Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum

Guy P. Nason
Department of Mathematics, University of Bristol, Bristol, England
E-mail: G.P.Nason@bristol.ac.uk

Rainer von Sachs
Institut de Statistique, UCL, Louvain-la-Neuve, Belgium
E-mail: rvs@stat.ucl.ac.be

and

Gerald Kroisandt
Fachbereich Mathematik, Universität Kaiserslautern, Germany
E-mail: kroisandt@itwm.uni-kl.de

[27th October 1998]

Abstract.
This article defines and studies a new class of non-stationary random processes constructed from discrete non-decimated wavelets which generalizes the Cramér (Fourier) representation of stationary time series. We define an evolutionary wavelet spectrum (EWS) which quantifies how process power varies locally over time and scale. We show how the EWS may be rigorously estimated by a smoothed wavelet periodogram and how both these quantities may be inverted to provide an estimable time-localized autocovariance. We illustrate our theory with a pedagogical example based on discrete non-decimated Haar wavelets and also a real medical time series example.

Keywords: Non-stationary time series; Wavelet processes; Wavelet spectrum;
Wavelet periodogram; Local stationarity; Non-linear wavelet shrinkage

Financial support from the contract "Projet d'Actions de Recherche Concertes" no. 93/98-164 of the Belgian Government is gratefully acknowledged.

1 Introduction

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. This article introduces representations of non-stationary time series, i.e. data with a *time–varying second order structure*, in terms of discrete non-decimated wavelets. This gives us a tool that permits quantification of the autocovariance of a non-stationary time series as the series evolves through time. Time series can be non-stationary in many ways and several methods already exist. This article contributes methodology which is "optimal" for a certain class of non-stationary processes but also presents useful interpretable information for many others. Our work provides an additional complementary tool rather than overthrowing existing methodology.

Stationary stochastic processes X_t, $t \in \mathbb{Z}$, may be written

$$X_t = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t)\, d\zeta(\omega), \tag{1}$$

where $d\zeta(\omega)$ is an orthonormal increment processes (see Priestley (1981)). Further, the autocovariance of X_t has a well-known Fourier representation in terms of the spectrum:

$$c_X(\tau) = \int_{-\pi}^{\pi} f(\omega) \exp(i\omega\tau)\, d\omega. \tag{2}$$

In this article we concentrate on processes whose second-order structure *changes over time*, for example we could introduce a time-dependency into the autocovariance or spectrum. Alternatively, we could replace the amplitude, $A(\omega)$, in (1) by a time-varying version, $A_t(\omega)$ (e.g. the *oscillatory* and *locally stationary* processes of Priestley (1981) and Dahlhaus (1997) respectively).

Our approach is different in that we replace the set of harmonics $\{\exp(i\omega t) | \, \omega \in [-\pi, \pi]\}$ by a set of discrete non-decimated wavelets (see Nason and Silverman (1994) or Strang (1993) for introductions to wavelets). 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 (1991), Flandrin (1993)). The question immediately arises: is it possible and meaningful to use such atomic decompositions to represent, not necessarily stationary, stochastic processes?

Figure 1 shows a heart rate (ECG) recording of a infant. It is unlikely that this will be a stationary time series. One reason for this, of interest to paediatricians, is that the heart rate varies considerably over time and changes significantly between periods of sleep and waking. For time series such as these Section 2 introduces a possible underlying models based on discrete non-decimated wavelets, a time-localized spectrum to be estimated using a wavelet periodogram and a time-localized autocovariance. Section 3 covers rigorous estimation of these quantities. Section 4 shows how the smoothed wavelet periodogram of the ECG series can reveal important features of interest that cannot be elicited from the time series, standard periodogram, or some time-frequency methods.

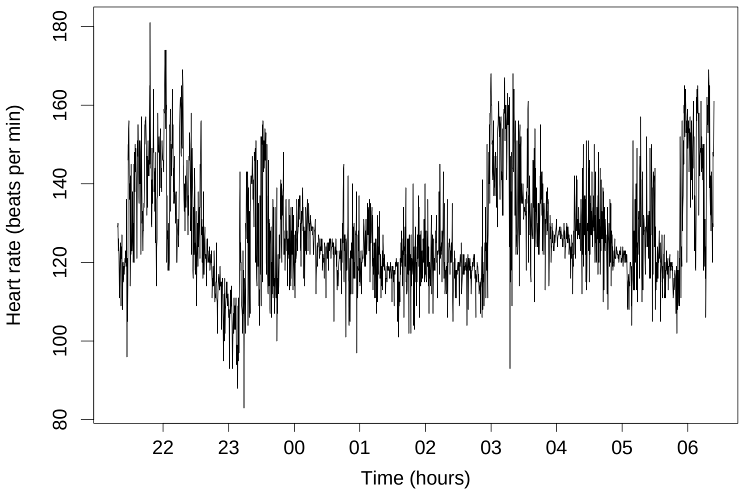

Figure 1: *Heart rate recording of 66 day old infant. Series is sampled at* $\frac{1}{16}$*Hz and is recorded from 21:17:59 to 06:27:18. There are* $T = 2048$ *observations.*

Our article formalizes the heuristic ideas of Nason and Silverman (1995) who suggested using non-decimated wavelets as a means for producing "local spectral density estimates". Our formalization permits proper identification and estimation of a wavelet spectrum from a single finite–length stretch of the process $\{X_t\}$ by restricting the time–variation of the second order structure of $\{X_t\}$ precisely as in the case of time–varying Fourier spectra, see Dahlhaus (1997), von Sachs and Schneider (1996), Neumann and von Sachs (1997).

Most existing work on wavelets with stochastic processes (Cambanis and Masry (1994), Cambanis and Houdré (1995), Kawasaki and Shibata (1995) and Cheng and Tong (1996)) does not aim to give a decomposition with respect to an (orthogonal) increment process in the time-scale plane. These papers focus on probabilistic approximations and do not cover estimation. Morettin and Chang (1995) develop a wavelet periodogram based on an orthonormal wavelet basis but only for stationary time series. Using localized representations with truly local basis functions is not completely new, see, for example Section 13.5 of Walter (1994) or Abry *et al.* (1995).

2 Theoretical foundations: The wavelet process model

This section introduces a process model based on discrete non-decimated wavelets (LSW process); an evolutionary wavelet spectrum that quantifies power in LSW processes at a particular time and scale and a time-localized autocovariance.

Discrete non-decimated wavelets are discretized versions of the continuous non-decimated wavelets (see Nason and Silverman (1995)). For example, the discrete non-decimated Haar wavelets at scales -1 and -2 respectively are

$$\psi_{-1} = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \text{ and } \psi_{-2} = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right) \tag{3}$$

and so on. The key point for discrete *non-decimated* wavelets is they can be shifted to any data location and not just by shifts by 2^{-j} as in Mallat's discrete wavelet transform (DWT). Discrete non-decimated wavelets are no longer orthogonal but an overcomplete collection of shifted vectors. We define the quantity ψ_{jk} to be the kth element in the vector ψ_j and $\psi_{jk}(\tau)$ is the kth element of the vector $\psi_{j(k-\tau)}$, i.e. ψ_{jk} shifted by integers τ.

2.1 Locally stationary wavelet (LSW) processes

We first define our process model constructed from genuine time-scale building blocks or "atoms", $\psi_{jk}(t)$, with random amplitudes, $w^0_{j,k;T}\ \xi_{jk}$, as follows.

Definition 2.1 *The* **locally stationary wavelet** *(LSW) processes are a sequence of doubly-indexed stochastic processes* $\{X_{t,T}\}_{t=0,\dots,T-1}$, $T = 2^J \geq 1$ *having the following representation in the mean–square sense*

$$X_{t,T} = \sum_{j=-J}^{-1} \sum_{k} w^0_{j,k;T}\ \psi_{jk}(t)\ \xi_{jk}, \tag{4}$$

where ξ_{jk} *is a random orthonormal increment sequence and where* $\{\psi_{jk}(t)\}_{jk}$ *is a discrete non-decimated family of wavelets for* $j = -1, -2, \dots, -J(T)$, $k = 0, \dots, T-1$ *based on a mother wavelet* $\psi(t)$ *of compact support.*

The quantities in representation (4) possess the following properties:

1. $E\xi_{jk} = 0$ *for all* j, k. *Hence* $EX_{t,T} = 0$ *for all* t *and* T.

2. $\mathrm{cov}(\xi_{jk}, \xi_{\ell m}) = \delta_{j\ell}\delta_{km}$.

3. *There exists for each* $j \leq -1$ *a Lipschitz–continuous function* $W_j(z)$ *for* $z \in (0, 1)$ *which fulfils the following properties:*

$$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1). \tag{5}$$

The Lipschitz constants L_j *are uniformly bounded in* j *and*

$$\sum_{j=-\infty}^{-1} 2^{-j}\ L_j < \infty\ . \tag{6}$$

There exists a sequence of constants C_j such that for each T

$$\sup_k \left| w^0_{j,k;T} - W_j\left(\frac{k}{T}\right) \right| \leq C_j/T \tag{7}$$

where for each $j = -1, \ldots, -J(T) = -\log_2(T)$ the $\sup$ is over $k = 0, \ldots, T-1$, and where $\{C_j\}$ fulfils

$$\sum_{j=-\infty}^{-1} C_j < \infty \; . \tag{8}$$

Remark 2.2 (Rescaled time) As with Dahlhaus (1997) we are not observing a fixed continuous time process on an increasingly finer mesh as $T \to \infty$. Instead assumption 3 uses *rescaled time* $z = k/T \in (0,1)$ which permits increasing amounts of data about the local structure of $W_j(z)$ to be collected as $T \to \infty$. Hence assumption 3 allows us to (asymptotically) identify the model coefficients determined by uniquely defined $W_j(z)$. In this sense we have a unique representation (4) given the fixed wavelet basis, although the $\{w^0_{j,k;T}\}$ themselves cannot be unique because of the overcompleteness of the non-decimated basis. Our method differs from Dahlhaus (1997) in that we track local power in the covariance decomposition of X_t with respect to *scales* instead of frequencies along time. The smoothness assumption on $W_j(z)$, as a function of rescaled time, z, controls the variation of each coefficient $w^0_{j,k}$ as a function of k so that it can not change too quickly. Rescaling time by T^{-1} and the conditions in Definition 2.1 (3) describe how the local structure becomes increasingly "stationary" along a growing number of local neighbourhoods with decreasing time variation.

Remark 2.3 (Example: Haar MA processes) Consider the following stationary moving average process: $X^1_t = 2^{-\frac{1}{2}}(\epsilon_t - \epsilon_{t-1})$, where $\{\epsilon_t\}$ is a purely random process with mean zero and variance of σ^2. The MA(1) process X^1_t is a LSW process with $w^0_{j,k;T}$ (and $W_j(z)$) equal to 1 for $j = -1$, $k \in \mathbb{Z}$ and zero for all other j, $\xi_{-1k} = \epsilon_k$ and $\psi_{jk}(t)$ are the Haar discrete non-decimated wavelets as defined by (3). The MA coefficients in X^1_t are just those of the scale -1 non-decimated Haar wavelets. The LSW process $X^2_t = 2^{-1}(\epsilon_t + \epsilon_{t-1} - \epsilon_{t-2} - \epsilon_{t-3})$ uses scale -2 discrete non-decimated Haar wavelets is MA(3) and has $w^0_{j,k;T}$ (and $W_j(z)$) equal to 1 for $j = -2$, $k \in Z$ and zero for all other j and $\xi_{-2k} = \epsilon_k$. Continuing we can build the MA($2^r - 1$) process X^r_t using scale $-r$ discrete non-decimated Haar wavelets. We call the collection $\{X^r_t\}^{\infty}_{r=1}$ the *Haar moving average* processes (Daubechies MA processes can be similarly constructed using Daubechies' compactly supported wavelets). Any MA process can be represented by a linear combination of Haar MA processes and often the representation is sparse (because any sequence in $l^2(\mathbb{Z})$ can be decomposed into a Haar basis). In these examples $W_j(z)$ is a constant function of z: stationary processes always have this property. Remark 2.6 shows a non-stationary example where $W_j(z)$ is not constant.

Remark 2.4 (Model interpretation) Roughly speaking, we expect the amplitude $w^0_{j,k;T}$ 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 τ that matches the "wavelength" of $\psi_{jk}(t)$ which is proportional to 2^{-j} (assuming that $\psi(t)$ is localized at time zero, which it nearly always is in practice). Model (4) 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. Assumption 1 forces LSW processes to have zero mean. In practice, a time series might need to be detrended using any of the available techniques including those based on wavelets (see von Sachs and MacGibbon (1997)).

We adopt the usual Meyer-Mallat scale numbering scheme although how it adapts to increasing numbers of data points T requires some explanation. 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. $J = J(T)$). 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 longer cycles can appear in the series and so the model includes increasingly coarser wavelets. In other words $-J$ should tend to $-\infty$ with increasing T.

Wavelet devotees will note that the model (4) has no scaling function coefficient. Asymptotically, it is not required. As in traditional wavelet multiresolution analysis, as the scale tends to $-\infty$, the coarse scale approximation will finally be included in the overall sum of the details.

More general families could be substituted for the non-decimated wavelets in (4) e.g. (discretized) continuous wavelet transforms (CWT) or non-decimated wavelet *packets*, see Nason *et al.* (1997). However, non-decimated wavelets seem to give the right balance between orthogonality and too much overcompleteness as they control the redundancy in the process representation (which CWT-based models do not). The DWT permits for a rigorous theory (see von Sachs *et al.* (1998)) but does not include traditional stationary processes in the model. Though assumption 2 appears to be somewhat restrictive, note that, because we use a shift-equivariant non-decimated wavelets the model includes a large class of correlated processes. In particular LSW processes include all stationary processes with $\sum_\tau |c(\tau)| < \infty$: a large class of processes with short-range dependence (see Proposition 2.17).

2.2 The evolutionary wavelet spectrum

The *evolutionary wavelet spectrum* measures the local power (contribution to variance) in a LSW process at a particular (rescaled) time, z and scale, j (scale can be loosely interpreted as the usual time series lag). The EWS is the analogue of the usual stationary process spectrum, $f(\omega)$.

Definition 2.5 *The sequence* $\{X_{t,T}\}_{t=0,\ldots,T-1}$, *for the infinite sequence* $T \geq 1$, *has* **evolutionary wavelet spectrum** *(EWS) defined by*

$$S_j(z) := |W_j(z)|^2,\ j = -1, \ldots, -J(T), z \in (0,1), \textit{with respect to} \{\psi_{jk}\}. \tag{9}$$

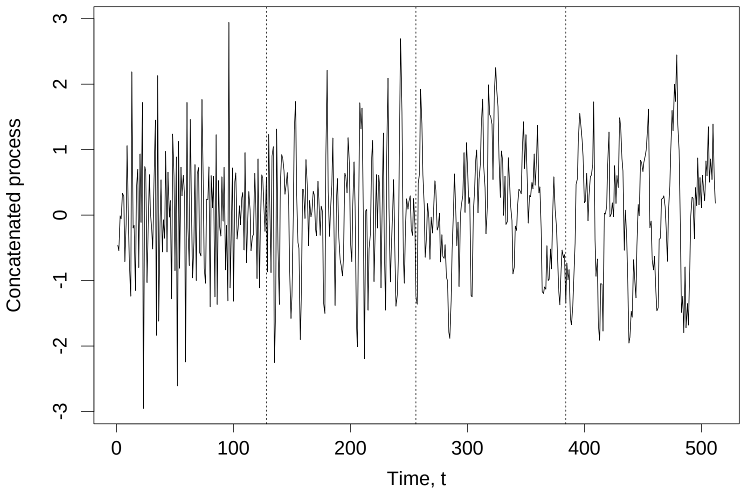

Figure 2: A concatenation of $n = 128$ observations from consecutive Haar moving average processes X^1, X^2, X^3 and X^4. The variance of the underlying purely random process was $\sigma^2 = 1$. The vertical dotted lines indicate where process X^r changes to process X^{r+1} for $r = 1, 2, 3$.

Using assumption 3 of Definition 2.1: $S_j(z) = \lim_{T\to\infty} |w^0_{j,[zT];T}|^2, \quad \forall z \in (0,1)$*, and thus fulfils* $\sum_{j=-\infty}^{-1} S_j(z) < \infty$ *uniformly in* $z \in (0,1)$.

Our LSW model (4) therefore delivers a time–scale decomposition which parallels the time–frequency decomposition of Dahlhaus' (1997). (See Remark 4.18 in von Sachs *et al.* (1997) for further connections). Our model is similar to Dahlhaus' model in two other ways: the EWS is defined only for $z \in (0,1)$, as boundaries do not make sense in this model and the EWS is uniquely defined (in terms of localized autocovariance, shown by Theorem 2.13).

Remark 2.6 (Example: non-stationary processes.) Define the EWS of the rth Haar MA process to be $S^r_j(z)$. Then $S^r_j(z) = \delta_{-jr}$ for all $z \in (0,1)$ since $|W_j(z)|^2$ is equal to 1 for $j = -r$ and zero otherwise. Suppose we 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. One realization of such a concatenated series is shown in Figure 2. The EWS for the concatenated series will simply be $S^1_j(z)$ followed successively by $S^2_j(z), S^3_j(z)$ and $S^4_j(z)$. An estimate of the EWS for the concatenated series appears in Figure 3 (so, for example, the bottom line of coefficients in Figure 3 estimates $S_{-1}(t)$ and shows that it is non-zero only when the MA(1) process, $X^1(t)$, is "active". Then, at time $t = 128$, the MA(3) process, $X^2(t)$, becomes active and this is reflected by the non-zero block of $S_{-2}(t)$ coefficients until time $t = 256$, and so on).

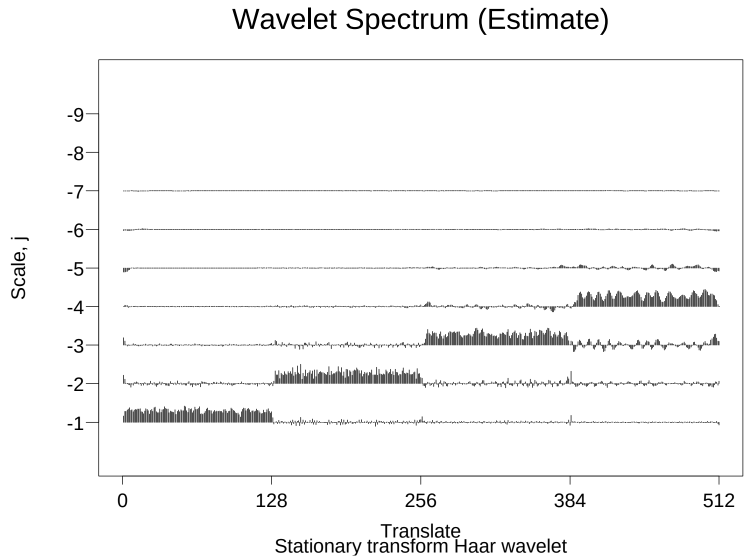

Figure 3: Mean of 100 corrected wavelet periodogram estimates of $S_j(t)$ for $j = -1, \ldots, -9$. (Each corrected wavelet periodogram (section 3.1) was computed from an independent simulation of the concatenated MA process described in the text).

2.3 Local autocovariances and autocorrelation wavelets

If $X_{t,T}$ is a LSW process then it is no surprise that its autocovariance $c_T(z, \tau) = \operatorname{cov}\left\{X_{[zT],T}; X_{[zT]+\tau,T}\right\}$, for $z \in (0,1)$ and $\tau \in \mathbb{Z}$, also has a "wavelet–type" representation. More precisely, Proposition 2.11 shows that c_T tends to a *local autocovariance*, c, defined in Definition 2.9, which itself can be represented by a series of *autocorrelation wavelets* with coefficients given by $S_j(z)$. This is analogous to the classical stationary case where autocovariance has a Fourier representation in terms of the spectrum as in (2).

Definition 2.7 (Autocorrelation wavelets)*:* $\Psi_j(\tau) := \sum_k \psi_{jk}(0)\, \psi_{jk}(\tau)\,,\ j < 0, \tau \in \mathbb{Z}$.

Remark 2.8 (Haar autocorrelation wavelets) Discrete non-decimated Haar wavelets are sampled versions of the continuous Haar mother wavelet at different scales. Thus the Haar autocorrelation wavelet $\Psi_j(\tau)$ is itself a sampled version of the continuous Haar autocorrelation wavelet, $\Psi_H(u)$, given by

$$\Psi_H(u) = \int_{-\infty}^{\infty} \psi_H(x)\psi_H(x-u)\,dx = \begin{cases} 1 - 3|u| & \text{for } |u| \in [0, 1/2], \\ |u| - 1 & \text{for } |u| \in (1/2, 1], \end{cases} \tag{10}$$

where $\psi_H(x)$ is the Haar mother wavelet. In general if the generating wavelets $\psi(x)$ are compactly supported then so is the corresponding autocorrelation wavelet $\Psi(u)$: the support of $\Psi_H(u)$ is $[-1, 1]$. For Haar the discrete autocorrelation wavelet, $\Psi_j(\tau)$ can be written in terms of $\Psi_H(u)$ for $j < 0$ by $\Psi_j(\tau) = \Psi_H\left(2^j|\tau|\right)$, for $\tau = -(n-1), \ldots, 0, \ldots, (n-1)$,

$n = 2^{-j}$, and zero for other values of τ. For small j, for Haar wavelets, the form of $\Psi_j(\tau)$ is very simple: e.g. $\Psi_{-1}(\tau) = 1, \frac{1}{2}, 0$ for $|\tau| = 0, 1$, otherwise.

Autocorrelation wavelets are related to the autocorrelation shell of Saito and Beylkin (1993) which they use in a multiresolution analysis. Our usage 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 as shown next.

Definition 2.9 *Define the* **local autocovariance** *(LACV) $c(z,\tau)$ of a LSW process with EWS $\{S_j(z)\}$ by*

$$c(z,\tau) \;=\; \sum_{j=-\infty}^{-1} S_j(z)\,\Psi_j(\tau)\;,\;\tau\in\mathbb{Z}\;,z\in(0,1)\,. \tag{11}$$

The following relates to stationary processes even though Definition 2.9 refers to the more general case of LSW processes.

Remark 2.10 (Haar MA process example) The Haar MA process $X^1(t)$ has autocovariance $c^1_X(\tau) = \delta_{\tau,0} - \frac{1}{2}\delta_{|\tau|,1}$ which is precisely the autocorrelation wavelet $\Psi_{-1}(\tau)$. Therefore $X^1(t)$ has an extremely sparse representation in terms of the autocorrelation wavelets: $c^1_X(\tau) = \Psi_{-1}(\tau)$ i.e. equation (11) with $S_j(z)$ equal to 1 for $j = -1$ and zero otherwise. Indeed, by construction, the autocovariance of $X^r(t)$ is always sparsely represented: $c^r(\tau) = \Psi_{-r}(\tau)$ for all $r \in \mathbb{Z}$. Similar sparse representations occur if we replace Haar by other compactly supported wavelets. Such representations are instructive as they tell us what processes are sparsely represented in $\{S_j(z)\}_{j=-\infty}^{-1}$ and thus which are likely to be well-estimated by our wavelet machinery.

The above remark considered a simple stationary case. However, Definition 2.9 applies for more general LSW processes. The following proposition shows how a LSW process autocovariance, c_T, asymptotically tends to $c(z,\tau)$.

Proposition 2.11 *As $T \to \infty$, uniformly in $\tau \in \mathbb{Z}$ and $z \in (0,1)$, $|c_T(z,\tau) - c(z,\tau)\,| = O\left(T^{-1}\right)$.*

If the process is stationary then the dependence on z in c_T, c and S_j disappears and our representation turns from a local into a global one (possible because of the shift-equivariant non-decimated wavelets).

2.4 Uniqueness of the autocovariance representation

In the stationary theory representation (2) is invertible: the spectrum is also the Fourier transform of the autocovariance. The following question naturally arises: is the EWS the inverse of the LACV? Inversion is possible if we introduce a bi-orthogonal (to the non-orthogonal $\{\Psi_j(\tau)\}$) dual set, $\{\kappa_j(\tau)\}$.

Definition 2.12 *Define the operator* $A = (A_{j\ell})_{j,\ell<0}$ *by*

$$A_{j\ell} := < \Psi_j, \Psi_\ell > = \sum_\tau \Psi_j(\tau)\, \Psi_\ell(\tau)\,, \tag{12}$$

and the J*–dimensional matrix* $A_J = (A_{j\ell})_{j,\ell=-1,\dots,-J}$*. Further define, for each* $J(T) = \log_2(T)$*, the vector* $\kappa(\tau) := \{\kappa_j(\tau)\}_{j=-1,-2,\dots,-J}$ *via* $\kappa(\tau) = A_J^{-1}\ \Psi(\tau)$ *, where* $\Psi(\tau) := \{\Psi_j(\tau)\}_{j=-1,-2,\dots,-J}$*.*

The following theorem, valid for all Daubechies' compactly supported wavelets, shows that A is an invertible operator and that for each J, the norm of A_J^{-1} is bounded from above by some constant C_J.

Theorem 2.13 *(a) The family* $\{\Psi_j(\tau)\}$ *is linearly independent, hence the operator* A *is invertible, with a bound* C_J *on* $\|A_J^{-1}\|$ *for each* J*.*
(b) Consequently, the EWS is uniquely defined given the corresponding LSW process.

Theorem 2.13(b) supplies an unambiguous representation of the EWS in terms of the LACV which can be seen by inserting equation (11) into the following equation (13).

Proposition 2.14 *Inverse formula of equation (11):*

$$S_j(z) \;=\; \sum_\tau c(z,\tau)\, \kappa_j(\tau) \;=\; \sum_\ell A_{j\ell}^{-1} \sum_\tau c(z,\tau)\, \Psi_\ell(\tau) \qquad \forall\ j\ < 0. \tag{13}$$

We emphasize that for each T the process representation (4) cannot be unique but the representation of the local autocovariance (11) actually is unique. Further, Corollary 2.16 below shows that if we replace c by the empirical c_T in (13) then its inverse representation converges to the EWS. This is a stronger result than (13) and requires a finer characterisation of the redundancy in A provided by the next theorem.

Theorem 2.15 *Let* $\lambda_{\min}(A)$ *denote the smallest eigenvalue of* A*. Then, for Haar and Shannon wavelet families, there exists a* $\delta > 0$ *such that* $\lambda_{\min}(A) \geq \delta$ *hence* $\|A^{-1}\| < \infty$*, i.e.* A *is positive–definite and has a bounded inverse.*

The Shannon wavelet is the limiting wavelet in the Daubechies' compactly supported series as the number of vanishing moments $N \to \infty$, see Chui (1997). We conjecture that Theorem 2.15 is valid for all Daubechies' compactly supported wavelets and provide strong evidence in the appendix.

Corollary 2.16 *The EWS* $\{S_j(z)\}$ *as defined in Definition 2.5 also arises as the asymptotic limit of*

$$\widetilde{T}_{j,T}(z) := \sum_{\ell=-J}^{-1} A_{j\ell}^{-1}\ \sum_\tau c_T(z,\tau)\, \Psi_\ell(\tau)\,, j = -1, \dots, -J(T) = \log_2(T). \tag{14}$$

That is,

$$\lim_{T\to\infty} \widetilde{T}_{j,T}(z) = S_j(z) \ = \ \lim_{T\to\infty} |w^o_{jk;T}|^2 \ , \ j \le -1 \ , \ z \in (0,1)\,, \tag{15}$$

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

Finally we inspect the links between LSW and stationary processes.

Proposition 2.17 (a) All stationary processes with absolutely summable autocovariance $\sum_\tau |c_X(\tau)| < \infty$ are LSW processes (with respect to wavelets fulfilling Theorem 2.15).
(b) Conversely, any LSW process with time independent EWS fulfilling the additional assumption $\sum_j 2^{-j}\ S_j < \infty$ is stationary with absolutely summable autocovariance.

3 Estimation theory

3.1 The wavelet periodogram

The wavelet periodogram is constructed using the wavelet family specified *a priori* in representation (4). The interesting situation of what happens when a different wavelet is used to estimate the EWS is left for future work.

Definition 3.1 *The* empirical wavelet coefficients *of an LSW process* $X_{t,T}$ *are given by*

$$d_{j,k;T} := \sum_{t=0}^{T-1} X_{t,T}\ \psi_{jk}(t). \tag{16}$$

For fixed j the number of summands in (16) does not change with T because the wavelet is compactly supported. The following key statistic is the analogue of the classical periodogram from stationary theory.

Definition 3.2 *The* **wavelet periodogram** *of a LSW process* $X_{t,T}$ *is given by* $I^j_{k,T} := |d_{j,k;T}|^2$.

As for classical periodograms the wavelet periodograms have asymptotically non-vanishing variance (Proposition 3.3) and need to be smoothed to obtain consistency. Various local smoothing methods could be used but we choose to use non-linear wavelet shrinkage as described below. Further, Figure 4 shows that the wavelet periodograms for scale j contain information from other scales $j' \neq j$ (e.g. power from scale -4 has leaked into scale -5). Writing $\mathbf{I}_k := (I^j_k)_{j=-1,\dots,-J}$ Proposition 3.3 shows that the corrected periodogram $\mathbf{L}_k = A_J^{-1}\mathbf{I}_k$, for $k = 0,\dots,T-1$ is asymptotically unbiased for the EWS. Figure 3 shows the superiority of the corrected periodogram compared to the raw one in Figure 4. Figure 3 correctly shows abrupt changes as one MA process changes into another. Figure 4 incorrectly exhibits significant power at levels -6 and -5 whereas the process only had power at scales of -1,-2,-3 and -4 (corresponding to X^r, $r = 1,2,3,4$ respectively).

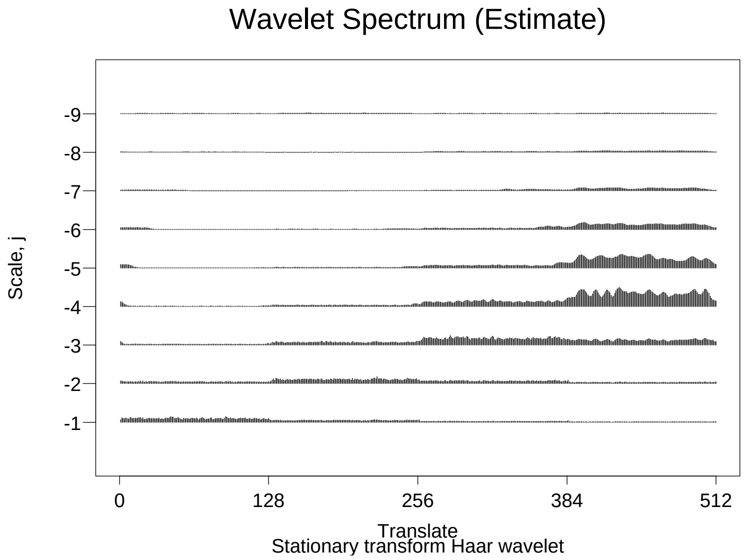

Figure 4: Mean of 100 uncorrected wavelet periodograms of $I^j_{k;512}$ for $j = -1, \ldots, -9$, for the simulated concatenated MA processes of Remark 2.6.

3.2 Asymptotics of the wavelet periodogram

For this section we assume that the conditions of Theorem 2.15 are fulfilled and that the ξ_{jk} in (4) are Gaussian, i.e. $\{X_{t,T}\}$ is Gaussian. The wavelet periodogram has analogous properties to those of the classical periodogram as follows.

Proposition 3.3 (expectation)

$$E\, I^j_{[zT],T} \;=\; \sum_\ell A_{j\ell}\, S_\ell(z) \;+ O\left(2^{-j}/T\right) \qquad \forall z \in (0,1). \tag{17}$$

Hence, for the vector of periodograms $\mathbf{I}(z) := \{I^\ell_{[zT],T}\}_{\ell=-1,\ldots,-J}$, *and the vector of the corrected periodograms* $\mathbf{L}(z) := \{L^j_{[zT],T}\}_{j=-1,\ldots,-J}$ *with* $\mathbf{L}(z) = A_J^{-1}\,\mathbf{I}(z)$

$$E\,\mathbf{L}(z) \;=\; E\, A_J^{-1}\,\mathbf{I}(z) \;=\; \mathbf{S}(z) \;+\; O\left(T^{-1}\right) \qquad \forall z \in (0,1)\,, \tag{18}$$

where $\mathbf{S}(z) := \{S_j(z)\}_{j=-1,\ldots,-J}$.

(variance) *The asymptotic variance of* $I^j_{[zT],T}$ *can be written as a square of linear combinations of the leading term of (17), i.e. appropriately weighted EWS.*

(covariance) *The correlation between two wavelet periodograms* $I^j_{k,T}$ *and* $I^\ell_{m,T}$ *decays with increasing distance between the location* k *on scale* j *and location* m *on scale* ℓ*: For example, within one scale* $j = \ell$*, it is zero as soon as* $|k-m|$ *exceeds the overlap of the corresponding wavelets support.*

The variance and covariance cannot be written in compact form so we omit them here. For further details see von Sachs *et al.* (1997), Proposition 5.3, Lemma 5.5 and Corollary 5.6.

3.3 Wavelet periodogram smoothing

Like the stationary case the wavelet periodogram is *not* a consistent estimator and needs to be smoothed. For each fixed scale j we smooth $I^j_{k,T}$ as a function of $z = k/T$ using DWT shrinkage or the translation-invariant (TI) denoising of Coifman and Donoho (1995). In practice we use the latter but only provide theoretical results only for the first (which parallels existing theory, e.g. Donoho (1995), Neumann and von Sachs (1997), Johnstone and Silverman (1997), and von Sachs and MacGibbon (1997). Our results also hold for the TI denoising). Since we know (below) how to do wavelet shrinkage on the χ^2–distributed wavelet periodogram we smooth first, then correct (by A^{-1}). This results in a simpler analysis than if we corrected before smoothing.

Von Sachs *et al.* (1997) describes in detail how to smooth using an orthonormal second-stage wavelet basis $\{\tilde{\psi}_{\ell m}\}$ of $L_2([0,1])$. The smoothing is carried out by non-linear thresholding of the empirical wavelet coefficients, $\widehat{v}_{\ell m}$, of $I^j(z)$ and then inverting to give the estimate $\tilde{S}_j(z)$. The appropriate threshold $\lambda = \lambda(j, \ell, m, T)$ can be determined from the following theorem.

Theorem 3.4 *For a Gaussian LSW processes and using a wavelet $\tilde{\psi}$ of bounded variation, the wavelet coefficients $\widehat{v}_{\ell m}$, with $2^\ell = o\,(T)$, obey uniformly in m,*

$$E\ \widehat{v}_{\ell m} - \int_0^1 \sum_n A_{jn}\ S_n(z)\ \tilde{\psi}_{\ell m}(z)\,dz \ = \ O\ \left(2^{\ell/2}/T\right), \tag{19}$$

and,

$$\mathrm{var}\ (\widehat{v}_{\ell m}) \ = \ 2\,T^{-1} \int_0^1 \left(\sum_n A_{jn}\ S_n(z)\right)^2 \ \tilde{\psi}^2_{\ell m}(z)\,dz \ + \ O\ \left(2^\ell\ T^{-2}\right). \tag{20}$$

Using process normality we proceed as in Gao (1993) and von Sachs and Schneider (1996) to show that with the following universal threshold the adaptive estimate $\widetilde{S}_j(z)$ attains the usual L_2 rate of convergence.

Theorem 3.5 *Under the assumptions of Theorem 3.4, with threshold given by $\lambda^2(l, m; j; T) = \mathrm{var}\ (\widehat{v}_{lm})\ \log^2(T)$, for each fixed j,*

$$\int_0^1 E\ \left(\widetilde{S}_j(z)\ -\ S_j(z)\right)^2\ dz \ = \ O\ \left(\log^2(T)/T^{\frac{2}{3}}\right)\ . \tag{21}$$

This theorem is based on existing results on quadratic forms of Gaussian variables, which are χ^2–distributed (see Neumann and von Sachs (1995), Theorem 3.1 A). For non–normality, techniques as in Neumann and von Sachs (1997) could also be applied. In practice some modification might be appropriate such as thresholding the log periodogram. This transform stabilizes the coefficient variance, pulls their distribution closer to normality (Priestley (1981)), and permits use of a universal threshold suitable for normally distributed data. Such transforms are well known in the classical periodogram and time-dependent case (e.g. von Sachs and Schneider (1996).)

3.4 Local variance and autocovariance estimation

Finally we address the problem of local autocovariance estimation by inversion of (smoothed) wavelet periodograms. A useful descriptive tool for estimating *local variance* can be obtained by summing the (smoothed) wavelet periodogram $\tilde{I}_k^j$ over scales j.

Proposition 3.6 *Let $\tilde{S}_j^*(z)$ denote the result of applying the inverse matrix A^{-1} to the smoothed wavelet periodogram $\widetilde{S}_j(z)$. Define $\widetilde{c}(z,\tau)$ by replacing $S_j(z)$ by $\tilde{S}_j^*(z)$ in equation (11) and replacing the lower sum limit $j = -\infty$ by $-J_0$. Let $T \to \infty$ and let $2^{J_0} = o(T)$. Then $\widetilde{c}(z,\tau)$ is a consistent estimator of $c(z,\tau)$ because for each fixed $\tau \in \mathbb{Z}$,*

$$E \int_0^1 (\, \widetilde{c}(z,\tau) \, - c(z,\tau)\,)^2 \, dz \, = \, o\,(1)\,.$$

Proof. The proof appears in von Sachs *et al.* (1997)

4 Examples and applications

4.1 Concatenated Haar example

Figure 3 shows the average of wavelet periodogram estimates from 100 realizations of the concatenated process of Remark 2.6. Figure 5 shows the corrected smoothed estimate $\tilde{S}_j^*(z)$ computed from one realization of the concatenated process: each level was smoothed by Coifman and Donoho's (1995) TI-denoising method using threshold $\lambda = \hat{\sigma} \log T$ from section 3.3 on levels 3 and finer using a MAD estimate of σ. Figure 5 is reasonable with power coming and going approximately where it is meant to in scales -1 to -4. The estimate is not perfect and several issues effect the estimate's quality. Firstly, which analyzing wavelet should the wavelet periodogram use? Even though the underlying process was Haar we used a smoother Daubechies extremal-phase wavelet $N = 3$. We advocate smoother analyzing wavelets to avoid "leakage" to surrounding scales because of their shorter support in the Fourier domain and they also help with the smoothing itself as the raw estimate looks less spiky and variable (indeed for the ECG data we use $N = 10$). Secondly, the smoothing is not perfect and the estimate is more variable in scale -4 and over-smoothed in scales -1 and -2 (a consequence of the raw wavelet periodogram being more variable in the coarser scales, because the raw periodogram coefficients are approximately χ^2 and the mean increases with coarser scale. The mean increase is because of the non-decimated transform and the decreasing diagonal of A^{-1} attempts to deal with this. However, the large *variation* remains). The question of how best to smooth needs to be investigated further.

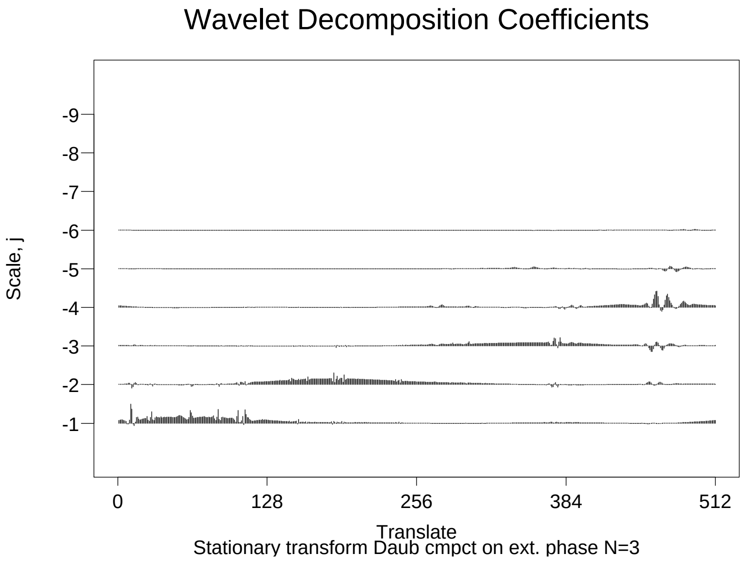

Figure 5: Corrected smoothed wavelet periodogram of one realization of the concatenated process given in Remark 2.6. The Daubechies extremal-phase wavelet $N = 3$ was used in the representation and the Daubechies least-asymmetric $N = 4$ wavelet was used for smoothing.

4.2 The infant ECG data (continued from Section 1)

Figure 6 shows the raw wavelet periodogram for the series of 2048 points in Figure 1, for all possible scales from -1 to $J = -11$. Figure 7 shows a smoothed corrected estimate of the EWS for the infant ECG series. The Daubechies' (1992) least–asymmetric wavelets of order $N = 10$ were used to form both the wavelet periodogram and to do TI wavelet smoothing of the log periodogram with a soft universal threshold $\lambda = \hat{\sigma}\sqrt{2 \log T}$ on scales 7 and finer using a MAD estimate of σ.

The information in Figure 7 is actually highly meaningful as is shown by the enlargement of scale $j = -1$ in Figure 8 (solid line). The dotted line in Figure 8 indicates the infant's sleep state as judged by a trained human observer (from brain wave measurements, EEG, and eye movements, EOG). The observer classifies the infant's sleep state as quiet (1), between quiet and active(2), active(3) and awake(4). There is a strong relationship between the estimate, $\tilde{S}^*_{-1}(z)$, and the sleep state. In particular, periods of activity and wakefulness occur whilst the estimate of $S_{-1}(z)$ is large and periods of quiet sleep occur when it is small. Our estimate is useful for feeding into models that can predict infant sleep state from heart rate (heart rate is easy and routine to measure, sleep state is more tricky and can be distressing for parent and infant. For further discussion of the modelling aspects for time series of these type see Nason *et al.* (1997)). In effect, $\tilde{S}^*_{-1}$ measures the local stochastic variability of the heart rate signal at that scale and it correlates fairly well with the sleep state variable. Figure 7 also indicates that scales -2 through to -5 also contain similar information to that presented in scale -1.

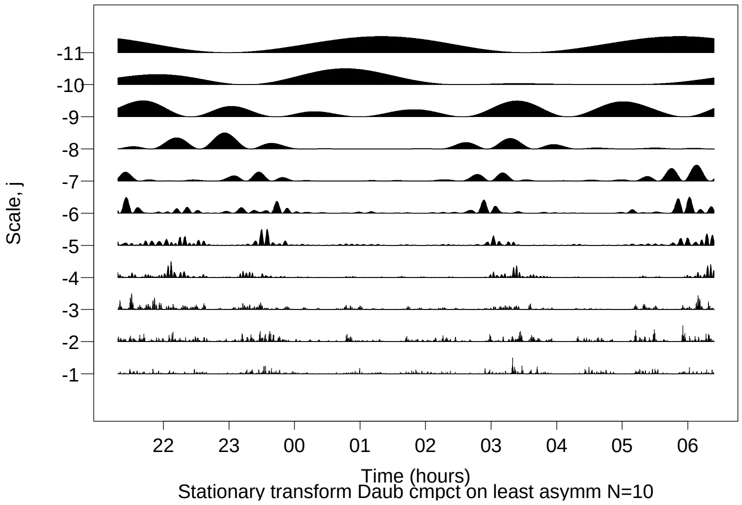

Figure 6: The raw wavelet periodogram $I^j_{k,2048}$ for the infant ECG data. Each level in the periodogram has been scaled independently so that detail at all levels can be seen (some of the larger scale levels, j more negative, are almost 10 times as large).

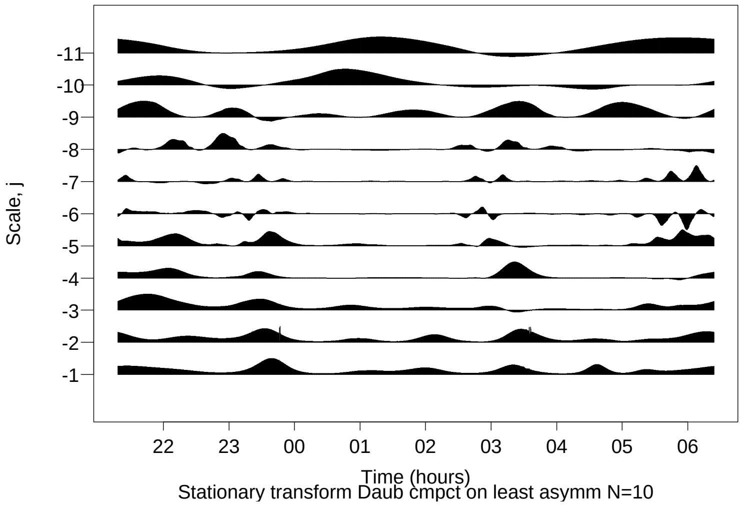

Figure 7: Estimate of EWS $\tilde{S}^*_j(z)$ for infant ECG data (levels scaled independently).

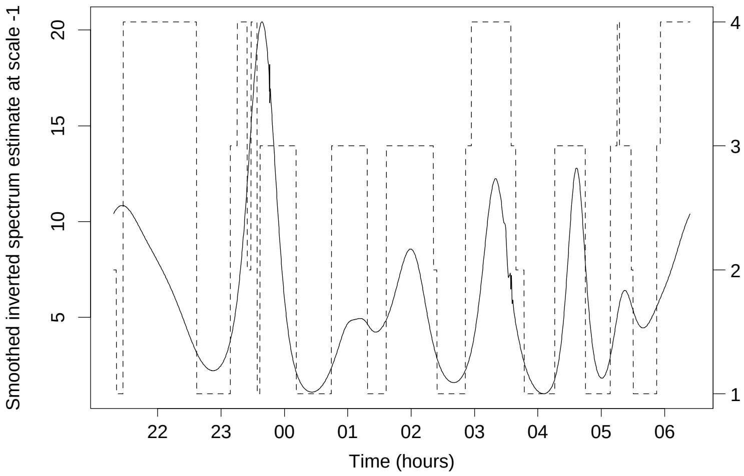

Figure 8: Solid line is EWS estimate at scale -1: $\tilde{S}^*_{-1}(z)$ for infant ECG data. The dotted line indicates infant's sleep state as determined by expert analysis of EEG and EOG (independent of heart rate). The dotted line value is indicated by the right-hand axis: 1=quiet sleep, 2=between 1 and 3, 3=active sleep, 4=awake.

5 Appendix: Proofs

Proof of Proposition 2.11

Using the representation of the process $X_{t,T}$ given in (4) the covariance is given by

$$\begin{aligned} c_T(z,\tau) = \operatorname{cov}\{X_{[zT],T}; X_{[zT]+\tau,T}\} &= \sum_{jk} |w^0_{j,k;T}|^2 \psi_{jk}([zT])\psi_{jk}([zT]+\tau) \\ &= \sum_{j\ell} |w^0_{j,[zT]+\ell;T}|^2 \psi_{j,[zT]+\ell}([zT])\psi_{j,[zT]+\ell}([zT]+\tau) \\ &= \sum_{jk} |w^0_{j,[zT]+k;T}|^2 \psi_{jk}(0)\ \psi_{jk}(\tau). \end{aligned}$$

With condition (8) $\left| |w^0_{j,[zT]+k;T}|^2 - S_j(z+k/T) \right| = O(C_j/T)$, and since S_j is Lipschitz $|S_j(z+k/T) - S_j(z)| = O(L_j|k|/T)$. Hence,

$$\begin{aligned} \left| \sum_{jk} |w^0_{j,[zT]+k;T}|^2 \psi_{jk}(0)\psi_{jk}(\tau) - c(z,\tau) \right| &= \left| \sum_{jk} |w^0_{j,[zT]+k;T}|^2 \psi_{jk}(0)\psi_{jk}(\tau) - \sum_j S_j(z)\Psi_j(\tau) \right| \\ &\le T^{-1} \sum_{jk} (L_j|k| + C_j)\ |\psi_{jk}(0)\psi_{jk}(\tau)| = O(T^{-1}), \end{aligned}$$

due to conditions (6) and (8). Note also that $\sum_j S_j(z)\Psi_j(\tau) < \infty$, by (5) and as $\Psi_j(\tau) = O(1)$ uniformly in τ.

Proof of Theorem 2.13

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 $i = 1, 2$,

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

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

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

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

$$0 = \sum_{j=-\infty}^{-1} \Delta_j(z)\Psi_j(\tau), \tag{22}$$

for each $z \in (0, 1)$ and each $\tau \in \mathbb{Z}$. This implies that $\Delta_j(z) = 0, \forall j < 0, z \in (0, 1)$, where $\Delta_j(z) := S^{(1)}_j(z) - S^{(2)}_j(z)$. We actually show that (22) implies that

$$\widetilde{\Delta}_j(z) = 0 \qquad \forall j < 0\ , z \in (0, 1), \tag{23}$$

where $\widetilde{\Delta}_j(z) := 2^j\Delta_j(z)$. To do so, observe that by Parseval's relation, starting from the definition in (12),

$$A_{j\ell} := \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau) = \frac{1}{2\pi}\int d\omega \widehat{\Psi_j}(\omega)\widehat{\Psi_\ell}(\omega), \tag{24}$$

with

$$\widehat{\Psi_j}(\omega) = |\widehat{\psi_j}(\omega)|^2 = 2^{-j}|m_1(2^{-(j+1)}\omega)|^2 \prod_{\ell=0}^{-(j+2)} |m_0(2^\ell\omega)|^2, \tag{25}$$

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 (22) implies (23): assume $0 = \sum_{j=-\infty}^{-1} \widetilde{\Delta}_j(z)\Psi_j(\tau)$ hence, for all $\ell < 0$, all $\tau \in \mathbb{Z}$,

$$0 = \sum_\ell \sum_j \widetilde{\Delta}_j(z)\widetilde{\Delta}_\ell(z) \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau).$$

Using Parseval (24) we obtain $0 = \sum_\ell \sum_j \widetilde{\Delta}_j(z)\widetilde{\Delta}_\ell(z) \int d\omega \widehat{\Psi}_j(\omega)\widehat{\Psi}_\ell(\omega)$, i. e.,

$$0 = \int d\omega \left(\sum_j \widetilde{\Delta}_j(z)\widehat{\Psi}_j(\omega) \right)^2 . \tag{26}$$

With $\sum_j S_j(z) < \infty$ we infer that $\sum_j \widetilde{\Delta}_j(z)\widehat{\Psi}_j(\omega)$ is continuous in $\omega \in [-\pi, \pi]$, because every $2^j\widehat{\Psi}_j(\omega)$ is, (as it is a trigonometric polynomial uniformly bounded above by one), and because $\sum_j 2^{-j}|\widetilde{\Delta}_j(z)| < \infty$. Hence, equation (26) is equivalent to

$$0 = \sum_{j=-\infty}^{-1} \widetilde{\Delta}_j(z)\widehat{\Psi}_j(\omega) \quad \forall \omega \in [-\pi, \pi] \quad \forall z \in (0,1). \tag{27}$$

To show the pointwise implication of (23) by (27), we again use continuity arguments and successively insert zeros of $|m_0(2^{-(j+1)}\omega)|^2$ which are at $\pi/2^{-(j+1)}$, $j < 0$. Fix $z \in (0,1)$, and let $\widetilde{\Delta}_j := \widetilde{\Delta}_j(z)$ for this fixed z. First insert $\omega = \pi$ to show that $\widetilde{\Delta}_{-1} = 0$. This is due to $|m_0(\pi)|^2 = 0$, i.e. $\widehat{\Psi}_j(\pi) = 0, j = -2, -3, \ldots$ and $\widehat{\Psi}_{-1}(\pi) \neq 0$ (as $|m_1(\pi)|^2 = 1$). In order to show that $\widetilde{\Delta}_{-2} = 0$, observe that

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

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. Insertion of $\omega = \pi/2$ results into $\widetilde{\Delta}_{-2} = 0$, 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 (23).

Proof of Theorem 2.15 (for the Haar wavelet)
We show that there exists some $\delta > 0$ such that $\lambda_{\min}(A) \geq \delta$ by showing that $\lambda_{\min}(B) \geq \delta$, where $B = D' \cdot A \cdot D$ with diagonal matrix $D = \mathrm{diag}(2^{j/2})_{j<0}$, i.e., $B_{j\ell} = 2^{j/2}A_{j\ell}2^{\ell/2}$. This is sufficient by (i) of the following matrix properties and Toeplitz matrix theory:

(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 $\lambda_{\min}(A) \geq \lambda_{\min}(B) + \lambda_{\min}(C)$.

(ii) (Reichel and Trefethen (1992), Theorem 3.1(i)) Let T be Toeplitz (and Hermitian) with elements $\{t_0, t_1, \ldots\}$ Let $f(z) = \sum_{n=-\infty}^{\infty} t_n z^n$ for $z \in \mathbf{C}$ be the symbol of the operator associated with T. If $\sum_n |t_n| < \infty$, then $f(z)$ is analytic in the open unit disk D in the complex plane and continuous in the closed unit disk $\Delta = D \cup S$, where S denotes the unit circle. The spectrum Λ of the (Laurent) operator T is fully understood, as $\Lambda(T) = f(S)$. If, additionally, T is symmetric then an estimate of the smallest eigenvalue of T is

$$\min_{|z|=1} f(z) = \min_{|z|=1} t_0 + 2\mathrm{Re}\left(\sum_{n=1}^{\infty} t_n z^n\right). \tag{28}$$

For convenience our indices will now run from 1 to ∞ rather than -1 to $-\infty$. Using straightforward algebra we can explicitly derive formulae for entries of A from Ψ_j given by (10) (see von Sachs *et al.* (1997) for more detail). The elements of A are:

$$A_{jj} = \frac{2^{2j}+5}{3\cdot 2^j},\, A_{j\ell} = \frac{2^{2j-1}+1}{2^\ell}, \ell > j > 0. \tag{29}$$

Now, with $B_{j\ell} = 2^{-j/2}A_{j\ell}2^{-\ell/2}$ and using equation (29),

$$B_{jj} = 1/3 + 5/3\cdot 2^{-2j},\ B_{j\ell} = \frac{2^{3j/2-1}+2^{-j/2}}{2^{3\ell/2}},\ \ell > j > 0. \tag{30}$$

For $\ell = j+m, m > 0, j$ fixed $B_{j,j+m} = 2^{-3m/2-1}(1+2\cdot 2^{-2j})$. Note that B is symmetric because A is. However, Formulae (30) only refer to the upper triangular portion of the matrix B (you cannot switch indices in these formulae). To use properties (i)–(iii) to bound the smallest eigenvalue of B from below, decompose B as follows:

$$B = T + R = T + \widetilde{D}^T \widetilde{T} \widetilde{D},$$

where T is a symmetric Toeplitz with $t_0 = 1/3$ and $t_m = 2^{-3m/2-1}$, where $\widetilde{D} = D^2$ is diagonal with $\widetilde{d}_j = 2^{-j}, j > 0$, and where $\widetilde{T}$ is again symmetric Toeplitz with $\widetilde{t}_0 = 5/3$ and $\widetilde{t}_m = 2^{-m/2}$. Because by (ii), $\lambda_{\min}(B) \geq \lambda_{\min}(T) + \lambda_{\min}(R)$, it is sufficient to show that both $\lambda_{\min}(T) \geq \delta > 0$ and $\lambda_{\min}(\widetilde{T}) \geq \widetilde{\delta} \geq 0$, the latter implying $\lambda_{\min}(R) \geq 0$, using (i), as clearly $\lambda_{\min}(D) \geq 0$. Use (iii) to treat the two Toeplitz matrices as follows. We start by showing $\lambda_{\min}(T) \geq \delta > 0$. In (iii), $t_0 = 1/3$ and $t_n = 1/2t^{|n|}$ with $t = 2^{-3/2}$. In equation (28) $f(z)$ takes the form, for $|z| = 1$:

$$f(z) = 1/3 + Re\left(\sum_{n=1}^{\infty} 2^{-3n/2} z^n\right) = 1/3 + t\frac{Re(z) - t}{1 + t^2 - 2tRe(z)},$$

by some elementary algebra. This is a strictly monotonically increasing function in $-1 \leq Re(z) \leq 1$, $\min_{|z|=1} f(z) = f(-1)$ with $f(-1) = \frac{2(\sqrt{2}-1)}{3(2\sqrt{2}+1)} =: \delta > 0$. By exactly the same principle, now with $\widetilde{t}_0 = 5/3$ and $\widetilde{t}_n = (2^{-1/2})^n$,

$$\lambda_{\min}(\widetilde{T}) \geq 5/3 + 2Re\left(\sum_{n=1}^{\infty} (-1)^n 2^{-n/2}\right) = \frac{(9+4\sqrt{2})}{3(2\sqrt{2}+3)} =: \widetilde{\delta} > 0.$$

Hence, $\lambda_{\min}(B) \geq \delta$ which, by (i) implies that $\lambda_{\min}(A) \geq \lambda^2_{\min}(D^{-1})\lambda_{\min}(B) = 2\delta > 0$.

Remark: There is strong evidence that the above Haar proof can be extended to other Daubechies compactly supported wavelets with orders $N > 1$. We briefly outline this extension, see von Sachs *et al.* (1997)) for further details. The key to the extension is the relationship between A and B to equivalent A^* and B^* arising from the continuous autocorrelation function built from time–continuous Daubechies wavelets. The continuous Haar wavelets arise as the limit of $j \to \infty$ in the interpolation scheme described in Remark 2.8. Similar schemes exist for other Daubechies' wavelets: e.g. the interpolation schemes of Dubuc and Deslauriers (1989). By using the scaling relation for continuous wavelets it is straightforward to show that B^* is truly Toeplitz regardless of the order N. Estimates of the decay of its off–diagonals, increasing with N, can serve to generalize the result for $N = 1$, for which B^* equals the T derived above. (In von Sachs *et al.* (1997) we show that $t_m = O(2^{-\gamma_N m})$, with, by Daubechies (1992), page 239, $\gamma_2 > 3/2, \gamma_3 > 5/2, \gamma_4 > 7/2, \ldots$.) On the other hand, in the Haar case, as the number j of rows tends to infinity, the matrix B itself approaches this Toeplitz $T = B^*$ from above. Equations (25) and (24) help to show this for the general case $N > 1$. Finally, we observe that as N increases the limiting T becomes progressively more diagonal (as does the starting B), which is due to the increasing decay of the Fourier transform $|\widehat{\Psi}(\omega)|$ for $|\omega| \to 0$ and $|\omega| \to \infty$, as N increases. In the limiting case ($N = \infty$) of Shannon wavelets the matrices are diagonal because of the non-overlapping support in the Fourier domain (next proof). We leave rigorous completion of the above for future work.

Proof of Theorem 2.15 (for the Shannon wavelet)

We compute A using the Fourier domain formula given by (24). The formula for the non-decimated wavelets in the Fourier domain is given by (25) and the corresponding formulae for $m_0(\omega)$ and $m_1(\omega)$ for the Shannon wavelet can be obtained from the FT of the continuous time mother and father wavelets which can be found in Chui (1997), pages 46 and 64. Define the set $C_j = \left[-\frac{\pi}{2^{-j-1}}, -\frac{\pi}{2^{-j}}\right] \cup \left[\frac{\pi}{2^{-j}}, \frac{\pi}{2^{-j-1}}\right]$. Then after some algebra we obtain $\widehat{\psi_j}(\omega) = -2^{-j/2}e^{-2^{-j-1}i\omega}\chi_{C_j}(\omega)$, where $\chi_A(\omega)$ is the indicator function of the set A. Hence, from (25) we obtain $\widehat{\Psi}_j(\omega) = |\widehat{\psi_j}(\omega)|^2 = 2^{-j}\chi_{C_j}(\omega)$. Clearly $A_{j\ell} = 0$ for $j \neq \ell$ since the supports of different $\widehat{\Psi}_j(\omega)$ do not overlap. Simple integration shows that $A_{jj} = 2^{-j}$ for $j < 0$. Hence the B operator is the identity. The proof for the continuous time operators A^* and B^* is virtually the same and shows that they are equal to A and B respectively.

Proof of Corollary 2.16

a) Recall first that $B = D' \cdot A \cdot D$ where D is a diagonal matrix: $D = \mathrm{diag}(2^{j/2})_{j<0}$, i.e. $B_{j\ell} = 2^{j/2}A_{j\ell}2^{\ell/2}$. Replacing in (14) $c_T(z,\tau)$ 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 continue as follows, while using $\sum_\tau |\Psi_\ell(\tau)| = O(2^{-\ell})$.

$$
\begin{aligned}
\widetilde{T}_{j,T}(z) &= \sum_{\ell=-1}^{-J} A^{-1}_{j\ell} \sum_\tau c_T(z,\tau)\Psi_\ell(\tau) = \sum_{\ell=-1}^{-J} A^{-1}_{\ell j} \sum_\tau [c(z,\tau) + R_T]\, \Psi_\ell(\tau) \\
&= \sum_{\ell=-1}^{-J} A^{-1}_{\ell j} \sum_\tau c(z,\tau)\Psi_\ell(\tau) + \widetilde{R}_{j,T}
\end{aligned}
$$

Observe that, as $T \to \infty, J(T) \to \infty$, the first part tends to $S_j(z)$ because from Definition 2.9

$$
\begin{aligned}
\sum_\ell A^{-1}_{\ell j} \sum_\tau c(z,\tau)\Psi_\ell(\tau) &= \sum_\ell A^{-1}_{\ell j} \sum_\tau \sum_n S_n(z)\Psi_n(\tau)\Psi_\ell(\tau) \\
&= \sum_\ell A^{-1}_{\ell j} \sum_n S_n(z) A_{n\ell} = S_j(z).
\end{aligned}
$$

The remainder behaves like

$$
\begin{aligned}
\widetilde{R}_{j,T} &\le T^{-1} \sum_{\ell=-1}^{-J} A^{-1}_{\ell j} \sum_\tau |\Psi_\ell(\tau)| = T^{-1} \sum_{\ell=-1}^{-J} 2^{j/2} B^{-1}_{\ell j} 2^{\ell/2} O(2^{-\ell}) \\
&= T^{-1} \sum_{\ell=-1}^{-J} 2^{j/2} B^{-1}_{\ell j} O(2^{-\ell/2}) = O(2^{J/2} 2^{j/2} T^{-1})
\end{aligned}
$$

because the norm of B^{-1} is bounded by Theorem 2.15. Further, we need to check the summability condition: $\sum_j \widetilde{T}_{j,T}(z) = \sum_j S_j(z) + \sum_j \widetilde{R}_{j,T}$, with $\sum_j S_j(z) < \infty$ from definition (2.5) and

$$
\sum_j \widetilde{R}_{j,T} = \sum_{j=-1}^{-J} \widetilde{R}_{j,T} + \sum_{j>-J} \widetilde{R}_{j,T} = O(2^{J/2} T^{-1}) + O(T^{-1})
$$

as $\sum_{j>-J} 2^{j/2} = O(2^{-J/2})$. Observing that $2^{J/2} = T^{1/2}$ ends the proof.

Proof of Proposition 2.17

a) For stationary processes with $\sum_\tau |c(\tau)| < \infty$ we observe that under the conditions of Theorem 2.15, $\sum_j |\kappa_j(\tau)| < \infty$ uniformly in τ, hence $\sum_j S_j < \infty$.

b) is an immediate consequence of the following corollary to Proposition 2.11.

Corollary 5.1 *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 $T \to \infty$, uniformly in z.*

Proof of Corollary 5.1

Define the following approximation: $c_{J_0}(z,\tau) = \sum_{j=-J_0}^{-1} S_j(z)\Psi_j(\tau)$. As a continuation of the proof of Proposition 2.11, using that $\sum_\tau |\Psi_j(\tau)| = O(2^{-j})$, we observe that

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

Further, $\sum_\tau |c_{J_0}(z,\tau) - c_T(z,\tau)| = O(2^{J_0}/T)$ 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 Proposition 3.3
Both (17) and (18) are consequences of (15) and its proof, in Corollary 2.16. For the variance, the wavelet periodograms are χ^2–distributed, and they have the same asymptotic behaviour as their Fourier analogues, (variance is asymptotically proportional to the expectation squared).
Proof of Theorem 3.4
With our model S_j and S_j^2 are both Lipschitz, too. Then (19) is an immediate consequence of Proposition 3.3, part 1, and equation (17) with the rates there. For the variance note that this formula is very similar to the variance of empirical wavelet coefficients of time–dependent "Fourier" periodograms, where the expectation limit is the spectrum and the variance limit is the squared spectrum (see, e.g., von Sachs and Schneider (1996), Thm 4.3, equation (4.7). Note that ours is for *fixed* "frequency".) We just substitute the squared spectrum by the squared expectation limit of the wavelet periodogram in (19).

6 Acknowledgements

We would like to thank P. Fleming, A. Sawczenko and J. Young of the Institute of Child Health, Royal Hospital for Sick Children, Bristol for supplying the ECG data. We would also like to thank Rainer Dahlhaus, Ingrid Daubechies, Lutz Duembgen, Martin Hanke, Jonathan Raz, and Kai Schneider for helpful discussions and would like to acknowledge the kindness of Dept. Statistics, Stanford University, and our respective departments. Nason was supported in part by EPSRC Grant GR/K70236 and von Sachs by DFG Grant Sa 474/3-1 and by EPSRC Visiting Fellowship GR/L52673.

References

Abry, P., Goncalves, P. and Flandrin, P. (1995) Wavelets, spectrum analysis and $1/f$-processes. *Lect. Notes Statist.,* **103**, 15–29.

Cambanis, S. and Houdré, C. (1995) On the continuous wavelet transform of second-order random processes. *IEEE Trans. Inf. Theor.,* **41**, 628–642.

Cambanis, S. and Masry, E. (1994) Wavelet approximation of deterministic and random signals: convergence properties and rates. *IEEE Trans. Inf. Theor.,* **40**, 1013–1029.

Cheng, B. and Tong, H. (1996) A theory of wavelet representation and decomposition for a general stochastic process. *Lect. Notes Statist.,* **115**, 115–129.

Chui, C.K. (1997) *Wavelets: a mathematical tool for signal analysis*. Philadelphia: SIAM.

Coifman, R.R. and Donoho, D.L. (1995) Translation-invariant de-noising. *Lect. Notes Statist.,* **103**, 125–150.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Ann. Statist.,* **25**, 1–37.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: SIAM.

Deslauriers, G. and Dubuc, S. (1989) Symmetric iterative interpolation. *Constr. Approx.,* **5**, 49–68.

Donoho, D.L. (1995) Denoising via soft thresholding. *IEEE Trans. Inf. Theor.,* **41**, 613–627.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc. B*, **57**, 301–337.

Flandrin, P. (1993) *Temps-fréquence*. Paris: Hérmes.

Gao, H.-Y. (1993) Wavelet estimation of spectral densities in time series analysis. *PhD thesis*, University of California, Berkeley.

Johnstone, I.M. and Silverman, B.W. (1997) Wavelet threshold estimators for data with correlated noise. *J. R. Statist. Soc. B*, **59**, 319–351.

Kawasaki, S. and Shibata, R. (1995) Weak stationarity of a time series with wavelet representation. *Japan J. Indust. Appl. Math.,* **12**, 37–45.

Morettin, P. and Chang, Ch. (1995) Wavelet analysis for stationary processes. *Technical report*. Departamento de Estatistica, Instituto de Matemática e Estatística, São Paulo, Brasil.

Nason, G.P., Sapatinas, T. and Sawczenko, A. (1997) Statistical modelling of time series using non-decimated wavelet representations. (submitted for publication).

Nason, G.P. and Silverman, B.W. (1994) The discrete wavelet transform in S. *J. Comput. Graph. Statist.,* **3**, 163–191.

Nason, G.P. and Silverman, B.W. (1995) The stationary wavelet transform and some statistical applications. *Lect. Notes Statist.,* **103**, 281–300.

Neumann, M.H. and von Sachs, R. (1995) Wavelet thresholding: beyond the Gaussian i.i.d. *Lect. Notes Statist.,* **103**, 301–329.

Neumann, M.H. and von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.,*, **25**, 38–76.

Priestley, M.B. (1981) *Spectral Analysis and Time Series.* London: Academic Press.

Reichel, L. and Trefethen, L.N. (1992) Eigenvalues and pseudo-eigenvalues of Toeplitz matrices. *Linear algebra and its applications*, **162–164**, 153–185.

Rioul, O. and Vetterli, M. (1991) Wavelets and signal processing. *IEEE Sig. Proc. Mag.,* **8**, 14–38.

Saito, N. and Beylkin, G. (1993) Multiresolution representations using the autocorrelation functions of compactly supported wavelets. *IEEE Trans. Sig. Proc.,* **41**, 3584–3590.

Strang, G. (1993) Wavelet transforms versus Fourier transforms. *Bull. (New Series) Am. Math. Soc.,* **28**, 288–305.

von Sachs, R. and MacGibbon, B. (1997) Non-parametric curve estimation by wavelet thresholding with locally stationary errors. Preprint “Berichte der AGTM” 179, Universität Kaiserslautern, (submitted for publication).

von Sachs, R., Nason, G.P., and Kroisandt, G. (1997) Adaptive estimation of the evolutionary wavelet spectrum. *Technical Report 516*, Statistics Department, Stanford University, Stanford.

von Sachs, R., Nason, G.P. and Kroisandt, G. (1998) Spectral representation and estimation for locally-stationary wavelet processes. *Proceedings of the workshop “Spline functions and wavelets”: Montreal*, (to appear).

von Sachs, R. and Schneider, K. (1996) Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Appl. Comput. Harm. Anal.,* **3**, 268–282.

Walter, G. (1994) *Wavelets and other Orthogonal Systems with Applications.* Boca Raton, FL: CRC Press.