation remain as before, except that we replace t with its rescaled-time version z, as was done for the evolutionary wavelet spectrum above. To emphasize the dependence on T, we now denote the autocovariance by $c_{X_T}(t, \tau)$.

Proposition 1 highlighted the discrepancy between the regular autocovariance from (12) and the wavelet autocovariance from (11) for the theoretical process with equality in the stationary case. In the sampled case, the autocovariances become close asymptotically.

PROPOSITION 5. *Let* $\{X_T(t)\}_{t\in\mathcal{T}}$ *be a sampled continuous-time locally stationary wavelet process,* $c_{X_T}(t,\tau)$ *be its autocovariance function,* $c(z,\tau)$ *be the local autocovariance and* $z = t/T$. *Then*

$$|c_{X_T}(t,\tau) - c(z,\tau)| = O(T^{-1}).$$

In other words, the autocovariance converges to the local autocovariance as $T \to \infty$.

4.2. *Estimation theory*

The wavelet coefficients are defined as in (15), while the raw wavelet periodogram is now defined in rescaled time $z = v/T$ as

$$I(u,z) = I\left(u, \frac{v}{T}\right) = |d(u,v)|^2 = |d(u,zT)|^2.$$

As in Definition 7, we again use the notation $\beta(u,z) = E[I(u,z)]$. This leads to the following result.

THEOREM 3. *Let* $I(u,z)$ *be the raw wavelet periodogram of the sampled continuous-time locally stationary wavelet process* $X_T(t)$, $S(u,z)$ *be its corresponding evolutionary wavelet spectrum and* $A(u,x)$ *be the inner product kernel. Then*

$$\left|\beta(u,z) - \int_0^{J_\psi(T)} A(u,x)S(x,z)\,\mathrm{d}x\right| = O(T^{-1}) \tag{20}$$

and

$$\left|\mathrm{var}\{I(u,z)\} - 2\left\{\int_0^{J_\psi(T)} A(u,x)S(x,z)\,\mathrm{d}x\right\}^2\right| = O(uT^{-1})$$

for all $u \in (0, J_\psi(T)]$ *and* $z \in (0,1)$.

Equation (20) suggests estimating spectrum S by solving the linear integral equation

$$\beta(u,z) = \int_0^{J_\psi(T)} A(u,x)S(x,z)\,\mathrm{d}x = \{T_A S(\cdot,z)\}(u), \tag{21}$$

using estimates $\hat{\beta}$ of β, on which we elaborate in the Supplementary Material, where

$$(T_A f)(x) = \int_0^{J_\psi(T)} A(x,y)f(y)\,\mathrm{d}y.$$

The same properties hold for this linear integral operator as for its theoretical counterpart defined in (17). The use of an estimate of β naturally introduces noise into (21). Hence, to compute an estimate $\hat{S}$ of spectrum S, we use the iterative soft-thresholding algorithm

developed by Daubechies et al. (2004) and described in § 2. This involves reformulating (21) as a minimization problem with the aim of minimizing the discrepancy

$$\Delta S(\cdot, z) = ||T_A S(\cdot, z) - \hat{\beta}(\cdot, z)||_2^2 + ||\mu(\cdot) S(\cdot, z)||_1 \tag{22}$$

for each $z \in (0, 1)$, where $||\cdot||_1$ and $||\cdot||_2$ denote the $L^1(\mathbb{R}^+)$ and $L^2(\mathbb{R}^+)$ norms, respectively, and $\mu > 0$ is the regularization function. By definition, the evolutionary wavelet spectrum is a nonnegative quantity and so we use asymmetric soft thresholding. Hence, in this setting, we reformulate the iterative scheme as

$$\hat{S}^n(\cdot, z) = \mathcal{S}_\mu[\hat{S}^{n-1}(\cdot, z) + T_A\{\hat{\beta}(\cdot, z) - T_A \hat{S}^{n-1}(\cdot, z)\}], \qquad n = 1, \ldots, N, \tag{23}$$

where N is the number of iterations and

$$(\mathcal{S}_\mu f)(u) = \begin{cases} f(u) - \mu(u)/2 & \text{if } f(u) \geqslant \mu(u)/2, \\ 0 & \text{if } f(u) < \mu(u)/2, \end{cases}$$

for $u \in \mathbb{R}^+$, which, as shown in § 2 of Daubechies et al. (2004), converges strongly to the unique minimizer of (22). Unfortunately, an exact closed-form equation for estimating the evolutionary wavelet spectrum from the raw wavelet periodogram does not exist, which is a key difference from the discrete-time setting of Nason et al. (2000) and other later estimators in the literature. More details on the convergence of this algorithm in practice can be found in the Supplementary Material.

Remark 4. We have implicitly chosen to penalize the function values directly when imposing the constraint in (22) and, subsequently, when defining the $\mathcal{S}_\mu$ operator. This makes guaranteeing the nonnegativity of the spectrum straightforward and is equivalent to selecting the standard basis in (5). The regularization in (22) takes the form of a function μ, allowing us to penalize scales differently if necessary. We have replaced $T_A^\dagger$ with T_A in (23), as T_A is, by definition, self-adjoint.

4.3. *Smoothing the discretized raw wavelet periodogram across time*

As for the discrete setting (Nason et al., 2000), the raw wavelet periodogram is not a consistent estimator of β. A key advantage of the continuous wavelet transform is that we can compute the wavelet coefficients $d(u, v)$ on a regularly spaced grid, regardless of the time series' original structure. Hence, we assume that we have computed the continuous wavelet transform at the evenly spaced discrete locations v_j, $j = 1, \ldots, M_v = 2^\eta$ for some $\eta \in \mathbb{R}^+$, such that $M_v \to \infty$ as $T \to \infty$, and the corresponding estimate of the raw wavelet periodogram $\beta(u, z_j)$ for $z_j = v_j/T$. Full details can be found in the Supplementary Material. Standard wavelet smoothing techniques using the discrete wavelet transform (see the 1993 University of California, Berkeley, PhD thesis by H. Gao, Donoho & Johnstone, 1994; Neumann & von Sachs, 1995; Neumann, 1996; von Sachs & Schneider, 1996; Nason, 2008) can then be applied to each level of $\hat{\beta}$ across locations in a straightforward manner, as in Nason et al. (2000).

Remark 5. In practice, one discretizes the scales as well; however, since we smooth each level of β independently, we omit this here. Full details can be found in the Supplementary Material.

To apply wavelet shrinkage to the discretized periodogram, we introduce a separate orthonormal discrete wavelet basis of $L^2(0,1)$. We refer to these as the wavelet shrinkage basis and coefficients to distinguish them from the continuous wavelets defined in (7) and used throughout the article.

DEFINITION 12. *Let* $\{\tilde{\phi}_{\ell_0,m}, \tilde{\psi}_{\ell,m}\}$ *be an orthonormal discrete wavelet basis of* $L^2(0,1)$*, and define the mother and father wavelet coefficients of the discretized raw wavelet periodogram as*

$$\tilde{d}^u_{\ell,m} = \frac{1}{M_v}\sum_{j=1}^{M_v} \hat{\beta}(u, z_j)\tilde{\psi}_{\ell,m}(z_j), \qquad \ell = \ell_0, \ldots, \eta,\ m = 0, \ldots, 2^\ell - 1, \tag{24}$$

and

$$\tilde{c}^u_{\ell_0,m} = \frac{1}{M_v}\sum_{j=1}^{M_v} \hat{\beta}(u, z_j)\tilde{\phi}_{\ell_0,m}(z_j), \qquad m = 0, \ldots, 2^{\ell_0} - 1, \tag{25}$$

respectively, where $\tilde{\psi}_{\ell,m}(z) = 2^{\ell/2}\tilde{\psi}(2^\ell z - m)$*,* $\tilde{\psi}$ *is a wavelet function satisfying (6),* $\ell = \ell_0, \ldots, \eta$*,* $m = 0, \ldots, 2^\ell - 1$ *and* ℓ_0 *is the coarsest scale included in the scheme. The superscript* u *is used to indicate that each level* u *is smoothed separately as a function of* z*.*

Equations (24) and (25) parallel equation (39) of an unpublished technical report from Stanford University by R. von Sachs, G. P. Nason and G. Kroisandt. Following Nason (2008), we set $\ell_0 = 3$, since the finer scales are those most susceptible to noise. We can then apply a soft threshold, which we define below, to the wavelet coefficients $\tilde{d}^u_{\ell,m}$ to obtain estimates $\tilde{d}^{u,S}_{\ell,m}$. The resulting nonlinear estimate of the raw wavelet periodogram is then

$$\tilde{\beta}(u, z_j) = \sum_{m=0}^{2^{\ell_0}-1} \tilde{c}^u_{\ell_0,m}\tilde{\phi}_{\ell_0,m}(z_j) + \sum_{\ell=\ell_0}^{\eta}\sum_{m=0}^{2^\ell-1} \tilde{d}^{u,S}_{\ell,m}\tilde{\psi}_{\ell,m}(z_j).$$

The following theorem parallels that of Nason et al. (2000).

THEOREM 4. *For locally stationary wavelet processes with homogeneous Gaussian Lévy bases, the wavelet coefficients* $\tilde{d}^u_{\ell,m}$ *are asymptotically normally distributed with*

$$\left| E(\tilde{d}^u_{\ell,m}) - \int_0^1 \int_0^\infty A(u,x)S(x,z)\,\mathrm{d}x\,\tilde{\psi}_{\ell,m}(z)\,\mathrm{d}z \right| = O(2^{\ell/2}T^{-1})$$

and

$$\left| \mathrm{var}(\tilde{d}^u_{\ell,m}) - 2T^{-1}\int_0^1 \left\{\int_0^\infty A(u,x)S(x,z)\,\mathrm{d}x\right\}^2 \tilde{\psi}^2_{\ell,m}(z)\,\mathrm{d}z \right| = O(2^\ell u T^{-2}).$$

Hence, Theorem 5 below follows from H. Gao's 1993 University of California, Berkeley, PhD thesis, von Sachs & Schneider (1996) and Nason et al. (2000).

THEOREM 5. *Define the threshold*

$$\lambda(\ell, m, u, T) = \tilde{\sigma}^u_{\ell,m}\log(T), \qquad \tilde{\sigma}^u_{\ell,m} = \{\mathrm{var}(\tilde{d}^u_{\ell,m})\}^{1/2}. \tag{26}$$

Then, under the assumptions of Theorem 4, for each $u \in \mathbb{R}^+$,

$$E\left\{\int_0^1 |\tilde{\beta}(u,z) - \beta(u,z)|^2 \,\mathrm{d}z\right\} = O\{T^{-2/3}\log^2(T)\}.$$

The above relies on the assumption that the underlying Lévy basis is Gaussian. In other settings, one could apply the techniques described by Neumann & von Sachs (1995). The methods relying on the Haar-Fisz transform, as developed by Fryzlewicz & Delouille (2005) and Fryzlewicz & Nason (2006) for the discrete setting, could be extended to work in the continuous case as well.

In general, the smoothing techniques described here are applied to the logarithm of the raw wavelet periodogram instead, which tends to stabilize the variance of the process; see Remark 4.11 of von Sachs & Schneider (1996) or Nason et al. (2000). This mitigates the varying spectral variance and also pulls the distribution closer to normality. Therefore, for the remainder of this article, we apply any smoothing to the log raw wavelet periodogram.

To estimate the noise level $\tilde{\sigma}^u_{\ell,m}$ in threshold (26), Donoho & Johnstone (1994) suggested using the median absolute deviation of the wavelet shrinkage coefficients at the finest scale, divided by 0.6745, since the finest scale coefficients are generally just noise. The scaling term of 0.6745 is used to ensure that the median absolute deviation is a consistent estimator in the normal case, but often used for other symmetric distributions. Alternatively, Nason (2008) proposed computing the median absolute deviation on the wavelet coefficients ranging from a scale of 3 to the finest scale. Both methods estimate a threshold that does not depend on ℓ and m, but does depend on u, as the smoothing is performed separately at each scale.

4.4. *Estimating the spectrum from the smoothed raw wavelet periodogram*

The smoothed raw wavelet periodogram $\tilde{\beta}$ is then used to compute an estimate of the evolutionary wavelet spectrum $\hat{S}$ using the iterative soft-thresholding algorithm of Daubechies et al. (2004). The initial value in (23) is set to $S^0(u,z) = \hat{\beta}(u,z)$, although this can be chosen arbitrarily. As shown in §4 of Daubechies et al. (2004), one should select $\mu = \mu(\varepsilon)$ such that

$$\lim_{\varepsilon\to 0}\mu_\varepsilon = 0 \quad \text{and} \quad \lim_{\varepsilon\to 0}\varepsilon^2/\mu_\varepsilon = 0,$$

to ensure convergence to the true solution as the noise level ε approaches zero. Therefore, we set $\mu_\varepsilon = \varepsilon$. In practice, we estimate this quantity using the same method as for noise estimation in wavelet shrinkage, as described by Donoho & Johnstone (1994), enabling us to penalize each scale individually. Moreover, the iterative scheme in (23) operates on each location separately, adding greater flexibility by allowing μ to vary by location.

4.5. *Computational complexity*

Our method uses the continuous wavelet transform, which, in a naive implementation, expends $O(n^2)$ computational effort per scale, where n is the number of data points. However, several methods have been developed that can achieve $O(n)$ instead (Wang & He, 2023). Smoothing the raw wavelet periodogram adds another $O(n)$ operations per scale. The iterative scheme (23) requires $O(m^2)$ operations per iteration and location, where m is the number of scales. This can be reduced to $O(m)$ if the inner product operator T_A

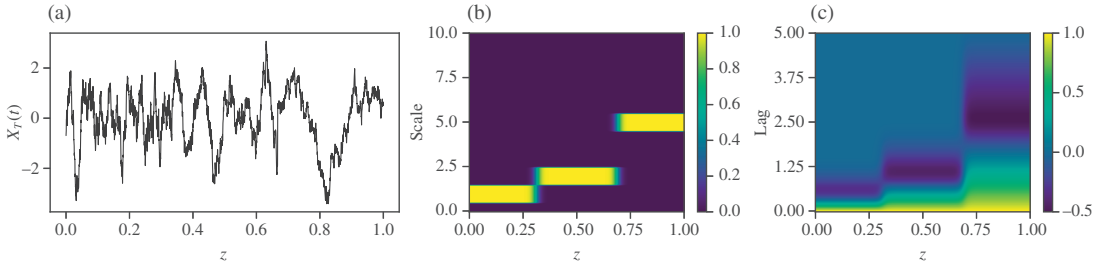

Fig. 1. (a) A single realization of a continuous-time nonstationary Haar MA process with $n(T) = 1500$ equally spaced samples. (b) The underlying evolutionary wavelet spectrum. (c) The corresponding local autocorrelation.

is sparse (Daubechies et al., 2004), which is often the case. On a standard laptop it takes approximately 30 s to simulate 1000 realizations, compute the raw wavelet periodograms and average the results, for example the realizations in § 4.6 below. It takes a further 5–10 s to smooth those estimates and compute the evolutionary wavelet spectrum estimates.

The earlier qualitative and approximate methods of Knight et al. (2012) require the application of P lifting wavelet transforms, where P corresponds to the number of lifting algorithm trajectories, with each lifting (wavelet) transform being $O(n)$ computational effort. Hence, the total computational effort for the methods of Knight & Nason (2009) is $O(nP)$, where P is typically of the order of thousands. The spectral estimation methods of Knight et al. (2012) build on Knight & Nason (2009), resulting in a method that is considerably more computationally intensive, but whose precise order cannot be established and implementation code can take many hours to run.

4.6. *Nonstationary Haar moving average example*

We now apply our methods to realizations from a nonstationary continuous-time Haar moving average process. Figure 1(b) shows the true underlying evolutionary wavelet spectrum. See the Supplementary Material for a complete functional specification and considerably more detail. We also display $n(T) = 1500$ samples drawn from a single realization and the true local autocorrelation, which, in this case, is equal to the local autocovariance.

The estimates of the evolutionary wavelet spectrum from a single realization of $X_T(t)$ are shown on the far left of Fig. 2. The top row corresponds to running the iterative soft-thresholding algorithm of Daubechies et al. (2004) for $N = 100$ iterations and the bottom row for $N = 10\,000$.

Figure 2 also shows the results of drawing R realizations each with $n(T) = 1500$ samples, estimating a raw wavelet periodogram for each, averaging them and then using the iterative soft-thresholding algorithm of Daubechies et al. (2004) to obtain a spectral estimate. Section S4.2 of the Supplementary Material defines this mathematically in equation (S8) and discusses the practical importance of this averaging when replicated samples are available. The second, third and fourth columns of Figure 2 show estimates averaged over $R = 10$, 100 and 1000 realizations.

Each estimate $\hat{\beta}$ is smoothed, as described in § 4.3 using the wavelets of Daubechies (1992) with three vanishing moments. As in Nason (2008), the threshold in (26) is estimated using, and applied to, the coefficients ranging from level three to the finest scale. The local autocorrelations obtained according to equation (S9) in the Supplementary Material are displayed in Fig. 3.

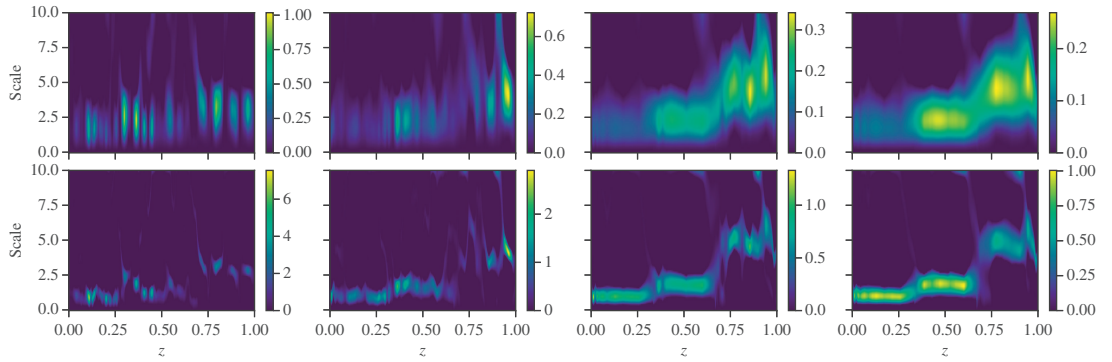

Fig. 2. Haar moving average evolutionary wavelet spectral estimates using the iterative soft-thresholding algorithm of Daubechies et al. (2004). Top row: $N = 100$ iterations. Bottom row: $N = 10\,000$ iterations. From left to right: estimates computed using $R = 1, 10, 100$ and 1000 realizations.

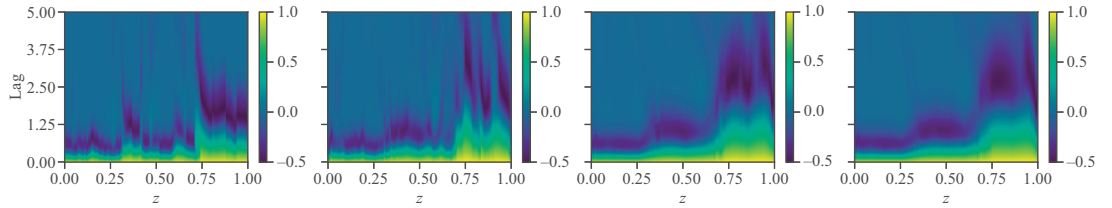

Fig. 3. Haar moving average local autocorrelation estimates. From left to right: estimates computed using $R = 1$, 10, 100 and 1000 realizations and $N = 10\,000$.

As expected, results improve as we increase both the number of samples used to estimate the raw wavelet periodogram and the number of iterations in the iterative soft-thresholding algorithm of Daubechies et al. (2004). We observe that finer-scale information requires a larger number of iterations to become as visible in the estimate as information at larger scales. Spectral information also tends to deviate more from its true value at larger scales, aligning with the results in Theorem 3. Even when only a single realization is employed, it is still possible to distinguish between different frequency regions. However, all estimates display a higher degree of variability in such cases, and so it may be beneficial to further investigate more appropriate smoothing techniques, an avenue left for future research.

5. Real-world applications

5.1. *Infant electrocardiogram*

We apply our methods to the electrocardiogram recording of an infant, consisting of 2048 regularly spaced observations sampled at 1/16 Hz, used in the original discrete-time work of Nason et al. (2000) and displayed in Fig. 4. The `BabyECG` time series used to construct this plot is freely available from the `wavethresh` (Nason, 2024) package on CRAN. To compare our approach to that of Knight et al. (2012), we replace 25% of the observations with missing values.

Nason et al. (2000) and Knight et al. (2012) used the least asymmetric wavelets of Daubechies (1992) of order 10 for estimation and we use the similar Ricker wavelets; see the Supplementary Material. We estimate our spectrum for scales ranging from 1 to $1024 = 2^{10}$,

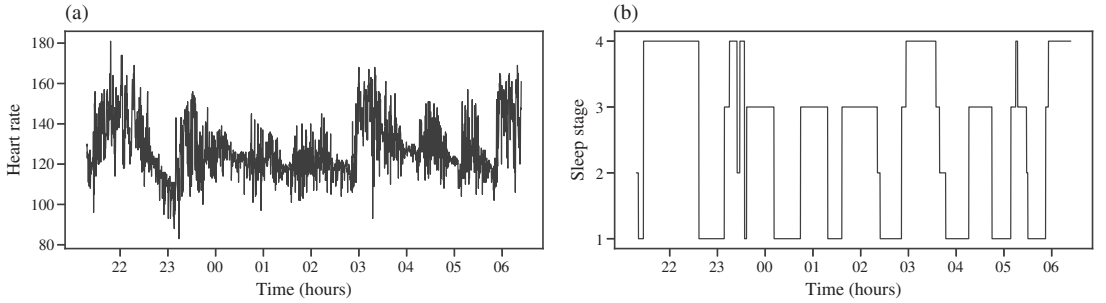

Fig. 4. (a) Electrocardiogram recording of a sleeping infant and (b) the corresponding sleep stages.

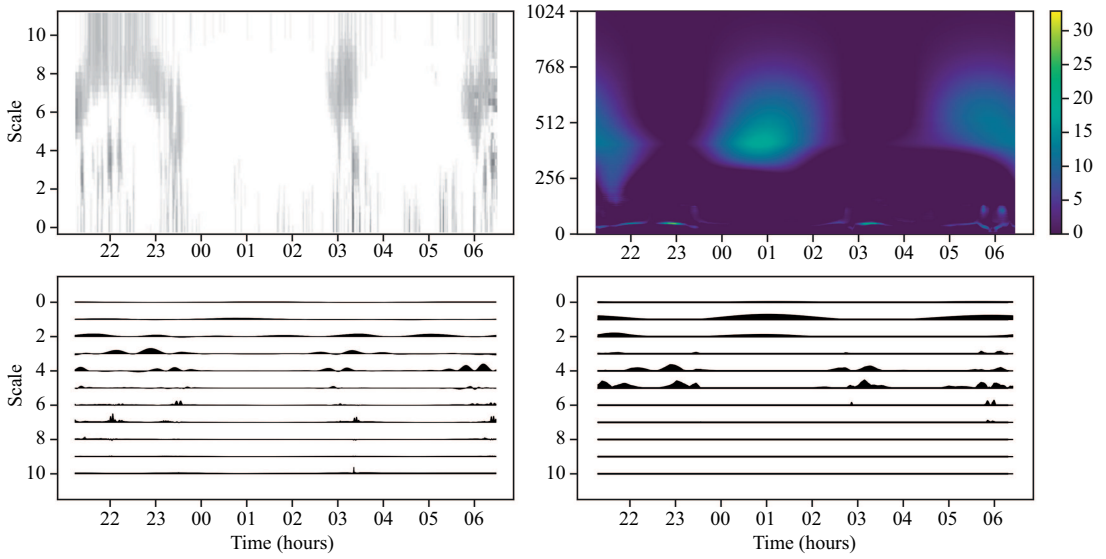

Fig. 5. Sleeping infant heart rate spectral estimates. Top left: wavelet periodogram, reproduced with permission from Knight et al. (2012). Top right: continuous wavelet spectral estimate. Bottom left: wavelet spectrum estimated with the methods of Nason et al. (2000). Bottom right: our discretized continuous wavelet spectrum.

smoothing the log periodogram using the wavelets of Daubechies (1992) with four vanishing moments. We also discretize our spectral estimate into dyadic bins, but only to compare it with that of Nason et al. (2000) estimated on the complete time series. Both continuous and discretized estimates are displayed in the right column of Fig. 5. Since different methods display scale numbering differently, all spectra in Fig. 5 are displayed so that the scales become coarser when moving up the y axis.

Another appeal of our approach over the lifting-based method of Knight et al. (2012) lies in the natural scale representation afforded by our construction. In contrast, the scales associated with the lifting periodogram are discrete and unevenly distributed, as dictated by the time-sampling regime that governs the scale construction. Therefore, Knight et al. (2012) must carry out regression over scale to produce a continuous-like lifting spectral representation. We contrast the lifting periodogram in Fig. 5 (top left) to its continuous spectral counterpart (top right).

Reassuringly, all estimators capture the activity burst time localization. Our new spectrum captures relevant frequency behaviour in similar regions to Nason et al. (2000), which

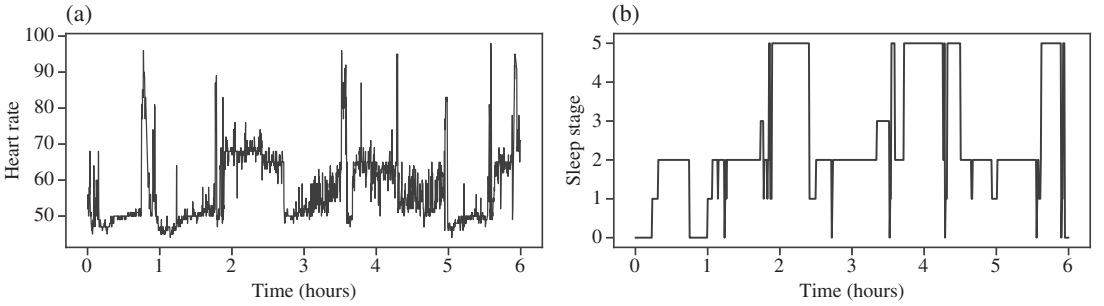

Fig. 6. (a) Heart rate of a sleeping individual and (b) the corresponding sleep stage. Data from Walch (2019).

benefits from complete data, whereas ours does not, whilst also providing a continuous scale representation. However, these features are less clearly elicited by the lifting scale periodogram mapping, possibly because of scale-smoothing artefacts. For example, the top left plot of Fig. 5 does not seem to have any noticeable power at any scale around time 0100, whereas the estimate from Nason et al. (2000) and our methods (right column) do so, particularly at mid-to-coarser scales.

5.2. *Application to heart rate series*

The data consist of $n(T) = 4007$ irregularly spaced observations spanning 21 569 s (approximately 6 h). Mostly, the intervals between consecutive observations range from 1–10 s, though a few points are separated by longer gaps, up to 130 s. The discrete-time methods of Nason et al. (2000) simply cannot be used, while existing methods (Knight et al., 2012), besides being computationally demanding, are unable to maintain a theoretical frame of reference for truly irregular data, as opposed to sampled with missingness. Our methods are the only ones to our knowledge that can produce a genuine estimate in reasonable time.

Walch et al. (2019) used wearable devices to collect heart rate data from several individuals for a study aimed at predicting sleep stages. The data are available from Walch (2019). The sleep stages are labelled 0 to 5, representing progressively deeper stages of sleep: stage 0 corresponds to wakefulness, stages 1 to 3 correspond to nonrapid eye movement sleep and stage 5 corresponds to rapid eye movement sleep. Figure 6 displays the recordings of the chosen individual.

Because of the notable nonstationarity exhibited by the heart rate time series, Haar wavelets are used in our analysis. Whilst our approach directly models the series using the locally stationary wavelet framework, an alternative approach might involve separating a local mean component from the series first, for example by extending the methods of von Sachs & MacGibbon (2000), Vogt (2012) and McGonigle et al. (2022) to continuous time. The raw wavelet periodogram is estimated at evenly spaced locations every 5 s, across scales ranging from 10 to 10 000 with symmetric boundary conditions. The log periodogram is smoothed using the least asymmetric wavelets of Daubechies (1992) with two vanishing moments. Finally, the iterative soft-thresholding algorithm of Daubechies et al. (2004) computes a spectral estimate, from which we obtain the local autocorrelation.

Figure 7 shows that the spectrum is concentrated primarily at coarser scales in regions where the heart rate corresponds to deeper sleep stages, such as the nonrapid eye movement stage. We can see three distinct peaks, which approximately align with the sleep stage. A less clear relationship might be seen in the local autocorrelation estimate. This suggests that spectral information at wider scales can be associated with deeper sleep stages.

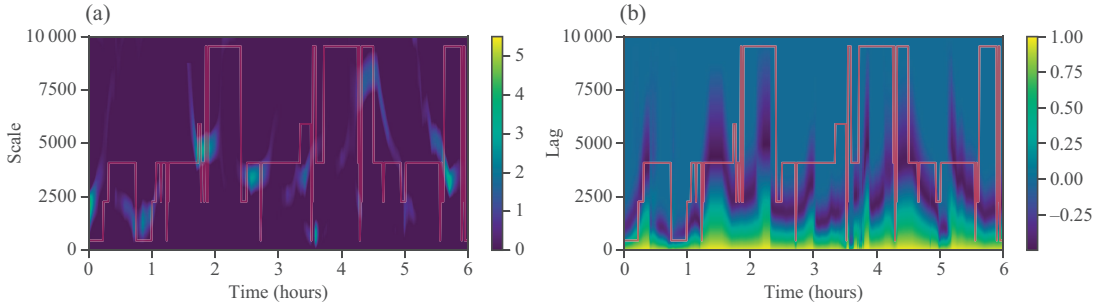

Fig. 7. Estimates for the heart rate data, with the sleep stage series overlaid on top in red. (a) Evolutionary wavelet spectrum. (b) Local autocorrelation.

6. Discussion

For future work, promising directions could be improved smoothing techniques for the raw wavelet periodogram, or alternative approaches to evolutionary wavelet spectrum estimation. This could involve extending the framework in § S4.3 of the Supplementary Material or adapting other algorithms, such as those of Daubechies et al. (2008, 2016), Beck & Teboulle (2009) and Loris et al. (2009). One could investigate the impact of discrete approximations on continuous-time models, particularly in relation to aliasing. Aliasing occurs when high-frequency components contaminate lower frequencies due to a sampling rate that is too low. Locally stationary wavelet processes have already shown promise in detecting its absence (Eckley & Nason, 2018; Palasciano & Nason, 2023). Continuous-time models could yield further insights into these behaviours, especially since the implementation of such methods inherently requires some form of downsampling.

Acknowledgement

We would like to thank the associate editors, reviewers and Professor Almut Veraart of Imperial College London for helpful discussions and feedback on an earlier version of this work.

Supplementary material

The Supplementary Material provides background material, autocorrelation wavelet and inner product kernels, a comparison of local and standard autocovariances on a simple example, practical considerations, an extension of the Haar moving average example and proofs. The code can be found at `https://github.com/henrypalasciano/CLSWP`.

References

Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*. Cambridge: Cambridge University Press.

Barndorff-Nielsen, O. E., Espen Benth, F. & Veraart, A. E. D. (2018). *Ambit Stochastics*. Cham: Springer.

Beck, A. & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2**, 183–202.

Bertero, M., Boccacci, P. & De Mol, C. (2021). *Introduction to Inverse Problems in Imaging*. Boca Raton, FL: CRC Press.

Beylkin, G. & Saito, N. (1992). Wavelets, their autocorrelation functions, and multiresolution representations of signals. In *Intelligent Robots and Computer Vision XI: Biological, Neural Net, and 3D Methods*, vol. **1826**, pp. 39–50. Washington, DC: International Society for Optics and Photonics.

Bitter, A., Stelzer, R. & Ströh, B. (2023). Continuous-time locally stationary time series models. *Adv. Appl. Prob*. **55**, 965–98.
Brockwell, P. J. (2001). Continuous-time ARMA processes. In *Handbook of Statistics*, vol. **19**, pp. 249–76. Amsterdam: Elsevier.
Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer.
Brockwell, P. J., Davis, R. A. & Yang, Y. (2007). Continuous-time Gaussian autoregression. *Statist. Sinica* **17**, 63–80.
Brockwell, P. J., Hyndman, R. J. & Grunwald, G. K. (1991). Continuous time threshold autoregressive models. *Statist. Sinica* **1**, 401–10.
Calvello, E., Reich, S. & Stuart, A. M. (2024). Ensemble Kalman methods: a mean field perspective. *arXiv:* 2209.11371v3.
Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.
Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics*, vol. **30**, Ed. Rao T. S., Rao S. S. and Rao C. R., pp. 351–413. Amsterdam: Elsevier.
Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.
Daubechies, I., Defrise, M. & De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math*. **57**, 1413–57.
Daubechies, I., Defrise, M. & De Mol, C. (2016). Sparsity-enforcing regularisation and ISTA revisited. *Inverse Probl*. **32**, 104001.
Daubechies, I., Fornasier, M. & Loris, I. (2008). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *J. Fourier Anal. Appl*. **14**, 764–92.
de Menezes, L. M., Houllier, M. A. & Tamvakis, M. (2016). Time-varying convergence in European electricity spot markets and their association with carbon and fuel prices. *Energ. Policy* **88**, 613–27.
Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaption via wavelet shrinkage. *Biometrika* **81**, 425–55.
Eckley, I. A. & Nason, G. P. (2018). A test for the absence of aliasing or local white noise in locally stationary wavelet time series. *Biometrika* **105**, 833–48.
Embleton, J., Knight, M. I. & Ombao, H. (2022). Multiscale spectral modelling for nonstationary time series within an ordered multiple-trial experiment. *Ann. Appl. Statist*. **16**, 2774–803.
Fryzlewicz, P. (2005). Modelling and forecasting financial log-returns as locally stationary wavelet processes. *J. Appl. Statist*. **32**, 503–28.
Fryzlewicz, P. & Delouille, V. (2005). A data-driven HAAR-FISZ transform for multiscale variance stabilization. In *IEEE/SP 13th Workshop on Statistical Signal Processing*, pp. 539–44. Piscataway, NJ: IEEE Press.
Fryzlewicz, P. & Nason, G. P. (2006). Haar–Fisz estimation of evolutionary wavelet spectra. *J. R. Statist. Soc. B* **68**, 611–34.
Hargreaves, J., Knight, M. I., Pitchford, J., Oakenfull, R., Chawla, S., Munns, J. & Davis, S. (2019). Wavelet spectral testing: Application to nonstationary circadian rhythms. *Ann. Appl. Statist*. **13**, 1817–46.
Killick, R., Knight, M. I., Nason, G. P. & Eckley, I. A. (2020). The local partial autocorrelation function and some applications. *Electron. J. Statist.* **14**, 3268–314.
Knight, M. I. & Nason, G. P. (2009). A 'nondecimated' lifting transform. *Statist. Comp*. **19**, 1–16.
Knight, M. I., Nunes, M. A. & Nason, G. P. (2012). Spectral estimation for locally stationary time series with missing observations. *Statist. Comp*. **22**, 877–95.
Kress, R. (1999). *Linear Integral Equations*. New York: Springer.
Loris, I., Bertero, M., De Mol, C., Zanella, R. & Zanni, L. (2009). Accelerating gradient projection methods for ℓ_1-constrained signal recovery by steplength selection rules. *App. Comp. Harm. Anal*. **27**, 247–54.
Lucchese, L., Pakkanen, M. S. & Veraart, A. E. D. (2023). Estimation and inference for multivariate continuous-time autoregressive processes. *arXiv:* 2307.13020v1.
McGonigle, E. T., Killick, R. & Nunes, M. A. (2022). Trend locally stationary wavelet processes. *J. Time Ser. Anal*. **43**, 895–917.
Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philos. Trans. R. Soc. A* **209**, 415–46.
Mikosch, T. & Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Rev. Econ. Statist*. **86**, 378–90.
Nason, G. P. (2008). *Wavelet Methods in Statistics with R*. New York: Springer.
Nason, G. P. (2024). wavethresh: Wavelets statistics and transforms. R package version 4.7.3.
Nason, G. P., von Sachs, R. & Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. B* **62**, 271–92.
Neumann, M. H. (1996). Spectral density estimation via non-linear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal*. **17**, 601–33.
Neumann, M. H. & von Sachs, R. (1995). Wavelet thresholding: beyond the Gaussian I.I.D. situation. In *Wavelets and Statistics*, Ed. Antoniadis A. and Oppenheim G., pp. 301–29. New York: Springer.

Nowotarski, J., Tomczyk, J. & Weron, R. (2013). Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energ. Econ*. **39**, 13–27.
Palasciano, H. A. & Nason, G. P. (2023). A test for the absence of aliasing or white noise in two-dimensional locally stationary wavelet processes. *Statist. Comp*. **33**, 108.
Percival, D. B. & Walden, A. T. (2020). *Spectral Analysis for Univariate Time Series*. Cambridge: Cambridge University Press.
Priestley, M. B. (1983). *Spectral Analysis and Time Series*, vols. **I–II**. London: Academic Press.
Rajput, B. S. & Rosinski, J. (1989). Spectral representations of infinitely divisible processes. *Prob. Theory Rel. Fields* **82**, 451–87.
Romano, J. D. & Cornish, N. J. (2017). Detection methods for stochastic gravitational-wave backgrounds: a unified treatment. *Living Rev. Relativ.* **20**, 1–223.
Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40**, 2601–33.
von Sachs, R. & MacGibbon, B. (2000). Non–parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist*. **27**, 475–99.
von Sachs, R. & Schneider, K. (1996). Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *App. Comp. Harm. Anal*. **3**, 268–82.
Walch, O. (2019). Motion and heart rate from a wrist-worn wearable and labeled sleep from polysomnography (version 1.0.0). *PhysioNet*, https://doi.org/10.13026/hmhs-py35.
Walch, O., Huang, Y., Forger, D. & Goldstein, C. (2019). Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep* **42**, zsz180.
Walsh, J. B. (1986). *An Introduction to Stochastic Partial Differential Equations*. Berlin: Springer.
Wang, Y. & He, P. (2023). Comparisons between fast algorithms for the continuous wavelet transform and applications in cosmology: the 1D case. *RAS Techniques and Instruments* **2**, 307–23.
Winkelmann, L. (2016). Forward guidance and the predictability of monetary policy: a wavelet-based jump detection approach. *Appl. Statist*. **65**, 299–314.

[*Received on* 20 *May* 2024. *Editorial decision on* 27 *January* 2025]