Ann Inst Stat Math (2010) 62:943–966
DOI 10.1007/s10463-008-0195-z

Wavelet variance analysis for gappy time series

Debashis Mondal · Donald B. Percival

Received: 31 October 2007 / Revised: 22 May 2008 / Published online: 9 September 2008

Abstract The wavelet variance is a scale-based decomposition of the process variance for a time series and has been used to analyze, for example, time deviations in atomic clocks, variations in soil properties in agricultural plots, accumulation of snow fields in the polar regions and marine atmospheric boundary layer turbulence. We propose two new unbiased estimators of the wavelet variance when the observed time series is 'gappy,' i.e., is sampled at regular intervals, but certain observations are missing. We deduce the large sample properties of these estimators and discuss methods for determining an approximate confidence interval for the wavelet variance. We apply our proposed methodology to series of gappy observations related to atmospheric pressure data and Nile River minima.

Keywords Cumulant · Fractionally differenced process · Local stationarity · Nile River minima · Semi-variogram · TAO data

1 Introduction

The wavelet variance (also called the wavelet power spectrum) decomposes the variance of a time series on a scale by scale basis and provides a time- and scale-based

---

Supported by NSF Grant DMS-02-22115.

D. Mondal (✉)
Department of Statistics, University of Chicago, 5734 South University Avenue,
Chicago, IL 60637, USA
e-mail: mondal@galton.uchicago.edu

D. B. Percival
Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195, USA
e-mail: dbp@apl.washington.edu

analysis of variance. The wavelet variance is particularly useful as an exploratory tool to identify important scales, to assess properties of long memory processes, to detect inhomogeneity of variance in time series and to estimate time-varying power spectra (thus complementing classical Fourier analysis). See, for example, the book by Percival and Walden (2000) or the review paper by Nason and von Sachs (1999). Applications include the analysis of time series related to electroencephalographic sleep state patterns of infants (Chiann and Morettin 1998), the El Niño–Southern Oscillation (Torrence and Compo 1998), soil variations (Lark and Webster 2001), solar coronal activity (Rybák and Dorotovič 2002), the relationship between rainfall and runoff (Labat et al. 2001), ocean surface waves (Massel 2001), surface albedo and temperature in desert grassland (Pelgrum et al. 2000), heart rate variability (Pichot et al. 1999) and the stability of the time kept by atomic clocks (Greenhall et al. 1999).

However, in practice, time series collected in various fields often deviate from regular sampling by having missing values ('gaps') amongst otherwise regularly sampled observations. As is also the case with the classical Fourier transform, the usual discrete wavelet transform is designed for regularly sampled observations and cannot be applied directly to time series with gaps. In this paper we develop a new methodology that can handle gaps in the data and provide valid statistical inference for wavelet variances. In geophysics, gaps are often handled by interpolating the data (see e.g., Vio et al. 2000), but such schemes are faced with problems of bias and of deducing what effect interpolation has had on any resulting statistical inference. On the other hand, there are various definitions for nonstandard wavelet transforms that could be applied to gappy data, with the 'lifting' scheme being a prominent example (Sweldens 1997). Wavelet analysis has also been discussed in the context of irregular time series (Foster 1996), and in the context of signals with continuous gaps (Frick et al. 1998). The general problem with these approaches is that the wavelet coefficients are not truly associated with particular scales, thus making it hard to draw meaningful scale-dependent inferences. The methodologies developed in this paper overcome these problems for gappy time series and reduce to the usual estimators of the wavelet variance for gap-free data. (Related works address the problem of the spectral analysis of gappy data, e.g., Stoica et al. 2000. The statistical properties of some of these methodologies are unknown and not easy to derive. We indicate in Sect. 8 how we can use our wavelet variance estimator to estimate the spectrum for gappy data.)

This paper is laid out as follows. In Sect. 2 we review estimation of the wavelet variance for gap-free time series based upon Daubechies wavelet filters. In Sects. 3 and 4 we describe confidence intervals for the wavelet variance based upon asymptotically normal unbiased estimators of this variance that are appropriate for gappy time series. In Sect. 5 we compare various estimates and perform some simulation studies on autoregressive and fractionally differenced processes, while Sect. 6 describes schemes for estimating the wavelet variance for time series with stationary dth order backward differences. We consider two examples involving gappy time series related to atmospheric pressure and Nile River minima in Sect. 7. Finally we end with some discussion in Sect. 8.

2 Wavelet variance estimation for non-gappy time series

Let $h_{1,l}$ denote a unit level Daubechies wavelet filter of width L normalized such that $\sum_l h_{1,l}^2 = 1/2$ (Daubechies 1992). The transfer function for this filter, i.e., its discrete Fourier transform (DFT),

$$H_1(f) = \sum_{l=0}^{L-1} h_{1,l} \mathrm{e}^{-\mathrm{i}2\pi f l},$$

has a corresponding squared gain function by definition satisfying

$$\mathcal{H}_1(f) = |H_1(f)|^2 = \sin^L(\pi f) \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2} - 1 + l}{l} \cos^{2l}(\pi f). \tag{1}$$

We note that $h_{1,l}$ can be expressed as the convolution of $L/2$ first difference filters and a single averaging filter that can be obtained by performing $L/2$ cumulative summations on $h_{1,l}$. The jth level wavelet filter $h_{j,l}$ is defined as the inverse DFT of

$$H_j(f) = H_1(2^{j-1} f) \prod_{l=0}^{j-2} \mathrm{e}^{-\mathrm{i}2\pi 2^l f (L-1)} H_1\left(\tfrac{1}{2} - 2^l f\right). \tag{2}$$

The width of this filter is given by $L_j = (2^j - 1)(L - 1) + 1$. We denote the corresponding squared gain function by $\mathcal{H}_j$. Since $H_j(0) = 0$, it follows that

$$\sum_{l=0}^{L_j - 1} h_{j,l} = 0. \tag{3}$$

For a nonnegative integer d, let X_t $(t \in \mathcal{Z})$ be a process with dth order stationary increments, which implies that

$$Y_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k} \tag{4}$$

is a stationary process. Let S_X and S_Y represent the spectral density functions (SDFs) for X_t and Y_t. These SDFs are defined over the Fourier frequencies $f \in [-1/2, 1/2]$ and are related by $S_Y(f) = [2\sin(\pi f)]^{2d} S_X(f)$. We can take the wavelet variance at scale $\tau_j = 2^{j-1}$ to be defined as

$$\nu_X^2(\tau_j) = \int_{-1/2}^{1/2} \mathcal{H}_j(f) S_X(f)\, \mathrm{d}f. \tag{5}$$

Roughly speaking, $\nu_X^2(\tau_j)$ is a measure of how much a weighted average of X_t over an interval of τ_j differs from a similar average in an adjacent interval. A plot of $\nu_X^2(\tau_j)$ against τ_j thus reveals which scales are important contributors to the process variance. For a process with infinite variance the wavelet variance at individual scales τ_j can exist and be finite, thus serving as a meaningful scale-based description of the process's variability. By virtue of (1) and (2), the wavelet variance is well defined for $L \geq 2d$. When $d = 0$ so that X_t is a stationary process with autocovariance sequence (ACVS) $s_{X,k} = \text{cov}\,\{X_t, X_{t+k}\}$, then we can rewrite the above as

$$\nu_X^2(\tau_j) = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} s_{X,l-l'}. \tag{6}$$

When $d = 1$, the increment process $Y_t = X_t - X_{t-1}$ rather than X_t itself is stationary, in which case the above equation can be replaced by one involving the ACVS for Y_t and the cumulative sum of $h_{j,l}$ (Craigmile and Percival 2005). Alternatively, let $\gamma_{X,k} = \text{var}\,(X_0 - X_k)/2$ denote the semi-variogram of X_t. Then the wavelet variance can be expressed as

$$\nu_X^2(\tau_j) = -\sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \gamma_{X,l-l'}. \tag{7}$$

The above equation also holds when X_t is stationary. The fundamental property of the wavelet variance is that it breaks up the process variance into pieces, each of which represents the contribution to the overall variance due to variability on a particular scale. In mathematical notation,

$$\text{var}\,(X_t) = \sum_{j=1}^{\infty} \nu_X^2(\tau_j).$$

Given an observed time series that can be regarded as a realization of $X_0, \ldots, X_{N-1}$ and assuming the sufficient condition $L > 2d$, an unbiased estimator of $\nu_X^2(\tau_j)$ is given by

$$\hat{\nu}_X^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} W_{j,t}^2, \quad \text{where } W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}$$

and $M_j = N - L_j + 1$. The wavelet coefficient process $W_{j,t}$ is stationary with mean zero, an SDF given by $\mathcal{H}_j(f)S_X(f)$ and an ACVS to be denoted by $s_{j,k}$. The following theorem holds (Percival 1995).

Theorem 1 *Let $W_{j,t}$ be a mean zero Gaussian stationary process satisfying the square integrable condition*

$$A_j = \int_{-1/2}^{1/2} \mathcal{H}_j^2(f) S_X^2(f)\,\mathrm{d}f = \sum_{k=-\infty}^{\infty} s_{j,k}^2 < \infty.$$

Then $\hat{\nu}_X^2(\tau_j)$ is asymptotically normal with mean $\nu_X^2(\tau_j)$ and large sample variance $2A_j/M_j$.

In practical applications, A_j is estimated by

$$\hat{A}_j = \tfrac{1}{2}\hat{s}_{j,0}^2 + \sum_{k=1}^{M_j-1} \hat{s}_{j,k}^2, \quad \text{where } \hat{s}_{j,k} = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1-|k|} W_{j,t} W_{j,t+|k|}$$

is the usual biased estimator of the ACVS for a process whose mean is known to be zero. Theorem 1 provides a simple basis for constructing confidence intervals for the wavelet variance $\nu_X^2(\tau_j)$.

3 Wavelet variance estimation for gappy time series

We consider first the case $d = 0$, so that X_t itself is stationary with ACVS $s_{X,k}$ and semi-variogram $\gamma_{X,k}$. Consider a portion $X_0, \ldots, X_{N-1}$ of this process. Let δ_t be the corresponding gap pattern, assumed to be a portion of a binary stationary process independent of X_t. The random variable δ_t assumes the values of 0 or 1 with nonzero probabilities, with zero indicating that the corresponding realization for X_t is missing. Define

$$\beta_k^{-1} = \Pr(\delta_t = 1 \text{ and } \delta_{t+k} = 1),$$

which is necessarily greater than zero. For $0 \le l, l' \le L_j - 1$, let

$$\hat{\beta}_{l,l'}^{-1} = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \delta_{t-l}\delta_{t-l'}.$$

We assume that $\hat{\beta}_{l,l'}^{-1} > 0$ for all l and l'. For a fixed j, this condition will hold asymptotically almost surely, but it can fail for finite N for a time series with too many gaps, a point that we return to in Sect. 8. By the weak law of large numbers for dependent processes (Feller 1966, p. 240, ex. 9) $\hat{\beta}_{l,l'}^{-1}$ is a consistent estimator of $\beta_{l-l'}^{-1}$ as $N \to \infty$.

Consider the following two statistics:

$$\hat{u}_X(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \hat{\beta}_{l,l'} X_{t-l} X_{t-l'} \delta_{t-l} \delta_{t-l'} \tag{8}$$

and

$$\hat{v}_X(\tau_j) = -\frac{1}{2M_j} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \hat{\beta}_{l,l'} \left(X_{t-l} - X_{t-l'}\right)^2 \delta_{t-l}\delta_{t-l'}. \tag{9}$$

When $\delta_t = 1$ for all t (the gap-free case), both statistics collapse to $\hat{\nu}_X^2(\tau_j)$. Conditioning on the observed gap pattern $\delta = (\delta_0, \ldots, \delta_{N-1})$, it follows that

$$E\{\hat{u}_X(\tau_j) \mid \delta\} = E\{\hat{v}_X(\tau_j) \mid \delta\} = \nu_X^2(\tau_j)$$

and hence that both statistics are unconditionally unbiased estimators of $\nu_X^2(\tau_j)$; however, whereas $\hat{\nu}_X^2(\tau_j) \geq 0$ necessarily in the gap-free case, these two estimators can be negative.

Remark 1 In the gappy case, the covariance-type estimator $\hat{u}_X(\tau_j)$ does not remain invariant if we add a constant to the original process X_t, whereas the semi-variogram type estimator $\hat{v}_X(\tau_j)$ does. In practical applications, this fact becomes important if the sample mean of the time series is large compared to its sample standard deviation, in which case it is important to use $\hat{u}_X(\tau_j)$ only after centering the series by subtracting off the sample mean.

4 Large sample properties of $\hat{u}_X(\tau_j)$ and $\hat{v}_X(\tau_j)$

For a fixed j, define the following stochastic processes:

$$Z_{u,j,t} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \beta_{l-l'} X_{t-l} X_{t-l'} \delta_{t-l}\delta_{t-l'} \tag{10}$$

and

$$Z_{v,j,t} = -\tfrac{1}{2} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \beta_{l-l'} (X_{t-l} - X_{t-l'})^2 \delta_{t-l}\delta_{t-l'}. \tag{11}$$

The processes $Z_{u,j,t}$ and $Z_{v,j,t}$ are both stationary with mean $\nu_X^2(\tau_j)$, and both collapse to $W_{j,t}^2$ in the gap-free case. Our estimators $\hat{u}_X(\tau_j)$ and $\hat{v}_X(\tau_j)$ are essentially sample means of $Z_{u,j,t}$ or $Z_{v,j,t}$, with $\beta_{l-l'}$ replaced by $\hat{\beta}_{l,l'}$. At this point we assume the following technical condition about our gap process.

Assumption 1 For fixed j, let $V_{p,t} = \delta_{t-l}\delta_{t-l'}$ for $p = (l, l')$ and $l, l' = 0, \ldots, L_j - 1$. We assume that the covariances of $V_{p_1,t}$ and $V_{p_2,t}$ are absolutely summable and the higher order cumulants satisfy

$$\sum_{t_1=0}^{N-1} \cdots \sum_{t_n=0}^{N-1} |\mathrm{cum}(V_{p_1,t_1}, \ldots, V_{p_n,t_n})| = o(N^{n/2}) \tag{12}$$

for $n = 3, 4, \ldots$ and for fixed $p_1, \ldots, p_n$.

Remark 2 Assumption 1 holds for a wide range of binary processes. For example, if δ_t is derived by thresholding a stationary Gaussian process whose covariances are absolutely summable, then the higher order cumulants of $V_{p,t}$ are absolutely summable. Assumption 1 is equivalent to saying that $M_j^{1/2}(\hat{\beta}_{l,l'}^{-1} - \beta_{l-l'}^{-1})$ is asymptotically normal with mean zero and finite variance. Thus Assumption 1 also holds for any stationary binary Markov chain of finite order. Note that Assumption 1 is weaker than the assumption that the cumulants are absolutely summable. This latter assumption is standard in the time series literature and has been used to prove central limit theorems in other contexts; see, e.g., Assumption 2.6.1 of Brillinger (1981).

The following central limit theorems (Theorems 2 and 3) provide the basis for inference about the wavelet variance using the estimators $\hat{u}_X(\tau_j)$ and $\hat{v}_X(\tau_j)$. We defer proofs to the Appendix, but we note that they are based on calculating mixed cumulants and require a technique sometimes called a diagram method. This method has been used widely to prove various central and non-central limit theorems involving functionals of Gaussian random variables; see e.g., Breuer and Major (1983), Giraitis and Surgailis (1985), Giraitis and Taqqu (1998), Fox and Taqqu (1987), Ho and Sun (1987) and the references therein. While building upon previous works, the proofs involve some unique and significantly different arguments that can be used to strengthen asymptotic results in other contexts, e.g., wavelet covariance estimation.

Theorem 2 *Suppose X_t is a stationary Gaussian process whose SDF is square integrable, and suppose δ_t is a strictly stationary binary process* (*independent of X_t*) *such that Assumption 1 holds Then $\hat{u}_X(\tau_j)$ is asymptotically normal with mean $\nu_X^2(\tau_j)$ and large sample variance $S_{u,j}(0)/M_j$, where $S_{u,j}$ is the SDF for $Z_{u,j,t}$.*

Remark 3 The Gaussian assumption on X_t can be dropped if we add appropriate mixing conditions, an approach that has been taken in the gap-free case (Serroukh et al. 2000). Since our estimators are essentially averages of stationary processes (10) and (11), asymptotic normality for the estimators (8) and (9) will follow if both X_t and the gap process δ_t possess appropriate mixing conditions. Moreover, construction of confidence intervals for the wavelet variance when X_t is non-Gaussian is the same as described below. This incorporates a certain degree of robustness into the methods developed in this paper.

Given a consistent estimator of $S_{u,j}(0)$, the above theorem can be used to construct an asymptotically correct confidence interval for $\nu_X^2(\tau_j)$. We use a multitaper spectral approach (Serroukh et al. 2000). Let

$$\widetilde{Z}_{u,j,t} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j,l'} \hat{\beta}_{l,l'} X_{t-l} X_{t-l'} \delta_{t-l} \delta_{t-l'}, \quad t = L_j - 1, \ldots, N-1.$$

Let $\lambda_{k,t}, t = 0, \ldots, M_j - 1$, for $k = 0, \ldots, K-1$ be the first K orthonormal Slepian tapers, where K is an odd integer. Define

$$J_{u,j,k} = \sum_{t=0}^{M_j-1} \lambda_{k,t} \widetilde{Z}_{u,j,t+L_j-1}, \quad \lambda_{k,+} = \sum_{t=0}^{M_j-1} \lambda_{k,t}$$

and

$$\widetilde{u}_j = \frac{\sum_{k=0,2,\ldots}^{K-1} J_{u,j,k}\lambda_{k,+}}{\sum_{k=0,2,\ldots}^{K-1} \lambda_{k,+}^2}.$$

We estimate $S_{u,j}(0)$ by

$$\hat{S}_{u,j}(0) = \frac{1}{K}\sum_{k=0}^{K-1}(J_{u,j,k} - \widetilde{u}_j\lambda_{k,+})^2.$$

Following the recommendation of Serroukh et al. (2000), we choose $K = 5$ and set the bandwidth parameter so that the Slepian tapers are band-limited to the interval $[-7/(2M_j), 7/(2M_j)]$. Previous Monte Carlo studies show that $\hat{S}_{u,j}(0)$ performs well (Serroukh et al. 2000). We now turn to the large sample properties of the second estimator $\hat{\upsilon}_X(\tau_j)$, which closely resemble those for $\hat{u}_X(\tau_j)$.

Theorem 3 *Suppose X_t or its increments is a stationary Gaussian process whose SDF is such that $\sin^2(\pi f)S_X(f)$ is square integrable. Assume the same conditions on δ_t as in Theorem 2. Then $\hat{\upsilon}_X(\tau_j)$ is asymptotically normal with mean $\upsilon_X^2(\tau_j)$ and large sample variance $S_{\upsilon,j}(0)/M_j$, where $S_{\upsilon,j}$ is the SDF for $Z_{\upsilon,j,t}$.*

Based upon

$$\widetilde{Z}_{\upsilon,j,t} = -\tfrac{1}{2}\sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_j-1} h_{j,l}h_{j,l'}\hat{\beta}_{l,l'}(X_{t-l} - X_{t-l'})^2\delta_{t-l}\delta_{t-l'},$$

we can estimate $S_{\upsilon,j}(0)$ using the same multitaper approach as before.

4.1 Efficiency study

The estimators $\hat{u}_X(\tau_j)$ and $\hat{\upsilon}_X(\tau_j)$ both work for stationary processes, whereas the latter can also be used for nonstationary processes with stationary increments. If $\hat{\upsilon}_X(\tau_j)$ performed better than $\hat{u}_X(\tau_j)$ in the stationary case, then the latter would be an unattractive estimator because it is restricted to just stationary processes. To address this issue, consider the asymptotic relative efficiency of the two estimators, which is given by the ratio of $S_{\upsilon,j}(0)$ to $S_{u,j}(0)$. For selected cases, this ratio can be computed to sufficient accuracy using the relationships

$$S_{u,j}(0) = \sum_{k=-\infty}^{\infty} s_{u,j,k} \quad \text{and} \quad S_{\upsilon,j}(0) = \sum_{k=-\infty}^{\infty} s_{\upsilon,j,k},$$

where $s_{u,j,k}$ and $s_{\upsilon,j,k}$ are the ACVSs corresponding to SDFs $S_{u,j}$ and $S_{\upsilon,j}$. We consider two cases, in both of which we use a level $j = 3$ Haar wavelet filter and assume that δ_t is a sequence of independent and identically distributed Bernoulli

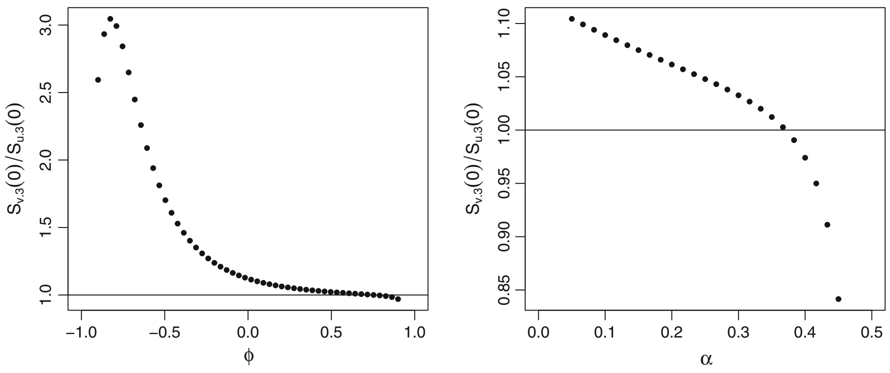

Fig. 1 Plot of asymptotic efficiency of $\hat{u}_X(\tau_3)$ with respect to $\hat{v}_X(\tau_3)$ under autoregressive (*left*) and fractionally differenced (*right*) models

random variables with $\Pr(\delta_t = 1) = 0.9$. In the first case, we let X_t to be a first order autoregressive (AR(1)) process with $s_{X,k} = \phi^{|k|}$. The left-hand plot of Fig. 1 shows the asymptotic relative efficiency as a function of ϕ. Except for ϕ close to unity, $\hat{u}_X(\tau_j)$ outperforms $\hat{v}_X(\tau_j)$. When ϕ is close to unity, the differencing inherent in $\hat{v}_X(\tau_j)$ makes it a more stable estimator than $\hat{u}_X(\tau_j)$, which is intuitively reasonable because the AR(1) process starts to resemble a random walk. For the second case, let X_t to be a stationary fractionally differenced (FD) process with $s_{X,k}$ satisfying

$$s_{X,0} = \frac{\Gamma(1-2\alpha)}{\Gamma(1-\alpha)\Gamma(1-\alpha)} \quad \text{and} \quad s_{X,k} = s_{X,k-1}\frac{k+\alpha-1}{k-\alpha}$$

for $k = 1, 2, \ldots$, see, e.g., Granger and Joyeux (1980) and Hosking (1981). Here $\alpha < 1/2$ is the long memory parameter, with $\alpha = 0$ corresponding to white noise and α close to $1/2$ corresponding to a highly correlated process whose ACVS damps down to zero very slowly. The right-hand plot of Fig. 1 shows the asymptotic relative efficiency as a function of α. As α approaches $1/2$, the semi-variogram type estimator $\hat{v}_X(\tau_j)$ outperforms $\hat{u}_X(\tau_j)$. These two cases tell us that $\hat{u}_X(\tau_j)$ is not uniformly better than $\hat{v}_X(\tau_j)$ for stationary processes and that, even for these processes, differencing can help stabilize the variance. Experimentation with other Daubechies filters leads us to the same conclusions.

5 Monte Carlo study

The purpose of this Monte Carlo study is to assess the adequacy of the normal approximation in Theorems 2 and 3 for simple situations. We also look at the performance of the estimates of $S_{u,j}(0)$ and $S_{v,j}(0)$.

5.1 Autoregressive process of order 1

In the first example, we simulate 1,000 time series of length 1,024 from an AR(1) process with $\phi = 0.9$. For each time series, we simulate δ_t independent and identically from a Bernoulli distribution with $\Pr(\delta_t = 1) = p = 0.9$. For each simulated gappy series, we estimate wavelet variances at scales indexed by $j = 1, \ldots, 6$ using $\hat{u}_X(\tau_j)$ and $\hat{v}_X(\tau_j)$ with the Haar wavelet filter. We also estimate the variance of the wavelet variances by using the multitaper method described in Sect. 4 and also from the sample variance of the Monte Carlo estimates. We then compare estimated values with the corresponding large sample approximations. In addition, we compare the performance of our estimators with the usual wavelet variance estimator after missing values are replaced with the mean of the data. Table 1 summarizes this experiment. Let $\hat{u}_{X,r}(\tau_j)$ and $\hat{v}_{X,r}(\tau_j)$ be the wavelet variance estimates for the rth realization, and let $\hat{S}_{u,j,r}(0)$ and $\hat{S}_{v,j,r}(0)$ be the corresponding multitaper estimates of $S_{u,j}(0)$ and $S_{v,j}(0)$. Let $\hat{\nu}^2_{X,r}(\tau_j)$ be the usual wavelet variance estimate for the rth realization when we replace gaps by the mean of the data. We note from Table 1 that the sample means of $\hat{u}_{X,r}(\tau_j)$ and $\hat{v}_{X,r}(\tau_j)$ are in excellent agreement with the true wavelet variance $\nu^2_X(\tau_j)$. The sample standard deviations of $\hat{u}_{X,r}(\tau_j)$ and $\hat{v}_{X,r}(\tau_j)$ are also in good agreement with $M_j^{-1/2} S_{u,j}^{1/2}(0)$ and $M_j^{-1/2} S_{v,j}^{1/2}(0)$. In particular, the ratios of the standard deviation of the $\hat{u}_{X,r}(\tau_j)$'s to their large sample approximations are quite close to unity, ranging between 0.884 and 1.005. The corresponding ratios for $\hat{v}_{X,r}(\tau_j)$ range between 0.926 and 1.002.

We also consider the performance of the multitaper estimates. In particular, we find the sample means of $M_j^{-1/2} \hat{S}_{u,j,r}^{1/2}(0)$ and $M_j^{-1/2} \hat{S}_{v,j,r}^{1/2}(0)$ to be close to their

Table 1 Summary of Monte Carlo results for AR(1) process

	j					
	1	2	3	4	5	6
$\nu^2_X(\tau_j)$	0.0500	0.0689	0.1079	0.1585	0.1907	0.1710
Mean of $\hat{u}_{X,r}(\tau_j)$	0.0502	0.0690	0.1084	0.1593	0.1911	0.1716
Mean of $\hat{v}_{X,r}(\tau_j)$	0.0503	0.0692	0.1085	0.1592	0.1910	0.1715
Mean of $\hat{\nu}^2_{X,r}(\tau_j)$ with gaps replaced by mean	0.0851	0.0781	0.0987	0.1343	0.1572	0.1402
$M_j^{-1/2} S_{u,j}^{1/2}(0)$	0.0087	0.0057	0.0104	0.0230	0.0347	0.0429
SD of $\hat{u}_{X,r}(\tau_j)$	0.0076	0.0055	0.0101	0.0204	0.0338	0.0431
Mean of $M_j^{-1/2} \hat{S}_{u,j,r}^{1/2}(0)$	0.0071	0.0047	0.0086	0.0175	0.0288	0.0340
$M_j^{-1/2} S_{v,j}^{1/2}(0)$	0.0027	0.0047	0.0102	0.0207	0.0345	0.0428
SD of $\hat{v}_{X,r}(\tau_j)$	0.0025	0.0044	0.0099	0.0205	0.0337	0.0428
Mean of $M_j^{-1/2} \hat{S}_{v,j,r}^{1/2}(0)$	0.0022	0.0039	0.0085	0.0173	0.0285	0.0339
SD of $\hat{\nu}^2_{X,r}(\tau_j)$ with gaps replaced by mean	0.0088	0.0059	0.0091	0.0175	0.0278	0.0343

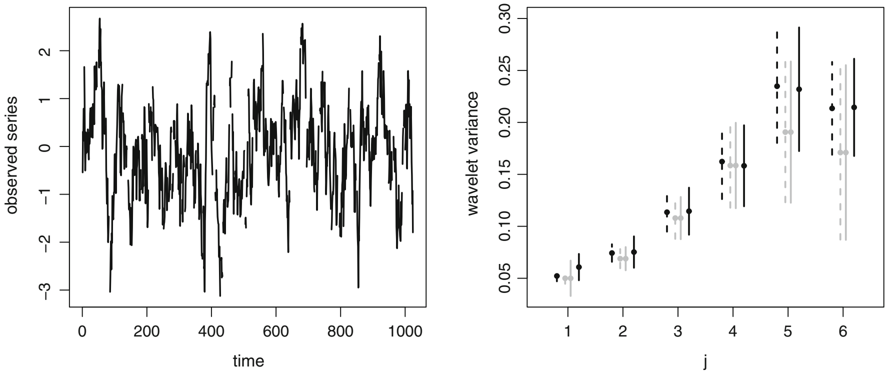

Fig. 2 Plot of a typical simulated gappy AR(1) time series and wavelet variances at various scales

respective theoretical values, but with a slight downward bias. Finally we note that estimates of wavelet variances as well as that of their standard deviations are significantly biased when we replace missing observations by the mean of the data. Figure 2 plots the realization of the time series for which the sum of squares of errors $\sum_j \{\hat{u}_{X,r}(\tau_j) - \nu_X^2(\tau_j)\}^2$ is closest to the average sum of squares of errors, namely, $1,000^{-1} \sum_r \sum_j \{\hat{u}_{X,r}(\tau_j) - \nu_X^2(\tau_j)\}^2$. For this typical realization, we also plot the estimated and theoretical wavelet variances with corresponding 95% confidence intervals. The black (gray) solid line in Fig. 2 gives the estimated (theoretical) confidence intervals based on $\hat{u}_X(\tau_j)$, with the dotted lines indicating corresponding intervals based upon $\hat{v}_X(\tau_j)$. We see reasonable agreement between the theoretical and estimated values.

5.2 Kolmogrov turbulence

In the second example, we generate 1,000 time series of length 1,024 from an FD(5/6) process, which is a nonstationary process that has properties very similar to Kolmogorov turbulence and hence is of interest in atmospheric science and oceanography. For each time series, we simulate the gaps δ_t as before. In this example increments of X_t rather X_t itself are stationary. Therefore we employ only $\hat{v}_X(\tau_j)$ and consider how well its variance is approximated by the large sample result stated in Theorem 3. We also compare our method with the usual wavelet variance estimator by replacing missing values with the mean of the data. Table 2 summarizes the results of this experiment using the Haar wavelet filter. Again we find that, for each level j, the average $\hat{v}_{X,r}(\tau_j)$ is in excellent agreement with the true $\nu_X^2(\tau_j)$; the sample standard deviation of $\hat{v}_{X,r}(\tau_j)$ is in good agreement with its large sample approximation; and the sample mean of $M_j^{-1/2}\hat{S}_{v,j}^{1/2}(0)$ is close to $M_j^{-1/2}S_{v,j}^{1/2}(0)$, with a slight downward bias. However, the estimates and their standard deviations are markedly inferior when we employ the usual wavelet variance estimator with gaps replaced by the mean of the data. Figure 3 has the same format as Fig. 2 and again indicates reasonable agreement between theoretical and estimated values.

Table 2 Summary of Monte Carlo results for FD(5/6) process

	j					
	1	2	3	4	5	6
$\nu_X^2(\tau_j)$	0.2594	0.3078	0.4427	0.6831	1.0762	1.7050
Mean of $\hat{\upsilon}_{X,r}(\tau_j)$	0.2599	0.3081	0.4421	0.6832	1.0771	1.7179
Mean of $\hat{\upsilon}^2_{X,r}(\tau_j)$ with gaps replaced by mean	1.4826	0.8854	0.6793	0.7181	0.9598	1.4326
$M_j^{-1/2} S_{\upsilon,j}^{1/2}(0)$	0.0141	0.0203	0.0399	0.0857	0.1899	0.4281
SD of $\hat{\upsilon}_{X,r}(\tau_j)$	0.0129	0.0186	0.0386	0.0847	0.1877	0.4275
Mean of $M_j^{-1/2} \hat{S}_{\upsilon,j,r}^{1/2}(0)$	0.0119	0.0168	0.0330	0.0704	0.1567	0.3489
SD of $\hat{\upsilon}^2_{X,r}(\tau_j)$ with gaps replaced by mean	0.8572	0.4308	0.2234	0.1406	0.1806	0.3637

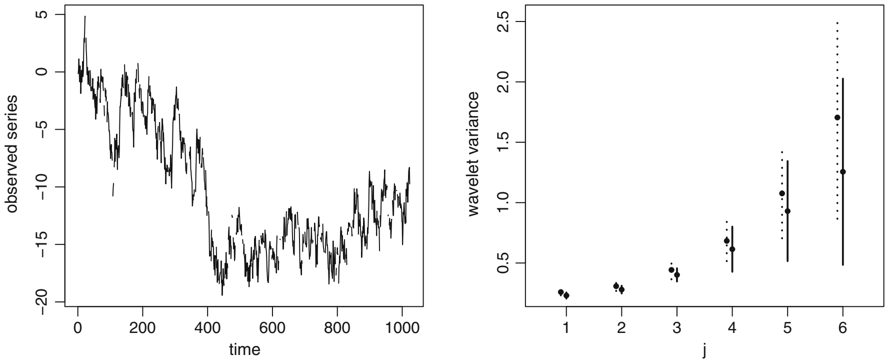

Fig. 3 Plot of a typical simulated gappy FD(5/6) time series and wavelet variances at various scales. *Solid lines* indicate the estimated intervals while dotted lines indicate the true intervals

6 Generalizations of basic theory

6.1 Non-Daubechies wavelet filters

Although we formulated Theorems 2 and 3 in terms of the Daubechies wavelet filters, in fact they are valid for a wider class of filters. In particular, both theorems continue to hold for any filter $h_{j,l}$ that has finite width and sums to zero (if the original process X_t has mean zero, Theorem 2 only requires $h_{j,l}$ to be of finite width). This provides us with an estimation theory for wavelet variances other than those defined by a Daubechies wavelet filter. For example, at the unit scale, we can entertain the filter $\{-1/4, 1/2, -1/4\}$, which can be considered to be a discrete approximation of the Mexican hat wavelet.

6.2 Gappy dth order stationary increment processes

Here we extend the basic theory to estimate the wavelet variance for a process X_t, $t \in \mathcal{Z}$, with dth order stationary increments Y_t. Let μ_Y be the mean, $s_{Y,k}$ the ACVS and $\gamma_{Y,k}$ the semi-variogram of Y_t. For $L \geq 2d$, an expression for the Daubechies wavelet variance that is analogous to (6) is

$$\nu_X^2(\tau_j) = \sum_{l=0}^{L_j-d-1} \sum_{l'=0}^{L_j-d-1} b_{j,l,d} b_{j,l',d} s_{Y,l-l'}, \tag{13}$$

where $b_{j,l,r}$ is the rth order cumulative summation of the Daubechies wavelet filter $h_{j,l}$, i.e.,

$$b_{j,l,0} = h_{j,l}, \qquad b_{j,l,k} = \sum_{r=0}^{l} b_{j,r,k-1},$$

for $l = 0, \ldots, L_j - k - 1$ (see Craigmile and Percival 2005). Moreover, if $L > 2d$, we obtain the alternative expression

$$\nu_X^2(\tau_j) = -\sum_{l=0}^{L_j-d-1} \sum_{l'=0}^{L_j-d-1} b_{j,l,d} b_{j,l',d} \gamma_{Y,l-l'}. \tag{14}$$

We can now proceed to estimate $\nu_X^2(\tau_j)$ as follows. First we carry out dth order differencing of the observed X_t to obtain an observed Y_t. This will generate a new gap pattern that has more gaps than the old gap structure, but the new gap pattern will still be stationary and independent of Y_t. We then mimic the stationary ($d = 0$) case described as in Sect. 3 with $b_{j,l,d}$ replacing $h_{j,l}$, the new gap pattern replacing δ_t, and Y_t replacing X_t in the estimators (8) and (9). As a simple illustration of this scheme, consider the case $d = 2$. For $t = 2, 3, \ldots$, compute $Y_t = X_t - 2X_{t-1} + X_{t-2}$ whenever $\delta_t = \delta_{t-1} = \delta_{t-2} = 1$. Let $\eta_t = 1$ if $\delta_t = \delta_{t-1} = \delta_{t-2} = 1$ and $= 0$ otherwise.

Let

$$\hat{\rho}_{l,l'}^{-1} = \frac{1}{M_j} \sum_{t=L_j-3}^{N-1} \eta_{t-l} \eta_{t-l'},$$

where now M_j is redefined to be $N - L_j + 3$. Again $\hat{\rho}_{l,l}^{-1}$ is a consistent estimator of $\rho_{l-l'}^{-1} = \Pr(\eta_{t-l} = 1,\ \eta_{t-l'} = 1)$. As before, assume $\hat{\rho}_{l,l'}^{-1} > 0$ for $l, l' = 0, \ldots, L_j - 3$. The new versions of the estimators of $\nu_X^2(\tau_j)$ are then given by

$$\hat{u}_X(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-3}^{N-1} \sum_{l=0}^{L_j-3} \sum_{l'=0}^{L_j-3} b_{j,l,2} b_{j,l',2} \hat{\rho}_{l,l'} Y_{t-l} Y_{t-l'} \eta_{t-l} \eta_{t-l'},$$

and

$$\hat{\hat{v}}_X(\tau_j) = -\frac{1}{2M_j}\sum_{t=L_j-3}^{N-1}\sum_{l'=0}^{L_j-3}\sum_{l'=0}^{L_j-3} b_{j,l,2}b_{j,l',2}\hat{\rho}_{l,l'}(Y_{t-l}-Y_{t-l'})^2\eta_{t-l}\eta_{t-l'}.$$

The large sample properties of these estimators are given by obvious analogs to Theorems 2 and 3.

Theorem 4 *Suppose X_t is a process whose dth order increments Y_t are a stationary Gaussian process with square integrable SDF, and suppose δ_t is a strictly stationary binary process (independent of X_t) such that the derived binary process η_t satisfies Assumption 1. Then, if $L \geq 2d$, $\hat{\hat{u}}_X(\tau_j)$ is asymptotically normal with mean $\nu_X^2(\tau_j)$ and large sample variance $S_{d,u,j}(0)/M_j$, where $S_{d,u,j}$ is the SDF for $\sum_l \sum_{l'} b_{j,l,d}b_{j,l',d}\rho_{l,l'}Y_{t-l}Y_{t-l'}\eta_{t-l}\eta_{t-l'}$.*

Theorem 5 *Suppose X_t is a process whose increments of order $d+1$ are a stationary Gaussian process with square integrable SDF, and suppose δ_t is as in the previous theorem. Then, if $L > 2d$, $\hat{\hat{v}}_X(\tau_j)$ is asymptotically normal with mean $\nu_X^2(\tau_j)$ and large sample variance $S_{d,v,j}(0)/M_j$, where $S_{d,v,j}$ is the SDF for $-\sum_l \sum_{l'} b_{j,l,d}b_{j,l',d}\rho_{l,l'}$ $(Y_{t-l}-Y_{t-l'})^2\eta_{t-l}\eta_{t-l'}/2$.*

The proofs of Theorems 4 and 5 are similar to those of, respectively, Theorems 2 and 3 and thus are omitted.

Remark 4 Since each extra differencing produces more gaps, an estimate that requires less differencing will be more efficient. This is where the semi-variogram type estimator $\hat{v}_X(\tau_j)$ comes in handy. Let C_t denote the backward differences of order $d-1$ for X_t. Then C_t is not stationary but its increments are. Let the semi-variogram of C_t be denoted by $\gamma_{C,k}$. Then by virtue of (14), we can write for $L \geq 2d$

$$\nu_X^2(\tau_j) = -\sum_{l=0}^{L_j-d}\sum_{l'=0}^{L_j-d} b_{j,l,d-1}b_{j,l',d-1}\gamma_{C,l-l'}. \tag{15}$$

Thus alternatively we can proceed as follows. We carry out $d-1$ successive differences of X_t to obtain C_t and then use the semi-variogram type estimator with the new gap structure and with the Daubechies filter replaced by $b_{j,l,d-1}$. Unlike the stationary case, this estimator often outperforms the covariance-type estimator that requires one more order of differencing.

6.3 Systematic gaps

We have focused on geophysical applications where gaps tend to occur at random. When systematic gaps occur, e.g., in financial time series when no trading takes place on weekends, we note that our estimates (8) and (9) produce valid unbiased estimate of the true wavelet variance as long as $\hat{\beta}^{-1}_{l,l'} > 0$ for $l, l' = 0, \ldots, L_j - 1$ (for the financial example, this condition on $\hat{\beta}$ holds when the length of the time series N is sufficiently large); moreover, our large sample theory can be readily adjusted to handle those gaps. First, we redefine the theoretical β by taking the deterministic limit of $\hat{\beta}$ as N tends to infinity. Next we observe that the processes $Z_{u,j,t}$ and $Z_{v,j,t}$ defined via (10) and (11) are no longer stationary under this systematic gap pattern. To see this consider $j = 2$ and the Haar wavelet filter for which $L_2 = 4$. Then $Z_{u,2,t}$ for a Friday depends on the observations obtained from Tuesday to Friday while $Z_{u,2,t}$ for a Monday depends only on values of the time series observed on Monday and the previous Friday. As a consequence we can not invoke Theorems 2 or 3 directly. However, because the gaps have a period of a week, we can retrieve stationarity by summing $Z_{u,j,t}$ and $Z_{v,j,t}$ over 7 days; i.e., $\sum_{m=0}^{6} Z_{u,j,t+m}$ and $\sum_{m=0}^{6} Z_{v,j,t+m}$ are stationary processes. For large M_j the summations of t in estimators (8) and (9) are essentially sums over these stationary processes, plus terms that are asymptotically negligible. Thus we can prove asymptotic normality of (8) and (9) from the respective asymptotic normality of the averages of $\sum_{m=0}^{6} Z_{u,j,t+m}$ and $\sum_{m=0}^{6} Z_{v,j,t+m}$. The proofs are similar to those for Theorems 2 and 3, with some simplification because the gaps are deterministic (an alternative approach is to use Theorem 1 of Ho and Sun 1987). Large sample confidence intervals can be constructed using the multitaper procedure described in Sect. 4.

7 Examples

7.1 Analysis of TAO data

We apply our techniques to daily atmospheric pressure data (Fig. 4) collected over a period of 578 days by the Tropical Atmospheric Ocean (TAO) buoy array operated by the National Oceanic and Atmospheric Administration. There were 527 days of observed values and 51 days during which no observations were made. Shorts gaps in this time series are mainly due to satellite transmission problems. Equipment

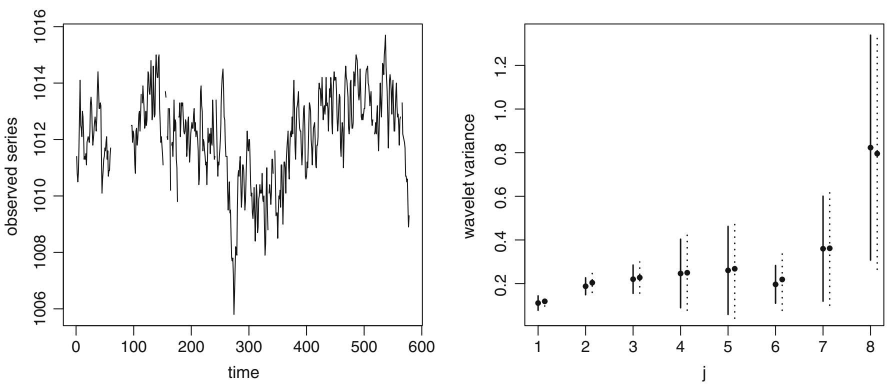

Fig. 4 Atmospheric pressure data (*left*) from NOAA's TAO buoy array and Haar wavelet variance estimates (*right*) for scales indexed by $j = 1, \ldots, 8$

malfunctions that require buoy repairs result in longer gaps. It is reasonable to assume that the gaps are independent of the pressure values and are a realization of a stationary process. Of particular interest are contributions to the overall variability due to different dynamical phenomena, including an annual cycle, interseasonal oscillations and a menagerie of tropical waves and disturbances associated with small time scales. We employ wavelet variance estimators (8) and (9) using the Haar wavelet filter.

Estimated wavelet variances for levels $j = 1, \ldots 8$ are plotted in Fig. 4 along with the 95% confidence intervals (solid and dotted lines for, respectively, $\hat{u}_X(\tau_j)$ and $\hat{v}_X(\tau_j)$). We see close agreement between these two estimation procedures. Note that the wavelet variance is largest at scales τ_7 and τ_8, which correspond to periods of, respectively, 128–256 and 256–512 days. Large variability at these scales is due to a strong yearly cycle in the data. Apart from this, we also see a much weaker peak at scale τ_5, which corresponds to a period of 32–64 days and captures the interseasonal variability. Note also that there is hardly any variability at scale τ_1, although there is some at scales τ_2, τ_3 and τ_4, indicating relatively important contributions to the variance due to disturbances at all but the very smallest scale. (We obtained similar results using the Daubechies $L = 4$ extremal phase and $L = 8$ least asymmetric wavelet filters.)

7.2 Nile River minima

This time series (Fig. 5) consists of measurements of minimum yearly water level of the Nile River over the years 622–1921, with 622–1284 representing the longest segment without gaps (Toussoun 1925). The rate of gaps is about 43% after year 1285. Several authors have previously analyzed the initial gap-free segment (see, e.g., Beran 1994, and Percival and Walden 2000). The entire series, including the gappy part, has been analyzed based on a parametric state space model (Palma and Del Pino 1999), in contrast to our nonparametric approach. Historical records indicate a change around 715 in the way the series was measured. For the gap-free segment, there is more variability at scales τ_1 and τ_2 before 715 than after (Whitcher et al. 2002). Therefore

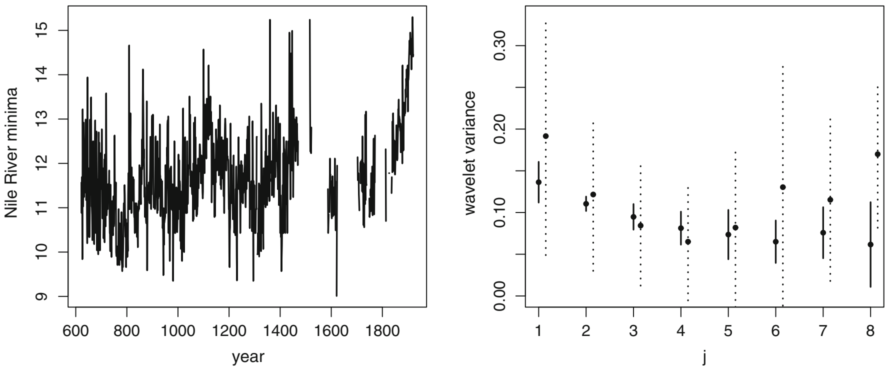

Fig. 5 Nile River minima (*left*) and Haar wavelet variance estimates (*right*) for scales indexed by $j = 1, \ldots, 8$

we restrict ourselves to the period 716–1921. Figure 5 plots wavelet variance estimates up to scale τ_8 along with 95% confidence intervals using $\hat{\upsilon}_X(\tau_j)$ with the Haar wavelet filter. Here solid lines stand for the gap-free segment 716–1284, and dotted lines for the gappy segment 1286–1921. Except at scales τ_1, τ_6 and τ_8, we see reasonably good agreement between estimates from the two segments. Substantial uncertainties due to the large number of gaps are reflected in the larger confidence intervals for the gappy segment. Under the assumption that the statistical properties of the Nile River were the same throughout 716–1921, we could combine the two segments to produce overall estimates and confidence intervals for the wavelet variances; however, this assumption is questionable at certain scales. Over the years 1286–1470, there are only six gaps. Separate analysis of this segment suggests more variability at scales τ_1 and τ_2 than what was observed in 716–1284. In addition, construction of the first Aswan Dam starting in 1899 changed the nature of the Nile River in the subsequent years. However, a wavelet variance analysis over 1286–1898 (omitting the years after the dam was built) does not differ much from that of 1286–1921. Thus the apparent increase in variability at the largest scale from segment 716–1284 to 1286–1921 cannot be attributed just to the influence of the dam.

8 Discussion

In Sect. 3 we made the crucial assumption that, for a fixed j, $\hat{\beta}^{-1}_{l,l'} > 0$ when $l, l' = 0, \ldots, L_j - 1$. For small sample sizes, this condition might fail to hold. This situation arises mainly when half or more of the observations are missing and can be due to systematic periodic patterns in the gaps. For example, if δ_t alternates between zero and one, then $\hat{\beta}^{-1}_{0,1}$ is zero, reflecting the fact that the observed time series does not contain relevant information about $\nu^2_X(\tau_1)$. A methodology different from what we have discussed might be able to handle some gap patterns for which $\hat{\beta}^{-1}_{l,l'} = 0$. In particular, generalized prolate spheroidal sequences have been used to handle spectral estimation of irregularly sampled processes (Bronez 1988). This approach in essence corresponds to the construction of special filters and could be used to construct approximations to the Daubechies filters when $\hat{\beta}^{-1}_{l,l'} = 0$.

Estimation of the SDF for gappy time series is a long-standing difficult problem. In Sect. 1 we noted that the wavelet variance provides a simple and useful estimator of the integral of the SDF over a certain octave band. In particular, the Blackman–Tukey pilot spectrum (Blackman and Tukey 1958, Sec. 18) coincides with the Haar wavelet variance. Recently Tsakiroglou and Walden (2002) generalized this pilot spectrum by utilising the (maximal overlap) discrete wavelet packet transform. The result is an SDF estimator that is competitive with existing estimators. With a similar generalization, our wavelet variance estimator for gappy time series can be adapted to serve as an SDF estimator. Moreover, Nason et al. (2000) used shrinkage of squared wavelet coefficients to estimate spectra for locally stationary processes. In the same vein, we can apply wavelet shrinkage to the $Z_{u,j,t}$ or $Z_{v,j,t}$ processes to estimate time-varying spectra when the original time series is observed with gaps.

Finally we note a generalization of interest in the analysis of multivariate gappy time series. Given two time series $X_{1,t}$ and $X_{2,t}$, the wavelet cross covariance yields a scale-based analysis of the cross covariance between the two series in a manner similar to wavelet variance analysis (for estimation of the wavelet cross covariance, see Whitcher et al. (2000), and the references therein). The methodology described in this paper can be readily adapted to estimate the wavelet cross covariance for multivariate time series with gaps.

Acknowledgments This research was supported by the U.S. National Science Foundation under Grant No. DMS 0222115. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors thank Peter Guttorp and Chris Bretherton for discussions and the associate editor and two referees for comments that improved the paper (in particular, the idea of including the standard estimator with gaps filled by the mean as part of the Monte Carlo experiments is due to one of the referees).

A Proofs

Here we sketch proofs of Theorems 2 and 3 (full details are given in Mondal and Percival 2008). First we state the followings propositions and lemmas. To avoid a triviality, we assume throughout that var $\{X_t\} > 0$.

Lemma 1 *Let $U_{l,l',t}$ and $V_{l,l',t}$ be stationary processes that are independent of each other for any choice of k, k', l and l'. Let*

$$U_{l,l',t} = \psi_{l,l'} + \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f t}\, \mathrm{d}\mathcal{U}_{l,l'}(f)$$

and

$$V_{l,l',t} = \omega_{l,l'} + \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f t}\, \mathrm{d}\mathcal{V}_{l,l'}(f)$$

be their respective spectral representations. For any k, k', l and l', let $S_{k,k',l,l'}$ and $G_{k,k',l,l'}$ denote the respective cross spectrum between $U_{k,k',t}$ and $U_{l,l',t}$ and between $V_{k,k',t}$ and $V_{l,l',t}$. Let $a_{l,l'}$ be fixed real numbers. Define

$$Q_t = \sum_{l,l'} a_{l,l'}(U_{l,l',t}V_{l,l',t} - \psi_{l,l'}\omega_{l,l'}).$$

Then Q_t is a second order stationary process whose SDF is given by

$$S_Q(f) = \sum_{k,k'}\sum_{l,l'} a_{k,k'}a_{l,l'}\big[\psi_{k,k'}\psi_{l,l'}G_{k,k',l,l'}(f) + \omega_{k,k'}\omega_{l,l'}S_{k,k',l,l'}(f) \\ + S * G_{k,k',l,l'}(f)\big], \tag{16}$$

where

$$S * G_{k,k',l,l'}(f) = \int_{-1/2}^{1/2} G_{k,k',l,l'}(f - f')S_{k,k',l,l'}(f')\,\mathrm{d}f'.$$

Proposition 1 *Let X_t be a real-valued zero mean Gaussian stationary process with ACVS $s_{X,m}$ and SDF S_X that is square integrable over $[-1/2, 1/2]$. Let δ_t be a binary-valued strictly stationary process that is independent of X_t and satisfies Assumption 1. Let $Z_{u,j,t}$ be as in equation (10). Then $Z_{u,j,t}$ is a second order stationary process whose SDF at zero is strictly positive.*

Proposition 2 *Let X_t be a real-valued Gaussian stationary process with zero mean, and SDF S_X that satisfies*

$$\int_{-1/2}^{1/2} \sin^4(\pi f) S_X^2(f)\, \mathrm{d}f < \infty.$$

Let δ_t be a binary-valued strictly stationary process that is independent of X_t and satisfies Assumption 1. Let $Z_{v,j,t}$ be as in Eq. (11). Then $Z_{v,j,t}$ is a second order stationary process whose SDF at zero is strictly positive.

The following theorem is from Brillinger (1981, p. 21).

Theorem 6 *Consider a two way array of random variables (RVs) $\Theta_{i,j}$, $j = 1, \ldots, J_i$ and $i = 1, \ldots, n$. Consider the n RVs $\Upsilon_i = \prod_{j=1}^{J_i} \Theta_{i,j}$ for $i = 1, \ldots, n$. Then the joint cumulant of $\Upsilon_1, \ldots, \Upsilon_n$ is given by the formula*

$$\mathrm{cum}(\Upsilon_1, \ldots, \Upsilon_n) = \sum_{\chi} \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_1) \cdots \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_r),$$

where the summation is over all indecomposable partitions $\chi = \chi_1 \cup \cdots \cup \chi_r$ of the (not necessarily rectangular) two way table

$$\begin{array}{ccc} (1,1) & \cdots & (1,J_1) \\ \vdots & & \vdots \\ (n,1) & \cdots & (n,J_n). \end{array} \tag{17}$$

Lemma 2 *Assume that X_t satisfies the conditions stated in Theorem 2. Let $U_{p,t} = X_{t-l} X_{t-l'}$ and $\psi_p = \mathrm{E}\, U_{p,t}$, where $p = (l, l')$. Then for $n \geq 3$ and fixed $p_1, \ldots, p_n$,*

$$\sum_{t_1,\ldots t_n} |\mathrm{cum}(U_{p_1,t_1} - \psi_{p_1}, \ldots, U_{p_n,t_n} - \psi_{p_n})| = o(M^{n/2}), \tag{18}$$

where each t_i ranges from 0 to $M - 1$ (here and below M is shorthand for M_j in the main text).

Lemma 3 *Assume that X_t satisfies the conditions stated in Theorem 3. Let $U_{p,t} = -(X_{t-l} - X_{t-l'})^2/2$ and $\psi_p = \mathrm{E}\, U_{p,t}$, in which $p = (l, l')$. Then for $n \geq 3$ and fixed $p_1, \ldots, p_n$,*

$$\sum_{t_1,\ldots t_n} |\mathrm{cum}(U_{p_1,t_1} - \psi_{p_1}, \ldots, U_{p_n,t_n} - \psi_{p_n})| = o(M^{n/2}),$$

where each t_i ranges from 0 to $M - 1$.

Lemma 4 *Let $U_{p,t}$ be either as in Lemma 2 or as in Lemma 3. Assume*

$$\kappa_n(p_1, \ldots, p_n, t_1, \ldots, t_n) = \mathrm{cum}(U_{p_1,t_1} - \psi_{p_1}, \ldots, U_{p_n,t_n} - \psi_{p_n}).$$

Define for $i = 1, 2, \ldots, n-1$

$$\kappa_n(p_1, \ldots, p_n, t_1, \ldots, t_i) = \sum_{t_{i+1},\ldots,t_n} M^{-(n-i-1)/2} \kappa_n(p_1, \ldots, p_n, t_1, \ldots, t_n),$$

where the summation in t_j ranges from 0 to $M-1$. Then $\kappa_n(p_1, \ldots, p_n, t_1, \ldots, t_i)$ is bounded and satisfies

$$\sum_{t_1,\ldots,t_i} \kappa_n(p_1, \ldots, p_n, t_1, \ldots, t_i) = o\left(M^{(i+1)/2}\right), \quad i = 1, 2, \ldots, n. \tag{19}$$

Proof of Theorem 2 Take $U_{p,t} = X_{t-l}X_{t-l'}$, $V_{p,t} = \delta_{t-l}\delta_{t-l'}$ and $a_p = h_{j,l}h_{j,l'}\beta_{l-l'}$, where $p = (l, l')$. Take $Q_t = \sum_p a_p(U_{p,t}V_{p,t} - \psi_p\omega_p)$ as in Lemma 1. Note that $\hat{u}_X(\tau_j) - \nu_X^2(\tau_j)$ is the average of Q_t over $L_j - 1 \leq t \leq N - 1$ with $\beta_{l-l'}$ replaced by its consistent estimate $\hat{\beta}_{l,l'}$. Since Q_t is stationary, we first prove a central limit theorem for $R = M^{-1/2}\sum_{t=0}^{M-1} Q_t$ and then invoke Slutsky's theorem to complete the proof that $\hat{u}_X(\tau_j)$ is asymptotically normal. We use page 2 of Žurbenko (1986) to write the log of the characteristic function of R as

$$\log F(\lambda) = \sum_{n=1}^{\infty} \frac{\mathrm{i}^n\lambda^n}{n!} \sum_{t_1,\ldots,t_n} \frac{B_n(t_1, \ldots, t_n)}{M^{n/2}},$$

where B_n is the nth order cumulant of Q_t, and each t_i ranges from 0 to $M-1$. Since Q_t is centered, $B_1(t_1) = 0$. By Proposition 1, the autocovariances $s_{Q,\tau}$ of Q_t are absolutely summable and $M^{-1}\sum_{t_1}\sum_{t_2} B_2(t_1, t_2) \to \sum_\tau s_{Q,\tau} = S_Q(0) > 0$. In order to prove the central limit theorem for R, it suffices to show that $\sum_{t_1,\ldots,t_n} M^{-n/2}B_n(t_1, \ldots, t_n) \to 0$ for $n = 3, 4, \ldots$.

First using page 19 of Brillinger (1981), we break up the nth order cumulant as follows:

$$B_n(t_1, \ldots, t_n) = \sum_{p_1} \cdots \sum_{p_n} a_{p_1} \cdots a_{p_n}$$
$$\mathrm{cum}(U_{p_1,t_1}V_{p_1,t_1} - \psi_{p_1}\omega_{p_1}, \ldots, U_{p_n,t_n}V_{p_n,t_n} - \psi_{p_n}\omega_{p_n}).$$

Let $D_{1,p,t} = (U_{p,t} - \psi_p)(V_{p,t} - \omega_p)$, $D_{2,p,t} = \omega_p(U_{p,t} - \psi_p)$ and $D_{3,p,t} = \psi_p(V_{p,t} - \omega_p)$. Then $U_{p,t}V_{p,t} - \psi_p\omega_p = D_{1,p,t} + D_{2,p,t} + D_{3,p,t}$. Using page 19 of Brillinger (1981) again, we have

$$\mathrm{cum}(U_{p_1,t_1}V_{p_1,t_1} - \psi_{p_1}\omega_{p_1}, \ldots, U_{p_n,t_n}V_{p_n,t_n} - \psi_{p_n}\omega_{p_n})$$
$$= \sum_{c_1,\ldots,c_n} \mathrm{cum}(D_{c_1,p_1,t_1}, \ldots, D_{c_n,p_n,t_n}),$$

where each c_i ranges from 1 to 3. Therefore, it suffices to show that, for fixed $p_1, \ldots, p_n$ and $c_1, \ldots, c_n$, $\mathrm{cum}(D_{c_1,p_1,t_1}, \ldots, D_{c_n,p_n,t_n}) = o(M^{n/2})$. Since the cumulant of n RVs is invariant under a reordering of the RVs, assume $c_1 = c_2 = \cdots = c_m = 1, c_{m+1} = c_{m+2} = \cdots = c_{m'} = 2, c_{m'+1} = c_{m'+2} = \cdots = c_n = 3$, and consider a two way table $\Theta_{i,j}$ with n rows. Rows $i = 1, \ldots, m$ each contain exactly two RVs, namely, $U_{p_i,t_i} - \psi_{p_i}$ and $V_{p_i,t_i} - \omega_{p_i}$ (note that the product of the RVs in row i is D_{1,p_i,t_i}). The remaining $n-m$ rows contain one RV each, namely, $U_{p_i,t_i} - \psi_{p_i}$ (which is proportional to D_{2,p_i,t_i}) for $i = (m+1), \ldots, m'$, and $V_{p_i,t_i} - \omega_{p_i}$ (proportional to D_{3,p_i,t_i}) for $i = m'+1, \ldots, n$. Theorem 6 says $\mathrm{cum}(D_{c_1,p_1,t_1}, \ldots, D_{c_n,p_n,t_n})$ is proportional to $\sum_\chi \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_1) \cdots \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_r)$. We complete the proof by showing that for any fixed χ

$$\sum_{t_1,\ldots,t_n} \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_1) \cdots \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_r) = o(M^{n/2}). \tag{20}$$

We prove the above in the following steps.

Step 1 Since $\Theta_{i,j}$ is centered, its first order cumulant is zero, so we can restrict ourselves to cases where $|\chi_k| \geq 2$ for all k. If any group of RVs in $\Theta_{i,j} : (i,j) \in \chi_k$ is independent of the remaining RVs in that set, then $\mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi_k) = 0$. Since the $U_{p_i,t_i} - \psi_{p_i}$'s and $V_{p_i,t_i} - \omega_{p_i}$'s are independent, we need only consider χ_k containing either just $U_{p_i,t_i} - \psi_{p_i}$'s or just $V_{p_i,t_i} - \omega_{p_i}$'s.

Step 2 Consider $m = 0$. In this case each row in $\Theta_{i,j}$ has only one RV, and thus all of $\Theta_{i,j}$ together form the only indecomposable partition $\chi = \chi_1$. Now if $m' = 0$, then by Assumption 1

$$\sum_{t_1,\ldots,t_n} \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi)$$
$$= \sum_{t_1,\ldots,t_n} \mathrm{cum}(V_{p_1,t_1} - \omega_{p_1}, \ldots, V_{p_n,t_n} - \omega_{p_n}) = o(M^{n/2}).$$

On the other hand if $m' = n$, then by Lemma 2

$$\sum_{t_1,\ldots,t_n} \mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi)$$
$$= \sum_{t_1,\ldots,t_n} \mathrm{cum}(U_{p_1,t_1} - \omega_{p_1}, \ldots, U_{p_n,t_n} - \omega_{p_n}) = o(M^{n/2}).$$

Finally we rule out the case $1 \leq m' < n$ because then χ contains both $U_{p_i,t_i} - \psi_{p_i}$'s and $V_{p_i,t_i} - \omega_{p_i}$'s and hence $\mathrm{cum}(\Theta_{i,j} : (i,j) \in \chi) = 0$.

Step 3 Finally consider $m \geq 1$. Assume that $\chi_1, \ldots, \chi_q$ partition the RVs $\{\Theta_{1,1}, \ldots, \Theta_{m',1}\}$ (these are all $U_{p_i,t_i} - \psi_{p_i}$) and that $\chi_{q+1}, \ldots, \chi_r$ partition the random variables $\{\Theta_{1,2}, \ldots, \Theta_{m,2}, \Theta_{m'+1,1}, \ldots, \Theta_{n,1}\}$ (these are all $V_{p_i,t_i} - \omega_{p_i}$). To check that (20) holds, we need to consider five cases.

Case 1 When $m' > m$ we sum over $t_{m+1}, \ldots, t_{m'}$ in the left hand side of (20) and use (19). In order to keep track of all the individual t_i for which $(i, 1)$ belongs to χ_k for $k = 1, \ldots, q$, we set $0 = \rho_0 \le \rho_1 \le \cdots \le \rho_q = m$, $m = \sigma_0 \le \sigma_1 \le \cdots \le \sigma_q = m'$ and assume, for $k = 1, \ldots, q$, $\chi_k = \{(\rho_{k-1} + 1, 1), \ldots, (\rho_k, 1), (\sigma_{k-1} + 1, 1), \ldots, (\sigma_k, 1)\}$. Then, for $k = 1, \ldots, q$, we obtain by Lemma 4

$$\begin{aligned}
&\sum_{t_{\sigma_{k-1}+1}, \ldots, t_{\sigma_k}} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_k) \\
&\quad = \sum_{t_{\sigma_{k-1}+1}, \ldots, t_{\sigma_k}} \kappa_{\rho_k + \sigma_k - \rho_{k-1} - \sigma_{k-1}}(p_i, t_i : (i, 1) \in \chi_k) \\
&\quad = M^{(\sigma_k - \sigma_{k-1} - 1)^+/2} \kappa_{\rho_k + \sigma_k - \rho_{k-1} - \sigma_{k-1}}(p_i, : (i, 1) \in \chi_k, t_{\rho_{k-1}+1}, \ldots, t_{\rho_k}).
\end{aligned}$$

Now boundedness of $\kappa_{\rho_k + \sigma_k - \rho_{k-1} - \sigma_{k-1}}(p_i, (i, 1) \in \chi_k, t_{\rho_{k-1}+1}, \ldots, t_{\rho_k})$ yields

$$\begin{aligned}
&M^{-(m'-m-1)/2} \sum_{t_1, \ldots, t_n} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_1) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_r) \\
&\propto \sum_{t_1, \ldots, t_m} \sum_{t_{m'+1}, \ldots, t_n} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_{q+1}) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_r) = o(M^{n/2}).
\end{aligned}$$

The last equality follows from Assumption 1.

Case 2 If $m' = m$ and $|\chi_k| > 2$ for some k in $q+1, \ldots, r$, then using the boundedness of $\operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_{k'})$ for $k' = 1, \ldots, q$, we obtain

$$\begin{aligned}
&\sum_{t_1, \ldots, t_n} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_1) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_r) \\
&\quad \le C_0 \sum_{t_1, \ldots, t_n} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_{q+1}) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_r) = o(M^{n/2}).
\end{aligned}$$

In the above C_0 is a constant and the last equality follows from Assumption 1.

Case 3 Consider $|\chi_k| = 2$ for $k = q + 1, \ldots, r$ and assume $m' = m$. Clearly $2m > n > m$ and $r - q = n - m$. Let $(m + i, 1)$ be contained in χ_{q+i} for $i = 1, \ldots, n$. We sum over $t_{m+1}, \ldots, t_n$ to obtain

$$\begin{aligned}
&\sum_{t_1, \ldots, t_n} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_1) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_r) \\
&\le \text{constant} \sum_{t_1, \ldots, t_m} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_1) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_q) \\
&\qquad \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_{q+n-m+1}) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_r) \\
&\le \text{constant} \sum_{t_1, \ldots, t_m} \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_1) \cdots \operatorname{cum}(\Theta_{i,j} : (i, j) \in \chi_q) = o(M^{n/2}).
\end{aligned}$$

In the above derivation we need the fact that $\sum_{t_i} \mathrm{cum}(V_{p_i,t_i}, V_{p_\tau,t_\tau})$ is bounded and note that the constant is changing from line to line.

Case 4 Consider the case $n = m$. Again if any $|\chi_k| > 2$ for $k = 1, \ldots, q$, we are done by using (18) along with the fact that cumulants of $V_{p_i,t_i} - \omega_{p_i}$ are bounded.

Case 5 The last case is $n = m$ and $|\chi_k| = 2$ for all k. The proof requires Theorem 6 to write down the left hand side of (20) in terms of covariances of U_{p_i,t_i} and V_{p_i,t_i} and hinges on the fact that ACVS of U_{p_i,t_i} and V_{p_i,t_i} are absolutely summable.

Proof of Theorem 3 Take $U_{p,t} = -(X_{t-l} - X_{t-l'})^2/2$, $V_{p,t} = \delta_{t-l}\delta_{t-l'}$ and $a_p = h_{j,l}h_{j,l'}\beta_{l-l'}$, where $p = (l, l')$. Use Proposition 2 and Lemma 3 in respective places of Proposition 1 and Lemma 2 and complete the proof as in Theorem 2 by checking all the steps. □

References

Beran, J. (1994). *Statistics for long-memory processes*. New York: Chapman and Hall.

Blackman, R.B., Tukey, J.W. (1958). *The measurement of power spectra*. New York: Dover Publications.

Breuer, P., Major, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *Journal of Multivariate Analysis 13*, 425–441.

Brillinger, D.R. (1981). *Time series*. Oakland: Holden-Day Inc.

Bronez, T.P. (1988). Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences. *IEEE Transactions on Acoustics, Speech, and Signal Processing 36*, 1862–1873.

Chiann, C., Morettin, P.A. (1998). *A wavelet analysis for time series. Nonparametric Statistics 10*, 1–46.

Craigmile, P.F., Percival, D.B. (2005). Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Transactions on Information Theory 51*, 1039–1048.

Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia: SIAM.

Feller, W. (1966). *An introduction to probability theory and its applications (Vol. II)*. New York: John Wiley.

Foster, G. (1996). Wavelets for period analysis of unevenly sampled time series. *The Astronomical Journal 112*, 1709–1729

Fox, R., Taqqu, M.S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probability Theory and Related Fields 74*, 213–240.

Frick, P., Grossmann, A., Tchamitchian, P. (1998). Wavelet analysis of signals with gaps. *Journal of Mathematical Physics 39*, 4091–4107.

Giraitis, L., Surgailis, D. (1985). CLT and other limit theorems for functionals of Gaussian processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 70*, 191–212.

Giraitis, L., Taqqu, M.S. (1998). Central limit theorems for quadratic forms with time-domain conditions. *The Annals of Probability 26*, 377–398.

Granger, C.W.J., Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–29.

Greenhall, C.A., Howe, D.A., Percival, D.B. (1999). Total variance, an estimator of long-term frequency stability. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 46*, 1183–1191.

Ho, H.C., Sun, T.C. (1987). A central limit theorem for noninstantaneous filters of a stationary Gaussian process. *Journal of Multivariate Analysis 22*, 144–155.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika 68*, 165–176.

Labat, D., Ababou, R., Mangin, A. (2001). Introduction of wavelet analyses to rainfall/runoffs relationship for a karstic basin: the case of licq–atherey karstic system (France). *Ground Water 39*, 605–615.

Lark, R.M., Webster, R. (2001). Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet transform. *European Journal of Soil Science 52*, 547–562.

Massel, S.R. (2001). Wavelet analysis for processing of ocean surface wave records. *Ocean Engineering 28*, 957–987.

Mondal, M., Percival, D.B. (2008). Wavelet analysis of variance for time series with missing values. Technical Report no. 535, Department of Statistics, University of Washington.

Nason, G.P., von Sachs, R. (1999). Wavelets in time series analysis. *Philosophical Transactions of the Royal Society London Series A 357*, 2511–2526.

Nason, G.P., von Sachs, R., Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society Series B Methodological 62*, 271–292.

Palma, W., Del Pino, G. (1999). Statistical analysis of incomplete long-range dependent data. *Biometrika 86*, 965–972.

Pelgrum, H., Schmugge, T., Rango, A., Ritchie, J., Kustas, B. (2000). Length-scale analysis of surface albedo, temperature, and normalized difference vegetation index in desert grassland. *Water Resources Research 36*, 1757–1766.

Percival, D.B. (1995). On estimation of the wavelet variance. *Biometrika 82*, 619–631.

Percival, D.B., Walden, A.T. (2000). *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press.

Pichot, V., Gaspoz, J.M., Molliex, S., Antoniadis, A., Busso, T., Roche, F., Costes, F., Quintin, L., Lacour, J.R., Barthélémy, J.C. (1999). Wavelet transform to quantify heart rate variability and to assess its instantaneous changes. *Journal of Applied Physiology 86*, 1081–1091.

Rybák, J., Dorotovič, I. (2002). Temporal variability of the coronal green-line index (1947–1998). *Solar Physics 205*, 177–187.

Serroukh, A., Walden, A.T., Percival, D.B. (2000). Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. *Journal of the American Statistical Association 95*, 184–196.

Stoica, P., Larsson, E.G., Li, J. (2000). Adaptive filter-bank approack to restoration and spectral analysis of gapped data. *The Astronomical Journal 163*, 2163–2173.

Sweldens, W. (1997). The lifting scheme: a construction of second generation wavelets. *SIAM Journal on Mathematical Analysis 19*, 511–546.

Torrence, C., Compo, G.P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society 79*, 61–78.

Toussoun, O. (1925). Mémoire sur l'histoire du Nil. *Mémoires a l'Institut d'Égypte 9*, 366–404.

Tsakiroglou, E., Walden, A.T. (2002). From Blackman–Tukey pilot estimators to wavelet packet estimators: a modern perspective on an old spectrum estimation idea. *Signal Processing 82*, 1425–1441.

Vio, R., Strohmer, T., Wamsteker, W. (2000). On the reconstruction of irregularly sampled time series. *Publications of the Astronomical Society of the Pacific 112*, 74–90.

Whitcher, B.J., Guttorp, P., Percival, D.B. (2000). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research 105*, 14941–14962.

Whitcher, B.J., Byers, S.D., Guttorp, P., Percival, D.B. (2002). Testing for homogeneity of variance in time series: long memory, wavelets and the Nile river. *Water Resources Research 38*, 1054–1070.

Žurbenko, I.G. (1986). *The spectral analysis of time series*. Amsterdam: North-Holland.