ADAPTIVE ESTIMATION OF THE EVOLUTIONARY WAVELET SPECTRUM

by

Rainer von Sachs, Guy P. Nason and Gerald Kroisandt

TECHNICAL REPORT NO. 516
October 1997

Partially supported by NSF DMS-95-05151

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Adaptive Estimation of the Evolutionary Wavelet Spectrum

by

Rainer von Sachs, Guy P. Nason and Gerald Kroisandt

TECHNICAL REPORT NO. 516
October 1997

Partially supported by NSF DMS-95-05151

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Adaptive estimation of the evolutionary wavelet spectrum

Rainer von Sachs
University of Kaiserslautern, Germany

Guy P. Nason
University of Bristol, UK

and

Gerald Kroisandt
University of Kaiserslautern, Germany

October 15, 1997

Abstract

In this work we define and study a new class of non-stationary random processes which are characterized by a representation with respect to a family of localized basis functions. Using non-decimated or "stationary" wavelets this generalizes the Cramér (Fourier) spectral representation of stationary time series. We provide a time-scale instead of a time-frequency decomposition and, hence, instead of thinking as scale in terms of "inverse frequency" we start from genuine time-scale building blocks or "atoms".

Using our new model of "locally stationary wavelet" processes, we develop a theory how to define and estimate an "evolutionary wavelet spectrum". Our asymptotics are based on rescaling in time-location which allows us to perform rigorous estimation starting from a single stretch of observations of the process. This wavelet spectrum measures the local power in the variance-covariance decomposition of the process at a certain scale and a (rescaled) time location.

To estimate the wavelet spectrum we use (corrected and appropriately smoothed) "wavelet periodograms". Further, we suggest an inverse transformation of the smoothed wavelet periodogram that estimates a local autocovariance of the original stochastic process. Some numerical simulations and an application to a medical time series, which shows a particular non-stationary behaviour, indicate the usefulness of our new approach.

1 Introduction

This article introduces a way of representing non-stationary time series, i.e. data with a time-varying second order structure, in terms of building blocks (wavelets) that are localized in time and scale. If a time series is stationary then classical theory provides optimal and well-tested means for its analysis. However, we would submit that the majority of actual time series are, in fact, not stationary. We present some actual non-stationary time series from biomedical research although non-stationary series turn up in all kinds of places such as sound and speech processing, geophysics, neurophysiology, bioacoustics, sonar, radar, mobile radio communication and econometrics.

Statisticians are already familiar with the concept of scale. For stationary processes scale is nothing more than the usual lag of the autocovariance function. Any person who has examined the autocovariance or autocorrelation function of a time series at different lags has analysed that time series at different scales.

Our objective is the development of a rigorous theory for the modelling of time series using time-scale localized building blocks. This theory gives us a tool that permits us to quantify how the autocovariance of a non-stationary time series changes as the series evolves through time.
(i.e. a time-varying autocovariance). We obtain the time and scale localisation by modelling our time series in terms of wavelets.

We stress that ours is by no means the only method for the analysis of non-stationary time series. Time series can be non-stationary in many ways and several methods already exist. This article contributes methodology which will be "optimal" for a certain class of non-stationary processes but which will also present useful interpretable information for many others. Our methodology is intended to be complementary to these other methods and to be another tool in the toolbox rather than overthrowing existing methodology.

The remainder of Section 1 sets the scene for the rest of the paper: it reiterates the concepts from stationary time series theory that we shall adapt later for non-stationary processes and reviews recent work on non-stationary time series and processes based on wavelets. Section 2 provides an overview of the main ingredients of our time-scale approach and informally introduces our process representation, the wavelet spectrum, the wavelet periodogram and a localized autocovariance (these are our analogues to the Cramér representation, spectrum, periodogram and autocovariance from stationary theory). Section 3 shows how each of these new quantities can be used on real data. Sections 4, 5 and 6 concentrate on the theoretical foundations of the main ingredients and show how to estimate them in a statistically rigorous way. In Section 4 we develop our model of locally stationary wavelet processes to allow for a rigorous theory. Evolutionary wavelet spectrum and local autocovariances are defined and discussed as well as connections to stationary processes and related concepts of local stationarity. Section 5 investigates the wavelet periodogram and gives some relation to the classical periodogram. In Section 6, we develop asymptotic results for smoothing wavelet periodograms, obtain in particular the rate of the $L_2$-risk between a wavelet threshold estimator and the wavelet spectrum, and define a consistent estimator for the local autocovariance. Proofs are deferred to an Appendix.

1.1 Stationary processes and divergence from stationarity

The classical Cramér decomposition of a stationary stochastic process $X_t$, $t \in \mathbb{Z}$, may be written

$$X_t = \int_{-\pi}^{\pi} \exp(i\omega t) \, dZ(\omega) = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t) \, d\zeta(\omega),$$  

(1)

where $dZ(\omega)$ and $d\zeta(\omega)$ are orthogonal and orthonormal increment processes respectively (see Priestley [25]). In recent years there have been several generalizations away from the assumption of stationarity. In this article we concentrate on generalizations where the second-order structure of a process changes over time. To motivate our theory recall that the autocovariance function, $c_X(\tau)$, of the stationary stochastic process, $(X_t)$, may be written as

$$c_X(\tau) = \int_{-\pi}^{\pi} f(\omega) \exp(i\omega \tau) \, d\omega,$$  

(2)

for $\tau \in \mathbb{Z}$ and where the spectral density function $f(\omega)$ is linked to $Z(\omega)$ and the spectral distribution function $F(\omega)$ by

$$E|dZ(\omega)|^2 = dF(\omega) = f(\omega) \, d\omega,$$  

(3)

(if $dF(\omega)$ is absolutely continuous with respect to Lebesgue measure).

To permit the second-order structure of $X_t$ to change over time we could introduce a time-dependency into the autocovariance or spectral density functions. Alternatively, we could replace the amplitude function, $A(\omega)$, from (1) by a time-varying version, $A_t(\omega)$ (and likewise replace the spectrum $f(\omega)$ by $f_t(\omega)$). More specific developments of this kind are oscillatory and locally stationary processes (see Priestley [25] and Dahlhaus [8] respectively).

Our approach is somewhat different in that we replace the set of harmonics $\{\exp(i\omega t)\, | \, \omega \in [-\pi,\pi]\}$ by a set of locally supported wavelets. (Wavelets are small oscillations that are localized in the time and Fourier domains, see Nason and Silverman [22] or Strang [29] for an introduction
to wavelets. See Daubechies [10], Meyer [19] or Chui [6] for more authoritative expositions. Recently, local atomic decompositions (wavelets, wavelet libraries) have become popular for the analysis of deterministic signals as alternatives to non-local Fourier representations (Rioul and Vetterli [27], Flandrin [14]). The question immediately arises: is it possible and meaningful to use such atomic decompositions to represent, not necessarily stationary, stochastic processes?

1.2 A biomedical time series example

Figure 1 shows a subsample of a heart rate recording of an 66 day old baby sampled at 0.25Hz from 21:17:59 to 06:27:18. It is unlikely that this will be a stationary time series. For one thing, of interest to paediatricians, the “heart rate” varies considerably over time and changes significantly between periods of sleep and waking. For time series such as these the next sections introduce:

1. possible underlying models based on wavelet representations;

2. a time-localized spectrum to be estimated using a wavelet periodogram;

3. a time-localized autocovariance.

Indeed, all of these concepts are similar to the classical process, spectrum, periodogram and autocovariance except that they are time-localized with the localization provided by wavelets.

For the heart rate example we shall show that a wavelet periodogram smoothed using non-linear wavelet shrinkage can reveal important features of interest that cannot be elicited from the time series or standard periodogram techniques alone. Indeed, some time-frequency methods were also used and these did not reveal much of interest. We return to this example in section 3.

1.3 Related work

Nason and Silverman [23] suggested using the stationary wavelet transform as an exploratory technique for producing “local spectral density estimates” of time series data. This article formalizes the heuristic ideas from Nason and Silverman and provides an underlying and rigorous theory of models, estimation and local autocovariance.
Later, in order to both properly identify and estimate a wavelet spectrum from a single finite-length stretch of the process \( \{ X_t \} \) we need to restrict the time-variation of the second order structure of \( \{ X_t \} \). This is exactly as in the situation of time-varying Fourier spectra \( f_t(\omega) \), where no rigorous (asymptotic) estimation theory is possible without control of an otherwise arbitrary time variation of the spectrum (see Dahlhaus [8], von Sachs and Schneider [33], Neumann and von Sachs [24]).

Most of the existing work on wavelets with stochastic processes (Campanis and Masry [3], Campanis and Houdré [2], Kawasaki and Shibata [17] and Cheng and Tong [5]) does not aim to give a decomposition with respect to an (orthogonal) increment process in the time-scale plane. The first three papers focus on probabilistic approximations and do not cover estimation. Indeed, Kawasaki and Shibata [17] develop a wavelet process representation where they study conditions on the properties of a general measure to provide a weakly (Fourier) stationary process. Also, the paper by Morettin and Chang [20] uses the wavelet periodogram approach based on an orthonormal wavelet basis but only for stationary time series. We should also mention that the idea of using localized representations with truly local basis functions is certainly not completely new. For example section 13.5 of Walter [34] gives an example of a cyclostationary process that can be expanded into a series of box-car (Haar scaling) functions with uncorrelated coefficients. Another related field is the analysis and synthesis of \( 1/f \)–processes with wavelets. In the work of Abry et al. [1], in the particular situation of a stationary increment process, the time marginal of the scalogram was used in order to estimate the spectral exponent of the fractional Brownian motion. One can certainly think about generalizations to other non-stationary situations which call for the more localized wavelet approaches being introduced here.

2 The main ingredients of our approach

In this section we give a brief overview of all the main components of our theory. We do not go into detail as we wish to give the reader a feel for how the components fit together. To give the reader some guidance we appeal to analogous concepts from the theory of stationary time series. Each concept presented here is dealt with rigorously in later sections.

2.1 Preliminaries

We deal with non-stationary processes where the covariance \( c_X(s,t) = \text{cov}(X_s, X_t) \) is no longer a function of \( t - s \) but of both \( s \) and \( t \). To do this we switch from Fourier-based decompositions consisting of complex exponentials, \( e^{i\omega t} \), to a decomposition in terms of a set of locally supported wavelets \( \{ \psi_{j,k}(t) \mid j \in \mathbb{Z}, k \in \mathbb{Z} \} \), where

\[
\psi_{j,k}(t) = 2^{j/2} \psi \left\{ 2^j(t - k) \right\},
\]

for a sufficiently well-concentrated mother wavelet \( \psi(t) \). Note that here we are using the non-decimated version of dilated and shifted \( \psi_{j,k} \)'s where at each scale \( j \) there is the same number of shifted locations \( k \).

By using wavelets we move from the arena of Fourier based time-frequency decompositions to a genuine time-scale decomposition. We avoid thinking of scale as "inverse-frequency" as our processes will be constructed from genuine time-scale building blocks or "atoms", \( \psi_{j,k}(t) \), with random amplitudes \( \eta_{j,k} \):

\[
X_t = \sum_{j=\infty}^{\infty} \sum_{k=\infty}^{\infty} \eta_{j,k} \psi_{j,k}(t), \quad t = 0, \ldots, T - 1,
\]

where the representation is understood in the mean-square sense. Roughly speaking, we expect the amplitude \( \eta_{j,k} \) to be large if at time \( t = k \) there is high correlation of \( X_k \) with \( X_{k-\tau} \) or \( X_{k+\tau} \), for some \( \tau \), that matches the "wavelength" of the atom \( \psi_{j,k}(t) \) which is proportional
to $2^{-j}$ (assuming that $\psi(t)$ is localized at time zero, which it nearly always is in practice). We observe that model (5) permits a local representation by taking advantage of the standard wavelet property that fast (high-frequency) oscillations can change quickly, and slow oscillations can change slowly.

More generally, we could substitute various different families of localized atoms for $\psi(t)$, which by no means need be orthogonal, including versions which are based on (appropriately discretized) continuous wavelet transforms (CWT). Also, we could use, for example, stationary wavelet packets which have the extra index of number of oscillations (see Nason et al. [21]).

In this article, however, the wavelets we use in (4) are indexed by scale and position, only. Furthermore they follow the layout of the non-decimated or stationary wavelet transform (SWT) rather than the discrete orthogonal wavelet transform (DWT) in that at each time point there is a wavelet at each scale (and hence the wavelets defined in (4) differ from the usual wavelet definition). See, again, Nason and Silverman [23] on using the SWT as an exploratory technique for producing "local spectral density estimates" of time series data.

This choice seems to be just the right one if we look for an appropriate discretization of continuously time-scale indexed models which give the compromise between orthogonality and too much overcompleteness. The use of the DWT, though it does allow for a rigorous theory (see [31], and Proposition 5.3 in this article for some motivating properties), does not include traditional stationary processes in the model. CWT-based modelling includes by far too much redundancy to get any meaningful representation with uncorrelated coefficients (see below). As will be explained in more detail below, the SWT, however, allows for rich enough a class of processes with still a sufficient control of redundancy, and leads to a representation which is unique, in a certain sense.

2.2 Wavelet processes, wavelet spectra and autocorrelation wavelets: a simplified illustration

Wavelet processes. The specification of model (5) is not complete as it stands. To obtain a useful theory we make some assumptions about the model (again, we do this rigorously in section 4, in Definition 4.1). We assume that the random amplitudes $\eta_{jk}$ have zero mean and are uncorrelated, i.e. $E\eta_{jk} = 0$ and $\text{cov}(\eta_{jk}, \eta_{lm}) = \delta_{jl}\delta_{km}$, where $\delta_{pq}$ is the Kronecker delta and equal to zero unless $p = q$ when it is 1.

As we consider only zero-mean processes, in practice some trend should be included in the model to adjust for non-zero mean processes. This trend could be modelled and estimated again by wavelet methods (cf. [30]). So here, as always for data with non-zero mean the sample mean should be removed before analysis.

We have also departed from the usual wavelet numbering scheme. The data live on scale zero, scale $-1$ is the scale which contains the finest resolution wavelet detail and scale $-J$ the coarsest (in practice determined by the length $T$ of data, i.e. $J = J(T)$). Note also that there is no scaling coefficient on this coarsest scale; we shall comment on this in more detail later. The advantage of the altered numbering scheme is that we keep the support of the wavelets on the finest scale fixed and constant with respect to the length $T$ of the observed time series. However, as $T$ increases it is possible to observe longer and longer cycles within the series and so the model includes increasingly coarser wavelets. In other words $-J$ should tend to $-\infty$ with increasing $T$ (indeed, with our asymptotic theory later we have $J(T) = \log_2(T)$. For readers familiar with wavelets this is nothing more than the number of levels in a wavelet transform increasing by one as the number of data points doubles).

We cannot have an arbitrary asymptotic increase in coefficients and still have representation (5) converging in mean-square. So we also ensure that the expected value of the squared coefficients, $E\eta^2_{jk}$, decays as $j \to -\infty$.

As a na"ive illustration of this subtle point consider $T = 16$ points sampled every $\frac{1}{16}$ seconds. Then the $-1$ scale points "cover" every $\frac{1}{8}$s, the scale $-2$ points "cover" every $\frac{1}{4}$s and eventually scale $-J$ "covers" the whole series over $1$s. Now suppose we double the number of samples $T = 32$
but still sample every $\frac{1}{15}$ s. Clearly the scale $- J$ now “covers” 2s, not 1s. So the interpretation of
the coarsest scale $- J$ changes as do all scales above which are fixed relative to this scale. This
makes intuitive sense as longer and longer cycles may be observed in the series with growing $T$.

**Wavelet spectrum.** The wavelet spectrum of the process defined in (5) is given by $E\eta_{jk}^2$. The
wavelet spectrum is a measure of power of the process at a particular scale and location. A more
precise definition appears in Definition 4.5. The classical analogue to the wavelet spectrum
is the spectrum $f(\omega)$ of a stationary time series which measures power (contribution to variance)
at frequency $\omega$.

**Autocovariance function.** Just as in classical stationary theory the wavelet spectrum
has a complementary counterpart in the form of a localized autocovariance function (see
Definition 4.8). For the purposes of the current informal discussion suppose for the moment
that $X_t$ is second-order stationary and has zero mean. Our localized autocovariance function
will not depend on location (because of the stationarity) and indeed can be represented as

$$c_X(\tau) = \sum_{j=-\infty}^{-1} S_j \Psi_j(\tau), \quad \tau \in \mathbb{Z},$$  \hspace{1cm} (6)

where

$$\Psi_j(\tau) := \sum_k \psi_{jk}(0) \psi_{jk}(\tau)$$  \hspace{1cm} (7)

and with $S_j := W_{jk}^2 = E\eta_{jk}^2$, independently of the location index $k$, because for stationary $X_t$, $E\eta_{jk}^2$ will be the same for all $k$. (The wavelets in (7) are actually symbols for discrete wavelet
basis functions rather than the actual continuous wavelets themselves. We shall make this more
precise later, in Section 4.1, but for an example for Haar wavelets:

$$\psi_{-1,0}(t) = \frac{1}{\sqrt{2}} I_{(t=0)} \text{ and } \psi_{-1,1} = -\frac{1}{\sqrt{2}} I_{(t=1)}.$$  \hspace{1cm} (8)

and, similarly,

$$\psi_{-2,0}(t) = \psi_{-2,1}(1) = \frac{1}{2} \text{ and } \psi_{-2,2}(2) = \psi_{-2,3}(3) = -\frac{1}{2}.$$  \hspace{1cm} (9)

and so on. The functions $\Psi_j(\tau)$ are actually the autocorrelation functions of these discrete
wavelet symbols, and the collection \{\Psi_j(\tau)\} is the family of autocorrelation wavelets. Thus
our localized autocovariance is a (back-) transform of the wavelet spectrum with respect to these
autocorrelation wavelets, just as the classical autocovariance is the Fourier (back-) transform of the
classical spectrum.

With stationarity the a priori local representation (5) needs to turn into a global
representation: i.e. the representation needs to be shift-equivariant. This is one important
reason why the representation relies on the stationary (non-decimated) wavelets in (4). The
autocorrelation function of the stationary wavelets, $\Psi_j(\tau)$ provides a new family, which is not
localized, but is averaged over all stationary wavelet locations, $k$, at scale $j$. This makes sense
as the corresponding spectral coefficients $E\eta_{jk}$ are constant over $k$ for stationary processes. We
shall illustrate how our method works for stationary processes in an example below.

Stationary processes are useful for providing simplified explanations for our theory but
our raison d'être is to deal with processes that are not stationary. Later we shall consider
covariance function $c(t,\tau) = \text{cov}(X_t, X_{t+\tau})$ and spectral coefficients $S_j(t)$ that are time-
dependent. However, the time-dependence is obtained by considering a rescaled time, $z = t/T$, rather than actual time $t$. The introduction of rescaled time allows a rigorous asymptotic theory
in which estimation of (rescaled) time-dependent quantities such as $c(t, z)$ and $S_j(z)$ is possible.
This time-rescaling theory follows the ideas of Dahlhaus's [8] locally stationary processes but
here we track the local power in the covariance decomposition of the process, $X_t$, with respect
to scales (instead of frequencies) along time.
Before we go any further with our description of the statistical aspect of our work we have to provide a little more of the necessary mathematical detail on how we are going to use the stationary wavelets. In Definition 4.12 we shall introduce the (infinite) matrix \( A = (A_{jk})_{j,k=-1,-2,\ldots}, \) where \( A_{jk} = \langle \Psi^*_j(\tau), \Psi_k(\tau) \rangle \), which measures the redundancy in the non-orthogonal autocorrelation wavelet family. The \( A \) matrix is symmetric, defines an elliptic operator (possesses a positive minimum eigenvalue) and hence has bounded inverse \( A^{-1} \) and, in particular, shows that the collection \( \{ \Psi_j(\tau) \}_{j<0} \) is linearly independent. Moreover, it will be shown later (Theorem 4.13 and Proposition 4.14) that in (6) that equality of two autocovariance sequences implies equality of the two corresponding sequences of wavelet spectra. That is, though the process representation (with respect to an overcomplete set of basis functions) itself cannot be unique in the traditional sense, the one of the autocovariance (with respect to the set of autocorrelation wavelets) actually is. Each autocovariance sequence gives rise to a unique set of wavelet spectra through an inverse transform of (6) (which, for the general non-stationary situation, will be given in Proposition 4.14),

\[
S_j = \sum_{\tau} c_x(\tau) \kappa_j(\tau), \quad j = -1, -2, \ldots, \tag{10}
\]

where (the vector) \( \kappa(\tau) = \{ \kappa_j(\tau) \}_{j=-1,-2,\ldots} \) is defined via \( \kappa(\tau) = A^{-1} \Psi(\tau) \). Observe that (10) is unambiguous because application of the inverse of the operator \( A \) on the vector \( \{ \Psi_j(\tau) \}_{j=-1,-2,\ldots} \) factors out the redundancy in this non-orthogonal autocorrelation wavelet "basis". Note also that \( (\Psi(\tau), \kappa(\tau)) \) can be seen as a pair of dual "bi-orthogonal bases", with the operator \( A \) playing the role of the "kernel" of \( \Psi(\tau) \) and the operator \( A^{-1} \) the one of \( \kappa(\tau) \) (see Chui [6], Theorem 3.20). (The dual-basis analogy to \( L^2 \)-Hilbert space theory is only for illustration and should be considered with caution). We could even switch to a symmetrized pair of "forward" and "inverse" transformation in (6) and (10) by application of \( A^{-1/2} \) to \( \{ \Psi_j(\tau) \} \), which is in fact its orthogonalization. \( \tilde{\kappa}_j(\tau) = A^{-1/2} \Psi_j(\tau) \) (as vectors in \( \tau \)) is a set of orthogonal vectors (over \( j \)), and the wavelet spectrum \( S_j \) could be seen as the projection coefficients of the autocovariances \( c_x(\tau) \) onto this set. Nevertheless, for our approach, we choose the basically equivalent first normalization (with \( A^{-1} \)), simply because we want to keep some of the nice intuitive properties of the family \( \{ \Psi_j(\tau) \} \) (which we will discuss below).

### 2.3 A motivating example: stationary moving average processes

This section provides a motivating example that shows how our representation (6) can be useful for stationary processes (the next section looks at a motivating non-stationary example).

The aim of this section is to show that there are processes for which representation (6) delivers a much sparser representation when compared to the corresponding Fourier one given by (2). Sparse representations are generally easier to estimate as well as being appealing in their own right.

The processes for which our representation is sparse are a special collection of MA\((q)\) processes that have the discrete wavelet basis function symbols as coefficients. Each different compactly supported wavelet determines a different class of MA\((q)\) processes that can be sparsely represented using the autocorrelation family \( \{ \Psi_j(\tau) \} \) (the following discussion concerns Haar but it could equally apply to any other compactly supported wavelet). So, for example, let

\[
X^1_t = \frac{1}{\sqrt{2}} \epsilon_t - \frac{1}{\sqrt{2}} \epsilon_{t-1},
\]

where \( \{ \epsilon_t \} \) is a purely random process with mean zero and variance of \( \sigma^2 \). The MA\((1)\) process \( X^1(t) \) has an autocovariance function \( c^1_X(\tau) = \delta_{\tau,0} - \frac{1}{2} \delta_{|\tau|,1} \). For Haar wavelets \( \Psi_{-1}(\tau) = c^1_X(\tau) \) for all \( \tau \) and so that in representation (6) only the coefficient \( S_{-1} \) will be different from zero. The next process in this scheme is the MA\((2)\) process, \( X^2_t \), with coefficients motivated from (9). Its covariance function has representation (6) where only the \( S_{-2} \) coefficient is non-zero. This
Figure 2: A concatenation of \( n = 128 \) observations successively from the Haar moving average processes \( X^1, X^2, X^3 \) and \( X^4 \). The variance of the underlying purely random process was \( \sigma^2 = 1 \). The vertical dotted lines indicate where process \( X^r \) changes to process \( X^{r+1} \) for \( r = 1, 2, 3 \).

The scheme can be continued for \( X^2, X^4 \) and so on, allowing for representations by only \( S_{-3}, S_{-4} \) and so on. We shall call the collection of processes \( \{X^r_t\}_{r=1}^{\infty} \) the Haar moving average processes.

In fact, any MA process can be represented by a linear combination of the Haar MA processes and in quite a number of cases a sparse representation. This is simply a consequence of the fact that any sequence in \( L^2(\mathbb{Z}) \) can be decomposed into a set of Haar basis functions. In parallel, the (Haar) autocorrelation wavelets \( \Psi_{-j}(\tau) \), which for each scale \( j \) are themselves positive semidefinite, are well-suited to represent autocovariances with a sufficiently fast decay. Note also that by using other compactly supported wavelets, say in the Daubechies series, it would be possible to find the “best” representation for a given signal. However, the main point of this toy example is that it is possible to decompose processes \( X_t \) into very localized building blocks with random coefficients which results in a sparse autocovariance representation of the form (6). This sort of representation is not really of much use for stationary series but this section shows that it can be done.

### 2.4 Non-stationary processes: A second example

The real benefit of our model shows up if we move on to non-stationary examples. Suppose we now concatenate a set of different moving average processes. For definiteness, suppose we use the Haar moving average processes and concatenate \( n \) observations from each of \( X^1_t, X^2_t, X^3_t \) and \( X^4_t \). Within each of the segments of \( n \) observations the process is stationary but as a process of \( 4n \) observations it is non-stationary. The change from segment to segment over time will be picked up as a change in the wavelet spectrum \( S_j(t) \) (formally it should be \( S_j(z) \) but we permit an abuse of notation here). One realization of such a concatenated series is shown in Figure 2. In practice we can not observe the wavelet spectrum but only an estimate: the (corrected) wavelet periodogram. Figure 3 actually shows the mean of 100 (corrected) wavelet periodogram of 100 simulations of the concatenated MA process. The mean of the 100 simulations provides an estimate of the wavelet spectrum, \( S_j(t) \), over time (so, for example, the bottom line of coefficients in Figure 3 estimates \( S_{-1}(t) \) and shows that \( S_{-1}(t) \) is non-zero only when the MA(1) process, \( X^1(t) \), is “active”. Then, at time \( t = 128 \), the MA(2) process, \( X^2(t) \), becomes active and this
Wavelet Spectrum (Estimate)

![Wavelet Spectrum Diagram]

Figure 3: Mean of 100 corrected wavelet periodogram estimates of $S_j(t)$ for $j = -1, \ldots, -9$. (Each corrected wavelet periodogram (section 2.5) was of an independent simulation of the concatenated MA process described in the text).

is reflected by the non-zero block of $S_{-2}(t)$ coefficients until time $t = 256$, and so on.) For this simulation we have access to as many realizations of the process as we please and can estimate $S_j(t)$ to whatever accuracy we desire. In practice, we would only have access to one realization of the time series and so we smooth the periodogram to obtain good estimates.

For comparison, Figure 4 shows the square of the stationary wavelet coefficients — the uncorrected wavelet periodogram. Nason and Silverman [23] proposed using a smoothed version of Figure 4 as a “spectral estimate”. Comparing the figures demonstrates that our new corrected wavelet periodogram is superior. Figure 3 shows the abrupt changes as one MA process changes into another whereas in Figure 4 the changes are spread out and smoothing will not improve matters. Further, Figure 4 exhibits significant power at levels -6 and -5 which is incorrect as the process only had power at scales -1,-2,-3 and -4 (corresponding to $X^1$, $X^2$, $X^3$ and $X^4$ respectively). Obviously, we can think about more complicated examples of this kind where, e.g., the regions of piecewise stationarity overlap. When we did this we actually again saw a correct picture of how the energy is sharply distributed over different scales at different times.

Clearly classical stationary time series methods are not appropriate for non-stationary processes. This raises the question: is our theory here appropriate for all non-stationary processes? The answer is no, there are many ways in which a time series can fail to be stationary. For some forms of non-stationarity time-frequency methods will be appropriate: for example, the time-varying Fourier spectra as proposed by Dahlhaus [7]. Loosely speaking the time-frequency methods will be of most use where the processes are highly concentrated in frequency (e.g., modulated pure sinusoids) whereas ours will be beneficial where the processes are concentrated in scale (e.g., certain types of non-stationary moving averages of finite order). Our time-scale approach and other approaches for non-stationary series should be seen as complementary as unless the types of processes are known a priori it could be that any, some or none of the approaches will be of any use in practice.

Finally, for rigorous modelling and estimation the model in (5), although simple and
Figure 4: Mean of 100 uncorrected wavelet periodogram of $S_j(t)$ for $j = -1, \ldots, -9$. (Each estimate consisted of the square of SWT coefficients of a simulation of the concatenated MA process described in the text).

appealing, is actually too simple. The reason is that it does not permit consistent estimation of its defining coefficients: as more and more data $\{X_t\}_{t=1,\ldots,T}$ are observed with growing $T$ more and more coefficients $\{\eta_{jk}\}_{k=1,\ldots,T}$ (per scale $J$) enter the model. We need to ensure that at least part of the data is used to collect more statistical information on each coefficient, to reduce uncertainty whilst trying to estimate them from the data. The mechanism we use and which governs our rigorous definitions of Section 4 is rescaled time as introduced by Dahlhaus [8] for locally stationary processes.

2.5 Estimation: Wavelet periodograms and smoothing

Now we turn to the problem of how to estimate a wavelet spectrum (with respect to a given non-decimated wavelet basis) from a single stretch of observations from the process. For the purposes of this informal discussion we shall consider stationary processes but return in detail to the non-stationary case in in Section 5.

As in Nason and Silverman [23] our spectrum estimator is based on a wavelet periodogram: the squared empirical coefficients from the stationary wavelet transform of the data of length $T$ on scales $j = -1, \ldots, -J$:

$$ I_k^j := \left| \sum_{t=0}^{T-1} X_t \psi_{jk}(t) \right|^2, \quad j = -1, \ldots, -J, \quad k = 0, \ldots, T - 1. $$

Here, the largest (i.e., coarsest) possible scale is determined by $J = J(T) = \log_2(T)$. This formula is analogous to the one for the classical periodogram (see Chatfield [4] Chapter 7, for example). As for classical periodograms the wavelet periodograms have asymptotically non-vanishing variance (see Propositions 5.3 and 5.4) and need to be smoothed to obtain consistency. Various local smoothing methods could be used but we choose to use non-linear wavelet shrinkage (e.g. Donoho et al. [13]) with respect to another wavelet system (DWT or SWT), see Section 6.1,
equations 38 and 39. Indeed work by Gao [15] and von Sachs and Schneider [33], for Gaussian time series data, provides the necessary theory for denoising \( \chi^2 \)-distributed coefficients necessary for smoothing the wavelet periodogram. We call the resulting smoothed squared SWT coefficients \( \tilde{S}_j \): again a sequence of functions in rescaled time. Theorem 6.2 shows that \( \tilde{S}_j \) converges in mean-square to the corresponding wavelet spectrum \( S_j \).

However, as shown by Figure 4 the wavelet periodograms considered for fixed scale \( j \) contain information from other scales \( j' \neq j \) (in Figure 4 power from scale 5 has leaked into scale 4). To correct for this we can use the inverse of the finite dimensional matrix \( A_J := (A_{j\ell})_{j,\ell=-1,...,-J} \) built from the matrix \( A \) defined in Section 2.2. Writing \( \hat{I}_k^j \) as a vector: \( I_k := (\hat{I}_k^j)_{j=-1,...,-J} \) we consider the corrected periodogram

\[
L_k = A_J^{-1} I_k
\]

for \( k = 0, \ldots, T - 1 \). This is an asymptotically unbiased estimator of the wavelet spectrum \( S := (S_j)_{j=-1,...,-J} \), as, again for stationary processes independent of \( k \), we have

\[
E L = A_J^{-1} E I \to S, \quad \text{as} \quad T \to \infty,
\]

because \( E I \to A S \) (see Proposition 5.4).

For good estimates we have to:

- smooth the wavelet periodogram (for consistency);
- apply the inverse of \( A_J \) (to correct for the redundancy created by using the stationary wavelet transform).

In principle, these two steps could be performed in either order. However, we know a great deal about how to smooth the \( \chi^2 \)-distributed wavelet periodogram and so we smooth first and then correct by multiplying by the inverse of \( A_J \). The resulting smoothed estimator is denoted \( \tilde{S}_j \).

Finally, to obtain a mean-square consistent estimator of the (local) autocovariance \( c \) we replace the unknown wavelet spectrum \( S_j \) in equation (6) by \( \tilde{S}_j \) (see Proposition 6.3).

3 Application to baby heart beat example

This example continues from the discussion given in section 1.2. Figure 5 shows the raw wavelet periodogram for the time series shown in Figure 1. The series has \( T = 8192 \) points and so \( J = 13 \). Figure 6 shows an estimate of the evolutionary wavelet spectrum for the baby heart beat time series. This was obtained by smoothing the raw wavelet periodogram using non-linear wavelet thresholding and then correcting using \( A^{-1} \) as described in Section 2.5.

The Daubechies' [9] least-asymmetric wavelets \( N = 10 \) were used to form the SWT. The logarithms of the squared coefficients \( I_{8192}^j \) were then directly subjected to non-linear wavelet shrinkage using SUREshrink soft thresholding on scales 5 to 12 with the noise level estimated using the standard sample variance. Daubechies' least-asymmetric \( N = 10 \) wavelets were also used for the non-linear wavelet shrinkage.

Scales \( j = -1 \) and \( j = -2 \) in Figure 6 both contain a discontinuity (a small spike) in the smoothed estimate just before 0200 hours. With this discontinuity removed they look like scale \( j = -3 \). Scale \( j = -3 \) is shown in more detail in Figure 7. The dotted line in Figure 7 refers to the baby's sleep state as judged by a trained human observer (from measurements of brain waves, EEG, and eye movements, EOG). The observer classifies the baby's state of sleep as quiet sleep(1), between quiet and active(2), active sleep(3) and awake(4). It is clear that there is some, if somewhat imperfect, relationship between the estimate, \( \tilde{S}_{-3}(z) \), and the sleep state. In particular, periods of activity and wakefulness occur whilst the estimate of \( S_{-3}(z) \) is large and periods of quiet sleep when it is small. It is especially interesting that the spectral estimate distinguishes between quiet and active sleep as this is known to be a particularly difficult problem (for further discussion of the modelling aspects for time series of these type see Nason et al. [21]).
Figure 5: The raw wavelet periodogram $H_{k,81b2}^j$ for baby heart rate data. Each level $j$ in the periodogram has been scaled independently of the others so that the detail at all levels can be seen (some of the larger scale levels are almost 10 times as large).

Figure 6: Estimate of evolutionary wavelet spectrum $\tilde{S}_j(z)$ for baby heart rate data. Like Figure 5 each level has been scaled independently.
Figure 7: Solid line is estimate of evolutionary wavelet spectrum at scale \(-3\): \(\tilde{S}_{-3}(z)\) for baby heart rate data. The dotted line indicates the sleep state of the baby as determined by expert analysis of EEG and EOG (independent of heart rate — ECG). The value of the dotted line is indicated by the right-hand axis: 1 = baby quiet sleep, 2 = between 1 and 3, 3 = baby in active sleep, 4 = baby awake. See text for further description.

In effect, the \(\tilde{S}_{-3}\) quantity is measuring the local stochastic variability of the signal at that scale and it correlates fairly well with the sleep state variable.

4 Theoretical foundations: The particular wavelet process model

This section defines the building blocks of our processes: the discrete stationary (non-decimated) wavelets. Then we define our special locally stationary wavelet (LSW) process model. The wavelet spectrum of an LSW model may be identified and estimated from a single stretch of observations.

4.1 Discrete stationary (non-decimated) wavelets

For the remainder of the paper we use discrete stationary (non-decimated) wavelets as our basis functions which are discretized versions of the continuous scaled and shifted mother wavelets defined as follows.

Let \(\{h_k\}\) and \(\{g_k\}\) be the usual finite impulse response quadrature mirror filters associated with a real compactly supported wavelet (see Mallat [18] or Nason and Silverman [22]) for a description. Adopting the notation of Nason and Silverman [23] let \(H\), \(G\) and \(D_0\) be the usual low-pass, high-pass and dyadic decimation operators i.e. for \((u_n)_{n \in \mathbb{Z}}\)

\[
(H \, u)_k = \sum_{n \in \mathbb{Z}} h_{n-k}u_n,
\]

similarly for \(G\) but using coefficients \(\{g_k\}\) and

\[
(D_0 \, u)_k = u_{2k}.
\]
Then the level \( j \) discrete wavelet coefficients of the data \( \{ u_n \} \) are given by
\[
d_j = \mathcal{D}_0 \mathcal{G} (\mathcal{D}_0 \mathcal{H})^{j-1} u := W_j u,
\]
treating both the data \( u = (u_n) \) and coefficients \( d_j \) as vectors. The operator \( W_j \) is a matrix with each row corresponding to a discretized wavelet at scale \( j \). For the DWT successive rows would be identical except for a dyadic shift \( 2^{-j} \). For compactly supported wavelets each row has only a finite number of non-zero coefficients. Define \( \psi_j \) to be the set of non-zero coefficients present in each row.

For example, for Haar wavelets the only non-zero coefficients in each row in \( W_{-1} \) are \( \psi_{-1} = \left( \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right) \) and in each row in \( W_{-2} \) are \( \psi_{-2} = \left( \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2} \right) \) and so on.

The key point for stationary discrete wavelets is that the wavelets can be shifted to any location and not just by shifts by \( 2^{-j} \). Indeed, here we use this hence no more orthogonal but overcomplete collection of shifted vectors. So, we define the quantity \( \psi_{jk}(t) \) to be the \( k \)th element in the vector \( \psi_j \) and \( \psi_{jk}(\tau) \) is the \( k \)th element of the vector \( \psi_{j(k-\tau)} \) shifted by integers \( \tau \).

### 4.2 Locally stationary wavelet processes

**Definition 4.1** Define a class of locally stationary wavelet (LSW) processes, a sequence of doubly-indexed stochastic processes \( \{ X_t, T \}_{t=0,\ldots,T-1}, T = 2^J \geq 1 \). The LSW processes have the following representation in the mean-square sense:

\[
X_{t,T} = \sum_{j=-J}^{1} \sum_{k} \sum_{j,k:T}^0 \psi_{jk}(t) \xi_{jk}, \tag{11}
\]

where \( \xi_{jk} \) is a random orthonormal sequence of increments and where \( \{ \psi_{jk}(t) \}_{jk} \) is a discrete stationary family of wavelets for \( j = -1, -2, \ldots, -J(T), k = 0, \ldots, T - 1 \) based on a mother wavelet \( \psi(t) \) of compact support (which can generate an orthonormal basis of \( L_2(\mathbb{R}) \)).

The quantities in representation (11) possess the following properties:

1. \( E \xi_{jk} = 0 \) for all \( j, k \). Hence \( E X_{t,T} = 0 \) for all \( t \) and \( T \).
2. \( \text{cov}(\xi_{jk}, \xi_{lm}) = \delta_{j,k} \delta_{k,m} \).
3. There exists for each \( j \leq -1 \) a Lipschitz-continuous function \( W_j(z) \) for \( z \in (0,1) \) which fulfills the following properties:
   - \[
   \sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1), \tag{12}
   \]
   - The Lipschitz constants \( L_j \) are uniformly bounded in \( j \) and
     \[
     \sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty. \tag{13}
     \]
   - There exists a sequence of constants \( C_j \) such that for each \( T \)
     \[
     \sup_k \left| w_{j,k:T}^0 - W_j \left( \frac{k}{T} \right) \right| \leq C_j / T \tag{14}
     \]
     where for each \( j = -1, \ldots, -J(T) = -\log_2(T) \) the sup is over \( k = 0, \ldots, T - 1 \), and where \( \{ C_j \} \) fulfills
     \[
     \sum_{j=-\infty}^{-1} C_j < \infty. \tag{15}
     \]
Remark 4.2 The second assumption of uncorrelated increments is certainly somewhat restrictive. In practice, a broader class of processes which allows for correlation within or even between scales could probably be more useful. However, because we use a shift-equivariant SWT representation we in fact include quite a large class of processes with correlation. In particular our LSW processes includes all stationary processes with \( \sum \tau \mid c(\tau) \mid < \infty \): a large class of processes with short-range dependence (see remark 4.17).

Assumption 3 is crucial. It permits us to asymptotically identify the model coefficients and leads to a in this sense unique representation (11) given the fixed wavelet basis. That is, the sequence of \( \{ W_j \} \) is actually uniquely determined in contrast to the \( \{ w_{j,k,T} \} \) which, obviously, cannot be. The smoothness assumption on \( W_j(z) \) as a function of renormalized time \( z = k/T \in (0,1) \), controls the variation of each coefficient \( w_{j,k} \) as a function of \( k \) so that it can change only slowly enough. The smoothness ensures a neighbourhood of locations that determine the variation and this neighbourhood becomes asymptotically arbitrarily small in the rescaled time \( z \). The smoothness of \( W_j(z) \) and the rescaling ratio \( T^{-1} \) control the scale ("speed") of decay of the variation. The rate, \( C_j/T \), controls how much the variation of \( w_{j,k,T} \) is allowed to differ from the smooth locally stationary "amplitude" function \( W_j(z) \).

Assumption 3 guarantees both the convergence of \( \sum w_{j,k,T}^2 \) in the limit as \( T \to \infty \), and the finiteness and convergence of the variance of the process itself. (This last condition is verified in Proposition 4.9 below which shows that, in particular, \( \text{var} \{ X_{[z,T],T} \} \to c(z,0) < \infty \) for all \( z \in (0,1) \).)

Remark 4.3 Representation (11) does not include a scaling function coefficient as, asymptotically, it is not required. As in traditional multiresolution analysis with wavelets, as the scale tends to \( -\infty \), the coarse scale approximation will finally be included in the overall sum of the details. Strictly speaking, with this model, if we were to compare with classical spectral analysis, for finite \( T \) we might miss the spectral contribution of the "subseries" \( X_{i,T} \) at the very zero frequency (because all wavelets integrate to zero). However, asymptotically, this constraint disappears.

Remark 4.4 As with Dahlhaus [8] we are not observing a fixed continuous time process on an increasingly finer mesh as \( T \to \infty \). We rather use renormalization in time (location) \( k \) to introduce rescaled time \( z = k/T \in (0,1) \) such that the model in Definition 4.1 allows for (theoretically) obtaining more and more data about the local structure of \( W_j(z) \) as \( T \to \infty \). It describes how the local structure becomes more and more "stationary" with respect to the given wavelet basis.

4.3 The evolutionary wavelet spectrum

We now define the evolutionary wavelet spectrum (EWS) of an LSW process: the localized density of a wavelet spectral measure in the (rescaled-) time-scale plane.

Definition 4.5 Let

\[
S_j(z) := |W_j(z)|^2, \ z \in (0,1).
\]  \hfill (16)

We call \( \{ S_j(z) \}_{j=1,...,J(T)} \) the "evolutionary wavelet spectrum" (EWS) of the sequence \( \{ X_{i,T} \}_{i=0,...,T-1} \) with respect to \( \{ \psi_j \} \).

Using assumption 3 of Definition 4.1,

\[
S_j(z) = \lim_{T \to \infty} |w_{j,[z,T],T}|^2, \ \forall z \in (0,1),
\]  \hfill (17)

and fulfills,

\[
\sum_{j=-\infty}^{-1} S_j(z) < \infty \text{ uniformly in } z \in (0,1).
\]  \hfill (18)
The EWS measures the local power of the variance/covariance decomposition of the process $\{X_{t,T}\}$ at scale $j$ and an (arbitrary) location $z \in (0,1)$. Hence equation (11) delivers a time-scale decomposition which parallels a time-frequency decomposition (Dahlhaus’ localized Cramér spectral model described in Remark 4.18 below). Indeed, our model is similar to Dahlhaus’ model in two other ways:

1. the evolutionary wavelet spectrum $S_j(z)$ is defined only for $z \in (0,1)$, as boundaries do not make sense in this model.

2. the EWS is a uniquely defined quantity. The uniqueness is important and is discussed in some detail later in Corollary 4.15.

**Remark 4.6** The EWS, $S_j(z)$, inherits Lipschitz continuity from $W_j(z)$, with a Lipschitz constant which fulfills (13). Together with condition (12), which implies the uniform summability of $S_j(z)$, and with (15), this set of conditions guarantees the uniform decay of the wavelet model coefficients $w_{j,k}$ for coarser and coarser scales.

### 4.4 Local autocovariances and autocorrelation wavelets

Recall that representation (11) represents the process $X_{t,T}$ in terms of discrete stationary wavelets. As a consequence the autocovariance function,

$$c_T(z, \tau) = \text{cov}\{X_{[zT],T}, X_{[zT]+\tau, T}\}$$

for $z \in (0,1)$ and $\tau \in \mathbb{Z}$, also has a “wavelet–type representation”. This is analogous to the classical stationary case where the autocovariance is the inverse Fourier transform of the spectrum. Proposition 4.9 shows that $c_T$ tends to a local autocovariance, $c$ defined in Definition 4.8, which itself can be represented by a series of autocorrelation wavelets with coefficients given by $S_j(z)$.

**Definition 4.7 (“Autocorrelation wavelets”)**

$$\Psi_j(\tau) := \sum_k \psi_{jk}(0) \psi_{jk}(\tau), \; j < 0, \tau \in \mathbb{Z}.$$

The autocorrelation wavelets are related to the autocorrelation shell of Saito and Beylkin [28]. However, Saito and Beylkin use scaled and shifted versions of the autocorrelation function of a wavelet and use these to represent functions. Ours however averages over all locations within one scale $j$ and provides a family of symmetric, compactly supported and positive semi-definite functions on $\tau \in \mathbb{Z}$ which are well-suited for the construction of autocovariance functions of quasi–stationary processes (for example, with the motivating examples in Section 2.3).

**Definition 4.8** Define the local autocovariance (LACV) $c(z, \tau)$ of a given LSW process with EWS $\{S_j(z)\}$ by

$$c(z, \tau) = \sum_{j=-\infty}^{\infty} \; S_j(z) \; \Psi_j(\tau), \; \tau \in \mathbb{Z}, \; z \in (0,1) \; . \tag{19}$$

Now we show that the process autocovariance, $c_T$, asymptotically tends to the function, $c(z, \tau)$, which has an autocorrelation wavelet representation.

**Proposition 4.9** As $T \to \infty$, uniformly in $\tau \in \mathbb{Z}$ and $z \in (0,1)$,

$$c_T(z, \tau) = \text{cov}\{X_{[zT],T}, X_{[zT]+\tau, T}\} \to c(z, \tau) \; .$$

Moreover, using conditions (15) and (13),

$$|c_T(z, \tau) - c(z, \tau)| = O(T^{-1}) \; ,$$

uniformly in both arguments.
Corollary 4.10 Replacing condition (12) by the slightly stronger condition
\[
\sup_z S_j(z) \leq C 2^j , \quad j = -1, -2, \ldots ,
\] (20)
allows for relaxing conditions (15) and (13) to
\[
\sup_j C_j < \infty .
\] (21)
and
\[
\sup_j L_j < \infty ,
\] (22)
to show that
\[
|c_T(z, \tau) - c(z, \tau)| = O(T^{-1/2})
\] (23)
uniformly in both arguments.

However, we need the stronger conditions (15) and (13) in order to show the uniqueness of the wavelet spectrum in Corollary 4.15 and for the assertion of Proposition 4.14 to hold for all \( j < 0 \).

For stationary processes, representation (19) turns into the simpler form given in (6) and the EWS \( S_j \) becomes constant in \( z \) (see equation (33) for equivalent behaviour for the periodogram). For stationary processes our representation turns from a local into a global one. It is for this reason that we use a shift–equivariant basis.

Finally, it is interesting to observe how the (local) decay properties of \( c(z, \tau) \), for fixed \( z \), are related to the decay of the EWS \( S_j(z) \) for \( j \rightarrow -\infty \):

Corollary 4.11 Let \( \sum_j 2^{-j} S_j(z) < \infty , \) uniformly in \( z \). Then, \( \sum_{\tau} |c(z, \tau)| < \infty , \) uniformly in \( z \) and
\[
\sum_{\tau} |c_T(z, \tau) - c(z, \tau)| = o(1),
\]
as \( T \rightarrow \infty \), uniformly in \( z \).

4.5 Inversion of the autocovariance representation

Now we derive a way of inverting representation (19) to find an expression for the EWS in terms of the LACV (analogous to the classical stationary theory which says that the spectrum is the Fourier transform of the autocovariance). However, the representing functions, \( \{\Psi_j(\tau)\} \), in (19) are not orthogonal and thus we introduce a dual set, \( \{\kappa_j(\tau)\} \), in the following definition which permits us to invert the representation.

Definition 4.12 Define the operator \( A = (A_{jt})_{j, t \in \mathbb{Z}} \) by
\[
A_{jt} := \langle \Psi_j, \Psi_t \rangle = \sum_{\tau} \Psi_j(\tau) \Psi_t(\tau) ,
\] (24)
and the \( J \)-dimensional matrix \( A_j = (A_{jt})_{j, t = -1, \ldots , -J} \).
Further let (the vector) \( \kappa(\tau) := \{\kappa_j(\tau)\}_{j = -1, \ldots , -J} \) be defined via \( \kappa(\tau) = A^{-1} \Psi(\tau) \).

The \( \kappa(\tau) \) is well–defined because the operator \( A \) which describes the degree of redundancy in \( \{\Psi_j(\tau)\} \), has a bounded inverse \( A^{-1} \) as shown by the following theorem.

Theorem 4.13 Let \( \lambda_{\min}(A) \) denote the smallest eigenvalue of \( A \). Then there exists a \( \delta > 0 \) such that
\[
\lambda_{\min}(A) \geq \delta , \quad \text{hence} \quad \|A^{-1}\| < \infty .
\]
i.e. \( A \) is positive–definite and has a bounded inverse.
For future reference define the bounded operator $B = D' \cdot A \cdot D$ where $D$ is a diagonal matrix $D = \text{diag}(2^{j/2})_{j \in \mathbb{N}_0}$. The matrix $B$ also has a bounded inverse $B^{-1}$ as shown in the proof of Theorem 4.13.

The fact that $A$ is invertible means that the claimed inverse of equation (19) exists and coincides with the EWS, which is consequently uniquely defined, as shown by the following Proposition and Corollary 4.15.

**Proposition 4.14** As an inverse formula of equation (19),

$$ S_j(z) = \sum_{\tau} c(z, \tau) \kappa_j(\tau) = \sum_{\ell} A^{-1}_{j \ell} \sum_{\tau} c(z, \tau) \Psi_{\ell}(\tau) \quad \forall j < 0. \quad (25) $$

The above representation of the EWS in terms of the LACF is unambiguous as is shown by part (b) of the following Corollary. The Corollary also shows the connection between the wavelet spectrum, $S_j(z) = \lim_{T \to \infty} |w_{j,k}^0|_{T}|^2$, as a "postulated" asymptotic limit (a smooth function in $z$ with some decay in $j$) and the process data $\{X_{t,T}\}$ which are generated by the sequence $\{w_{j,k}^0\}_{k,T}$ and the given wavelet family.

**Corollary 4.15** (a) The EWS $\{S_j(z)\}$ as defined in Definition 4.5 also arises as the asymptotic limit of

$$ \tilde{T}_{j,T}(z) := \sum_{\ell=-J}^{-1} A^{-1}_{j \ell} \sum_{\tau} c_T(z, \tau) \Psi_{\ell}(\tau), \quad j = -1, \ldots, -J, \quad J = J(T) = \log_2(T), \quad (26) $$

That is,

$$ \lim_{T \to \infty} \tilde{T}_{j,T}(z) = S_j(z) = \lim_{T \to \infty} |w_{j,k}^0|_{T}|^2, \quad j \leq -1, \quad z \in (0, 1), \quad (27) $$

with

$$ \lim_{T \to \infty} \sum_{j=-J(T)}^{-1} \tilde{T}_{j,T}(z) < \infty. $$

(b) The EWS is uniquely defined given the corresponding LSW process.

We have shown that asymptotically the spectral representation of our LSW process is unique in the following sense. Though clearly, for each finite $T$, the process representation (11) itself cannot be a unique one: the one of the local autocovariance (19), as an asymptotically defined quantity, actually is. If we started from a LSW process $\{X_{t,T}\}$ with model coefficients $w_{j,k}^0$ we know by Proposition 4.9 and Corollary 4.15(a) that the limit of $\tilde{T}_{j,T}(z)$ tends to the defined EWS $S_j(z) = \lim_{T \to \infty} |w_{j,k}^0|_{T}|^2$ through the localized autocovariance $c_T$. In other words the quantity defined in Definition 4.5 is indeed the (asymptotic limit of) an inverse "wavelet-type" transform of the autocovariance. This theory completely parallels the uniqueness of the evolutionary spectrum for the Dahlhaus class of locally stationary processes (see Dahlhaus [7] Proposition 2.1).

Note also that $(\Psi(\tau), \kappa(\tau))$ can be seen as a pair of dual "bi-orthogonal bases" which provide the forward and backward transform for the pair of dual quantities $(c(z, \tau), S_j(z))$ for fixed $z \in (0, 1)$. Here the operators $A$ and $A^{-1}$ play the role of a "kernel" for forward and backward transformation, respectively.

**Remark 4.16** A more intuitive interpretation of the pair of transforms between local autocovariances and their EWSs is the following. If we were to rotate $\{\Psi_j(\tau)\}$ by $A^{-1/2}$, then we actually would orthogonalityize this family and could use the same set of (over $j$) orthogonal vectors $\tilde{\kappa}_j(\tau) = A^{-1/2} \Psi_j(\tau)$ for the forward and backward transform. Hence, the EWS $\{S_j(z)\}$ then would be the coefficients in an orthogonal series representation of $c(z, \tau)$, and hence obviously unique. However, as we want to keep some of the interpretative properties of the set $\{\Psi_j(\tau)\}$, as described in Sections 2.3 and 2.4, we remain with the normalization of $A^{-1}$ (this is a matter of taste).
Remark 4.17 We observe that by equation (25), we can relate the (local) autocovariance of any given zero mean process, which allows for a decomposition like (11 to a sequence of possibly valid $S_j(z)$: If condition $\sum_j S_j(z) < \infty$ is satisfied than the given process falls into our class of LSW processes with an EWS $S_j(z)$ with respect to the wavelet basis $\{\psi_{jk}\}$ underlying $\{\kappa_j\}$. Hence, all stationary processes with absolutely summable autocovariance $\sum_\tau |c_{X}(\tau)| < \infty$ are LSW processes, because automatically $\sum_j \kappa_j(\tau) < \infty$, uniformly in $\tau$, and so $\sum_j S_j(z) < \infty$ is fulfilled.

Conversely, as a consequence of Corollary 4.11, any LSW process with time independent EWS $\{S_j(z)\}$ fulfilling the additional assumption $\sum_j 2^{-j} S_j(z) < \infty$ is a stationary process with absolutely summable autocovariance $c_X(\tau)$.

We finish this section with a discussion of the relation of LSW processes to the Dahlhaus class of locally stationary processes.

Remark 4.18 Assumption 3 in Definition 4.1 is motivated by Dahlhaus' definition [8] of "locally stationary processes" with the following time-frequency or generalised (time-varying) Cramér spectral representation:

$$X_{t,T} = \int_{-\pi}^{\pi} \mathcal{A}_{t,T}(\lambda) \exp i\lambda t \, d\xi(\lambda),$$

(28)

where:

(a) $\{d\xi(\lambda)\}_\lambda$ is an orthonormal increment process on $[-\pi, \pi]$;

(b) $\sup_{t,\lambda} |\mathcal{A}_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right)| = O(T^{-1})$,

where $A(u, \lambda)$, $u \in (0,1)$ is a smooth function, $\lambda \in [-\pi, \pi]$ and the uniquely defined "evolutionary spectrum" is $f(u, \lambda) = |A(u, \lambda)|^2$. For further details, see [8].

In continuation of earlier comments we observe the following relation between this class of locally stationary "Fourier" processes (which we like to call LSF, for short) and our class of LSW processes. In general, a LSF process is in the class of LSW if it fulfills $\sum_\tau |\tilde{c}(z, \tau)| < \infty$, uniformly in $z \in (0,1)$, where $\tilde{c}(z, \tau) := \int_{-\pi}^{\pi} \tilde{f}(z, \lambda)d\lambda$ (cf. Remark 4.17 above). Conversely, by Corollary 4.11, a LSW process with EWS fulfilling $\sum_j 2^{-j} S_j(z) < \infty$, uniformly in $z$, is a LSF process with sufficiently smooth evolutionary spectrum $\tilde{f}(z, \lambda)$ (in the slightly more general sense of [24], Theorem 3.1) fulfilling $|\sum_\tau \tilde{c}(z, \tau)| < \infty$. For all these relations, note that the regularity in local time $z$ in our LSW class was assumed to be comparitively high (Lipschitz). So, strictly speaking, the evolutionary spectrum of an LSF process had to be Lipschitz, too, to be in this LSW class. However, we actually believe that we can relax this regularity of the EWS in $z$ which we were not ambitious to do in order to keep the technical level of proving low. Note also that explicit formulae relating both types of evolutionary spectra can be found in the next section, in Corollary 5.6.

5 Estimation

We only consider the situation of a fixed underlying wavelet basis which is a priori given for the representation (11). We do not consider the problem of choice of best basis fitting to a given set of observations $X_1, \ldots, X_T$ nor the question of what happens if a different wavelet basis is chosen. Both are interesting questions for future work.

Recall that in the following definition of the empirical wavelet coefficients the symbol $\psi_{jk}(t)$ stands for an element of the shifted vector $\psi_{j(k-\ell)}$, as introduced in Section 4.1.

Definition 5.1 Let

$$d_{j,k:T} := \sum_{t=1}^{T} X_{t,T} \psi_{jk}(t)$$
be the empirical wavelet coefficients, where $X_{t,T}$ is a LSW process as given by (11).

In fact, in this definition the sum over $t$ is finite. For fixed $j$ the actual number of summands does not change with $T$; it is always proportional to $M_j = M \cdot 2^{-j}, j = -1, \ldots, -J,$ where $M$ denotes the length of the support of $\psi(t)$ on the finest scale 0.

Now we define our key statistic which plays the same role in wavelet spectral analysis as the classical periodogram does in traditional (Fourier) spectral analysis (see also Nason and Silverman [23] whose definition coincides with ours).

**Definition 5.2** Let

$$I_{k,T}^j := |d_{j,k,T}|^2$$

be the wavelet periodogram of $\{X_{t,T}\}$.

We now further assume that the $\xi_{jk}$ are normally distributed. Therefore $\{X_{t,T}\}$ is also Gaussian. The wavelet periodogram satisfies properties analogous to those satisfied by the classical Fourier periodogram as follows:

For reasons of clear presentation we will first discuss the asymptotics of a wavelet periodogram which would be based on the use of the DWT in our model of LSW processes, for which we use the notation $\psi_{jk}^D(t) = 2^{j/2} \psi(2^{j}t - k)$ for distinction. We cite from [31], Proposition 3.3:

**Proposition 5.3**

In an LSW model based on using an orthogonal wavelet basis $\psi_{jk}^D(t) = 2^{j/2} \psi(2^{j}t - k)$, we had:

(expectation)

$$E I_{2^j [z T], T}^j = S_j(z) + O(T^{-1}) \quad \forall z \in (0,1).$$

(variance)

$$\text{var} I_{2^j [z T], T}^j = 2 S_j^2(z) + O(T^{-1}) \quad \forall z \in (0,1).$$

(covariance)

$$\text{cov}(I_{k,T}^j; I_{m,T}^l) = 2 \delta_{jk} \delta_{km} \left\{ S_j \left( \frac{k}{2^j T} \right)^2 + O(T^{-1}) \right\} .$$

Now we give the analogous results for the wavelet periodogram based on the non–decimated wavelet representation, as in Definition 4.1 of our model. Note that now neither the wavelets are orthogonal nor the periodograms are uncorrelated any longer.

**Proposition 5.4**

In an LSW model based on using a non–decimated wavelet basis $\psi_{jk}(t) = 2^{j/2} \psi(2^{j}t - k)$, we have:

(expectation)

$$E I_{[z T], T}^j = \sum_t A_{jt} S_t(z) + O(2^{-j}/T) \quad \forall z \in (0,1).$$

Hence, for the vector of periodograms $\{I_{[z T], T}^j\}_{t=-1,\ldots,-J}$, for each fixed $j < 0$,

$$E \sum_t A_{jt}^{-1} I_{[z T], T}^j = S_j(z) + O(T^{-1}) \quad \forall z \in (0,1).$$
(variance) The asymptotic variance of $I_{1,T}^{2}$ can be written as a square of linear combinations of the leading term of (30), i.e., of appropriately weighted wavelet spectra $S_\ell(z)$.

(covariance) The correlation between two wavelet periodograms $I_{k,T}^{\ell}$ and $I_{m,T}^{\ell}$ decays with an increasing distance between the relative positions of the location $k$ on scale $j$ to the location $m$ on scale $\ell$. For example, within one scale $j = \ell$, it is zero as soon as $|k - m|$ exceeds the length of the overlap of the supports of the corresponding wavelets.

As the actual equations for variance and covariance cannot be written in compact form, we decided to omit them from the proposition above. In Section 6.1 we will only make use of the asymptotically leading term of the variance.

For completeness we also cite the following lemma from [31] which gives some useful connection between time-scale and time-frequency analysis (it is of no originality, and can be found in more generality in, for example, [14], equation (3-70)). Note that, in contrast to [31] and [24], we will introduce non-symmetric lags in the definition of the occurring Wigner expressions, to match our notational convention of the previous sections.

**Lemma 5.5**

$$I_{k,T}^{\ell} = \int \sum_{t=1}^{T} I_{t,T}^{\ell}(\omega) \ W_{jk}(t, \omega) \ d\omega,$$

where

$$I_{t,T}^{\ell}(\omega) = \sum_{s: t \leq t \leq t+s \leq T} X_{t,T} X_{t+s,T} \ exp(-i\omega s),$$

is the localized “pre-periodogram” (defined in [24]), which is a Wigner–Ville local spectral estimate, and

$$W_{jk}(t, \omega) = \sum_{\tau} \psi_{jk}(t) \ psi_{jk}(t + \tau) \ exp(-i\omega \tau)$$

is the Wigner–Ville distribution of the wavelet $\psi_{jk}$.

This lemma, which is simply based on Parseval’s equation, will be useful for the following corollary which collects some relations between the wavelet periodogram and the classical Fourier spectral density for a stationary process. Also it relates to the evolutionary Fourier spectrum for a locally stationary process (28) defined by Dahlhaus [8], cf. Remark 4.18.

**Corollary 5.6**

1. If $\{X_{t,T}\}$ is a classical second-order stationary process $\{X_{t}\}$ with spectral density $f(\omega)$, then

$$\lim_{T \to \infty} E I_{k,T}^{\ell} = \int_{-\pi}^{\pi} f(\omega) \ |\psi_{jk}(\omega)|^2 \ d\omega$$

which is independent of $k$ (as $|\psi_{jk}(\omega)|^2$ is) and where

$$\psi_{jk}(\omega) := \sum_{s=-}\infty^{\infty} \psi_{jk}(s) \ exp(i\omega s).$$

The quantity $\psi_{jk}$ is not independent of $k$ but is a phased-shifted version of $\psi_{j0}$. However, taking absolute values cancels the phase shift. For notational simplicity we define $\psi_{j} = \psi_{j0}$.

Combining equations (33) and (31), for stationary processes,

$$S_{j}(z) = \sum_{\ell} A_{j,\ell}^{1} \int f(\omega) \ |\psi_{\ell}(\omega)|^2 \ d\omega,$$
which is now independent of \( z \) as expected for stationary processes. Note that in order to get a \( S_j(z) \) which is the wavelet spectrum of a LSW-process it has to fulfill condition (12). This depends, of course, in general both on the smoothness of the chosen basis wavelet \( \psi \) and of the given spectrum \( f(\omega) \) of the stationary process. A sufficient condition, however, is the absolute summability of the autocovariances \( c(\tau) \) (cf. Remark 4.17) which is a common smoothness condition of the spectrum \( f(\omega) \).

The same applies to the following assertion:

2. If \( \{X_t,T\} \) is locally stationary in the sense of (28) with evolutionary Fourier spectrum \( f(u,\omega), \)

\[
\lim_{T \to \infty} E I_{[uT],T}^j = \int f(u,\omega) |\hat{\psi}_j(\omega)|^2 d\omega.
\]

In this case, the evolutionary wavelet spectrum \( S_j(z) \) is related to the evolutionary Fourier spectrum by the following equation:

\[
S_j(z) = \sum_{\ell} A_{j\ell}^{-1} \int f(z,\omega) |\hat{\psi}_\ell(\omega)|^2 d\omega,
\]

and is, again, a valid EWS of a LWS process if \( \sum_j S_j(z) < \infty \) for all \( z \in (0,1) \).

The proof is completely analogous to the proof of the corresponding Corollary 3.5 in [31], together with equation (31).

For completeness we cite the proof of Corollary 3.5 in [31] in the Appendix.

6 Smoothing and inverting the wavelet periodogram

6.1 Smoothing by non-linear wavelet thresholding

Propositions 5.3 and 5.4 demonstrate that the wavelet periodogram is not a consistent estimator of the wavelet spectrum and needs to be smoothed. Nason and Silverman [23] developed a heuristic method based on linear Fourier smoothing. However, alternative methods could be used, such as, for example, kernel smoothing or a moving average. Of course, the question of how to choose the smoothing parameter then arises.

Here we will discuss non-linear thresholding of \( H_{k,T}^j \) as a function of \( z = k/T \) with respect to either a traditional discrete transform (DWT) with an orthonormal wavelet basis \( \{\hat{\psi}_{tm}(z),\hat{\psi}_{tm}(z)\} \) of \( L_2([0,1]) \). Here \( \hat{\psi}_{tm}(z) = 2^{t/2} \psi(2^t z - m) \), where \( t_0 \) is the coarsest scale in the scheme and \( \ell \geq t_0, m = 0,\ldots,2^\ell - 1 \). An alternative is to use again a stationary wavelet transform (SWT), see equation (39) below, which we actually pursue in practice in order to benefit from translation–invariant denoising (see below).

Applying non-linear thresholding to wavelet periodograms almost completely parallels existing wavelet curve estimation techniques, as in [12], [24], [16], and [30], if we proceed as follows.

We first smooth the wavelet periodograms, then we apply the inverse of the matrix \( A \) to achieve an asymptotically unbiased estimator of the wavelet spectrum (cf. Proposition 5.4).

We considerably simplify the asymptotics of the secondary–stage wavelet smoothing because we know how to do wavelet thresholding of \( \chi^2 \)-distributed random variables. Compare what we observed in the previous section, in Proposition 5.4.

We describe in detail how to smooth by using orthogonal secondary–stage wavelets for smoothing, and our asymptotic results are formulated for those only. It is, however, no problem to extend this to the use of a SWT, which in practice turns out to be beneficial (cf. Coifman and Donoho). All the asymptotic results which will follow carry over to translation–invariant denoising of curves, which is not surprising as the SWT can be pictured as an average of T DWT's applied on T shifted versions of the original signal (sampled curve).
Empirical wavelet coefficients of a DWT are provided by projecting, for fixed \( j \) the wavelet periodogram \( I_{k,T}^j \) onto the wavelets \( \tilde{\psi}_{tm}(z) \). Note the similarity to smoothing squared empirical wavelet coefficients per scale \( j \) in the work of [30] on nonparametric regression with locally stationary errors, which was introduced to estimate the unknown scale of variation of the empirical wavelet coefficients, i.e. to find the regions of similar local variation of these.

Hence, using this secondary–stage smoothing the discrete orthogonal DWT we define the following. For levels \( j = -1, \ldots, -J \) and \( k = 0, \ldots, T - 1 \), let

\[
I_{k,T}^j = \sum_{\ell} \sum_m \tilde{\psi}_{tm} \left( \frac{k}{T} \right), \quad k = 0, \ldots, T - 1, \quad (37)
\]

where

\[
\tilde{\psi}_{tm} = T^{-1} \sum_{n=1}^{T} I_{n,T}^j \tilde{\psi}_{tm} \left( \frac{n}{T} \right), \quad \text{for} \quad \ell = \ell_0, \ldots, \log_2(T), \quad m = 0, \ldots, 2^\ell - 1. \quad (38)
\]

Note that \( \tilde{\psi}_{tm} = \tilde{\psi}_{tm,T}^j \) depends on \( j \) and \( T \), also. With slight abuse of notation, \( \tilde{\psi}_{t_0 - 1,m} := \tilde{\psi}_{t_0 m} \), in order to include the scaling coefficient on the coarsest scale \( \ell_0 \) of this second wavelet scheme. Note also that on each level \( j \), due to the use of the non–decimated wavelet basis in the definition of the wavelet periodogram, we really have access to \( T \) values of \( I_{k,T}^j \), as a function of \( k = 0, \ldots, T - 1 \).

Completely analogously, we build coefficients with respect to the SWT for smoothing wavelet periodograms. Let \( \tilde{\psi}_t(z) = 2^{\ell/2} \tilde{\psi}(2^\ell z) \) and let

\[
\tilde{v}_{tm} = T^{-1} \sum_{n=1}^{T} I_{n,T}^j \tilde{\psi}_t \left( \frac{n - m}{T} \right), \quad \text{for} \quad \ell = \ell_0, \ldots, \log_2(T), \quad m = 0, \ldots, T - 1. \quad (39)
\]

In order to do denoising we now apply non–linear thresholding to the wavelet coefficients \( \{\tilde{v}_{tm}\} \). The resulting reconstruction (denoised estimator) is obtained by inverting the transform using only those coefficients which remain after soft or hard thresholding ([13]).

\[
\hat{s}_j(z) = \sum_{\ell} \sum_m \tilde{v}_{tm} \tilde{\psi}_{tm}(z), \quad z \in (0, 1), \quad (40)
\]

where \( \tilde{v}_{tm} = \delta^S_H(\tilde{v}_{tm}) \) (i.e. soft or hard thresholding) and where the sum exceeds over the set of indices given in Theorem 6.1 below.

Note that, for fixed \( j \), \( \hat{s}_j(z) = \hat{I}_k^j \) for \( z = \frac{k}{T}, \quad k = 0, \ldots, T - 1 \), that is a set of denoised wavelet periodograms on the original grid of locations.

The threshold \( \lambda = \lambda(j, \ell, m, T) \) can be appropriately determined from the following theorem, which we formulate for using DWT–coefficients given by (38), only.

**Theorem 6.1** For LSW processes as given in Definition 4.1, under the assumption of Gaussianity and with a wavelet \( \psi \) of bounded variation, the wavelet coefficients \( \tilde{v}_{tm} \) in (38), with \( 2^\ell = o(T) \), obey uniformly in \( m \),

\[
E \tilde{v}_{tm} - \int_0^1 \sum_j A_{jj'} S_j(z) \tilde{\psi}_{tm}(z) \, dz = O \left( 2^{\ell/2}/T \right), \quad (41)
\]

and,

\[
\text{var} (\tilde{v}_{tm}) = 2^{-1} \int_0^1 \left( \sum_n A_{jn} S_n(z) \right)^2 \tilde{\psi}_{tm}(z) \, dz + O \left( 2^\ell T^{-2} \right). \quad (42)
\]
Of course, in this theorem, the rates depend on the smoothness assumptions on $S_j(z)$, and also on the smoothness of the secondary-stage wavelet $\tilde{\psi}$. And note also, that as always in wavelet curve estimation, one has to keep away from the finest scale $l$, i.e. $2^l = o(T)$.

Assuming Gaussianity of the process, we can proceed as in [15], [33] to show that with a "universal" threshold $\lambda = \lambda(l, m; j; T)$ fulfilling

$$\lambda^2(l, m; j; T) = 2 T^{-1} \log^2(T) \var(\epsilon_{lm}), \quad (43)$$

the adaptive estimate $\tilde{S}_j(z)$ of $S_j(z)$ attains the asymptotically near-optimal minimax rate of the $L_2$-risk between estimator and evolutionary wavelet spectrum. Note that our considered function class is comparatively small here, as we deal with Lipschitz functions. However, we believe that we can generalize to less regular wavelet spectra by now traditional means of proving (cf. [24]).

**Theorem 6.2** Under the assumptions of Theorem 6.1, with a threshold as given in equation (43), for each fixed $j$,

$$\int_0^1 E\left(\tilde{S}_j(z) - S_j(z)\right)^2 dz = O\left(\log(T)/T^{4}\right). \quad (44)$$

This theorem is based on existing results on quadratic forms of Gaussian variables, which are $\chi^2$-distributed. For non-Gaussianity, techniques as in [24] should be applied, replacing $\log(T)$ by $\sqrt{\log(T)}$ in the universal threshold.

For completeness, we add the following remark, related to the work of [31]. This non-linear smoothing of wavelet periodograms can even be rigorously performed if in the first place only a DWT were used to build the wavelet periodograms (cf. Proposition 5.3). The only modification is that asymptotically one has to keep away from the coarsest scale $-J = \log_2(T)$, i.e. $2^{-J} = o(T)$, as on this scale there is no asymptotically increasing number of "data" $I_{k,T}^j$. However, there is a growing number of coefficients $I_{k,T}^j$ if we are on a fixed scale $j$, fixed relatively to the finest scale. Then the number of coefficients grows with order $O(T)$ on this finest scale as well as on all the fixed scales below, which are bounded away from the coarsest scale. Hence, $S_j$ can be estimated with increasing accuracy as $T$ increases.

We like to mention that in practice some modification might be appropriate as for this approach the thresholds depend on the unknown wavelet spectrum. This is exactly as it is for stationary Fourier spectrum estimation with wavelet thresholding of periodograms and also for the time-dependent version of this, see [33]. In both situations we experienced quite good results with thresholding of the logarithm of the periodogram, as this transformation stabilizes the variance of the empirical wavelet coefficients. Hence application of a universal threshold (see, e.g., [12]) might be appropriate.

### 6.2 Local variance and autocovariance estimation

There are certainly two motivations why we might be interested in estimation of local variances and autocovariances of the underlying process, in this specific situation. First, if we sum in the table of (smoothed) wavelet periodograms $I_k^j$ over scales $j$, then we observe that this sum can be considered as a local variance estimate as it measures the power of the signal content over local time $k$. Also, apart from being an intuitive descriptive tool, this quantity quite generally can be used to derive a statistical test on stationarity of the underlying time series. Here, the null hypothesis of stationarity could be given in terms of the respective theoretical quantity $\sum_j S_j(z)$ being a constant over $z$. Hence we could run a test on whether the empirical sum is a constant over time $k$, which we will keep for further research. Compare also the similarity to the approach of [32] on an analogous test for locally stationary "Fourier" processes, based on traditional (segmented) periodograms.

The second motivation is, of course, the more rigorous approach of how to estimate the (in Section 4.4 defined) local autocovariance $c(z, r)$ by inversion of (smoothed) wavelet periodograms.
Proposition 6.3 Let \( \tilde{S}_j(z) \) denote the result of applying the inverse matrix \( A^{-1} \) to the smoothed wavelet periodogram \( \tilde{S}_j(z) \) of the previous section. Then, replacing in equation (19), i.e. in

\[
c(z, \tau) = \sum_j S_j(z) \Psi_j(\tau)
\]

the wavelet spectrum \( S_j(z) \) by \( \tilde{S}_j(z) \) allows us to consistently estimate the local autocovariance \( c(z, \tau) \), in the following \( L_2([0,1]) \) sense in \( z \).

Let \( 2^{j_0} = o(T) \). Then

\[
\tilde{c}(z, \tau) = \sum_{j=-j_0}^{-1} \tilde{S}_j(z) \Psi_j(\tau)
\]

is a consistent estimator of \( c(z, \tau) \) as \( T \to \infty \), in that, for each fixed \( \tau \in \mathcal{Z} \),

\[
E \int_0^1 \left( \tilde{c}(z, \tau) - c(z, \tau) \right)^2 \, dz = o(1)
\]

In particular, for \( \tau = 0 \) we observe that \( \tilde{c}(z, 0) \) estimates the local variance \( \sum_j S_j(z) \Psi_j(0) = \sum_j S_j(z) \) (as \( \Psi_j(0) = 1 \) for all \( j \)).

7 Appendix: Proofs

Proof of Proposition 4.9

Using the representation of the process \( X_t,T \) given in (11) the covariance is given by

\[
c_T(z, \tau) = \text{cov}\{X_{t,T}; X_{t,T+\tau}z \}
\]

\[
= \sum_{jk} |w^0_{jk,T}|^2 \psi_jk([zT]) \psi_jk([zT]+\tau)
\]

\[
= \sum_{jk} |w^0_{jk,[zT]+\tau,T}|^2 \psi_jk([zT]+\tau) \psi_jk([zT]+\tau)
\]

\[
= \sum_{jk} |w^0_{jk,[zT]+k,T}|^2 \psi_jk(0) \psi_jk(\tau)
\]

With condition (15),

\[
\left| |w^0_{jk,[zT]+k,T}|^2 - S_j(z + k/T) \right| = O(C_j/T),
\]

and since \( S_j \) is Lipschitz

\[
|S_j(z + k/T) - S_j(z)| = O(L_j |k|/T).
\]

Hence,

\[
\left| \sum_{jk} |w^0_{jk,[zT]+k,T}|^2 \psi_jk(0) \psi_jk(\tau) - c(z, \tau) \right|
\]

\[
= \left| \sum_{jk} |w^0_{jk,[zT]+k,T}|^2 \psi_jk(0) \psi_jk(\tau) - \sum_j S_j(z) \Psi_j(\tau) \right|
\]

\[
\leq T^{-1} \sum_{jk} (L_j |k| + C_j) |\psi_jk(0)| \psi_jk(\tau)|
\]

\[
= O(T^{-1}),
\]
due to conditions (13) and (15). Note also that $\sum_j S_j(z) \Psi_j(\tau) < \infty$, by (12) and as $\Psi_j(\tau) = O(1)$ uniformly in $\tau$.

**Proof of Corollary 4.10**

We define the following approximation

$$c_{J_0}(z,\tau) = \sum_{j=-J_0}^{\infty} S_j(z) \Psi_j(\tau).$$

The assertion of the corollary follows immediately by the same arguments as in the proof of Proposition 4.9 by observing that

$$c_T(z,\tau) - c_{J_0}(z,\tau) = O\left(\frac{2^{J_0}}{T}\right),$$

and

$$c_{J_0}(z,\tau) - c(z,\tau) = O\left(2^{-J_0}\right),$$

which is sufficient by balance of both rates for $J_0 \to \infty$.

**Proof of Corollary 4.11**

Essentially the proof works like the proof of Corollary 4.10 (however, under the stronger conditions). As a continuation of the proof of Proposition 4.9 we observe that both

$$\sum_\tau |c(z,\tau)| \leq \sum_\tau \left| \sum_j S_j(z) \Psi_j(\tau) \right| \leq C \sum_j 2^{-j} S_j(z) < \infty,$$

and that

$$\sum_\tau |c_{J_0}(z,\tau) - c_T(z,\tau)| = O\left(\frac{2^{J_0}}{T}\right)$$

and

$$\sum_\tau |c_{J_0}(z,\tau) - c(z,\tau)| \leq \sum_{j<-J_0} 2^{-j} S_j(z) = o(1).$$

**Proof of Theorem 4.13**

We will show that there exists some $\delta > 0$ such that $\lambda_{\min}(A) \geq \delta$ by showing that $\lambda_{\min}(B) \geq \delta$, where, again, $B = D' \cdot A \cdot D$ with diagonal matrix $D = \text{diag}(2^{\frac{j}{2}})_{j<\infty}$, i.e., $B_{ij} = 2^{\frac{j}{2}} A_{ij} 2^{\frac{i}{2}}$. This is sufficient by property (i) of the following three known properties of matrix algebra and the theory of Toeplitz matrices which we shall use in our proof:

(i) If $A$ is a Hermitian (symmetric) matrix with $A = D' \cdot C \cdot D$, $D$ diagonal, then $\lambda_{\min}(A) \geq \lambda_{\min}(D) \lambda_{\min}(C) \lambda_{\min}(D)$.

(ii) (Weyl) If $A = B + C$, where $B$ and $C$ are Hermitian (symmetric), then $\lambda_{\min}(A) \geq \lambda_{\min}(B) + \lambda_{\min}(C)$.

(iii) (Theorem 3.1 (i) of Reichel and Trefethen [26]) Let $T$ be Toeplitz (and Hermitian) and define $t_i$ to be the elements on diagonal $i$ for $i \geq 0$ (so the diagonal elements are $t_0$, the first off-diagonal elements are $t_1$ and so on). Introduce $f(z)$ as the symbol of the operator associated with $T$, i.e.

$$f(z) = \sum_{n=-\infty}^{\infty} t_n z^n, \quad z \in \mathbb{C}.$$
\( f(z) \) is an analytic function in the open unit disk \( D \) in the complex plane and continuous in the closed unit disk \( \Delta = D \cup S \), where \( S \) denotes the unit circle, if \( \sum_{n} |t_{n}| < \infty \). In this case, the spectrum \( \Lambda \) of the (Laurent) operator \( T \) is fully understood, as \( \Lambda(T) = f(S) \). Hence, with the additional property of a symmetric \( T \), an estimate of the smallest eigenvalue of \( T \) is achieved by

\[
\min_{|z|=1} f(z) = \min_{|z|=1} t_{0} + 2 \Re \left( \sum_{n=1}^{\infty} t_{n} z^{n} \right). \tag{47}
\]

**Proof of Theorem 4.13 for Haar wavelets**

We first prove the theorem for Haar wavelets. As Haar wavelets have the shortest possible compact support in the time domain, their Fourier transform decays only slowly. Wavelets with longer support in the time domain show a faster decay in the Fourier domain. Consequently, for compactly supported wavelets other than Haar the decay off the diagonal of the matrix \( A \) is faster, as can be observed using Parseval's relation in the definition of the matrix \( A \) in (24),

\[
A_{j\ell} := \sum_{\tau} \Psi_{j}(\tau) \Psi_{\ell}(\tau) = \int d\omega \ \widehat{\Psi}_{j}(\omega) \ \widehat{\Psi}_{\ell}(\omega),
\]

where the definition of the Fourier transform \( \widehat{\Psi}_{j}(\omega) \) of \( \Psi_{j}(\tau) \) is explicitly given below in (55). This phenomenon could actually be observed in numerically calculating the entries of \( A \) for some members of the Daubechies' compactly supported wavelets of high enough order \( N \). Eventually, with growing \( N \), the (observed) matrix \( A \) became diagonally dominant. So, in some sense, the investigation of \( A \) for Haar wavelets covers the most "extreme" case and will be theoretically dealt with in the sequel with rigor. Nevertheless, we shall indicate later how the proof may be generalised to other members of the Daubechies compactly supported wavelets with order \( N > 1 \). For convenience our indices will now run from 1 to \( \infty \) rather than from \(-1\) to \(-\infty\). For the Haar wavelets we can explicitly compute the formulae for entries of the matrix \( A \): the inner product matrix of the autocorrelation wavelets in the time domain.

**Lemma** In the Haar case the elements of the autocorrelation wavelet inner product matrix \( A \) have the following form:

\[
A_{jj} = \frac{2^{2j} + 5}{3 \cdot 2^{j}} \quad , \quad A_{j\ell} = \frac{2^{j-1} + 1}{2^{\ell}} \quad , \quad \ell > j > 0. \tag{48}
\]

Proof: In general, the discrete autocorrelation wavelets \( \Psi_{j}(\tau) \) can be derived from discretizing the continuous wavelet autocorrelation function. For Haar wavelets this is given by

\[
\Psi(\tau) = \int_{-\infty}^{\infty} \psi_{H}(x) \psi_{H}(x-\tau) \ dx
\]

\[
= \begin{cases} 
1 - 3|\tau| & |\tau| \in [0,1/2] \\
|\tau| - 1 & |\tau| \in (1/2,1],
\end{cases}
\]

\[
\Psi_{j}(\tau) = \Psi\left(\frac{|\tau|}{2^{j}}\right), \tag{49}
\]

for \( \tau = -(n-1), \ldots, 0, \ldots, (n-1) \), \( n = 2^{j} \), and zero for other values of \( \tau \). So our (discrete) autocorrelation wavelet coefficients are the interpolants of \( \Psi \) on nested increasing dyadic grids.

The matrix \( A \) has elements

\[
A_{j\ell} = \sum_{\tau} \Psi_{j}(\tau) \Psi_{\ell}(\tau)
\]

for \( j, \ell = 1, \ldots, \infty \).

27
Case i. For $\ell = j$ we may write

$$A_{jj} = \sum_{\tau = -(n-1)}^{n-1} \Psi_j^2(\tau)$$

$$= 1 + 2 \left\{ \sum_{\tau = 1}^{2^j-1} \Psi^2 \left( \frac{\tau}{2^j} \right) + \sum_{\tau = 2^j-1+1}^{2^j-1} \Psi^2 \left( \frac{\tau}{2^j} \right) \right\}$$

$$= 1 + 2 \left\{ A + B \right\},$$

because $\Psi_j(0) = 1$ for all $j$. The sum is split into two components $A$ and $B$ because $\Psi(\tau)$ is composed of two linear parts and for each set of indices in the sum the component will be the same (i.e. for the first sum $\tau = 1, \ldots, 2^j-1$ and $\frac{\tau}{2^j} \in [0, 1/2)$ and so the $1 - 3|\tau|$ linear component would be used).

After some algebra:

$$A = 2^{j-3} - \frac{3}{8} + \frac{3}{2^{j+2}},$$

and

$$B = \frac{2^{-2j}}{6} (2^j - 1) 2^{j-1} (2^j - 1),$$

which results in

$$A_{jj} = \frac{2^{2j} + 5}{3 \cdot 2^j}$$

as required.

Case ii. For $j \neq \ell$ and without loss of generality $\ell > j$. Let $n = 2^j$. Then

$$A_{j\ell} = \sum_{\tau = -(n-1)}^{n-1} \Psi_j(\tau) \Psi_\ell(\tau)$$

$$= 1 + 2 \sum_{\tau = 1}^{n-1} \Psi \left( \frac{\tau}{2^j} \right) \Psi \left( \frac{\tau}{2^\ell} \right).$$

The largest $\tau$ can be in the above sum is $n-1 = 2^j - 1$. So the maximum value of the argument of the second $\Psi$ is

$$\frac{2^j - 1}{2^\ell} \leq \frac{2^j}{2^{\ell+m}} = 2^{j-m} \leq \frac{1}{2},$$

for $m = \ell - j \geq 1$. So since the argument of the second $\Psi$ is always less than 1/2 we can substitute the $1 - 3|\tau|$ part of (49) into the summation to obtain

$$A_{j\ell} = 1 + 2 \sum_{\tau = 1}^{n-1} \Psi \left( \frac{\tau}{2^j} \right) \left( 1 - \frac{3\tau}{2^k} \right),$$

After a considerable amount of algebra we obtain

$$A_{j\ell} = 2^{-\ell} \left( 2^{2j-1} + 1 \right)$$

as required. \textit{End of Proof of Lemma}

\textbf{Continuation of Proof of Theorem 4.13 for Haar wavelets}

Hence,

$$B_{jj} = 1/3 + 5/3 \cdot 2^{-2j}, \quad B_{j\ell} = \frac{2^{3j/2-1} + 2^{-j/2}}{2^{3\ell/2}}, \quad \ell > j > 0.$$  \hfill (50)
That is, for \( \ell = j + m, m > 0, j \) fixed,  
\[
B_{j,j+m} = 2^{-3m/2-1} \left( 1 + 2 \cdot 2^{-2j} \right).
\]

Note that \( B \) is symmetric matrix because \( A \) is. Formulae (50) only refer to the upper triangular portion of the matrix (i.e. you cannot switch the indices in \( B_{\ell \ell} \) to obtain the lower triangle).

To use properties (i)–(iii) to bound the smallest eigenvalue of \( B \) from below, we decompose \( B \) as follows:
\[
B = T + R = T + \tilde{D}'\tilde{T}\tilde{D},
\]
where \( T \) is a symmetric Toeplitz with \( t_0 = 1/3 \) and \( t_m = 2^{-3m/2-1} \), where \( \tilde{D} = D^2 \) is diagonal with \( d_j = 2^{-j}, j > 0 \), and where \( \tilde{T} \) is again symmetric Toeplitz with \( \tilde{t}_0 = 5/3 \) and \( \tilde{t}_m = 2^{-3m/2} \), i.e. \( \tilde{T} = I + 2T \).

Because by (ii), \( \lambda_{\min}(B) \geq \lambda_{\min}(T) + \lambda_{\min}(R) \), it is sufficient to show that \( \lambda_{\min}(T) \geq \delta > 0 \), implying \( \lambda_{\min}(\tilde{T}) > 0 \), hence \( \lambda_{\min}(R) \geq 0 \), using (i), as clearly \( \lambda_{\min}(D) \geq 0 \).

We use (iii) to show what remains to, i.e. \( \lambda_{\min}(T) \geq \delta > 0 \).

In (iii), \( t_0 = 1/3 \) and \( t_n = 1/2 \) \( n^{\text{th}} \) with \( t = 2^{-3/2} \). That is, in equation (47) \( f(z) \) takes the form, for \( |z| = 1 \):
\[
f(z) = 1/3 + \text{Re} \left( \sum_{n=1}^{\infty} 2^{-3n/2} z^n \right) = 1/3 + t \frac{\text{Re}(z) - t}{1 + t^2 - 2t \text{Re}(z)},
\]
by some elementary algebra. As this is a strictly monotonically increasing function in \(-1 \leq \text{Re}(z) \leq 1 \), \( \min_{|z|=1} f(z) = f(-1) \) with
\[
f(-1) = \frac{2 \left( \sqrt{2} - 1 \right)}{3 \left( 2\sqrt{2} + 1 \right)} =: \delta > 0.
\]

Hence, \( \lambda_{\min}(B) \geq \delta \) which, by (i) implies that \( \lambda_{\min}(A) \geq \lambda_{\min}^2(D) \lambda_{\min}(B) = 2\delta > 0 \).

**Extension of Proposition 4.13 to Daubechies compactly supported wavelets**

The previous proof for Haar wavelets depends on explicit knowledge of the autocorrelation wavelet \( \psi_j(\tau) \) which, in the case of Haar, can be obtained from the continuous autocorrelation wavelet given in (49).

For Daubechies compactly supported wavelets there is no known closed form expression for the mother wavelet \( \psi \) and therefore a direct approach is not likely. The Fourier transform of the continuous autocorrelation function is simply \( |\hat{\psi}(\omega)|^2 \) and can be written as
\[
|\hat{\psi}(\omega)|^2 = \frac{1}{2\pi} M_0 \left( \frac{\omega}{2} + \pi \right) \prod_{j=1}^{\infty} M_0(2^{-j-1}\omega) \tag{51}
\]
where \( M_0(\omega) = |m_0(\omega)|^2 \) and \( m_0(\omega) \) is the discrete Fourier transform of the low-pass quadrature mirror filter \( \{h_k\} \) (defined and used later in the proof of Corollary 4.15b, see (55)). Even for the simplest non-Haar Daubechies compactly supported wavelet of order \( N = 2 \) substituting the quantity
\[
M_0(\omega) = \frac{1}{2} + \frac{9}{16} \cos \omega - \frac{1}{16} \cos 3\omega
\]
into (51) results in an infinite product of trigonometric polynomials and then we would have to find the inverse Fourier transform of that.

However, Dubuc and Deslauriers [11] have devised a scheme that is capable of generating the continuous autocorrelation function. Our proof above only required the value of the continuous
acf on an increasingly fine dyadic grid. The Dubuc and Deslauriers scheme uses known values of the acf on a grid and produces values of the acf at the mid-points of that grid and could thus could be employed to numerically generate the $A$ matrix. Further investigation of the scheme might also provide a method for analytically computing the entries of $A$.

**Proof of Proposition 4.14**

The proof is simply a consequence of part a) of Corollary 4.15, in (26), i.e.

$$S_j(z) = \lim_{\tau \to \infty} \widetilde{F}_{j,T}(z) = \sum_{\tau} c(z, \tau) \kappa_j(\tau)$$

with

$$\widetilde{F}_{j,T}(z) = \sum_{\ell=-J}^{-1} A_{j}^{-1} \sum_{\tau} c_T(z, \tau) \Psi_{\ell}(\tau)$$

and Proposition 4.9.

**Proof of Corollary 4.15**

a) Recall first that, $B = D' \cdot A \cdot D$ where $D$ is a diagonal matrix: $D = \text{diag}(2^{j/2})_{j \in \mathbb{N}}$, i.e. $B_{\ell j} = 2^{j/2} A_{\ell j} 2^{\ell/2}$.

Replacing in (26) $c_T(z, \tau)$ by its asymptotic limit $c(z, \tau)$ with $R_T = |c_T(z, \tau) - c(z, \tau)| = O(T^{-1})$ uniformly in both arguments, we get, using $\sum_{\tau} |\Psi_{\ell}(\tau)| = O(2^{-\ell})$:

Observe that $\widetilde{F}_{j,T}(z), j = -1, \ldots, -J$ is actually based on the finite–dimensional $A_{\ell}^{-1}$, with $J = J(T) = \log_2(T)$. That is,

$$\widetilde{F}_{j,T}(z) = \sum_{\ell=-J}^{-1} A_{\ell}^{-1} \sum_{\tau} c_T(z, \tau) \Psi_{\ell}(\tau)$$

$$= \sum_{\ell=-J}^{-1} A_{\ell j}^{-1} \sum_{\tau} [c(z, \tau) + R_T] \Psi_{\ell}(\tau)$$

$$= \sum_{\ell=-J}^{-1} A_{\ell j}^{-1} \sum_{\tau} c(z, \tau) \Psi_{\ell}(\tau) + \tilde{R}_{j,T}$$

Observe that, as $T \to \infty, J(T) \to \infty$, the first part tends to $S_j(z)$ because from Definition 4.8

$$\sum_{\ell} A_{\ell j}^{-1} \sum_{\tau} c(z, \tau) \Psi_{\ell}(\tau) = \sum_{\ell} A_{\ell j}^{-1} \sum_{\tau} \sum_{n} S_n(z) \Psi_{n}(\tau) \Psi_{\ell}(\tau)$$

$$= \sum_{\ell} A_{\ell j}^{-1} \sum_{n} S_n(z) \Psi_{n}(\tau) \Psi_{\ell}(\tau)$$

$$= S_j(z) .$$

The remainder behaves like

$$\tilde{R}_{j,T} \leq T^{-1} \sum_{\ell=-J}^{-1} \sum_{\tau} |\Psi_{\ell}(\tau)|$$

$$\leq T^{-1} \sum_{\ell=-J}^{-1} 2^{\ell/2} B_{\ell j}^{-1} 2^{\ell/2} O(2^{-\ell})$$

$$\leq T^{-1} \sum_{\ell=-J}^{-1} 2^{\ell/2} B_{\ell j}^{-1} O(2^{-\ell/2})$$

$$= O(2^{j/2} 2^{\ell/2} T^{-1})$$
because the norm of $B^{-1}$ is bounded.
Further, we need to check the summability condition:

$$\sum_j \tilde{T}_{j,T}(z) = \sum_j S_j(z) + \sum_j \tilde{R}_{j,T},$$

with $\sum_j S_j(z) < \infty$ by condition (18) and

$$\sum_j \tilde{R}_{j,T} = \sum_{j=-1}^{-J} \tilde{R}_{j,T} + \sum_{j>J} \tilde{R}_{j,T}$$

$$= O \left( 2^{J/2} T^{-1} \right) + O \left( T^{-1} \right)$$

as $\sum_{j>-J} 2^{j/2} = O \left( 2^{-J/2} \right)$.

Observing that $2^{J/2} = O \left( T^{1/2} \right)$ ends the proof.

b) The uniqueness of the representation follows from Theorem 4.13. However, the following explicit proof of uniqueness does not depend on Theorem 4.13, shows the linear independency of the family $\{\Psi_j\}$ and works irrespective of the wavelet used.

Suppose there were two spectral representations of the same LSW process, i.e. there existed $w^{(1)}_{j,k,T}$ and $w^{(2)}_{j,k,T}$ with, for $i = 1, 2$,

$$\sup_k \left| w^{(i)}_{j,k,T} - W^{(i)} (\frac{k}{T}) \right| = O \left( T^{-1} \right)$$

and with the same covariance. This means that, for each $z \in (0, 1)$ and each $\tau \in Z$,

$$c(z, \tau) = \sum_{j=-\infty}^{-1} S_j^{(1)}(z) \Psi_j(\tau) = \sum_{j=-\infty}^{-1} S_j^{(2)}(z) \Psi_j(\tau) ,$$

with $S_j^{(i)}(z) = |W_j^{(i)}(z)|^2$, $i = 1, 2$. So we have to show that

$$0 = \sum_{j=-\infty}^{-1} \Delta_j(z) \Psi_j(\tau), \quad (52)$$

for each $z \in (0, 1)$ and each $\tau \in Z$. This implies that

$$\Delta_j(z) = 0 \quad \forall \ j < 0 , \ z \in (0, 1), \quad (53)$$

where $\Delta_j(z) := S_j^{(1)}(z) - S_j^{(2)}(z)$.

We actually show that (52) implies that

$$\tilde{\Delta}_j(z) = 0 \quad \forall \ j < 0 , \ z \in (0, 1), \quad (54)$$

where $\tilde{\Delta}_j(z) := 2^j \Delta_j(z)$.

To do so, observe that by Parseval's relation, starting from the definition in (24),

$$A_{jt} := \sum_\tau \Psi_j(\tau) \Psi_\ell(\tau) = \int d\omega \bar{\Psi}_j(\omega) \bar{\Psi}_\ell(\omega),$$

with

$$\bar{\Psi}_j(\omega) = |\psi_j(\omega)|^2 = 2^{-j} |m_1(2^{-j+1}\omega)|^2 \prod_{\ell=0}^{-(j+2)} |m_0(2^{\ell}\omega)|^2 , \quad (55)$$
where $m_0(\omega) = 2^{-1/2} \sum_k h_k \exp(-i\omega k)$, for a sequence $\{h_k\}_k$ of wavelet filter weights with
$\sum_k h_k^2 = 1$, $1/\sqrt{2} \sum_k h_k = 1$, and $|m_1(\omega)|^2 = 1 - |m_0(\omega)|^2$.
Now we show that (52) implies (54):

$$0 = \sum_{j=-\infty}^{-1} \tilde{\Delta}_j(z) \tilde{\Psi}_j(\tau),$$

hence, for all $\ell < 0$, all $\tau \in \mathcal{Z}$, hence, for all $\ell < 0$, all $\tau \in \mathcal{Z},$

$$0 = \sum_{\ell} \sum_j \tilde{\Delta}_j(z) \tilde{\Delta}_\ell(z) \sum_\tau \tilde{\Psi}_j(\tau) \tilde{\Psi}_\ell(\tau),$$

and with Parseval’s equation (cf. the definition of the matrix $A$ above)

$$0 = \sum_{\ell} \sum_j \tilde{\Delta}_j(z) \tilde{\Delta}_\ell(z) \int d\omega \tilde{\Psi}_j(\omega) \tilde{\Psi}_\ell(\omega),$$

i.e.,

$$0 = \int d\omega \left( \sum_j \tilde{\Delta}_j(z) \tilde{\Psi}_j(\omega) \right)^2.$$ (56)

With $\sum_j S_j(z) < \infty$ we infer that $\sum_j \tilde{\Delta}_j(z) \tilde{\Psi}_j(\omega)$ is a continuous function in $\omega \in [-\pi, \pi]$, because every $2^j \tilde{\Psi}_j(\omega)$ is, (as it is a trigonometric polynomial which is uniformly bounded above by one), and because $\sum_j 2^{-j} \tilde{\Delta}_j(z) < \infty$.

Hence, equation (56) is equivalent to

$$0 = \sum_{j=-\infty}^{-1} \tilde{\Delta}_j(z) \tilde{\Psi}_j(\omega) \quad \forall \omega \in [-\pi, \pi] \quad \forall z \in (0, 1).$$ (57)

To show the pointwise implication of (54) by (57), we use again arguments of continuity and inserting successively the zeros of $|m_1(2^{-(j+1)}\omega)|^2$ which are at $\pi/2^{-j}, j < 0$.

Fix $z \in (0, 1)$, and let $\tilde{\Delta}_j := \tilde{\Delta}_j(z)$ for this fixed $z$:
First insert $\omega = \pi$ to show that $\tilde{\Delta}_{-1} = 0$: This is due to $|m_0(\pi)|^2 = 0$, i.e. $\tilde{\Psi}_j(\pi) = 0, j = -2, -3, \ldots$ and $\tilde{\Psi}_{-1}(\pi) \neq 0$ (as $|m_1(\pi)|^2 = 1$). In order to show that $\tilde{\Delta}_{-2} = 0$, observe that

$$0 = \sum_{j=-2}^{-\infty} \tilde{\Delta}_j \tilde{\Psi}_j(\omega) = |m_0(\omega)|^2 \left( \sum_{j=-2}^{-\infty} \tilde{\Delta}_j 2^{-j} |m_1(2^{-j+1}\omega)|^2 \prod_{\ell=1}^{-(j+2)} |m_0(2^{\ell}\omega)|^2 \right).$$

As $|m_0(\omega)|^2$ is analytic and $m_0(\omega)$, as a trigonometric polynomial, has only finitely many zeros, the function in brackets, which is again continuous, must vanish identically (i.e. for all $\omega$ except on a set of Lebesgue–measure zero). Insertion of $\omega = \pi/2$ results into $\tilde{\Delta}_{-2} = 0$, as $|m_1(2 \cdot \pi/2)|^2 = 1 \neq 0$ and $|m_0(2 \cdot \pi/2)|^2 = 0$.

Iteration of this scheme for $j = -3, -4, \ldots$ leads to the assertion (54).

**Proof of Proposition 5.3**

We cite this proof from [31], Proposition 3.3. First observe that the following holds in mean square sense:

$$d_{j,k,T} = \sum_{t=1}^{T} X_{t,T} \psi^P_{j,k}(t)$$
\[
\sum_{t=1}^{T} \left\{ \sum_{l,m} w_{l,m,T}^0 \psi_{l,m}^0(t) \xi_{lm} \right\} \psi_{jk}^D(t) = \sum_{l,m} w_{l,m,T}^0 \xi_{lm} \left\{ \sum_{t=1}^{T} \psi_{l,m}^0(t) \psi_{jk}^D(t) \right\} = w_{j,k:T}^0 \xi_{jk}.
\]

Hence, since the \( \xi_{lm} \) are uncorrelated

\[
\text{cov}\{d_{j,k:T}; d_{l,m:T}\} = |w_{l,m,T}^0|^2 \delta_{jl} \delta_{km}.
\]

This implies that as \( T \to \infty \)

\[
E \ I_{j,T}^j = |w_{j,2^j:2^j,T}^0|^2 = S_j(z) + O(T^{-1}),
\]

by assumption 3 of Definition 4.1. Further,

\[
\text{var} \ I_{j,T}^j = \text{var} \ |d_{j,2^j:2^j,T}|^2 \\
= |w_{j,2^j:2^j,T}^0|^4 \text{var} (\xi_{j,2^j:2^j,T}^2) \\
= 2 |w_{j,2^j:2^j,T}^0|^4 \left( \sum_{j=1}^{2^j} \xi_{j,T}^2 \right)^2 + O(T^{-1}),
\]

where we use that \( \xi_{j,2^j:2^j,T}^2 \) is \( \chi^2 \)-distributed because \( \xi_{jk} \) is Gaussian. Finally, due to the Theorem of Isserlis for Gaussian variables

\[
\text{cov}\{I_{k,T}^j; I_{l,m,T}^l\} = 2 |w_{j,k:T}^0|^4 \delta_{jl} \delta_{km} = 2 \delta_{jl} \delta_{km} \left\{ S_j \left( \frac{k}{2^T} \right)^2 + O(T^{-1}) \right\}.
\]

**Proof of Proposition 5.4**

Both (30) and (31) are consequences of (27) and its proof, in Corollary 4.15.

For the variance, we observe that the wavelet periodograms are \( \chi^2 \)-distributed, and show the same asymptotic behaviour as for the Fourier analogue, i.e. the variance is asymptotically proportional to the square of the expectation.

**Proof of Corollary 5.6**

Proof of 1.: By the spectral representation of the stationary \( X_t = \int \exp(i\omega t) \, dZ(\omega) \),

\[
\lim_{T \to \infty} E I_{k,T}^j = \lim_{T \to \infty} E \left| \sum_{t=1}^{T} X_t \psi_{jk}(t) \right|^2 \\
= \lim_{T \to \infty} E \left| \sum_{t=1}^{T} \int \exp(i\omega t) \, dZ(\omega) \psi_{jk}(t) \right|^2 \\
= \lim_{T \to \infty} \int \left( \sum_{t=1}^{T} \psi_{jk}(t) \exp(i\omega t) \right) dZ(\omega) \\
= \lim_{T \to \infty} \int \left| \sum_{t=1}^{T} \psi_{jk}(t) \exp(i\omega t) \right|^2 f(\omega) \, d\omega \\
= \int \left| \overline{\psi_{jk}}(\omega) \right|^2 f(\omega) \, d\omega.
\]

Proof of 2.: This proof is based on two approximations:

First,

\[
E I_{t,T}(\omega) = \sum_s c_T(t/T, s) \exp(-i\omega s),
\]

33
where, with Remark 4.18 and equation (28),
\[
c_T(u, s) := \text{cov} \{ X_{[u T], T}; X_{[u T + s], T} \} = \int_{-\pi}^{\pi} A([u T]/T, \lambda) \overline{A([u T + s]/T, \lambda)} \exp(i\lambda s) \, d\lambda.
\]
Here the idea is, that as \( T \to \infty \), due to the smoothness of \( A \) in the first argument,
\[
c_T(u, s) \to c(u, s) := \int_{-\pi}^{\pi} |A(u, \lambda)|^2 \exp(i\lambda s) \, d\lambda.
\]
The second approximation leads to the remainder of
\[
\sum_t \int d\omega \ E I_{1, T}(\omega) \ W_{jk}^*(t, \omega) = \sum_t \int d\omega \ E I_{t+k, T}(\omega) \ W_j^*(t, \omega) = \int d\omega \ f(u, \omega) |\overline{\psi}_j(\omega)|^2 + O \left( \frac{M_j}{T} \right),
\]
if \( \sup_u \sum_\tau |\tilde{A}(u, \tau)| < \infty \) (cf. Assumption (A3) of [24]).
To show this, use \( f \left( \frac{t+k}{T}, \omega \right) = f \left( \frac{t}{T}, \omega \right) + O \left( \frac{t}{T} \right) \), and
\[
\int d\omega \sum_t M_j \ W_j^*(t, \omega) = O (M_j),
\]
which – for less regular spectra – needs techniques as in [24], proof of Lemma 3.2.

**Proof of Theorem 6.1**

Using our model of local wavelet stationarity, it is obvious that \( S_j \) and \( S_j^2 \) are both Lipschitz, too. Then (41) is an immediate consequence of Proposition 5.4, part 1, and equation (30) with the rates given there.

For the variance we restrict ourselves to note that this formula is completely similar to what we are used to from the variance of empirical wavelet coefficients of traditional or time-dependent "Fourier" periodograms, where the limit of the expectation is simply the spectrum and the limit of the variance the square of the spectrum. (Cf., e.g., [33], equation (4.7) in Theorem 4.3, which gives the asymptotic variance as an integral of the squared time-dependent spectrum times the incorporated wavelets. Note however, that ours is for fixed "frequency", i.e. scale \( j \), hence only wavelets over local time are incorporated). We just have to substitute the square of the spectrum by the square of the limit of the expectation of the wavelet periodogram in (41).

**Proof of Proposition 6.3**

First, for equation (45) to be well defined, we observe that, as \( 2^{J_0} = o(T) \), in equation (31) the components
\[
E \ L_{k, T}^j = \sum_{t=-J_0}^{J_0} A_{j t}^{-1} E \ I_{k, T}^t
\]
are summable over \( j \),
\[
\sum_j E \ L_{k, T}^j < \infty,
\]
which is a consequence of the proof of Corollary 4.15 (with \( J = J_0 \)).
As we are not ambitious to get the fastest possibly rate of mean–square consistency for \( \overline{c}(z, \tau) \),

\[34\]
we simply proceed like follows:

\[
E \int_0^1 (\bar{c}(z, \tau) - c(z, \tau))^2 \, dz = E \int_0^1 \left( \sum_{j = -J_0}^{-1} (\bar{S}_j(z) - S_j(z)) \, \psi_j(\tau) \right)^2 \, dz + R_{J_0},
\]

where 

\[
R_{J_0} = O(2^{-J_0}),
\]

at least (as \( \sum_{j < -J_0} S_j(z) \, \psi_j(\tau) = O(2^{-J_0}) \). 

For the leading term we continue like

\[
E \int_0^1 \left( \sum_{j = -J_0}^{-1} (\tilde{S}_j(z) - S_j(z)) \, \psi_j(\tau) \right)^2 \, dz \\
\leq \sum_j \sum_{\ell} E \left( \int (\tilde{S}_j(z) - S_j(z))^2 \, dz \right)^{1/2} \left( \int (\tilde{S}_{\ell}(z) - S_{\ell}(z))^2 \, dz \right)^{1/2} \psi_j(\tau) \, \psi_{\ell}(\tau) \\
\leq \left( \sum_{j = -J_0}^{-1} \psi_j(\tau) \left( E \int (\tilde{S}_j(z) - S_j(z))^2 \, dz \right)^{1/2} \right)^2 \\
= o \left( \log(T) \cdot \left( \frac{\log(T)}{T^{2/3}} \right)^{1/2} \right)^2,
\]

by equation (44), holding for \( \tilde{S}_j(z) \) as well, and as \( 2^{-k} = o(T) \) by assumption. Balancing both rates leads to the assertion of mean-square consistency.

8 acknowledgements

The authors would like to thank Prof. Peter Fleming, Dr Andrew Sawczenko and Jeanine Young of the Institute of Child Health, Royal Hospital for Sick Children, Bristol for supplying the ECG data. Guy Nason was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) Grant GR/K70236. Von Sachs was supported in part by Deutsche Forschungsgemeinschaft Grant Sa 474/3-1 and by an EPSRC Visiting Fellowship Research Grant GR/L52673.

Parts of the paper were written while Rainer von Sachs visited the Department of Statistics, Stanford University. It is also based on intensive work during his visits to the Department of Mathematics, University of Bristol. Von Sachs would like to thank a lot of people at both institutions for providing a warm and fruitful atmosphere of hospitality and work. Also, many thanks to Kai Schneider, who was the first to discuss related ideas with von Sachs long before this work got started, and to Rainer Dahlhaus and Jonathan Raz who followed the development of our work with continuous interest and helpful comments. Finally, we would like to acknowledge the help of Ingrid Daubechies, Lutz Duembgen and Martin Hanke which we received over discussions on the technical aspects of proving Theorem 4.13 and Corollary 4.15.

References


