# T E C H N I C A L R E P O R T

## 0336

# ON ADAPTIVE ESTIMATION FOR LOCALLY STATIONARY WAVELET PROCESSES AND ITS APPLICATIONS

S. VAN BELLEGEM and R. VON SACHS

I A P S T A T I S T I C S N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

# ON ADAPTIVE ESTIMATION FOR LOCALLY STATIONARY WAVELET PROCESSES AND ITS APPLICATIONS*

Sébastien Van Bellegem†
*Institut de statistique, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.*
*vanbellegem@stat.ucl.ac.be*

Rainer von Sachs
*Institut de statistique, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.*
*rvs@stat.ucl.ac.be*

## Abstract

The class of locally stationary wavelet processes is a wavelet-based model for covariance nonstationary zero mean time series. This paper presents an algorithm for the pointwise adaptive estimation of their time-varying spectral density. The performance of the procedure is evaluated on simulated and real time series. Two applications of the procedure are also presented and evaluated on real data. The first is a test of local significance for the coefficients of the so-called wavelet periodogram. The second is a new test of covariance stationarity.

*Keywords*: Local stationarity; wavelet spectrum; autocorrelation wavelet; wavelet periodogram; pointwise adaptive estimation; change-point; test of stationarity

AMS 2000 Subject Classification: Primary 62M15; secondary 62G05, 60G15, 65T60

*Financial support from the contract 'Projet d'Actions de Recherche Concertes' nr 98/03–217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.
†Research Fellow of the National Fund for Scientific Research (F.N.R.S.)

## 1. A wavelet-based model for locally stationary processes

### 1.1. *Introduction*

It is a common observation that many time series in the applied sciences, such as seismological, meteorological, economical or biomedical data, are not covariance stationary and show a time-varying second-order structure. That is, variance and covariance can change over time. For instance, the baby heart rate data in Figure 1 or the tremor data in Figure 2 are likely to have an inhomogeneous second-order structure. Over the past decade a variety of models that control deviation from second-order stationarity have been proposed in the literature[5,6,10,9], among which models of *local stationarity* have proven useful in order to develop a consistent theory of estimation of the model quantities[3]. Roughly speaking the change over time of a covariance which behaves locally as stationary is supposed to be slow, or at least can be controlled by some usual regularity condition (such as the existence of bounded derivatives as a function of time). This is obviously a generalisation of the idea of piecewise stationary processes, the most straightforward model of local stationarity.

Among these models of local stationarity, Nason, von Sachs and Kroisandt[14] (hereafter NvSK[14]) proposed a wavelet-based model of local stationarity for zero-mean processes with a time-varying covariance. Wavelets are known to well describe structure which is localised over time and (resolution) scale (such as the changing oscillatory behaviour of the baby EEG over certain time scales). The core of the NvSK[14] model of *locally stationary wavelet* (LSW) processes is a wavelet-type representation of the time-varying covariance, by introducing so-called *autocorrelation wavelets*. The coefficients of this representation are called *evolutionary wavelet spectrum* (EWS), their change over time controls the localised deviation from stationarity. This model has already proved useful in a variety of situations for modelling and analysing meteorological data[4], solar irradiance[20] or climate reconstruction[15] to name but a few.

In this article we use an extension of the definition of LSW processes which is due to Van Bellegem and von Sachs[18], cited hereafter as VBvS[18]. This generalisation allows to include a less smoothly over time varying covariance structure by imposing a very mild regularity condition (of bounded total variation) on the time-change. In VBvS[18] a test of significance for the coefficients of the wavelet periodogram has been developped. With this approach a pointwise adaptive estimator of the EWS has been constructed, following the method of local adaptivity of Lepski and Spokoiny[7,8]. The practical use of this theory leads to a set of non trivial new questions which are addressed in this paper. We develop (almost) fully data driven algorithms for the test of sparsity and the pointwise adaptive estimator of the EWS, and study its performance on some simulated examples. We further illustrate the procedures on the Baby Heart Rate data (Figure 1) and provide finally a further application of our methodology, a new test of stationarity of a given time series. This is also illustrated by means of application to the Tremor data (Figure 2).

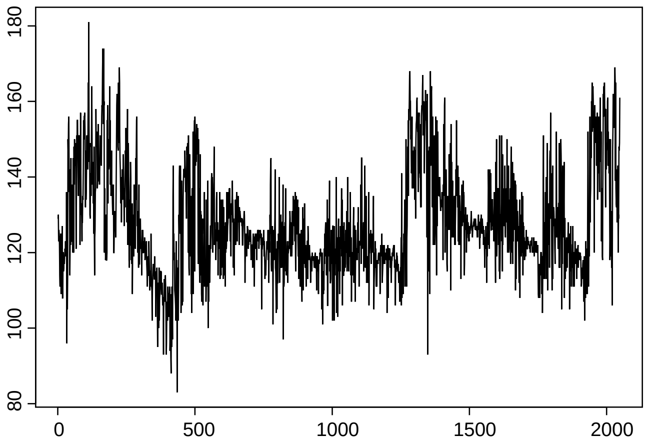



Fig. 1. ECG recording of a 66-day-old infant. Series is sampled at 1/16 Hz and is recorded from 21:17:59 to 06:27:18, $T = 2048$ observations. (Data courtesy Institute of Child Health, Royal Hospital for Sick Children, Bristol, and Guy P. Nason)

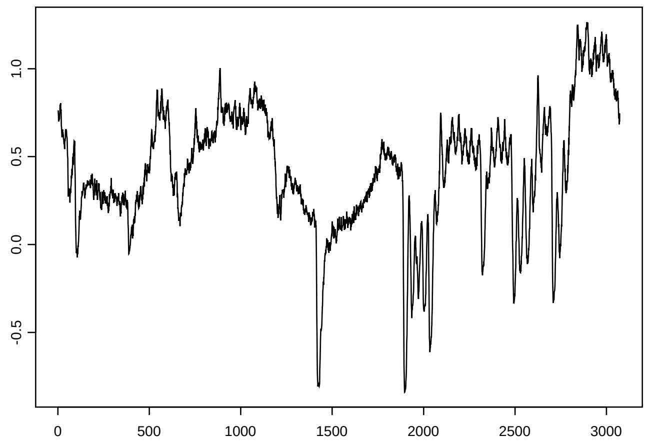



Fig. 2. Tremor data, $T = 3072$ data (Data courtesy Cognitive Neuroscience Laboratory of the University of Quebec, Anne Beuter and Roderick Edwards).

### 1.2. *The LSW model*

We now cite the definition of our model from VBvS[18].

**Definition 1.1.** A sequence of doubly-indexed stochastic processes $X_{t,T}$ ($t = 0, \ldots, T-1, T > 0$) with mean zero is in the class of locally stationary wavelet processes (LSW processes) if there exists a representation in the mean-square sense

$$X_{t,T} = \sum_{j=-\infty}^{-1} \sum_{k=0}^{T-1} w_{jk;T} \; \psi_{jk}(t) \; \xi_{jk}, \tag{1.1}$$

where $\{\psi_{jk}(t) = \psi_{j0}(t-k)\}_{jk}$ with $j < 0$ is a discrete non-decimated family of wavelets based on a mother wavelet $\psi(t)$ of compact support, and such that:

(1) $\xi_{jk}$ is a Gaussian orthonormal increment sequence with $\mathrm{E}\xi_{jk} = 0$ and $\mathrm{Cov}\,(\xi_{jk}, \xi_{\ell m}) = \delta_{j\ell}\;\delta_{km}$ for all $j, \ell, k, m$, where $\delta_{j\ell} = 1$ if $j = \ell$ and 0 if not;

(2) For each $j = -1, -2, -3, \ldots$, there exists a function $W_j(z)$ on $(0,1)$ possessing the following properties:

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

(b) There exists a sequence of constants $C_j$ such that for each $T$

$$\sup_{k=0,\ldots,T-1} \left| w_{jk;T} - W_j\left(\frac{k}{T}\right) \right| \leqslant \frac{C_j}{T} \;, \tag{1.2}$$

(c) $W_j^2(z)$ is bounded by $L_j$ in the total variation norm, i.e.

$$\mathrm{TV}_{[0,1]}\left(W_j^2\right) \leqslant L_j, \tag{1.3}$$

where

$$\mathrm{TV}_{[0,1]}(f) := \sup \left\{ \sum_{i=1}^{I} |f(a_i) - f(a_{i-1})| \;\; : 0 < a_0 < \ldots < a_I < 1, \; I \in \mathbb{N} \right\} \;.$$

(d) The constants $C_j$ and $L_j$ are such that

$$\sum_{j=-\infty}^{-1} N_j (N_j L_j + C_j) \leqslant \rho < \infty \tag{1.4}$$

where $N_j = |\,\mathrm{supp}\,\psi_{j0}| = (2^{-j} - 1)(N_{-1} - 1) + 1$.

LSW processes use wavelets to decompose a stochastic process with respect to an orthogonal increment process in the time-scale plane. (Gaussianity of the process is only assumed here for sake of convenience for the subsequent theory of testing.) Moreover, NvSK[14] and VBvS[18] developped a theory which ensures the existence of a unique wavelet spectrum. This property is a consequence of the local stationarity setting which introduces a *rescaled time* $z = t/T \in (0,1)$ on which $W_j(z)$ is defined. The rescaled time permits increasing amounts of data about the local structure of $W_j(z)$ to be collected as the observed time $T$ tends to infinity. Even though

LSW processes are not uniquely determined by the sequence $\{w_{jk;T}\}$, the model allows to identify (asymptotically) the model coefficients determined by uniquely defined $W_j(z)$. Then, the *evolutionary wavelet spectrum* (EWS) of an LSW process $\{X_{t,T}\}_{t=0,\ldots,T-1}$, with respect to $\psi$, is given by

$$S_j(z) = |W_j(z)|^2\,,\ z \in (0,1) \tag{1.5}$$

and is such that, by definition of the process, $S_j(z) = \lim_{T\to\infty} |w_{j,[zT];T}|^2$ for all $z \in (0,1)$, and by Definition 1.1, $\sum_{j=-\infty}^{-1} S_j(z) < \infty$ uniformly in $z \in (0,1)$.

The evolutionary wavelet spectrum $S_j(z)$ is related to the time-depending autocorrelation function of the LSW process. Observe that the autocovariance function of an LSW process can be written as

$$c_{X,T}\,(z,\tau) = \mathrm{Cov}\left(X_{[zT],T}, X_{[zT]+\tau,T}\right)$$

for $z \in (0,1)$ and $\tau$ in $\mathbb{Z}$, and where $[\,\cdot\,]$ denotes the integer part of a real number. The next result shows that this autocovariance converges asymptotically to a *local autocovariance* defined by

$$c_X\,(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j\,(\tau) \tag{1.7}$$

where $\Psi_j(\tau)$ is the autocorrelation wavelet function defined as follows.

**Definition 1.2.** Set

$$\Psi_j(\tau) = \sum_{k=-\infty}^{\infty} \psi_{jk}(0)\psi_{jk}(\tau)\ ,$$

where $\tau \in \mathbb{Z}$ and $j = -1,-2,-3,\ldots$. The function $\Psi_j$ is called the *discrete autocorrelation wavelet function at scale* $j$ (ACW in short).

Obviously the ACW inherits localisation properties from wavelets. However, it is symmetric about $\tau = 0$, that is $\Psi_j(\tau) = \Psi_j(-\tau)$ for all scales $j$ and for all $\tau$.

We now observe (from VBvS[18]) the announced property.

**Proposition 1.1.** *Under the assumptions of Definition 1.1, if* $T \to \infty$

$$\sum_{\tau=-\infty}^{\infty} \int_0^1 dz\ \left|c_{X,T}\,(z,\tau) - c_X\,(z,\tau)\right| = O\left(T^{-1}\right)$$

*for all LSW process.*

We further observe that equation (1.7) is a multiscale decomposition of the autocovariance structure of the process over time: The larger the wavelet spectrum $S_j(z)$ is at a particular scale $j$ and point $z$ in the rescaled time, the more dominant is the contribution of scale $j$ in the variance at time $z$. Thus, the evolutionary wavelet spectrum describes the distribution of the (co)variance at a particular scale and time location.

We finally note that an LWS process is a (second-order) stationary process if its EWS $S_j(z)$ does not depend on time $z$, for all scales $j$. We will exploit this crucial property in our development of a test of stationarity in Section 3.2 further below.

## 2. The adaptive estimation procedure

In this section we first recall from NvSK[14] a preliminary estimator of the EWS which serves as the basis to construct the adaptive estimator of VBvS[18]. It is basically a wavelet scalogram (called "wavelet periodogram" here, because of its statistical similarities to the classical Fourier-based periodogram). This preliminary estimator, based on the non-decimated and hence highly redundant wavelet transform, does not only need a bias-correction, but it is not a consistent estimator and needs to be smoothed (scale by scale) as a function of time. We explain how to use the theory of Lepski and Spokoiny[8] to adaptively smooth the wavelet periodogram by some local average over an interval in time ("histogram") using as key result an exponential inequality on this histogram. The basic idea is to test on homogeneity along this interval in time, increasing its length as long as the test does not reject the null hypothesis of homogeneity over time. As genuinely original contribution of this paper, we derive the concrete adaptive procedure and apply our algorithm to estimate the EWS of some simulated examples. The following Section 3 derives, as applications of this adaptive procedure, the test of significance and the test of stationarity, the second part of our newly developped material.

### 2.1. *Wavelet periodogram*

The wavelet periodogram $I_{j;T}(z)$, a preliminary estimator of the EWS, is constructed by taking the squared empirical coefficients from the non-decimated transform of our process data $X_{t,T},\ t = 1, \ldots, T$. As the EWS, it is defined, for each fixed scale $j$, as a function of rescaled time:

$$I_{j;T}\left(\frac{k}{T}\right) = \left(\sum_{t=0}^{T-1} X_{t,T}\psi_{jk}(t)\right)^2 \qquad j = -1, \ldots, -[\log_2 T]; k = 0, \ldots, T-1 \ .$$

Some asymptotic properties of this estimator have been studied by NvSK[14] who showed that the wavelet periodogram is *not* an asymptotic unbiased estimator of the wavelet spectrum. Indeed, Proposition 4 of NvSK[14] states that, for all fixed scales $j < 0$,

$$\mathrm{E}[I_{\ell;T}(z)] \ = \sum_{\ell=-[\log_2 T]}^{-1} A_{j\ell}\ S_j(z) \ + \ O\left(T^{-1}\right) \ , \qquad (2.11)$$

uniformly in $z \in (0,1)$, where the matrix $A$ is the Gram matrix of inner products of the autocorrelation wavelets:

$$A_{j\ell} = \sum_{\tau} \Psi_j(\tau)\Psi_\ell(\tau). \qquad (2.12)$$

Equation (2.11) motivates the definition of a *corrected wavelet periodogram* (CWP)

$$L_{j;T}\left(\frac{k}{T}\right) = \sum_{\ell=-[\log_2 T]}^{-1} A_{j\ell}^{-1} \left(\sum_{t=0}^{T-1} X_{t,T}\psi_{\ell k}(t)\right)^2 \tag{2.13}$$

as a preliminary estimator of the EWS. However, this CWP is not a consistent estimator of the EWS (its asymptotic variance is proportional to $S_j^2(z)$) and needs (pointwise) smoothing over time.

**Remark 2.1.** Invertibility of this matrix $A$ has been established in NvSK[14] when $\{\Psi_j\}$ is constructed using Haar or Shannon wavelets. If other compactly supported wavelets are used, numerical results suggest that the invertibility of $A$ still holds, but a complete proof of this result has not been established yet.

### 2.2. *The key result*

In order to estimate $S_j(z_0)$ at a fixed point $z_0$, we proceed as follows. Our construction of local averages over time of the CWP $L_{j;T}(z)$ for a fixed scale $j$ is a histogram over a given segment in time $\mathcal{R} = (s_1, s_2) \subseteq (0,1)$ containing the fixed time point $z_0$:

$$Q_{j,\mathcal{R};T} = |\mathcal{R}T|^{-1} \sum_{k\in\mathcal{R}T} L_{j;T}\left(\frac{k}{T}\right) \ , \tag{2.14}$$

where $k \in \mathcal{R}T$ means $k/T \in \mathcal{R}$. Obviously this is a natural estimator of the analogous theoretical quantity, the averaged wavelet spectrum

$$Q_{j,\mathcal{R}} = |\mathcal{R}|^{-1} \int_{\mathcal{R}} dz\ S_j(z) \ . \tag{2.15}$$

Some important statistical properties of $Q_{j,\mathcal{R};T}$ have been studied in VBvS[18] under a set of assumptions that we recall now.

**(A.1)** The autocovariance function $c_{X,T}$ and the local autocovariance function $c_X$ of the LSW process are such that

$$\|c_{X,T}\|_{1,\infty} \ := \sum_{\tau=-\infty}^{\infty} \sup_{t=0,\dots,T-1} \left|c_{X,T}\left(\frac{t}{T},\tau\right)\right|$$

is uniformly bounded in $T$, and

$$\|c_X\|_{1,\infty} := \sum_{\tau=-\infty}^{\infty} \sup_{z\in(0,1)} |c_X(z,\tau)| < \infty.$$

**(A.2)** There exists $\varepsilon > 0$ such that, for all $z \in (0,1)$, $\sum_{j=-\infty}^{-1} S_j(z) \geqslant \varepsilon$.

**(A.3)** The evolutionary wavelet spectrum $S_j(z)$ defined in (1.5) is such that

$$\sum_{\ell=-\infty}^{-[\log_2(T)]-1} \sup_{z\in(0,1)} S_\ell(z) = O\left(T^{-1}\right).$$

Assumption A.1 guarantees uniform absolute summability of the autocovariance of the process (which is a classical short memory property). Moreover, as the sum over scales of $S_j(z)$ is the local variance of the process at time $[zT]$, Assumption A.2 imposes that the local variance of the process is nowhere zero. The last assumption A.3 is necessary in order to control the difference between the EWS and the CWP at lower scales. Recall that our definition of the LSW processes involves the scales $j = -1$ up to $-\infty$, while the CWP is defined for scales $j = -1$ to $j = -[\log T]$ only.

Under these assumptions, VBvS[18] show that $Q_{j,\mathcal{R};T}$ is an asymptotically unbiased and consistent estimator of $Q_{j,\mathcal{R}}$. Obviously, in order to estimate the EWS $S_j(z_0)$ at the point $z_0$, the question arises how to choose the best interval $\mathcal{R}$ around $z_0$ in (2.14). We choose this interval by a method, related to the pointwise adaptive estimation theory of Lepski[7], such that the EWS at scale $j$ is nearly constant around $z_0$. In other words, the method aims to find an interval around $z_0$ where the EWS is *homogeneous*. We recall this well-known concept of homogeneity according to Lepski[7], calling $S_j(z)$ homogeneous on $\mathcal{R}$ if

$$b(\mathcal{R}) := \sup_{z \in \mathcal{R}} |S_j(z) - Q_{j,\mathcal{R}}| \ \leq \ C_j \ \sigma_{j,\mathcal{R},T} \ \log_2^2(T) \ ,$$

with some constant $C_j$ and with $\sigma^2_{j,\mathcal{R},T}$ denoting the variance of $Q_{j,\mathcal{R};T}$. This inequality expresses that, on $\mathcal{R}$, we cannot detect a variation of the EWS that is smaller than the natural variation of the estimator multiplied by $\log_2^2 T$. The need of such a log term is well-known from the theory of Lepski[7] on pointwise adaptive estimation in the iid setting.

Now we precise the basic idea of the procedure for concretely choosing $\mathcal{R}$ as follows. Suppose that $S_j(z_0)$ is well approximated by the averaged spectrum $Q_{j,\mathcal{U}}$ for a given interval $\mathcal{U}$ containing the reference point $z_0$. We then consider a second interval $\mathcal{R}$ containing $\mathcal{U}$ and we test if $Q_{j,\mathcal{R}}$ differs significantly from $Q_{j,\mathcal{U}}$. If so, then we reject the hypothesis of homogeneity of the wavelet spectrum $S_j(z)$ on $z \in \mathcal{R}$. Finally, the adaptive estimator corresponds to the largest interval $\mathcal{R}$ such that the hypothesis of homogeneity is not rejected.

A detailed description of the algorithm is presented in the next section. As we explained above, the key point of the procedure is to test if $Q_{j,\mathcal{R}}$ differs significantly from $Q_{j,\mathcal{U}}$, where $\mathcal{R}$ is an interval containing $z_0$ and $\mathcal{U}$ is a subset of $\mathcal{R}$. Let us now explain how we can test this hypothesis.

The standardised test statistic we use is

$$\mathcal{T}_T(\mathcal{R},\mathcal{U}) := \frac{|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}|}{\log_2^2(T)\sqrt{\mathrm{Var}\,|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}|}} \ .$$

Again, observe that the standardisation of the test statistic contains a $\log_2^2(T)$ term. This is in accordance with the standard theory of pointwise adaptive estimation in the iid setting, from which it is known that the loss of a minimax estimator, being normalised, will not be asymptotically degenerated[2,7].

The key result for the estimation gives an upper bound for the deviation of this statistic. Under the hypothesis of homogeneity, i.e. when the difference $|Q_{j,\mathcal{R}} - Q_{j,\mathcal{U}}|$

is small, we can show (VBvS[18]) that

$$\Pr(\mathcal{T}_T(\mathcal{R},\mathcal{U}) > \eta) \leqslant g(\mathcal{R},\mathcal{U},\eta)\ , \tag{2.21}$$

where $g$ is given by

$$g(\mathcal{R},\mathcal{U},\eta) := \exp\left\{ -\frac{1}{8} \cdot \frac{\eta^2 k_T^2}{1 + c \cdot 2^{j/2} \eta k_T (|\mathcal{R}|^{-1} + |\mathcal{U}|^{-1}) \sigma_{\mathcal{R},\mathcal{U}}^{-1} T^{-1/2})} \right\}$$

with $k_T = \log_2^2 T$ and $\sigma_{\mathcal{R},\mathcal{U}} = \sqrt{\mathrm{Var}\,|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}|}$. In the definition of $g$, $c$ is a constant depending on $\|c_X\|_{1,\infty}$, on $\rho$ (given in Definition 1.1) and on the wavelet $\psi$.

This key result allows to test the homogeneity of the spectrum, as explained in the next section.

### 2.3. *The adaptive procedure*

In this section, we give the algorithm which selects the interval of homogeneity around a given point $z_0$, at a fixed scale $j < 0$.

We first choose a set $\Lambda$ of interval-candidates $\mathcal{R}$. Then, for each candidate $\mathcal{R}$, we apply the homogeneity test with respect to a given set $\wp(\mathcal{R})$ of subintervals $\mathcal{U}$ of $\mathcal{R}$.

***Initialization.*** Select the smallest interval $\mathcal{R}$ in $\Lambda$.

***Iteration.*** Select the next interval $\mathcal{R}$ and calculate the corresponding estimate $\mathcal{T}_T(\mathcal{R},\mathcal{U})$. Calculate an estimator and the estimated variance $\tilde{\sigma}^2_{\mathcal{R},\mathcal{U},T}$ of its variance $\sigma^2_{\mathcal{R},\mathcal{U}}$, following the description of the next subsection 2.4.

***Testing homogeneity.*** Reject $\mathcal{R}$ if there exists one $\mathcal{U} \in \wp(\mathcal{R})$ such that

$$g\left(\mathcal{R},\mathcal{U},\mathcal{T}_T(\mathcal{R},\mathcal{U})\right) > g_0\ .$$

***Loop.*** If $\mathcal{R}$ is not rejected, then iterate using a larger interval. Otherwise, select the latest non rejected interval.

This procedure requires the preselection of the sets $\Lambda$ and $\wp(\mathcal{R})$, but also the choice of a constant $g_0$. These are discussed in the following remarks.

**Remark 2.2 (Choice of $\Lambda$ and $\wp(\mathcal{R})$).** Several propositions have been proposed in the literature for choosing these two sets $\Lambda$ and $\wp(\mathcal{R})$ [11,17]. In our computations, we use the following sets. For each scale $j < 0$, the CWP (2.13) is evaluated on a grid $k/T$, $k = 0, \dots, T-1$ in time. We define the set $\mathcal{K} = \{iT/K, T : i = 0, \dots, (K-1)\}$ which depends on a constant $K$, the choice of which we discuss below. Then we choose the set $\Lambda$ as

$$\Lambda = \{[r_0/T, r_1/T] : r_0, r_1 \in \mathcal{K} \text{ and } r_0 < [z_0 T] < r_1\}\,.$$

Next, for every interval $\mathcal{R} = [r_m/T, r_n/T]$ in $\Lambda$, we define the set $\wp(\mathcal{R})$ of subintervals $\mathcal{U}$ by taking all smaller subintervals $[r_k/T, r_n/T]$ with the right end point $r_n/T$ and similarly all smaller intervals $[r_m/T, r_\ell/T]$ with the left end-point $r_m/T$:

$$\wp(\mathcal{R}) = \{\mathcal{U} = [r_k/T, r_n/T] \text{ or } \mathcal{U} = [r_m/T, r_\ell/T] : m < \ell, k < n\}$$

if $\mathcal{R} = [r_m/T, r_n/T]$.

**Remark 2.3 (Choice of $g_0$ and $K$).** This procedure requires the choice of a constant $g_0$ (specifying the level of the test), as well as the parameter $K$ which defines the set of subintervals in the test of homogeneity. Note that these two parameters $K$ and $g_0$ are global, in the sense that they do not depend on the time point $z_0$ where we are testing the homogeneity. They no longer depend on $\mathcal{R}$ or $\mathcal{U}$, and they are fixed by the model only. They can be chosen following some preliminary study of the CWP. In a different context, Mercurio and Spokoiny[11] propose to select these nuisance parameters by minimisation of the mean square prediction error. This could also be possible in our context, since the prediction theory of LSW processes has been recently developped[4].

### 2.4. *Preliminary estimation of the variance*

The computation of the test statistics $\mathcal{T}_T$ needs the preliminary estimation of the unknown variance $\sigma^2_{\mathcal{R},\mathcal{U}} = \mathrm{Var}\,|Q_{j,\mathcal{R};T} - Q_{j,\mathcal{U};T}|$. A straightforward computation shows that this variance may be decomposed as follows:

$$\sigma^2_{\mathcal{R},\mathcal{U}} = 2\|U'_{\mathcal{R},\mathcal{U}}\Sigma_T\|_2^2$$

where $\|\cdot\|_2$ denotes the Euclidean norm of a matrix, $U_{\mathcal{R},\mathcal{U}}$ is a known matrix depending only on the intervals $\mathcal{R},\mathcal{U}$, scale $j$ and wavelet $\psi$, and $\Sigma_T$ is the covariance matrix of the LSW process $(X_{0,T},\ldots,X_{T-1,T})'$. This covariance matrix is of course unknown in practice since it depends on the unknown EWS. If $\sigma_{s,s+u}$ denotes the entry $(s, s+u)$ of the matrix $\Sigma_T$, VBvS[18] have proposed to estimate $\sigma_{s,s+u}$ by the plug-in estimator

$$\tilde{\sigma}_{s,s+u} = \sum_{j=-[\log_2 T]}^{-1} Q_{j,\mathcal{R}_T(s);T}\Psi_j(u)\mathbb{I}_{|u|\leqslant M_T} \tag{2.27}$$

where $\mathcal{R}_T(s)$ denotes an interval $\mathcal{R}_T$ that contains the time point $s/T$. The length of this interval depends on $T$ such that it shrinks to zero when $T$ tends to infinity. An appropriate theoretical rate is $|\mathcal{R}_T| \sim \log_2^{-3} T$. In the estimator (2.27), we note also that the indicator $\mathbb{I}\{|u| \leqslant M_T\}$ sets to zero all $\tilde{\sigma}_{s,s+u}$ with $|u| \geqslant M_T$. This is in accordance with the short-memory property stated in Assumption (A.1) above. An appropriate rate for $M_T$ is $\log_2^{\alpha} T$ with $\alpha > 0$.

VBvS[18] showed that the convergence in probability of the resulting estimator $\tilde{\sigma}^2_{\mathcal{R},\mathcal{U},T}$ is fast enough to ensure that the exponential inequality (2.21) is a correct approximation which can be used to construct the test of homogeneity. However, the theoretical rates proposed in VBvS[18] for $|\mathcal{R}_T|$ and $M_T$ are derived from an asymptotic theory, and they are not directly useful for practical purposes.

As a matter of fact, an intensive simulation study shows that, even if the choice of $|\mathcal{R}_T|$ and $M_T$ is important for the preliminary estimation of the covariance matrix $\Sigma_T$, their impact on the quality of the test of homogeneity is limited. In practice, we recommend to set $|\mathcal{R}_T| = [\log_2 T]$. In our applications, we use the following rule for the choice of $M_T$. We start with a large value for $M_T$ (around 10) and examine the decreasing (or increasing) of the off-diagonals of the covariance matrix. Very often, in our applications, the behaviour of these off-diagonals starts with an abrupt decreasing, sometimes followed by an increasing trend. When this behaviour is observed in several off-diagonals of the matrix, we recommand to clip the matrix just before the abrupt decreasing. This procedure has been followed for choosing $M_T$ in the numerical studies below.

We end this section by mentioning that the pre-estimation of the variance is not computationaly expensive. The key point here is that we only need to estimate the covariance matrix $\Sigma_T$ by $\tilde{\Sigma}_T$ of the whole process and to store it. Then the estimated variance of $\mathcal{T}_T(\mathcal{R},\mathcal{U})$ is obtained by computing $2\|U'_{\mathcal{R},\mathcal{U}}\tilde{\Sigma}_T\|_2^2$, where the matrix $U_{\mathcal{R},\mathcal{U}}$ is purely deterministic.

An exhaustive study of the impact of the preliminary estimator on the results of the test is out of the scope of this article, but can be found in Van Bellegem[19].

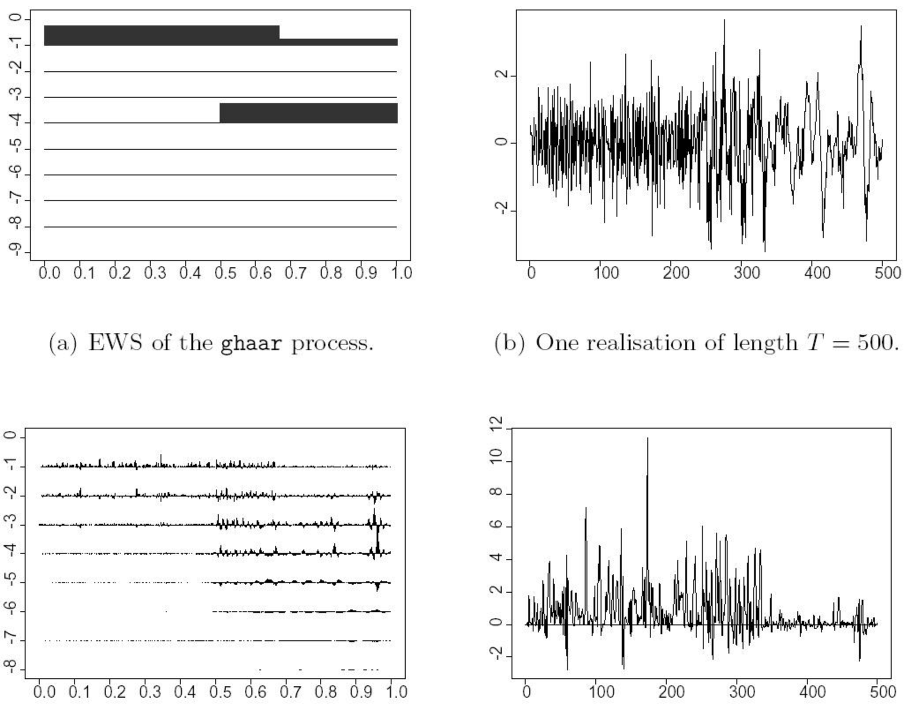
### 2.5. *Simulated example*

To illustrate the pointwise adaptive estimation procedure, we consider a simulated process of length $T = 500$ with theoretical spectrum given in Figure 3(a). This process has two active scales on $j = -1$ and $-4$. The scale $j = -4$ is active only from the middle of the time series, while the scale $j = -1$ is active at each time point, but with a breakpoint occuring at time point $2T/3$.

In the following, we have applied the pointwise estimator on 19 equidistant time points to furnish the estimation on (a grid of) the whole scale $j = -1$. In this situation it may happen that some homogeneity tests are computed several times for the same intervals $\mathcal{R}$ and $\mathcal{U}$, e.g. if we estimate the EWS at two points included in one true interval of homogeneity. For computational efficiency we store the results of all tests of homogeneity for each $\mathcal{R}$ and $\mathcal{U}$, such that each test is computed only once.

The results presented in Figure 4 correspond to the following choices of parameters: $K = 20$, $g_0 = 0.4$. In this simulation, these two parameters are chosen by hand, but the result is robust to this choice. Note that here we estimate the variances using (2.27) by taking $|\mathcal{R}_T| = [\log_2(500)]$ and $M_T = 1$. The above figure plots the true EWS at scale $j = -1$ (solid line) together with our estimators at 19 equidistant time points (dots). The constancy of the EWS on $(0, 2/3)$ and $(2/3, 1)$ is well detected by the estimator, except at one time point near the breakpoint. If we consider the CWP at scale $j = -1$, see Figure 3(d), one may observe that the CWP has a short period with lower amplitudes before the time break. We believe that the single bad pointwise estimator comes from this phenomenon. Once again, it is important to note that the CWP is a highly variable quantity, and our estimation is provided with only 500 data.

In the estimation plotted in Figure 4, all the pointwise estimators near the value




(a) EWS of the `ghaar` process. (b) One realisation of length $T = 500$.

(c) CWP of the `ghaar` process. (d) Scale $j = -1$ of the CWP.

Fig. 3. The `ghaar` process uses the spectrum in (a) with nondecimated Haar wavelets and Gaussian increments in the Definition of the locally stationary wavelet process.

1 of the EWS on $(0, 2/3)$ select the same interval of homogeneity $\tilde{\mathcal{R}}$, as expected. This interval is showed together with the original time series at the bottom of Figure 4. Similarly, the estimators near 0.25, the value of the EWS constant on $(2/3, 1)$, select the same interval of homogeneity, which is plotted in the second plot.

## 3. Further applications of the adaptive procedure

### 3.1. *Test of local significance*

As motivated in our real examples below, it is of interest to be able to decide, perhaps even before estimating the EWS, for the regions in the time-scale where the EWS needs to be modelled to be different from zero. This can be done by the test of homogeneity by choosing as null hypothesis $Q_{j,\mathcal{R}} = 0$ for the specific interval $\mathcal{R}$ on the specific scale $j$ of interest.

The null hypothesis of this test is

$$\mathrm{H}_0 : \mathrm{E}(Q_{j,\mathcal{R};T}) = 0 \quad \text{for a fixed scale } j < 0 \text{ and for all } z \in \mathcal{R}.$$

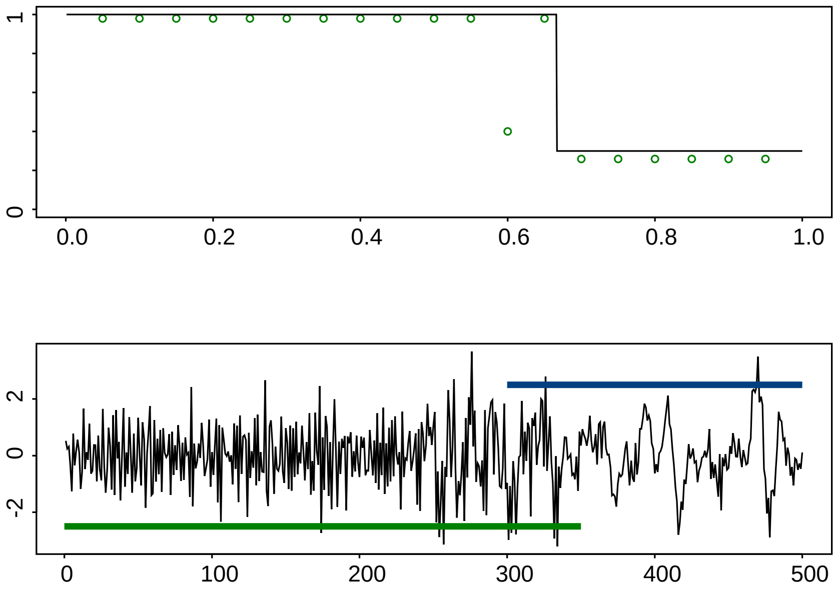



Fig. 4. Above is the result of the pointwise adaptive estimator at scale $j = -1$ for the `ghaar` process (the solid line is the true spectrum, and the dots are the pointwise adaptive estimators). The estimation is obtained from the realisation of the process plotted at the bottom. The two horizontal lines on the time series correspond to some intervals of homogeneity selected by the adaptive procedure (see text).

The test statistic here is

$$\mathcal{Q}_T(\mathcal{R}) = \frac{Q_{j,\mathcal{R};T}}{\sqrt{\operatorname{Var} Q_{j,\mathcal{R};T}}}$$

and the test is now be based on the inequality

$$\Pr\left(\mathcal{Q}_T(\mathcal{R}) > \eta \middle| \mathrm{H}_0\right) \leqslant h(\mathcal{R}, \eta)$$

where

$$h(\mathcal{R}, \eta) := \exp\left\{-\frac{1}{8} \cdot \frac{\eta^2}{1 + c \cdot 2^{j/2}\eta 2^{j/2}|\mathcal{R}|^{-1}(T \operatorname{Var} Q_{j,\mathcal{R};T})^{-1/2}}\right\} \tag{3.31}$$

which is derived similarly to the exponential inequality (2.21).

### 3.2. *Test of stationarity*

A final application of the previous theory is given by a new test of covariance stationarity for time series.

A covariance stationary process is characterised by an EWS which is constant over time, i.e. $S_j(z) = S_j$ for each scale $j$ (see Section 1.2). The key idea for the test of stationarity is to test if the EWS is constant over time at each scale or not. With this idea, testing the stationarity of a time series is equivalement to test the homogeneity of the EWS at each scale over $\mathcal{R} = (0, 1)$. This procedure is illustrated in the real data example of Section 4.2.

## 4. Application on real data

### 4.1. *Baby heart rate*

In this study of a first real data example, we illustrate the possibility to combine the proposed test of local significance and the pointwise adaptive estimator although they are originally devoted to different statistical problems. The key idea is to test the significance of some whole scales over the whole time, before performing the estimation procedure on the scales which are significantly different from zero.

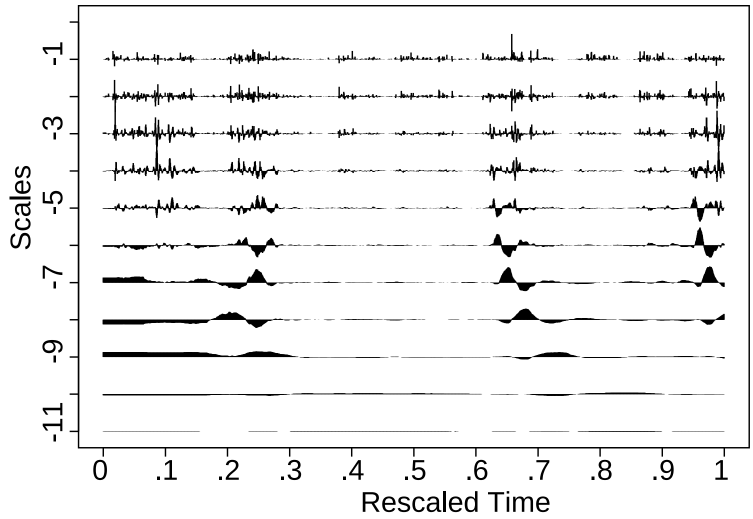



Fig. 5. Corrected wavelet periodogram (CWP) of the data, computed with Haar wavelets.

Our study concerns a heart rate (electrocardiogram (ECG)) recording of a 66-day-old infant. Figure 1 plots the series, sampled at 1/16 Hz and recorded from 21:17:59 to 06:27:18 ($T = 2048$ observations). This series is considered in previous studies[13,14] as a motivating example for the exploratory analysis using the LSW model and may be obtained from the web[12]. First of all, it is unlikely that this time series will be a stationary time series. The heart rate varies considerably over time and changes significantly between periods of sleep and waking. These changes are of a big interest for the paediatricians. These are related to other variables of interest which are not easily observable nor easy to quantify. For instance, the paediatricians are interessed to measure the sleep states (quiet, active, awake,...) using some objective measures, and the question is how the heart rate may be used as a tool for measuring the sleep state. We shall come back later to this question of the link between the ECG and the sleep states.

The approach using the LSW model leads to a multiscale representation of the nonstationary process. Figure 5 shows the CWP of the heart rate, computed with nondecimated Haar wavelets. The CWP is not smoothed and highly variable, and our goal now is to extract some useful information from it.

As a first step, we need a pre-estimator of $\Sigma_T$ and, for this, need to choose an appropriate parameter $M_T$ and segment $\mathcal{R}_T$. From the conclusions of Section 2.4 above, we choose $\mathcal{R}_T(z)$ centered in $z$ and of length $[\log_2 T]$. The selection of $M_T$ is provided as described in Section 2.4. We first compute $\tilde{\Sigma}_T$ with $M_T = 10$ and then analyse the behaviour of its off-diagonals. Figure 6 shows the values of 10 different off-diagonals, that is we superimpose $\tilde{\sigma}_{s,s+u}$ for $u = 0, \ldots, 10$ and for 10 different $s$. This shows a similar behaviour between the off-diagonals, which decrease quickly to $M_T = 2$ then vary slowly. We then choose $M_T = 2$ in the pre-estimation of $\tilde{\Sigma}_T$.

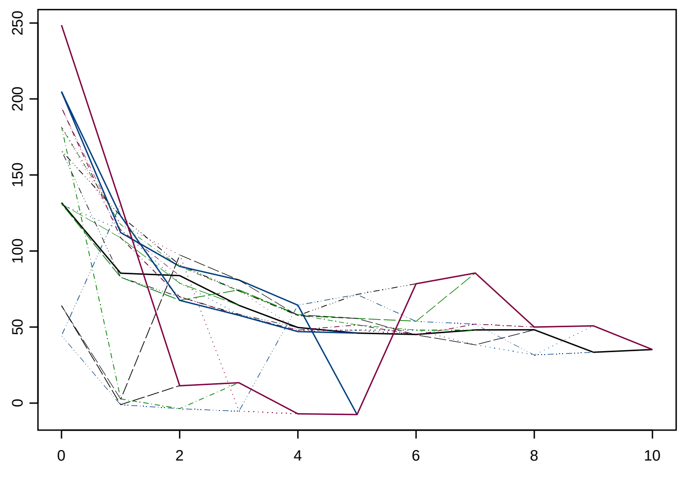


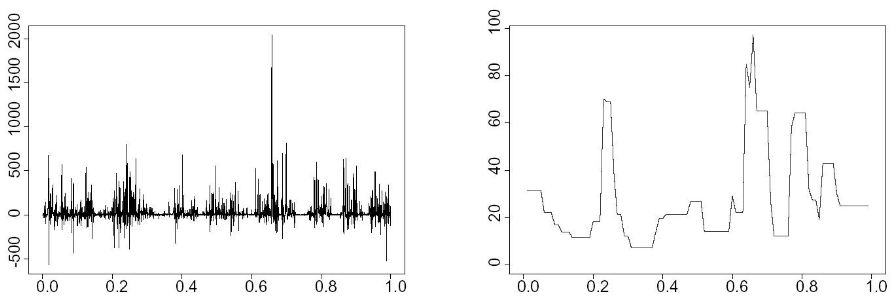

Fig. 6. Ten off-diagonals of the estimated matrix $\tilde{\Sigma}_T$.

To start the analysis, we want to detect if some scales of the CWP are not

| Scale | $Q_{j,\mathcal{R};T}$ | $\tilde{\sigma}^2_{j,\mathcal{R};T}$ | approximated $p$-value |
|---|---|---|---|
| -1 | 31.66 | 12.01 | $1.68 \cdot 10^{-4}$ |
| -2 | 9.71 | 22.50 | 0.60 |
| -3 | 9.31 | 101.98 | 0.90 |
| -4 | 19.69 | 104.83 | 0.63 |
| -5 | 17.25 | 66.95 | 0.57 |
| -6 | 20.14 | 35.90 | 0.24 |
| -7 | 44.66 | 18.08 | $1.44 \cdot 10^{-6}$ |
| -8 | 3.67 | 8.79 | 0.83 |
| -9 | 33.53 | 4.07 | $5.77 \cdot 10^{-15}$ |
| -10 | 17.56 | 1.69 | $5.66 \cdot 10^{-10}$ |

Table 1. Results of the test of sparsity over $\mathcal{R} = (0, 1)$ performed at each scale $j$ between $-1$ and $-10$ for the heart rate data.

significant. For this, we apply our test of significance over $\mathcal{R} = (0, 1)$ at each scale. The results of the test are given in Table 1. From this table, we conclude that the only active scales of the data are given by $j = -1, -7, -9$ and $-10$, and the other scales are not significantly different from zero. To our knowledge, such conclusion is new for these data, and also very helpful since it indicates that the analysis should focus on 4 active scales only.




(a) Scale $j = -1$ of the CWP. (b) The estimator ($j = -1$).

Fig. 7. Pointwise adaptive estimator performed at scale $j = -1$ for the baby ECG. The estimator is computed at 100 different points, and we line up two consecutive points. (a) is the scale $-1$ of the CWP, and (b) is the estimator based on (a).

We now focus on the significant scale $j = -1$ and apply our estimation procedure. The results are given in Figure 7. In our estimation, we estimate the EWS at $K = 100$ points. In the estimation procedure, we also set $g_0 = 0.95$, which is quite large. In terms of homogeneity tests, this means that we perhaps reject the homogeneity assumption very often. In our opinion, this is in fact very sensible, because this error (Type I error) is not as serious as the complementary type II error. Indeed, for the pointwise estimation, it is not such a problem if the homogeneity interval is too small, since the corresponding histogram will not differ too much from the histogram based on a possible larger truly homogeneous interval. In contrast, if we choose a too large interval, for instance if we choose an interval which contains some discontinuity, then the estimator will be significantly different. Once again, it could be possible to select $g_0$ automatically by minimisation of a criterion such as the mean square prediction error.

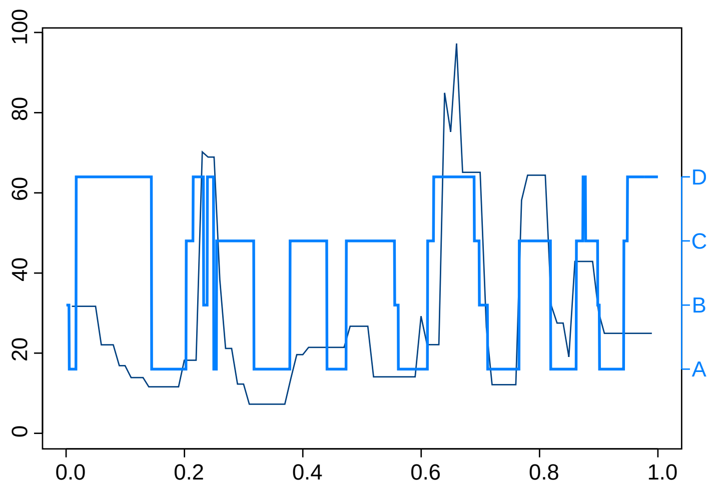



Fig. 8. Pointwise adaptive estimator of the EWS at scale $j = -1$ together with the sleep states (A, quiet sleep; B, state between A and C; C, active sleep; D, awake).

Simultaneously to the ECG recording, some experts on the analysis of brain-waves and eye movements recorded the sleep states of the infant. These states are recorded independently of the ECG. They are plotted in Figure 8 together with our estimator of the EWS at scale $j = -1$. The observers classifies the sleep states as quiet sleep (A), between quite and active sleep (B), active sleep (C) and awake (D).

It is clear that there is some relationship between our estimator and the sleep states. In particular, periods of activity occur whilst the estimate of $S_{-1}(z)$ is large, and periods of quiet sleep when it is small. It is worth mentioning that the ECG is easy to measure, while the sleep states is more tricky and less objective. With our estimator, we are also able to detect some changes in the sleep states, sometimes with a small delay. Then, we believe that our estimator may help to provide some objective measurement of the activity during sleep. Finally, we mention that a finer analysis of the fitting between the sleep states and our adaptive estimator is also one possible way to derive an automatic choice of the parameter $g_j$.

## 4.2. *Tremor data*

The aim of this last example is to apply our proposed test of stationarity to the following data example.

The data shown in Figure 2 are the first 3072 observations of a set of tremor data. The object of the study is to compare different regions of tremor activity coming from a subject with Parkinson's disease. These data have been considered by von Sachs and Neumann[16] (hereafter vSN[16]) who apply their test of stationarity over three consecutive segments of length 1024 of the first-order differenced series, shown in Figure 9. As in vSN[16], we have added a Gaussian white noise of standard deviation 0.01 to the original data, in order to break the discrete nature of the data.

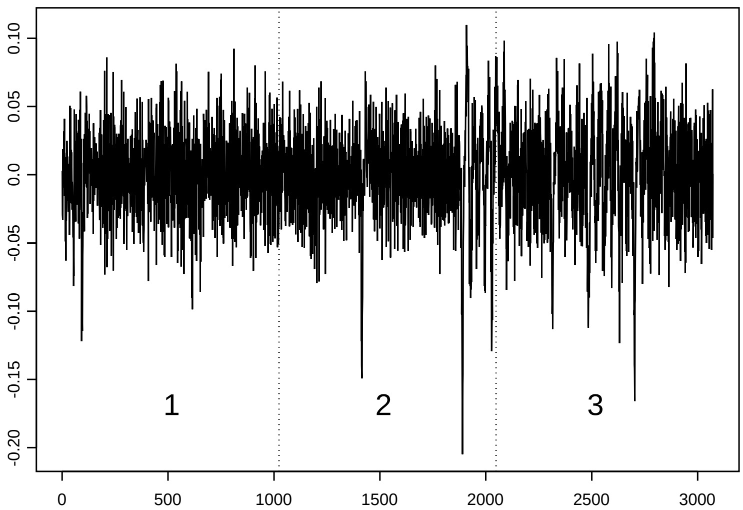



Fig. 9. First-order difference of the tremor data.

We apply the test of homogeneity over $\mathcal{R} = (0,1)$ scale by scale for the three segments (1), (2) and (3). The parameters for the pre-estimation of the variance are $|\mathcal{R}_T| = \log_2(1024)$ and $M_T = 2$. For each scale, we test the homogeneity between the EWS on $(0,1)$ and on 20 subintervals. Table 2 reports the results. The number

| Scale | (1) | (2) | (3) |
|---|---|---|---|
| -1 | 0.08 | 0.39 | 0.05 (*) |
| -2 | 0.35 | 0.44 | 0.46 |
| -3 | 0.74 | 0.42 | 0.68 |
| -4 | 0.34 | 0.001 (**) | 0.28 |
| -5 | 0.07 | $3 \cdot 10^{-21}$ (**) | 0.002 (**) |
| -6 | 0.54 | $8 \cdot 10^{-5}$ (**) | 0.24 |
| -7 | 0.72 | 0.02 (*) | 0.60 |
| -8 | 0.91 | 0.84 | 0.81 |
| -9 | 0.92 | 0.94 | 0.88 |
| -10 | 0.98 | 0.98 | 0.97 |

Table 2. Results of the test of stationarity for the three segments taken from the tremor data. (1), (2) and (3) refer to the first, second and third segment in Figure 9. At each scale, for each segment, we perform 20 tests of homogeneity between $\mathcal{R} = (0, 1)$ and 20 subintervals $\mathcal{U}$. The number reported in the table is the minimum probability value obtained among the 20 tests. This value is computed using $g$ in (2.21). (*) indicates a value less than or equal to 0.05 and (**) indicates a value less than 0.01.

reported in the table is the minimum probability value obtained between the 20 tests. This value is computed using $g$ in (2.21). (*) indicates a value less than or equal to 0.05 and (**) indicates a value less than 0.01.

The conclusion of this study is a lack of stationarity for segments (2) and (3) of the tremor data. The test of vSN[16] concludes also to a lack of stationarity for series (2). However, they do not detect any change of regime in the series (3), and our conclusion seems to be a new observation. A careful inspection of the time series shows that some changes of regime indeed occur in segment (2), and also in segment (3) (around the time point 2300). This is also in accordance with the findings of the neurologists who attributed two different regimes of tremor activity for this part of data. Our conclusion is that (at least) one change of regime occurs in segment (2) as well as in segment (3).

The difference with the conclusion of vSN[16] may be explained by Table 2. Indeed, the lack of stationarity for segment (3) is due to an inhomogeneity at scale $-5$ only (and perhaps also at scale $-1$). This is certainly a very subtle behaviour to be detected, and our multiscale approach succeeded to find this lack of homogeneity.

Our analysis offers a more precise interpretation of the nonstationarity of the tremor data. Moreover, we would like to recall that, unlike the test of vSN[16], our approach is not limited to time series with a length equal to a power of 2. Furthermore, it is of course possible with our method to detect the underlying intervals of homogeneity in the tremor data, such that we can say exactly where the changes of regime occur in the time series.

## 5. Conclusion

In this paper we have developped a series of new algorithms for various problems of adaptive estimation and tests, for a particular class of non-stationary processes. These Locally Stationary Wavelet processes have been shown useful to combine two goals: controlling the deviation from second-order stationarity, and giving a meaningful time-scale decomposition for a stochastic process (and its autocovariance) which is potentially globally non-stationary or just along certain predominant scales.

Our proposed methods derive all from the idea of "testing homogeneity", along certain intervals of certain scales of the process decomposition. The test of local significance is essentially a test to decide for significance of a (squared) wavelet coefficient in the local autocovariance representation. Consecutively, we can apply our proposed adaptive estimator for the wavelet spectrum on only those regions in the time-scale plane where there is significant contribution. Another application is testing for stationarity of the underlying process: scale by scale, testing whether the EWS depends on time, and globally whether there is one scale for which the EWS would not be constant over time. Furthermore, we can use our method to precisely identify some possible changes of regime (breakpoints) in the time series.

We have applied our suggested procedures to a variety of simulated and two real data examples. We found out that it is of high utility to have theoretically well understood methods to finer analyse non-stationary, transient behaviour in the correlation structure of the observations. Often there are physiological explanations (such as the sleep state activities for the heart rate data) for these time-changing phenomena being concentrated on certain predominant scales.

We believe that further research might be useful, e.g. on possibilities of better controlling the problem of multiple hypothesis testing. From our numerical examples it has become evident that more recent statistical methods, such as controlling the False Discovery Rate[1], might be promising further developments of our proposals.

## References

1. Y. Benjamini and Y. Hochberg, *Controlling the False Discovery Rate: A practical and powerful approach to multiple testing*, *J. Roy. Statist. Soc. Ser. B* **57** (1995) 289–300.
2. L. D. Brown and M. G. Low, *A constrained risk inequality with applications to non-parametric functional estimation*, *Ann. Statist.* **24** (1996) 2524–2535.
3. R. Dahlhaus, *Fitting time series models to nonstationary processes*, *Ann. Statist.* **25** (1997) 1–37.
4. P. Fryźlewicz, S. Van Bellegem and R. von Sachs, *Forecasting non-stationary time series by wavelet process modelling*, *Ann. Inst. Statist. Math.* (2003) To appear.
5. M. Hallin, *Mixed autoregressive moving average multivariate processes with time-dependent coefficients*, *J. Multivariate Anal.* **8** (1978) 567–572.
6. S. Kawasaki and R. Shibata, *Weak stationary of a time series with wavelet representation*, *Japan J. Industr. Appl. Math.* **12** (1995) 37–45.
7. O. Lepski, *On a problem of adaptive estimation in Gaussian White noise*, *Theory of Probab. Appl.* **35** (1990) 454–470.

8. O. Lepski and V. Spokoiny, *Optimal pointwise adaptive methods in nonparametric estimation*, *Ann. Statist.* **25** (1997) 2512–2546.
9. S. Mallat, G. Papanicolaou and Z. Zhang, *Adaptive covariance estimation of locally stationary processes*, *Ann. Statist.* **26** (1998) 1–47.
10. G. Mélard and A. Herteleer-De Schutter, *Contributions to the evolutionary spectral theory*, *J. Time Ser. Anal.* **10** (1989) 41–63.
11. D. Mercurio and V. Spokoiny, *Statistical inference for time-inhomogeneous volatility models*, *Ann. Statist.* **32** (2004) To appear.
12. G. P. Nason, *WaveThresh3 Software*, Department of mathematics, University of Bristol, Bristol, UK (1998).
13. G. P. Nason and R. von Sachs, *Wavelets in time series analysis*, *Phil. Trans. R. Soc. Lond. A* **357** (1999) 2511–2526.
14. G. P. Nason, R. von Sachs and G. Kroisandt, *Wavelet processes and adaptive estimation of evolutionary wavelet spectra*, *J. Roy. Statist. Soc. Ser. B* **62** (2000) 271–292.
15. H-S. Oh, C. M. Ammann, P. Naveau, D. Nychka and B. L. Otto-Bliesner, *Multiresolution time series analysis applied to solar irradiance and climate reconstructions*, *J. Atmos. Solar-Terr. Phys.* **65** (2003) 191–201.
16. R. von Sachs and M. H. Neumann, *A wavelet-based test for stationarity*, *J. Time Ser. Anal.* **21** (2000) 597–613.
17. V. Spokoiny, *Estimation of a function with discontinuities via local polynomial fit with an adaptive choice*, *Ann. Statist.* **26** (1998) 1356–1378.
18. S. Van Bellegem and R. von Sachs, *Locally adaptive estimation of sparse evolutionary wavelet spectra*, Discussion Paper # 0310, Institut de statistique, Univ. catholique de Louvain (2003).
19. S. Van Bellegem, *Adaptive methods for modelling, estimating and forecasting locally stationary processes*, PhD Thesis, Institut de statistique, Univ. catholique de Louvain (2003).
20. A. Woyte, R. Melmans and J. Nijs, *Fluctuations in instantaneous clearness index: Analysis and statistics*, Preprint, Katholieke Universiteit Leuven (2003).