A Wavelet Variance Primer

Donald B. Percival[1] and Debashis Mondal[2]

[1] Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195, USA, dbp@apl.washington.edu
[2] Department of Statistics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637, USA, mondal@galton.uchicago.edu

1 Introduction

The discrete wavelet transform (DWT), as formulated in the late 1980s by Daubechies (1988), Mallat (1988a, 1988b, 1988c) and others, has inspired extensive research into how to use this transform to study time series. One focus of this research has been on the wavelet variance (also called the wavelet spectrum). The wavelet variance decomposes the variance of a time series and hence provides an analysis of variance (ANOVA). The most widely used ANOVA technique in time series analysis is spectral analysis, which involves the Fourier-based spectral density function (SDF). The ANOVA that the wavelet variance provides is in many ways similar to that afforded by the SDF (Li and Oh, 2002); however, from a practitioner's point of view, there are key differences. The SDF is a decomposition of variance across a continuum of Fourier frequencies. Each component in the decomposition reflects the degree to which a time series resembles a sinusoid with a particular frequency. The wavelet variance differs in that it is a decomposition across a discrete set of scales. Roughly speaking, a scale is an interval (or span) of time over which a time series is averaged. The strength of each component in the decomposition measures how much variability there is between adjacently located averages associated with a particular scale. The concept of scale is distinct from that of period (the inverse of frequency). Both are measured in the same units, but period but does not involve averaging. Although it is possible to estimate an SDF indirectly via the wavelet variance (Tsakiroglou and Walden, 2002), the different interpretations that frequency and scale have make the ANOVA afforded by the wavelet variance more appealing than the one given by the SDF for interpreting certain time series. Examples of applications that have made use of the wavelet variance are extensive and include time series related to electroencephalographic sleep state patterns of infants (Chiann and Morettin, 1998), frequency instability of atomic clocks (Greenhall et al., 1999), rainfall/runoff relationships (Labat et al., 2001), variations in soil composition (Lark and Webster, 2001), ocean surface waves (Massel,

2001), surface albedo and temperature in desert grasslands (Pelgrum et al., 2000), heart rate variability (Pichot et al., 1999), stochastic fluctuations on accreting binary stars (Scargle et al. 1993), solar coronal activity (Rybák and Dorotovič, 2002) and the El Niño–Southern Oscillation (Torrence and Compo, 1998). In addition, in contrast to the Fourier transform, the DWT is localized in time, and hence the wavelet variance can be readily adapted for exploring processes that are locally stationary with time-varying SDFs (Nason et al., 2000) and for detecting inhomogeneities in time series (Whitcher et al., 2002).

The intent of this chapter is to provide a basic introduction to the wavelet variance, with an emphasis on its interpretation, its statistical properties and some recent extensions to the basic methodology. We start with an overview of the maximal overlap DWT (MODWT), which is the version of the DWT of most interest for formulating the wavelet variance (Sect. 2). The MODWT leads to a basic ANOVA of a time series, which we describe in Sect. 3. If we assume that the time series under analysis is a realization of an intrinsically stationary process (as defined in Sect. 4), we can define a theoretical wavelet variance and regard the descriptive statistics discussed in Sect. 3 as basic estimators of this variance. We discuss the fundamental statistical theory behind estimators of the wavelet variance in Sect. 5, following which we discuss estimators intended to handle special circumstances (gappy time series in Sect. 6.1 and series with aberrant observations in Sect. 6.2). We then describe two wavelet-based methodologies in Sect. 7, one for deducing the presence of power-law dependence in a time series, the other for defining a characteristic scale. In both cases the statistics that arise are qualitatively similar in that they combine wavelet variance estimates together across adjacent scales. We devote the penultimate section (8) to five real-world examples illustrating the methodology discussed in previous sections (two of these examples serve to briefly compare Fourier-based spectral analysis with the analysis afforded by the wavelet variance; see also Faÿ et al., 2009). We close with some concluding remarks in Sect. 9.

2 Maximal Overlap Discrete Wavelet Transform

The wavelet variance is based upon the maximal overlap wavelet transform (MODWT) of a time series, so we start with a discussion of this transform and its basic properties. The MODWT is closely related to transforms with a variety of names in the literature, including 'undecimated DWT' (Shensa, 1992, implemented via the 'à trous algorithm'), 'shift invariant DWT' (Beylkin, 1992; Lang et al., 1995), 'wavelet frames' (Unser, 1995), 'translation invariant DWT' (Coifman and Donoho, 1995; Liang and Parks, 1996; Del Marco and Weiss, 1997), 'stationary DWT' (Nason and Silverman, 1995), 'time invariant DWT' (Pesquet et al., 1996) and 'non-decimated DWT' (Bruce and Gao, 1996). For more details about the MODWT, we refer the reader to Percival and Walden (2000), from which we adopt the notation used below.

The starting point for the MODWT is a Daubechies wavelet filter $\{\tilde{h}_{1,l}, l = 0, 1, \ldots, L_1 - 1\}$, where we insist that $\tilde{h}_{1,0} \neq 0$ and $\tilde{h}_{1,L_1-1} \neq 0$ so that the filter has width L_1 (for technical reasons, this width must be even). We define $\tilde{h}_{1,l} = 0$ for $l < 0$ and $l \geq L_1$ for convenience. By definition, a Daubechies wavelet filter must satisfy three properties:

$$\sum_{l\in\mathbb{Z}} \tilde{h}_{1,l} = 0, \;\; \sum_{l\in\mathbb{Z}} \tilde{h}_{1,l}^2 = 1/2 \;\text{ and }\; \sum_{l\in\mathbb{Z}} \tilde{h}_{1,l}\tilde{h}_{1,l+2n} = 0, \quad n = \pm 1, \pm 2, \ldots, \quad (1)$$

where $\mathbb{Z}$ is the set of all integers. While it is easy to construct filters that satisfy the first two properties, the third (orthogonality to even shifts) is challenging. The simplest filter with all three properties is the Haar wavelet filter, which has width $L_1 = 2$ and filter coefficients $\tilde{h}_{1,0} = 1/2$ and $\tilde{h}_{1,1} = -1/2$. We denote the transfer function (i.e., discrete Fourier transform (DFT)) for $\{\tilde{h}_{1,l}\}$ by

$$\widetilde{H}_1(f) \equiv \sum_{l\in\mathbb{Z}} \tilde{h}_{1,l} e^{-i2\pi f l}, \quad -\infty < f < \infty,$$

and its associated squared gain function by $\widetilde{\mathcal{H}}_1(f) \equiv |\widetilde{H}_1(f)|^2$. For the Haar wavelet filter, we have

$$\widetilde{H}_1(f) = \tfrac{1}{2} - \tfrac{1}{2} e^{-i2\pi f} \;\text{ and }\; \widetilde{\mathcal{H}}_1(f) = \sin^2(\pi f). \quad (2)$$

The wavelet filter spawns a complementary filter known as the scaling filter, defined by

$$\tilde{g}_{1,l} = (-1)^{l+1} \tilde{h}_{1,L_1-l-1}, \quad l \in \mathbb{Z}.$$

In the Haar case, we have $\tilde{g}_{1,0} = 1/2$, $\tilde{g}_{1,1} = 1/2$ and $\tilde{g}_{1,l} = 0$ otherwise. We denote the corresponding transfer and squared gain functions by $\widetilde{G}_1(\cdot)$ and $\widetilde{\mathcal{G}}_1(\cdot)$. For the Haar scaling filter, we have

$$\widetilde{G}_1(f) = \tfrac{1}{2} + \tfrac{1}{2} e^{-i2\pi f} \;\text{ and }\; \widetilde{\mathcal{G}}_1(f) = \cos^2(\pi f).$$

While the wavelet filter is a high-pass filter with a nominal pass-band defined by $1/4 \leq |f| \leq 1/2$, the scaling filter is a low-pass filter with pass-band dictated by $0 \leq |f| \leq 1/4$. A fundamental (but far from obvious) consequence of conditions (1), is that the squared gain functions for the wavelet and scaling filters must satisfy

$$\widetilde{\mathcal{H}}_1(f) + \widetilde{\mathcal{G}}_1(f) = 1 \;\text{ for all } f. \quad (3)$$

An implication of the above is that applying both filters to a time series results in outputs that preserve the content of the original series over all Fourier frequencies. Note that the above relationship holds in the Haar case because well-known identity $\sin^2(x) + \cos^2(x) = 1$.

Let $\{X_t, t = 0, 1, \ldots, N-1\}$ represent a time series of N observations regularly sampled in time; i.e., the time associated with X_t is $t_0 + t\,\Delta$, where t_0 is the time at which X_0 is observed, and Δ is the sampling interval between

adjacent observations. Upon circularly filtering $\{X_t\}$ with the wavelet and scaling filters, we obtain the unit-level MODWT wavelet coefficients

$$\widetilde{W}_{1,t} \equiv \sum_{l=0}^{L_1-1} \tilde{h}_{1,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1,$$

and corresponding scaling coefficients

$$\widetilde{V}_{1,t} \equiv \sum_{l=0}^{L_1-1} \tilde{g}_{1,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1.$$

The modulo operator in the above is such that '$t-l \bmod N$' is equal to $t-l$ if $0 \le t-l \le N-1$; otherwise, it is equal to $t-l+nN$, where n is the unique integer such that $0 \le t-l+nN \le N-1$. This operator in effect ties the beginning and end of the time series together, which is why the filtering is referred to as circular. For the Haar case, we have

$$\widetilde{W}_{1,t} = \frac{X_t - X_{t-1}}{2} \text{ and } \widetilde{V}_{1,t} = \frac{X_t + X_{t-1}}{2} \text{ for } t = 1, 2, \ldots N-1,$$

whereas

$$\widetilde{W}_{1,0} = \frac{X_0 - X_{N-1}}{2} \text{ and } \widetilde{V}_{1,0} = \frac{X_0 + X_{N-1}}{2}. \tag{4}$$

With the exception of $t = 0$, each scaling coefficient is the average of two adjacent values from the time series. We associate these averages with a standardized scale $\lambda_1 = 2$ and a physical scale of $\lambda_1 \Delta$. By contrast, each wavelet coefficient is proportional to the difference between two adjacent values. If we take the point of view that each X_t spans the time interval Δ (as would be appropriate if X_t were to represent, e.g., the average annual temperature at a particular spot on the earth so that $\Delta = 1$ year), then we can regard each wavelet coefficient as being proportional to the difference of averages over a standardized scale of $\tau_1 = 1$ and a physical scale of $\tau_1 \Delta$.

We can also interpret the MODWT scaling and wavelet coefficients as averages and differences between adjacently located averages when the transform is based on other Daubechies wavelet filters besides the Haar. For these other filters, each $\widetilde{V}_{1,t}$ at indices $t = L_1 - 1, \ldots, N-1$ is related to a *weighted* localized average of the time series over a scale of 2Δ, where we now take 2Δ to be a measure of the effective width of the weighted average. The wavelet coefficients $\widetilde{W}_{1,t}$ are related to changes in adjacent *weighted* localized averages over a scale of Δ. Our ability to make these interpretations requires that $\{\tilde{h}_l\}$ satisfies certain conditions above and beyond those imposed by Equation (1). For example, the regularity conditions that lead to $\{\tilde{h}_{1,l}\}$ having a squared gain function of

$$\widetilde{\mathcal{H}}_1(f) = \sin^{L_1}(\pi f) \sum_{l=0}^{\frac{L_1}{2}-1} \binom{\frac{L_1}{2} - 1 + l}{l} \cos^{2l}(\pi f) \tag{5}$$

allow us to attach these interpretations to $\widetilde{V}_{1,t}$ and $\widetilde{W}_{1,t}$. Note that the above collapses to Equation (2) in the Haar case ($L_1 = 2$). The squared gain functions for the widely used 'least asymmetric' Daubechies wavelet filters take the above form.

It is important to note that the wavelet and scaling coefficients with indices $t = 0, 1, \ldots, L_1 - 2$ do *not* involve localized averages. Instead they combine values from both the beginning and end of the time series. We refer to these special cases as 'boundary' coefficients, to which we will need to pay special attention. In the Haar case, the unit-level boundary coefficients are shown in Equation (4).

Just as the unit-level wavelet coefficients are related to differences of averages at scale $\tau_1 = 1$ while the scaling coefficients extract averages from $\{X_t\}$ at scale $\lambda_1 = 2$, higher-level MODWT coefficients extract quantities with similar interpretations for larger scales $\tau_j = 2^{j-1}$ and $\lambda_j = 2^j$, where j is the level index. We define the level $j > 1$ coefficients in terms of the higher-level wavelet filter $\{\tilde{h}_{j,l}, l = 0, 1, \ldots, L_j - 1\}$ and scaling filter $\{\tilde{g}_{j,l}, l = 0, 1, \ldots, L_j - 1\}$, where $L_j \equiv (2^j - 1)(L_1 - 1) + 1$. The appropriate definitions are

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j - 1} \tilde{h}_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1, \tag{6}$$

and

$$\widetilde{V}_{j,t} \equiv \sum_{l=0}^{L_{J_0} - 1} \tilde{g}_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1. \tag{7}$$

For the Haar case, we have

$$\widetilde{W}_{j,t} = \frac{1}{2^j}\left(\sum_{l=0}^{2^{j-1}-1} X_{t-l} - \sum_{l=0}^{2^{j-1}-1} X_{t-l-2^{j-1}} \right) \text{ and } \widetilde{V}_{j,t} = \frac{1}{2^j} \sum_{l=0}^{2^j-1} X_{t-l}$$

for $t = 2^{j-1}, 2^{j-1} + 1, \ldots N - 1$. The scaling coefficients are averages over scale $\lambda_j = 2^j$, whereas the wavelet coefficients are proportional to differences of adjacent averages over scale $\tau_j = 2^{j-1}$. For wavelets other than the Haar, the higher-level filters depend upon just the basic wavelet and scaling filters $\{h_{1,l}\}$ and $\{g_{1,l}\}$ and are most easily described in terms of inverse DFTs of their transfer functions. The transfer functions for $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ are given by

$$\widetilde{H}_j(f) \equiv \widetilde{H}_1(2^{j-1} f) \prod_{l=0}^{j-2} \widetilde{G}_1(2^l f) \text{ and } \widetilde{G}_j(f) \equiv \prod_{l=0}^{j-1} \widetilde{G}_1(2^l f).$$

The higher-level wavelet filters are band-pass filters with nominal pass-bands dictated by $1/2^{j+1} \leq |f| \leq 1/2^j$, whereas $\{\tilde{g}_{j,l}\}$ is a low-pass filter with pass-band given by $0 \leq |f| \leq 1/2^{j+1}$. For future use, we let

$$\widetilde{\mathcal{H}}_j(f) \equiv |\widetilde{H}_j(f)|^2 \text{ and } \widetilde{\mathcal{G}}_j(f) \equiv |\widetilde{G}_j(f)|^2 \tag{8}$$

denote the corresponding squared gain functions.

Finally, we note that in practice the MODWT wavelet and scaling coefficients are not computed directly via (6) and (7), but rather via an efficient recursive procedure known as the pyramid algorithm (for pseudo-code describing this algorithm, see pp. 177–8 of Percival and Walden, 2000).

3 Analysis of Variance via the MODWT

Let $\boldsymbol{X}$, $\widetilde{\boldsymbol{W}}_j$ and $\widetilde{\boldsymbol{V}}_j$ be column vectors of dimension N whose tth elements are, respectively, X_t, $\widetilde{W}_{1,t}$ and $\widetilde{V}_{1,t}$. Let

$$\|\boldsymbol{X}\|^2 \equiv \sum_{t=0}^{N-1} X_t^2$$

be the square of the Euclidean norm of $\boldsymbol{X}$. We refer to $\|\boldsymbol{X}\|^2$ as the 'energy' in $\boldsymbol{X}$. A key point about the MODWT is that it is energy preserving, in the sense that

$$\|\boldsymbol{X}\|^2 = \|\widetilde{\boldsymbol{W}}_1\|^2 + \|\widetilde{\boldsymbol{V}}_1\|^2. \tag{9}$$

In general this decomposition of the energy into two parts follows from Parseval's theorem and Equation (3). In the Haar case it follows readily from

$$\widetilde{W}_{1,t}^2 + \widetilde{V}_{1,t}^2 = \frac{(X_t + X_{t-1})^2}{4} + \frac{(X_t - X_{t-1})^2}{4} = \frac{X_t^2 + X_{t-1}^2}{2},$$

$t = 1, \ldots, N-1$, along with a similar piece involving the boundary coefficients $\widetilde{W}_{1,0}$ and $\widetilde{V}_{1,0}$.

Letting $\overline{X} = \sum_{t=0}^{N-1} X_t/N$ represent the sample mean of $\boldsymbol{X}$, we can express the sample variance of our time series as

$$\begin{aligned}\hat{\sigma}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}\left(X_t - \overline{X}\right)^2 = \frac{1}{N}\|\boldsymbol{X}\|^2 - \overline{X}^2 &= \frac{1}{N}\|\widetilde{\boldsymbol{W}}_1\|^2 + \left(\frac{1}{N}\|\widetilde{\boldsymbol{V}}_1\|^2 - \overline{X}^2\right) \\ &\equiv \hat{\sigma}_{W_1}^2 + \hat{\sigma}_{V_1}^2,\end{aligned}$$

where $\hat{\sigma}_{W_1}^2$ and $\hat{\sigma}_{V_1}^2$ can be taken to be sample variances for $\widetilde{\boldsymbol{W}}_1$ and $\widetilde{\boldsymbol{V}}_1$ (the definition of the wavelet filter ensures that, if the X_t's have a population mean, then the population mean of the $\widetilde{W}_{1,t}$'s is zero under mild conditions; on the other hand, the sample mean of $\widetilde{\boldsymbol{V}}_1$ is always $\overline{X}$, as is easy to verify directly in the Haar case). We thus can break up the sample variance of $\boldsymbol{X}$ into two parts, one of which ($\hat{\sigma}_{W_1}^2$) is attributable to changes in the time series over standardized scale $\tau_1 = 1$, and the other ($\hat{\sigma}_{V_1}^2$), to averages in $\boldsymbol{X}$ over scale $\lambda_1 = 2\tau_1 = 2$; alternatively, we can think of $\hat{\sigma}_{W_1}^2$ and $\hat{\sigma}_{V_1}^2$ as capturing the parts of $\hat{\sigma}_X^2$ due to, respectively, high- and low-frequency fluctuations.

We can generalize the above scheme to define ANOVAs out to some maximum level $J_0 \geq 1$. Considering a level $J_0 = 2$ ANOVA first, the basic idea is to replace $\|\widetilde{\boldsymbol{V}}_1\|^2$ in Equation (9) with the sum of two values, namely, $\|\widetilde{\boldsymbol{W}}_2\|^2$ and $\|\widetilde{\boldsymbol{V}}_2\|^2$, the first of which is related to changes in adjacent weighted localized averages of $\{X_t\}$ over a scale of $\tau_2 = 2$, and the second, to weighted localized averages over a scale of $\lambda_2 = 2\tau_2 = 4$. By recursively replacing $\|\widetilde{\boldsymbol{V}}_{j-1}\|^2$ with $\|\widetilde{\boldsymbol{W}}_j\|^2 + \|\widetilde{\boldsymbol{V}}_j\|^2$, we are led to the level J_0 decomposition:

$$\|\boldsymbol{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\boldsymbol{W}}_j\|^2 + \|\widetilde{\boldsymbol{V}}_{J_0}\|^2 \text{ and } \hat{\sigma}_X^2 = \sum_{j=1}^{J_0} \hat{\sigma}_{W_j}^2 + \hat{\sigma}_{V_{J_0}}^2,$$

where

$$\hat{\sigma}_{W_j}^2 \equiv \frac{1}{N}\|\widetilde{\boldsymbol{W}}_j\|^2 \text{ and } \hat{\sigma}_{V_{J_0}}^2 \equiv \frac{1}{N}\|\widetilde{\boldsymbol{V}}_{J_0}\|^2 - \overline{X}^2. \tag{10}$$

We refer to $\hat{\sigma}_{W_j}^2$ as the jth level empirical wavelet variance and to $\sigma_{V_{J_0}}^2$ as level J_0 empirical scaling variance. The interpretation of $\hat{\sigma}_{W_j}^2$ is that it is related to changes in adjacent weighted localized averages of $\{X_t\}$ over a scale of $\tau_j = 2^{j-1}$, while $\sigma_{V_{J_0}}^2$ is associated with weighted localized averages over a scale of $\lambda_{J_0} = 2^{J_0}$. This generalization of Equation (9) again follows from Parseval's theorem, but this time in conjunction with a generalization of Equation (3), namely,

$$\sum_{j=1}^{J_0} \widetilde{\mathcal{H}}_j(f) + \widetilde{\mathcal{G}}_{J_0}(f) = 1 \text{ for all } f. \tag{11}$$

As a simple example of a wavelet-based ANOVA, consider a small segment of length $N = 192$ from a time series of subtidal sea-level fluctuations (Fig. 1(a)); see Percival and Mofjeld (1997) for details about these data. This segment is of interest because of several bumps, each spanning approximately 16 units of time. Figure 1(b) shows the empirical wavelet variances $\hat{\sigma}_{W_j}^2$ (circles) and empirical scaling variance $\hat{\sigma}_{V_{J_0}}^2$ (asterisk) based upon a level $J_0 = 7$ Haar MODWT. The sum of these eight variances is exactly equal to the sample variance $\hat{\sigma}_X^2 \doteq 258.6$ of the time series. The largest wavelet variance occurs at level $j = 5$, which corresponds to scale $\tau_5 = 2^4 = 16$. The peak at this scale quantifies what a visual inspection picks out, namely, features in the series (the bumps) with a characteristic span of 16 time units. The fact that the scaling variance at level $J_0 = 7$ is small relative to the displayed wavelet variances tells us that the bulk of the variance of the time series can be attributed to changes in averages over scales $\tau_7 = 2^6 = 64$ and smaller. The scale-based ANOVA given by the wavelet variance can thus offer an intuitively sensible explanation of how a time series is structured.

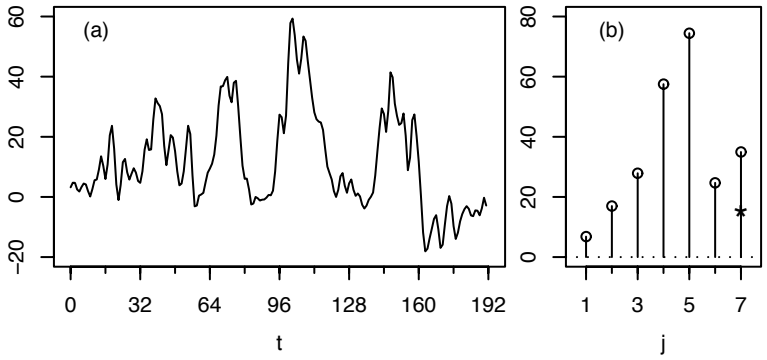

Fig. 1. Subtidal sea-level time series (a) and the Haar empirical wavelet variances (b) at levels $j = 1, \ldots, 7$ (circles) and empirical scaling variance for level 7 (asterisk). The seven levels correspond to scales $\tau_j = 2^{j-1}$, so the peak at level $j = 5$ is associated with changes on scale $\tau_5 = 16$. (Reflection boundary conditions were used in forming the MODWT – see Sect. 5.2 for details.)

4 Definition and Basic Properties of Wavelet Variance

In this section we formulate the wavelet variance for a dth order intrinsically stationary process $\{X_t : t \in \mathbb{Z}\}$, where $d \geq 0$ is an integer. The definition for such a process is as follows. If $d = 0$, then $\{X_t\}$ is just a second-order stationary process; i.e., its expected value $E\{X_t\}$ and covariances $\operatorname{cov}\{X_{t+\tau}, X_t\}$, $\tau \in \mathbb{Z}$, are finite and do not depend on t. For $d > 0$, let $\{X_t^{(d)}\}$ represent the dth order backward difference of $\{X_t\}$:

$$X_t^{(d)} \equiv \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}$$

(thus $X_t^{(1)} = X_t - X_{t-1}$, $X_t^{(2)} = X_t - 2X_{t-1} + X_{t-2}$ and so forth). Then $\{X_t^{(d)}\}$ is second-order stationary, but $\{X_t^{(d-1)}\}$, $\{X_t^{(d-2)}\}$, $\ldots$, $\{X_t^{(0)}\}$ are not, where we define $X_t^{(0)} = X_t$ for convenience. For example, suppose $\{Z_t\}$ is a white noise process, i.e., a sequence of uncorrelated random variables (RVs) with zero mean and finite variance. Then the nonstationary random walk process $X_t = \sum_{u=0}^{t} Z_t$ is first-order intrinsically stationary since its first difference is a stationary process. As a second example, suppose a and $b \neq 0$ are constants, and now let $X_t = a + bt + Z_t$ so that $\{X_t\}$ is nonstationary. Then $X_t^{(1)} = b + Z_t - Z_{t-1}$ is stationary, so $\{X_t\}$ is first-order intrinsically stationary. A special case of a dth order intrinsically stationary process is an ARIMA(p, d, q) process, which is the most widely used parametric model in time series analysis (see, e.g., Brockwell and Davis, 2002).

For use below, let $\{s_\tau^{(d)} : \tau \in \mathbb{Z}\}$ denote the autocovariance sequence (ACVS) for $\{X_t^{(d)}\}$:

$$s_\tau^{(d)} \equiv \operatorname{cov}\{X_{t+\tau}^{(d)}, X_t^{(d)}\} = E\{(X_{t+\tau}^{(d)} - \mu^{(d)})(X_t^{(d)} - \mu^{(d)})\},$$

where $\mu^{(d)} \equiv E\{X_t^{(d)}\}$. Under the assumption that $\{X_t^{(d)}\}$ has an SDF $S_{X^{(d)}}(\cdot)$, we can, when $d > 0$, define a generalized SDF for $\{X_t\}$ itself via

$$S_X(f) = \frac{S_{X^{(d)}}(f)}{[4\sin^2(\pi f)]^d} \tag{12}$$

(Yaglom, 1958); here $4\sin^2(\pi f)$ comes into play because it defines the squared gain function for a first-order backward difference filter (cf. Equation (2)).

Given a dth order intrinsically stationary process $\{X_t : t \in \mathbb{Z}\}$ with an SDF given by Equation (12) and a jth level wavelet filter $\{\tilde{h}_{j,l}, l = 0, 1, \ldots, L_j - 1\}$, we can define an associated wavelet coefficient process via

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}. \tag{13}$$

Under the assumptions that the associated unit-level wavelet filter has a squared gain function given by Equation (5) and that $L_1 \geq 2d$, the wavelet coefficient process is stationary with an SDF given by

$$S_j(f) \equiv \widetilde{\mathcal{H}}_j(f) S_X(f) = \frac{\widetilde{\mathcal{H}}_j(f) S_{X^{(d)}}(f)}{[4\sin^2(\pi f)]^d},$$

where $\widetilde{\mathcal{H}}_j(\cdot)$ is defined by Equation (8). The squared gain function $\widetilde{\mathcal{H}}_j(\cdot)$ depends upon $\widetilde{\mathcal{H}}_1(\cdot)$ of Equation (5), which can be interpreted as arising from an implicit cascade of two filters. The first is a backward difference filter of order $L_1/2$. The condition $L_1 \geq 2d$ thus ensures that the jth level wavelet filter has enough embedded differencing operations to transform $\{X_t\}$ into the stationary process $\{X_t^{(d)}\}$ (the second filter in the cascade transforms $\{X_t^{(d)}\}$ into $\overline{W}_{j,t}$). The jth level wavelet variance is just the variance of the stationary process $\{\overline{W}_{j,t}\}$:

$$\nu_X^2(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\} = \int_{-1/2}^{1/2} S_j(f)\, df.$$

If $d = 0$ so that $\{X_t\}$ is stationary, then

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

and hence the wavelet variance is a scale-based ANOVA for $\{X_t\}$, paralleling the empirical ANOVA for the sample variance described in Sect. 3. If $d > 0$,

the summation above diverges to infinity, which is a reasonable definition for the variance of certain (but not all) nonstationary processes with stationary differences.

For $d \geq 0$, the wavelet variance just depends on the ACVS $\{s_\tau^{(d)}\}$ for the stationary process $\{X_t^{(d)}\}$ at the heart of $\{X_t\}$:

$$\begin{aligned} \nu_X^2(\tau_j) &= \sum_{l=0}^{L_j-d-1} \sum_{m=0}^{L_j-d-1} \tilde{b}_{j,l}^{(d)} \tilde{b}_{j,m}^{(d)} s_{l-m}^{(d)} \\ &= s_0^{(d)} \sum_{l=0}^{L_j-d-1} \left(\tilde{b}_{j,l}^{(d)}\right)^2 + 2 \sum_{\tau=1}^{L_j-d-1} s_\tau^{(d)} \sum_{l=0}^{L_j-d-1-\tau} \tilde{b}_{j,l}^{(d)} \tilde{b}_{j,l+\tau}^{(d)}, \end{aligned} \tag{14}$$

where $\{\tilde{b}_{j,l}^{(d)}\}$ is the dth-order cumulative summation of $\{\tilde{h}_{j,l}\}$; i.e., with $\tilde{b}_{j,l}^{(0)} \equiv \tilde{h}_{j,l}$, we have, for $k = 1, \ldots, d$,

$$\tilde{b}_{j,l}^{(k)} = \sum_{n=0}^{l} \tilde{b}_{j,n}^{(k-1)}, \quad l = 0, 1, \ldots, L_j - k - 1$$

(Lemma 1, Craigmile and Percival, 2005). This expression for $\nu_X^2(\tau_j)$ follows from reexpressing Equation (13) in terms of $\{\tilde{b}_{j,l}^{(d)}\}$ and $\{X_t^{(d)}\}$:

$$\overline{W}_{j,t} = \sum_{l=0}^{L_j-d-1} \tilde{b}_{j,l}^{(d)} X_{t-l}^{(d)}. \tag{15}$$

The above leads directly to an expression for the ACVS for $\{\overline{W}_{j,t}\}$, namely,

$$s_{j,\tau} \equiv \operatorname{cov}\{\overline{W}_{j,t+\tau}, \overline{W}_{j,t}\} = \sum_{l=0}^{L_j-d-1} \sum_{m=0}^{L_j-d-1} \tilde{b}_{j,l}^{(d)} \tilde{b}_{j,m}^{(d)} s_{\tau+l-m}^{(d)}.$$

Equation (14) follows from the above since $\nu_X^2(\tau_j) = s_{j,0}$.

When $d = 0$ or 1, we can also express the wavelet variance in terms of the semi-variogram, defined as $\gamma_\tau = \frac{1}{2} \operatorname{var}\{X_\tau - X_0\}$. We then have

$$\nu_X^2(\tau_j) = -\sum_{l=0}^{L_j-1} \sum_{m=0}^{L_j-1} \tilde{h}_{j,l} \tilde{h}_{j,m} \gamma_{l-m}.$$

5 Basic Estimators of the Wavelet Variance

We now consider the problem of estimating the wavelet variance given a time series that can be regarded as a realization of a portion X_0, X_1, …, X_{N-1} of a dth order intrinsically stationary process. Under the assumption that

$L_1 \geq 2d$, we can base an estimator for $\nu_X^2(\tau_j)$ on the level j MODWT wavelet coefficients $\widetilde{W}_{j,t}$ of Equation (6). If we compare

$$\widetilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N} \text{ with } \overline{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l} \text{ for } t = 0, 1, \ldots, N-1,$$

we see that $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ when $t \geq L_j - 1$, but this equality need not hold when $0 \leq t < L_j - 1$, i.e., when $\widetilde{W}_{j,t}$ is a boundary coefficient. As discussed in the next two subsections, excluding the boundary coefficients leads us to an unbiased estimator of the wavelet variance, whereas, with certain modifications, we can form an attractive biased estimator that makes use of all available coefficients.

5.1 Unbiased estimators of the wavelet variance

Under the assumption that $E\{\overline{W}_{j,t}\} = 0$ and that $M_j \equiv N - L_j + 1 > 0$, an unbiased estimator of $\nu_X^2(\tau_j)$ is given by

$$\hat{\nu}_X^2(\tau_j) \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2. \tag{16}$$

An unbiased estimator is not possible in general without the condition $E\{\overline{W}_{j,t}\} = 0$. As can be seen from Equation (15), this condition will hold if $E\{X_t^{(d)}\} = 0$, which in general cannot be guaranteed; however, no matter what $E\{X_t^{(d)}\}$ is, the condition will hold as long as

$$\sum_{l=0}^{L_j-d-1} \tilde{b}_{j,l}^{(d)} = 0,$$

which we *can* guarantee by assuming $L_1 > 2d$ rather than just $L_1 \geq 2d$. The zero mean condition is thus easy to achieve by just increasing the length of the basic wavelet filter.

The large sample distribution for $\hat{\nu}_X^2(\tau_j)$ is tractable if we make some additional assumptions about the wavelet coefficient process. One pathway is to assume that $\{\overline{W}_{j,t}\}$ is a stationary Gaussian (normal) process with a square summable ACVS, i.e.,

$$A_j \equiv \sum_{\tau=-\infty}^{\infty} s_j^2 < \infty. \tag{17}$$

With this assumption, it follows that

$$\frac{M_j^{1/2}(\hat{\nu}_X^2(\tau_j) - \nu_X^2(\tau_j))}{(2A_j)^{1/2}} \stackrel{\text{d}}{=} \mathcal{N}(0,1) \tag{18}$$

asymptotically, where '$\stackrel{\mathrm{d}}{=}$' stands for 'is equal in distribution to', and $\mathcal{N}(0,1)$ is a standard Gaussian RV (Mondal, 2007, gives a succinct proof of the above based upon Theorem 5 of Giraitis and Surgailis, 1985, and discusses earlier – but more complicated and less general – proofs in Percival, 1983, for the Haar wavelet and in Percival, 1995, for general Daubechies wavelet filters). Accordingly the random interval

$$\left[\hat{\nu}_X^2(\tau_j) - \Phi^{-1}(1-p)\left(\frac{2A_j}{M_j}\right)^{1/2},\ \ \hat{\nu}_X^2(\tau_j) + \Phi^{-1}(1-p)\left(\frac{2A_j}{M_j}\right)^{1/2}\right] \quad (19)$$

constitutes an approximate $100(1-2p)\%$ confidence interval (CI) for $\nu_X^2(\tau_j)$, where $\Phi^{-1}(p)$ is the $p \times 100\%$ percentage point for the standard Gaussian distribution.

The lower limit of the CI displayed in (19) is not restricted to be positive even though the true wavelet variance is. This fact poses a problem if we adopt the common practice of plotting estimates $\hat{\nu}_X^2(\tau_j)$ and their associated CIs on a logarithmic scale. An alternative – but asymptotically equivalent – approach that yields CIs with positive lower limits is to assume that

$$\frac{\eta_j \hat{\nu}_X^2(\tau_j)}{\nu_X^2(\tau_j)} \stackrel{\mathrm{d}}{=} \chi_{\eta_j}^2, \quad (20)$$

where $\chi_{\eta_j}^2$ is a chi-square RV with η_j degrees of freedom. We can set η_j using a moment matching scheme. Recalling first that $E\{\chi_{\eta_j}^2\} = \eta_j$ and $\mathrm{var}\,\{\chi_{\eta_j}^2\} = 2\eta_j$ so that

$$\frac{2(E\{c\chi_{\eta_j}^2\})^2}{\mathrm{var}\,\{c\chi_{\eta_j}^2\}} = \eta_j \ \text{ for any constant } c\ ,$$

we use the facts that $E\{\hat{\nu}_X^2(\tau_j)\} = \nu_X^2(\tau_j)$ and $\mathrm{var}\,\{\hat{\nu}_X^2(\tau_j)\} \approx 2A_j/M_j$ to obtain

$$\eta_j = \frac{2\left(E\{\hat{\nu}_X^2(\tau_j)\}\right)^2}{\mathrm{var}\,\{\hat{\nu}_X^2(\tau_j)\}} = \frac{2\nu_X^4(\tau_j)}{\mathrm{var}\,\{\hat{\nu}_X^2(\tau_j)\}} \approx \frac{M_j \nu_X^4(\tau_j)}{A_j}. \quad (21)$$

The random interval

$$\left[\frac{\eta_j \hat{\nu}_X^2(\tau_j)}{Q_{\eta_j}(1-p)}, \frac{\eta_j \hat{\nu}_X^2(\tau_j)}{Q_{\eta_j}(p)}\right] \quad (22)$$

is then an approximate $100(1-2p)\%$ CI for $\nu_X^2(\tau_j)$, where $Q_{\eta_j}(p)$ is the $p \times 100\%$ percentage point for the $\chi_{\eta_j}^2$ distribution.

We must know A_j to form the CIs of Equations (19) and (22). If we regard $\widetilde{W}_{j,L_j-1}, \ldots, \widetilde{W}_{j,N-1}$ as a time series whose mean value is zero, then we can estimate its ACVS via

$$\hat{s}_{j,\tau} \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-|\tau|-1} \widetilde{W}_{j,t} \widetilde{W}_{j,t+|\tau|}, \ \ 0 \le |\tau| \le M_j - 1.$$

An approximately unbiased estimator of A_j is given by

$$\hat{A}_j \equiv \sum_{\tau=-(M_j-1)}^{M_j-1} \frac{\hat{s}^2_{j,\tau}}{2} = \frac{\hat{s}^2_{j,0}}{2} + \sum_{\tau=1}^{M_j-1} \hat{s}^2_{j,\tau} = \frac{\hat{\nu}^4(\tau_j)}{2} + \sum_{\tau=1}^{M_j-1} \hat{s}^2_{j,\tau} \qquad (23)$$

(comparison of $\hat{A}_j$ with the definition of A_j in Equation (17) indicates a counter-intuitive division by two in the above – this is due to the moments of the χ^2_2 distribution, as explained in Percival and Walden, 2000, Section 8.4). We can plug this estimator into Equation (19) to get Gaussian-based approximate CIs. We can also plug $\hat{A}_j$ along with $\hat{\nu}^4(\tau_j)$ into Equation (21) to estimate η_j, which in turn can be used in Equation (22) to produce $\chi^2_{\eta_j}$-based approximate CIs. Monte Carlo studies indicate that CIs based on estimating A_j are reasonably accurate as long as $M_j \geq 128$. If there aren't enough wavelet coefficients to get a decent estimate of A_j, a fall-back for getting an approximate CI for $\nu^2(\tau_j)$ is to use Equation (22) with η_j set to $\max\{M_j/2^j, 1\}$. This approach banks on the fact that the wavelet filter $\{\tilde{h}_{j,l}\}$ is an approximate bandpass filter and hence that $\{\overline{W}_{j,t}\}$ should resemble a band-limited process. If the SDF for $\{\overline{W}_{j,t}\}$ is relatively flat over the pass-band, this alternative approach is viable, but tends to produce conservative CIs.

The large-sample theory described above is based on the assumption that $\{\widetilde{W}_{j,t}\}$ is Gaussian. If $\{X_t\}$ itself is Gaussian, then $\{\widetilde{W}_{j,t}\}$ must be Gaussian since each $\widetilde{W}_{j,t}$ is a linear combination of RVs in $\{X_t\}$. For certain non-Gaussian processes $\{X_t\}$, the assumption that $\{\widetilde{W}_{j,t}\}$ is approximately Gaussian is viable (particularly for large j) because linear filtering tends to induce Gaussianity (Mallows, 1967). If $\{\widetilde{W}_{j,t}\}$ cannot be regarded as approximately Gaussian, we can obtain a large sample approximation to the distribution of $\hat{\nu}^2(\tau_j)$ if we are willing to make certain assumptions (Serroukh et al., 2000). Let $\mathcal{M}^0_{-\infty}$ and $\mathcal{M}^\infty_n$ denote the σ-algebras generated by, respectively, $\{\ldots, \widetilde{W}_{j,-1}, \widetilde{W}_{j,0}\}$ and $\{\widetilde{W}_{j,n}, \widetilde{W}_{j,n+1}, \ldots\}$. For $n > 0$ define the mixing coefficient

$$\alpha_n = \sup_{A \in \mathcal{M}^0_{-\infty}, B \in \mathcal{M}^\infty_n} |P(A \cap B) - P(A)P(B)|\,,$$

where $P(A)$ is the probability of the event A. If $\{\widetilde{W}_{j,t}\}$ is strictly stationary with $E\{|\widetilde{W}_{j,t}|^{4+2\delta}\} < \infty$ for some $\delta > 0$, if $\sum_{n=1}^\infty \alpha_n^{\delta/(2+\delta)} < \infty$ and if the SDF $S_{\widetilde{W}^2_{j,t}}(f)$ for the process $\{\widetilde{W}^2_{j,t}\}$ is positive at zero frequency, then

$$\frac{M_j^{1/2}(\hat{\nu}^2_X(\tau_j) - \nu^2_X(\tau_j))}{S^{1/2}_{\widetilde{W}^2_{j,t}}(0)} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1) \qquad (24)$$

approximately for large M_j. The condition on the mixing coefficients implies that $\alpha_n \to 0$ as $n \to \infty$, in which case $\{\widetilde{W}_{j,t}\}$ is said to have strong mixing; i.e., its dependence is short range (Rosenblatt, 1985, pp. 62–3). The assumptions

we need for (24) to hold are thus restrictive, but not overly so (Serroukh et al., 2000, give specific examples of processes for which these assumptions hold).

As in the Gaussian case, we can reexpress Equation (24) in terms of a chi-square distribution as per Equation (20), with the degrees of freedom now given by

$$\eta_j \approx \frac{2M_j \nu_X^4(\tau_j)}{S_{\widetilde{W}_{j,t}^2}(0)} \tag{25}$$

To form the CI of Equation (22), we need in practice to estimate $S_{\widetilde{W}_{j,t}^2}(0)$ based upon $\widetilde{W}_{j,L_j-1}^2, \ldots, \widetilde{W}_{j,N-1}^2$. Serroukh et al. (2000) advocate a multitaper SDF estimator of order $K = 5$ based upon Slepian (discrete prolate spheroidal) data tapers $\{v_{k,t}\}$, $k = 0, 1, \ldots, K-1$, with design bandwidth set to $7/M_j$ (Thomson, 1982; Percival and Walden, 1993). Define

$$J_k(0) = \sum_{t=L_j-1}^{N-1} v_{k,t} \widetilde{W}_{j,t}^2, \; V_k(0) = \sum_{t=L_j-1}^{N-1} v_{k,t} \text{ and } \check{\nu}_X^2(\tau_j) = \frac{\sum_{k=0}^{K-1} J_k(0) V_k(0)}{\sum_{k=0}^{K-1} V_k^2(0)}. \tag{26}$$

Then the required estimator takes the form

$$\hat{S}_{\widetilde{W}_{j,t}^2}(0) = \frac{1}{K} \sum_{k=0}^{K-1} \left(J_k(0) - V_k(0) \check{\nu}_X^2(\tau_j) \right)^2. \tag{27}$$

Two comments are in order. First, computation of the above can be simplified by noting that $V_k(0) = 0$ for odd k. Second, as suggested by its notation, we can regard $\check{\nu}_X^2(\tau_j)$ as an estimator of the wavelet variance alternative to $\hat{\nu}_X^2(\tau_j)$ (both estimators are unbiased, but $\check{\nu}_X^2(\tau_j)$ is not constrained to be nonnegative, even though the exponent '2' would suggest otherwise).

5.2 Biased estimators of the wavelet variance

The unbiased estimator $\hat{\nu}_X^2(\tau_j)$ of the wavelet variance makes use of just the nonboundary wavelet coefficients from the MODWT. The number M_j of such coefficients decreases drastically as the level index j increases. For example, for a wavelet filter of width $L_1 = 8$ and a time series of length $N = 1024$, we find $M_j = 1017$, 1003, 975, 919, 807, 583 and 135 for levels $j = 1, \ldots, 7$, and there are no nonboundary coefficients at levels $j \geq 8$ (these correspond to scales $\tau_j \geq 128$). This fact motivates us to consider estimators of $\nu_X^2(\tau_j)$ that make use of the $N - \max\{M_j, 0\}$ boundary wavelet coefficients. One obvious candidate is

$$\hat{\sigma}_{W_j}^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2,$$

which we introduced in the context of forming a scaled-based ANOVA for a time series (see Equation (10)). This estimator is in general biased because

$E\{\widetilde{W}_{j,t}^2\}$ need not equal $\nu_X^2(\tau_j)$ when $t = 0, \ldots, L_j - 2$. If the underlying time series $\{X_t\}$ is a stationary process, then $E\{\hat{\sigma}_{W_j}^2\} \to \nu_X^2(\tau_j)$ as $N \to \infty$; however, if $\{X_t\}$ is first-order intrinsically stationary, then $\hat{\sigma}_{W_j}^2$ is not in generally asymptotically unbiased. The basic reason is that $E\{(X_0 - X_{N-1})^2\} \to \infty$, which means that we can expect to see an increasing mismatch between the beginning and end of the time series. This mismatch adversely impacts the boundary wavelet coefficients because they join together X_t's from both ends of the time series.

We can create a biased estimator that is asymptotically equivalent to $\hat{\nu}_X^2(\tau_j)$ for zeroth- and first-order intrinsically stationary processes by using wavelet coefficients obtained from an augmented version of our original time series $X_0, \ldots, X_{N-1}$. To do so, we append a time-reversed version of the series to the original X_t's to create a series of length $2N$:

$$X_0, X_1, \ldots, X_{N-2}, X_{N-1}, X_{N-1}, X_{N-2}, \ldots, X_1, X_0.$$

Denoting this series by $X'_0, \ldots, X'_{2N-1}$, we first note that its sample variance is identical to that for the original series $X_0, \ldots, X_{N-1}$ so that an ANOVA of the X'_t's is a useful surrogate for an ANOVA of the original X_t's. The MODWT of the X'_t's is given by

$$\widetilde{W}'_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X'_{t-l \bmod 2N}, \quad t = 0, 1, \ldots, 2N-1,$$

which we also refer to as the MODWT of the X_t's based upon reflection boundary conditions. Greenhall et al. (1999) proposed the wavelet variance estimator

$$\overleftrightarrow{\nu}_X^2(\tau_j) \equiv \frac{1}{2N} \sum_{t=0}^{2N-1} (\widetilde{W}'_{j,t})^2,$$

but only explored its statistical properties through limited computer experiments. Aldrich (2005) showed that, while this estimator is generally biased, it is asymptotically equivalent to $\hat{\nu}_X^2(\tau_j)$ for zeroth- and first-order intrinsically stationary processes (but not for second-order and higher). He also compared the mean squared errors of $\overleftrightarrow{\nu}_X^2(\tau_j)$ and $\hat{\nu}_X^2(\tau_j)$ for representative processes through both exact expressions and computer experiments and found the biased estimator to be superior to the unbiased estimator, particularly in cases where M_j is small relative to N.

6 Specialized Estimators of the Wavelet Variance

The basic estimators of the wavelet variance discussed in the previous section are predicated on certain assumptions that might not hold in practical situations. Here we consider estimation of the wavelet variance when faced with

departures from one of two assumptions. The first assumption is that our time series consists of N contiguous values; i.e., there are no missing values in the time series. The second is that our series does not suffer from contamination unrelated to the process of interest (i.e., from corrupted observations).

6.1 Estimation of wavelet variance for gappy time series

Suppose we use an automatic measuring system to record the temperature every day at noon at some outdoor location, with the intent of eventually collecting a time series of N temperature measurements regularly sampled in time. In practise, we might not achieve this goal for a variety of reasons (power outages, sporadic instrumentation failure, vandalism, etc.) so that we end up with gaps (missing values) in the collected series. The wavelet variance estimators we discussed in Sect. 5 presume a regularly sampled series. If we want to use one of these estimators, we are faced with the task of filling in missing values. There are a variety of methods for doing so, ranging from the simple (using the sample mean of the existing observations as a surrogate for the missing ones) to the sophisticated (developing a statistical model for the time series and using the model to fill in the gaps via conditional expectations or a stochastic interpolation scheme). Interpolation can work well for certain gappy time series (particular if both the number of gaps and the gap sizes are small), but can be problematic for others. Here we discuss two specialized wavelet variance estimators proposed by Mondal and Percival (2010) to handle gappy time series without having to resort to interpolation.

Let $\{\delta_t\}$ be a strictly stationary binary-valued process such that δ_t is zero or one according to whether X_t is missing or present (we assume that $E\{\delta_t\} > 0$ and that $\{\delta_t\}$ and $\{X_t\}$ are independent). Define

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

and, for $0 \le l \le L_j - 1$ and $0 \le l' \le L_j - 1$, let

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

which is an estimator of $\beta_{l-l'}^{-1}$. While $\beta_k^{-1} > 0$ necessarily, we must assume that $\hat{\beta}_{l,l'}^{-1} > 0$ for all l, l'. This assumption is restrictive because it might not hold for a time series with too many gaps (it does hold asymptotically almost surely). Define a covariance-type estimator by

$$\hat{u}_X^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,l'}\hat{\beta}_{l,l'}X_{t-l}X_{t-l'}\delta_{t-l}\delta_{t-l'} \qquad (28)$$

and a semi-variogram-type estimator by

$$\hat{v}_X^2(\tau_j) = -\frac{1}{2M_j} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,l'}\hat{\beta}_{l,l'}(X_{t-l} - X_{t-l'})^2 \delta_{t-l}\delta_{t-l'}. \quad (29)$$

Both $\hat{u}_X^2(\tau_j)$ and $\hat{v}_X^2(\tau_j)$ collapse to the usual unbiased estimator $\hat{\nu}_X^2(\tau_j)$ when $\delta_t = 1$ for all t (the gap-free case). In the presence of gaps, the expected value of both estimators is $\nu_X^2(\tau_j)$, but it should be carefully noted that neither estimator is guaranteed to be nonnegative for a gappy realization of $X_0, \ldots, X_{N-1}$.

Let us now consider the large sample theory for $\hat{u}_X^2(\tau_j)$ and $\hat{v}_X^2(\tau_j)$. Mondal and Percival (2010) show that, if $\{X_t\}$ is a stationary Gaussian process whose SDF $S_X(\cdot)$ is square integrable and if $\{\delta_t\}$ satisfies certain technical conditions in addition to strict stationarity, then $\hat{u}_X^2(\tau_j)$ is asymptotically Gaussian with mean $\nu_X^2(\tau_j)$ and large sample variance given by $S_{U_{j,t}^2}(0)/M_j$, where the numerator is the SDF at zero frequency for the stationary process

$$U_{j,t}^2 \equiv \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,l'}\beta_{l-l'}X_{t-l}X_{t-l'}\delta_{t-l}\delta_{t-l'}.$$

This process has a mean of $\nu_X^2(\tau_j)$ and collapses to $\overline{W}_{j,t}^2$ in the gap-free case (again, note carefully that $U_{j,t}^2$ can be negative for certain realizations). If we let

$$\widetilde{U}_{j,t}^2 \equiv \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,l'}\hat{\beta}_{l,l'}X_{t-l}X_{t-l'}\delta_{t-l}\delta_{t-l'}, \;\; t = L_j - 1, \ldots, N-1,$$

we can estimate $S_{U_{j,t}^2}(0)$ using the multitaper approach of Equations (26) and (27) with $J_k(0)$ redefined to be $\sum_t v_{k,t}\widetilde{U}_{j,t}^2$. On the other hand, if $\{X_t\}$ is a zeroth- or first-order intrinsically stationary Gaussian process such that $\sin^2(\pi f)S_X(f)$ is square integrable and if we make the same assumptions as before about $\{\delta_t\}$, then $\hat{v}_X^2(\tau_j)$ is asymptotically Gaussian with mean $\nu_X^2(\tau_j)$ and large sample variance given by $S_{V_{j,t}^2}(0)/M_j$, which, as before, involves an SDF at zero frequency, but this time for the stationary process

$$V_{j,t}^2 \equiv -\frac{1}{2} \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,l'}\beta_{l-l'}(X_{t-l} - X_{t-l'})^2 \delta_{t-l}\delta_{t-l'}.$$

Again, the above process has a mean of $\nu_X^2(\tau_j)$, collapses to $\overline{W}_{j,t}^2$ in the gap-free case and can be negative for certain realizations. Letting

$$\widetilde{V}_{j,t}^2 \equiv \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,l'}\hat{\beta}_{l,l'}(X_{t-l} - X_{t-l'})^2 \delta_{t-l}\delta_{t-l'}, \;\; t = L_j - 1, \ldots, N-1,$$

we can estimate $S_{V_{j,t}^2}(0)$ via Equations (26) and (27) with $J_k(0)$ now redefined to be $\sum_t v_{k,t}\widetilde{V}_{j,t}^2$. Mondal and Percival (2010) note that the Gaussianity

assumption on $\{X_t\}$ can be dropped, and both estimators will still have the same limiting distribution if we assume mixing conditions similar to what was needed to obtain the result stated in Equation (24).

Both $\hat{u}^2_X(\tau_j)$ and $\hat{v}^2_X(\tau_j)$ can handle stationary processes, but $\hat{v}^2_X(\tau_j)$ works also for first-order intrinsically stationary processes. It might seem we could dispense with $\hat{u}^2_X(\tau_j)$ in favor of $\hat{v}^2_X(\tau_j)$; however, for certain – but not all – stationary processes, $\hat{u}^2_X(\tau_j)$ proves to be more efficient asymptotically than $\hat{v}^2_X(\tau_j)$ as measured by the ratio $S_{V^2_{j,t}}(0)/S_{U^2_{j,t}}(0)$. There is thus a role for both estimators. We note, however, one important practical distinction between them. The semi-variogram-type estimator $\hat{v}^2_X(\tau_j)$ is invariant if we add a constant to the observed time series, whereas the covariance-type estimator $\hat{u}^2_X(\tau_j)$ is not. It is thus important to center a time series by subtracting off its sample mean prior to computing $\hat{u}^2_X(\tau_j)$.

6.2 Robust estimation of wavelet variance

The usual unbiased estimator of the wavelet variance is the sample mean of squared wavelet coefficients. In general, sample means as an estimator of a population mean are particularly sensitive to contamination, i.e., a small number of large values that do not reflect the statistical properties of the underlying process of interest. This fact has motivated the quest for robust alternatives to sample means that perform better in the presence of contamination. A simple robust estimator of $\nu^2_X(\tau_j)$ is the sample median of $\widetilde{W}^2_{j,L_j-1}, \ldots, \widetilde{W}^2_{j,N-1}$ after an adjustment to take into account the difference between the population mean $\nu^2_X(\tau_j)$ and the population median of the $\widetilde{W}^2_{j,t}$'s (Stoev et al., 2006). We can develop an appropriate statistical theory for a median-type estimator of the wavelet variance by considering

$$\widetilde{Q}_{j,t} \equiv \log(\widetilde{W}^2_{j,t}).$$

Because the log of the median of the $\widetilde{W}^2_{j,t}$'s is the same as the median of the $\widetilde{Q}_{j,t}$'s, a large sample theory based on the latter is pertinent for an estimator based the sample median of the $\widetilde{W}^2_{j,t}$'s. The advantage of the log transform is that it recasts the median-type estimator as a special case of the M-estimators pioneered by Huber (1964). These estimators work with location parameters, whereas $\nu^2_X(\tau_j)$ is a scale parameter, but one that can be recast as a location parameter via the log transform. Focusing on the case where $\{\widetilde{W}_{j,t}\}$ is Gaussian, it follows from Bartlett and Kendall (1946) that

$$E\{\widetilde{Q}_{j,t}\} = \log(\nu^2_X(\tau_j)) + \psi(\tfrac{1}{2}) + \log(2) \equiv \mu_j \;\text{ and }\; \operatorname{var}\{\widetilde{Q}_{j,t}\} = \psi'(\tfrac{1}{2}) = \frac{\pi^2}{2},$$

where ψ and ψ' are the di- and tri-gamma functions. These facts allow us to write

$$\widetilde{Q}_{j,t} = \mu_j + \epsilon_{j,t}, \tag{30}$$

where $E\{\epsilon_{j,t}\} = 0$ and $\text{var}\{\widetilde{Q}_{j,t}\} = \pi^2/2$. We can thus manipulate a location estimator for $\widetilde{Q}_{j,t}$ so that it becomes an estimator of $\nu_X^2(\tau_j)$ since

$$\nu_X^2(\tau_j) = \exp\left(\mu_j - \psi(\tfrac{1}{2}) - \log(2)\right)$$

(we would need different manipulations to handle non-Gaussian processes).

In general, an M-estimator for μ_j of Equation (30) is based upon a real-valued function $\varphi(\cdot)$ that is defined over the real axis $\mathbb{R}$ and satisfies certain technical conditions (see Mondal and Percival, 2011a, for details). The M-estimator is

$$\hat{\mu}_j \equiv \arg\min_{x \in \mathbb{R}} \left| \sum_{t=L_j-1}^{N-1} \varphi(\widetilde{Q}_{j,t} - x) \right|.$$

Let us specialize to the case $\varphi(x) = \text{sign}(x)$, for which the M-estimator becomes the sample median of the $\widetilde{Q}_{j,t}$'s. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density and distribution functions for a standard Gaussian RV, and let $\Phi^{-1}(\cdot)$ be the inverse of $\Phi(\cdot)$. Under the assumption that $\{\widetilde{W}_{j,t}\}$ is a zero-mean Gaussian process with an SDF that is square integrable, Mondal and Percival (2011a) show that $\hat{\mu}_j$ is asymptotically Gaussian with mean

$$\mu_{0,j} = \log\left(\nu_X^2(\tau_j)\right) + 2\log\left(\Phi^{-1}(\tfrac{3}{4})\right)$$

and large sample variance $S_\varphi(0)/(M_j C)$, which involves a constant $C = 4\left[\phi(\Phi^{-1}(\frac{3}{4}))\Phi^{-1}(\frac{3}{4})\right]^2$ and the SDF at zero frequency for the stationary process $\{\varphi(\widetilde{Q}_{j,t} - \mu_{0,j})\}$. We can estimate $S_\varphi(0)$ via the multitaper approach of Equations (26) and (27) by redefining $J_k(0)$ to be $\sum_t v_{k,t}\varphi(\widetilde{Q}_{j,t} - \hat{\mu}_j)$. Denoting this estimator by $\hat{S}_\varphi(0)$, it can be shown that an approximately unbiased and robust estimator of $\nu_X^2(\tau_j)$ is given by

$$\hat{r}_X^2(\tau_j) = \frac{\text{median}\{\widetilde{W}_{j,t}^2\} \cdot \exp\left(-\hat{S}_\varphi(0)/[2M_j C]\right)}{\left(\Phi^{-1}(\frac{3}{4})\right)^2}. \tag{31}$$

The above estimator is asymptotically normal with mean $\nu_X^2(\tau_j)$ and large sample variance $\nu_X^4(\tau_j)S_\varphi(0)/(M_j C)$. We can use this large sample theory to form CIs for $\nu_X^2(\tau_j)$ based upon the median-type estimator $\hat{r}_X^2(\tau_j)$. (Mondal and Percival, 2011a, provide theory paralleling the above for M-estimators other than the median.)

The median-type estimator $\hat{r}_X^2(\tau_j)$ guards against data contamination but is a less efficient estimator of $\nu_X^2(\tau_j)$ than the unbiased mean-type estimator $\hat{\nu}_X^2(\tau_j)$ when in fact the $\widetilde{W}_{j,t}$'s are free of contamination. Mondal and Percival (2011a) found that, for moderate sample sizes, $\hat{r}_X^2(\tau_j)$ has approximately twice the variance of $\hat{\nu}_X^2(\tau_j)$ over a selection of stationary processes encompassing both short- and long-range dependence. Thus, if the $\widetilde{W}_{j,t}$'s are truly Gaussian, we can expect $\hat{r}_X^2(\tau_j)$ to perform markedly poorer than $\hat{\nu}_X^2(\tau_j)$, but the presence of contamination can lead to the median-type estimator being preferred.

7 Combining Wavelet Variance Estimators Across Scales

In the previous two sections, we have presented a variety of wavelet variance estimators, which, in conjunction with their sampling theory, can be used to form, say, 95% CIs for the true wavelet variance $\nu_X^2(\tau_j)$. Taking the unbiased estimator as an example, it is conventional to plot the estimates $\hat{\nu}_X^2(\tau_j)$ and associated CIs versus standardized scale τ_j (or physical scale $\tau_j\,\Delta$) on log/log axes, in part because the estimates and CIs can range over many orders of magnitude and in part because scales increase by factors of two as the level index j increases. In addition to telling us how the variance of a time series is partitioned out across different scales, two patterns warranting further analysis often emerge in plots of $\log(\hat{\nu}_X^2(\tau_j))$ versus $\log(\tau_j)$. The first pattern is a stretch of scales over which $\log(\hat{\nu}_X^2(\tau_j))$ varies approximately linearly with $\log(\tau_j)$. Two explanations for this linear pattern are that the intrinsically stationary process $\{X_t\}$ exhibits long-range dependence or fractal fluctuations, both of which are special cases of a power-law variation. The second pattern is a peak at, say, scale τ_j; i.e., we have both $\log(\hat{\nu}_X^2(\tau_j)) > \log(\hat{\nu}_X^2(\tau_{j-1}))$ and $\log(\hat{\nu}_X^2(\tau_j)) > \log(\hat{\nu}_X^2(\tau_{j+1}))$. Since the log transform preserves order, such a peak indicates a tendency of $\{X_t\}$ to have fluctuations over a so-called characteristic scale in the neighborhood of τ_j. In the subsections below, we look at methods for quantifying power-law variations and characteristic scales based upon the wavelet variance. Both methods involve statistics that combine $\log(\hat{\nu}_X^2(\tau_j))$ across adjacent scales. In preparation for delving into the sampling properties of these statistics, here we give some background on the statistical properties of $\log(\hat{\nu}_X^2(\tau_j))$ (for simplicity, we focus on the unbiased estimator $\hat{\nu}_X^2(\tau_j)$, but the other estimators that we have discussed can be used instead with appropriate adjustments).

Under the assumption that $\hat{\nu}_X^2(\tau_j)$ obeys a chi-square distribution when properly normalized as per Equation (20), we can write

$$\log(\hat{\nu}_X^2(\tau_j)) \stackrel{\mathrm{d}}{=} \log(\chi^2_{\eta_j}) + \log(\nu_X^2(\tau_j)) - \log(\eta_j).$$

Bartlett and Kendall (1946) show that

$$E\{\log(\chi^2_{\eta_j})\} = \psi(\tfrac{\eta_j}{2}) + \log(2),$$

where ψ is the digamma function. Hence we have

$$E\{\log(\hat{\nu}_X^2(\tau_j))\} = \log(\nu_X^2(\tau_j)) + \psi(\tfrac{\eta_j}{2}) + \log(2) - \log(\eta_j).$$

Assuming that the bivariate stationary processes $\{\widetilde{W}_{j,t}\}$ and $\{\widetilde{W}_{k,t}\}$ are jointly Gaussian with cross-covariance sequence $s_{j,k,\tau} \equiv \operatorname{cov}\{\widetilde{W}_{j,t+\tau}, \widetilde{W}_{k,t}\}$, we can approximate $\operatorname{cov}\{\log(\hat{\nu}_X^2(\tau_j)), \log(\hat{\nu}_X^2(\tau_k))\}$ by

$$\frac{\operatorname{cov}\{\hat{\nu}_X^2(\tau_j), \hat{\nu}_X^2(\tau_k)\}}{\nu_X^2(\tau_j)\nu_X^2(\tau_k)} + 2\frac{\operatorname{var}\{\hat{\nu}_X^2(\tau_j)\}\operatorname{var}\{\hat{\nu}_X^2(\tau_k)\} + (\operatorname{cov}\{\hat{\nu}_X^2(\tau_j), \hat{\nu}_X^2(\tau_k)\})^2}{\nu_X^4(\tau_j)\nu_X^4(\tau_k)}, \tag{32}$$

where, for $j \leq k$,

$$\operatorname{cov}\left\{\hat{\nu}_X^2(\tau_j), \hat{\nu}_X^2(\tau_k)\right\} \approx \frac{2}{M_j} \sum_{\tau=-\infty}^{\infty} s_{j,k,\tau}^2 \equiv \frac{2A_{j,k}}{M_j}.$$

(Keim and Percival, 2011). In practice, we can estimate $A_{j,k}$ using

$$\hat{A}_{j,k} = \frac{1}{2}\left(\hat{\nu}_X^2(\tau_j)\hat{\nu}_X^2(\tau_k) + 2\sum_{\tau=1}^{M_k-1} \hat{s}_{j,\tau}\hat{s}_{k,\tau}\right)$$

(note that, when $k = j$, the above becomes identical to $\hat{A}_j$ of Equation (23)).

7.1 Estimation of power-law exponents

As illustrated in Sect. 8 below, plots of the wavelet variance versus scale on log/log axes sometimes show stretches over which $\log\left(\hat{\nu}_X^2(\tau_j)\right)$ appears to vary linearly with $\log(\tau_j)$, i.e., that

$$\log\left(\hat{\nu}_X^2(\tau_j)\right) \approx \alpha + \beta \log(\tau_j) \text{ over levels } j \text{ such that, say, } J_1 \leq j \leq J_2.$$

This pattern is consistent with a hypothesis that the true wavelet variance obeys a power law over scales τ_{J_1} to τ_{J_2}:

$$\nu_X^2(\tau_j) = c\tau_j^{\beta}, \text{ where } c = e^{\alpha}.$$

The power-law exponent β manifests itself as the slope on a log-log plot and is amenable to various interpretations. For example, at small scales, a slope of $0 \leq \beta \leq 2$ might indicate that $\{X_t\}$ has a fractal dimension of $D = 2 - \frac{\beta}{2}$ (Gneiting et al., 2011); on the other hand, a log-log plot that is linear over large scales is indicative of an intrinsically stationary process with long-range dependence that might be well modeled by either a fractional Gaussian process with Hurst parameter $H = 1 + \frac{\beta}{2}$ when $-1 < \beta < 0$, a fractional Brownian motion with parameter Hurst parameter $H = \frac{\beta}{2}$ when $0 < \beta < 2$, or a fractionally differenced process with parameter $\delta = (\beta + 1)/2$ when $\beta > -1$ (Flandrin, 1992; Abry et al., 1993 and 1995; Abry and Veitch, 1998; Jensen, 1999; Stoev and Taqqu, 2003; Stoev et al., 2006; Coeurjolly, 2008; Faÿ et al., 2009). Metrologists studying fractional frequency deviates from atomic clocks and other high-performance oscillators would equate slopes of $\beta = -3, -2, -1$, 0 and 1 to five canonical noise processes known as, respectively, white phase, flicker phase, white frequency, flicker frequency and random-walk frequency noise (Stein, 1985; Percival, 2003).

To estimate the power-law exponent β based on wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$, $j = J_1, \ldots, J_2$, we first define

$$Y_j \equiv \log\left(\hat{\nu}_X^2(\tau_j)\right) - \psi(\tfrac{\eta_j}{2}) - \log(2) + \log(\eta_j)$$

and then form the linear regression model

$$Y_j = \alpha + \beta \log(\tau_j) + e_j,$$

for which the error term

$$e_j \equiv \log\left(\frac{\hat{\nu}_X^2(\tau_j)}{\nu_X^2(\tau_j)}\right) - \psi(\tfrac{\eta_j}{2}) - \log(2) + \log(\eta_j)$$

is equal in distribution to the RV $\log(\chi^2_{\eta_j}) - \psi(\frac{\eta_j}{2}) - \log(2)$. As a rule of thumb, if $\eta_j \geq 10$ for each j, we can regard the e_j's as approximately obeying a multivariate Gaussian distribution. In vector notation, we can write

$$\mathbf{Y} = \mathcal{A}\boldsymbol{\theta} + \boldsymbol{e},$$

where $\mathbf{Y} \equiv [Y_{J_1}, \ldots, Y_{J_2}]^T$, $\mathcal{A}$ is a $(J_2 - J_1 + 1) \times 2$ matrix whose first column consists just of ones and whose second column is $\log(\tau_{J_1}), \ldots, \log(\tau_{J_2})$; $\boldsymbol{\theta} \equiv [\alpha, \beta]^T$; and $\mathbf{e} \equiv [e_{J_1}, \ldots, e_{J_2}]^T$ is a random vector with zero means and a covariance matrix $\Sigma_{\mathbf{e}}$ that is symmetric with its (j,k)th element given by Equation (32) when $j \leq k$. The generalized least squares (GLS) estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = [\hat{\alpha}, \hat{\beta}]^T \equiv \left(\mathcal{A}^T \Sigma_{\mathbf{e}}^{-1} \mathcal{A}\right)^{-1} \mathcal{A}^T \Sigma_{\mathbf{e}}^{-1} \mathbf{Y} \tag{33}$$

(Draper and Smith, 1998). Under our working assumptions, the estimator $\hat{\boldsymbol{\theta}}$ is multivariate Gaussian with mean vector $\boldsymbol{\theta}$ and covariance matrix $\left(\mathcal{A}^T \Sigma_{\mathbf{e}}^{-1} \mathcal{A}\right)^{-1}$, whose lower right-hand corner is the variance associated with the GLS estimator $\hat{\beta}$ of the power-law exponent. We can in turn use $\hat{\beta}$ to estimate, e.g., the parameter δ for a fractionally differenced process via $\hat{\delta} = (\hat{\beta} + 1)/2$, noting that $\operatorname{var}\{\hat{\delta}\} = \operatorname{var}\{\hat{\beta}\}/4$.

A simpler way to estimate $\boldsymbol{\theta}$ is to use a weighted least squares (WLS) estimator. This estimator takes the same form as $\hat{\boldsymbol{\theta}}$ of Equation (33), but with $\Sigma_{\mathbf{e}}$ replaced by a diagonal matrix $\Lambda_{\mathbf{e}}$ whose diagonal elements are the same as those in $\Sigma_{\mathbf{e}}$; i.e., the jth diagonal element is $\operatorname{var}\{\log(\hat{\nu}_X^2(\tau_j))\}$. Using Equations (32) and (21), we can approximate this variance by

$$\frac{\operatorname{var}\{\hat{\nu}_X^2(\tau_j)\}}{\nu_X^4(\tau_j)} + \frac{4(\operatorname{var}\{\hat{\nu}_X^2(\tau_j)\})^2}{\nu_X^8(\tau_j)} = \frac{2}{\eta_j} + \frac{16}{\eta_j^2}.$$

The WLS estimator is attractive in that it just depends on the EDOFs η_j and does not make use of the covariances $\operatorname{cov}\{\log(\hat{\nu}_X^2(\tau_j)), \log(\hat{\nu}_X^2(\tau_k))\}$, $j \neq k$. The WLS estimator is suboptimal unless these covariances are close to zero, which becomes a better approximation as the wavelet filter width L_1 increases. Assuming the validity of this approximation, we can take the WLS estimator to be multivariate Gaussian with mean vector $\boldsymbol{\theta}$ and covariance matrix $\left(\mathcal{A}^T \Lambda_{\mathbf{e}}^{-1} \mathcal{A}\right)^{-1}$. (Section 9.5 of Percival and Walden, 2000, formulates a WLS estimator with the diagonal elements of $\Lambda_{\mathbf{e}}$ given by $\psi'(\frac{\eta_j}{2})$. The trigamma function enters into play because $\operatorname{var}\{e_j\} = \operatorname{var}\{\log(\chi^2_{\eta_j})\} = \psi'(\frac{\eta_j}{2})$, a result due to Bartlett and Kendall, 1946. This approach and the one presented here are essentially the same since $\psi'(\frac{\eta_j}{2}) \approx \frac{2}{\eta_j}$ for large η_j.)

7.2 Estimation of characteristic scale

The notion of characteristic scale pervades the physical sciences, but has no commonly accepted single definition (von Storch and Zwiers, 1999). Since the wavelet variance is scale-based, it is natural to entertain a definition in terms of peaks in plots of $\nu_X^2(\tau_j)$ versus τ_j (Keim and Percival, 2011). Accordingly, suppose $\{X_t\}$ is an intrinsically stationary process whose wavelet variance is such that $\nu_X^2(\tau_j) \geq \nu_X^2(\tau_{j-1})$ and $\nu_X^2(\tau_j) \geq \nu_X^2(\tau_{j+1})$ for some $j \geq 2$, with strict inequality holding in at least one of the two cases. We define a wavelet-based characteristic scale as the location $\tau_{c,j}$ at which a quadratic fit through the points $(x_k, y_k) \equiv (\log(\tau_k), \log(\nu_X^2(\tau_k)))$, $k = j-1$, j and $j+1$, is maximized:

$$\tau_{c,j} = 2^{-\beta_1/\beta_2}\tau_j, \text{ where } \beta_1 \equiv \frac{y_{j+1} - y_{j-1}}{2} \text{ and } \beta_2 \equiv y_{j+1} - 2y_j + y_{j-1}.$$

Note that this definition is based only on properties of the wavelet variance locally around scale τ_j, not on its properties at arbitrarily large scales. Other measures of correlation length in use break down in the face of long-range dependence, which is a large-scale property.

We can form an estimator of $\tau_{c,j}$ in an obvious manner by substituting $\hat{y}_k = \log(\hat{\nu}_k^2)$ for y_k in the above, thus yielding estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and hence $\hat{\tau}_{c,j}$. An approximate 95% CI for $\tau_{c,j}$ is given by

$$\left[2^{-1.96\sigma_{\hat{\kappa}}}\hat{\tau}_{c,j}, 2^{1.96\sigma_{\hat{\kappa}}}\hat{\tau}_{c,j}\right]$$

(Keim and Percival, 2011), which depends upon the quantity $\sigma_{\hat{\kappa}}$, whose square $\sigma_{\hat{\kappa}}^2$ can be computed through the following steps. Let Σ be the 3×3 covariance matrix whose elements are dictated by Equation (32); i.e., the (m,n)th element of Σ is gotten by setting (j,k) in (32) to $(j-2+m, j-2+n)$, where $m \leq n$ range over the values 1, 2 and 3. Let

$$H = \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 1 & -2 & 1 \end{bmatrix}.$$

The 2×2 covariance matrix for $\hat{\beta}_1$ and $\hat{\beta}_2$ is given by the symmetric matrix $H\Sigma H^T$. Using the elements of this matrix, we can form

$$\sigma_{\hat{\kappa}}^2 = \frac{\operatorname{var}\{\hat{\beta}_1\}}{\beta_2^2} + \frac{\beta_1^2 \operatorname{var}\{\hat{\beta}_2\}}{\beta_2^4} + \frac{\operatorname{var}\{\hat{\beta}_1\}\operatorname{var}\{\hat{\beta}_2\} + 2(\operatorname{cov}\{\hat{\beta}_1, \hat{\beta}_2\})^2}{\beta_2^4} + \frac{3\beta_1^2(\operatorname{var}\{\hat{\beta}_2\})^2}{\beta_2^6} - \frac{2\beta_1 \operatorname{cov}\{\hat{\beta}_1, \hat{\beta}_2\}}{\beta_2^3}.$$

In practice, we can estimate $\sigma_{\hat{\kappa}}^2$ in a 'plug-in' manner by replacing the elements of Σ with obvious estimators.

8 Examples

Here we present five examples of wavelet variance analysis to illustrate the methodology discussed in previous sections.

8.1 Fractional frequency deviates from an atomic clock

We first consider a time series that is derived from measurements of the difference in time as kept by two hydrogen masers. Phase differences ϕ_t between the two masers (directly related to time differences) were measured once per minute for 4000 minutes and converted into fractional frequency deviates by a proper scaling of the first differences $\phi_t - \phi_{t-1}$. After multiplication by 10^{12} (merely to facilitate plotting), we obtain the series $\{X_t\}$ shown in Figure 2(a). The plot shows several pairs of large positive and negative spikes, which are due to isolated glitches in the phase measurements ϕ_t. The plot of empirical quantiles versus theoretical Gaussian quantiles shown in Figure 2(b) demonstrates that the data are well modeled by a Gaussian distribution except for the glitches distorting the tails of the distribution. The glitches are rogue occurrences that do not reflect the inherent ability of the hydrogen masers to keep time.

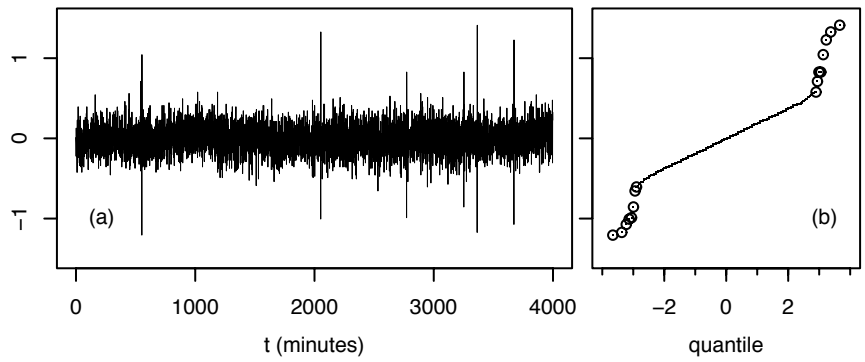

Fig. 2. Fractional frequency deviates (after multiplication by 10^{12}) derived from time differences between two hydrogen masers sampled one per minute, along with a plot of empirical quantiles versus theoretical Gaussian quantiles, with eight smallest and largest quantiles indicated by circles (data courtesy of Drs Lara Schmidt and Demetrios Matsakis, US Naval Observatory).

Scientists assessing the ability of atomic clocks to keep time have used the Allan variance as a performance measure since its introduction in the 1960s (Allan, 1966). The Allan variance is equal to twice the Haar wavelet variance of fractional frequency deviates, so we regard the Haar wavelet and Allan

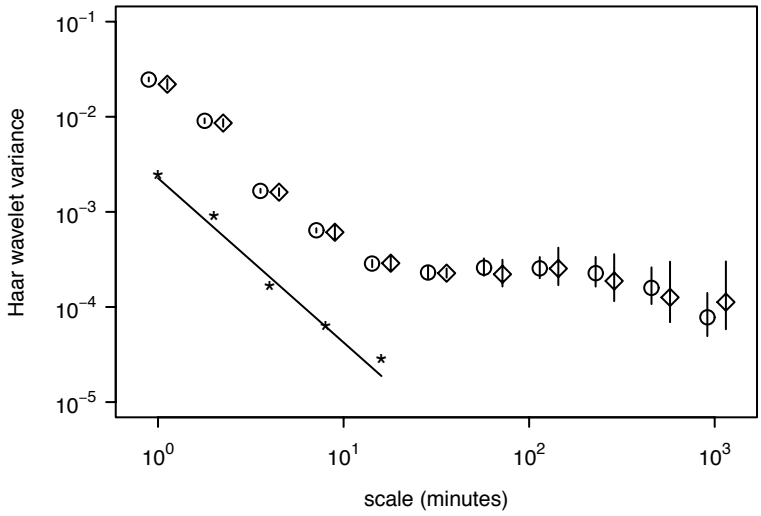

Fig. 3. Haar wavelet variance estimates of fractional frequency deviates along with 95% confidences intervals based on the unbiased estimator $\hat{\nu}_X^2(\tau_j)$ (circles) and median-type robust estimator $\hat{r}_X^2(\tau_j)$ (diamonds). An estimator of the Allan variance (routinely used to assess performance of atomic clocks) is given by $2\hat{\nu}_X^2(\tau_j)$. The asterisks show the unbiased wavelet variance estimates displaced downwards by an order of magnitude (i.e., $\hat{\nu}_X^2(\tau_j)/10$) along with a line determined by generalized least squares.

variances as equivalent in the discussion that follows. The popularity of the Allan variance is due in part to the ease with which it can be interpreted relative to SDF-based measures of clock performance (changes in averages over various scales are directly related to timing errors in clocks, whereas the SDF is related only indirectly). Figure 3 shows the Haar wavelet variance estimated using the unbiased estimator $\hat{\nu}_X^2(\tau_j)$ (circles) and the median-type robust estimator $\hat{r}_X^2(\tau_j)$ (diamonds). The conventional and robust estimates are in good agreement, indicating that the rogue values are not adversely affecting $\hat{\nu}_X^2(\tau_j)$.

Plots of the Allan variance have been traditionally used to identify the presence of power-law noise processes, with emphasis on certain canonical laws. Figure 3 makes it clear that no single power-law noise process can adequately model $\{X_t\}$ over all scales, but we can employ different processes over selected scales. For example, arguably the lowest five scales exhibit linear variation in Fig. 3, so we can make use of the methodology described in Sect. 7.1 to estimate a power-law exponent β over those scales based upon $\hat{\nu}_X^2(\tau_j)$ and Equation (33); however, the exponent so estimated is -1.73, with a corresponding 95% CI of $[-1.76, -1.70]$. This estimate lies between the exponents

associated with two canonical power-law processes, namely, flicker-phase noise ($\beta = -2$) and white-frequency noise ($\beta = -1$), but is not in good agreement with either. Prediction of $\nu_X^2(\tau_j)$ based upon the regression model is shown by the line in the lower left-hand portion of Fig. 3, along with asterisks depicting the $\hat{\nu}_X^2(\tau_j)$'s (both the line and the wavelet variance estimates are displaced down by an order of magnitude for display purposes).

8.2 Residual sea-ice thickness

Beginning in the late 1950s, the US Navy used submarines with upward-looking sonars to measure the underwater draft of Arctic sea ice, from which ice thickness can be inferred. The submarines collected these data by cruising under the ice in straight lines, resulting in profiles of thickness along transects that can be treated as a time series (with time being replaced with distance along a transect). These data are the most direct observational data we have documenting the evolution of sea-ice thickness over the past half century. Testing the hypothesis that there has been a significant decline in the average thickness of Arctic sea ice requires an understanding of the correlation properties of the profiles. Here we demonstrate how the wavelet variance can be used to assess these properties.

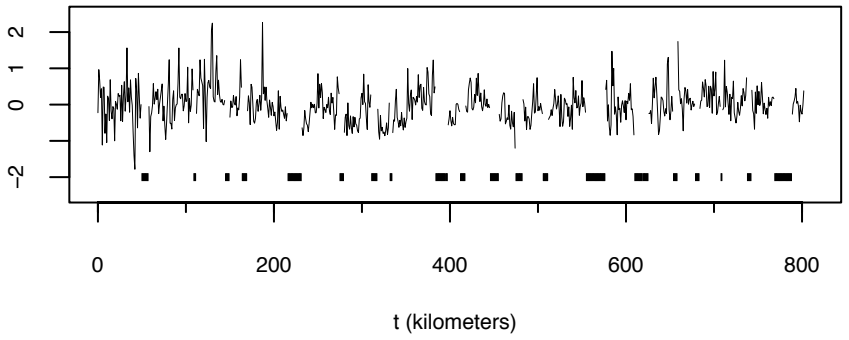

Fig. 4. Residual sea-ice thickness (based on data from a September 1997 Scientific Ice Expedition (SCICEX) cruise archived at the National Snow and Ice Data Center).

Figure 4 shows a thickness profile covering 802 kilometers after detrending by subtracting off a least squares line. This profile has a number of gaps whose positions are indicated by vertical hatch marks under the plot of the profile itself. As described in Percival et al. (2008), we can fill in these gaps using a stochastic interpolation scheme based upon either a first-order autoregressive or fractionally differenced Gaussian process. Doing so allows us

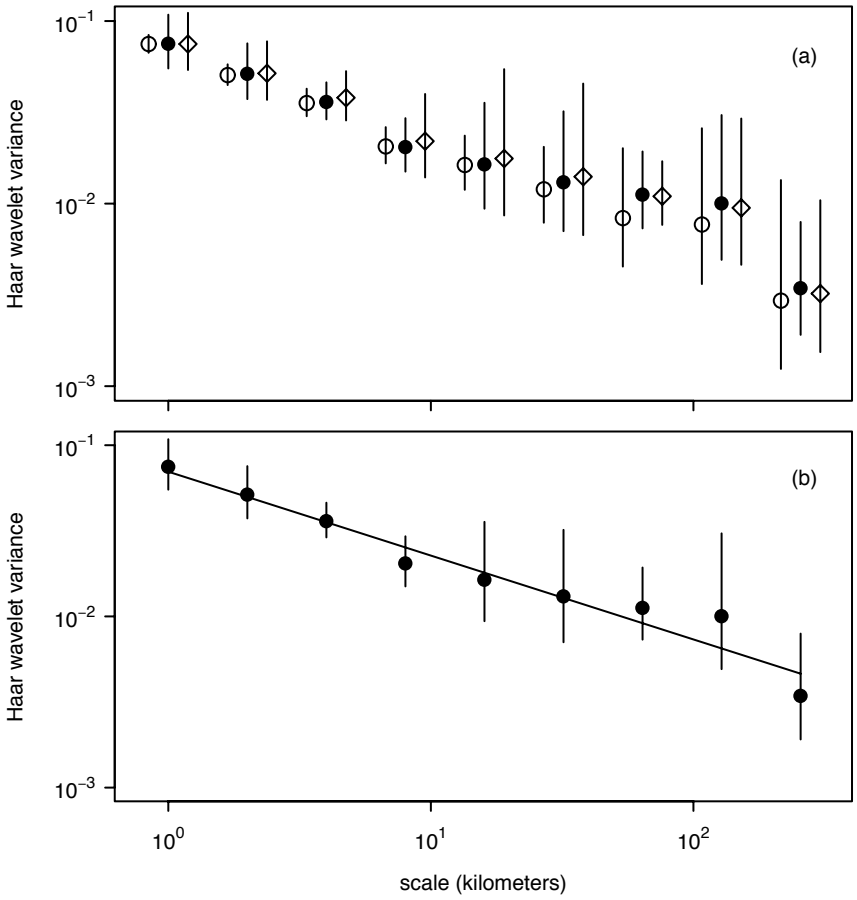

Fig. 5. Haar wavelet variance estimates of residual ice thickness along with 95% confidences intervals based on gap-filled data and the unbiased estimator $\hat{\nu}_X^2(\tau_j)$ (open circles) and on gappy data and the covariance-type and semi-periodogram-type estimators $\hat{u}_X^2(\tau_j)$ and $\hat{v}_X^2(\tau_j)$ (solid circles and diamonds, respectively). The lower plot shows a line from a weighted least squares fit of $\log(\hat{u}_X^2(\tau_j))$ versus $\log(\tau_j)$.

to compute the conventional unbiased Haar wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$ shown by the open circles in Fig. 5(a). Rather than using a gap-filled series, we can compute estimates based on the gappy series using the covariance-type estimator of Equation (28) and the semi-variogram-type estimator of Equation (29) – these are shown by, respectively, the solid circles and diamonds in Fig. 5(a). The three estimates at each scale agree well within one another, which provides some reassurance that the gap-filling procedure is not misrepresenting the correlation properties of the data. Figure 5(b) replicates the

covariance-type estimates. The approximate linear decay of the wavelet variance versus scale on this log-log plot suggests modeling the data as a process with long-range dependence. The WLS estimator described in Sect. 7.1 gives an estimated power-law exponent of $\hat{\beta} = -0.49$, which translates into an estimate of $\hat{\delta} = 0.26$ for the long-memory parameter for a fractionally differenced process. This value is consistent with an estimate (0.27) obtained using a computationally intensive maximum likelihood procedure (Percival et al., 2008) and is typical of results obtained for other thickness profiles. The wavelet variance thus contributes to our understanding of the correlation properties of sea-ice thickness, which in turn is a key component in assessing the significance of changes over the past quarter century in this important indicator of the Arctic climate (Rothrock et al., 2008).

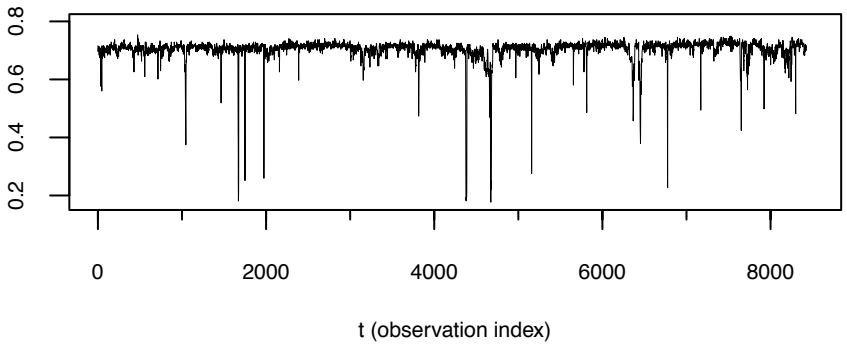

Fig. 6. Surface albedo of pack ice in the Beaufort Sea from a single line of a Landsat TM image obtained from channel 3 on April 16, 1992.

8.3 Albedo measurements of pack ice

Figure 6 shows a plot of surface albedo (a measure of proportion of incident light reflected) of spring ice in the Beaufort Sea as recorded by the Landsat satellite. The series consists of $N = 8428$ values spaced $\Delta_t = 25$ meter apart collected along a transect. Its distribution is highly non-Gaussian, with a lower tail dominated by spikes of low brightness attributable to open water and narrow cracks in thick ice. Lindsay et al. (1996) considered the wavelet variance for this series to investigate its potential for characterizing sea-ice variability. The circles in Fig. 7(a) show the Haar wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$ for physical scales ranging from 25 meters up to 25.6 kilometers (corresponding to standardized scales 1 to 1024). The broad peak in the wavelet variance curve indicates a characteristic scale between 200 and 400 meters (standardized scales 8 and 16). The solid curves above and below the circles

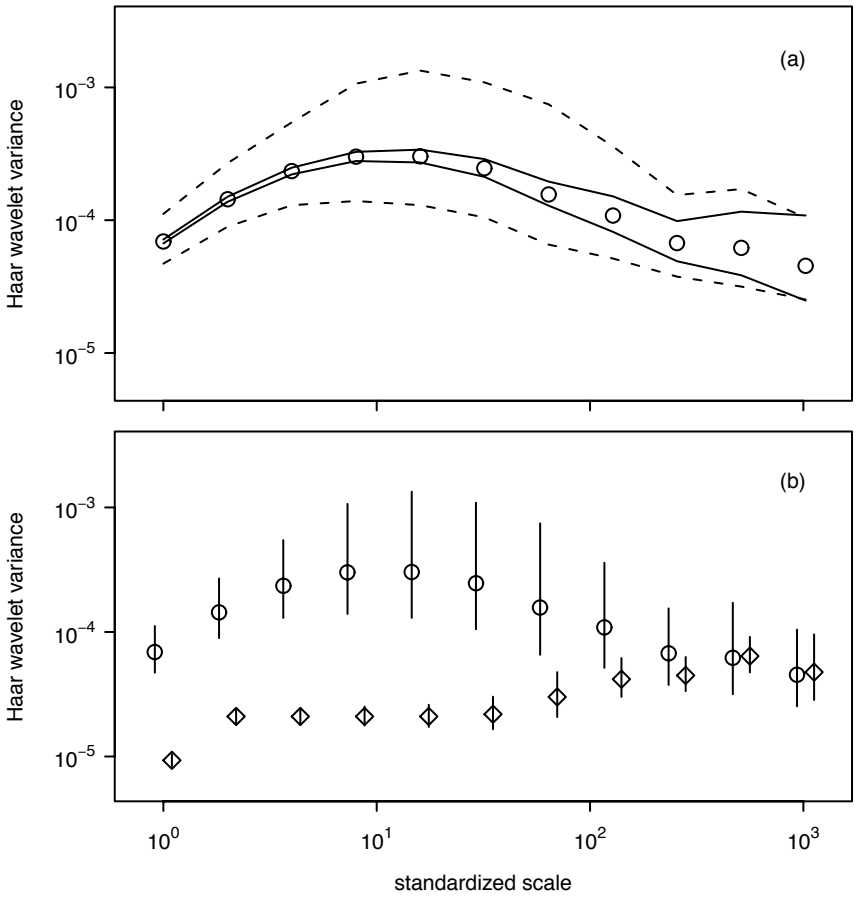

Fig. 7. Haar wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$ for albedo series (circles in both plots), along with Gaussian-based 95% confidence intervals (solid lines in upper plot) and non-Gaussian-based CIs (dashed lines). In addition to $\hat{\nu}_X^2(\tau_j)$ and its associated non-Gaussian-based CIs, the lower plot shows the robust median-type estimate $\hat{r}_X^2(\tau_j)$ (diamonds) with associated 95% CIs.

depict Gaussian-based 95% CIs formed via Equations (21), (23) and (22), while the dashed curves are corresponding CIs appropriate for non-Gaussian data based upon Equations (25), (27) and then (22) again. Note that, at small scales, the Gaussian-based CIs are considerably narrower than the ones based on non-Gaussian theory, but that the difference is less marked at larger scales, with the two CIs being virtually identical at the largest scale displayed. This example illustrates that there is a danger of underestimating the variability in wavelet variance estimates from an unwarranted assumption of Gaussianity.

Figure 7(b) again shows the estimates $\hat{\nu}_X^2(\tau_j)$ as circles along with the non-Gaussian-based 95% CIs. The diamonds show the robust median-type estimate $\hat{r}_X^2(\tau_j)$ of Equation (31) along with 95% CIs. This estimate deemphasizes the spikes in the series and hence reflects background properties of sea ice once the effect of open water and cracks has been downplayed. The robust estimate is quite a bit different from the usual estimate $\hat{\nu}_X^2(\tau_j)$ at small scales, suggesting that these scales are dominated by open water and cracks and that the characteristic scale between 200 and 400 meters is due mainly to these features. The spatial distribution of these features is of geophysical interest, and hence we cannot regard the spikes in Fig. 6 as rouge observations. Nonetheless the robust estimate is of interest because it tells us how much of the overall variability at each scale is due to background sea-ice processes that are interrupted by open water and cracks.

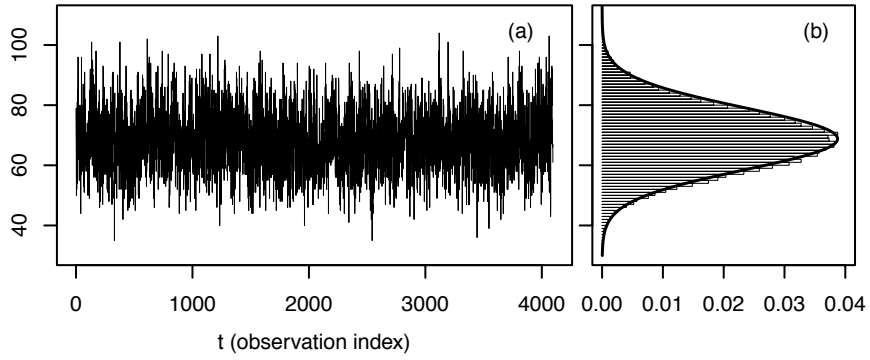

Fig. 8. X-ray fluctuations from a binary star system (first 4096 data values), along with histogram and fitted Gaussian probability density function.

8.4 X-ray fluctuations from a binary star system

Our fourth example is a time series of $N = 65,526$ counts from the X-ray binary system GX 5–1 as recorded by the Ginga satellite over a 512 second stretch of time (Norris et al., 1990; Hertz and Feigelson, 1997). Each observation X_t is the number of X-rays arriving within an interval (bin size) of $\Delta_t = 1/128$ seconds. Figure 8(a) shows the first 4096 values of the time series, while 8(b) shows a histogram for the entire series, along with a Gaussian probability density function whose mean and variance match those of the data.

Figure 9 shows a plot of the unbiased Haar wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$ along with 95% CIs, which, since the data are reasonably close to Gaussian, are based on Equations (21), (23) and (22). Because of the large

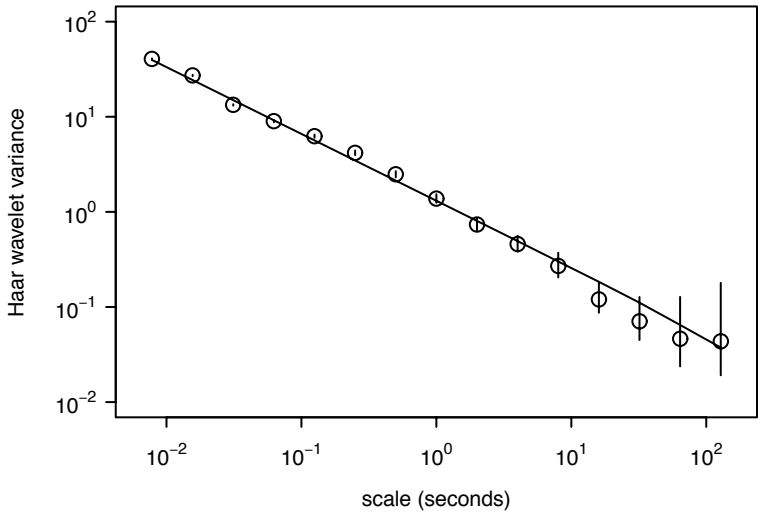

Fig. 9. Unbiased Haar wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$ for the time series of X-ray fluctuations (circles), along with Gaussian-based 95% confidence intervals (vertical lines) and a line with slope -0.7 based upon the generalized least square estimator of Equation (33).

sample size, the widths of the CIs are so small at the smaller scales as to be barely visible in the plot. Note that $\log(\hat{\nu}_X^2(\tau_j))$ decays roughly linearly with $\log(\tau)$ over all fifteen displayed scales. This fact suggests that a power-law model might be a simple description of the overall correlation properties for this time series. The GLS estimator of Equation (33) yields an estimate of $\hat{\beta} \doteq -0.702$ for the power-law exponent, with an associated 95% CI of $[-0.712, -0.693]$. Using the relationship $\delta = (\beta + 1)/2$, this exponent translates into an estimate of $\hat{\delta} = 0.149$ for the long-range parameter for a fractionally differenced process. Predicted values for the wavelet variances from the regression model are shown by the line in Fig. 9. Note that the CIs for the wavelet variances fail to trap the predicted values at a number of scales (particularly those below one second), an indication that the data have a more intricate correlation structure than what can be captured by a simple power-law model. The wavelet variance is thus able both to suggest a simple overall model for the X-ray fluctuations and to point out its limitations.

8.5 Coherent structures in river flow

Figure 10 shows a time series capturing so-called 'coherent structures' (such as boils or eddies) in river flows (Chickadel et al., 2009). The 4096 values shown in the plot are from a longer series of length $N = 29972$ that has a

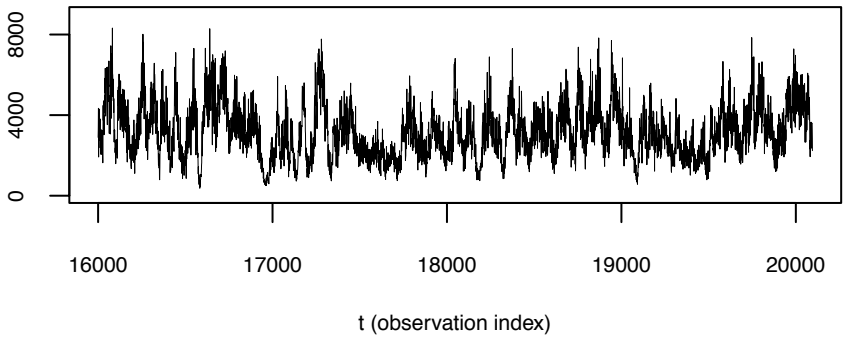

Fig. 10. Coherent structures in river flows (first 4096 values; data courtesy of Alex Horner-Devine and Bronwyn Hayworth, Department of Civil and Environmental Engineering, University of Washington).

sampling interval of $\Delta = 1/25$ second and spans a little less than 20 minutes (the subseries in the plot covers about 2.7 minutes). This time series is derived from measurements from three transducers and a velocity profiler set on the bottom of the Snohomish River Estuary in Washington State immediately downstream of a sill pointing upwards. The structures are essentially quasi-periodic upwellings from the river that appear as temporary 'blobs' on the surface of the river. Each blob dissipates within a second or so, and then another blob forms sometime later. As the tide increases, the water velocity increases, and the frequency at which the blobs occur appears to increase.

Videos of the river surface clearly show these boils qualitatively, but quantifying this little-understood phenomenon using standard Fourier-based spectral analysis is problematic because it appears as a small perturbation in a low-frequency rolloff. Figure 11 shows unbiased Haar wavelet variance estimates $\hat{\nu}_X^2(\tau_j)$ (circles) for this series, along with associated Gaussian-based 95% CIs (arguably non-Gaussian-based CIs would be more appropriate here). Here we see a peak at scale $\tau_6 \Delta = 1.28$ seconds, a clear indication of a characteristic scale in this vicinity. Using the methodology described in Sect. 7.2, we obtain an estimated characteristic scale of $\hat{\tau}_{c,6} \Delta = 1.6$ seconds, with an associated 95% CI of $[1.4, 1.9]$ seconds. The vertical dashed line in Fig. 11 indicates this estimated characteristic scale, while the thick horizonal line shows its associated CI. The estimated characteristic scale is based upon a quadratic fit through $\hat{\nu}_X^2(\tau_5)$, $\hat{\nu}_X^2(\tau_6)$ and $\hat{\nu}_X^2(\tau_7)$, which is shown by the curve passing through these estimates. In contrast to the SDF, the wavelet variance thus gives a readily interpretable quantification of the phenomenon of interest. We can study the time-evolving properties of the boils by estimating characteristic scales for time series spanning successive 20-minute time intervals.

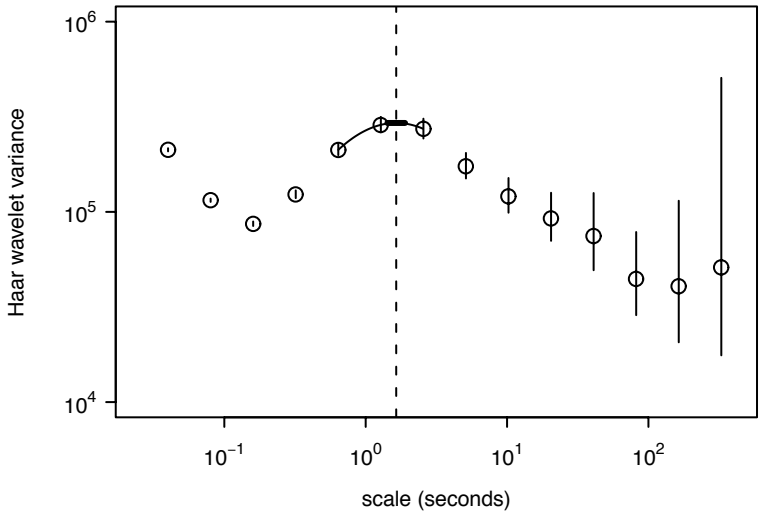

Fig. 11. Haar wavelet variance estimates from time series of coherent structures (circles), with wavelet-based characteristic scale indicated by vertical dashed line. A 95% confidence interval for the characteristic scale is shown by the thick horizontal line, while similar intervals for the wavelet variances are shown by the vertical lines emanating from the circles.

9 Concluding Remarks

We have presented a basic introduction to the wavelet variance of time series and its sampling theory. The five real-world examples in the previous section hopefully give the reader an idea of how the wavelet variance can be used as a tool for analyzing time series. Much has been left uncovered. Our statistical treatment has focused on time series that can be regarded as realizations of intrinsically stationary processes. Time series that fall outside of this framework can be fruitfully handled by breaking the series into subseries and analyzing each subseries separately *if* it is reasonable to assume that each subseries is itself a realization of an intrinsically stationary process (but with a possibly different process for each subseries). This procedure provides a simple way of using the wavelet-based methodology discussed in this chapter to handle certain nonstationary time series; however, the reader should be aware that there are other wavelet-based approaches designed to handle nonstationary time series. One approach is based on a 'locally stationary modeling' philosophy that facilitates certain asymptotic considerations. A good entry point to this body of literature (with an interesting application to seismic time series) is Fryzlewicz and Ombao (2009) and references therein.

Our presentation has focused on univariate time series. To study the bivariate relationships between multiple time series, Hudgins (1992) introduced the notion of the wavelet covariance (or wavelet cross spectrum) and wavelet cross correlation in terms of a continuous wavelet transform and applied these concepts to atmospheric turbulence in a subsequent paper (Hudgins et al., 1993). Whitcher et al. (2000) and Serroukh and Walden (2000b, 2000c) provide a statistical theory for wavelet covariance analysis of bivariate time series that parallels our treatment of the wavelet variance for single series. This theory presumes that the individual series are intrinsically stationary, but certain bivariate series whose relationships are evolving over time can be handled within this framework by breaking the series into subseries. Sanderson et al. (2010) describe an alternative approach to studying nonstationary bivariate time series that involves the use of wavelet-based locally stationary models. The notion of the wavelet variance can also be extended outside the context of time series to form a scale-based ANOVA for two-dimensional images. Unser (1995) is a pioneering work in this area, which also discusses wavelet-based texture analysis. Lark and Webster (2004) and Milne et al. (2010) document substantive applications of the two-dimensional wavelet variance in the analysis of soil variations. Mondal and Percival (2011b) develop a statistical theory for the two-dimensional wavelet variance that closely parallels the theory for the one-dimensional case presented in this chapter.

Finally we note that all of the computations and figures in this chapter were done in the statistical language `R` (R Development Core Team, 2011). Code for reproducing all of the numerical examples is available upon request from the authors.

Acknowledgments

Preparation of this chapter was supported in part by U.S. National Science Foundation Grants Nos. ARC 0529955 (Percival) and DMS 0906300 (Mondal). Any opinions, findings and conclusions or recommendations expressed in this chapter are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Abry, P., P. Gonçalvès and P. Flandrin, Wavelet-based spectral analysis of $1/f$ processes, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), **3**, 237–240, 1993.
2. Abry, P., P. Gonçalvès and P. Flandrin, Wavelets, spectrum analysis and $1/f$ processes, in *Wavelets and Statistics* (Lecture Notes in Statistics, Volume 103), edited by A. Antoniadis and G. Oppenheim, Springer–Verlag, New York, 15–29, 1995.

3. Abry, P., and D. Veitch, Wavelet analysis of long-range-dependent traffic, *IEEE Transactions on Information Theory*, **44**, 2–15, 1998.
4. Aldrich, E. M., *Alternative Estimators of Wavelet Variance*, MS dissertation, Department of Statistics, University of Washington, 2005.
5. Allan, D. W., Statistics of atomic frequency standards, *Proceedings of the IEEE,* **54**, 221–230, 1966.
6. Bartlett, M. S., and D. G. Kendall, The statistical analysis of variance-heterogeneity and the logarithmic transformation, *Supplement to the Journal of the Royal Statistical Society,* **8**, 128–138, 1946.
7. Beylkin, G., On the representation of operators in bases of compactly supported wavelets, *SIAM Journal on Numerical Analysis*, **29**, 1716–1740, 1992.
8. Brockwell, P. J., and R. A. Davis, *Introduction to Time Series and Forecasting* (Second Edition), Springer, New York, 2002.
9. Bruce, A. G., and H.-Y. Gao, *Applied Wavelet Analysis with S-PLUS*, Springer, New York, 1996.
10. Chiann, C., and P. A. Morettin, A wavelet analysis for time series, *Nonparametric Statistics*, **10**, 1–46, 1998.
11. Chickadel, C. C., A. R. Horner-Devine, S. A. Talke and A. T. Jessup, Vertical boil propagation from a submerged estuarine sill, *Geophysical Research Letters*, **36**, L10601, doi:10.1029/2009GL037278, 2009.
12. Coeurjolly, J.-F., Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles, *Annals of Statistics*, **36**, 1404–1434, 2008.
13. Coifman, R. R., and D. L. Donoho, Translation-invariant de-noising, in *Wavelets and Statistics* (Lecture Notes in Statistics, Volume 103), edited by A. Antoniadis and G. Oppenheim, Springer–Verlag, New York, 125–150, 1995.
14. Craigmile, P. F., and D. B. Percival, Asymptotic decorrelation of between-scale wavelet coefficients, *IEEE Transactions on Information Theory*, **51**, 1039–1048, 2005.
15. Daubechies, I., Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, **41**, 909–996, 1988.
16. Del Marco, S., and J. Weiss, Improved transient signal detection using a wavepacket-based detector with an extended translation-invariant wavelet transform, *IEEE Transactions on Signal Processing*, **45**, 841–850, 1997.
17. Draper, N. R., and H. Smith, *Applied Regression Analysis* (Third Edition), John Wiley & Sons, New York, 1998.
18. Faÿ, G., E. Moulines, F. Roueff and M. Taqqu, Estimators of long-memory: Fourier versus wavelets, *Journal of Econometrics*, **151**, 159–177, 2009.
19. Fryzlewicz, P., and H. Ombao, Consistent classification of nonstationary time series using stochastic wavelet representations, *Journal of the American Statistical Association*, **104**, 299–312, 2009.
20. Flandrin, P., Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Transactions on Information Theory*, **38**, 910–917, 1992.
21. Giraitis, L., and D. Surgailis, CLT and other limit theorems for functionals of Gaussian processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **70**, 191–212, 1985.
22. Gneiting, T., H. Ševčíková and D. B. Percival, Estimators of fractal dimension: assessing the roughness of time series and spatial data, under revision for *Statistical Science*, 2011.

23. Greenhall, C. A., D. A. Howe and D. B. Percival, Total Variance, an estimator of long-term frequency stability, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **46**, 1183–1191, 1999.
24. Hertz, P., and E. D. Feigelson, A sample of astronomical time series, in *Applications of Time Series Analysis in Astronomy and Meteorology*, edited by T. Subba Rao, M. B. Priestley and O. Lessi, Chapman & Hall, London, 340–356, 1997.
25. Huber, P. J., Robust estimation of a location parameter, *Annals of Mathematical Statistics*, **35**, 73–101, 1964.
26. Hudgins, L. H., *Wavelet Analysis of Atmospheric Turbulence*, PhD dissertation, Department of Physics & Astronomy, University of California, Irvine, 1992.
27. Hudgins, L. H., C. A. Friehe and M. E. Mayer, Wavelet transforms and atmospheric turbulence, *Physical Review Letters*, **70**, 3279–3282, 1993.
28. Jensen, M. J., Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter, *Journal of Forecasting*, **18**, 17–32, 1999.
29. Keim, M. J., and D. B. Percival, Assessing characteristic scales Using wavelets, submitted, 2011.
30. Labat, D., R. Ababou and A. Mangin, Introduction of wavelet analyses to rainfall/runoffs relationship for a karstic basin: the case of Licq–Atherey karstic system, *Ground Water*, **39**, 605–615, 2001.
31. Lang, M., H. Guo, J. E. Odegard, C. S. Burrus and R. O. Wells, Nonlinear processing of a shift invariant DWT for noise reduction, in *Wavelet Applications II* (Proceedings of the SPIE 2491), edited by H. H. Szu, SPIE Press, Bellingham, Washington, 640–651, 1995.
32. Lark, R. M., and R. Webster, 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, 2001.
33. Lark, R. M., and R. Webster, Analysing soil variation in two dimensions with the discrete wavelet transform, *European Journal of Soil Science*, **55**, 777–797, 2004.
34. Li, T.-H., and H.-S. Oh, Wavelet spectrum and its characterization property for random processes, *IEEE Transactions on Information Theory*, **48**, 2922–2937, 2002.
35. Liang, J., and T. W. Parks, A translation-invariant wavelet representation algorithm with applications, *IEEE Transactions on Signal Processing*, **44**, 225–232, 1996.
36. Lindsay, R. W., D. B. Percival and D. A. Rothrock, The discrete wavelet transform and the scale analysis of the surface properties of sea ice, *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 771–787, 1996.
37. Mallat, S. G., Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$, *Transactions of the American Mathematical Society*, **315**, 69–87, 1989a.
38. Mallat, S. G., A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693, 1989b.
39. Mallat, S. G., Multifrequency channel decompositions of images and wavelet models, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **37**, 2091–2110, 1989c.
40. Mallows, C. L., Linear processes are nearly Gaussian, *Journal of Applied Probability*, **4**, 313–329, 1967.

41. Massel, S. R., Wavelet analysis for processing of ocean surface wave records, *Ocean Engineering*, **28**, 957–987, 2001.
42. Milne, A. E., R. M. Lark and R. Webster, Spectral and wavelet analysis of gilgai patterns from air photography, *Australian Journal of Soil Research*, **48**, 309–325, 2010.
43. Mondal, D., *Wavelet Variance Analysis for Time Series and Random Fields*, PhD dissertation, Department of Statistics, University of Washington, 2007.
44. Mondal, D., and D. B. Percival, Wavelet variance analysis for gappy time series, *Annals of the Institute of Statistical Mathematics*, **62**, 943–966, 2010.
45. Mondal, D., and D. B. Percival, M-estimation of wavelet variance analysis, *Annals of the Institute of Statistical Mathematics*, in press, 2011a.
46. Mondal, D., and D. B. Percival, Wavelet variance analysis for random fields on a regular lattice, *IEEE Transactions on Image Processing*, under revision, 2011b.
47. Nason, G. P., and B. W. Silverman, The stationary wavelet transform and some statistical applications, in *Wavelets and Statistics* (Lecture Notes in Statistics, Volume 103), edited by A. Antoniadis and G. Oppenheim, Springer–Verlag, New York, 281–299, 1995.
48. Nason, G. P., R. von Sachs and G. Kroisandt, Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *Journal of the Royal Statistical Society. Series B. Methodological*, **62**, 271–292, 2000.
49. Norris, J. P., P. Hertz, K. S. Wood, B. A. Vaughan, P. F. Michelson, K. Mitsuda, and T. Dotani, Independence of short time scale fluctuations of quasi-periodic oscillations and low frequency noise in GX 5–1, *Astrophysical Journal*, **361**, 514–526, 1990.
50. Pelgrum, H., T. Schmugge, A. Rango, J. Ritchie and B. Kustas, Length-scale analysis of surface albedo, temperature, and normalized difference vegetation index in desert grassland, *Water Resources Research*, **36**, 1757–1766, 2000.
51. Percival, D. B., *The Statistics of Long Memory Processes*, PhD dissertation, Department of Statistics, University of Washington, 1983.
52. Percival, D. B., On estimation of the wavelet variance, *Biometrika*, **82**, 619–631, 1995.
53. Percival, D. B., Stochastic models and statistical analysis for clock noise, *Metrologia*, **40**, S289–S304, 2003.
54. Percival, D. B., and H. O. Mofjeld, Analysis of subtidal coastal sea level fluctuations using wavelets, *Journal of the American Statistical Association*, **92**, 868–880, 1997.
55. Percival, D. B., D. A. Rothrock, A. S. Thorndike and T. Gneiting, The variance of mean sea-ice thickness: effect of long-range dependence, *Journal of Geophysical Research – Oceans*, **113**, C01004, doi:10.1029/2007JC004391, 2008.
56. Percival, D. B., and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press, Cambridge, England, 1993.
57. Percival, D. B., and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, England, 2000.
58. Pesquet, J.-C., H. Krim, and H. Carfantan, Time-invariant orthonormal wavelet representations, *IEEE Transactions on Signal Processing*, **44**, 1964–1970, 1996.
59. Pichot, V., J.-M. Gaspoz, S. Molliex, A. Antoniadis, T. Busso, F. Roche, F. Costes, L. Quintin, J.-R. Lacour and J.-C. Barthélémy, Wavelet transform to quantify heart rate variability and to assess its instantaneous changes, *Journal of Applied Physiology*, **86**, 1081–1091, 1999.

60. R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011 (URL: `http://www.r-project.org/`).
61. Rosenblatt, M., *Stationary Sequences and Random Fields*, Birkhäuser, Boston, 1985.
62. Rothrock, D. A., D. B. Percival and M. Wensnahan, The decline in arctic sea-ice thickness: separating the spatial, annual, and interannual variability in a quarter century of submarine data, *Journal of Geophysical Research – Oceans*, **113**, C05003, doi:10.1029/2007JC004252, 2008.
63. Rybák, J., and I. Dorotovič, Temporal variability of the coronal green-line index (1947–1998), *Solar Physics*, **205**, 177–187, 2002.
64. Sanderson, J., P. Fryzlewicz, and M. W. Jones, Estimating linear dependence between nonstationary time series using the locally stationary wavelet model, *Biometrika*, **97**, 435–446, 2010.
65. Scargle, J. D., T. Steiman-Cameron, K. Young, D. L. Donoho, J. P. Crutchfield and J. Imamura, The quasi-periodic oscillations and very low frequency noise of Scorpius X–1 as transient chaos: a dripping handrail?, *The Astronomical Journal*, **411**, L91–L94, 1993.
66. Serroukh, A., and A. T. Walden, Wavelet scale analysis of bivariate time series I: motivation and estimation, *Nonparametric Statistics*, **13**, 1–36, 2000a.
67. Serroukh, A., and A. T. Walden, Wavelet scale analysis of bivariate time series II: statistical properties for linear processes, *Nonparametric Statistics*, **13**, 36–56, 2000b.
68. Serroukh, A., A. T. Walden and D. B. Percival, 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, 2000.
69. Shensa, M. J., The discrete wavelet transform: wedding the à trous and Mallat algorithms, *IEEE Transactions on Signal Processing*, **40**, 2464–2482, 1992.
70. Stein, S. R., Frequency and time – their measurement and characterization, in *Precision Frequency Control, Volume 2: Oscillators and Standards*, edited by E. .A. Gerber and A. Ballato, Academic Press, Orlando, 191–232, 1985.
71. Stoev, S., and M. S. Taqqu, Wavelet estimation of the Hurst parameter in stable processes, in *Processes with Long Range Correlations: Theory and Applications* (Lecture Notes in Physics, Volume 621), edited by D. Rangarajan and M. Ding, Springer, Berlin, 61–87, 2003.
72. Stoev, S., M. S. Taqqu, C. Park, G. Michailidis and J. S. Marron, LASS: a tool for the local analysis of self-similarity, *Computational Statistics & Data Analysis*, **50**, 2447–2471, 2006.
73. Thomson, D. J., Spectrum estimation and harmonic analysis, *Proceedings of the IEEE,* **70**, 1055–1096, 1982.
74. Torrence, C., and G. P. Compo, A practical guide to wavelet analysis, *Bulletin of the American Meteorological Society*, **79**, 61–78, 1998.
75. Tsakiroglou, E., and A. T. Walden, From Blackman–Tukey pilot estimators to wavelet packet estimators: a modern perspective on an old spectrum estimation idea, *Signal Processing*, **82**, 1425–1441, 2002.
76. Unser, M., Texture classification and segmentation using wavelet frames, *IEEE Transactions on Image Processing*, **4**, 1549–1560, 1995.
77. von Storch, H., and F. W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, England, 1999.

78. Whitcher, B. J, S. D. Byers, P. Guttorp and D. B. Percival, Testing for homogeneity of variance in time series: long memory, wavelets and the Nile River, *Water Resources Research*, **38**, 1054–1070, 2002.
79. Whitcher, B. J, P. Guttorp and D. B. Percival, Wavelet analysis of covariance with application to atmospheric time series, *Journal of Geophysical Research – Atmospheres*, **105**, 14,941–14,962, 2000.
80. Yaglom, A. M., Correlation theory of processes with random stationary nth increments, *American Mathematical Society Translations* (Series 2), **8**, 87–141, 1958.