Electronic Journal of Statistics
Vol. 16 (2022) 4398–4448
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2044

Modelling time-varying first and second-order structure of time series via wavelets and differencing*

Euan T. McGonigle[1,2]

[1]*STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, UK*

[2]*School of Mathematics, University of Bristol, Bristol, UK*
e-mail: euan.mcgonigle@bristol.ac.uk

Rebecca Killick[3]

[3]*Department of Mathematics and Statistics, Lancaster University, Lancaster, UK*
e-mail: r.killick@lancaster.ac.uk

Matthew A. Nunes[4]

[4]*Department of Mathematical Sciences, University of Bath, Bath, UK*
e-mail: m.a.nunes@bath.ac.uk

Abstract: Most time series observed in practice exhibit time-varying trend (first-order) and autocovariance (second-order) behaviour. Differencing is a commonly-used technique to remove the trend in such series, in order to estimate the time-varying second-order structure (of the differenced series). However, often we require inference on the second-order behaviour of the original series, for example, when performing trend estimation. In this article, we propose a method, using differencing, to jointly estimate the time-varying trend and second-order structure of a nonstationary time series, within the locally stationary wavelet modelling framework. We develop a wavelet-based estimator of the second-order structure of the original time series based on the differenced estimate, and show how this can be incorporated into the estimation of the trend of the time series. We perform a simulation study to investigate the performance of the methodology, and demonstrate the utility of the method by analysing data examples from environmental and biomedical science.

MSC2020 subject classifications: Primary 62M10; secondary 62G05.

Keywords and phrases: Differencing, locally stationary time series, trend estimation, wavelet thresholding, wavelet spectrum.

Received June 2021.

arXiv: 2108.07550

*E.T. McGonigle gratefully acknowledges financial support from EPSRC and Numerical Algorithms Group Ltd. via The Smith Institute i-CASE award No. EP/R511997/1. R. Killick gratefully acknowledges funding from EP/R01860X/1.

1. Introduction

Time series data can often possess complex and dynamic characteristics. Most commonly-encountered time series in practice are nonstationary – the mean and autocovariance of the series vary over time. Modelling how these properties change over time is crucial for making inference on the data. Nonstationary time series arise in a variety of applications, for example in environmental sciences [12], climatology [7], and financial time series [25]. In this article, we consider a time series model of the form

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \epsilon_{t,T}, \quad 0 \leq t < T, \tag{1}$$

where $\mu : [0,1] \to \mathbb{R}$ is a non-parametric deterministic trend function, and $\epsilon_{t,T}$ is a locally stationary wavelet (LSW) process with $\mathbb{E}(\epsilon_{t,T}) = 0$. This model accounts for nonstationarity in *both* the first and second-order structure of the time series: the time-varying mean function is encapsulated in the trend term, while the time-varying second-order behaviour is described by the LSW process term. A thorough explanation of Model (1), including necessary background on LSW processes, is given in Section 2.

First and second-order estimation of a time series are most commonly considered separately, rather than jointly within the one common methodological framework. However, there is an emerging literature where the mean function is estimated as well as parameters responsible for second-order nonstationarity. For example [28] consider time varying AR(p)-processes where the mean function and AR coefficients are estimated jointly within the same procedure. Other methods, such as [6] which focuses on the semi-parametric setting, consider results where the mean function is time-varying and/or estimated. [13] test for increases and decreases in trend in the presence of stationary time series errors, while [9] consider the problem of prediction in locally stationary time series.

The problem of performing inference on the mean function in the presence of nonstationary second-order structure is a highly challenging one. In [32], for example, the authors describe a method using wavelet thresholding, however the threshold used is data-driven and depends on the nonstationary second-order structure, which is ultimately unknown. [31] employs kernel-based methods to estimate a smooth non-parametric trend function in the presence of locally stationary errors. [10] test for relevant changes in the mean of nonstationary processes. Similarly, there is less attention in the literature on nonstationary second-order estimation in the presence of a non-zero mean function.

In order to estimate a nonstationary second-order structure a zero-mean process is often assumed. One of the most well-known methods for removing the trend in a time series is differencing: see for example [3] and [27]. The time series $\{X_t\}$ can be, for example, first-differenced to obtain a new time series, $\{\nabla X_t = X_t - X_{t-1}\}$. Differencing aims to remove a trend without the need to estimate any parameters (whose estimation often includes an assumption of stationary errors). In the time series literature, it is often the case that inference is made on the properties of the differenced time series. In contrast, this article

proposes a method to jointly estimate the time-varying trend and second-order structure of the original time series, by employing the commonly-used strategy of differencing a time series in order to remove the trend.

Our approach to this problem can be summarised as follows. By differencing the time series to remove the trend, we can estimate the appropriate second-order quantities of interest of the locally stationary wavelet part of Model (1). This is achieved by considering the effect of differencing on the second-order properties of the series in order to develop an appropriate estimation procedure. Expanding upon the rigorous theory developed in [21], we obtain results on the consistent estimation of the second-order structure using our modified estimation strategy for the original time series. Using this estimate, we discuss a wavelet thresholding technique to estimate the trend function $\mu(t/T)$ in a principled manner, by taking into account the time-varying second-order behaviour. Our methodology thus enables the joint estimation of the first and second-order structure of nonstationary time series. The applications demonstrate the added utility that estimating the second order structure of the original time series brings.

The rest of the article is organised as follows. In Section 2 we introduce the time series model which we focus on in this article, describe key assumptions, and discuss necessary background to LSW processes. In Section 3, we analyse the effect that differencing has on the spectral structure of a time series, and explain the intuition behind our methodology. Furthermore, we describe the methodology for consistent estimation of the second-order structure in the presence of trend, and in Section 4 we use this estimate in order to estimate the trend of the series. Simulation studies assessing the method's performance are given in Section 5. In Section 6, we apply our framework to a data example, demonstrating the utility of the method, while concluding remarks are given in Section 7. All proofs and additional simulation and data application results are contained within the supplementary material.

2. Wavelets and model formulation

In this section we introduce the modelling paradigm that we will use, as well as explaining the necessary background concepts.

In the LSW framework, wavelets act as building blocks, analogous to the Fourier exponentials in the classical Cramér representation for stationary processes. Briefly, wavelets are oscillatory basis functions which provide efficient (sparse) multiscale representations of signals. Wavelets are useful in estimating time-varying quantities, especially nonstationary characteristics, due to these attractive properties. For an overview of wavelet techniques, see for example [17] or [30]. For a function $f \in L^2(\mathbb{R})$, we have the expression

$$f(x) = \sum_{k\in\mathbb{Z}} c_J \phi_{J,k}(x) + \sum_{j\leq J} \sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x),$$

where the wavelet $\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)$ is a dilated and translated version

of a (mother) wavelet $\psi(x)$ and similarly for the father wavelet $\phi(x)$. The coefficients $d_{j,k}$ at location k and scale j represent the oscillatory behaviour of the signal f at a particular frequency, whereas the coefficients $c_{j,k}$ give information about the mean behaviour of the signal at different scales j.

Next, we define the discrete wavelets that make up the building blocks of the LSW model. Let ψ be a compactly supported wavelet, for example any within the Daubechies family [8]. Denote by $\{h_k, g_k\}$ the low- /high- pass filter pair associated with ψ. The pair $\{h_k, g_k\}$ filter a signal into low and high frequency components respectively and form a pair of quadrature mirror filters: the frequency response of $\{h_k\}$ is the mirror image, about the value $\pi/2$, of $\{g_k\}$. Letting N_h be the number of non-zero values of $\{h_k\}$, the filters are related by the equation $g_k = (-1)^k h_{N_h - k}$ [17, Equation (2.52)].

Define $L_j = (2^{-j} - 1)(N_h - 1) + 1$. The discrete wavelets at a given scale $j \in \mathbb{Z}^-$, as discussed in [21], are defined as the vectors $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,L_j - 1})$, where

$$\psi_{-1,n} = \sum_k g_{n-2k}\delta_{0k} = g_n, \qquad n = 0, \ldots, L_{-1} - 1, \tag{2}$$

$$\psi_{j-1,n} = \sum_k h_{n-2k}\psi_{j,k}, \qquad n = 0, \ldots, L_j - 1, \tag{3}$$

where δ_{0k} is the Kronecker delta. The discrete father wavelet is defined similarly using the associated low-pass filter $\{h_k\}$. The simplest example of a wavelet basis is the Haar wavelet, which is given by

$$\psi^H_{j,k} = 2^{-j/2}\mathbb{I}\left(0 \le k \le 2^{-j-1} - 1\right) - 2^{-j/2}\mathbb{I}\left(2^{-j-1} \le k \le 2^{-j} - 1\right),$$

where $j = \{-1, -2, -3, \ldots\}$ and $k \in \mathbb{Z}$.

2.1. Model definition

Below, we define the *Trend* Locally Stationary Wavelet (T-LSW) model, which is composed of a deterministic Lipschitz continuous trend component and a locally stationary wavelet component. Our T-LSW model, developing on the theory of locally stationary wavelet processes of [21], allows for simultaneous inference on the time-varying mean and autocovariance of a time series.

Definition 2.1. A *trend locally stationary wavelet (T-LSW) process* $\{X_{t,T}\}$, $t = 0, \ldots, T-1$, and $T = 2^J \ge 1$ for $J \in \mathbb{N}$ is a doubly-indexed stochastic process with the following representation in the mean-square sense:

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \sum_{j=-\infty}^{-1}\sum_{k\in\mathbb{Z}} w_{j,k;T}\psi_{j,k-t}\xi_{j,k}, \tag{4}$$

where $\{\xi_{j,k}\}$ is a random, uncorrelated, zero-mean orthonormal increment sequence, $\{w_{j,k;T}\}$ is a set of amplitudes, and $\{\psi_{j,k}\}_{j,k}$ is a set of discrete non-decimated wavelets defined using the discrete wavelets given in Equations (2) and (3). The quantities in representation (4) possess the following properties:

1. The function $\mu : [0,1] \in \mathbb{R}$ is Lipschitz continuous with constant $K > 0$, i.e
$$\left|\mu\left(\frac{t}{T}\right) - \mu\left(\frac{s}{T}\right)\right| \leq \frac{K}{T}, \ \forall\ s,t \in [0,T].$$
2. There exists, for each $j \leq -1$, a Lipschitz continuous function $W_j(z)$ for $z \in (0,1)$ which satisfies the following properties:
$$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1);$$
the Lipschitz constants L_j associated to W_j are uniformly bounded in j and satisfy $\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty$. There exists a sequence of constants C_j such that for each T
$$\sup_k \left|w_{j,k;T} - W_j\left(\frac{k}{T}\right)\right| \leq \frac{C_j}{T},$$
where for each $j \leq -1$ the supremum is over $k = 0, \ldots, T-1$, and where the sequence $\{C_j\}$ satisfies $\sum_{j=-\infty}^{-1} C_j < \infty$.

The model imposes the same assumptions on the LSW component as in [21], allowing for locally stationary second-order structure. Equation (4) uses discrete non-decimated wavelets, as opposed to standard discrete wavelets. Discrete non-decimated wavelets can be shifted to any location defined by the finest resolution level, not just by dyadic shifts of 2^{-j} as is the case for discrete decimated wavelets. Discrete non-decimated wavelets are no longer orthogonal, but an overcomplete collection of shifted vectors. The use of shift-equivariant non-decimated wavelets ensures that the representation in (4) includes a large class of correlated process, including all short memory stationary processes.

Model (4) also permits nonstationary first-order behaviour by incorporating a smooth mean function μ. Imposing a Lipschitz continuous trend assumption is not overly restrictive, given that trend functions are generally assumed to be smooth and slowly-evolving. In particular, polynomials (restricted to the interval $[0,1]$) are Lipschitz continuous, as are sinusoids. Such an assumption is commonly made in the literature, see for example [31] and [13]. Furthermore, in Section 3.6 we will discuss the case when the trend is not Lipschitz. Lastly, we note that the assumption that $T = 2^J$ is a theoretical one; in practice, using the boundary handling method described in [16], we can analyse time series of any length.

2.2. Background to LSW processes

In the original work of [21], the trend function $\mu(t/T)$ in Equation (4) is assumed to be everywhere zero, which forces the process mean $\mathbb{E}(X_{t,T})$ to be equal to zero for all t. Consequently, within the original LSW framework it is only possible to estimate time-varying second-order structure when the time series does not

exhibit a trend. In order to discuss our proposed methodology in the setting where a trend *is* present, we dedicate the rest of the section to recalling a number of definitions and results from [21], which we will expand upon and adapt to the setting where a trend is present.

The second-order structure of a LSW process is uniquely determined by its spectrum. The evolutionary wavelet spectrum (EWS) of a LSW process is defined as $S_j(z) := |W_j(z)|^2$ for rescaled time $z = k/T \in (0,1)$, and measures the contribution to variance at a particular rescaled time z and scale j. Since the W_j are assumed to be Lipschitz continuous, the spectrum S_j is also Lipschitz continuous, which ensures it evolves slowly over time. Alterations to the LSW model that use different assumptions on the EWS can be found in [11] and [29], which assume respectively that the EWS is piecewise constant and of bounded total variation.

For ease of notation we now drop the dependence on T in the subscripts of the model quantities. The EWS is estimated via the empirical wavelet coefficients of the time series, given by $d_{j,k} := \langle X_t, \psi_{j,k-t}\rangle = \sum_t X_t \psi_{j,k-t}$. As with Fourier approaches, the raw wavelet periodogram $I_k^j := |d_{j,k}|^2$ is a biased, inconsistent estimator of the EWS ([21], Proposition 4):

$$\mathbb{E}\left(I_k^j\right) = \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{5}$$

$$\mathrm{Var}(I_k^j) = 2\left(\sum_l A_{jl} S_l(k/T)\right)^2 + \mathcal{O}(2^{-j}T^{-1}), \tag{6}$$

where the operator $A = (A_{jl})_{j,l<0}$ is given by $A_{jl} := \langle \Psi_j, \Psi_l\rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau)$, and the autocorrelation wavelets are defined by $\Psi_j(\tau) := \sum_{k\in\mathbb{Z}} \psi_{j,k}\psi_{j,k-\tau}$, $j < 0, \tau \in \mathbb{Z}$. Hence, for the vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT\rfloor}\}_{l=-1,\ldots,-J}$, and the vector of corrected periodograms $\mathbf{L}(z) := \{L^j_{\lfloor zT\rfloor}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = A_J^{-1}\mathbf{I}(z)$, where the J-dimensional matrix $A_J := (A_{jl})_{j,l=-1,\ldots,-J}$,

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left(A_J^{-1}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\ z \in (0,1), \tag{7}$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$. The raw wavelet periodogram is first smoothed and then corrected to produce an asymptotically unbiased, consistent estimator. There are several approaches to smoothing for consistency, for example via a running mean as in [18] or wavelet thresholding as in [21]. The correction is performed by premultiplying the (smoothed) raw wavelet periodogram by A_J^{-1}, as motivated by Equation (7). The operator A is shown to have bounded inverse for all Daubechies compactly supported wavelets in [2].

The local autocovariance (LACV) function for a LSW process provides information about the covariance at a rescaled location $z = k/T \in (0,1)$. The LACV, $c(z,\tau)$, of a LSW process with EWS $\{S_j(z)\}$ is defined as $c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau)$, for $\tau \in \mathbb{Z}$, $z \in (0,1)$. The LACV can be thought of as a decomposition of the autocovariance of a process over scales and rescaled time locations. The LACV is estimated by plugging in the smoothed, corrected estimate for the EWS into the definition of the LACV. Using wavelet thresholding

of the EWS estimator, it is shown in [21] that the LACV estimator is consistent. The next section addresses the consistent estimation of these second-order quantities in the presence of first-order nonstationarity.

3. LSW estimation in the presence of trend via differencing

In this section, we discuss methodology for estimation of the evolutionary wavelet spectrum and local autocovariance function of a trend-LSW process. In order to estimate these quantities, we employ the ubiquitous practice of differencing to remove the trend, but crucially correct for the effect of this on the spectrum.

3.1. Using differencing to detrend a time series

One of the most common methods for removing the trend in a time series is differencing, see for example [3] and [27]. The time series $\{X_t\}$ can be, for example, first-differenced to obtain a new time series, $\{\nabla X_t = X_t - X_{t-1}\}$, upon which inference is then performed. One advantage of differencing is that no parameters are estimated in the differencing operation, which is not the case for detrending achieved using an estimator of the trend. An n-th order difference is capable of removing a polynomial trend of degree n from the data.

One of the consequences of differencing is that the second-order statistical properties of the time series in Model (4) will change, sometimes quite drastically. For example, differencing may induce spurious autocorrelation or lower the autocorrelation at certain lags, or change the sparsity structure of the underlying wavelet spectrum. Therefore, it is potentially problematic to directly use the differenced process for inference on the original process. In the context of ARIMA modelling, differencing is performed in order to induce stationarity, and estimation is then performed on the differenced series. However, this reasoning does not hold within our setting: if we difference a nonstationary LSW process, the result will still be a nonstationary process. Due to the potentially complex structure of the LSW process, properties of the differenced series are not necessarily representative of the original series. However, in order to estimate the trend component in (4), we require an estimate of the second-order structure of the original time series, not the differenced series.

It is straightforward to produce an example where spectral behaviour can be completely altered as a consequence of detrending using, for example, first-differencing. Consider the zero-mean LSW process of length $T = 2^{10} = 1024$, defined by the spectrum $S_j(z) = 1$ for $j = -1$, 0 otherwise. This time series is referred to as the scale -1 Haar moving average process, and can be written as $X_t = (\epsilon_t - \epsilon_{t-1})/\sqrt{2}$, where the ϵ_t are independent identically distributed (IID) random variables. Taking the first-difference, we obtain $\nabla X_t = X_t - X_{t-1} = (\epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2})/\sqrt{2}$. Computing the expectation of the raw wavelet periodogram of the differenced time series, we find that $\mathbb{E}(I_k^{-1}) = 5$, and for $j < -1$, $\mathbb{E}(I_k^j) = 3 \times 2^{j+1}$. Therefore, for all $z \in (0, 1)$, the expected value of the LSW estimator at time z is $\mathbb{E}(\mathbf{L}(z)) = \mathbb{E}(A_{10}^{-1}\mathbf{I}(z))$.

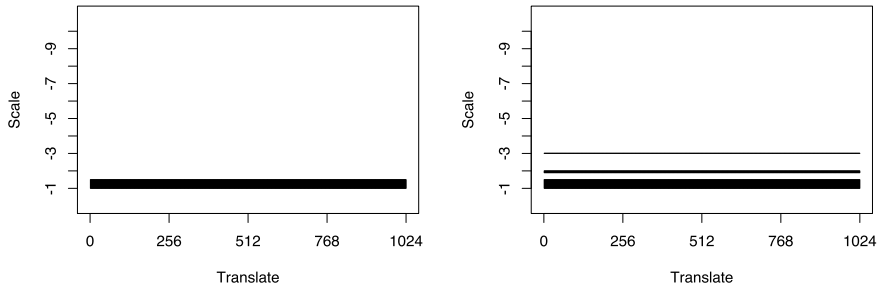

FIG 1. *Left: original spectrum of the scale* -1 *Haar moving average process. Right: expectation of spectral estimate of differenced time series.*

However, a problem arises since $\mathbb{E}(I_k^j) \neq \sum_l A_{jl} S_l(k/T) + \mathcal{O}(T^{-1})$, due to having differenced the time series. In particular, the expectation of the EWS estimate at scale -2 at any time point is given by $\mathbb{E}\left(\sum_l A_{-2,l}^{-1} I_k^l\right) = -0.79 < 0$. Therefore, the expectation of our estimate of the spectrum at level -2 is -0.79, while in the original time series we had $S_{-2}(z) = 0$ for all $z \in (0,1)$. In Figure 1 left, we see a plot of the original spectrum, while on the right, we see the expectation of the corrected periodogram estimate, showing a clear discrepancy.

Differencing can also cause vastly alter the sparsity structure of the spectrum of the error process. The white noise process $Y_t = \epsilon_t$, where the ϵ_t are IID random variables, can be represented as an LSW process using Haar wavelets and a spectrum defined by $S_j(z) = 2^j$ for all $z \in (0,1)$. Differencing this time series gives $\nabla X_t = X_t - X_{t-1} = \epsilon_t - \epsilon_{t-1}$, which is a Haar LSW process with spectrum $S_j(z) = 2$ for $j = -1$, $S_j(z) = 0$ otherwise. This induces autocorrelation in the time series at lag 1, which is similar to what is observed when over-differencing in classical stationary time series. Therefore, it can be seen that we must take into account the differencing of the time series if we wish to say something meaningful about the original series.

3.2. Intuition behind estimation procedure

Denote the LSW component of Model (4) by ϵ_t. Given that the trend component of the Model (4) is Lipschitz continuous, first-differencing the time series yields

$$\begin{aligned}\nabla X_t = X_t - X_{t-1} &= \mu\left(\frac{t}{T}\right) + \epsilon_t - \mu\left(\frac{t-1}{T}\right) - \epsilon_{t-1} \\ &= \epsilon_t - \epsilon_{t-1} + \mathcal{O}(T^{-1}).\end{aligned}$$

Hence, differencing the original series results in a differenced locally stationary wavelet process that is asymptotically zero-mean. We wish to estimate the evolutionary wavelet spectrum of the original time series $\{X_t\}$ using the differenced series $\{\nabla X_t\}$. Proceeding by using the standard estimation procedure of [21] by

taking the squared wavelet coefficients and premultiplying by the inverse of the A matrix, as in Equation (7), is not appropriate, as we have seen in Section 3.1. Denote the empirical non-decimated wavelet coefficients of the first-differenced series by

$$\tilde{d}_{j,k} := \sum_t \nabla X_t \psi_{j,k-t}.$$

We can compute the expectation of the raw wavelet periodogram $\tilde{I}_k^j := |\tilde{d}_{j,k}|^2$, which will yield an analogous result to Equations (5) and (6). Hence, our estimation strategy will follow that of [21]. However, the correction of the raw wavelet periodogram to achieve unbiasedness will require a different correction matrix, to account for the fact that the raw wavelet periodogram is calculated on the differenced series.

Note that, although wavelets themselves are differencing operators, coarser wavelet scales accumulate large bias due to the trend and large filter length. Bias in the coarse scales will also affect the bias in the finer scales, due to the necessity for correcting the estimate using a correction matrix. This can be particularly noticeable for trends that display cusps. Differencing first removes the trend immediately, ensuring that the wavelet transform does not accumulate large bias.

Remark 1. *We can write the differenced process, in the mean-square sense, as*

$$\epsilon_t - \epsilon_{t-1} = \sum_j \sum_k w_{j,k} \psi_{j,k-t} \tilde{\xi}_{j,k},$$

where $\tilde{\xi}_{j,k} = \xi_{j,k} - \xi_{j,k-1}$. *This process satisfies all of the required properties of a standard LSW process except one: the increments are no longer orthonormal. Instead, we have that*

$$\mathrm{Cov}(\tilde{\xi}_{j,k}, \tilde{\xi}_{l,m}) = \begin{cases} 2 & \text{for } j = l, k = m, \\ -1 & \text{for } j = l \text{ and } k = m+1 \text{ or } k+1 = m, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, there is a distinction between the observed time series, which we assume to have LSW errors, and the differenced one, which we do not. This is in contrast to existing approaches utilising differencing for second-order estimation, such as ARIMA models, which instead assume that the differenced series follows the model form rather than the original series.

3.3. Asymptotic behaviour of the differenced raw wavelet periodogram

As motivated by the discussion in the previous section, we estimate the spectrum by calculating the raw wavelet periodogram of the first-differenced time series. We return to the case of general n-th differencing in Section 3.6. For the purpose of theoretical results, we assume that the $\{\xi_{j,k}\}$ in Model (4) are Gaussian. In

practice, this assumption is not required, and in the simulation study in Section 5 we also investigate the case where the $\{\xi_{j,k}\}$ are exponential random variables. Before obtaining the result of the expectation of the raw wavelet periodogram, we define the following two operators, which involve the inner product of the autocorrelation wavelets at lag 1.

Definition 3.1. Let B be the backshift operator such that $B\Psi_j(\tau) = \Psi_j(\tau - 1)$. Define the operator $A^1 = (A^1_{jl})_{j,l<0}$ by

$$A^1_{jl} := \langle \Psi_j, B\Psi_l \rangle = \sum_\tau \Psi_j(\tau)\Psi_l(\tau - 1),$$

and the operator $D^1 = (D^1_{jl})_{j,l<0}$ by

$$\begin{aligned} D^1_{jl} &:= \langle (1-B)\Psi_j, (1-B)\Psi_l \rangle = 2A_{jl} - 2A^1_{jl} \\ &= 2\sum_\tau \Psi_j(\tau)\left(\Psi_l(\tau) - \Psi_l(\tau - 1)\right). \end{aligned}$$

Denote by $A^1_J := (A^1_{jl})_{j,l=-1,\ldots,-J}$ and $D^1_J := (D^1_{jl})_{j,l=-1,\ldots,-J}$ as the corresponding J-dimensional matrices of A^1 and D^1 respectively.

Proposition 3.2. *The matrix D^1_J is invertible.*

Intuitively, it is not surprising that the quantity D^1 will appear when we calculate the expectation of the squared wavelet coefficients of the first-differenced series. Indeed, we have the following result for the asymptotic behaviour of the raw wavelet periodogram of the first-differenced time series.

Proposition 3.3. *Let $\tilde{I}^j_k = |\tilde{d}_{j,k}|^2$ be the wavelet periodogram of the first-differenced time series. Under the assumptions of Model (4),*

$$\mathbb{E}(\tilde{I}^j_k) = \sum_l D^1_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}),$$

$$\mathrm{Var}(\tilde{I}^j_k) = 2\left(\sum_l D^1_{jl} S_l\left(\frac{k}{T}\right)\right)^2 + \mathcal{O}(2^{-j}T^{-1}).$$

Therefore, for the vector of periodograms $\mathbf{I}(z) := \{I^l_{\lfloor zT \rfloor}\}_{l=-1,\ldots,-J}$, and the vector of corrected periodograms $\mathbf{L}(z) := \{L^j_{\lfloor zT \rfloor}\}_{j=-1,\ldots,-J}$ with $\mathbf{L}(z) = (D^1_J)^{-1}\mathbf{I}(z)$,

$$\mathbb{E}\left(\mathbf{L}(z)\right) = \mathbb{E}\left((D^1_J)^{-1}\mathbf{I}(z)\right) = \mathbf{S}(z) + \mathcal{O}(T^{-1}) \qquad \forall\, z \in (0,1),$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$ is the EWS of the original process. Thus we can bias correct the raw wavelet periodogram of the first-differenced time series by using the inverse of D^1_J, instead of the inverse of A_J from [21].

Remark 2. *Proposition 3.3 can be generalised to the lag L differenced time series* $\{\nabla_L X_t := X_t - X_{t-L}\}_{t=L}^{T-1}$ *for any fixed integer L. The expectation of the raw wavelet periodogram of the lag L differenced series at location k is given by*

$$\sum_l D_{jl}^L S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}),$$

where $D_{jl}^L = 2\sum_\tau \Psi_j(\tau)(\Psi_l(\tau) - \Psi_l(\tau - L))$*. Then, the matrix* D_J^L *is invertible by the same argument used to show invertibiltiy of* D_J^1*. This enables the raw wavelet periodogram of the L-differenced series to be corrected using the matrix* $(D_J^L)^{-1}$*. This technique can be used to extend the method to allow for seasonal trends, as shown empirically in Section 5.3.*

3.4. Bounded invertibility of Haar and Shannon operators

In order to achieve an *asymptotically* unbiased estimator of the EWS, we require boundedness of the inverse operator used when performing the bias correction step in the estimation procedure. In the original LSW work of [21], boundedness of the inverse of the operator A was shown in the case of the Haar and Shannon wavelets. The members of the Daubechies compactly supported family of wavelets are characterised by the number of vanishing moments. The Haar wavelet is equivalently the Daubechies extremal phase wavelet with one vanishing moment, while the Shannon wavelet is the limiting wavelet in the Daubechies compactly supported family [4] as the number of vanishing moments tends to ∞. As noted in [33], Haar and Shannon wavelets can therefore be viewed as the lower and upper "extremes" of the family of Daubechies compactly support wavelets. Therefore, proving bounded invertibility in these two cases intuitively suggests that bounded invertibility should hold in the case of all Daubechies' compactly supported wavelets. It is for these two wavelets that we prove bounded invertibility here; extensions to other wavelets are left for future work.

Proposition 3.2 shows that the matrix D_J^1 is invertible with bounded inverse. Asymptotically, however, the inverse operator is unbounded, which mirrors a result in the setting of locally stationary Fourier time series [24, Equation 5]. Intuitively this is expected, since the differencing operator itself is asymptotically non-invertible. We can interpret the expectation result in Proposition 3.3 as a quantification of the effect of differencing on the spectral structure of the time series.

We can account for this theoretical issue by proving bounded invertibility for a related, rescaled operator $P : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$, where the entries of P are given by $P_{jl} = 2^{-j/2} D_{jl}^1 2^{-l/2}$. The use of P is due to the fact that the inverse of the operator D^1 is unbounded, so our approach is to work with the operator P, which we can show has a bounded inverse. Showing that P possesses a bounded inverse enables us to show theoretical consistency properties of the wavelet periodogram-based estimator. Hence, we show for the Haar and Shannon

wavelet families that P possesses a bounded inverse, which enables consistent estimation of the EWS for Model (4). Note that, in the practical implementation of the methodology, we still use the inverse of the D_J^1 matrix for correcting the raw wavelet periodogram as it is an invertible matrix (Proposition 3.2).

Theorem 3.4. *Let $\lambda_{min}(P)$ denote the smallest eigenvalue of P, where the entries of P are given by $P_{jl} = 2^{-j/2} D^1_{jl} 2^{-l/2}$. Then, for the Haar and Shannon wavelet families, there exists $\delta > 0$ such that $\lambda_{min}(P) \geq \delta$ and hence $||P^{-1}|| < \infty$. That is, P is positive-definite and has a bounded inverse.*

3.5. Smoothing and estimation theory

As in the original LSW model, the wavelet periodogram is not a consistent estimator and must be smoothed. Smoothing to achieve consistency can be performed, for example, via wavelet thresholding or a running mean. For brevity, we provide theoretical results for wavelet thresholding, building on known results in the literature.

In order to utilise the result on boundedness of the operator P^{-1} in Theorem 3.4, we rewrite the formula for the expectation of the wavelet periodogram. We can express the expectation in terms of P and a scaled version of S_j, given by $\tilde{S}_j = 2^{j/2} S_j$, by rescaling the periodogram appropriately. Concretely, consider the auxiliary process $\epsilon_t = \sum_{j,k} \tilde{w}_{jk} \tilde{\psi}_{j,k-t} \xi_{lm}$, where $\tilde{w}_{jk} = 2^{j/4} w_{jk}$ and $\tilde{\psi}_{j,k-t} = 2^{-j/4} \psi_{j,k-t}$. Then, the expectation of the raw wavelet periodogram (with respect to the rescaled wavelet $\tilde{\psi}_{j,k-t}$) is given by

$$\mathbb{E}(\tilde{I}^j_k) = \sum_l P_{jl} \tilde{S}_l(k/T) + \mathcal{O}(T^{-1}),$$

where $\tilde{S}_j(k/T) = 2^{j/2} S_j(k/T)$ and $P_{jl} = 2^{-j/2} D^1_{jl} 2^{-l/2}$.

Then, to achieve consistency we take a similar approach to [21]. For each fixed scale j, the rescaled periodogram $\tilde{I}^j_k$ of a Gaussian LSW process (which is scaled χ^2-distributed) is smoothed as a function of $z = k/T$ using, for example, discrete wavelet transform (DWT) shrinkage or translation invariant (TI) denoising of [5]. Using the DWT, smoothing is performed with respect to an orthonormal wavelet basis $\{\phi'_{r_0,s}(z), \psi'_{rs}(z)\}$ of $L^2([0,1])$. Here, $\psi'_{rs}(z) = 2^{r/2} \psi'(2^r z - s)$, where r_0 is the coarsest scale analysed and $s = 0, \ldots, 2^r - 1$. Smoothing is achieved using non-linear thresholding of the empirical wavelet coefficients $\hat{v}^j_{rs}$ of $\tilde{I}^j_k$.

Explicitly, as described in [33], for levels $j = -1, \ldots, -J$, the wavelet expansion of the scaled periodogram can be written as

$$\tilde{I}^j_{\lfloor zT \rfloor} = \sum_r \sum_s v^j_{rs} \psi'_{rs}(z),$$

where the "true" wavelet coefficients are given by $v^j_{rs} = \int_0^1 \tilde{I}^j_{\lfloor zT \rfloor} \psi'_{rs}(z)\, dz$. As in [33], we employ a slight abuse of notation, with $\psi'_{r_0-1,s} = \phi'_{r_0,s}$, in order

to include the scaling coefficient at the coarsest scale r_0 of the second wavelet scheme. The empirical analogues of the wavelet coefficients are given by

$$\hat{v}^j_{rs} = T^{-1} \sum_{n=0}^{T-1} \tilde{I}^j_{n,T} \psi'_{rs}(n/T), \quad \text{for } r = r_0, \ldots, \log_2(T), \quad s = 0, \ldots, 2^r - 1. \tag{8}$$

Analogously, we can build non-decimated wavelet coefficients for TI denoising. Let $\psi'_r(z) = 2^{r/2}\psi'(2^r z)$, and let

$$\hat{v}^j_{rs} = T^{-1} \sum_{n=0}^{T-1} \tilde{I}^j_{n,T} \psi'_{rs}((n-s)/T), \quad \text{for } r = r_0, \ldots, \log_2(T), \quad s = 0, \ldots, T-1.$$

Then, denoising is achieved by applying non-linear hard wavelet thresholding to the wavelet coefficients $\hat{v}^j_{rs}$. The resulting estimator is obtained by inverting the wavelet transform using only the coefficients which remain after thresholding:

$$\hat{I}^j_{\lfloor zT \rfloor} = \sum_r \sum_s \tilde{v}^j_{rs} \psi'_{rs}(z), \quad z \in (0,1),$$

where $\tilde{v}^j_{rs} = \hat{v}^j_{rs} \mathbb{I}(|\hat{v}^j_{rs}| > \lambda)$ are the hard thresholded wavelet coefficients with threshold λ. The particular threshold is specified in Theorem 3.5, as is the appropriate set of indices r over which to perform the summation.

The theoretical argument to consistently estimate $S_j(k/T)$ is thus as follows. First, smooth the rescaled wavelet periodogram at each scale j using wavelet thresholding. Next, use the (bounded) P-inverse matrix to correct the smoothed, rescaled periodogram. Finally, multiply the estimate at each scale by $2^{-j/2}$, since $S_j(k/T) = 2^{-j/2}\tilde{S}_j(k/T)$, yielding the final estimator, $\widehat{S}_j(k/T)$, of $S_j(k/T)$. Using the hard threshold $\lambda(j,r,s,T)^2 = \mathrm{Var}(\hat{v}^j_{rs}) \log^2(T)$ when smoothing the periodogram via wavelet thresholding, we can show that the smoothed, corrected estimate $\widehat{S}_j(z)$ is consistent with respect to the L_2 error.

Theorem 3.5. *Let ψ' be a wavelet of bounded variation, with $2^r = o(T)$ for wavelet coefficients $\hat{v}^j_{rs}$. For a Gaussian trend-LSW process and using the threshold $\lambda^2(j,r,s,T) = \mathrm{Var}(\hat{v}^j_{rs}) \log^2(T)$, for each fixed j,*

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = \mathcal{O}\left(2^{-j} T^{-2/3} \log^2(T)\right). \tag{9}$$

The rate obtained in Equation (9) is a consequence of known results on wavelet thresholding estimators, utilised in [22], and the multiplication of $2^{-j/2}$ that occurs in the estimation procedure. The rate highlights the fact that differencing the time series has resulted in an "information loss", with spectral estimation in coarser scales resulting in slower rates of convergence. As a consequence, we can only consistently estimate the wavelet spectrum for a proportion of the finest scales j. In particular, we have that we need $2^{-j} = \mathcal{O}(T^{-2/3-\delta})$ for some $\delta > 0$ in order for the mean-squared error of the EWS estimator to be $o(1)$. Next, we tackle estimation of the local autocovariance via the EWS estimate.

Proposition 3.6. *Define* $\hat{c}(z,\tau)$ *by replacing* $S_j(z)$ *by* $\widehat{S}_j(z)$ *in the equation for the local autocovariance and replacing the lower limit in the sum from* $j=-\infty$ *to* $j=-J_0$, *i.e.*

$$\hat{c}(z,\tau)=\sum_{j=-J_0}^{-1}\widehat{S}_j(z)\Psi_j(\tau).$$

Let $T\to\infty$ *and let* $J_0=\alpha\log_2 T$ *for* $\alpha\in(0,1)$. *Assume that* $S_j(z)\leq D2^{\gamma j}$ *for some positive constant* D, *where* $\gamma=(3\alpha)^{-1}-1/2$. *Then,*

$$\mathbb{E}\left[\int_0^1\left(\hat{c}(z,\tau)-c(z,\tau)\right)^2dz\right]=\mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right),$$

i.e. $\hat{c}(z,\tau)$ *is a consistent estimator of* $c(z,\tau)$ *for each fixed* $\tau\in\mathbb{Z}$, *provided that* $T^{\alpha-2/3}\log^2(T)\to 0$.

We reiterate that the rescaling argument used to achieve consistency in Theorem 3.5 and Proposition 3.6 is performed purely for theoretical reasons; practical considerations for estimation are discussed at the end of the section. The results in Theorem 3.5 and Proposition 3.6 show that in the case where the trend is Lipschitz continuous, we can consistently estimate both the EWS and LACV of the original process using first-order differences and a modified bias correction.

The assumption placed on the decay rate of the EWS in Proposition 3.6 is a purely technical one, utilised in order to ensure mean square consistency of the estimator. The assumption controls for the fact that the local autocovariance is estimated using the finest J_0 scales, instead of across infinite scales which are not available in practice. The specific form of the decay rate is calculated in order to balance with the error rate of the wavelet thresholding procedure.

3.6. n-th order differencing

In some cases, a first-difference may not be enough to remove a trend. Further differencing can be performed, although it is usually only necessary to at most second-difference a time series [1]. If we assume that the $(n-1)$-th derivative of μ is Lipschitz, then the n-th difference of the time series will be (asymptotically) free of trend. We denote the n-th difference of a time series as $\{\nabla^n X_t\}$.

To calculate the expectation of the squared non-decimated wavelet coefficients of the n-th differenced series, we can argue in a similar fashion to the case of first-differencing. Denote the operator $A^n=(A^n_{jl})_{j,l<0}$ by $A^n_{jl}:=\langle\Psi_j,B^n\Psi_l\rangle=\sum_\tau\Psi_j(\tau)\Psi_l(\tau-n)$, and the corresponding J-dimensional matrix $A^n_J:=(A^n_{jl})_{j,l=-1,\ldots,-J}$. The entries of the matrix A^n_J are given by the inner product of the autocorrelation wavelet with the autocorrelation wavelet at lag n. If we difference a time series n times, then the expectation of the squared wavelet coefficients will involve linear combinations of inner product autocorrelation wavelet matrices from lag 0 (i.e. the standard A-matrix) up to lag n (the matrix A^n). The result for the expectation of the wavelet periodogram of the n-th differenced time series mirrors that of the first-differenced series, and is given by the following proposition.

Proposition 3.7. *Let* $\tilde{d}_{j,k} = \sum_t \nabla^n X_t \psi_{j,k-t}$ *be the non-decimated wavelet coefficients of* $\Delta^n X_t$*, and let* $\tilde{I}_k^j := |\tilde{d}_{j,k}|^2$*. If the* $(n-1)$*-th derivative of* μ *is Lipschitz, then*

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l D_{jl}^n S_l(k/T) + \mathcal{O}(T^{-1}),$$

where

$$D_{jl}^n = \binom{2n}{n} A_{jl} + 2\sum_{\tau=1}^{n} (-1)^\tau \binom{2n}{n+\tau} A_{jl}^\tau.$$

Corollary 3.8. *For second differences,*

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l \left(6A_{jl} - 8A_{jl}^1 + 2A_{jl}^2\right) S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).$$

Note that Proposition 3.7 generalises Proposition 4 in [21], in which $n = 0$. As in the case of first-differences, the bias operator D^n does not possess a bounded inverse. Intuitively, for higher order differences, the eigenvalues of D^n decay to 0 at increasingly faster rates. As such, correcting the estimation in a similar fashion as was described in Theorem 3.5 yields much slower rates of convergence.

For the remainder of the section, we focus on the situation where a second-difference suffices to remove the trend. We can show that using first-differences suffices to obtain a consistent spectral estimate, even though the trend has not been fully removed. The first-differences of the trend are Lipschitz continuous, and as such the magnitude of the wavelet coefficients can be bounded using the wavelet characterisation of Hölder spaces [8]. Using this bound, we can bound the error of the first-differenced estimator in terms of the error due to estimation and error due to the squared wavelet coefficients of the first-differenced trend. Using this argument we are able to show consistency of both the spectrum and local autocovariance estimator as follows.

Theorem 3.9. *Assume that the first derivative of* μ *is Lipschitz, and let* $J_1 = \beta \log_2 T$ *for* $\beta \in (0,1)$*. Further assume that the smoothed raw wavelet periodogram is corrected across the finest* J_1 *scales only. Under the same conditions as Theorem 3.5, for each fixed* j*,* $\widehat{S}_j(z)$ *is a consistent estimator of* $S_j(z)$*, provided that* $T^{7\beta-4} \to 0$ *as* $T \to \infty$*, since*

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = \mathcal{O}\left(2^{-j}T^{7\beta-4}\right) + \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right) + \mathcal{O}\left(2^{-j}T^{-\beta}\right).$$

Define $\hat{c}(z,\tau)$ *by replacing* $S_j(z)$ *by* $\widehat{S}_j(z)$ *in the equation for the local autocovariance and replacing the lower limit in the sum from* $j = -\infty$ *to* $j = -J_0$*, i.e.*

$$\hat{c}(z,\tau) = \sum_{j=-J_0}^{-1} \widehat{S}_j(z)\Psi_j(\tau),$$

where $J_0 = \alpha \log_2 T$ *for* $\alpha < \beta \in (0,1)$. *Under the assumptions of Proposition 3.6, provided that* $T^{\alpha+7\beta-4} \to 0$ *and* $T^{\alpha-2/3}\log^2(T) \to 0$, $\hat{c}(z,\tau)$ *is a consistent estimator of* $c(z,\tau)$, *since for each fixed* $\tau \in \mathbb{Z}$,

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] = \mathcal{O}\left(T^{\alpha+7\beta-4}\right) + \mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right) + \mathcal{O}(T^{\alpha-\beta}).$$

Hence, we can still consistently estimate the EWS and LACV via first-order differences when the trend of the time series has a Lipschitz continuous first derivative. Therefore, we argue that in most practical scenarios, it is sufficient to only perform one difference in order to estimate the evolutionary wavelet spectrum and local autocovariance in the presence of a trend.

In practice, we set $J_0 = J_1$ when performing EWS and LACV estimation. That is, we use J_1 scales for estimating the EWS, and the same number of scales for estimating the LACV. Performing estimation in this fashion ensures that the spectral estimate is well-behaved, and is an approach commonly adopted in the LSW literature. We suggest using $J_1 = \lfloor \beta \log_2(T) \rfloor$, with $\beta = 7/10$, motivated by extensive numerical results given in Appendix A.5. Furthermore, this choice is in agreement with other discussion in the literature, see for example [26]. Finally, an algorithmic description of the spectral estimation procedure, where smoothing is carried out either via wavelet thresholding or using a running mean, is detailed in Algorithm 1.

4. Trend estimation using the spectral estimate

In this section, we discuss a wavelet thresholding approach for the estimation of the trend component of Model (4). If a first (or second) difference is capable of removing the trend from a time series, we have shown that we can consistently estimate the time-varying evolutionary wavelet spectrum using the smoothed, corrected raw wavelet periodogram of the differenced time series. We now wish to use this estimate in order to estimate the trend of the time series.

4.1. Wavelet thresholding estimator

The approach we take is to use the spectral estimate directly within a wavelet thresholding estimation procedure. In [32], the authors describe a wavelet thresholding methodology for consistent curve estimation in the presence of locally stationary errors, subject to mild regularity conditions. The authors propose to use a local median absolute deviation pre-estimate for the variance of the wavelet coefficients, which is used in the threshold. We instead incorporate the consistent spectral estimate into the thresholding procedure. This yields an analogous version of Theorem 1 in [32], for the specific case of Lipschitz continuous trend functions.

With a slight abuse of notation, denote the estimated wavelet coefficients of the time series by $\hat{v}_{rs}$. The results in this section apply to the commonly used

Algorithm 1: Spectral Estimation Procedure

Input: Data $\{X_t\}_{t=0}^{T-1}$, spectral estimation wavelet ψ^0, maximum scale J_1, smoothing wavelet ψ' or smoothing bin width parameter W

1. Compute wavelet periodogram of first-differenced time series:

$$\tilde{I}_k^j = \left(\sum_t \nabla X_t \psi_{j,k-t}^0 \right)^2, \; j = -1, \ldots, -J_1.$$

2. Compute the smoothed wavelet periodogram $\hat{I}_k^j$ for $j = -1, \ldots, -J_1$, either by

(A) Wavelet thresholding:

$$\hat{I}_k^j = \sum_r \sum_s \tilde{v}_{rs}^j \psi'_{rs}(k/T),$$

where $\tilde{v}_{rs}^j = \hat{v}_{rs}^j \mathbb{I}(\hat{v}_{rs}^j > \lambda(j,r,s,T))$, with $\hat{v}_{rs}^j$ given in Equation (8), and $\lambda(j,r,s,T) = \hat{\sigma}_j \log(T)$, with $\hat{\sigma}_j$ computed using the median absolute deviation estimator of the $\hat{v}_{rs}^j$.

(B) Running mean :

$$\hat{I}_k^j = \frac{1}{2W+1} \sum_{w=-W}^{W} \tilde{I}_{k+w}^j.$$

3. Compute the corrected, smoothed wavelet periodogram as

$$\hat{\boldsymbol{S}}(k/T) = (D_{J_1}^1)^{-1} \hat{\boldsymbol{I}}_k,$$

where $\hat{\boldsymbol{S}}(k/T) = \{\hat{S}_j(k/T)\}_{j=-1}^{-J_1}$, $\hat{\boldsymbol{I}}_k = \{\hat{I}_k^l\}_{l=-1}^{-J_1}$, and $D_{J_1}^1$ is defined in Definition 3.1.

Output: Spectrum estimate $\{\hat{S}_j(k/T)\}_{j=-1}^{-J_1}$ for $k = 0, \ldots, T-1$.

soft and hard thresholding rules, given respectively by

$$\hat{v}_{rs}^S = \operatorname{sgn}(d_{r,s}) \left(|d_{r,s}| - \lambda \right) \mathbb{I}(|d_{r,s}| > \lambda),$$
$$\hat{v}_{rs}^H = d_{r,s} \mathbb{I}(|d_{r,s}| > \lambda),$$

where $\lambda = \lambda(r,s,T)$ is the threshold, and $\mathbb{I}$ is the indicator function. Asymptotic normality of the empirical wavelet coefficients, established in [32], permits the use of a coefficient-dependent universal threshold $\lambda(r,s,T) = \sigma_{rs}\sqrt{2\log(T)}$, where σ_{rs}^2 is the variance of the wavelet coefficients. This yields the following result for the wavelet thresholding estimator $\hat{\mu}$ obtained using the DWT and the threshold $\lambda(r,s,T)$, with either soft or hard thresholding, calculated in the same way as described in Section 3.5. We note here however, that the result also holds for TI wavelet denoising (outlined in Section 3.5) since the non-decimated wavelet coefficients can be seen as a set of DWT coefficients computed from cyclic shifts of the data.

Proposition 4.1. *Let $\tilde{\psi}$ be a wavelet of bounded variation, with $2^r = o(T)$ for wavelet coefficients $\hat{v}_{rs}$. For a trend-LSW process with Lipschitz continuous trend, and using the threshold $\lambda(r,s,T) = \sigma_{rs}\sqrt{2\log(T)}$, the estimator $\hat{\mu}$*

satisfies

$$\mathbb{E}\left[\int_0^1 \left(\hat{\mu}(z) - \mu(z)\right)^2 dz\right] = \mathcal{O}\left(\left(\frac{\log(T)}{T}\right)^{2/3}\right).$$

Note that this result is subject to mild regularity assumptions on the LSW component of the model, all of which are satisfied. The innovations $\{\xi_{j,k}\}$ are not restricted to be Gaussian, and can for example be exponential, gamma, or inverse Gaussian distributed. To estimate the variance σ^2_{rs}, which is necessary to choose the threshold $\lambda(r, s, T)$, we use an estimate of the variance of the empirical wavelet coefficients. If the LSW process is generated by wavelet ψ^0, and the wavelet used for thresholding is denoted ψ^1, the variance of the empirical wavelet coefficients $d_{j,k} := \sum_t X_t \psi^1_{j,k-t}$ is given by

$$\mathrm{Var}(d_{j,k}) = \sum_l C^{1,0}_{jl} S_l(k/T) + \mathcal{O}(T^{-1}), \tag{10}$$

where $C^{(1,0)}_{jl} = \sum_\tau \Psi^0_j(\tau)\Psi^1_l(\tau)$, and where $\Psi^0_j(\tau)$ and $\Psi^1_j(\tau)$ are autocorrelation wavelets with respect to wavelets ψ^0 and ψ^1. By plugging in the estimate $\widehat{S}_j(z)$, obtained in Section 3.5, into the expression (10), this yields the universal-type threshold $\lambda(r, s, T) = \hat{\sigma}_{r,s}\sqrt{2\log(T)}$, where $\hat{\sigma}^2_{r,s} = \sum_l C^{1,0}_{rl} \widehat{S}_l(s/T)$.

4.2. Practical considerations

In alignment with discussion in [32], in practice we analyse approximately the finest 7/10 scales of the time series, the same as in the spectral estimation procedure. In practice, we have found that applying hard thresholding yields better performance. We recommend the use of translation invariant (TI) thresholding over a standard discrete wavelet transform. We have found that it offers stronger practical performance, in terms of the mean squared error of the estimator. As noted in [17], use of a non-decimated transform ensures that there is "more chance" of the wavelet coefficients picking up the signal of the time series.

Note that it is possible to obtain negative estimates of the variance of the wavelet coefficients, although based upon our empirical analyses this is rare. In this case we replace the negative values by the nearest neighbouring positive value, which was found to have no discernible impact on the trend estimation procedure. We recommend the use of the Daubechies Least Asymmetric wavelet with 4 vanishing moments, as we have found empirically that it works well for estimation purposes and also helps to minimise the number of negative variance estimates. Note that Proposition 4.1 holds for non-Gaussian time series, while theoretical results concerning the second-order estimation require an assumption of normality. In practice, our approach still performs well in the presence of non-Gaussian noise, as we show in the simulation study.

To complete the discussion, we provide algorithmic pseudo code for the trend estimation procedure using TI thresholding in Algorithm 2.

Algorithm 2: Trend Estimation Procedure

Input: Data $\{X_t\}_{t=0}^{T-1}$, spectrum estimate $\{\hat{S}_j(k/T)\}_{j=-1}^{-J_1}$ for $k = 0, \ldots, T-1$, spectrum estimation wavelet ψ^0, trend estimation wavelet ψ^1

1. Compute non-decimated wavelet transform of data:

$$v_{r,s} = \sum_t X_t \psi^1_{r,s-t}, \; j = -1, \ldots, -\log_2(T).$$

2. Compute variance of wavelet coefficients as

$$\hat{\sigma}^2_{r,s} = \mathrm{Var}(v_{r,s}) = \sum_{l=-1}^{-J_1} C^{1,0}_{rl} \widehat{S}_l(s/T).$$

3. Obtain the thresholded wavelet coefficients $\hat{v}_{r,s}$ using the soft or hard thresholding rules

$$\hat{v}^S_{rs} = \mathrm{sgn}(d_{r,s}) \left(|d_{r,s}| - \lambda(r,s,T)\right)\mathbb{I}(|d_{r,s}| > \lambda(r,s,T)),$$
$$\hat{v}^H_{rs} = d_{r,s}\mathbb{I}(|d_{r,s}| > \lambda(r,s,T)),$$

where

$$\lambda(r,s,T) = \hat{\sigma}_{r,s}\sqrt{2\log(T)}, s = -1, \ldots, -J_1.$$

4. Invert the thresholded wavelet coefficients $\hat{v}_{r,s}$ using basis averaging to obtain $\hat{\mu}(t/T)$.

Output: Trend estimate $\hat{\mu}(t/T)$

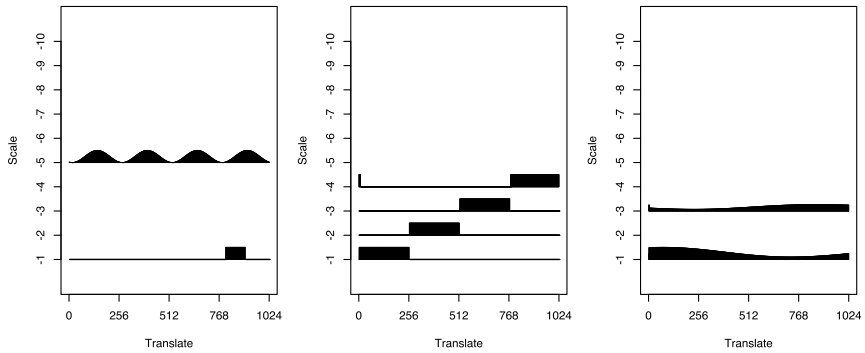

FIG 2. *Left: spectrum S^1, sinusoid with "burst". Centre: S^2, concatenated moving average process. Right: S^3, slowly-evolving fine-scale power.*

5. Simulation study

In this section we illustrate the ability of our proposed methodology to jointly estimate the mean and EWS of a trend-LSW process by performing a simulation study. For each set of simulations, we use the three EWS shown in Figure 2, which represent spectra with distinct characteristics. The spectra are explicitly defined in the supplementary material. Spectrum S^1, studied in [17], displays coarse-scale, slowly-evolving sinusoidal behaviour with a fine-scale burst

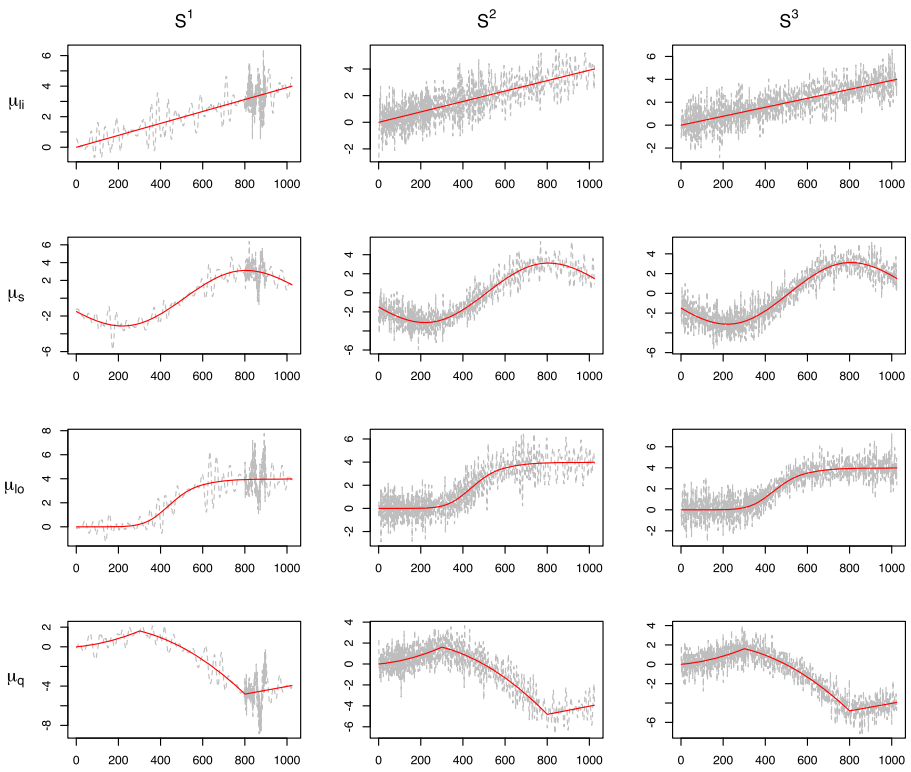

FIG 3. *Example realisations from each trend and spectrum scenario. Dashed line shows time series with true underlying trend shown in solid line. Left: spectrum S^1, sinusoid with "burst". Centre: S^2, concatenated moving average process. Right: S^3, slowly-evolving fine-scale power.*

in power at time point 800. Spectrum S^2 is a concatenation of moving average processes and contains power moving from fine to coarser scales, and was examined in [21]. Spectrum S^3 contains slowly-evolving power at fine scales.

We simulate 100 realisations of time series $\{X_t\}_{t=0}^{T-1}$ of length $T = 2^{10} = 1024$ from LSW processes with those spectra, with different additive trend functions. The trends used in the simulation study are a linear, sinusoidal, logistic and piecewise quadratic trend, denoted in Figure 3 by μ_{li}, μ_s, μ_{lo} and μ_q respectively, and defined explicitly in the supplementary material. These functions are Lipschitz continuous with varying degrees of smoothness, with μ_q also being non-differentiable at two time-points. All LSW processes were simulated using the Daubechies Extremal Phase wavelet with 4 vanishing moments. Similar results, which can be found in the supplementary material, were reported for other wavelets. Example realisations from the simulation study are shown in Figure 3. All simulations were performed in `R` with estimation procedures implemented using modifications to code in the `locits` [19] and `wavethresh` [20] packages.

5.1. Spectral estimation performance

In this simulation we show that by first-differencing to remove the trend, we can obtain an unbiased EWS estimate, which in turn can be used to obtain a trend estimate that performs well. For each realisation, the un-smoothed estimate of the EWS was calculated, which was then used to obtain an averaged estimate for the EWS across the 100 realisations. In alignment with the discussion in Section 3.5, we correct the raw wavelet periodogram across the finest 7 scales.

We report the mean squared error of the averaged spectrum compared with the true spectrum in Table 1. We compare these values with the mean squared error obtained by using the standard LSW estimation procedure of [21]. In this case, no trend is present, the estimate is performed using the original A matrix inverse for bias correction, and no differencing is performed. This is calculated using the `ewspec3` command in the `locits R` package. This is reported in the "None" row in Table 1, and represents the 'best-case' performance with which to compare. We see that despite the presence of a trend, differencing enables us to approximately remove it in order to accurately estimate the spectrum. The estimation error is smaller using our methodology than the "None" case for spectrum 2, while it is marginally worse for the other two spectra. Visual inspection of the resulting estimators shows they are generally satisfactory, with examples given in Figure 4. Note that, while the spectral estimation could also have been performed using a second difference instead of first, this over-differencing results in higher estimation error. A numerical comparison between first and second differences, showing the effect of over-differencing, is given in Appendix A.3.

TABLE 1
Mean squared error comparison for the averaged spectrum estimate, multiplied by 10^3*, across the spectrum and trend scenarios.*

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	3.13	4.88	1.87
Linear	3.32	4.63	2.76
Sine	3.32	4.63	2.76
Logistic	3.32	4.63	2.76
Piece. Quad.	3.32	4.67	2.79

5.2. Trend estimation performance

We next assess the trend estimation procedure. The trend estimate is computed using the Daubechies Least Asymmetric wavelet with 4 vanishing moments, analysing the finest 6 scales. The spectrum estimate used in the thresholding procedure is smoothed with a running mean of bin width size 128. We compute the average mean squared error across the 100 realisations, as well as the standard deviation of the errors. For all simulations, a hard threshold of the form $\hat{\sigma}_{r,s}\sqrt{2\log(1024)}$ is used.

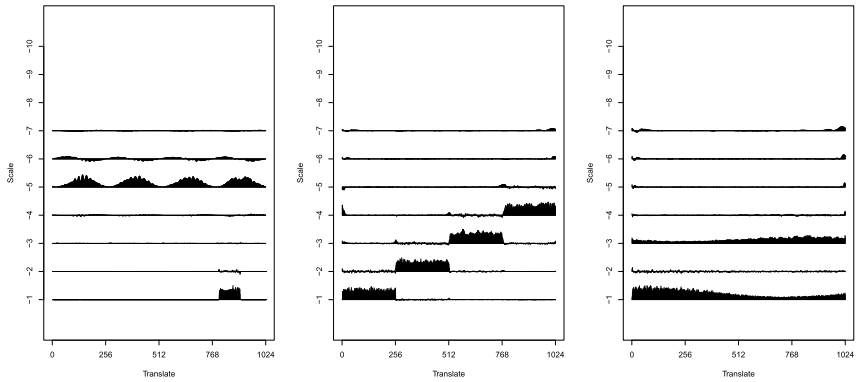

FIG 4. *Example averaged spectral estimates. Left: spectrum 1, linear trend. Middle: spectrum 2, logistic trend. Right: spectrum 3, piecewise quadratic trend.*

We compare our method to two other wavelet-based trend estimation methods. In order to make a fair comparison, we use the same hyper-parameters where applicable. Firstly, we compare to the standard wavelet thresholding method based upon a global, time-independent universal threshold given by $\hat{\sigma}_r\sqrt{2\log(1024)}$, computed using the `wavethresh` package in R. This estimator assumes that the error process is second-order stationary, and is referred to as stationary wavelet thresholding (SWT) in the tables. Secondly, we compare to the wavelet-based trend estimation procedure in [32]. No code is publicly available for this method, and so we have implemented the method utilising `wavethresh` and following the description of the computation of the threshold in Section 2.5 of [32]. In the tables, the abbreviation LSWT refers to our method of locally stationary wavelet thresholding, and MVSWT refers to the wavelet-based method of [32].

We repeat this simulation twice; firstly using Gaussian innovations in the LSW process, and secondly using exponentially distributed innovations. Further results examining the performance in other, non-LSW error scenarios can be found in the supplementary material. The results of the simulation for Gaussian LSW processes are reported in Table 2, while the results for exponentially distributed innovations are shown in Table 3. Values in bold are the lowest values for each trend and spectrum value across the three methods.

From the tables we observe that the performance of our estimator in the presence of exponentially distributed random innovations is comparable to the Gaussian case. The error is consistent across the various time series scenarios, and the standard deviation is low. We see that our method outperforms the two other wavelet-based methods across all trend and spectrum scenarios.

5.3. Extension to seasonal time series

Our approach can be extended empirically to time series that display an additive deterministic seasonal component. From the discussion in Remark 2, if the

TABLE 2
Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using Gaussian innovations.

Trend	Spectrum	Method		
		LSWT	SWT	VSMWT
Linear	1	**0.024** (0.012)	0.519 (0.112)	0.303 (0.100)
	2	**0.030** (0.019)	0.762 (0.078)	0.225 (0.068)
	3	**0.028** (0.024)	0.441 (0.040)	0.160 (0.038)
Sine	1	**0.022** (0.010)	0.526 (0.110)	0.286 (0.109)
	2	**0.026** (0.014)	0.744 (0.075)	0.215 (0.068)
	3	**0.022** (0.017)	0.447 (0.043)	0.156 (0.037)
Logistic	1	**0.023** (0.015)	0.494 (0.109)	0.261 (0.093)
	2	**0.033** (0.019)	0.753 (0.083)	0.226 (0.071)
	3	**0.027** (0.024)	0.449 (0.037)	0.164 (0.035)
Piece. quad.	1	**0.022** (0.012)	0.517 (0.127)	0.290 (0.105)
	2	**0.032** (0.018)	0.720 (0.087)	0.216 (0.072)
	3	**0.028** (0.024)	0.366 (0.046)	0.167 (0.039)

TABLE 3
Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using exponential innovations.

Trend	Spectrum	Method		
		LSWT	SWT	VSMWT
Linear	1	**0.030** (0.018)	0.521 (0.148)	0.305 (0.127)
	2	**0.035** (0.024)	0.779 (0.108)	0.328 (0.092)
	3	**0.040** (0.026)	0.497 (0.052)	0.318 (0.057)
Sine	1	**0.027** (0.025)	0.520 (0.119)	0.308 (0.102)
	2	**0.033** (0.020)	0.782 (0.110)	0.322 (0.085)
	3	**0.037** (0.022)	0.495 (0.060)	0.309 (0.064)
Logistic	1	**0.030** (0.018)	0.508 (0.126)	0.291 (0.108)
	2	**0.036** (0.022)	0.788 (0.098)	0.332 (0.085)
	3	**0.044** (0.031)	0.506 (0.061)	0.327 (0.058)
Piece. quad.	1	**0.031** (0.016)	0.524 (0.116)	0.319 (0.106)
	2	**0.038** (0.022)	0.748 (0.098)	0.326 (0.100)
	3	**0.045** (0.031)	0.432 (0.055)	0.318 (0.055)

time series possesses a (known) number of seasons K, lag K-differencing can be performed to remove the seasonality in order to estimate the EWS of the original time series. To illustrate this, we investigated the performance of the method in the presence of seasonal and smooth trends.

We simulated 100 realisations of time series according to the various trend and spectrum scenarios described in Section 5.1. To each of these scenarios, we added a (monthly) stationary seasonal component with $K = 12$, where the value of each of the 12 seasonal components $K_1, \ldots, K_{12}$ is generated from a uniform random variable on the interval $[0, 10]$. That is, the time series is generated by

$$X_t = \mu\left(\frac{t}{T}\right) + s_t + \epsilon_t,$$

where μ is the smooth trend function, the seasonal component $s_t = K_1$ for $t \mod 12 = 1, \ldots, s_t = K_{12}$ for $t \mod 12 = 0$, and ϵ_t is the LSW component.

Lastly, we considered the case where the seasonal component can be time-

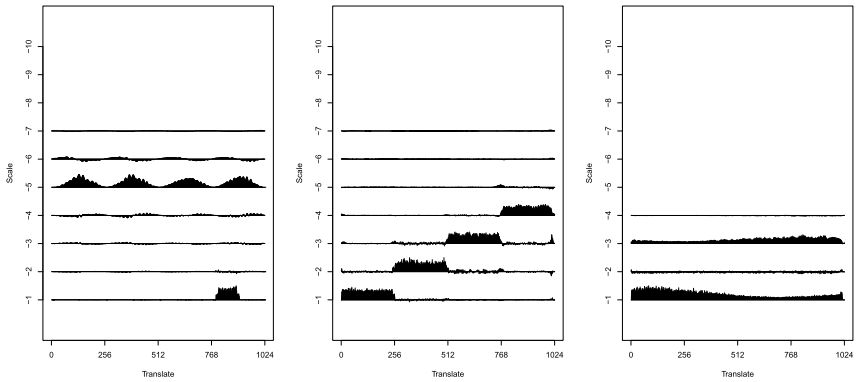

FIG 5. *Example averaged spectral estimates. Left: spectrum 1, seasonal trend only. Middle: spectrum 2, seasonal and sinusoidal trend. Right: spectrum 3 and time-varying seasonal trend only.*

varying, with the smooth trend component $\mu_t = 0$. In this case, the seasonal component is given by a linear trend where the initial values are simulated from the uniform distribution on the interval $[0, 10]$, with slope parameters generated as a uniform random variable on the interval $[-0.05, 0.05]$.

TABLE 4
Mean squared error comparison for the averaged spectrum estimate, multiplied by 10^3*, across the spectrum, trend, and seasonality scenarios.*

Trend	Spectrum 1	Spectrum 2	Spectrum 3
Seasonal + No Trend	4.76	8.44	2.54
Seasonal + Linear	4.76	8.44	2.54
Seasonal + Sine	4.76	8.43	2.54
Seasonal + Logistic	4.76	8.44	2.54
Seasonal + Piece. Quad.	4.79	8.46	2.55
Time-Varying Seasonal + No Trend	4.76	8.43	2.53

The results of these experiments are given in Table 4, which can be compared to the results for the lag 1 differences in Table 1. In particular, the performance is actually slightly better for spectrum 3; the smoothest spectrum. The other 2 spectra yield worse results, however: the performance drop-off is most noticeable for spectrum 2, which contains piecewise stationary components. This is due to the increased bias caused in estimating the time-varying spectra using lag 12 differences instead of lag 1 differences. Examples of the averaged spectrum estimates are given in Figure 5, qualitatively showing strong similarities to the examples in Figure 4.

6. Data application: Baby electrocardiogram data

Next, we discuss a data application that highlights the benefits of our proposed methodology. We re-analyse the baby electrocardiogram (ECG) data set first

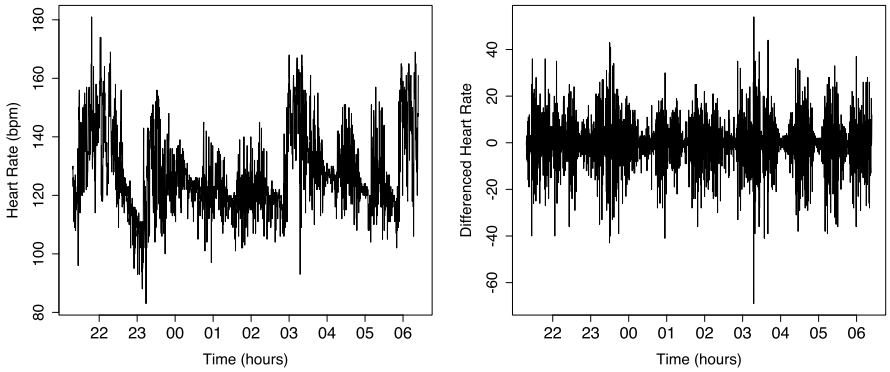

FIG 6. *Left: original ECG data. Right: first-differenced ECG data.*

examined in [21]. Furthermore, in the supplementary material, we analyse a wave height data set collected from a buoy in the Atlantic Ocean that was studied in [15].

In Figure 6 left, we see a time series for the ECG reading of a 66-day-old infant. The series is available in the `wavethresh` package and was collected by the Institute of Child Health at the University of Bristol. The series can be seen to exhibit both nonstationary trend and second-order behaviour. As [21] note, one reason for the nonstationarity is that the ECG varies considerably over time and changes significantly between periods of sleep and waking. In [21], the authors first-difference the series to remove the trend, as shown in Figure 6 right, resulting in an approximately zero-mean time series. Then, LSW analysis is performed on the differenced series.

We instead perform analysis on the original series, and note the similarities and differences to the original analysis in [21]. We use the Daubechies Least Asymmetric wavelet with 10 vanishing moments to analyse both series and the same bin width of size 128 for the running mean smoother, but as discussed we use a different matrix to correct for bias. In Figure 7 left, we see our analysis on the original series. The top plot is not scaled, while the bottom plot is individually scaled at each level. On the right, we see (our version of) the original analysis of [21], again with top plot unscaled and bottom plot scaled individually. In line with previous discussion, only the finest 7 scales were used for analysis. From the top plots, we can see the effect that differencing has had on the resulting spectral estimate. In our estimate based on the original time series, we see that the power is spread fairly evenly across scales. However, in the spectral estimate of the differenced series, power is concentrated in the finest scales. From the bottom plots, we see that both estimates exhibit similar overall behaviour, with both evolving over time in a similar fashion.

Due to the difference in magnitude between the estimated spectra at different scales, some features of the original series can be identified more easily in our analysis. In [21], an accompanying data set, the sleep state of the observed

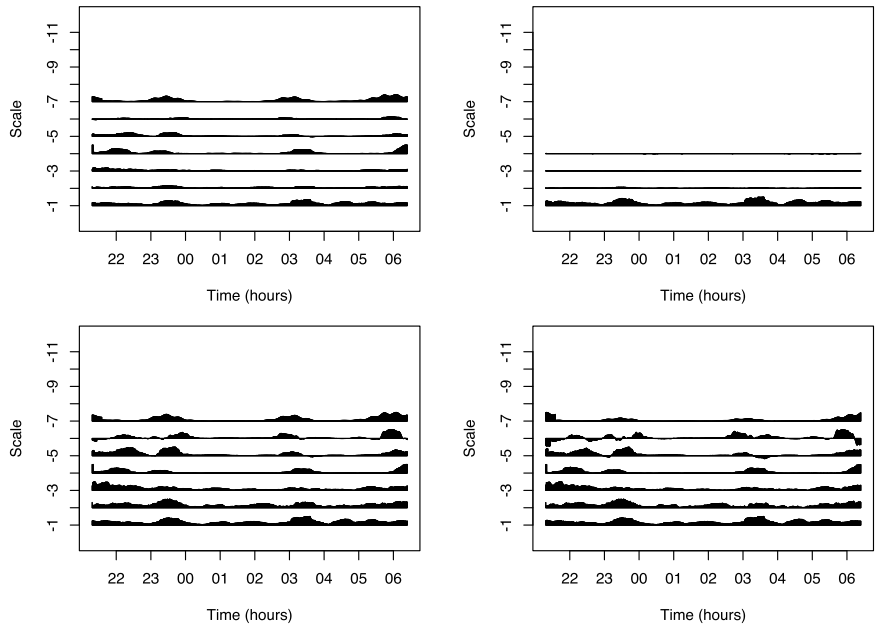

FIG 7. *Comparison of the spectral estimates. Top left: spectrum estimate using our methodology, unscaled, bottom left: scaled individually. Top right: spectrum estimate of the differenced series, unscaled, bottom right: scaled individually.*

infant, is also used in the analysis. The sleep state is judged by a trained human observer, and is measured as either quiet (1), between quiet and active (2), active (3), or awake (4). A strong association between sleep state and spectral value is observed for scales -1 to -5 in [21]. However, using the original series instead of the differenced series can allow this association to become more apparent: in particular, consider the spectrum estimate at scale -4, enlarged in Figure 8. In solid line is the spectral estimate of the original series, while the dashed line shows the spectral estimate for the differenced series. The dotted line shows the sleep state of the infant. Our estimate correlates strongly with the sleep state, especially in highlighting periods of being awake (sleep state 4). In general, the "signal" of variability appears stronger in our analysis, which is due to the differencing lowering the level of autocorrelation within the series. Furthermore, our estimate contains fewer negative spectral values, perhaps suggesting that the original time series is better represented as an LSW process as opposed to the first-differenced series.

We perform TI wavelet thresholding using the Daubechies Least Asymmetric wavelet with 4 vanishing moments to estimate the trend of the Baby ECG series. We analyse the finest 7 scales of the series, using a hard universal threshold of $\hat{\sigma}_{r,s}\sqrt{2\log(2048)}$, where $\hat{\sigma}_{r,s}$ is calculated using the spectral estimate in Figure 7 left. The trend estimate is shown in Figure 9 in the solid line. We see that in general the estimate is quite smooth, with more rapid changes in mean occurring

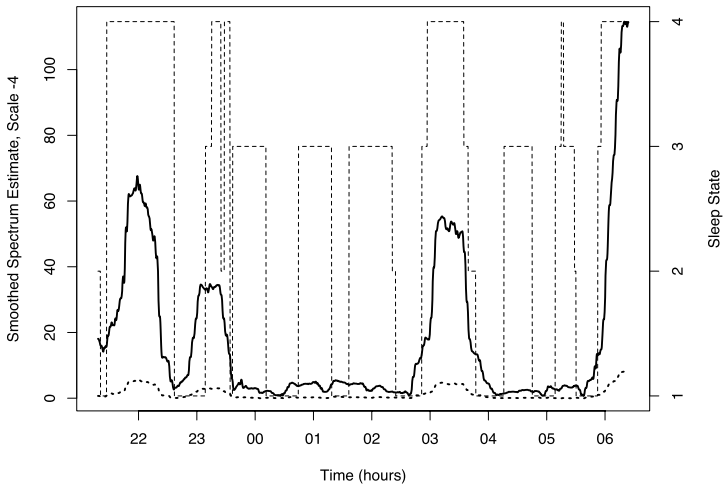

FIG 8. *Estimate of the EWS at scale −4: solid line is for the original series, while the dashed line is for the differenced series. Sleep state shown in dotted line.*

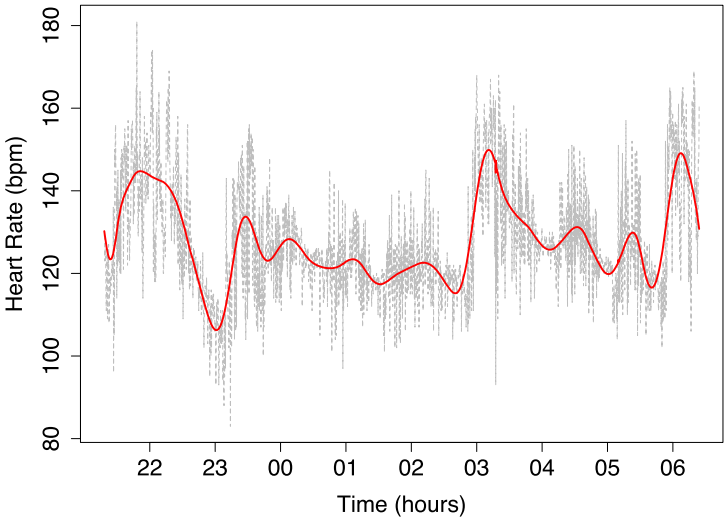

FIG 9. *Trend estimate for the Baby ECG data shown in solid line, with data shown in dashed line.*

at approximately 23:00, 03:00, and 06:00, corresponding to changes in sleep state. The estimate again correlates with the underlying sleep state, highlighting the benefit of performing a joint first and second-order analysis of the data.

7. Concluding remarks

In this article, we have considered the problem of jointly modelling time-varying first and second-order properties of nonstationary time series. Our model employs the locally stationary wavelet model of [21], used for modelling second-order nonstationarity, and adapts it to incorporate a nonstationary trend component. Using the common statistical technique of differencing, we have shown that we can consistently estimate the evolutionary wavelet spectrum and local autocovariance of the original series. Using these results, we have proposed a wavelet thresholding approach for the nonparametric estimation of the trend of the time series. The trend estimation methodology benefits from the information provided by the ability to consistently estimate the second-order structure of the time series, despite the presence of the trend component.

Using our methodology we have analysed two time series that highlight the strength of the method to provide information about the original time series, which can become masked when only considering the differenced time series. In particular, our analysis of the Baby ECG data set highlights the benefit of modelling the stochastic component of the original series for the purpose of identifying sleep states more readily.

Appendix A: Additional numerical results

A.1. Spectra and trend functions used

In this section we provide additional information for the discussion in Section 5 of the main text. The three evolutionary wavelet spectra used in the simulation study in Section 5 are defined as

$$S_j^1(z) = \begin{cases} \sin^2(4\pi z) & \text{for } j = -5, z \in (0,1), \\ 1 & \text{for } j = -1, z \in (800/1024, 900/1024), \\ 0 & \text{otherwise,} \end{cases}$$

$$S_j^2(z) = \begin{cases} 1 & \text{for } j = -1, z \in (0/1024, 256/1024), \\ 1 & \text{for } j = -2, z \in (256/1024, 512/1024), \\ 1 & \text{for } j = -3, z \in (512/1024, 768/1024), \\ 1 & \text{for } j = -4, z \in (768/1024, 1), \\ 0 & \text{otherwise,} \end{cases}$$

$$S_j^3(z) = \begin{cases} \frac{1}{2} + \frac{1}{4}\sin(\pi z) - \frac{1}{2}\cos(3\pi z/2) & \text{for } j = -1, z \in (0,1), \\ \frac{1}{2} - \frac{1}{8}\sin(2\pi z) - \frac{1}{4}\cos(\pi z/2) & \text{for } j = -3, z \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

The four trend functions used in the simulation study are defined as

$$\mu_{li}(z) = 4z,$$

TABLE 5
Mean squared error, multiplied by 10^3, comparing across the spectrum and trend scenarios.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
None	1.25	3.32	1.30
Linear	1.29	4.09	2.34
Sine	1.29	4.09	2.34
Logistic	1.29	4.09	2.34
Piece. Quad.	1.29	4.11	2.34

$$\mu_s(z) = -2\sin(2\pi z) - \frac{3}{2}\cos(\pi z),$$
$$\mu_{lo}(z) = \frac{4}{1+\exp(4-7\log 4z)},$$
$$\mu_q(z) = \begin{cases} 12z^2 + 2z & \text{for } z \in (0, 300/1024), \\ 1.81 - 16z^2 + 4z & \text{for } z \in (300/1024, 800/1024), \\ 4z - 7.94 & \text{for } z \in (800/1024, 1), \end{cases}$$

where $z = t/T$.

A.2. Results for Daubechies LA10 wavelet LSW processes

In this section we report the results of the simulations when the Daubechies Least Asymmetric wavelet with 10 vanishing moments (LA10) is used to simulate the LSW processes. Table 5 reports the mean squared error comparison for the averaged spectral estimates, where the LSW process has been generated using the LA10 wavelet. The estimation error for spectrum 1 is almost identical comparing to "None" using our methodology with a trend present, while for spectrum 2 and 3 there is only a marginal increase. The results show that the method performs well irrespective of the wavelet that generates the LSW process.

Table 6 reports the average mean squared error and standard deviation for the trend estimate where the 100 LSW processes have been generated using Gaussian innovations and the LA10 wavelet. Lastly, Table 7 reports the average mean squared error and standard deviation for the trend estimate where the 100 LSW processes have been generated using exponential innovations and the LA10 wavelet. For both cases, a hard threshold of $\hat{\sigma}_{r,s}\sqrt{2\log(1024)}$ is used. Again, the error is consistent across the various time series scenarios. We see that the trend estimate performs well no matter the generating wavelet of the process.

A.3. Effect of over-differencing

In the main text, we recommended using a first difference for removing the trend. Here, we compare the estimation performance of the spectral estimate computed using first- and second-differences, to highlight the effect of over-differencing the time series. We perform the same simulations as in Section 5.1 of the main

TABLE 6

Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using Gaussian innovations and the Daubechies LA10 wavelet.

Trend	Spectrum	Method		
		LSWT	SWT	VSMWT
Linear	1	**0.019** (0.013)	0.516 (0.117)	0.275 (0.091)
	2	**0.038** (0.024)	0.746 (0.083)	0.207 (0.0644)
	3	**0.029** (0.027)	0.494 (0.042)	0.171 (0.039)
Sine	1	**0.021** (0.012)	0.547 (0.130)	0.287 (0.111)
	2	**0.031** (0.016)	0.752 (0.077)	0.214 (0.067)
	3	**0.024** (0.018)	0.485 (0.045)	0.156 (0.040)
Logistic	1	**0.022** (0.013)	0.544 (0.112)	0.306 (0.107)
	2	**0.042** (0.028)	0.764 (0.088)	0.214 (0.066)
	3	**0.028** (0.025)	0.491 (0.038)	0.170 (0.041)
Piece. quad.	1	**0.025** (0.014)	0.537 (0.127)	0.280 (0.110)
	2	**0.041** (0.024)	0.736 (0.076)	0.211 (0.080)
	3	**0.034** (0.028)	0.414 (0.043)	0.162 (0.042)

TABLE 7

Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated using exponential innovations and the Daubechies LA10 wavelet.

Trend	Spectrum	Method		
		LSWT	SWT	VSMWT
Linear	1	**0.027** (0.016)	0.547(0.140)	0.312 (0.107)
	2	**0.050** (0.028)	0.793 (0.113)	0.337(0.096)
	3	**0.040** (0.029)	0.529 (0.066)	0.311 (0.068
Sine	1	**0.027** (0.014)	0.540 (0.138)	0.311 (0.121)
	2	**0.041** (0.021)	0.788 (0.118)	0.340 (0.103)
	3	**0.036** (0.019)	0.542 (0.059)	0.308 (0.067)
Logistic	1	**0.026** (0.017)	0.544 (0.143)	0.315 (0.118)
	2	**0.046** (0.026)	0.801 (0.112)	0.254 (0.091)
	3	**0.045** (0.030)	0.541 (0.053)	0.327 (0.059)
Piece. quad.	1	**0.032** (0.020)	0.515 (0.122)	0.296 (0.102)
	2	**0.044** (0.028)	0.774 (0.111)	0.347 (0.100)
	3	**0.042** (0.029)	0.473 (0.057)	0.319 (0.061)

TABLE 8

Relative mean squared error comparison for the averaged spectrum estimate, across the spectrum and trend scenarios.

Trend	Spectrum 1	Spectrum 2	Spectrum 3
Linear	1.562	1.251	1.461
Sine	1.562	1.251	1.461
Logistic	1.561	1.251	1.461
Piece. Quad.	1.562	1.251	1.459

text, and compute the averaged spectral estimate using both first- and second-differences. To compare the two results, we compute the relative mean squared error (RMSE) of the first-differenced estimate and second-differenced estimate, in each of the trend and spectrum scenarios. To control for boundary effects, we only compute the RMSE on the non-boundary wavelet coefficients. The results are reported in Table 8. We see that over-differencing yields a higher MSE for spectral estimation. The RMSE is above 1 in all scenarios, showing that the first-difference estimator achieves a uniformly stronger performance.

TABLE 9
Average mean squared error and standard deviation in brackets of trend estimate over 100 realisations generated from Models A – D.

Trend	Model	Method		
		LSWT	SWT	VSMWT
Linear	A	**0.140** (0.061)	0.376 (0.108)	0.482 (0.139)
	B	**0.147** (0.040)	0.265 (0.070)	0.500 (0.086)
	C	**0.129** (0.048)	0.435 (0.080)	0.532 (0.095)
	D	**0.216** (0.059)	0.349 (0.160)	0.840 (0.129)
Sine	A	**0.124** (0.069)	0.389 (0.134)	0.494 (0.130)
	B	**0.134** (0.033)	0.241 (0.091)	0.499 (0.087)
	C	**0.122** (0.040)	0.421 (0.083)	0.516 (0.100)
	D	**0.174** (0.045)	0.256 (0.146)	0.806 (0.135)
Logistic	A	**0.129** (0.048)	0.370 (0.091)	0.474 (0.100)
	B	**0.150** (0.033)	0.267 (0.077)	0.495 (0.092)
	C	**0.129** (0.047)	0.431 (0.093)	0.531 (0.098)
	D	**0.214** (0.051)	0.357 (0.147)	0.837 (0.143)
Piece. quad.	A	**0.141** (0.051)	0.362 (0.094)	0.467 (0.115)
	B	**0.141** (0.040)	0.275 (0.073)	0.500 (0.099)
	C	**0.131** (0.046)	0.420 (0.084)	0.525 (0.098)
	D	**0.221** (0.057)	0.332 (0.138)	0.834 (0.145)

A.4. Further trend estimation simulations

Next, we assess the performance of the trend estimation procedure in the presence of non LSW-type error structures to highlight the versatility of our methodology. In particular, we simulate 100 realisations of time series using the previously defined trends, with (Gaussian) errors simulated from the following models:

(A) Time-varying AR(2) model with parameters given by $\phi_1(z) = 0.8\cos(1.5 - \cos(4\pi z))$, $\phi_2(z) = -0.2 + 0.4z$, $z \in (0, 1)$.
(B) AR(1) model with parameter $\phi = 0.6$.
(C) Time-varying AR(1) model with parameter given by $\phi(z) = 0.7$ for $z \in (0, 600/1024)$ and $\phi(z) = -0.3$ for $z \in (600/1024, 1)$.
(D) ARMA(1,3) model with AR parameter $\phi = 0.4$ and MA parameters $\theta_1 = 0.8$, $\theta_2 = -0.3$, and $\theta_3 = 0.4$.

These scenarios represent some common stationary and nonstationary error structures. Model A is an AR(2) process with slowly-evolving autoregressive parameters, and is a variant of a process studied in [32]. Model C represents a piecewise stationary AR(1) process with a single changepoint in the autoregressive parameter. The results of these simulations are given in Table 9, again with bold values showing the lowest reported mean squared error. We see that our methodology is able to perform well across the four scenarios, working well in both stationary and nonstationary second-order settings. It outperforms the other methods in all of the trend and error scenarios. Overall, we see that our proposed methodology offers strong practical performance across a variety of time series models.

A.5. Choice of α and β Parameters

In this section we investigate the effect that the choice of the parameters α and β have on estimation quality, in order to inform the best choice of the parameters in practice. Recall that β is the proportion of scales used when estimating the spectrum, whilst α is the number of scales used for estimating the LACV.

Firstly, the choice of β was investigated. We simulated 1000 realisations of LSW processes, for three wavelets: the Haar, Daubechies EP4, and Daubechies EP10 wavelet. For each realisation the spectrum was given by $S_j = U_j 2^{\gamma j}$, where $U_j \sim \text{Uniform}(0, 5)$ and $\gamma > 0$ is varied. This provides a more complete picture of the impact of the choice of β than only using the three previously defined spectra. The decaying spectral structure is used in order to control for the fact that estimation of the EWS is more difficult in coarse scales, and mirrors the assumption in Proposition 3.6 in the main text. In addition, we use a global mean to smooth the raw wavelet periodogram at each level, in order to focus on the effect of β, rather than the bin width parameter.

We calculated the smoothed EWS estimate, while varying the number of scales used to perform EWS correction, for $\gamma = 1/2, 1/4, 1/8$, and time series length $T = 512, 1024, 2048$. We computed the average mean squared error over the 1000 realisations between the estimated and true spectra. For sake of brevity, the results presented here are for the case where the time series has no trend: further simulations in each of the trend scenarios considered in the main text produced almost identical results. The results are shown in Figure 10. Note that the results in Figure 10 suggest that the choice of β is dependent on the wavelet used: the higher the number of vanishing moments, the higher the optimal choice of β. Across all (27) scenarios, the mean value of β that produces the lowest average mean squared error in each scenario is 0.721. This motivates setting $J_1 = \lfloor \frac{7}{10} \log_2(T) \rfloor$ as a flexible rule-of-thumb. In practice, different values of β can also be compared visually to assess the fit.

Next, given the above choice of $\beta = \lfloor \frac{7}{10} \log_2(T) \rfloor$, we perform the same set of simulations as above, this time with varying α, the number of scales used to estimate the LACV function. Note that necessarily $\alpha \leq \beta$. We compute the average MSE (averaged across lags) of the estimated autocovariance function, computed at the first 2^{J-3} lags where $J = \log_2(T)$. The results of the simulation are shown in Figure 11. These show that taking $\alpha = \beta$ yields the best practical performance in all scenarios.

A.6. Canadian wave height data

Here we examine publicly available wave height data for a location in the North Atlantic. The data we use are obtained from Fisheries and Oceans Canada, East Scotian Slop buoy ID C44137. The data measures wave heights collected at hourly intervals, from approximately mid-June 2005 to mid-May 2006, giving a time series of length $2^{13} = 8192$, and is available in the `changepoint` package in `R` [14]. [15] also perform nonstationary analysis on the series, using

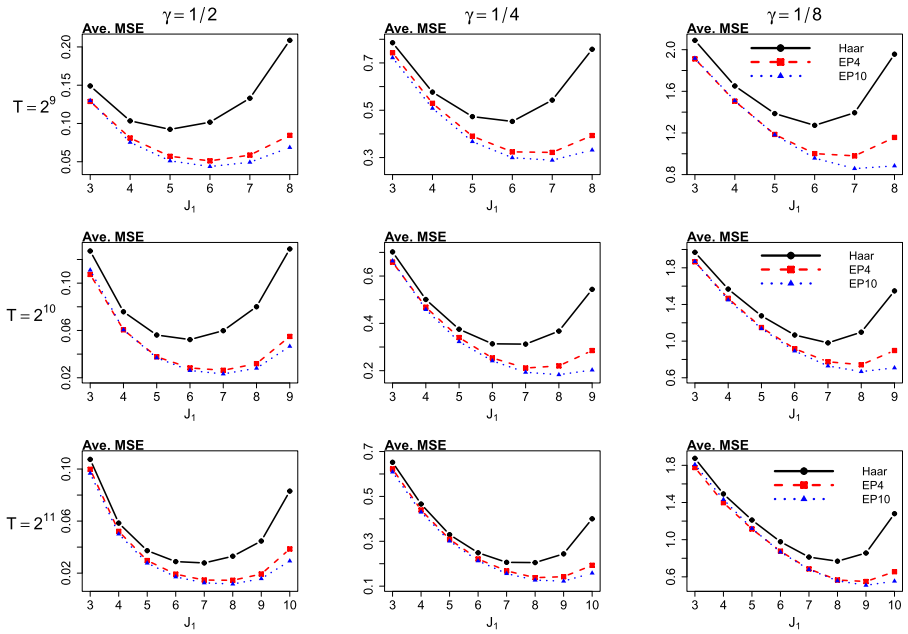

FIG 10. *Average mean squared error (MSE) for different choices of J_1, for varying values of γ and T, using the Haar, EP4, and EP10 wavelets.*

the first-differences of a longer version of the series to detect changepoints in variance only. The data are plotted in Figure 12 left, from which we see larger wave heights in the winter and smaller wave heights in the summer, as well as increased variability in the winter. Figure 12 right shows the first-differenced time series. [15] take first-differences to remove the trend, and we also find this is sufficient.

For the spectral estimate, we use the Daubechies Least Asymmetric wavelet with 10 vanishing moments. For simplicity, the periodogram is smoothed using a running mean with bin width 512, corresponding to a time length of roughly 3 weeks. The estimate is shown in Figure 13, where each level in the plot is scaled individually for clarity. We can see strong nonstationarity in the spectrum, with more variability during the winter months and less in the summer, as found in [15] using a changepoint approach.

To estimate the trend of the series, we perform TI wavelet thresholding using the Daubechies Least Asymmetric wavelet with 4 vanishing moments. In line with previous discussion, we analyse the finest 9 scales of the series. We use a soft universal threshold of $\hat{\sigma}_{r,s}\sqrt{2\log(8192)}$ to account for likely non-Gaussianity, where $\hat{\sigma}_{r,s}$ is calculated using the spectral estimate in Figure 13. The trend estimate is shown in the solid line in Figure 14. We see that the estimated trend function is relatively smooth with occasional sharp changes, with the mean wave height larger during the winter and smaller in the summer. There are several

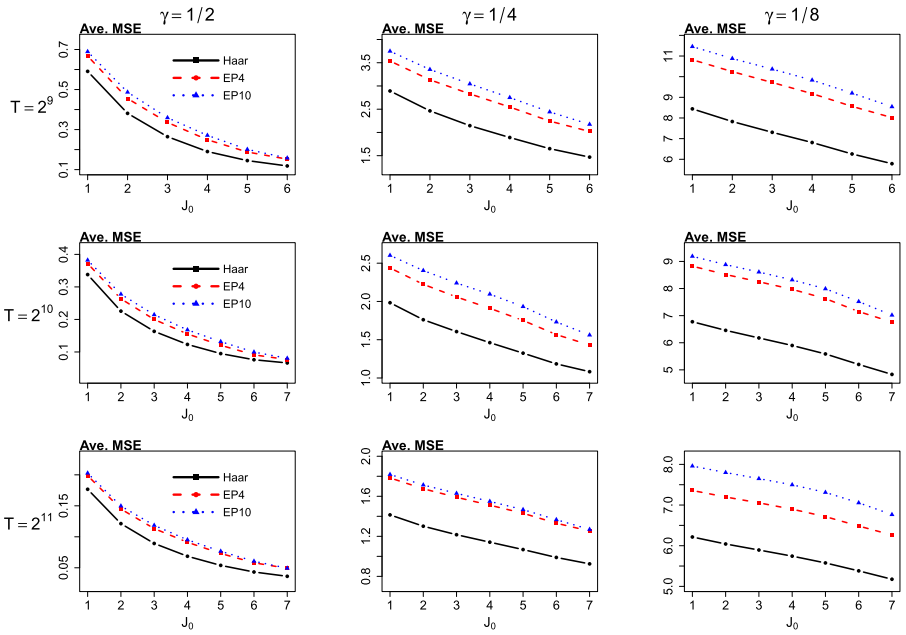

FIG 11. *Average mean squared error (MSE) for different choices of J_0, for varying values of γ and T, using the Haar, EP4, and EP10 wavelets.*

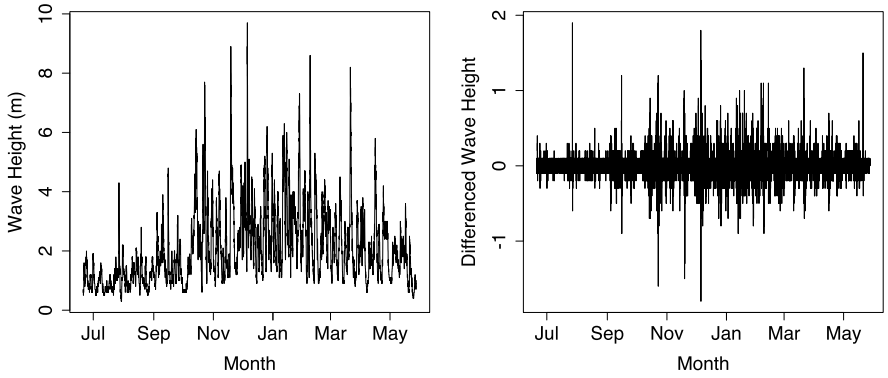

FIG 12. *Left: North Atlantic wave heights recorded hourly between June 2005 and May 2006. Right: first-differenced wave heights.*

locations where perhaps the wavelet coefficients were not thresholded correctly, leading to the sharp changes. Using a scale-dependent smoothing of the raw wavelet periodogram is one possible way to remedy this.

Using the spectral estimate in Figure 13, we calculated the local autocovariance estimate of the series. In Figure 15, we see in the solid line the estimate using our methodology for the local variance function of the time series. The

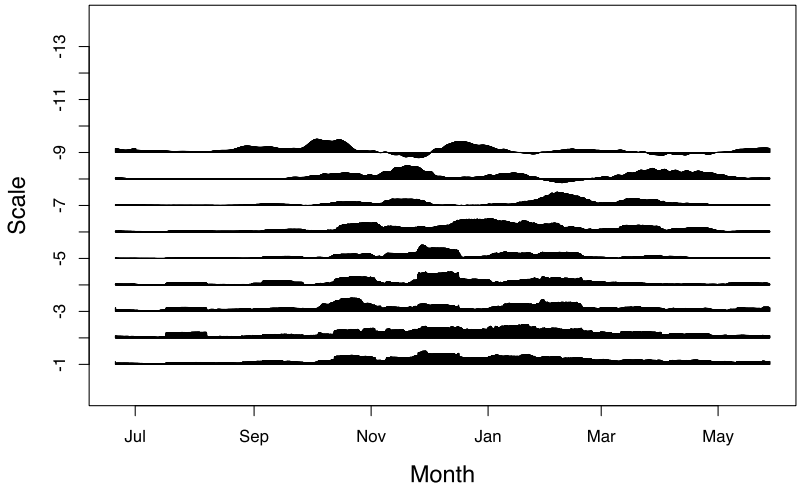

FIG 13. *Smoothed spectral estimate for the wave height data in Figure 12, performed using first-differencing of the data.*

nonstationary nature of the variance is clear to see, with the summer months containing low variability, while winter months display much higher variability. As one way of determining the performance of our method, we can compare the variance estimate with the estimate obtained by using the detrended wave height data. The detrended wave height data is obtained by subtracting the wavelet thresholding trend estimate from the data. We then perform standard LSW inference on this series, using the same parameters as in our approach. In dashed line, we see the local variance estimate obtained in this way. We see that the two estimates agree, which is reassuring on two counts: first that our spectral estimate obtained using the differenced data is accurate, and secondly that the trend estimate obtained was accurate.

Finally, we plot the autocorrelation function across 4 time points, which highlights the second-order nonstationary nature of the data, and shows how the structure of the autocorrelation varies considerably over time. In Figure 16, we see that estimate for the local autocorrelation function for the first observation recorded in the months of July 2005, October 2005, January 2006 and April 2006. These are shown in the solid, dashed, dotted, and dashed and dotted lines respectively. We see that in general the series exhibits strong autocorrelation, which we may expect due to the observations being recorded at hourly intervals. Furthermore, we note that the shape of the autocorrelation function changes across the four months plotted.

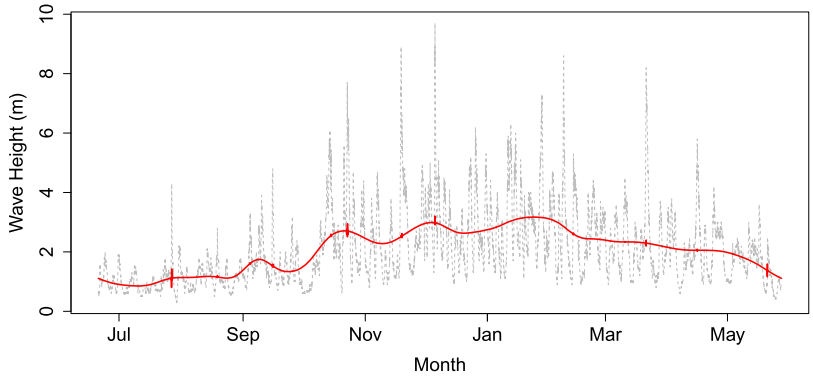

FIG 14. *Trend estimate for the North Atlantic wave data shown in solid line, with data shown in dashed line.*

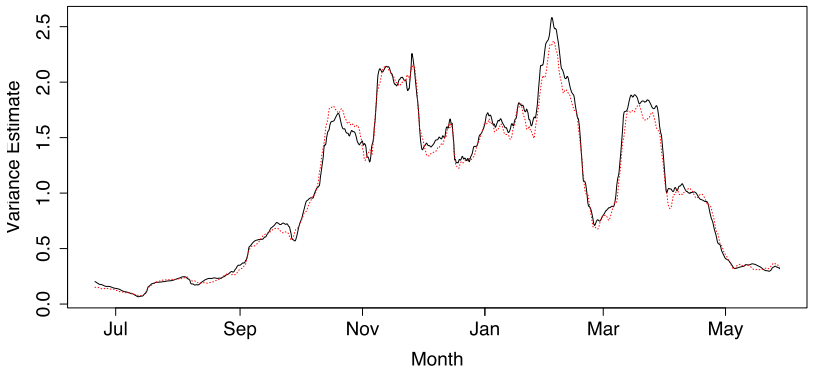

FIG 15. *Local variance estimate for the wave height data. Solid line: obtained using our methodology. Dashed line: obtained from the detrended data.*

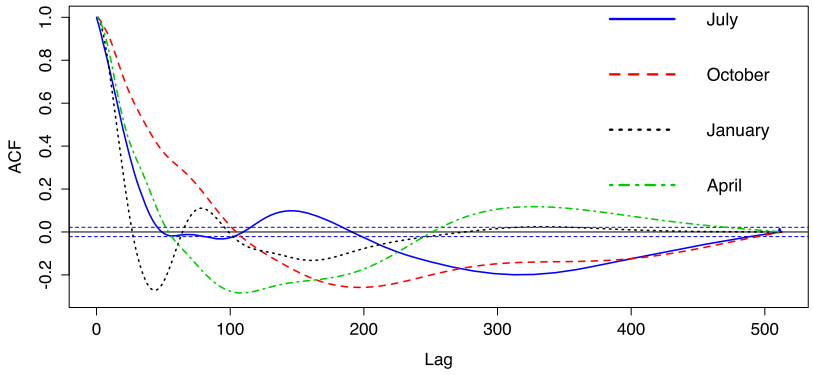

FIG 16. *Local autocorrelation function estimate for the wave height data, at 4 different time points. Solid: July, dashed: October, dotted: January, dotted and dashed: April.*

Appendix B: Proofs of results

B.1. Proof of Proposition 3.2

The matrix D_J^1 is symmetric and positive semi-definite, since it can be expressed as a Gram matrix of vector inner products:

$$D_J^1 = \langle \nabla\Psi_j, \nabla\Psi_l \rangle,$$

where Ψ_j is the autocorrelation wavelet vector at scale j, and ∇ is the first-differencing matrix with diagonal entries 1, above the main diagonal entries equal to -1, and all other entries equal to zero. The matrix ∇ is invertible (with inverse given by the upper triangular matrix with non-zero entries equal to 1). By [21] Theorem 1, the family of vectors $\{\Psi_j\}_{j=-J}^{-1}$ is linearly independent. This implies that $\{\nabla\Psi_j\}_{j=-J}^{-1}$ is also linearly independent, since this family is given by the invertible matrix transform of a family of linearly independent vectors. Therefore, D_J^1 is invertible since it is the Gram matrix of a set of linearly independent vectors.

B.2. Proof of Proposition 3.3

The expectation is given by

$$\begin{aligned}
\mathbb{E}(\tilde{I}_k^j) &= \mathbb{E}\left[\left(\sum_t \left[\epsilon_t - \epsilon_{t-1} + \mathcal{O}(T^{-1})\right]\psi_{j,k-t}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_t \epsilon_t \psi_{j,k-t}\right)^2\right] + \mathbb{E}\left[\left(\sum_t \epsilon_{t-1}\psi_{j,k-t}\right)^2\right] \\
&\quad - 2\mathbb{E}\left(\sum_s \sum_t \epsilon_t \epsilon_{s-1}\psi_{j,k-t}\psi_{j,k-s}\right) + \mathcal{O}(T^{-1}) \\
&:= I + II + III + \mathcal{O}(T^{-1}),
\end{aligned}$$

where the remainder term can come out of the inner bracket since, for fixed j, the sum is finite as the wavelet is compactly supported. Now we evaluate each expectation individually. The term I is simply the expectation of the raw wavelet periodogram of the original LSW model. Hence,

$$I = \sum_l A_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).$$

Next, the term II is equal to $\mathbb{E}(I_{k-1}^j)$, since $d_{j,k-1} = \sum_s X_s \psi_{j,k-1-s}$, and by substituting $s = t-1$, we obtain $\sum_t X_{t-1}\psi_{j,k-t} = \sum_s X_s \psi_{j,k-1-s}$. Next, setting

$u = k - 1$, we obtain

$$\mathbb{E}(I_u^j) = \sum_l \sum_m w_{lm}^2 \left(\sum_t \psi_{l,m-t} \psi_{j,u-t} \right)^2 .$$

Now, following a similar argument to [21], we obtain

$$\begin{aligned} II &= \sum_l S_l \left(\frac{u+1}{T} \right) \sum_n \sum_t \sum_s \psi_{j,-s} \psi_{j,-t} \psi_{l,n-s} \psi_{l,n-t} + \mathcal{O}(T^{-1}) \\ &= \sum_l A_{jl} S_l \left(\frac{k}{T} \right) + \mathcal{O}(T^{-1}). \end{aligned}$$

Finally,

$$\begin{aligned} III &= -2\mathbb{E} \left(\sum_s \sum_t \epsilon_t \epsilon_{s-1} \psi_{j,k-t} \psi_{j,k-s} \right) \\ &= -2 \sum_s \sum_t \psi_{j,k-t} \psi_{j,k-s} \mathbb{E}(\epsilon_t \epsilon_{s-1}) \\ &= -2 \sum_s \sum_t \psi_{j,k-t} \psi_{j,k-s} \sum_l S_l \left(\frac{t+s-1}{2T} \right) \Psi_l(s-t-1) + \mathcal{O}(T^{-1}), \end{aligned}$$

since

$$\mathbb{E}(\epsilon_t \epsilon_{s-1}) = c \left(\frac{t+s-1}{2T}, s-t-1 \right) + \mathcal{O}(T^{-1}).$$

The remainder of the proof is similar to that of the proof of Theorem 1 of [18]. Let $u = k - t$ and $v = k - s$, substituting into the above to obtain

$$\begin{aligned} III &= -2 \sum_l \sum_u \sum_v \psi_{j,u} \psi_{j,v} S_l \left(\frac{k}{T} - \frac{u+v-1}{2T} \right) \Psi_l(u-v-1) + \mathcal{O}(T^{-1}) \\ &= -2 \sum_l \sum_u \sum_v \psi_{j,u} \psi_{j,v} \left[S_l \left(\frac{k}{T} \right) + \mathcal{O} \left(\frac{|u+v-1|}{2T} \right) \right] \Psi_l(u-v-1) \\ &\quad + \mathcal{O}(T^{-1}). \end{aligned}$$

The remainder term in the above expression can be shown to be $\mathcal{O}(T^{-1})$ (A.2.1 in [18]). Hence, we are left with

$$III = -2 \sum_l S_l \left(\frac{k}{T} \right) \sum_u \sum_v \psi_{j,u} \psi_{j,v} \Psi_l(u-v-1) + \mathcal{O}(T^{-1}).$$

Finally, substituting $r = u - v$, we obtain

$$C = -2 \sum_l S_l \left(\frac{k}{T} \right) \sum_u \sum_r \psi_{j,u} \psi_{j,u-r} \Psi_l(r-1) + \mathcal{O}(T^{-1})$$

$$= -2\sum_l S_l\left(\frac{k}{T}\right)\sum_r \Psi_j(r)\Psi_l(r-1) + \mathcal{O}(T^{-1})$$

$$= -2\sum_l A^1_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).$$

Hence, we obtain

$$\begin{aligned}\mathbb{E}(\tilde{I}^j_k) &= I + II + III \\ &= \sum_l A_{jl} S_l\left(\frac{k}{T}\right) + \sum_l A_{jl} S_l\left(\frac{k}{T}\right) - 2\sum_l A^1_{jl} S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}) \\ &= 2\sum_l \left(A_{jl} - A^1_{jl}\right) S_l\left(\frac{k}{T}\right) + \mathcal{O}(T^{-1}).\end{aligned}$$

For the variance part, we follow the same argument as in the proof of Proposition 4 of [21]. The wavelet coefficients of the differenced series are asymptotically Gaussian. Hence, the wavelet periodograms are asymptotically scaled χ^2-distributed, and hence the variance is asymptotically proportional to the expectation squared.

B.3. Proof of Theorem 3.4

(Proof for the Haar wavelet). The proof follows the same strategy to that of the proof of Theorem 2 in [21]. We show that there exists $\delta > 0$ such that $\lambda_{\min}(P) \geq \delta$, by using the following property from Toeplitz matrix theory. Let T be Toeplitz (and Hermitian) with elements $\{t_0, t_1, \ldots\}$. Let $f(z) = \sum_{n=-\infty}^{\infty} t_n z^n$ for $z \in \mathbb{C}$ be the symbol of the operator associated with T. If $\sum_n |t_n| < \infty$, then $f(z)$ is analytic in the open unit disc D in the complex plane and continuous in the closed unit disc $\Delta = D \cup S$, where S denotes the unit circle. The spectrum Λ of the (Laurent) operator T is $\Lambda(T) = f(S)$. If T is symmetric then an estimate of the smallest eigenvalue of T is

$$\min_{|z|=1}\{f(z)\} = \min_{|z|=1}\left\{t_0 + 2\mathrm{Re}\left(\sum_{n=1}^{\infty} t_n z^n\right)\right\}.$$

([23], theorem 3.1 (i)). For ease of notation, indices will now run from 1 to ∞ instead of -1 to $-\infty$. Using straightforward algebra, we can derive explicit formulae for the entries of P, which are given by the following lemma.

Lemma B.1. *In the case of the Haar wavelet, the elements of the matrix P are given by*

$$P_{jj} = 10,$$

$$P_{j,j+m} = 6 \times 2^{-m/2}, \text{ for } l = j + m, m > 0.$$

Proof. In general, we can derive the discrete autocorrelation wavelets $\Psi_j(\tau)$ by discretising the continuous wavelet autocorrelation function $\Psi(\tau)$, using the relationship

$$\Psi_j(\tau) = \Psi\left(\frac{|\tau|}{2^j}\right).$$

For Haar wavelets, the continuous wavelet autocorrelation function is given by

$$\Psi(\tau) = \int_{-\infty}^{\infty} \psi_H(x)\psi_H(x-\tau)dx = \begin{cases} 1-3|\tau| & \text{for } |\tau| \in [0,1/2], \\ |\tau|-1 & \text{for } |\tau| \in (1/2,1], \end{cases}$$

where $\psi_H(x)$ is the Haar mother wavelet. The discretisation formula holds for $\tau = -(2^j-1), \ldots, 0, \ldots, (2^j-1)$, and is equal to zero for all other values of τ. By [21], we have that the elements of A are given by

$$A_{jj} = \frac{1}{3}2^j + \frac{5}{3}2^{-j}, \qquad A_{jl} = 2^{2j-l-1} + 2^{-l}, \quad l > j > 0.$$

We must therefore calculate the elements of A^1, from which we can obtain the elements of D^1, and hence P. Note that replacing the $(\tau-1)$ terms with $(\tau+1)$ results in an equivalent definition of the operators A^1 and D^1. When $l = j$, we have that

$$\begin{aligned} A^1_{jj} &= \sum_{\tau=-(2^j-1)}^{2^j-1} \Psi_j(\tau)\Psi_j(\tau+1) \\ &= 2\left[\sum_{\tau=1}^{2^{j-1}-1} \Psi_j(\tau)\Psi_j(\tau+1) + \sum_{\tau=2^{j-1}+1}^{2^j-1} \Psi_j(\tau)\Psi_j(\tau+1)\right] \\ &\quad - 2^{-j-1} + \frac{1}{4} - 2^{-j-1} + \frac{1}{4} \\ &= 2\left[\sum_{\tau=1}^{2^{j-1}-1} \left(1-\frac{3\tau}{2^j}\right)\left(1-\frac{3(\tau+1)}{2^j}\right) + \sum_{\tau=2^{j-1}+1}^{2^j-1} \left(\frac{\tau}{2^j}-1\right)\left(\frac{\tau+1}{2^j}-1\right)\right] \\ &\quad - 2^{-j} + \frac{1}{2} \\ &= \frac{1}{3}2^j - \frac{10}{3}2^{-j}. \end{aligned}$$

Hence $D^1_{jj} = 2A_{jj} - 2A^1_{jj} = 10 \times 2^{-j}$, and therefore $P_{jj} = 10$. When $j \neq l$, without loss of generality assume $l > j$. We have that

$$A^1_{jl} = \sum_{\tau=-(2^j-1)}^{2^j-1} \Psi_j(\tau)\Psi_l(\tau+1) = \sum_{\tau=-(2^j-1)}^{2^j-1} \Psi_j(\tau)\left(1-\frac{3|\tau+1|}{2^l}\right)$$

$$= 1 - 3 \times 2^{-l} + \sum_{\tau=-(2^j-1)}^{-1} \Psi_j(\tau) \left(1 - \frac{3|\tau+1|}{2^l}\right) + \sum_{\tau=1}^{2^j-1} \Psi_j(\tau) \left(1 - \frac{3(\tau+1)}{2^l}\right),$$

since the value of $|\tau+1|$ is always less than or equal to $1/2$, and so we use the $1-3|\tau|$ part of the autocorrelation wavelet formula. The first sum is equal to

$$\begin{aligned}
&\sum_{\tau=-(2^j-1)}^{-2^{j-1}-1} \Psi_j(\tau) \left(1 + \frac{3(\tau+1)}{2^l}\right) + \sum_{\tau=-2^{j-1}}^{-1} \Psi_j(\tau) \left(1 + \frac{3(\tau+1)}{2^l}\right) \\
&= \sum_{\tau=-(2^j-1)}^{-2^{j-1}-1} \left(-1 - \frac{\tau}{2^j}\right) \left(1 + \frac{3(\tau+1)}{2^l}\right) + \sum_{\tau=-2^{j-1}}^{-1} \left(1 + \frac{3\tau}{2^j}\right) \left(1 + \frac{3(\tau+1)}{2^l}\right) \\
&= \left(-2^{j-l-2} - 2^{j-3} + 2^{-l-1} + \frac{1}{4} - 2^{j-l-1} + 2^{2j-l-2}\right) \\
&\quad + \left(3 \times 2^{j-l-2} + 2^{j-3} - 3 \times 2^{-l-1} - \frac{3}{4}\right) \\
&= 2^{2j-l-2} - 2^{-l} - \frac{1}{2}.
\end{aligned}$$

The second sum is given by

$$\begin{aligned}
&\sum_{\tau=1}^{2^{j-1}} \Psi_j(\tau) \left(1 - \frac{3(\tau+1)}{2^l}\right) + \sum_{\tau=2^{j-1}+1}^{2^j-1} \Psi_j(\tau) \left(1 - \frac{3(\tau+1)}{2^l}\right) \\
&= \sum_{\tau=1}^{2^{j-1}} \left(1 - \frac{3\tau}{2^j}\right) \left(1 - \frac{3(\tau+1)}{2^l}\right) + \sum_{\tau=2^{j-1}+1}^{2^j-1} \left(\frac{\tau}{2^j} - 1\right) \left(1 - \frac{3(\tau+1)}{2^l}\right) \\
&= \left(2^{j-3} + 3 \times 2^{-l} - \frac{3}{4}\right) + \left(2^{2j-l-2} - 2^{j-3} - 2^{-l} + \frac{1}{4}\right) \\
&= 2^{2j-l-2} + 2^{1-l} - \frac{1}{2}.
\end{aligned}$$

From this, we finally obtain that

$$\begin{aligned}
A_{jl}^1 &= \left(2^{2j-l-2} - 2^{-l} - \frac{1}{2}\right) + \left(2^{2j-l-2} + 2^{1-l} - \frac{1}{2}\right) + \left(1 - 3 \times 2^{-l}\right) \\
&= 2^{2j-l-1} - 2^{1-l}.
\end{aligned}$$

Therefore, we have that $D_{jl}^1 = 2A_{jl} - 2A_{jl}^1 = 6 \times 2^{-l}$, from which the required form for P follows. □

Returning to the main proof, we have that P is symmetric, however the formula above only refers to the upper triangular portion of the matrix P. Now, P is a Toeplitz matrix with $p_0 = 10$ and $p_m = 6 \times 2^{-m/2}$. We can now show that $\lambda_{\min}(P) \geq \delta > 0$. Substituting in the formula for the symbol of the symmetric Toeplitz P, we obtain

$$\begin{aligned}\min_{|z|=1}\{f(z)\} &= \min_{|z|=1}\left\{10 + 2\mathrm{Re}\left(\sum_{n=1}^{\infty} 6 \times 2^{-n/2} z^n\right)\right\} \\ &= 10 + 12 \min_{|z|=1}\left\{\mathrm{Re}\left(\frac{-\sqrt{2}z}{\sqrt{2}z - 2}\right)\right\} \\ &= 10 + 12 \min_{|z|=1}\left\{\frac{2\sqrt{2}\mathrm{Re}(z) - 2}{6 - 4\sqrt{2}\mathrm{Re}(z)}\right\}.\end{aligned}$$

This function is strictly monotonically increasing on $-1 \leq \mathrm{Re}(z) \leq 1$, therefore it follows that $\min_{|z|=1}\{f(z)\} = f(-1)$. Hence,

$$\lambda_{\min}(P) \geq f(-1) = 10 + \frac{12(-2\sqrt{2} - 2)}{6 + 4\sqrt{2}} = \frac{18 + 8\sqrt{2}}{3 + 2\sqrt{2}} > 0.$$

(Proof for the Shannon wavelet). Note that the indices now run over the negative integers. We can compute the entries of P using a variant of the Fourier domain formula shown in Equation (11) below, which is a consequence of Parseval's relation.

$$A_{jl} = \sum_{\tau} \Psi_j(\tau)\Psi_l(\tau) = \frac{1}{2\pi}\int \hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega)d\omega, \tag{11}$$

where $\hat{\Psi}_j(\omega)$ denotes the Fourier transform of $\Psi_j(\tau)$, which is equal to the squared modulus of the Fourier transform of the non-decimated wavelet coefficients, $\hat{\psi}_j(\omega)$. Explicitly,

$$\hat{\Psi}_j(\omega) = \left|\hat{\psi}_j(\omega)\right|^2 = 2^{-j}\left|m_1\left(2^{-(j+1)}\omega\right)\right|^2 \prod_{l=0}^{-(j+2)} \left|m_0\left(2^l\omega\right)\right|^2, \tag{12}$$

where $m_0(\omega) = 2^{-1/2}\sum_k h_k \exp(-i\omega k)$ is the transfer function; $\{h_k\}$ is is the high-pass quadrature mirror filter with $\sum_k h_k^2 = 1$ and $\sum_k h_k = 2^{1/2}$; and $|m_1(\omega)|^2 = 1 - |m_0(\omega)|^2$.

The formula for the Fourier transform of the non-decimated wavelets given in Equation (12) and the corresponding formulae for $m_0(\omega)$ and $m_1(\omega)$ for the Shannon wavelet can be found using the Fourier transform of the continuous time mother and father wavelets, which can be found in [4], pages 46 and 64. Define the set C_j, for $j < 0$, to be

$$C_j = \left[-\frac{\pi}{2^{-j-1}}, -\frac{\pi}{2^{-j}}\right] \cup \left[\frac{\pi}{2^{-j}}, \frac{\pi}{2^{-j-1}}\right].$$

As in [21], the Fourier transform of the non-decimated Shannon wavelets is given by

$$\hat{\psi}_j(\omega) = -2^{-j/2}\exp(-2^{-j-1}i\omega)\mathbb{I}_{C_j}(\omega),$$

where $\mathbb{I}_{C_j}$ is the indicator function on the set C_j. From this the Fourier transform of the autocorrelation wavelets can be obtained as

$$\hat{\Psi}_j(\omega) = \left|\hat{\psi}_j(\omega)\right|^2 = 2^{-j}\mathbb{I}_{C_j}(\omega).$$

Using Parseval's relation and the shifting property of the Fourier transform, we have that

$$\begin{aligned} D^1_{jl} &= 2\sum_\tau \Psi_j(\tau)(\Psi_l(\tau) - \Psi_l(\tau+1)) \\ &= \frac{1}{2\pi}\int \hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega)d\omega - \frac{1}{2\pi}\int e^{-i\omega}\hat{\Psi}_j(\omega)\hat{\Psi}_l(\omega)d\omega. \end{aligned}$$

The first term in the sum is exactly the entries of the original A-matrix which are given by $A_{jj} = 2^{-j}$ for $j < 0$, $A_{jl} = 0$ for $j \neq l$. When $j \neq l$, the second term is equal to zero since the supports of different $\hat{\Psi}_j(\omega)$ do not overlap. Thus, the matrix is diagonal, and when $j = l$, the second term is given by

$$\begin{aligned} -\sum_\tau \Psi_j(\tau)\Psi_j(\tau+1) &= -\frac{1}{2\pi}\int e^{-i\omega}\hat{\Psi}_j(\omega)^2 d\omega \\ &= -\frac{2^{-2j-1}}{\pi}\int e^{-i\omega}\mathbb{I}_{C_j}(\omega)d\omega \\ &= -\frac{2^{-2j-1}}{\pi}\left(\int_{-2^{j+1}\pi}^{-2^j\pi} e^{-i\omega}d\omega + \int_{2^j\pi}^{2^{j+1}\pi} e^{-i\omega}d\omega\right) \\ &= -\frac{2^{-2j-1}}{\pi}\left(\left[ie^{-i\omega}\right]_{-2^{j+1}\pi}^{-2^j\pi} + \left[ie^{-i\omega}\right]_{2^j\pi}^{2^{j+1}\pi}\right) \\ &= -\frac{2^{-2j-1}i}{\pi}\left(e^{2^j\pi i} - e^{2^{j+1}\pi i} + e^{-2^{j+1}\pi i} - e^{-2^j\pi i}\right). \end{aligned}$$

Now, expanding the complex exponential terms into trigonometric functions, we obtain

$$\begin{aligned} -\sum_\tau \Psi_j(\tau)\Psi_j(\tau+1) = -\frac{2^{-2j-1}i}{\pi}\big\{&\cos(2^j\pi) + i\sin(2^j\pi) - \cos(2^{j+1}\pi) \\ &-i\sin(2^{j+1}\pi) + \cos(-2^{j+1}\pi) + i\sin(-2^{j+1}\pi) \\ &-\cos(-2^j\pi) - i\sin(-2^j\pi)\big\}. \end{aligned}$$

After much simplification in which the cosine terms will vanish, we obtain

$$-\sum_\tau \Psi_j(\tau)\Psi_j(\tau+1) = \frac{2^{-2j}}{\pi}\left(\sin(2^j\pi) - \sin(2^{j+1}\pi)\right).$$

Hence, the diagonal entries of the matrix D^1 are given by

$$D^1_{jj} = 2^{-j+1} + \frac{2^{-2j+1}}{\pi}\left(\sin(2^j\pi) - \sin(2^{j+1}\pi)\right).$$

Hence, $P_{jj} = 2^{-2j+1} + 2^{-3j+1}\left[\sin(2^j\pi) - \sin(2^{j+1}\pi)\right]/\pi$, while the off-diagonal terms are zero. Next, approximate the diagonal terms using a Maclaurin series:

$$\begin{aligned} P_{jj} &\approx 2^{-2j+1} + \frac{2^{-3j+1}}{\pi}\left(2^j\pi - \frac{(2^j\pi)^3}{6} - 2^{j+1}\pi + \frac{(2^{j+1}\pi)^3}{6}\right) \\ &= 2^{-2j+1} + 2^{-2j+1} - 2^{-2j+2} - \frac{\pi^2}{3} + \frac{8\pi^2}{3} \\ &= \frac{7\pi^2}{3}. \end{aligned}$$

Therefore, the matrix is diagonal, with all diagonal entries being uniformly bounded away from 0, with $P_{jj} \geq P_{-1,-1} \approx 13.09$ for all j. Hence, we have shown that there exists some $\delta > 0$ such that $\lambda_{\min}(P) \geq \delta$.

B.4. Proof of Theorem 3.5

As $T \to \infty$, the eigenvalues of D^1 tend to zero and hence, when viewed as an operator acting on the sequence space $\ell^2(\mathbb{N})$, its inverse is unbounded. In the locally stationary Fourier time series setting, a similar relationship is found, as given in Equation (5) of [24]. Loosely speaking, the original spectrum at frequency ω is related to the differenced one through multiplication of the term $|1 - e^{-i\omega}|^{-2}$. As $\omega \to 0$ (corresponding to low frequencies) the equation blows up. This mirrors our scenario, where, as the correction matrix grows in size – and we consider coarse-scale (low frequency) behaviour – the inverse matrix norm becomes larger.

We can account for the unboundedness of the inverse of D^1 by using a rescaling of the LSW process itself. Consider the auxiliary process defined by $\epsilon_t = \sum_{j,k} \tilde{w}_{jk}\tilde{\psi}_{j,k-t}\xi_{lm}$, where $\tilde{w}_{jk} = 2^{j/4}w_{jk}$ and $\tilde{\psi}_{j,k-t} = 2^{-j/4}\psi_{j,k-t}$. Then, the expectation of the raw wavelet periodogram (with respect to the rescaled wavelet) is given by

$$\mathbb{E}(\tilde{I}^j_k) = \sum_l P_{jl}\tilde{S}_l(k/T) + \mathcal{O}(T^{-1}), \tag{13}$$

where $\tilde{S}_j(k/T) = 2^{j/2}S_j(k/T)$ and $P_{jl} = 2^{-j/2}D^1_{jl}2^{-l/2}$. We can therefore use the (bounded) P-inverse matrix to correct the smoothed, rescaled periodogram, and then multiply by $2^{-j/2}$, since $S_j(k/T) = 2^{-j/2}\tilde{S}_j(k/T)$. To determine the appropriate threshold for the wavelet-based estimator, we use the following lemma:

Lemma B.2. *For a Gaussian trend LSW process and using a wavelet ψ' of bounded variation, the wavelet coefficients $\hat{v}^j_{rs}$, with $2^r = o(T)$, obey uniformly*

in s,

$$\mathbb{E}(\hat{v}_{rs}) - \int_0^1 \sum_n P_{jn}\tilde{S}_n(z)\psi'_{rs}(z)dz = \mathcal{O}\left(2^{r/2}T^{-1}\right),$$

and

$$Var(\hat{v}_{rs}) = 2T^{-1}\int_0^1 \left(\sum_n P_{jn}\tilde{S}_n(z)\right)^2 \psi'^2_{rs}(z)dz + \mathcal{O}\left(2^r T^{-2}\right).$$

Lemma B.2 is analogous to Theorem 3 of [21]. The result of mean square consistency follows due to a combination of Equation (13) and Theorem 4 of [21]. The mean squared error of the smoothed, corrected wavelet periodogram is given by

$$\begin{aligned}\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] &= 2^{-j}\mathbb{E}\left[\int_0^1 \left(\hat{\tilde{S}}_j(z) - \tilde{S}_j(z)\right)^2 dz\right] \\ &\leq 2^{-j+1}\mathbb{E}\left[\int_0^1 \left(\sum_{l=-J}^{-1}\left(\hat{I}^l_{\lfloor zT\rfloor} - \Lambda_l(z)\right)P_{jl}^{-1}\right)^2 dz\right] \\ &\quad + 2^{-j+1}\int_0^1 \left(\sum_{l<-J}\Lambda_l(z)P_{jl}^{-1}\right)^2 dz,\end{aligned}$$

where $\hat{I}^l_{\lfloor zT\rfloor}$ is the smoothed estimate of the rescaled raw wavelet periodogram and $\Lambda_l(z) = \sum_n P_{nl}\tilde{S}_n(z)$. The first term can be bounded as

$$\begin{aligned}I &\leq 2^{-j+1}\left(\sum_{l=-J}^{-1} P_{jl}^{-1}\left(\mathbb{E}\left[\int_0^1 \left(\hat{I}^l_{\lfloor zT\rfloor} - \Lambda_l(z)\right)^2 dz\right]\right)^{1/2}\right)^2 \\ &\leq 2^{-j+1}\left(\sum_{l=-J}^{-1} P_{jl}^{-1}\mathcal{O}\left(T^{-2/3}\log^2(T)\right)^{1/2}\right)^2 = \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right),\end{aligned}$$

which follows since P possesses a bounded inverse with exponentially decaying entries, and using the rate of convergence of the mean squared error of the wavelet thresholding estimator derived in [22], Theorem 3.1 A. The second term is asymptotically dominated by the first, since it can be bounded as

$$\begin{aligned}II &\leq 2^{-j+1}\left(\sum_{l<-J}\Lambda_l(z)P_{jl}^{-1}\right)^2 \leq 2^{-j+1}\left(\sum_{l<-J}P_{jl}^{-1}\sum_{n=-\infty}^{-1}2^{-l/2}D^1_{\ln}S_n(z)\right)^2 \\ &\leq 2^{-j+1}\left(\sum_{l<-J}\mathcal{O}(2^{l/2})\sum_{n=-\infty}^{-1}S_n(z)\right)^2 = \mathcal{O}(2^{-j}\times 2^{-J}) = \mathcal{O}(2^{-j}T^{-1}),\end{aligned}$$

which follows since P_{jl}^{-1} is bounded, $D^1_{\ln} = \mathcal{O}(2^l)$, $\sum_n S_n(z) < \infty$, and $T = 2^J$. Hence, the mean squared error is given by $\mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right)$.

B.5. Proof of Proposition 3.6

We can write the mean squared error as

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] \leq 2\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J_0}^{-1} \left(\widehat{S}_j(z) - S_j(z)\right)\Psi_j(\tau)\right)^2 dz\right] + 2R_{J_0},$$

where R_{J_0} can be bounded as

$$R_{J_0} = \left(\sum_{j<-J_0} S_j(z)\Psi_j(\tau)\right)^2 \leq \left(\sum_{j<-J_0} S_j(z)\right)^2$$
$$\leq \left(\sum_{j<-J_0} \mathcal{O}(2^{\gamma j})\right)^2 = \mathcal{O}(2^{-2\gamma J_0}) = \mathcal{O}(T^{-2\gamma\alpha}) = \mathcal{O}(T^{\alpha-2/3}).$$

For the first term, we obtain

$$\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J_0}^{-1} \left(\widehat{S}_j(z) - S_j(z)\right)\Psi_j(\tau)\right)^2 dz\right]$$
$$\leq \left(\sum_{j=-J_0}^{-1} \Psi_j(\tau)\left(\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right]\right)^{1/2}\right)^2$$
$$\leq \left(\sum_{j=-J_0}^{-1} \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right)^{1/2}\right)^2$$
$$= \mathcal{O}\left(\left(T^{\alpha-2/3}\log^2(T)\right)^{1/2}\right)^2 = \mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right),$$

by Equation (6) from the main text, and using that $\Psi_j(\tau) \leq 1$ for all j and τ. Hence, provided that $T^{\alpha-2/3}\log^2(T) \to 0$ as $T \to \infty$, the estimator is mean square consistent.

B.6. Proof of Proposition 3.7

In order to derive the formula for the expectation of the squared wavelet coefficients of a general n-th difference, we require the formula the n-th difference itself, which is a well-known result. Denote by $\nabla^n X_t$ the n-th difference of the time series $\{X_t\}$ at time t. Then,

$$\nabla^n X_t = \sum_{k=0}^{n}(-1)^k \binom{n}{k} X_{t-k} = \sum_{k=0}^{n}(-1)^k \binom{n}{k}\epsilon_{t-k} + \mathcal{O}(T^{-1}), \qquad (14)$$

which follows from the differentiability assumption of the trend μ. Now, observe that, for an n-th difference, the expectation will involve the sum of the spectrum over all scales, multiplied by a linear combination of lagged inner product A-matrix entries, denoted A^τ, from lag 0 to lag n. To calculate the coefficient in front of each of the A^τ, we simply calculate the sum of the coefficients of the $\{\epsilon_t \epsilon_s\}$ for each particular lag in the expansion obtained by squaring the n-th difference of the time series, for which we can use Equation (14).

For example, to calculate the coefficient of A, we add together the coefficients of the squared terms in the square of the differenced series, i.e. add the coefficients of $\epsilon_t^2, \epsilon_{t-1}^2, \ldots \epsilon_{t-n}^2$. For the coefficient of A^1, we add together the coefficients of the terms in the square of the difference that differ in index by 1, i.e. we add the coefficients of $\epsilon_t\epsilon_{t-1}, \epsilon_{t-1}\epsilon_{t-2}, \epsilon_{t-1}\epsilon_t, \ldots, \epsilon_{n-1}\epsilon_n$. In particular, if we were interested in the third difference, then the coefficient in front of A would be $1+9+9+1$, and the coefficient in front of A^1 would be $-3-3-9-9-3-3 = -30$, and so on. Hence the formula for the coefficient in front of A is given by

$$\sum_{r=0}^{n} \binom{n}{r}^2 = \binom{2n}{n},$$

and similarly, the formula for the coefficient in front of A^τ, for $\tau \geq 1$, is given by

$$2(-1)^\tau \sum_{r=0}^{n-\tau} \binom{n}{r}\binom{n}{r+\tau} = 2(-1)^\tau \binom{2n}{n+\tau},$$

where the multiplication by 2 arises due to symmetry (for example we must add both the coefficients of $\epsilon_{t-1}\epsilon_{t-2}$ and $\epsilon_{t-2}\epsilon_{t-1}$). The equality on the right follows from a counting argument. The number of ways to choose $n+\tau$ objects from $2n$ choices is the same as the number of ways of choosing r objects from the first n and choosing $r+\tau$ from the remaining n objects, for $0 \leq r \leq n-\tau$. The expectation is accurate up to order $\mathcal{O}(T^{-1})$ by the same argument as in the proof of Proposition 3.3. Hence, the squared expectation of the wavelet coefficients of the n-th differenced series are given by

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l S_l\left(\frac{k}{T}\right)\left[\binom{2n}{n}A_{jl} + 2\sum_{\tau=1}^{n}(-1)^\tau \binom{2n}{n+\tau}A_{jl}^\tau\right] + \mathcal{O}(T^{-1}).$$

B.7. Proof of Theorem 3.9

By the Daubechies characterisation of Hölder spaces, rescaling of the LSW process, and Proposition 3.3, the expectation of the rescaled raw wavelet periodogram of the differenced time series is given by

$$\mathbb{E}(\tilde{I}_k^j) = \sum_l P_{jl}\tilde{S}_l(k/T) + \mathcal{O}(2^{-7j/2}T^{-2}) + \mathcal{O}(T^{-1}). \tag{15}$$

Hence, the mean squared error of the smoothed wavelet periodogram is given by

$$\mathbb{E}\left[\int_0^1 \left(\widehat{S}_j(z) - S_j(z)\right)^2 dz\right] = 2^{-j}\mathbb{E}\left[\int_0^1 \left(\hat{\tilde{S}}_j(z) - \tilde{S}_j(z)\right)^2 dz\right]$$
$$\leq 2^{-j+1}\mathbb{E}\left[\int_0^1 \left(\sum_{l=-J_1}^{-1} \left(\hat{I}^l_{\lfloor zT \rfloor} - \Lambda_l(z)\right) P^{-1}_{jl}\right)^2 dz\right]$$
$$+ 2^{-j+1}\int_0^1 \left(\sum_{l<-J_1} \Lambda_l(z) P^{-1}_{jl}\right)^2 dz$$

where $\hat{I}^l_{\lfloor zT \rfloor}$ is the smoothed estimate of the raw wavelet periodogram and $\Lambda_l(z) = \sum_n P_{nl}\tilde{S}_n(z)$. The first term can be bounded as

$$I \leq 2^{-j+1}\left(\sum_{l=-J_1}^{-1} P^{-1}_{jl}\left(\mathbb{E}\left[\int_0^1 \left(\hat{I}^l_{\lfloor zT \rfloor} - \Lambda_l(z)\right)^2 dz\right]\right)^{1/2}\right)^2$$
$$\leq 2^{-j+1}\left(\sum_{l=-J_1}^{-1} P^{-1}_{jl}\left(\mathcal{O}\left(2^{-7l}T^{-4}\right) + \mathcal{O}\left(T^{-2/3}\log^2(T)\right)\right)^{1/2}\right)^2$$
$$= \mathcal{O}\left(2^{-j}T^{7\beta-4}\right) + \mathcal{O}\left(2^{-j}T^{-2/3}\log^2(T)\right),$$

which follows since P possesses a bounded inverse with exponentially decaying entries, and using Equation (15). The second term can be bounded in the same fashion as in the proof of Theorem 3.5, and is of order $\mathcal{O}(2^{-j}T^{-\beta})$, which gives the stated consistency result. Similarly, the mean squared error of the LACV estimator is calculated as

$$\mathbb{E}\left[\int_0^1 \left(\hat{c}(z,\tau) - c(z,\tau)\right)^2 dz\right] \leq 2\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J_0}^{-1} \left(\widehat{S}_j(z) - S_j(z)\right)\Psi_j(\tau)\right)^2 dz\right]$$
$$+ 2R_{J_0},$$

where R_{J_0} is asymptotically negligible by the argument in Proposition 3.6. For the first term, we obtain

$$\mathbb{E}\left[\int_0^1 \left(\sum_{j=-J_0}^{-1} \left(\hat{S}_j(z) - S_j(z)\right)\Psi_j(\tau)\right)^2 dz\right]$$
$$\leq \left(\sum_{j=-J_0}^{-1} 2^{-j/2}\left(\mathcal{O}\left(T^{7\beta-4}\right) + \mathcal{O}\left(T^{-2/3}\log^2(T)\right) + \mathcal{O}(T^{-\beta})\right)^{1/2}\right)^2$$

$$= \left(\mathcal{O}\left(T^{7\beta-4}\right) + \mathcal{O}\left(T^{-2/3}\log^2(T)\right) + \mathcal{O}(T^{-\beta})\right)\left(\sum_{j=-J_0}^{-1} 2^{-j/2}\right)^2$$

$$= \mathcal{O}\left(T^{\alpha+7\beta-4}\right) + \mathcal{O}\left(T^{\alpha-2/3}\log^2(T)\right) + \mathcal{O}(T^{\alpha-\beta}).$$

B.8. Proof of Proposition 4.1

The appropriate threshold is derived by using an analogous result to B.2, from which we obtain

$$\mathbb{E}(\hat{v}_{rs}) - \int_0^1 \mu(z)\psi'_{rs}(z)dz = \mathcal{O}\left(2^{r/2}T^{-1}\right),$$

and

$$\mathrm{Var}(\hat{v}_{rs}) = T^{-1}\int_0^1 \sum_n S_n(z)\psi'^2_{rs}(z)dz + \mathcal{O}\left(2^r T^{-2}\right).$$

The mean squared error rate is obtained by using Theorem 1 of [32], with the specific case of a Lipschitz continuous trend.

Acknowledgments

The authors thank the two anonymous reviewers for helpful suggestions which led to an improved manuscript.

References

[1] Brockwell, P. J., Davis, R. A. and Fienberg, S. E. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media. MR1093459
[2] Cardinali, A. and Nason, G. P. (2017). Locally stationary wavelet packet processes: basis selection and model fitting. *Journal of Time Series Analysis* **38** 151–174. MR3611738
[3] Chan, K. H., Hayya, J. C. and Ord, J. K. (1977). A note on trend removal methods: the case of polynomial regression versus variate differencing. *Econometrica (pre-1986)* **45** 737.
[4] Chui, C. K. (1997). *Wavelets: A Mathematical Tool for Signal Analysis*. SIAM. MR1443204
[5] Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant denoising. In *Wavelets and Statistics* 125–150. Springer.
[6] Dahlhaus, R. and Neumann, M. H. (2001). Locally adaptive fitting of semiparametric models to nonstationary time series. *Stochastic processes and their applications* **91** 277–308. MR1807680
[7] Das, S. and Politis, D. N. (2020). Predictive inference for locally stationary time series with an application to climate data. *Journal of the American Statistical Association* 1–16. MR4270034

[8] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM. MR1162107
[9] Dette, H. and Wu, W. (2022). Prediction in locally stationary time series. *Journal of Business & Economic Statistics* **40** 370–381. MR4356579
[10] Dette, H., Wu, W. et al. (2019). Detecting relevant changes in the mean of nonstationary processes – a mass excess approach. *The Annals of Statistics* **47** 3578–3608. MR4025752
[11] Fryzlewicz, P. and Nason, G. P. (2006). Haar–Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 611–634. MR2301011
[12] Hu, L., Huang, T. and You, J. (2019). Two-step estimation of time-varying additive model for locally stationary time series. *Computational Statistics & Data Analysis* **130** 94–110. MR3860531
[13] Khismatullina, M. and Vogt, M. (2020). Multiscale inference and long-run variance estimation in non-parametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82** 5–37. MR4060975
[14] Killick, R. and Eckley, I. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software* **58** 1–19.
[15] Killick, R., Fearnhead, P. and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107** 1590–1598. MR3036418
[16] McGonigle, E. T., Killick, R. and Nunes, M. A. (2022). Trend locally stationary wavelet processes. *Journal of Time Series Analysis (to appear)*.
[17] Nason, G. (2008). *Wavelet Methods in Statistics with R*. Springer Science & Business Media. MR2445580
[18] 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. MR3124795
[19] Nason, G. (2016). locits: Tests of stationarity and localized autocovariance R package version 1.7.3.
[20] Nason, G. (2016). wavethresh: Wavelets statistics and transforms R package version 4.6.8.
[21] Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62** 271–292. MR1749539
[22] Neumann, M. H. and Von Sachs, R. (1995). Wavelet thresholding: beyond the Gaussian IID situation. In *Wavelets and Statistics* 301–329. Springer. MR1364677
[23] Reichel, L. and Trefethen, L. N. Eigenvalues and pseudo-eigenvalues of Toeplitz matrices. *Linear Algebra and its Applications*. MR1148398
[24] Roueff, F. and Von Sachs, R. (2011). Locally stationary long memory estimation. *Stochastic Processes and Their Applications* **121** 813–844. MR2770908
[25] Roueff, F., Von Sachs, R. et al. (2019). Time-frequency analysis of

locally stationary Hawkes processes. *Bernoulli* **25** 1355–1385. MR3920375

[26] Sanderson, J., Fryzlewicz, P. and Jones, M. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika* **97** 435–446. MR2650749

[27] Shumway, R. H. and Stoffer, D. S. (2010). *Time Series Analysis and its Applications: With R Examples*. Springer Science & Business Media. MR2721825

[28] Tunyavetchakit, S. (2010). *On the optimal segment length for tapered Yule-Walker estimates for time-varying autoregressive processes*. Heidelberg.

[29] Van Bellegem, S., Von Sachs, R. et al. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *The Annals of Statistics* **36** 1879–1924. MR2435459

[30] Vidakovic, B. (2009). *Statistical Modeling by Wavelets*. John Wiley & Sons. MR1681904

[31] Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics* **40** 2601–2633. MR3097614

[32] Von Sachs, R. and MacGibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics* **27** 475–499. MR1795776

[33] von Sachs, R., Nason, G. P. and Kroisandt, G. (1997). Adaptive estimation of the evolutionary wavelet spectrum. *University of Kaiserslautern, Germany, and University of Bristol, UK*.