Long-Memory Processes, the Allan Variance and Wavelets

by

D. B. Percival[†] and P. Guttorp[*]

[†] Applied Physics Laboratory, HN–10, University of Washington, Seattle, WA 98195

[*] Department of Statistics, GN–22, University of Washington, Seattle, WA 98195

Abstract – Long term memory has frequently been observed in physical time series. Statistical theory for long term memory stochastic processes is radically different from the standard time series analysis, which assumes short term memory. The Allen variance is a particular measure of variability developed for long term memory processes. This variance can be interpreted as a Haar wavelet coefficient variance, suggesting an approach towards assessing the variability of general wavelet classes. The theory is applied to a 'time' series of vertical ocean shear measurements for which some drawbacks with the Haar wavelets are observed.

1. Introduction

In a variety of applications, time series analysts have noticed that the estimated autocovariance sequence for their data tends to decrease rather slowly, indicating that the series has 'long memory' in the sense that changes in the remote past continue to affect the present value of the series. Beran (1992) gives a good review of statistical and historical aspects of long memory processes. Time series that are well modelled by long memory processes have been observed, for example, by Newcomb (1895) in astronomy, by Gosset (Student, 1927) in chemistry, and by Smith (1938) in agriculture. In geophysics a famous early example is Hurst's (1951) study of the minimum annual height of the river Nile. This series has a sample autocovariance sequence (acvs) $\hat{s}_\tau$ that is approximately proportional to $|\tau|^{-0.3}$; i.e., the sequence decays hyperbolically, thus ruling out such standard time series models as ARMA models that have exponentially decaying autocovariance sequences (here $\hat{s}_\tau$ is an estimate of $s_\tau \equiv \mathrm{cov}\,\{X_t, X_{t+\tau}\}$ when $\{X_t\}$ is a stationary process). The applicability of long memory processes to climate data has been recently discussed by Raftery and Haslett (1989) and Smith (1992).

The statistical properties of a long memory process can be quite different from those of a set of independent and identically distributed (iid) observations. For example, the familiar variability properties of sample averages of iid observations are far from valid for a long memory process. In fact, Smith (1938) observed that sample averages in agricultural uniformity trials had variances that did not decrease at the rate of the number of terms in the average, and deduced that there must be substantial spatial correlation. To see the effect of long memory on the variability of averages, consider a simple example of a long memory process, namely, a fractional Gaussian process $\{X_t\}$ with self-similarity parameter $1/2 \le H < 1$ (Mandelbrot and Wallis, 1969). By definition, such a process has an acvs given by

$$s_\tau = \frac{s_0}{2}\left(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}\right), \quad \tau = \pm 1, \pm 2, \ldots,$$

where $s_0 = \mathrm{var}\{X_t\}$ (note that the case $H = 1/2$ corresponds to the iid case because then $s_\tau = 0$ for $\tau \neq 0$, whereas $\{X_t\}$ have long term memory if $1/2 < H < 1$). Then $\mathrm{var}\{\bar{X}\} = s_0 N^{2H-2}$, where $\bar{X}$ is the sample mean of N observations (i.e., $\sum_{t=1}^{N} X_t/N$). Note that, for values of H close to 1, the rate of decrease of variability in $\bar{X}$ is markedly different from the $1/N$ rate of the iid case. Naive application of iid statistics to the sample mean of a long memory process can thus be very misleading. For example, if we fit a fractional Gaussian process to the Nile data, we obtain an estimate of $H = 0.85$, implying that the variance of the sample mean decreases like $1/N^{0.3}$ instead of the usual $1/N$ rate. Numerically, obtaining 100 observations from this fractional Gaussian process is equivalent to obtaining only 4 observations from an iid process!

In spite of its slow rate of decay, the sample average of a long memory process is still a surprisingly efficient estimator of the mean level of the process (Beran and Künsch, 1985; Percival, 1985). Unfortunately, the same cannot be said for other standard statistics. In particular, the sample variance is a poor estimator of the process variance for a long memory process because it has both severe bias and low efficency (Beran, 1992). Due to these problems in estimation, Allan (1966) criticised the use of the sample variance as a meaningful estimator of variability for stationary processes with long memory and for nonstationary processes with infinite variance. He proposed an alternative theoretical measure of variability that is now known as the Allan variance (see Section 2 below). In terms of filtering theory, the Allan variance can be interpreted as the variance of a process after it has been subjected to an approximate band-pass filter of 'constant Q' (i.e., the ratio of the center frequency of the pass-band and the width of the pass-band is a constant). The chief advantages of the Allan variance for long memory processes are two-fold: first, this variance can be estimated without bias and with good efficiency for such processes, and, second, estimates of the Allan variance can in turn be used to estimate the parameters of the long memory process (as we note in Section 2, these statements also hold for certain

nonstationary power-law processes). The Allan variance has been applied for over twenty-five years as the routine time domain measure of frequency stability in high-performance oscillators (such as atomic clocks).

In a recent article Flandrin (1992) briefly noted that the Allan variance can be interpreted in terms of the coefficients of a Haar wavelet transform of a time series. We explore this connection in detail in Section 2, where we note that the notion of the Allan variance can be generalized to other wavelets to define a wavelet variance. The wavelet variance is a useful way of summarizing the properties of the wavelet transform for certain processes. In particular, the parameters of long memory processes can be deduced from the wavelet variance. In addition we note that standard estimators for the Allan variance can be easily generalized to these other wavelet variances.

In terms of computational complexity, the Allan variance is the simplest of the wavelet variances. Unfortunately, as we demonstrate in Section 3, it can be misleading for certain processes of interest in geophysics. In the particular example we consider in Section 3, however, we find that the wavelet variance for Daubechies's 'least asymmetric' wavelet of order 8 (Daubechies, 1992) yields markedly better results. This example demonstrates the usefulness for data analysis of classes of wavelets beyond the simple Haar wavelet. Our analysis also demonstrates the importance of conducting a parallel spectral analysis to validate the interpretation of a wavelet variance in terms of the parameters for a long memory process. Finally, in Section 4 we discuss some possible extensions, including some general thoughts on assessing the variability of wavelet coefficients.

2. The Allan Variance and Wavelets

Suppose that we have a time series of length N that can be regarded as one portion of one realization from the stochastic process $\{Y_t, t = 0, \pm 1, \ldots\}$ (for convenience, we assume that the sampling interval between consecutive observations is fixed at unity). Let

$$\overline{Y}_t(k) \equiv \frac{1}{k} \sum_{j=0}^{k-1} Y_{t-j} \tag{1}$$

represent the sample average of k consecutive observations, the latest one of which is Y_t. The Allan variance at scale k is denoted by $\sigma_Y^2(k)$ and is defined to be half the mean square difference between adjacent nonoverlapping $\overline{Y}_t(k)$'s; i.e.,

$$\sigma_Y^2(k) \equiv \frac{1}{2} E\left\{[\overline{Y}_t(k) - \overline{Y}_{t-k}(k)]^2\right\}.$$

In order for $\sigma_Y^2(k)$ to be independent of the index t, we must impose a stationarity condition on the process $\{Y_t\}$, namely, that either it or its first backward difference $Z_t \equiv Y_t - Y_{t-1}$ is a stationary process. Note that the Allan variance at scale k is a measure of how much averages of length k change from one time period of length k to the next.

To see why the Allan variance might be of interest for long memory processes, suppose momentarily that $\{Y_t\}$ is a fractional Gaussian process with self-similarity parameter $1/2 < H < 1$. If we let $S_Y(\cdot)$ denote the spectral density function (sdf) for $\{Y_t\}$ defined for frequencies f between $-1/2$ and $1/2$, then we have

$$S_Y(f) = L_1(f)|f|^{1-2H},$$

where $L_1(\cdot)$ is a slowly varying function for $|f| \to 0$ (Beran, 1992). Thus, if we plot $\log(S_Y(f))$ versus $\log(f)$ for positive values of f close to zero, we will observe (to a good approximation) a line with a slope of $1-2H$. A similar result holds for the Allan variance: we have

$$\sigma_Y^2(k) = L_2(k)k^{2H-2},$$

where $L_2(\cdot)$ is a slowly varying function for $k \to \infty$ (Percival, 1983, Theorem 2.19). Thus, if we plot $\log(\sigma_Y^2(k))$ versus $\log(k)$ for large k, we will observe (to a good approximation) a line with a slope of $2H-2$. In the analysis of frequency stability, it is common practice to produce plots of the so-called 'σ/τ curve,' which is just a plot of an estimate of $\log(\sigma_Y(k))$ versus $\log(k)$. On such a plot, a slope of β with $-1/2 \leq \beta < 0$ would be indicative of a stationary process with a self-similarity parameter $H = \beta + 1$.

As we mentioned above, the Allan variance is also well-defined when $\{Y_t\}$ is not itself a stationary process but its associated first backward difference $\{Z_t\}$ is. In such cases, we can *define* an sdf $S_Y(\cdot)$ for $\{Y_t\}$ in terms of the sdf $S_Z(\cdot)$ for $\{Z_t\}$ via the relationship

$$S_Y(f) \equiv \frac{S_Z(f)}{4\sin^2(\pi f)}$$

(this definition is motivated by the theory of linear filters). Suppose momentarily that

$$S_Y(f) = L_3(f)|f|^\alpha,$$

where $L_3(\cdot)$ is a slowly varying function for $|f| \to 0$, and $-3 < \alpha < 0$; i.e., $S_Y(\cdot)$ is a 'red' power-law process with exponent α greater than -3. Then we have

$$\sigma_Y^2(k) = L_4(k)k^{-\alpha-1},$$

where $L_4(\cdot)$ is a slowly varying function for $k \to \infty$. Thus, in principle, the Allan variance can be used to deduce the parameters for certain 'red' power-law processes.

Let us now consider the Haar wavelet transform of the time series $Y_1, \ldots, Y_N$, where we now assume that the sample size N is a power of 2 so that $N = 2^p$ for some positive integer p. By definition, this transform consists of $N-1$ 'detail' coefficients and one 'smooth' coefficient $s_1 \equiv \sum_{t=1}^N Y_t/N$. The detail coefficients $d_{j,k}$ are defined for scales $k = 1, 2, 4, \ldots, N/2$ and – within the kth scale – for indices $j = 1, 2, 3, \ldots, N/2k$ as

$$d_{j,k} \equiv \frac{1}{\sqrt{2k}} \left[\sum_{l=0}^{k-1} Y_{2jk-l} - \sum_{l=0}^{k-1} Y_{2jk-k-l} \right]. \tag{2}$$

For example, we have

$$\begin{aligned}
d_{j,1} &\equiv [Y_{2j} - Y_{2j-1}]/\sqrt{2}, \quad j = 1, \ldots, N/2 \\
d_{j,2} &\equiv [Y_{4j} + Y_{4j-1} - Y_{4j-2} - Y_{4j-3}]/2, \quad j = 1, \ldots, N/4 \\
d_{j,4} &\equiv [Y_{8j} + \cdots + Y_{8j-3} - Y_{8j-4} - \cdots - Y_{8j-7}]/2\sqrt{2}, \quad j = 1, \ldots, N/8 \\
&\vdots \\
d_{j,N/2} &\equiv [Y_N + \cdots + Y_{N/2+1} - Y_{N/2} - \cdots - Y_1]/(\sqrt{2})^p.
\end{aligned}$$

We can now state the relationship between the Allan variance and the Haar wavelet coefficients $d_{j,k}$. Using Equations (1) and (2), we have

$$d_{j,k} = \left(\frac{k}{2}\right)^{1/2} [\overline{Y}_{2jk}(k) - \overline{Y}_{2jk-k}(k)].$$

If either $\{Y_t\}$ or its first backward difference $\{Z_t\}$ is a stationary process with zero mean, we then have

$$\text{var}\,\{d_{j,k}\} = E\{d_{j,k}^2\} = \frac{k}{2} E\{[\overline{Y}_{2jk}(k) - \overline{Y}_{2jk-k}(k)]^2\} = k\sigma_Y^2(k). \tag{3}$$

Thus the Allan variance at scale k is directly related to the variance of the Haar wavelet coefficient at that same scale. (Violation of the seemingly innocuous assumption that $E\{Z_t\} = 0$ can seriously impact our ability to make sense of estimates of the Allan variance – see Percival, 1983, for further discussion.)

In view of Equation (3), an obvious unbiased estimator for the Allan variance is

$$\tilde{\sigma}_Y^2(k) \equiv \frac{2}{N} \sum_{j=1}^{N/2k} d_{j,k}^2 = \frac{k}{N} \sum_{j=1}^{N/2k} [\overline{Y}_{2jk}(k) - \overline{Y}_{2jk-k}(k)]^2. \tag{4}$$

The above estimator is known in the frequency stability literature as the 'nonoverlapped' estimator of the Allan variance because each value Y_t in the time series contributes to exactly one term in each summation in Equation (4). The estimator $\tilde{\sigma}_Y^2(k)$ is not commonly used because of the superior statistical properties of a closely related estimator known as the maximal-overlap estimator (Greenhall, 1991). This estimator is defined as

$$\hat{\sigma}_Y^2(k) \equiv \frac{1}{2(N-2k+1)} \sum_{t=2k}^{N} \left(\overline{Y}_t(k) - \overline{Y}_{t-k}(k)\right)^2. \tag{5}$$

We can interpret the above estimator in terms of a filtering operation from which we can obtain the Haar wavelet coefficients $d_{j,k}$ by subsampling (i.e., downsampling). At scale k, the filter that yields the $d_{j,k}$'s is of length $2k$ and has coefficients

$$g_{l,k} \equiv \begin{cases} 1/\sqrt{2k}, & l = 0, \ldots, k-1; \\ -1/\sqrt{2k}, & l = k, \ldots, 2k-1 \end{cases}$$

(cf. Equation (2)). Note that $\sum_{l=0}^{2k-1} g_{l,k}^2 = 1$. If we let

$$W_{t,k} = \sum_{l=0}^{2k} g_{l,k} Y_{t-l}, \quad t = 2k, \ldots, N, \tag{6}$$

then we have $d_{j,k} = W_{2jk,k}$; i.e., the $d_{j,k}$'s are obtained by appropriately subsampling every $2k$th value of the $W_{t,k}$'s. We can now easily see the difference between the nonoverlapped and maximal-overlap estimators:

$$\tilde{\sigma}_Y^2(k) = \frac{2}{N} \sum_{j=1}^{N/2k} W_{2jk,k}^2, \text{ whereas } \hat{\sigma}_Y^2(k) = \frac{1}{k(N-2k+1)} \sum_{t=2k}^{N} W_{t,k}^2; \tag{7}$$

i.e., $\tilde{\sigma}_Y^2(k)$ makes use of just the $N/2k$ subsampled $W_{t,k}$'s, whereas $\hat{\sigma}_Y^2(k)$ makes use of all $N - 2k + 1$ of the $W_{t,k}$'s. Note that use of $\hat{\sigma}_Y^2(k)$ rather than $\tilde{\sigma}_Y^2(k)$ imparts a certain degree of independence in the choice of the origin (i.e., if we form a new time series by discarding Y_1 and adding Y_{N+1}, only one of the $W_{t,k}$'s in the maximal-overlap estimator changes whereas all of them change in the nonoverlapped estimator). Note also that the definition of $\hat{\sigma}_Y^2(k)$ in Equation (5) holds for all sample sizes N (i.e., N need not be a power of 2).

For computational purposes, it is more efficient to evaluate the right-hand side of Equation (5) by computing the summations needed to obtain the $\overline{Y}_t(k)$'s just once. To do so, let $X_0 \equiv 0$, and let $X_t \equiv X_{t-1} + Y_t$ for $t = 1, \ldots, N$. Since $X_t = \sum_{u=1}^{t} Y_u$, we have

$$\overline{Y}_t(k) = \frac{1}{k} \sum_{j=0}^{k-1} Y_{t-j} = (X_t - X_{t-k})/k,$$

and hence we have

$$\overline{Y}_t(k) - \overline{Y}_{t-k}(k) = (X_t - 2X_{t-k} + X_{t-2k})/k \text{ and } W_t = (X_t - 2X_{t-k} + X_{t-2k})/\sqrt{2k}.$$

We can thus rewrite Equation (5) as

$$\hat{\sigma}_Y^2(k) = \frac{1}{2k^2(N-2k+1)} \sum_{t=2k}^{N} (X_t - 2X_{t-k} + X_{t-2k})^2 .$$

It is easy to extend the above ideas to wavelets other than the Haar wavelet. To do so, we must first generate the appropriate filter $\{h_{l,k}\}$ for each scale k based upon the wavelet for scale 1. This can be done easily by taking the inverse discrete wavelet transform of a properly placed pulse (for details, see the subsection 'What Do Wavelets Look Like?' in Section 13.10 of Press *et al.*, 1992). Let L_k be the length of the wavelet of scale k so that, for example, $L_1 = 2$ for the Haar wavelet while $L_1 = 8$ for the Daubechies's least asymmetric wavelet of order 8 (the values of L_k for scales $k = 2, 4, \ldots$, can be obtained via the recursive formula $L_{2k} = 2L_k + L_1 - 2$). For each scale k we then filter our time series to obtain

$$V_{t,k} = \sum_{l=0}^{L_k-1} h_{l,k} Y_{t-l}, \quad t = L_k, \ldots, N$$

(cf. Equation (6)). Note that, in contrast to the Haar wavelet, we can obtain only some of the wavelet coefficients by subsampling the $V_{t,k}$'s – the missing coefficients are those that involve circularly wrapping the time series (see the discussion in Press *et al.*, 1992). Under the same assumptions on $\{Y_t\}$ that we used to define the Allan variance, we can define the wavelet variance at scale k as

$$\nu_Y^2(k) \equiv E\{V_{t,k}^2\}/k$$

(cf. Equation (3)). The equivalent of the maximial-overlap estimator for this wavelet variance would be given by

$$\hat{\nu}_Y^2(k) \equiv \frac{1}{k(N - L_k + 1)} \sum_{t=L_k}^{N} V_{t,k}^2$$

(cf. Equation (7)). A plot of the square root of this quantity versus scale k on a log/log scale yields a generalization of the σ/τ curve, from which we can infer from regions of linearity that a power law process might be a good model for our data.

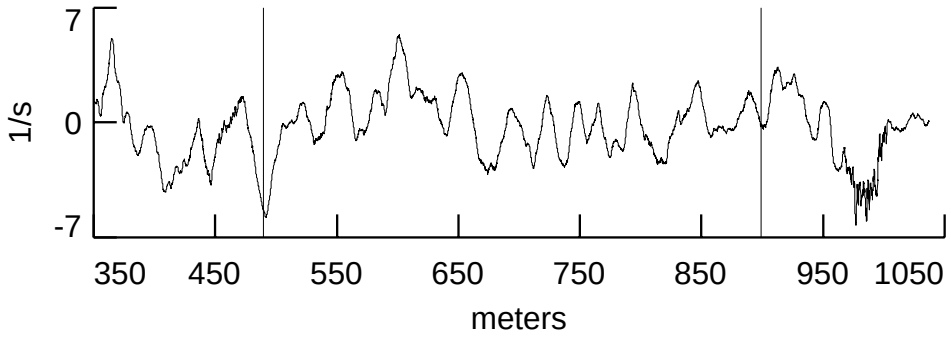

Figure 1. Plot of vertical shear measurements versus depth.

3. An Application to Vertical Shear Measurements

Here we illustrate the ideas of the previous section by examining a 'time' series of vertical ocean shear measurements (collected by Mike Gregg, Applied Physics Laboratory). The data were collected by an instrumental probe that is dropped over the side of a ship and is designed to then descend vertically down into the ocean. As it descends, the probe collects measurements on various aspects of the ocean as a function of depth. The ordering variable of our 'time' series is thus depth (measured in meters). One of the measurements that the probe collects is the x component of the velocity of water. This velocity is collected every 0.1 meters, first differenced over an interval of 10 meters, and then low-pass filtered to obtain a series that is related to the vertical shear in the ocean. These vertical shear measurements are thought to obey a power-law process over certain ranges of spatial frequency, so they are a useful candidate for examining how well the Allan variance and other wavelet variances can deduce the presence of such a process.

Figure 1 shows the series of vertical shear measurements used in this study. The series extends from a depth of 350.0 meters down to 1037.4 meters in increments of 0.1 meters (there are 6875 data values in all). There are two thin vertical lines marked on the plot, between which there are 4096 values ranging from 489.5 meters to 899.0 meter. In what follows, we will assume that this subseries can be regarded as a portion of one realization

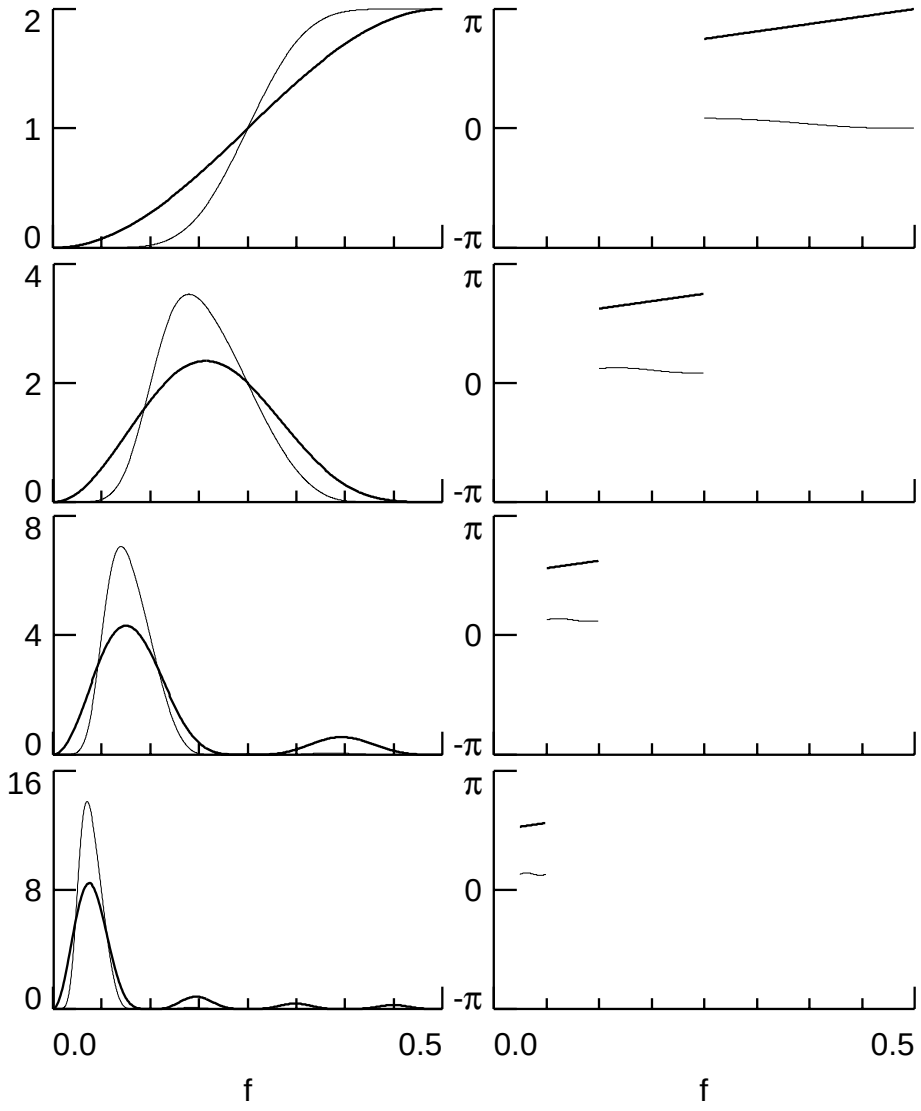

Figure 2. Squared modulus of transfer functions (left-hand column of plots) and phase functions (right-hand column) versus frequency for the Haar wavelet (thick curves) and LA(8) wavelet (thin curves) for scales 1, 2, 4 and 8 (top to bottom rows).

of either a stationary process or a process whose first backward difference is a stationary process (we need this assumption to apply meaningfully the methodology of Section 2).

The thick curves in the four rows of side-by-side plots in Figure 2 show, respectively, the squared modulus of the transfer function for the Haar wavelet (left-hand column) and the associated phase function over the nominal pass-band of the wavelet (right-hand

column) for the four scales 1, 2, 4 and 8 (top to bottom rows). The thin curves in each plot show the corresponding functions for Daubechies's 'least asymmetric' wavelet of order 8 (Daubechies, 1992; henceforth called the LA(8) wavelet). Note that the transfer functions for both wavelets roughly define a set of octave band filters; i.e., the transfer functions for scale k are approximately concentrated between frequencies and $1/4k$ and $1/2k$ (the spacing between the minor tick marks on the frequency axis is 1/16). The plots show that the transfer functions for the LA(8) wavelet are a better approximation to a set of octave band filters than those for the Haar wavelet. The phase functions for the Haar wavelet are approximately linear over the nominal pass-bands, whereas those for the LA(8) wavelet are approximately constant and fairly close to zero. Thus the output from the Haar wavelet filters will be phase-shifted with respect to the input, making it difficult to line up events at various scales with the original time series. In contrast, because the filters for the LA(8) wavelet are approximately zero phase, we can more easily line up events at various scales with the original time series. Note, however, that the spans of Haar wavelet filters for scales 1, 2, 4 and 8 are, respectively, 2, 4, 8 and 16, whereas the corresponding spans for the LA(8) wavelet are 8, 22, 50 and 106. If these wavelets are used as noncircular filters as discussed in Section 2, the output from the LA(8) wavelet will become increasingly shorter compared to that of the Haar wavelet as the scale increases. In fact, for the larger scales the length of the LA(8) wavelet will exceed the length of the time series, whereas the reverse will be true for the Haar wavelet.

Figure 3 shows (from bottom to top) the outputs from the Haar wavelet filters for scales 1, 2, 4 and 8. Each of these filtered series is drawn with the same vertical scale, so we see that the variability gets progressively larger as we move from lower to higher scales (the distance between minor tick marks on the vertical scale is 1 s^{-1}). The usual Haar wavelet transform for these scales can be obtained by appropriately subsampling these filtered series. Note that all four of these filtered series appear to be approximately rescaled versions of each other. We will see below that this is caused by a form of 'leakage'

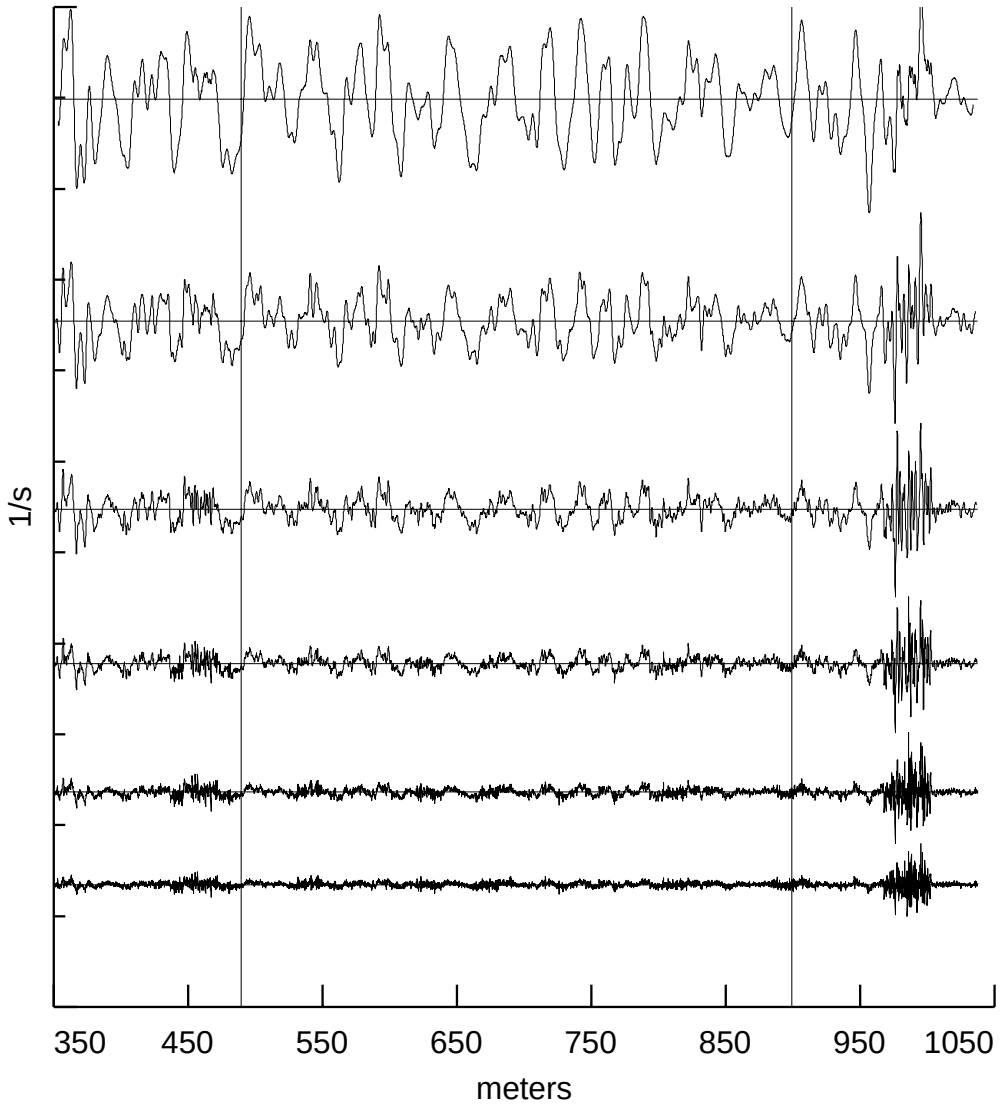

Figure 3. Outputs of Haar wavelet filters for scales 1, 2, 4 and 8 (bottom to top).

evidently due to the fact that the Haar approximation to pass-band filters is not good enough for this series and that in effect higher scale fluctuations are 'leaking' into these lower scale fluctuations. Note that there are two bursts in each of the filtered series in this figure, one near 450 meters and the other near 1000 meters.

Figure 4 is a repetition of Figure 3, with the LA(8) wavelet used instead of the Haar wavelet. The vertical scales on Figures 3 and 4 are identical, so it is evident that there is much less variability at the smaller scales for the LA(8) wavelet than for the Haar

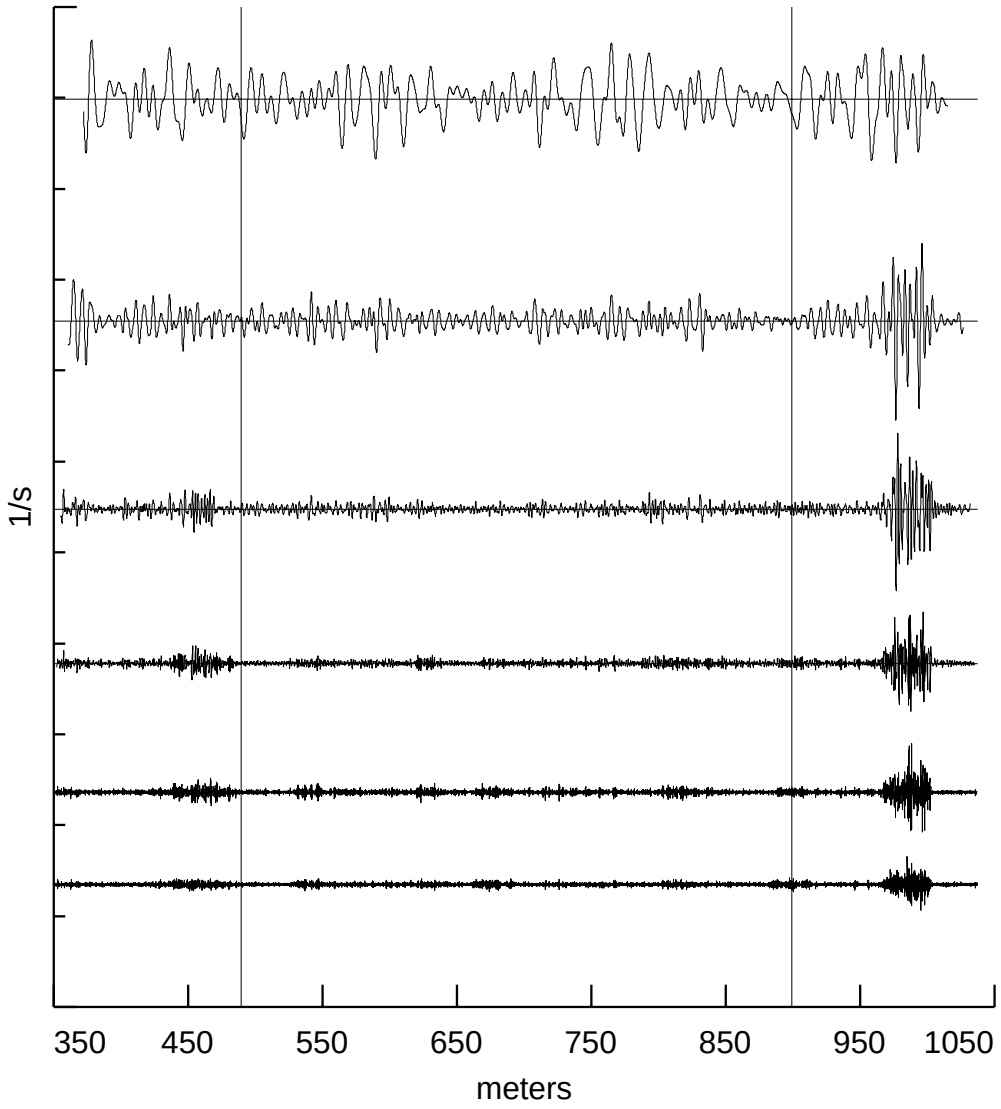

Figure 4. Outputs of LA(8) wavelet filters for scales 1, 2, 4 and 8 (bottom to top).

wavelet. While there is some correspondence between the filtered series at different scales, the correspondence is much less marked for the LA(8) wavelet than for the Haar wavelet. Each of the filtered series in this figure is slightly shorter than the corresponding one in Figure 3 due to the longer span of the filters for the LA(8) wavelet. Note that only part of the usual LA(8) wavelet transform for these scales can be obtained by subsampling these filtered series because we are *not* filtering our series as if it were circular (an assumption that would make no sense at all for a series of ocean depth measurements!). Again we can

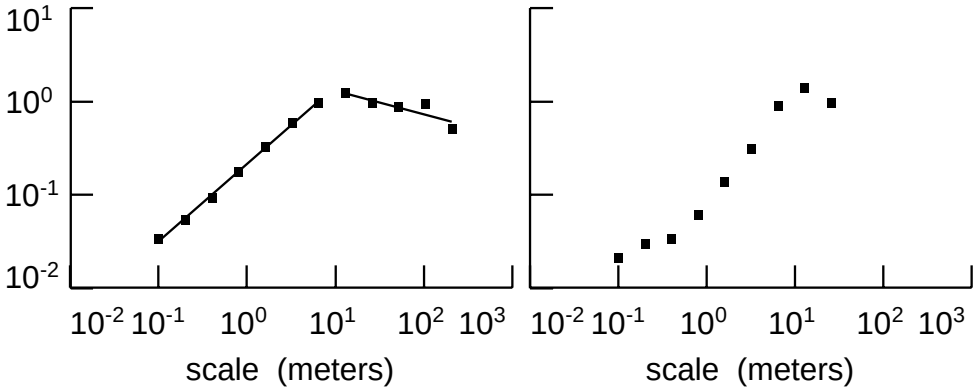

Figure 5. 'σ/τ' curves for the Haar wavelet (left-hand plot) and the LA(8) wavelet (right-hand).

see the bursts in each of the filtered series near 450 and 1000 meters. An examination of the filtered series for scales higher than 8 indicates that neither burst is evident at scales greater than 16. In the region marked by the thin vertical lines between the two bursts, the filtered series for all scales appear to be consistent with a stationarity assumption, so we have chosen this subseries of 4096 values as a candidate for analysis using the ideas discussed in Section 2.

The left-hand plot of Figure 5 shows the 'σ/τ' curve, a popular analysis tool in the frequency stability literature. This plot shows the square root of the Allan variance (estimated using the maximal-overlap estimator) at different scales plotted versus scale on a log-log plot. Theory suggests that regions of linearity correspond to a power-law process over a particular region of frequencies, with the exponent of the power-law process being related to the slope of the line (in log-log space). The first 7 values of the σ/τ curve fall on such a line almost perfectly. The line drawn between them on the plot was calculated via least squares and has a slope of 0.83. Since $\sigma_Y(k)$ (the square root of the Allan variance) varies approximately as $k^{-(\alpha+1)/2}$ for a power-law process with exponent α (see Section 2),

the σ/τ curve strongly suggests the presence of a power law process over scales of 0.1 to 6.4 meters with an exponent of $\alpha \doteq -2.66 \doteq -8/3$. The right-hand plot is the σ/τ curve corresponding to the LA(8) wavelet for the first 8 scales (recall that filtered series for the higher scales cannot be obtained due to the span of the filters for this wavelet), and it tells quite a different story. The values are not aligned in an obvious straight line as in the case of the Allan variance, and the slope that we found using the first 7 scales of the Allan variance certainly does not look reasonable for portions of the LA(8) wavelet variance.

In Figure 6 we translate the σ/τ curves of the previous figure into estimates of the sdf over different octave bands (these have constant spacing on a log frequency scale). For example, the wavelet variance for a scale of 0.1 meters maps onto the highest octave band; the one for a scale of 0.2 meters maps onto the second highest octave band; and so forth. The thick 'staircase' on this plot corresponds to the Allan variance, while the thin staircase corresponds to the LA(8) wavelet variance. In general, the LA(8) wavelet variance yields an sdf estimate that is lower than the estimate obtained from the Haar wavelet (i.e., the Allan variance). The difference between the two sdf estimates is almost one order of magnitude in three of the octave bands. The dots in this figure show an unsmoothed direct sdf estimate using a discrete prolate spheroidal sequence data taper with time/bandwidth product of 4 (a comparison of the periodogram with direct sdf estimates using a nontrivial data taper indicated that the periodogram suffers from leakage and that tapering is required here). Whereas the LA(8) wavelet sdf estimate agrees fairly well with the direct spectral estimate, the Haar wavelet sdf estimate is too high in several of the octave bands, indicating some form of leakage evidently due to the fact that the Haar wavelet forms a fairly crude set of octave band filters.

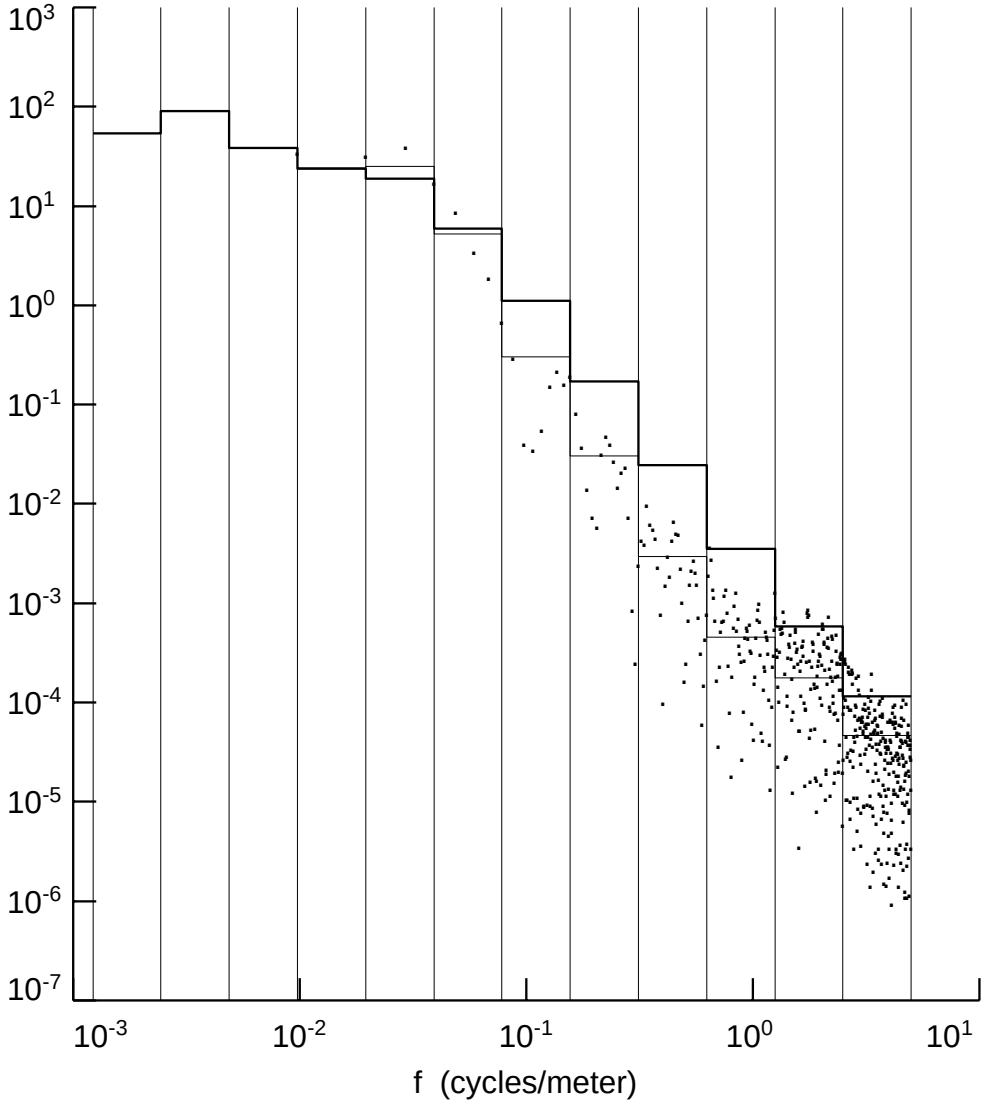

Figure 6. Comparison of spectral estimates implied by σ/τ curves with a direct spectral estimate.

4. Concluding Remarks

Here we make several remarks concerning the results of the previous sections. First, a wavelet analysis of a time series that can be modelled as a power-law process can be used to deduce the properties of the underlying process, but it is dangerous to do so without a careful look at a traditional sdf estimate with good prevention against leakage. For example, if we had computed just the left-hand σ/τ curve of Figure 5, we might have been

badly fooled by the degree to which the points line up as theory suggests they should in the presence of a power-law process.

Second, as others have noted, the wavelet variance is a tool that is well-adapted for studying power-law processes; however, the wavelet variance can lead us to find power-law processes in data that might be best modelled in other ways. Consider, for example, the octave band centered at 0.1 cycles/meter in Figure 6. The wavelet-based estimates of the power in this band are both quite a bit higher than what the direct sdf estimate suggests is reasonable. This octave band corresponds to a scale of 3.2 meters in the σ/τ curves of Figure 5. For both of these curves, the value at 3.2 meters can be obtained almost exactly by linearly interpolating (on a log/log scale) between the points at scales 1.6 and 6.4 meters. This suggests that the wavelet-based estimates might be biased in the sense that regions that do *not* agree with a nominal power-law behaviour will tend to be 'filled in' in a manner consistent with the power-law assumption. What is vitally needed is a careful study of the bias and variance of wavelet-based estimates of the sdf for processes that deviate from a power-law in at least some octave bands.

Third, for purposes of estimating the wavelet variance, results obtained by the frequency stability community using the Haar wavelet (i.e., the Allan variance) strongly favor the maximal overlap estimator. While it seems intuitively reasonable that this result holds for wavelets other than the Haar wavelet, more work is needed to verify that the maximal overlap estimator is indeed superior to the nonoverlapped estimator that would naturally fall out from the discrete wavelet transform (if we ignore those parts of the transform that correspond to circularly filtering a time series).

Fourth, in spite of its poor properties at small scales, the Haar wavelet seems to do quite well at long scales where in fact we cannot use the LA(8) wavelet (in a noncircular fashion).

Fifth, procedures have been worked out in the frequency stability community for placing confidence limits on the Allan variance under the fairly stringent assumption that

the exponent of the power law is known *a priori* (see Greenhall, 1991, for details). In principle, we can obtain similar confidence intervals for other wavelet variances. Also, by defining an estimator for the wavelet variance starting from a prewhitened direct spectral estimate, we can in fact easily obtain statistically valid confidence intervals based upon the sampling properties of direct spectral estimates. (Such an estimator would also allow us to get a handle on the covariance between estimates at different scales, a problem that has yet to be addressed even in the case of the Allan variance.)

Finally, Figure 4 points out that the real potential value of wavelets is in the area of time series with transient events. In the case of vertical shear measurements, wavelet analysis shows that these transients happen at just a few low-order scales. Figure 4 shows that the transient near 450 meters evidently started on scale 8 (i.e., 0.8 meters) and then dissipated latter on into smaller scales.

Acknowledgments

R. Spindel and J. Harlett of the Applied Physics Laboratory, University of Washington, graciously provided discretionary funding to support Percival's work on this manuscript. Guttorp's work was partially funded by NSF grant DM S–9115756.

References

[1] D. W. Allan, 'Statistics of Atomic Frequency Clocks,' *Proceedings of the IEEE*, vol. 54, pp. 221–30, 1966.

[2] J. Beran, 'Statistical Methods for Data with Long-Range Dependence,' *Statistical Science*, vol. 7, pp. 404–16, 1992.

[3] J. Beran and H. R. Künsch, 'Location Estimators for Processes with Long-Range Dependence,' Research Report 40, Seminar für Statistik, ETH, Zürich, 1985.

[4] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[5] P. Flandrin, 'Wavelet Analysis and Synthesis of Fractional Brownian Motion,' *IEEE Transactions on Information Theory*, vol. 38, pp. 910–7, 1992.

[6] C. A. Greenhall, "Recipes for Degrees of Freedom of Frequency Stability Estimators," *IEEE Transactions on Instrumentation and Measurement*, vol. 40, pp. 994–9, 1991.

[7] H. E. Hurst, 'Long-Term Storage Capacity of Reservoirs,' *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–9, 1951.

[8] B. B. Mandelbrot and J. R. Wallis, 'Computer Experiments with Fractional Gaussian Noises,' *Water Resources Research*, vol. 4, pp. 909–918, 1968.

[9] S. Newcomb, 'Astronomical Constants (The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy,' *Supplement to the American Ephemeris and Nautical Almanac for 1897*, U.S. Government Printing Office, Washington, 1895.

[10] D. B. Percival, 'The Statistics of Long Memory Processes,' unpublished Ph.D. dissertation, Department of Statistics, University of Washington, 1983.

[11] D. B. Percival, 'On the Sample Mean and Variance of a Long Memory Process,' Technical Report 69, Department of Statistics, University of Washington, 1985.

[12] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Second Edition), Cambridge University Press, Cambridge, 1992.

[13] A. E. Raftery and J. Haslett, 'Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource (with Discussion),' *Journal of the Royal Statistical Society, Series C*, vol. 38, pp. 1–21, 1989.

[14] H. F. Smith, 'An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops,' *Journal of Agricultural Science*, vol. 28, pp. 1–23, 1938.

[15] R. L. Smith, 'Comment [on Beran, 1992],' *Statistical Science*, vol. 7, pp. 422–5, 1992.

[16] Student, 'Errors in Routine Analysis,' *Biometrika*, vol. 19, pp. 151–64, 1927.