Wavelets in time series analysis

By Guy P. Nason and Rainer von Sachs
Department of Mathematics, University of Bristol
University Walk, Bristol BS8 1TW, UK
and
Institute of Statistics, Catholic University of Louvain,
Louvain-la-Neuve, Belgium.

This article reviews the role of wavelets in statistical time series analysis. We survey work that emphasises scale such as estimation of variance and the scale exponent of a process with a specific scale behaviour such as 1/f processes. We present some of our own work on locally stationary wavelet (LSW) processes which model both stationary and some kinds of non-stationary processes. Analysis of time series assuming the LSW model permits identification of an evolutionary wavelet spectrum (EWS) that quantifies the variation in a time series over a particular scale and at a particular time. We address estimation of the EWS and show how our methodology reveals phenomena of interest in an infant electrocardiogram series.

Keywords: Allan variance, locally stationary time series, long-memory processes, time-scale analysis, wavelet processes, wavelet spectrum.

1. Introduction

Reviewing the role of wavelets in statistical time series analysis (TSA) appears to be quite an impossible task. For one thing, wavelets have become so popular that such a review never could be exhaustive. Another reason, more pertinent, is that there is no such thing as *one* statistical time series *analysis* as the very many different fields encompassed by TSA are in fact so different that the choice of a particular methodology must naturally vary from area to area. Examples for this are numerous: think about the fundamentally different goals of treating comparatively short correlated biomedical time series to explain, for example, the impact of a set of explanatory variables on a response variable, or of analysing huge inhomogeneous data sets in sound, speech or electrical engineering, or finally, building models for a better understanding and possibly prediction of financial time series data.

Hence, here we can only touch upon some aspects of where and why it can be advantageous to use wavelet methods in some areas of statistical TSA. We consider the common situation where the trend and (co-)variation of autocorrelated data are to be modelled and estimated. Classically, the time series data are assumed to be *stationary*: their characterising quantities behave homogeneously over time. Later in this paper, we shall consider situations where some controlled deviation from stationarity is allowed.

We stress that this article does not provide an exhaustive review of this area. In particular we shall not go into detail concerning the time series aspects of wavelet denoising, nor the role of wavelets in *deterministic* time series but briefly summarise these next. Later on, we shall concentrate on seeing how wavelets can provide information about the *scale* behaviour of a series: i.e. what is happening at different scales in the series. This type of analysis, as opposed to a frequency analysis, can be appropriate for certain types of time series as well as provide interpretable and insightful information which is easy to convey to the non-specialist.

Wavelet denoising. In this volume Silverman (1999) and Johnstone (1999) describe the method and rationale behind *wavelet shrinkage* a general technique for curve denoising problems. The specific connection with the analysis of correlated time series data $\{Y_t\}$ is that in a general regression-like model

$$Y_t \;=\; m(X_t) \;+ \sigma(X_t)\,\epsilon_t \qquad t = 1, \ldots, T,$$

nonparametric estimation of the trend function m and/or the function σ, which measures the variability of the data, can be performed in exactly the same framework of non–linear wavelet shrinkage as for the original simple situation of Gaussian i.i.d. data (see Silverman, 1999). In typical biomedical applications the errors (and hence the Y_t themselves) need neither be uncorrelated nor even stationary (for one example, in the case of a deterministic equidistant design $\{X_t\} = \{t/T\}$ see von Sachs and MacGibbon, 1998.) In financial time series analysis, where the choice of $\{X_t\} = \{(Y_{t-1}, \ldots, Y_{t-p})\}$ leads to a nonparametric autoregressive model (of order p), the task is estimation of a conditional mean and variance under the assumption of *heteroscedastic* stationary errors (see, e.g., Hoffmann (1999), for the case $p = 1$). Both examples, but particularly the second one, call for *localised* methods of estimation. Problems of this sort typically show regimes of comparatively smooth behaviour, which from time to time may be disrupted by break points or other discontinuities. A subordinate task would then be to detect and estimate the precise location of these break points and here wavelet methods again prove useful. Finally, for this little review, we mention that wavelet shrinkage can be used for the estimation of spectral densities of stationary time series (Neumann, 1996; Gao, 1997) and of time–varying spectra using localised periodogram based estimators (von Sachs and Schneider, 1996; Neumann and von Sachs, 1997).

Deterministic time series. A completely different use of wavelets in statistical time series analysis was motivated by how wavelets originally entered the field of (deterministic) time–frequency analysis (see, e.g., Rioul and Vetterli, 1991, or Flandrin, 1998). Here wavelets, being *time–scale* representation methods, deliver a tool complementary to both classical and localised (e.g. windowed) Fourier analyses. We will focus next on this aspect for *stochastic* signals.

Stochastic time series. Recently, the search for localised time–scale representations of stochastic correlated signals lead to both analysing and synthesising (i.e., modelling) mainly *non-stationary* processes with wavelets or wavelet-like bases. By non-stationarity we mean here two different types of deviation from stationarity: first, but not foremost, we will address wavelet analyses of certain long–memory processes which show a specific (global or local) scale behaviour in Section 2 including the well-known $1/f$ (or power law) processes. However, our main emphasis is on signals with a possibly time–changing probability distribution or characterising quantities. Overall, from the modelling point of view, wavelets offer clear advantages mainly for the two types of non-stationary data mentioned above.

Statistical time–scale analysis performs a statistical wavelet *spectral analysis* of time series analogously to the classical *Fourier* spectral analysis where the *second–order* structure of the series such as variance, autocovariance (dependence structure within the series) and spectral densities are to be estimated (see Priestley, 1981). The analyses depend fundamentally on processes possessing a representation in terms of random coefficients with respect to some localising basis – in classical spectral analysis: the Fourier basis, which is perfectly localised in frequency; in time–scale analysis: a basis which is localised in time and scale. Then, the respective second-order quantities of interest (e.g. variance or autocovariance) can be represented by a superposition of the (Fourier or the wavelet) spectra. Our specific model, in Section 3(b) uses a particular set of basis functions: discrete non-decimated wavelets.

However, before turning to scale *and* time localised models the next section reviews the basic ideas about using wavelets for the *analysis* of statistical phenomena with characteristic behaviour living on certain global scales. For example, a variance decomposition on a scale by scale basis has considerable appeal to scientists who think about physical phenomena in terms of variation operating over a range of different scales (a classical example being $1/f$-processes). We will review a very simple example to demonstrate how insights from *analysis* can be used to derive models for the *synthesis* of stochastic processes.

Both in analysis and synthesis it is, of course, possible to *localise* these scale specific phenomena in time as well. For the specific example of $1/f$-processes we refer to Gonçalvès and Flandrin (1993) and more recent work of, e.g., Wang *et al.* (1997) and the overview on wavelet analysis, estimation and synthesis of scaling data by Abry *et al.* (1999). The common paradigm of all these approaches is the separation of the scale on which the process data are sampled from the scale(s) where the respective behaviour is observed.

2. Estimation of "global" scale behaviour

This section gives examples on estimating the scale characteristics of processes which do not show a location dependency. In fact, we restrict our discussion to the utility of wavelets for the analysis and synthesis of long-memory processes. We begin with a brief introduction to scalograms using the example of the *Allan variance* originally developed by Allan (1966) as a time domain measure of frequency stability in high frequency oscillators (McCoy and Walden, 1996). Then we turn, more specifically to the problem of estimation of the scale exponent of $1/f$-processes, in the specific case of fractional Brownian motion.

(*a*) *Long-Memory processes, Allan and wavelet variance*

For pedagogical reasons we will concentrate on Percival and Guttorp (1994), one of the earlier papers in the vast literature in this field, and also, for a simple exposition, we will only concentrate on the Haar wavelet although other wavelets may be equally, if not more, useful.

Consider a stretch of length T of a given zero mean stochastic process $\{X_t\}_{t\in Z}$.

The *Allan variance* $\sigma_X^2(\tau)$ at a particular scale $\tau \in Z$ is a measure of how averages, over windows of length τ, change from one period to the next. If $\overline{X}_t(\tau) = 1/\tau \sum_{n=0}^{\tau-1} X_{t-n}$, then

$$\sigma_X^2(\tau) \;:=\; 1/2 \;\mathrm{E}\;(|\overline{X}_t(\tau) - \overline{X}_{t-\tau}(\tau)|^2). \tag{2.1}$$

In order to have a meaningful quantity independent of time t, X_t must be stationary or at least have stationary increments of order 1.

In fact, the Allan variance turns out to be proportional to the Haar wavelet variance, another "scale variance" based on the standard discrete Haar wavelet transform as follows. Let $\{\hat{d}_{jk}\}$ denote the (empirical) wavelet coefficients of the signal $\{X_t\}_{t=0,\ldots,T-1}$ where we deviate from standard notation in that our scales $j = -1, \ldots, -\log_2(T)$ become more negative the coarser the level of the transform, and hence the location index k runs from 0 to $T/2^{-j} - 1$. This proportionality can easily be observed by writing the Haar coefficients as successive averages of the data with filters $1/\sqrt{2}$ and $-1/\sqrt{2}$ (which are the high-pass filter coefficients $\{g_k\}$ of the Haar transform, see Silverman, 1999). For example, for $j = -1$ at scale $\tau_j = 2^{-j-1} = 1$,

$$\hat{d}_{-1,k} \;=\; (1/\sqrt{2})\;(X_{2k+1} - X_{2k})\;, \quad k = 0, \ldots, T/2 - 1\;. \tag{2.2}$$

and it is easy to see that $\mathrm{var}\{\hat{d}_{-1,k}\} = \sigma_X^2(1)$. More generally,

$$\mathrm{var}\{\hat{d}_{jk}\} \;=\; \mathrm{E}\;\hat{d}_{jk}^2 \;=\; \tau_j\;\sigma_X^2(\tau_j)\;. \tag{2.3}$$

Motivated by this last equation, an unbiased estimator for the Allan variance, the so-called "non-overlapped" estimator, is the appropriately normalised sum of the squared wavelet coefficients

$$\widehat{\sigma}_X^2(\tau_j) \;:=\; 2/T \sum_{k=0}^{T/2^{-j}-1} \hat{d}_{jk}^2\;. \tag{2.4}$$

This estimator has the property that each data point X_t contributes to exactly one coefficient $\hat{d}_{jk}$. Again only considering the finest scale, $j = -1$, formula (2.4) can be written in terms of the data as

$$\widehat{\sigma}_X^2(1) = 1/T \sum_{k=0}^{T/2-1} (X_{2k+1} - X_{2k})^2\;.$$

We observe immediately that one can improve upon the above estimator by summing over not just $T/2$ values of these time series differences but over all $T-1$ possible ones. The resulting estimator clearly will have a smaller variance and also possesses independence with respect to the choice of the origin of the series $\{X_t\}$. This "maximal-overlap" estimator, denoted by $\widetilde{\sigma}_X^2(\tau_j)$, is based on the "non-decimated" (Haar) wavelet transform (NDWT), with wavelet coefficients $\widetilde{d}_{jk}$ (see appendix for a description). The NDWT amounts to a uniform sampling in k instead of the inclusion of subsampling or decimation in each step of the standard discrete (decimated) wavelet transform (DWT)

Figure 1 shows an estimator for the Allan variance of the MA(3) process $X_t^{(2)} =$

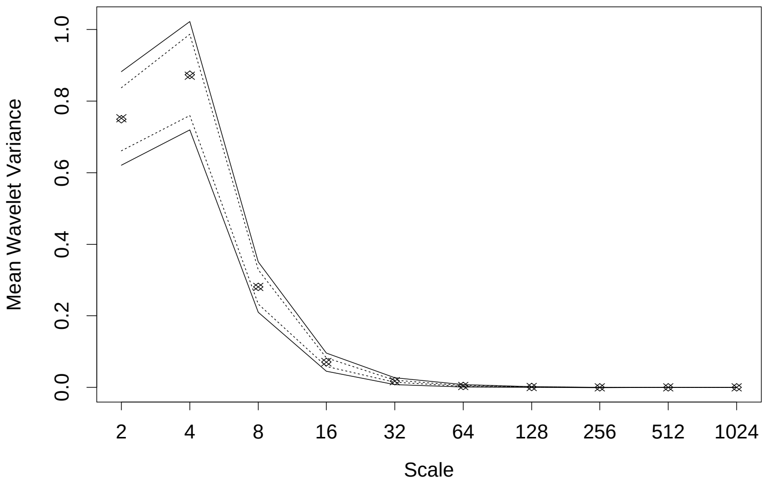


Figure 1. Mean of estimated Allan variances over 1000 simulations of the MA(3) process computed using Haar wavelet variances (see text). The mean of the estimates is shown by the symbols, the mean plus or minus twice the standard deviation of the estimate is shown by lines. For the "non-overlap" estimator the symbols are $\times$ with solid error lines; for the "maximal-overlap" estimator the symbols are $\diamond$ and the error lines are dotted. The theoretical values of the Allan variance at scales 2, 4 and 8 are $3/4, 7/8$ and $36/128$ respectively.

$(\varepsilon_t + \varepsilon_{t-1} - \varepsilon_{t-2} - \varepsilon_{t-3})/2$, with standard Gaussian white noise ε_t, using both the "non-overlap" and "maximal-overlap" estimators. This MA process is one of a class which we will meet in Section 3. The figure was produced by simulating 1024 observations from the process and computing the estimated Allan variance. This simulation procedure was repeated 1000 times and the figure actually shows the mean of all the estimated Allan variances with lines showing the accuracy of the estimates. It is clear that the "maximal-overlap" estimator has a smaller variance for a wide range of scales τ_j. It is clear from the formula of the MA(3) process that it "operates" over scale 4 and thus the Allan variance at scale 4 ($j = -2$) is largest in figure 1. Further, the Allan variance indicates that there is variation at scale 2 — this is not surprising because the process formula clearly links quantities over that scale as well. However, at scales 8 and larger the process has insignificant variation and so the Allan variance decays for large scales. Using (2.3) and the orthogonality of the DWT it is easy to see that the Allan variance for white noise is $\sigma^2(\tau_j) = 1/\tau_j = 2^{j+1}$, $j < 0$.

More general wavelets could be used in place of Haar in the wavelet variance estimators given above. In any case the use of the NDWT will be beneficial as is clear from the example above.

Why is the concept of a "scale variance" useful at all? The "scale variance" permits a new decomposition of the process variance which is different (but related) to the classical spectral decomposition using the (Fourier) spectral density. That is, suppose we have a stationary process $\{X_t\}$, then we can decompose its total variance $\mathrm{var}\{X_t\}$ into quantities which measure the fluctuation separately

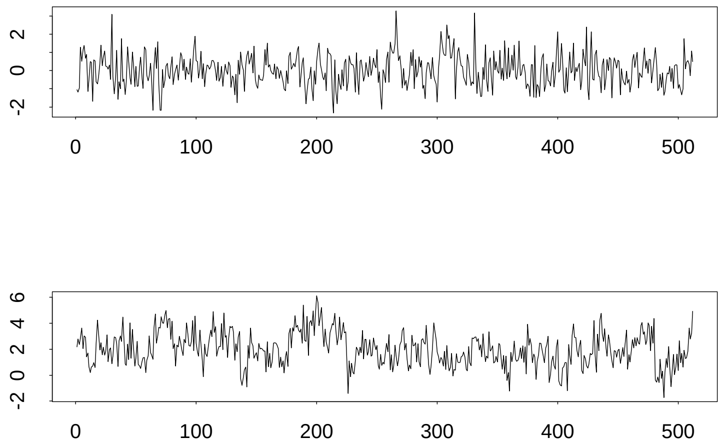


Figure 2. Top: realization of a fGn process with very little "long-memory". Bottom: realization of a fGn process with evident long-memory. In the lower picture you can "see" a "slow oscillation" underlying the process.

scale by scale:

$$\mathrm{var}\{X_t\} = 1/2 \sum_{j=-\infty}^{-1} \sigma_X^2(\tau_j) = \sum_j \mathrm{var}\{\widetilde{d}_{jk}\}/2\tau_j. \tag{2.5}$$

Here, on the right hand side the scale-dependent quantities play the role of a "wavelet spectrum" $S_j := \mathrm{var}\{\widetilde{d}_{jk}\}/2\tau_j$ which in the stationary case is independent of time k. For white noise $S_j = 2^j$. We will define a time-dependent wavelet spectrum in (3.4) for more general nonstationary processes in Section 3.

The Allan (or wavelet) variance is of additional interest for the study of long-memory processes. These are processes where the autocorrelation of the process decays at a very slow rate such that it is possible for effects to persist over long time scales (see Beran, 1994). Consider a 'fractional Gaussian noise' (fGn) process with self-similarity parameter $1/2 < H < 1$, or more precisely the first-order increments $\{Y_t\}$ of a fractional Brownian motion (fBm) $B = B_H$ (with $B_0 = 0$). This process is characterized by having normally distributed stationary and self-similar increments $Y_t := (B_{s+t} - B_s)/t \sim (B_t - B_0)/t \sim |t|^{H-1}\, B_1 \sim N(0, \sigma^2 |t|^{2H-2})$ (see, e.g., Abry *et al.*, 1995). Figure 2 shows sample paths for two different simulated fGn processes. Try and guess which one has the long-memory before you study the caption!

It can be shown that the Allan variance of $\{Y_t\}$ follows a power law, i.e.

$$\sigma_Y^2(\tau) = L(\tau)\ |\tau|^{2H-2}\ , |\tau| >> t_0\ ,$$

where $L(\cdot)$ is a slowly varying function for $\tau \to \infty$. Hence, a plot of $\log\{\sigma_Y^2(\tau)\}$ versus $\log(\tau)$ (or more precisely a least-squares or maximum-likelihood estimate of this log-linear relationship based on one of the estimators for σ_Y^2) can reveal an estimator of the parameter H, and in general, for large enough τ one can observe to a good approximation a line with slope $2H - 2$. Here we now see that it can be useful to use wavelets other than Haar when investigating a potential relationship because different wavelets cover slightly different frequency ranges and possess differing phase behaviour.

Some example applications can be found in Percival and Guttorp (1994) on the analysis of the scales of variation within vertical ocean shear measurements which obey a power law process over certain ranges. Serroukh, Walden and Percival (1998) investigate the wavelet variance of the surface albedo of pack ice which happens to be a strongly non-Gaussian series. This paper also derives statistical properties of the estimators used for obtaining confidence intervals.

(*b*) *Wavelet spectral analysis of* $1/f$*-processes*

We now slightly change our point of view, and instead of variances over certain scales we now examine the more general quantity: the spectral density or spectrum. The spectrum is the Fourier transform of the autocovariance function of a stationary process. For its proper definition in case of the (non-stationary) fBm we again must use the fractional Gaussian noise (fGn) which has a power law spectral density as follows:

$$f_Y(\omega) \;=\; L_f(\omega)\; |\omega|^{1-2H} \;, 0 < |\omega| << t_0^{-1} \;,$$

where again, $L_f(.)$ is a slowly varying function, now for $\omega \to 0$. If $H > 1/2$, we observe a singularity in zero frequency which is again an indicator for a strongly correlated time series, i.e. one with long memory.

As before, from the statistical point of view, there is a linear relation between $\log\{f_Y(\omega)\}$ and $\log(|\omega|)$ for small enough $|\omega|$, which allows us to base an estimator for the spectral exponent $\alpha = 1 - 2H$ on an estimator for the spectrum f_Y. This will in fact be one of the estimators $\widehat{\sigma}_Y^2$ or $\widetilde{\sigma}_Y^2$ for the wavelet variance σ_Y^2 which is related to the spectrum f_Y by the equation (2.6) below. We summarize section 2 of Abry, Gonçalvès and Flandrin (1995) who motivate why wavelet is superior to traditional Fourier spectral analysis for power law processes. The two methods may be compared by examining the expectation of the wavelet variance estimator $\widehat{\sigma}_Y^2(\tau_j)$ and the expectation of the average of short-time Fourier periodograms over segments of equal length of $\{Y_t\}$. The Fourier estimator is constructed in a similar way to the wavelet variance estimator except that Gabor-like basis functions (appropriately weighted exponentials) instead of wavelets are used. In other words the Fourier estimator is the time-marginal of a particular bilinear time–frequency distribution, the *spectrogram*, which is the squared modulus of a Gabor or short-time Fourier transform. In this context the wavelet variance estimator is the time marginal of the *scalogram*, i.e. the squared wavelet coefficients. The properties of the two estimators are different because of the way energy is distributed over the two different configurations of atoms (Fourier: rectangular equal area boxes centred at equispaced nodes; wavelets: the famous constant-Q tiling with centres located on the usual wavelet hierarchy).

The expectation of the time marginal of the spectrogram can be written as the convolution of the Fourier spectrum $f_Y(\omega)$ of $\{Y\}$ with the squared Fourier transform of the moving window. Similarly the expectation of the time marginal of the scalogram, i.e. of $\widehat{\sigma}_Y^2(\tau_j)$ (for reasons of simplicity we only refer to the inferior estimated based on the DWT), is the convolution of $f_Y(\omega)$ with the squared modulus of the Fourier transform $\widehat{\psi}_{jk}(\omega)$ of the wavelets in the DWT, i.e.

$$\sigma_Y^2(\tau_j) \;=\; \mathrm{E}\,\widehat{\sigma}_Y^2(\tau_j) = 2/T \sum_k \mathrm{E}\,\widehat{d}_{jk}^2 \;=\tau_j^{-1} \int_{-\pi}^{\pi} f_Y(\omega)\,|\widehat{\psi}_{jk}(\omega)|^2\,d\omega\,. \qquad (2.6)$$

Abry, Gonçalvès and Flandrin (1995) show foremost that in a log-log relationship of equation (2.6) the bias for estimating the spectral parameter α of f_Y, becomes frequency *independent* when using averaged scalograms instead of averaged spectrograms. This is a consequence of the fact that in the Fourier domain wavelets scale multiplicatively with respect to frequency, a property which the fixed window spectrograms do not enjoy. Further considerations in Abry, Gonçalvès and Flandrin (1995), such as those pertaining to the efficiency of the estimators, support the wavelet based approach for these processes.

Equation (2.6) also helps to further interpret the variation of the considered estimators of $\sigma_Y^2(\tau_j)$ in the example given in figure 1. As the spectrum $f_Y(\omega)$ of the MA(3) process of this example has its power concentrated near to high frequencies and as the variance of these estimators are approximately proportional to the square of their mean, it is clear from the integral in (2.6) that this variance increases with frequency, i.e. if we go to finer scales in the plots of figure 1.

Further examples for processes with such a singular power law behaviour near zero frequency can be found (see Flandrin, 1998), e.g., in the areas of atmospheric turbulence (e.g., Farge, 1999), hydrology, geophysical and financial data, and telecommunications traffic (Abry *et al.*, 1999), just to name few. Whitcher *et al.*, (1998) detect, test and estimate time series variance changes. Their work can identify the scale at which the change occurs. Generalizations of these kind of tests to time-varying autocovariance and spectral density for short-memory non-stationary processes can be found in von Sachs and Neumann (1998).

In Section 3 we discuss ideas of how to localise both the analysis and synthesis of the global scale behaviour discussed in this section.

(c) *Synthesis of long-memory processes using wavelets*

In the above considerations on *analysis* of $1/f$-processes we saw that wavelets form a key role. In reverse, it is not surprising that they are useful also for *synthesis*, i.e. in the theory of modelling $1/f$-processes (see again, Abry *et al.* 1999). Indeed, it is possible to, for example, simulate $1/f$-processes using wavelets (see Wornell and Oppenheim, 1993). One method for simulating fractional Gaussian noise (fGn) is given by McCoy and Walden (1996) as follows: (a) compute the variances, S_j, of the required fGn processes by integrating its spectrum over dyadic intervals $[-2^j, -2^{j-1}] \cup [2^{j-1}, 2^j]$; (b) for each scale j draw a sequence of 2^{-j} independent and identically distributed normal random variables d_{jk}; (c) apply the inverse DWT to the $\{d_{jk}\}$ coefficients to obtain an approximate realization of a $1/f$ process. Figure 2 shows two realizations from fGn processes using the McCoy and Walden (1996) methodology.

3. Wavelet processes: a particular time–scale model

(*a*) *Local stationarity*

Suppose, that we have a time series $\{X_t\}_{t \in \mathbb{Z}}$ and that we wish to estimate the variance $\sigma_t^2 = \mathrm{var}(X_t)$ over time. If the series is stationary then σ_t^2 will be constant and equal to σ^2 and we can use the usual sum of squared deviations from the mean estimator on a single stretch of the observed time series $X_1, \dots, X_T$. As more data become available (as T increases) the estimate of the variance $\hat{\sigma}^2$ improves. Alternatively, suppose that we know that the variance of the series changes at each time point t i.e. assume that $\mathrm{var}(X_t) = \sigma_t^2$ for all $t \in \mathbb{Z}$ where none of the σ_t^2 are the same. Here the series is non-stationary and we do not have much hope in obtaining a good estimate of σ_t^2 since the *only* information we can obtain about σ_t^2 comes from the single X_t. As a third alternative suppose the variance of a time series changes *slowly* as a function of time t. Then the variance around a particular time t^* could be estimated by pooling information from X_t close to t^*.

A similar situation occurs if the long-memory parameter H in the previous section changes over time, i.e. $H = H(t)$ then the Allan variance would also change over time

$$\sigma_Y^2(t, \tau) = \mathrm{E}\left(|\overline{Y}_t(\tau) - \overline{Y}_{t-\tau}(\tau)|^2\right) \sim \tau^{2H(t)-2}.$$

For a series with such a structure we would hopefully observe $\widetilde{d}_{j,k}^2 \approx \widetilde{d}_{j,k+1}^2$ for the NDWT coefficients. To estimate the Allan variance we would need to construct local averages of the data. In other words we would not sum over *all* empirical NDWT coefficients but perform adaptive averaging of the $\widetilde{d}_{jk}^2$ over k for fixed scale j.

Time series whose statistical properties are slowly varying over time are called *locally stationary*. Loosely speaking, if you examine them at close range they appear to be stationary and if you can collect enough data in their region of local stationarity then you can obtain sensible estimates for their statistical properties such as variance (or autocovariance or the frequency spectrum).

One possibility for modelling time series such as these is to *assume* that

$$\mathrm{var}\left(\widetilde{d}_{j,k}\right) \approx \mathrm{var}\left(\widetilde{d}_{j,k+1}\right) \tag{3.1}$$

so that we have some chance of identifying/estimating coefficients from one realization of a time series. More generally, an early idea for "local stationarity", due to Silverman (1957), proposes

$$\mathrm{cov}(X_t, X_s) = c(t, s) \approx m\left(\frac{s+t}{2}\right)\gamma(s-t) := m(k)\gamma(\tau), \tag{3.2}$$

where $k = (s+t)/2$ and $\tau = t - s$. This model says that the covariance behaves locally as a typical stationary autocovariance but then varies from place to place depending on k. The Silverman model reflects our own wavelet-specific model given in (3.5) except ours decomposes γ over scales using a wavelet-like basis. Other early important work in this area can be found in Page (1952) and Priestley (1965). More recently Dahlhaus (1997) introduced an interesting model which poses estimation of time series statistical properties (such as variance) as

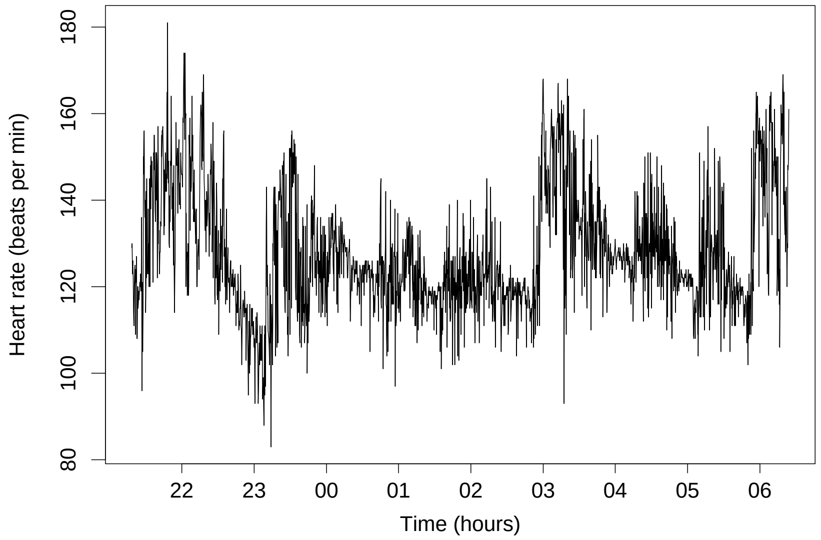


Figure 3. ECG recording of 66 day old infant. Series is sampled at $\frac{1}{16}$Hz and is recorded from 21:17:59 to 06:27:18. There are $T = 2048$ observations.

a curve estimation problem (which bestows great advantages when considering the performance of estimators because it assumes a *unique* spectrum).

Figure 3 shows an example of a time series that is not stationary. It shows the electrocardiogram (ECG) recording of a 66 day old infant. There are a number of interesting scientific and medical issues concerning such ECG data e.g. building and interpreting models between ECG and other covariates such as infant sleep state (see Nason, Sapatinas and Sawczenko 1999). However, for the purposes of this article we shall confine ourselves to examining how the variance of the series changes as a function of time and scale. Further analyses of this sort can be found in Nason, von Sachs and Kroisandt (1998).

(*b*) *Locally stationary wavelet processes*

(i) The processes model

A time-domain model for encapsulating localised scale activity was proposed by Nason, von Sachs and Kroisandt (1999). They define the *locally stationary wavelet* (LSW) process by

$$X_t = \sum_{j=-J}^{-1} \sum_{k \in \mathbb{Z}} w_{j,k}\, \psi_{jk}(t)\, \xi_{jk}, \quad \text{for } t = 0, \ldots, T-1, \tag{3.3}$$

where the $\{\xi_{jk}\}$ are mutually orthogonal zero mean random variables, the $\psi_{jk}(t)$ are discrete non-decimated wavelets (as described in the appendix) and the $w_{j,k}$ are amplitudes that quantify the energy contribution to the process at scales j and location k. Informally, the process X_t is built out of *wavelets with random amplitudes*. The LSW construction is similar in some ways to the well-known construction of stationary processes out of sinusoids with random amplitudes.

(ii) Evolutionary wavelet spectrum and local variance

To quantify how the size of $w_{j,k}$ changes over time we embed our model (3.3) into the Dahlhaus (1997) framework. To model locally stationary processes we use our assumption (3.1) to insist that $w_{j,k}^2 \approx w_{j,k+1}^2$ which forces $w_{j,k}^2$ to change slowly over k. Stationary processes can be included in this model by ensuring that $w_{j,k}^2$ is constant with respect to k. A convenient measure of the variation of $w_{j,k}^2$ is obtained by introducing rescaled time, $z \in (0,1)$ and defining the *evolutionary wavelet spectrum* (EWS) by

$$S_j(z) = S_j(k/T) \approx w_{j,k}^2 \tag{3.4}$$

for $k = 0, \ldots, T-1$. In the stationary case we lose the dependence on z (k) and obtain the "wavelet spectrum" or Allan variance given just after formula (2.5). One can see that as more time series observations are collected (as T increases) one obtains more information about $S_j(z)$ on a grid of values k/T which makes the estimation of $S_j(z)$ a standard statistical problem.

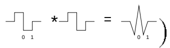
The important thing to remember about the EWS is

$\boxed{S_j(z) \text{ quantifies the contribution to process variance at scale } j \text{ and time } z.}$

In other words a large value of $S_j(z)$ indicates that there is a large amount of oscillatory power operating at scale j around location z. For examples of this see figures 4 and 5.

For a nonstationary process we would expect the variance of the process to vary over time and so we would expect our model to exhibit a time-localised version of (2.5), i.e. something like $\mathrm{var}(X_k) = \sum_j w_{j,k}^2$, or more precisely (3.7) below. If one takes our process model (3.3) and forms the autocovariance of X_t then one (asymptotically) obtains an expression in terms of the *autocorrelation* function of the *wavelets*. Let $c(z,\tau)$ define this *localised autocovariance*

$$\lim_{T\to\infty} \mathrm{cov}(X_{[zT]-\tau}, X_{[zT]+\tau}) = c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau), \tag{3.5}$$

where $\Psi_j(\tau)$ is the autocorrelation function of the discrete non-decimated wavelets defined by (for Haar )

$$\Psi_j(\tau) = \sum_{k=-\infty}^{\infty} \psi_{jk}(0)\psi_{jk}(\tau) \tag{3.6}$$

for $j < 0$. The representation in (3.3) is not unique because of the nature of the overdetermined non-decimated wavelet system. However, the autocovariance representation in (3.5) is unique. Relation (3.5) is reminiscent of the Silverman (1957) idea with $S_j(\cdot)$ playing the role of $m(\cdot)$ and $\Psi_j(\tau)$ the role of $\gamma(\tau)$ except that our model separates the behaviour over scales. Relation (3.5) also allows us to define the *localised variance* at time z by

$$v(z) = c(z,0) = \sum_{j=-\infty}^{-1} S_j(z) \tag{3.7}$$

i.e. the promised localised version of (2.5), note $\Psi_j(0)$ is always 1 for all scales j. For stationary series the localised autocovariance collapses to the usual autocovariance $c(\tau)$.

For estimation of the EWS we implement the idea of local averaging expressed in Section 3a. The EWS may be estimated by a corrected *wavelet periodogram* which is formed by taking the NDWT coefficients, $\widetilde{d}_{jk}$, of the sample realization, squaring them and then performing a correction to disentangle the effect of the overdetermined NDWT. Like the classical periodogram (see Priestley, 1981) the corrected wavelet periodogram is a noisy estimator of the EWS and needs to be smoothed to provide good estimates. The smoothing could be carried out by any number of methods but since we wish to be able to capture sharp changes in the EWS we adopt wavelet shrinkage techniques (Donoho *et al.* 1995, Coifman and Donoho 1995). Figure 5 shows a smoothed corrected wavelet periodogram for the infant ECG data. For further detailed analyses on this data set see Nason, von Sachs and Kroisandt (1998).

(iii) Motivating example: Haar MA processes

The MA(1) process $X_t^{(1)} = (\varepsilon_t - \varepsilon_{t-1})/\sqrt{2}$ is a LSW process as in (3.3) where the amplitudes are equal to 1 for $j = -1$ and zero otherwise and the constructing wavelets ψ_{jk} are Haar non-decimated wavelets as given in the appendix. The autocovariance function of $X_t^{(1)}$ is $c(\tau) = 1, -1/2, 0$ for $\tau = 0, \pm 1,$ otherwise and this c is precisely the finest scale Haar autocorrelation wavelet $\Psi_{-1}(\tau)$. So this special Haar MA process satisfies (3.5) with $S_{-j}(z) = 0$ for all $j < -1$ and $S_{-1}(z) = 1$ (and this agrees with (3.4) since the amplitudes are 1 for $j = -1$ only, and zero otherwise). We already met the MA process $X_t^{(2)}$ in section 2 (its Allan variance was plotted in figure 1). By the same argument its autocovariance function is this time the next finest scale Haar autocorrelation wavelet $\Psi_{-2}(\tau)$. Similarly, we can continue in this way defining the rth order Haar MA$(2^r - 1)$ process $X_t^{(r)}$ which has $\Psi_{-r}(\tau)$ for its autocovariance function for integers $r > 0$. Each of the Haar MA processes is stationary but we can construct a nonstationary process by concatenating the Haar MA processes. For example, suppose we take 128 observations from each of $X_t^{(1)}, X_t^{(2)}, X_t^{(3)}$ and $X_t^{(4)}$ and concatenate them (a realization from such a process is shown in Nason, von Sachs and Kroisandt, 1998). As a time series of 512 observations the process will not be stationary. The Haar MA processes have $S_{-j}(z) = 0$ for $-j \neq r$ and all z and $S_{-r}(z) = 1$ for $z \in ([r-1]/4, r/4)$ a plot of which appears in figure 4 (remember $z = k/512$ is rescaled time). The plot clearly shows that from time 1 to 128 the Haar MA(1) process is active with variation active at scale -1 (scale $2^{-j} = 2$), then at time 128 the MA(1) process, $X_t^{(1)}$, changes to the MA(3) process, $X_t^{(2)}$, until time 256 and so on.

Indeed, the section from 128 to 256 ($z \in (1/4, 1/2)$) should be compared to figure 1 which shows the Allan variance of $X_t^{(2)}$ which would be equivalent to averaging over the 128 to 256 time period. However, the EWS plot above does not show any power at scales $j = -1$ ($\tau_j = 2$) and $j = -3$ ($\tau_j = 8$) unlike the Allan variance plot. The absence of power in the EWS plot is because of the

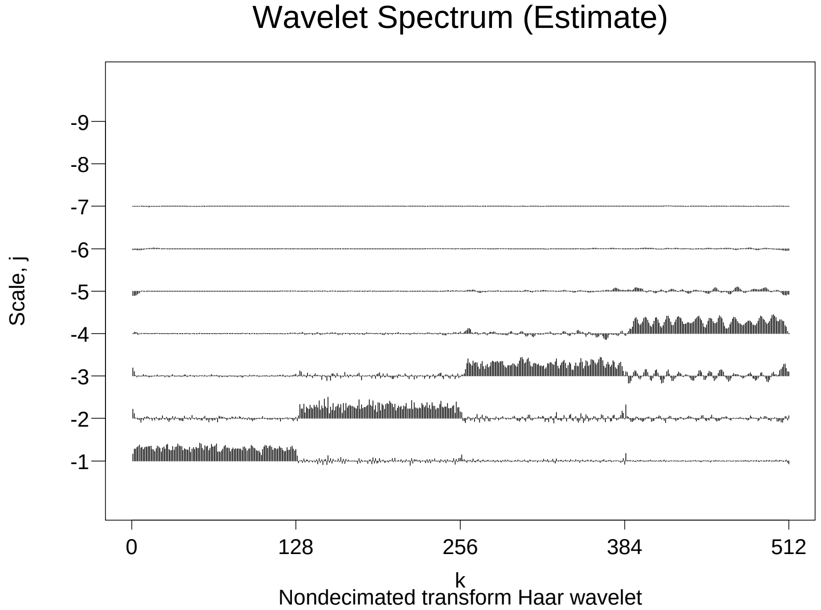


Figure 4. (Simulation generated) plot showing (estimated) EWS for concatenated Haar process. The horizontal axis shows "normal" time k but could be labelled in rescaled time $z = k/512$.

disentanglement mentioned in (ii): the EWS only shows variation in a scale but only that scale.

(iv) Application to infant ECG

Figure 5 shows an estimate of the evolutionary wavelet spectrum for the infant ECG data. It can be useful to aggregate information over scales. Figure 6 displays the variation in the ECG series at fine scales. The fine scale plot was obtained by summing over contributions from scales -1 thru -4 in the smoothed corrected wavelet periodogram (approximately 32s, 1, 2 and 4 minute time-scales). The dashed line in figure 6 shows another covariate: the sleep state of the infant as judged by a trained observer. It is clear that there is a correlation between the ECG and sleep state which can be exploited. There is no real clear link between the sleep state and the EWS at coarse and medium scales. However, a plot of an estimate of the localised variance, v, (not shown here) gives some idea of overall sleep behaviour.

The EWS is a useful tool in that it provides insight into the time-scale behaviour of a time series (much in the same way as a periodogram gives information about power in a stationary time series at different frequencies). However, the EWS has additional uses: Nason, Sapatinas and Sawczenko (1999) have also used the EWS to build models between sleep state (expensive and intrusive to measure) and ECG (cheap and easy) which allows estimates of future sleep state values to be predicted from future ECG values.

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. Nason was supported in part by EPSRC Grants GR/K70236 and GR/M10229 and von Sachs by DFG Grant Sa 474/3-1 and by EPSRC Visiting Fellowship GR/L52673.

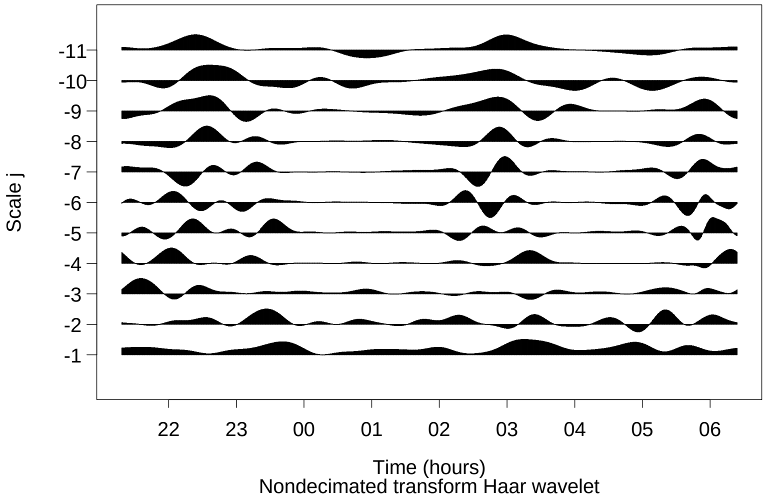


Figure 5. Estimate of the evolutionary spectrum $S_j(z)$ for infant ECG series shown in figure 3 (just as figure 4 does for the simulated concatenated Haar process). The original time axis is shown although it is really rescaled time in the interval $(0, 1)$.

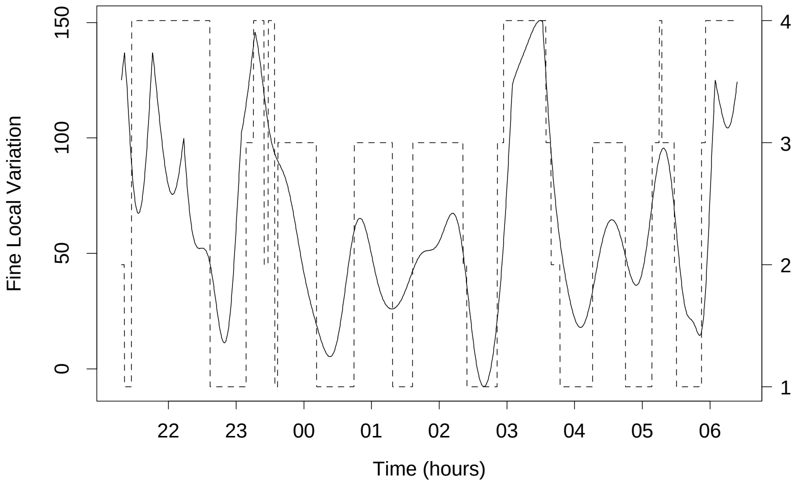


Figure 6. Estimate of local variation at fine scales for infant ECG series shown in figure 3 (solid line). The dashed line shows the sleep state of the infant (visually determined by a trained observer) ranging from 1=quiet sleep, 2=between 1 and 3, 3=active sleep, 4=awake. The original time axis is shown although z is really rescaled time in the interval $(0, 1)$.

Appendix A. Discrete non-decimated wavelets

Discrete non-decimated wavelets are merely the vectors of "filter" coefficients that appear in the matrix representation of the DWT. We shall define those that appear at scale $-j$ to be ψ_j. So, for example, with Haar $\psi_{-1} = (1/\sqrt{2}, -1/\sqrt{2})$

and $\psi_{-2} = (1/2, 1/2, -1/2, -1/2)$ and so on for other scales. The ψ_j for $j = -1, \ldots, -J$ can be obtained for any Daubechies compactly supported wavelet using the formulae:

$$\tilde{h}_{j-1,n} = \sum_k h_{n-2k} \tilde{h}_{j,k}, \qquad \psi_{j,n} = \sum_k g_{n-2k} \tilde{h}_{j,k}$$

where $\{h_k\}$ and $\{g_k\}$ are the usual Daubechies quadrature mirror filters and $\tilde{h}_{-1,k} = h_k$. For non-decimated wavelets $\psi_{jk}(\tau)$ is the kth element of the vector $\psi_{j(k-\tau)}$, i.e. ψ_{jk} shifted by integers τ. The key point for *non-decimated* discrete wavelets is they can be shifted to any location and not just by shifts by 2^{-j} (as in the DWT). Hence non-decimated discrete wavelets are no longer orthogonal but an overcomplete collection of shifted vectors. The NDWT can be computed through a fast algorithm (similar to the Mallat pyramid algorithm described in Silverman, 1999) which takes a computational effort of order $n \log n$ for a data set of length n. See Nason and Silverman (1995) for a more detailed description of the NDWT.

References

Abry, P. Flandrin, P., Taqqu, M. & Veitch, D. 1999 Wavelets for the analysis, estimation and synthesis of scaling data. in: Self Similar Network Traffic Analysis and Performance Evaluation. (eds. K. Park and W. Willinger). Wiley.

Abry, P., Gonçalvès, P. & Flandrin, P. 1995 Wavelets, spectrum analysis and 1/f processes. In *Wavelets and Statistics* (ed. A. Antoniadis & G. Oppenheim). Springer Lecture Notes in Statistics, no. 103, pp. 15–29.

Allan, D.W. 1966 Statistics of atomic frequency clocks. *Proceedings of the IEEE*, **31**, 221–230.

Beran, J. 1994 Statistics for long-memory processes. London: Chapman and Hall.

Coifman, R. R. & Donoho, D. L. 1994 Translation-invariant denoising. In *Wavelets and Statistics* (ed. A. Antoniadis & G. Oppenheim). Springer Lecture Notes in Statistics, no. 103, pp. 125–150.

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

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. 1995 Wavelet shrinkage: Asymptopia? (with discussion). *J. Roy. Statist. Soc.* B **57**, 301–369.

Farge, M. 1999 Intermittency and coherent votices in fully-developed turbulence. *Phil. Trans. R. Soc. Lond.* A **357** (to appear).

Flandrin, P. 1998. *Time–Frequency / Time–Scale Analysis* (Wavelet Analysis and Its Applications, Vol 10). Academic Press, San Diego, CA.

Gao, H.-Ye 1997. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Series Anal.*, **18**, 231–251.

Gonçalvès, P. & Flandrin, P. 1993 Bilinear Time-Scale Analysis Applied to Local Scaling Exponents Estimation. in: Progress in Wavelet Analysis and Applications (Y. Meyer and S. Roques, eds.), pp. 271–276, Editions Frontieres, Gif-sur-Yvette.

Johnstone, I.M. 1999 Wavelets and the theory of nonparametric function estimation. *Phil. Trans. R. Soc. Lond.* A **357** (to appear).

Hoffmann, M. 1999. On nonparametric estimation in nonlinear AR(1)-models. *Statistics and probability letters* (to appear).

McCoy, E.J. & Walden, A.T. 1996 Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics*, **5**, 26–56.

Nason, G.P., Sapatinas, T. & Sawczenko, A. 1999 Statistical modelling of time series using non-decimated wavelet representations. Preprint, University of Bristol.

Nason, G. P. & Silverman, B. W. 1995 The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics* (ed. A. Antoniadis & G. Oppenheim). Springer Lecture Notes in Statistics, no. 103, pp. 281–300.

Nason, G.P., von Sachs, R. & Kroisandt, G. 1998 Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. Discussion Paper 98/22, Institut de Statistique, UCL, Louvain-la-Neuve.

Neumann, M. 1996. Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.*, **17**, 601–633.

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

Page, C.H. 1952 Instantaneous power spectra. *J. Appl. Phys.* , **23**, 103–106.

Percival, D.B. & Guttorp, P. 1994 Long-memory processes, the Allan variance and wavelets. In *Wavelets in Geophysics* (ed. E. Foufoula-Georgiou & P. Kumar), pp. 325–344, Academic Press.

Priestley, M.B. 1965 Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc.* B **27**, 204–237.

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

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

Serroukh, A., Walden, A.T. & Percival, D.B. 1998 Statistical properties of the wavelet variance estimator for non-Gaussian/non-linear time series. Statistics Section *Technical Report*, TR-98-03. Department of Mathematics, Imperial College, London.

von Sachs, R. & MacGibbon B. 1998 Nonparametric curve estimation by wavelet thresholding with locally stationary errors (submitted for publication).

von Sachs, R. & Neumann, M. 1998 A wavelet-based test for stationarity (submitted for publication).

von Sachs, R. & Schneider, K. 1996 Wavelet smoothing of evolutionary spectra by non-linear thresholding. *Appl Comp. Harmonic Anal.* , **3**, 268–282.

Silverman, B.W. 1999 Wavelets in statistics: beyond the standard assumptions. *Phil. Trans. R. Soc. Lond.* A **357** (to appear).

Silverman, R.A. 1957 Locally stationary random processes. *IRE Trans. Information Theory*, **IT-3**, 182–187.

Wang, Y., Cavanaugh, J. and Song, Ch. 1997 Self-similarity index estimation via wavelets for locally self-similar processes. Preprint, Dept. Stat., University of Missouri.

Whitcher, B., Byers, S.D., Guttorp, P. & Percival, D.B. 1998 Testing for homogeneity of variance in time series: long memory, wavelets and the Nile river. Submitted for publication.

Wornell, G.W. and Oppenheim, A.V. 1992 Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on signal processing*, **40**, 611–623.