On Estimation of the Wavelet Variance

By DONALD B. PERCIVAL

Applied Physics Laboratory, HN–10, University of Washington,

Seattle, Washington 98195, U.S.A.

Summary

The wavelet variance decomposes the variance of a time series into components associated with different scales. We consider two estimators of the wavelet variance, the first based upon the discrete wavelet transform and the second, called the maximal-overlap estimator, based upon a filtering interpretation of wavelets. We determine the large sample distribution for both estimators and show that the maximal-overlap estimator is more efficient for a class of processes of interest in the physical sciences. We discuss methods for determining an approximate confidence interval for the wavelet variance. We demonstrate through Monte Carlo experiments that the large sample distribution for the maximal-overlap estimator is a reasonable approximation even for the moderate sample size of 128 observations. We apply our proposed methodology to a series of observations related to vertical shear in the ocean.

Some key words: Confidence interval; Fractional difference; Time series analysis; Wavelet transform

1. Introduction

The use of wavelets as a tool for time series analysis and signal processing has increased in recent years due to their potential for solving a number of practical problems; for background on wavelets, see, e.g., Mallat (1989), Strang (1989), Rioul & Vetterli (1991), Daubechies (1992), Press et al. (1992), Donoho (1993), Meyer (1993), Strang (1993), Strichartz (1993) and Vaidyanathan (1993). In particular, wavelets can decompose the variance of a physical process across different scales and have been used in this way in a number of scientific and engineering disciplines; see Gamage (1990), Bradshaw & Spies (1992), Flandrin (1992), Gao & Li (1993), Hudgins, Friehe & Mayer (1993), Kumar & Foufoula-Georgiou (1993), Tewfik, Kim & Deriche (1993) and Wornell (1993). The widespread interest in wavelet-based analysis of variance can be explained by comparing the wavelet decomposition of variance with a similar decomposition given by the spectrum S_Y of a real-valued stochastic process Y_t, $t = 0, \pm 1, \ldots$, with variance $\mathrm{var}\,(Y_t)$. A fundamental property of S_Y is that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} S_Y(f)\, df = \mathrm{var}\,(Y_t); \tag{1}$$

i.e., S_Y decomposes the process variance with respect to a continuous independent variable f, which is known as frequency and has units of, say, cycles per second. The analog of the above equation for the wavelet decomposition is

$$\sum_{l=0}^{\infty} \nu_Y^2(2^l) = \mathrm{var}\,(Y_t), \tag{2}$$

where $\nu_Y^2(\lambda)$ is the wavelet variance associated with the discrete independent variable $\lambda = 2^l$; see Equation (10) for a precise definition. Thus, just as the spectrum decomposes $\mathrm{var}\,(Y_t)$ across frequencies, the wavelet variance decomposes $\mathrm{var}\,(Y_t)$

with respect to λ, a variable known as scale and having units of, say, seconds. Roughly speaking, $\nu_Y^2(\lambda)$ is a measure of how much a weighted average with bandwidth λ of the process $\{Y_t\}$ changes from one time period of length λ to the next. A plot of $\nu_Y^2(\lambda)$ versus λ indicates which scales are important contributors to the process variance; see Figure 2 for an example. If we specialize to the simplest example of a wavelet variance, namely, one based upon the Haar wavelet filter of length 2, the wavelet variance is equal to half the Allan variance, a well-known measure of the performance of atomic clocks (Allan, 1966; Flandrin, 1992; Percival & Guttorp, 1994). Plots of the Allan variance versus λ have been used routinely for nearly 30 years to characterize how well clocks keep time over various time periods; however, the Allan variance can be misleading for interpreting certain geophysical processes, for which wavelet variances based upon a higher order wavelet filter are more appropriate (Percival & Guttorp, 1994).

The wavelet variance is also of interest because it provides a way of regularizing the spectrum. The notions of frequency and scale are closely related so that, under certain reasonable conditions,

$$\nu_Y^2(\lambda) \approx 2 \int_{\frac{1}{4\lambda}}^{\frac{1}{2\lambda}} S_Y(f)\, df; \tag{3}$$

see Equation (10) for the precise relationship between $\nu_Y^2(\lambda)$ and S_Y. The wavelet variance summarizes the information in the spectrum using just one value per octave frequency band and is particularly useful when the spectrum is relatively featureless within each octave band. For example, a model that commonly arises in the physical sciences is that the spectrum obeys the power law $S_Y(f) \propto |f|^\alpha$ over a certain interval of frequencies (Beran, 1992), which, using the above approximation, translates into the statement that $\nu_Y^2(\lambda) \propto \lambda^{-\alpha-1}$ over a corresponding set of scales. A region of linear variation on a plot of log $\nu_Y^2(\lambda)$ versus log λ indicates the existence of a power law behavior, and the slope of the line can be used to deduce the exponent α. For this simple model, there is no information lost in using the summary given by the wavelet variance. If we again specialize to the Haar wavelet variance, the pilot spectrum analysis of Blackman & Tukey (1958, Sec. 18) is identical to using (3) with this wavelet variance. Higher order wavelet variances are a useful generalization because the approximation in (3) improves as the length of the wavelet filter increases.

Because the wavelet variance is a regularization of the spectrum, estimation of the wavelet variance is more straightforward than nonparametric estimation of the spectrum. Suppose for the moment that we have a time series of length $N = 2^K$ that can be regarded as a realization of a portion $Y_1, \ldots, Y_N$ of the stochastic process Y_t. The discrete Fourier transform produces a basic estimator of S_Y called the periodogram, which is given by

$$\hat{S}_Y(f_j) \equiv \frac{1}{N} \left| \sum_{t=1}^{N} (Y_t - \overline{Y}) e^{-i2\pi f_j t} \right|^2 \quad \text{with} \;\; f_j \equiv \frac{j}{N} \;\; \text{and} \;\; \overline{Y} \equiv \frac{1}{N} \sum_{t=1}^{N} Y_t.$$

The periodogram satisfies a sampling version of Equation (1), namely,

$$\frac{1}{N} \sum_{j=-\frac{N}{2}+1}^{\frac{N}{2}} \hat{S}_Y(f_j) = \frac{1}{N} \sum_{t=1}^{N} (Y_t - \overline{Y})^2.$$

A fast Fourier transform algorithm can compute $\hat{S}_Y$ using just $O(N \log_2(N))$ multiplications (Strang, 1993), but the periodogram is not a useful estimator of S_Y because it can be badly biased and is an inconsistent estimator. To deal with these deficiencies, a practitioner must decide if bias is present and, if so, compensate for it using prewhitening and/or tapering, after which the resulting approximately unbiased estimator must be smoothed across frequencies to produce a consistent estimator of S_Y. In contrast, the discrete wavelet transform of $Y_1, \ldots, Y_N$ produces a useful estimator of the wavelet variance. As defined, e.g., in Press et al. (1992, Ch. 13.10), this transform uses a wavelet filter $h_0, \ldots, h_{L-1}$ of even length L to produce K new series, say, $D_{t,\lambda}$, $t = 1, \ldots, \frac{N}{2\lambda}$ with $\lambda = 1, 2, 4, \ldots, 2^{K-1}$. The wavelet variance for scale λ can be estimated using

$$\tilde{\nu}_Y^2(\lambda) \equiv \frac{1}{N} \sum_{t=1}^{\frac{N}{2\lambda}} D_{t,\lambda}^2,$$

which leads to a sampling version of Equation (2) given by

$$\sum_{l=0}^{K-1} \tilde{\nu}_Y^2(2^l) = \frac{1}{N} \sum_{t=1}^{N} (Y_t - \overline{Y})^2.$$

The discrete wavelet transform can be computed 'faster than the fast Fourier transform' in the sense of requiring just $O(N)$ multiplications (Strang, 1993). Whereas the periodogram can be badly biased, an unbiased estimator of $\nu_Y^2(\lambda)$ can easily be constructed based upon the $D_{t,\lambda}$ terms uninfluenced by boundary conditions; moreover, whereas the periodogram is inconsistent, Theorem 2 below can be used to establish consistency for this unbiased estimator. For processes with relatively featureless spectra, the wavelet variance is an attractive alternate characterization that is easy to interpret and estimate.

The purpose of this paper is show how confidence intervals for the wavelet variance can be produced based upon estimators of the wavelet variance. We consider two such estimators, both of which are unbiased. The first is based upon the recognition that $\tilde{\nu}_Y^2(\lambda)$ is a biased estimator of $\nu_Y^2(\lambda)$ due to a fixed number of terms ($\leq L-2$) in the series $D_{t,\lambda}$ that are influenced by boundary conditions. Let $V_{t,\lambda}$ represent the subseries of $D_{t,\lambda}$ uninfluenced by boundary conditions; for example, if $\lambda = 1$, there will be $\frac{N}{2} - \frac{L}{2} + 1$ terms in $V_{t,1}$ as compared to $\frac{N}{2}$ terms in $D_{t,1}$. If the process Y_t can be assumed to have stationary increments of a certain order, the series $V_{t,\lambda}$ is a portion of a stationary process whose variance is proportional to the wavelet variance. We refer to the estimator of $\nu_Y^2(\lambda)$ based upon $V_{t,\lambda}$ as the wavelet-transform estimator. In Section 3, however, we find that the wavelet variance can be more efficiently estimated by another estimator that make uses of a nonsubsampled version of the discrete wavelet transform. The motivation for this estimator, which we call the maximal-overlap estimator, is based upon the fact that the series $V_{t,\lambda}$ can be obtained by filtering Y_t with a wavelet filter $h_{l,\lambda}$, $l = 0, \ldots, (2\lambda - 1)(L - 1)$, and then subsampling every 2λth value of the filter output; here $h_{l,1} \equiv h_l$, and the wavelet filters for $\lambda > 1$ depend on just h_l. The maximal-overlap estimator is based upon the sample variance of the output from the wavelet filter without any subsampling. Because the wavelet filter for scale λ can be regarded as an approximation to a band-pass filter with a passband defined by the set of frequencies f such that $\frac{1}{4\lambda} \leq |f| \leq \frac{1}{2\lambda}$, a heuristic argument can be made that nothing much is to be gained by using the maximal-overlap estimator instead of the easily computed wavelet-transform estimator; see the discussion immediately following Equation (4). A main thrust of this paper is that in fact the asymptotic relative efficiency of the wavelet-transform estimator with respect to the maximal-overlap estimator is always less than unity and can in fact approach one half for certain processes of interest in the physical sciences. Moreover, there exists a 'pyramid' algorithm for computing the terms needed for the maximal-overlap estimator (Percival & Guttorp, 1994). This algorithm requires $O(N \log_2(N))$ multiplications and is not restricted to sample sizes N that are powers of two, so computation of the maximal-overlap estimator is certainly feasible.

Section 3 gives the large sample distribution of the maximal-overlap estimator, while Section 4 discusses four ways to obtain confidence intervals for the wavelet variance based upon this estimator. Section 5 summarizes some Monte Carlo experiments that indicate only a moderate sample size of 128 observations is needed for the large sample theory of Section 3 to be a reasonable approximation. To simplify and focus our discussion, Sections 2 to 5 discuss the wavelet variance for scale $\lambda = 1$ only, so in Section 6 we indicate how the unit scale results can be adapted to larger scales. Finally we demonstrate in Section 7 how our results can be used to attach a measure of uncertainty to estimates of the wavelet variance for a time series related to vertical shear in the ocean. Proofs of Theorems 1 and 2 below are omitted to conserve space, but can be obtained upon request from the author via traditional mail or electronic mail to the Internet address `dbp@apl.washington.edu`.

2. The wavelet variance

Suppose that Y_t is a stochastic process whose dth order backward difference

$$Z_t \equiv (1 - B)^d Y_t = \sum_{j=0}^{d} \binom{d}{j} (-1)^j Y_{t-j}$$

is a second-order stationary process with zero mean and spectrum S_Z; here d is a nonnegative integer, and B is the backward shift operator defined by $BY_t \equiv Y_{t-1}$ so that $B^j Y_t = Y_{t-j}$. If Y_t were itself stationary with spectrum S_Y, the theory of linear filters says that S_Y and S_Z would be related by $S_Y(f) = S_Z(f)/\mathcal{D}^d(f)$, where $\mathcal{D}(f) \equiv 4\sin^2(\pi f)$ is the squared modulus of the transfer function for a first order backward difference filter; if Y_t is not stationary, then $S_Y(f)$ is defined as $S_Z(f)/\mathcal{D}^d(f)$ (Yaglom, 1958). Note that, if $d = 0$, then Y_t is necessarily stationary, in which case the processes Y_t and Z_t are identical.

Let $h_0, \ldots, h_{L-1}$ denote the coefficients of a compactly supported Daubechies wavelet filter of even length L (Daubechies, 1992, Ch. 6). We assume the normalization $\sum h_l^2 = 1$. Let

$$H(f) \equiv \sum_{l=0}^{L-1} h_l e^{-i2\pi f l}$$

be the transfer function for h_l. The wavelet filter h_l can be regarded as an approximation to a high-pass filter with a passband defined by $\frac{1}{4} < |f| \leq \frac{1}{2}$. The modulus squared of H can be written explicitly as

$$\mathcal{H}(f) \equiv |H(f)|^2 = \mathcal{D}^{\frac{L}{2}}(f)\mathcal{C}(f) \quad \text{where} \quad \mathcal{C}(f) \equiv \frac{1}{2^{L-1}} \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \cos^{2l}(\pi f)$$

(Daubechies, 1992, Ch. 6.1). We can thus regard h_l as a two-stage filter, the first stage of which is an $\frac{L}{2}$th order backward difference filter, and the second of which uses a filter whose modulus squared transfer function is $\mathcal{C}$. Different factorizations of $\mathcal{C}$ lead to wavelet filters with necessarily the same modulus squared transfer function but with different phase properties (Daubechies, 1992, Ch. 6.4 and 8.1.1).

Let

$$W_t \equiv \sum_{l=0}^{L-1} h_l Y_{t-l}$$

represent the output obtained from filtering Y_t using the wavelet filter.

Theorem 1. *If $L \geq 2d$, then W_t is a stationary process with zero mean and spectrum defined by $S_W(f) = \mathcal{H}(f)S_Y(f)$.*

The wavelet variance of unit scale is just half the variance of W_t, i.e.,

$$\nu^2 \equiv E(W_t^2)/2.$$

Since the variance of a stationary process is equal to the integral of its spectrum, we have

$$\nu^2 = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} S_W(f)\,df = \frac{1}{4^d} \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos^{2l}(\pi f)\sin^{L-2d}(\pi f) S_Z(f)\,df.$$

The condition $L \geq 2d$ of Theorem 1 ensures that the product $\cos^{2l}(\pi f)\sin^{L-2d}(\pi f)$ in the integrand is bounded by unity and hence that the integral is finite.

3. Estimation of the wavelet variance

Suppose now that we are given a time series that can be regarded as a realization of one portion $Y_1, \ldots, Y_N$ of the process Y_t and that we want to estimate the wavelet variance ν^2 initially for unit scale only. We consider two estimators, both of which are based upon W_t, $t = L, \ldots, N$. The first estimator is the maximal-overlap estimator

$$\hat{\nu}_W^2 \equiv \frac{1}{2N_W} \sum_{t=L}^{N} W_t^2 \quad \text{with} \quad N_W \equiv N - L + 1,$$

i.e., the sample variance of the W_t's under the assumption that $E(W_t) = 0$. The terminology 'maximal-overlap' was used by Greenhall (1991) in a study of the Allan variance. The second estimator is the wavelet-transform estimator

$$\hat{\nu}_V^2 \equiv \frac{1}{2N_V} \sum_{t=\frac{L}{2}}^{\lfloor \frac{N}{2} \rfloor} V_t^2, \quad \text{where } V_t \equiv W_{2t} \text{ and } N_V \equiv \lfloor \tfrac{N}{2} \rfloor - \tfrac{L}{2} + 1,$$

i.e., the sample variance of the W_t's after subsampling every other observation; here $\lfloor x \rfloor$ refers to the greatest integer less than or equal to x. As discussed in Section 1, $\hat{\nu}_V^2$ is the 'natural' estimator of the wavelet variance that we would obtain from the discrete wavelet transform after discarding all terms influenced by boundary conditions.

Under the assumption that W_t and hence V_t are Gaussian processes, we wish to compare the variances $\operatorname{var}(\hat{\nu}_W^2)$ and $\operatorname{var}(\hat{\nu}_V^2)$ of $\hat{\nu}_W^2$ and $\hat{\nu}_V^2$ for large samples. A standard result in spectral analysis (Anderson, 1971, p. 388) tells us that the subsampled process V_t is a stationary process with spectrum S_V given by

$$S_V(f) \equiv \frac{S_W(\frac{f}{2}) + S_W(\frac{f}{2} + \frac{1}{2})}{2}, \quad -\tfrac{1}{2} \leq f \leq \tfrac{1}{2},$$

where the spectrum S_W is defined for $|f| > \frac{1}{2}$ by periodic extension.

Theorem 2. *If S_W is finitely square integrable and strictly positive almost everywhere, then the estimators $\hat{\nu}_W^2$ and $\hat{\nu}_V^2$ are asymptotically normally distributed with mean ν^2 and large sample variances $A_W/(2N_W)$ and $A_V/(2N_V)$, respectively, where*

$$A_W \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} S_W^2(f)\, df \quad \textit{and} \quad A_V \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} S_V^2(f)\, df.$$

To compare var $(\hat{\nu}_W^2)$ and var $(\hat{\nu}_V^2)$, we use the asymptotic relative efficiency of $\hat{\nu}_V^2$ with respect to $\hat{\nu}_W^2$, which by definition is

$$\mathcal{E} \equiv \lim_{N\to\infty} \frac{\text{var}\,(\hat{\nu}_W^2)}{\text{var}\,(\hat{\nu}_V^2)} = \frac{A_W}{2A_V} = \left(1 + \frac{\int_{-\frac{1}{2}}^{\frac{1}{2}} S_W(\frac{f}{2}) S_W(\frac{f}{2} + \frac{1}{2})\, df}{\int_{-\frac{1}{2}}^{\frac{1}{2}} S_W^2(f)\, df} \right)^{-1}. \qquad (4)$$

The last expression for $\mathcal{E}$ tells us that $\mathcal{E} < 1$ because, under the assumptions for Theorem 2, S_W is strictly positive almost everywhere and hence

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} S_W(\tfrac{f}{2}) S_W(\tfrac{f}{2} + \tfrac{1}{2})\, df > 0.$$

Heuristic use of the above also indicates why it might be argued there is little to be gained in using the maximal-overlap estimator $\hat{\nu}_W^2$ in place of the wavelet-transform estimator $\hat{\nu}_V^2$. Recall that the wavelet filter is regarded as an approximation to a high-pass filter with passband defined by $\frac{1}{4} < |f| \leq \frac{1}{2}$. If it were perfectly so, we would have

$$S_W(f) = \begin{cases} 0, & |f| \leq \frac{1}{4}; \\ S_Y(f), & \frac{1}{4} < |f| \leq \frac{1}{2}; \end{cases}$$

so that $\int S_W(\frac{f}{2}) S_W(\frac{f}{2} + \frac{1}{2})\, df = 0$ and hence $\mathcal{E} = 1$; however, as shown in Table 1 below, $\mathcal{E}$ is in fact considerably less than unity for certain processes, indicating that the shortness of the wavelet filter yields a rather imperfect high-pass filter.

We can evaluate $\mathcal{E}$ analytically for certain choices of S_Y. As a simple example, suppose that $S_Y(f) \propto |\sin(\pi f)|^\alpha$ so that S_Y varies as $|f|^\alpha$ for frequencies close to zero. Processes with such spectra occur in a wide range of applications (Beran, 1992). Note that the process Y_t corresponding to S_Y is stationary if $\alpha > -1$; if in addition $\alpha < 1$, then Y_t corresponds to a stationary and invertible fractional difference process (Granger & Joyeux, 1980; Hosking, 1981). Table 1 shows $\mathcal{E}$ for three 'blue noise' processes $\alpha = 1$, $\frac{1}{2}$ and $\frac{1}{4}$, a white noise process $\alpha = 0$, and two stationary and three nonstationary 'red noise' processes $\alpha = -\frac{1}{4}$, $-\frac{1}{2}$, -1, -2 and -3, all in combination with wavelet filters of length $L = 2$, i.e., the Haar wavelet filter, 4, 6 and 8. The dash in the '$\alpha = -3$, $L = 2$' entry indicates that the assumptions of Theorem 2 do not hold. The tabulated values have been obtained via straightforward, but tedious, analytical computations and verified using numerical integration. Note that the efficiency decreases as α decreases because the proportion of variance attributable to frequencies outside the passband $\frac{1}{4} < |f| \leq \frac{1}{2}$ increases as α decreases, whereas the efficiency increases as L increases because the wavelet filter becomes a better approximation to a high-pass filter as L increases.

4. Confidence intervals for the wavelet variance

In order to use Theorem 2 in practical applications to determine a confidence interval for ν^2 based upon $\hat{\nu}_W^2$, we must estimate A_W, which is the integral of S_W^2. Since A_W will be dominated by large values of S_W, we can just use the periodogram $\hat{S}_W$ of the W_t's as our estimator of S_W since spectral leakage is not a concern. Standard statistical theory suggests that, for large N, the ratio $2\hat{S}_W(f)/S_W(f)$ is distributed as a χ^2 random variable with two degrees of freedom if $0 < |f| < \frac{1}{2}$, from

Table 1. *Asymptotic relative efficiencies $\mathcal{E}$ of the wavelet-transform estimator $\hat{\nu}_V^2$ with respect to the maximal-overlap estimator $\hat{\nu}_W^2$ for processes with a spectrum proportional to $|\sin(\pi f)|^\alpha$ and compactly supported Daubechies wavelet filters of length $L = 2$, 4, 6 and 8. Note that $\mathcal{E} < 1$ implies that $\hat{\nu}_W^2$ has smaller large sample variance than $\hat{\nu}_V^2$.*

α	$L = 2$	$L = 4$	$L = 6$	$L = 8$
1	0·85	0·89	0·91	0·92
$\frac{1}{2}$	0·81	0·86	0·89	0·90
$\frac{1}{4}$	0·78	0·84	0·87	0·89
0	0·75	0·82	0·85	0·87
$-\frac{1}{4}$	0·72	0·80	0·83	0·86
$-\frac{1}{2}$	0·68	0·77	0·81	0·84
-1	0·61	0·72	0·77	0·80
-2	0·50	0·61	0·67	0·71
-3	—	0·52	0·58	0·62

which we obtain $E(\hat{S}^2_W(f)) \approx 2S^2_W(f)$. Since the contribution due to the special frequencies $f = 0$ and $\pm\frac{1}{2}$ becomes insignificant as N get large, we can take

$$\hat{A}_W \equiv \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{S}^2_W(f)\, df$$

to be an approximately unbiased estimator of A_W for large N. Finally we can use Parseval's theorem to obtain the convenient computational formula

$$\hat{A}_W = \frac{\hat{s}^2_{0,W}}{2} + \sum_{\tau=1}^{N_W-1} \hat{s}^2_{\tau,W},$$

where $\hat{s}_{\tau,W}$ is the usual biased estimator of the autocovariance sequence $s_{\tau,W}$ corresponding to S_W (Priestley, 1981, p. 322). Under the restrictive assumption that the estimator $\hat{A}_W$ is close to A_W, an approximate $100(1-2p)\%$ confidence interval for ν^2 would be given by

$$\left[\hat{\nu}^2_W - \Phi^{-1}(1-p)(\hat{A}_W/2N_W)^{\frac{1}{2}}, \hat{\nu}^2_W + \Phi^{-1}(1-p)(\hat{A}_W/2N_W)^{\frac{1}{2}}\right],$$

where $\Phi^{-1}(p)$ is the $p \times 100\%$ percentage point for the standard Gaussian distribution. Alternatively, we can use an 'equivalent degrees of freedom' argument (Priestley, 1981, p. 466) to claim that $\eta\hat{\nu}^2_W/\nu^2$ is approximately equal in distribution to a χ^2 random variable with η degrees of freedom for large N, where

$$\eta \equiv \frac{2\left\{E(\hat{\nu}^2_W)\right\}^2}{\text{var}\,(\hat{\nu}^2_W)} \approx \frac{4N_W\nu^4}{A_W}. \tag{5}$$

An approximate $100(1-2p)\%$ confidence interval for ν^2 would be given by

$$\left[\frac{\eta\hat{\nu}^2_W}{Q_\eta(1-p)}, \frac{\eta\hat{\nu}^2_W}{Q_\eta(p)}\right], \tag{6}$$

where $Q_\eta(p)$ is the $p \times 100\%$ percentage point for the χ^2_η distribution. The degrees of freedom η would be estimated using $4N_W\hat{\nu}^4_W/\hat{A}_W$.

Another approach to obtaining a confidence interval for ν^2 is to assume that S_Y, and hence S_W, is known to within a multiplicative constant; i.e., we suppose that, say, $S_W(f) = hS_0(f)$, where S_0 is a known function and h is an unknown constant. This assumption is used to obtain the confidence intervals for the Allan variance discussed in Greenhall (1991). By Parseval's theorem we have

$$\hat{\nu}_W^2 = \frac{1}{2N_W} \sum_{k=0}^{N_W-1} \hat{S}_W(f_k) \approx \frac{1}{N_W} \sum_{k=1}^{M} \hat{S}_W(f_k) + \frac{1}{2N_W} \hat{S}_W(\tfrac{1}{2}) I_{N_W}, \tag{7}$$

where $f_k \equiv k/N_W$; $M \equiv \lfloor \frac{N_W}{2} - \frac{1}{2} \rfloor$; and I_{N_W} is unity if N_W is even and zero if N_W is odd. The approximation above merely says that $\hat{S}_W(0)/N_W = \overline{W}^2$ is negligible, where $\overline{W}$ is the sample mean of the W_t's. Under the usual large sample approximations that $2\hat{S}_W(f_k)/S_W(f_k)$ with $0 < f_k < \frac{1}{2}$ and $\hat{S}_W(\frac{1}{2})/S_W(\frac{1}{2})$ are equal in distribution to χ^2 random variables with, respectively, two and one degrees of freedom and that the random variables in the right-hand side of Equation (7) are independent, we can again use an equivalent degrees of freedom argument to claim that $\eta\hat{\nu}_W^2/\nu^2$ is approximately equal in distribution to a χ^2 random variable with η degrees of freedom for large N, where now

$$\eta = \frac{\left(2\sum_{k=1}^{M} S_0(f_k) + S_0(\frac{1}{2}) I_{N_W}\right)^2}{2\sum_{k=1}^{M} S_0^2(f_k) + S_0^2(\frac{1}{2}) I_{N_W}}. \tag{8}$$

An approximate $100(1-2p)\%$ confidence interval for ν^2 would again be given by (6).

A further simplification is to recall that the wavelet filter h_l can be regarded as an approximate band-pass filter with passband defined by $\frac{1}{4} < |f| \leq \frac{1}{2}$. This fact suggests that it might be reasonable to assume in certain practical problems that S_W is band-limited and flat over its nominal passband. The validity of this assumption can readily be assessed by examining an estimate of S_W. The equation for the equivalent degrees of freedom then simplifies to

$$\eta = 2(M - \lfloor \tfrac{N_W}{4} \rfloor) + I_{N_W} \approx \tfrac{N_W}{2}. \tag{9}$$

If the sample size N_W is large enough (the next section suggests that $N_W = 128$ is often sufficient), a confidence interval based upon (6) with η estimated by $4N_W \hat{\nu}_W^4 / \hat{A}_W$ is likely to be reasonably accurate, and we would recommend this as the method of choice. For smaller sample sizes, this method can yield overly optimistic confidence intervals in some instances, in which case a confidence interval based on η from (8) or (9) is a useful check and should be preferred if it is markedly wider than the one based on $4N_W \hat{\nu}_W^4 / \hat{A}_W$. Use of (8) requires a reasonable guess at the shape of S_W; if such a guess is not available, η should be based upon (9).

5. Monte carlo experiments

We performed Monte Carlo experiments to assess whether the large sample variance stated for $\hat{\nu}_W^2$ in Theorem 2 is reasonably accurate for sample sizes of interest in practical applications. We generated 10^5 realizations of lengths $N_W = 128$ for each of the 9 processes indexed by α in Table 1. For stationary processes Y_t, each realization was produced by multiplying the Cholesky factorization of the inverse of the covariance matrix for the Y_t's times a vector containing a realization of a Gaussian white noise process; for nonstationary processes, the stationary differenced process Z_t was so produced, and then Y_t was generated via cumulative summation. The wavelet variance was estimated for each of these realizations using the maximal-overlap estimator $\hat{\nu}_W^2$ with the Daubechies extremal phase wavelet filters of lengths $L = 2$, 4, 6 and 8. Let $\hat{\nu}_{j,W}^2$ be the estimate from the jth realization. The ratio of the sample variance of the $\hat{\nu}_{j,W}^2$'s to the large sample approximation for var $(\hat{\nu}_W^2)$, i.e., $2N_W \sum(\hat{\nu}_{j,W}^2 - \nu^2)^2 / (10^5 A_W)$, was found to quite close to unity in all cases, with the smallest ratio being 0·982 for $\alpha = \frac{1}{4}$ and $L = 8$, and the largest being 1·017 for $\alpha = -2$ and $L = 4$. This result indicates that the large sample variance quoted in Theorem 2 is a reasonable approximation even for the moderate sample size $N_W = 128$.

We also considered how well we can estimate A_W using $\hat{A}_W$ from a given realization. Let $\hat{A}_{j,W}$ be the estimate from the jth realization. The ratio of the sample mean of the $\hat{A}_{j,W}$'s to A_W, i.e., $\sum \hat{A}_{j,W} / (10^5 A_W)$, was again quite close to unity, with the smallest ratio being 0·994 for $\alpha = 0$ and $L = 8$, and the largest being 1·005 for $\alpha = -2$ and $L = 2$. This result indicates that $\hat{A}_W$ is an approximately unbiased estimator of A_W. Finally we considered how well we can estimate the equivalent degrees of freedom η of Equation (5). Let $\hat{\eta}_j$ represents the estimate from the jth realization. Even though η varies from 68 for $\alpha = 1$ and $L = 8$ to 128 for $\alpha = -2$ and $L = 2$, the ratio of the sample mean of the $\hat{\eta}_j$ to η was fairly constant and indicates a small positive bias in $\hat{\eta}$; e.g., the ratio ranged from 1·02 to 1·07 for $L = 2$ and from 1·05 to 1·06 for $L = 8$. The coefficient of variation, i.e., ratio of the sample standard deviation to the sample mean of the $\hat{\eta}_j$'s, was also fairly constant, with values ranging from 0·09 to 0·15 for $L = 2$ and from 0·12 to 0·14 for $L = 8$. These values indicate that some caution must be exercised in interpreting confidence intervals based upon estimation of η from moderate sample sizes.

6. Extension to higher scales

Here we sketch briefly how the material in Sections 2 to 4 can be extended to handle a higher scale $\lambda = 2^\Lambda$, where Λ is a positive integer. Given the wavelet filter h_l of unit scale, let g_l be the corresponding so-called scaling filter, which is defined as $g_l \equiv (-1)^{l+1} h_{L-l-1}$ for $l = 0, \ldots, L-1$. The scaling filter g_l can be regarded as an approximation to a low-pass filter with a passband defined by $-\frac{1}{4} \leq |f| \leq \frac{1}{4}$. Let G denote the transfer function for g_l, and let $\mathcal{G}$ be the squared modulus of G, which can be written explicitly as

$$\mathcal{G}(f) = 2\cos^L(\pi f) \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \sin^{2l}(\pi f)$$

(Daubechies, 1992, Ch. 6.1). The wavelet and scaling filters $h_{l,\lambda}$ and $g_{l,\lambda}$ for scale λ are both of length $L_\lambda \equiv (2\lambda - 1)(L-1) + 1$ and have transfer functions H_λ and G_λ satisfying

$$G_\lambda(f) = \prod_{l=0}^{\Lambda} G(2^l f) \text{ and } H_\lambda(f) = H(\lambda f) \prod_{l=0}^{\Lambda-1} G(2^l f) = H(\lambda f) G_{\frac{\lambda}{2}}(f),$$

where $G_1 \equiv G$. The wavelet filter $h_{l,\lambda}$ for scale λ can be regarded as an approximation to a band-pass filter with passband given by $\frac{1}{4\lambda} < |f| \leq \frac{1}{2\lambda}$, whereas the scaling filter $g_{l,\lambda}$ approximates a low-pass filter with passband $-\frac{1}{4\lambda} \leq f \leq \frac{1}{4\lambda}$. The squared moduli of $H_\lambda(f)$ and $G_\lambda(f)$ obey the relationships $\mathcal{H}_\lambda(f) = \mathcal{H}(\lambda f)\mathcal{G}_{\frac{\lambda}{2}}(f)$ and $\mathcal{G}_\lambda(f) = \prod_{l=0}^{\Lambda} \mathcal{G}(2^l f)$ with $\mathcal{G}_1 \equiv \mathcal{G}$.

Let

$$W_{t,\lambda} \equiv \sum_{l=0}^{L_\lambda - 1} h_{l,\lambda} Y_{t-l}$$

represent the output obtained from filtering Y_t using the wavelet filter of scale λ. The analogy of Theorem 1 for scale λ is that, if $L \geq 2d$, then $W_{t,\lambda}$ is a stationary process with zero mean and spectrum defined by $S_{W_\lambda}(f) \equiv \mathcal{H}_\lambda(f) S_Y(f)$. The wavelet variance at scale λ for the process Y_t is defined as

$$\nu_Y^2(\lambda) \equiv \frac{E(W_{t,\lambda}^2)}{2\lambda} = \frac{1}{2\lambda} \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{H}_\lambda(f) S_Y(f)\, df. \tag{10}$$

Given $Y_1, \ldots, Y_N$, the maximal-overlap estimator of $\nu_Y^2(\lambda)$ is defined as

$$\hat{\nu}_Y^2(\lambda) \equiv \frac{1}{2\lambda N_{W_\lambda}} \sum_{t=L_\lambda}^{N} W_{t,\lambda}^2 \quad \text{with} \quad N_{W_\lambda} \equiv N - L_\lambda + 1.$$

Under the same conditions as given in Theorem 2, the estimator $\hat{\nu}_Y^2(\lambda)$ is asymptotically normal with mean $\nu_Y^2(\lambda)$ and large sample variance $A_{W_\lambda}/(2\lambda^2 N_{W_\lambda})$, where $A_{W_\lambda} \equiv \int S_{W_\lambda}^2(f)\,df$. Similar results can be stated for the wavelet-transform estimator. Limited calculations to date indicate that, at higher scales, the asymptotic relative efficiency $\mathcal{E}$ of the wavelet-transform estimator with respect to the maximal-overlap estimator assumes the same range of values as displayed in Table 1; i.e., the maximal-overlap estimator is the more efficient of the two estimators, with $\mathcal{E}$ dropping close to 0·5 for some processes. Finally the methods given in Section 4 for generating a confidence interval for the wavelet variance can be readily adapted to the scale λ case, with Equations (5) and (9) now becoming, respectively, $\eta = \max\{1, 4\lambda^2 N_{W_\lambda} \nu^4 / A_{W_\lambda}\}$ and $\eta \approx \max\{1, N_{W_\lambda}/(2\lambda)\}$.

7. An example

Here we illustrate the methodology of the previous sections by considering a 'time' series related to vertical ocean shear. This series was collected by an instrument that was dropped over the side of a ship and then descended vertically into the ocean. As it descended, the probe collected measurements concerning the ocean as a function of depth, one of which is the x component of the velocity of water. This velocity was measured every $\Delta = 0{\cdot}1$ meter, first differenced over an interval of 10 meters, and then low-pass filtered to obtain the series of $N = 4096$ values extending from a depth of 489·5 meters down to 899·0 meters shown in Figure 1.

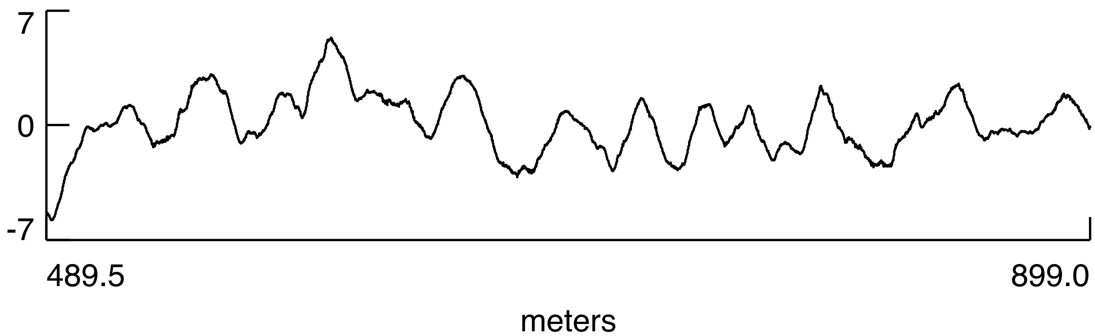

Fig. 1. Plot of measurements related to vertical shear in the ocean versus depth in meters. This series was collected and supplied by Mike Gregg, Applied Physics Laboratory, University of Washington. As of 1995, this series could be obtained via electronic mail by sending a message with the single line '`send lmpavw from datasets`' to the Internet address `statlib@lib.stat.cmu.edu`, which is the address for StatLib, a statistical archive maintained by Carnegie Mellon University.

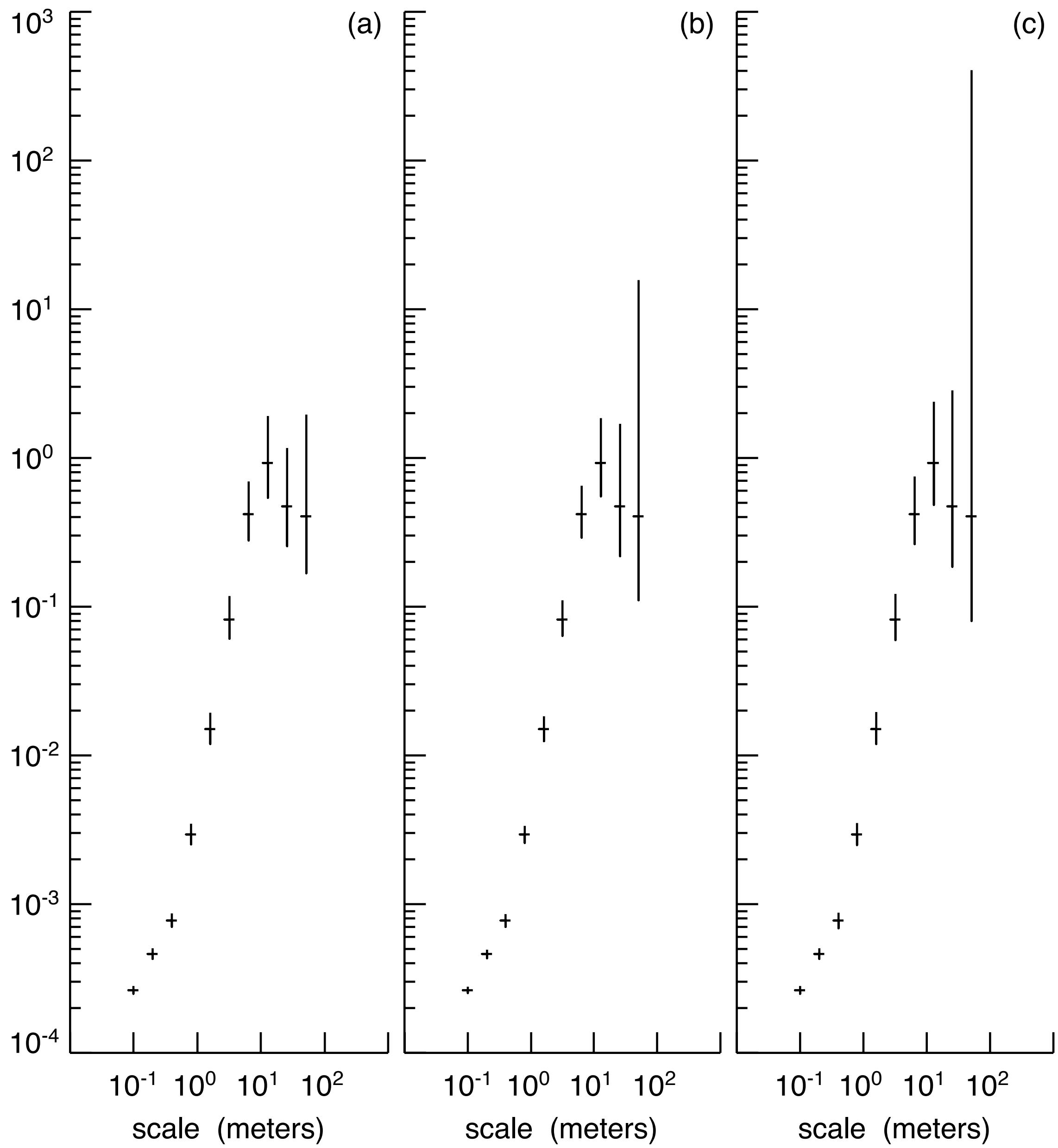

Fig. 2. 95% confidence intervals for the wavelet variance of series in Figure 1 based upon the maximal-overlap estimate using a χ^2 approximation with degrees of freedom determined by (a) estimation of $4\lambda^2 N_{W_\lambda} \nu^4 / A_{W_\lambda}$, (b) the nominal model $S_Y(f) \propto |f|^{-\frac{8}{3}}$ and (c) the simple approximation $N_{W_\lambda}/(2\lambda)$.

The wavelet variance was estimated for scales $\lambda = 1, 2, 4, \ldots, 512$ using the Daubechies extremal phase D_4 wavelet filter, for which $h_0 \propto 1 - \sqrt{3}$, $h_1 \propto -3 + \sqrt{3}$, $h_2 \propto 3 + \sqrt{3}$ and $h_3 \propto -1 - \sqrt{3}$ with the constant of proportionality $4\sqrt{2}$ insuring that $\sum h_l^2 = 1$. Figure 2 shows these estimates plotted versus physical scale $\lambda\Delta$ up to 51·2 meters, along with three 95% confidence intervals for the true wavelet variance. This figure indicates that the variance in the series is mainly due to fluctuations at scales 6·4 meters and longer, which can be associated with deep jets and internal waves in the ocean. Note that $\log \hat{\nu}_Y^2(\lambda)$ versus $\log \lambda$ varies approximately linearly over low scales 0·1 to 0·4 meter and also over intermediate scales 0·8 to 6·4 meters. The low scales are influenced mainly by turbulence, and the slope of $\log \hat{\nu}_Y^2(\lambda)$ versus $\log \lambda$ indicates that turbulence rolls off at a rate consistent with a power law of exponent $\alpha = -1{\cdot}8$, a result that can be compared to physical models. The power law rolloff of $\alpha = -3{\cdot}4$ at intermediate scales can be interpreted as a transition region between the internal wave and turbulent regions.

The three confidence intervals in Figure 2 are based upon the maximal-overlap estimate and a χ^2 approximation with degrees of freedom η determined by (a) estimation of $4\lambda^2 N_{W_\lambda} \nu^4 / A_{W_\lambda}$, (b) Equation (8) with the nominal model $S_Y(f) \propto |f|^{-\frac{8}{3}}$ suggested by a very crude spectral analysis for the time series (Percival & Guttorp, 1994), and (c) the simple approximation $N_{W_\lambda}/(2\lambda)$. At scales 6·4 meters and below, the confidence intervals for the three methods are interchangeable from a practitioner's point of view, but, not surprisingly, the agreement breaks down at larger scales. The equivalent degrees of freedom for methods (b) and (c) are, respectively, only 22·0 and 13·0 at scale 12·8 meters; 8·3 and 5·0 at scale 25·6 meters; and 2·0 and 1·0 at scale 51·2 meters. Because the degrees of freedom are so small for these scales, the large sample approximation (a) cannot be trusted, but, whereas the lengths of the confidence intervals for methods (b) and (c) are within a factor of two of each other for scales 12·8 and 25·6 meters, the same cannot be said at 51·2 meters. Thus, the three methods yield quite similar confidence intervals when the number of equivalent degrees of freedom is large, with approximations (b) and (c) being more valuable for smaller degrees of freedom. The confidence intervals can be used to assess whether or not fluctuations at, e.g., scale 25·6 meters for this particular series agree with other sets of measurements taken at different locations in the ocean.

Acknowledgments

The author wishes to thank Ron Lindsay for many helpful discussions; Chuck Greenhall for the suggestion to estimate the degrees of freedom in Equation (5); Mike Gregg for supplying the data and for discussions concerning it; and the referees and editors for their very helpful criticisms. This research was supported by an Office of Naval Research grant entitled 'Surface Heat Flux from AVHRR Ice Surface Temperature' (Drew Rothrock and Ron Lindsay, Co-Principal Investigators).

References

Allan, D. W. (1966). Statistics of atomic frequency clocks. *Proc. IEEE* **54**, 221–30.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley & Sons.

Blackman, R. B. & Tukey, J. W. (1958). *The Measurement of Power Spectra.* New York: Dover.

Bradshaw, G. A. & Spies, T. A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *J. Ecology* **80**, 205—15.

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statist. Sci.* **7**, 404–16.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* Philadelphia: SIAM.

Donoho, D. L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Different Perspectives on Wavelets: Proceedings of Symposia in Applied Mathematics, Vol. 47*, Ed. I. Daubechies, 173–205. Providence: American Mathematical Society.

Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Info. Theory* **38**, 910–7.

Gamage, N. K. K. (1990). Detection of coherent structures in shear induced turbulence using wavelet transform methods. In *Ninth Symposium on Turbulence and Diffusion*, 389–92. Boston: American Meteorological Society.

Gao, W. & Li, B. (1993). Wavelet analysis of coherent structures at the atmosphere-forest interface. *J. Appl. Meteor.* **32**, 1717—25.

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

Greenhall, C. A. (1991). Recipes for degrees of freedom of frequency stability estimators. *IEEE Trans. Instr. Meas.* **40**, 994–9.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.

Hudgins, L., Friehe, C. A. & Mayer, M. E. (1993). Wavelet transforms and atmospheric turbulence. *Phys. Rev. Letters* **71**, 3279—82.

Kumar, P. & Foufoula-georgiou, E. (1993). A multicomponent decomposition of spatial rainfall fields: 1. Segregation of large- and small-scale features using wavelet transforms. *Water Resour. Res.* **29**, 2515—32.

MALLAT, S. G. (1989). Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoust. Speech Sig. Proc.* **37**, 2091–110.

MEYER, Y. (1993). *Wavelets: Algorithms and Applications.* Philadelphia: SIAM.

PERCIVAL, D. B. & GUTTORP, P. (1994). Long-memory processes, the Allan variance and wavelets. In *Wavelets in Geophysics*, Ed. E. Foufoula-Georgiou and P. Kumar, 325–57. New York: Academic Press.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1992). *Numerical Recipes: The Art of Scientific Computing* (Second Edition). Cambridge: Cambridge University Press.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series.* New York: Academic Press.

RIOUL, O. & VETTERLI, M. (1991). Wavelets and signal processing. *IEEE Sig. Proc. Magazine* **8**, 14–38.

STRANG, G. (1989). Wavelets and dilation equations: a brief introduction. *SIAM Rev.* **31**, 614–27.

STRANG, G. (1993). Wavelet transforms versus Fourier transforms. *Bull. Amer. Math. Soc.* **28**, 288–305.

STRICHARTZ, R. S. (1993). How to make wavelets. *Amer. Math. Mon.* **100**, 539–56.

TEWFIK, A. H., KIM, M. & DERICHE, M. (1993). Multiscale signal processing techniques: a review. In *Signal Processing and its Applications: Handbook of Statistics, Vol. 10*, Ed. N. K. Bose and C. R. Rao, 819–81. Amsterdam: North-Holland.

VAIDYANATHAN, P. P. (1993). *Multirate Systems and Filter Banks.* Englewood Cliffs, New Jersey: Prentice-Hall.

WORNELL, G. W. (1993). Wavelet-based representations for the $1/f$ family of fractal processes. *Proc. IEEE* **81**, 1428–50.

YAGLOM, A. M. (1958). Correlation theory of processes with stationary random increments of order n. *Amer. Math. Soc. Transl., Ser. 2* **8**, 87–141.