ROBUST ESTIMATION OF WAVELET VARIANCE

Debashis Mondal and Donald B. Percival

University of Chicago and University of Washington

Abstract: The wavelet variance provides a scale-based decomposition of the process variance for a time series or a random field and has been used to analyze various multi-scale processes. Examples of such processes include atmospheric pressure, deviations in time as kept by atomic clocks, soil properties in agricultural plots, snow fields in the polar regions and brightness temperature maps of South Pacific clouds. In practice, data collected in the form of a time series or a random field often suffer from contamination that is unrelated to the process of interest. This paper describes robust estimation of the wavelet variance that can act as a guard against such contamination. A new M-estimation procedure that works for both time series and random fields is proposed, and its large sample theory is deduced. As an example, the robust procedure is applied to cloud data obtained from a satellite.

Key words and phrases: Asymptotic normality, Daubechies wavelet filter, Hermite expansion, M-estimation, Multi-scale process, Multi-taper methods, Cloud data, Wavelet cross-covariance.

1. Introduction

Wavelets decompose a stochastic process (e.g, a time series or a random field) with respect to a set of basis functions, each one of which is associated with a particular scale. The wavelet variance is formed by calculating the variance of the wavelet coefficients (i.e., the transformed process) at a given scale and quantifies the amount of variation present at the particular scale. Many physical processes are multi-scale in nature, where the overall observed process is an ensemble of various sub-processes, each operating at a characteristic scale, and each having its own time-dependent or time-independent variance. The wavelets variance provides a way to analyze the variability of such processes by permitting a scale-based analysis of variance; see Percival and Walden (2000) and the references therein.

The wavelet variance has been applied to a variety of time series and random fields and is particularly useful as an exploratory tool to study power law processes (Stoev and Taqqu, 2003), to detect inhomogeneity (Whitcher et al., 2002), to estimate spectral densities indirectly (Tsakiroglou and Walden, 2002), and to handle processes that are locally stationary with time- and space-varying spectra (Nason et al., 2000). Applications include the analysis of textures (Unser, 1995), electroencephalographic sleep state patterns of infants (Chiann and Morettin, 1998), the El Niño–Southern Oscillation (Torrence and Compo, 1998), soil variations (Lark and Webster, 2001), DNA analysis (Vanucci and Liò, 2001), solar coronal activity (Rybák and Dorotovič, 2002), the relationship between rainfall and runoff (Labat et al., 2001), ocean surface waves (Massel, 2001), surface albedo and temperature in desert grassland (Pelgrum et al., 2000), heart rate variability (Pichot et al., 1999) and the stability of the time kept by atomic clocks (Greenhall et al., 1999).

In practice, data collected in the form of a time series or a random field often suffer from contamination. For example, the satellite cloud data we consider in section 7 can be contaminated in at least three ways: at the satellite itself, during transmission of the signal through the atmosphere, and due to the presence of aberrant cloud types in a region where a single cloud type is dominant. The overall effect of outliers on conventional statistical inference for the wavelet variance can be quite large even if they are sparse, since their magnitude may differ very much from that of an uncontaminated data. In classical statistics, presence of outliers are handled in various ways. One approach is to remove questionable data points and then apply conventional methods; however, locating contamination in a time series or in a random field is complicated because of the dependence in the data (Fox, 1972). Discovering outliers requires precise distributional characterization of the underlying process, which is quite far from the goal of nonparametric statistics, namely, to devise statistical procedures that perform well for a broad class of underlying distributions. A second approach that obviates the need to characterize distributions is to use the median of squared wavelet coefficients rather than their mean to estimate the wavelet variance. Stoev et al. (2006) touch on the usefulness of median-type estimators of wavelet spectra to guard against contamination. The disadvantage of median-type esti-

mators is their lack of efficiency. Here we develop full M-estimation procedure for the wavelet variance and derive its large sample theory when the underlying process is Gaussian. We then discuss construction of approximate robust confidence intervals for the wavelet variance and study the efficiency of robust versus conventional estimators. We also present robust theory for non-Gaussian data. Finally we apply the robust procedure to satellite cloud data.

2.1. Daubechies Wavelet Filter

Let $\{h_{1,0}, \ldots, h_{1,L-1}\}$ be a unit level Daubechies wavelet filter (Daubechies, 1992, Section 6.2) of width $L = L_1$, which by definition satisfies three conditions:

$$\sum h_{1,l}^2 = \tfrac{1}{2}; \quad \sum h_{1,l} h_{1,l+2n} = 0,$$

for all nonzero integers n, where $h_{1,l} = 0$ for $l < 0$ and $l \geq L$; and $\sum i^l h_{1,l} = 0$ for $i = 1, \ldots, L/2$. The unit level scaling filter $\{g_{1,l} : l = 0, 1, \ldots, L_1 - 1\}$ is defined via the 'quadrature mirror' relationship $g_{1,l} = (-1)^{l+1} h_{1,L-1-l}$. Let $H_1(f)$ and $G_1(f)$ denote the transfer functions (Fourier transforms) of the filters $\{h_{1,l}\}$ and $\{g_{1,l}\}$. The squared gain function for $\{h_{1,l}\}$ is defined by

$$\mathcal{H}_1(f) = |H_1(f)|^2 = \sin^L(\pi f) \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \cos^{2l}(\pi f). \qquad (2.1)$$

The jth level wavelet filter $\{h_{j,l}\}$ is defined as the inverse discrete Fourier transform (DFT) of

$$H_j(f) = H_1(2^{j-1}f) \prod_{l=0}^{j-2} G_1(2^l f). \qquad (2.2)$$

The width of this filter is given by $L_j = (2^j - 1)(L - 1) + 1$. Similarly the jth level scaling filter $\{g_{j,l}\}$ is defined by the inverse DFT of

$$G_j(f) = \prod_{l=0}^{j-1} G_1(2^l f).$$

We denote the squared gain functions for H_j and G_j by $\mathcal{H}_j$ and $\mathcal{G}_j$.

2.2. Wavelet Variance for Time Series

Let $\{X_t, t \in \mathbb{Z}\}$ be an intrinsically stationary process of order d, where d is a nonnegative integer; i.e., its dth order increments $(1 - \mathcal{B})^d X_t$ are stationary,

where $\mathcal{B}X_t = X_{t-1}$. Let S_X denote the spectral density function (SDF) of the process. The jth level wavelet coefficient process is then given by

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l},$$

which corresponds to the changes on scale $\tau_j = 2^{j-1}$. The jth level wavelet variance is defined as the variance of $W_{j,t}$. Under the assumption $L \geq 2d$, $\{W_{j,t}\}$ is a stationary process with SDF $\mathcal{H}_j(f)S_X(f)$, and the wavelet variance is

$$\nu_X^2(\tau_j) = \int_{-1/2}^{1/2} \mathcal{H}_j(f)S_X(f)\,df \tag{2.3}$$

and is well-defined by virtue of (2.1) and (2.2). When $L = 2$, we have the Haar wavelet variance, for which the wavelet coefficients at level j are proportional to the difference of simple averages of 2^{j-1} consecutive observations. For $L > 2$, the wavelet coefficients are contrasts between localized weighted averages. When X_t is a stationary process with SDF S_X, Percival (1995) obtained the wavelet variance decomposition

$$\mathrm{var}\,(X_t) = \sum_{j=1}^{\infty} \nu_X^2(\tau_j) \tag{2.4}$$

as an alternative to the classical decomposition

$$\mathrm{var}\,(X_t) = \int_{-1/2}^{1/2} S_X(f)\,df.$$

The decomposition of $\mathrm{var}\,(X_t)$ offered by the wavelet variance complements that of the SDF by focusing directly on scale-based variations, which are often more interpretable and of more interest in the geosciences than frequency-based variations.

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

$$\hat{\nu}_X^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} W_{j,t}^2, \tag{2.5}$$

where $M_j = N - L_j + 1 > 0$. See Percival (1995) and Percival and Walden (2000) for large sample properties of this estimator and construction of confidence intervals.

2.3 Wavelet Variance for Random Fields

Let $X_{u,v}$, $u, v = 0, \pm 1, \pm 2, \ldots$ be a stationary Gaussian random field on the two-dimensional integer lattice $\mathcal{Z}^2$ with SDF $S_X(f, f')$. Its wavelet transform is defined by filtering the random field using the four possible combinations of wavelet and scaling filters along its rows and columns, yielding wavelet-wavelet, scaling-wavelet, wavelet-scaling and scaling-scaling coefficients:

$$W_{j,j',u,v} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} h_{j',l'} X_{u-l,v-l'},$$

$$U_{j,j',u,v} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} g_{j,l} h_{j',l'} X_{u-l,v-l'}, \quad V_{j,j',u,v} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} h_{j,l} g_{j',l'} X_{u-l,v-l'}$$

and

$$Z_{j,j',u,v} = \sum_{l=0}^{L_j-1} \sum_{l'=0}^{L_j-1} g_{j,l} g_{j',l'} X_{u-l,v-l'}.$$

The wavelet-wavelet variance is the variance of the wavelet-wavelet coefficients:

$$\nu_{X,h,h}^2(\tau_j, \tau_{j'}) = \text{var}\,(W_{j,j',u,v})$$

If $X_{u,v}$ is intrinsically stationary of order d, then $S_X(f, f')$ has a pole of order d at the origin and $\nu_{X,h,h}^2(\tau_j, \tau_{j'})$ is well defined if $L \geq d$. The wavelet-wavelet variance decomposes the process variance since

$$\text{var}\,(X_{u,v}) = \sum_{j=1}^{\infty} \sum_{j'=1}^{\infty} \nu_{X,h,h}^2(\tau_j, \tau_{j'}). \tag{2.6}$$

This generalizes the result for time series stated in (2.4) and provides a scale-based analysis of variance for random fields. When we have a realization of an intrinsically stationary random field $X_{u,v}$ on a finite array $\{(u, v) : u = 0, \ldots, N-1 \,\&\, v = 0, \ldots, M-1\}$, we can then estimate the wavelet-wavelet variance by the unbiased estimator

$$\hat{\nu}_{X,h,h}^2(\tau_j, \tau_{j'}) = \frac{1}{N_j M_{j'}} \sum_{u=L_j-1}^{N-1} \sum_{v=L_{j'}-1}^{M-1} W_{j,j',u,v}^2, \tag{2.7}$$

where $N_j = N - L_j + 1$ and $M_{j'} = M - L_{j'} + 1$. See Mondal (2007) for statistical inference based on this type of estimator.

We can also define wavelet-scaling and scaling-wavelet variances via the variances of the wavelet-scaling and scaling-wavelet coefficients:

$$\nu_{X,g,h}^2(\tau_j, \tau_{j'}) = \operatorname{var}(U_{j,j',u,v}), \quad \nu_{X,h,g}^2(\tau_j, \tau_{j'}) = \operatorname{var}(V_{j,j',u,v})$$

Although the wavelet-scale and scaling-wavelet variances do not appear in (2.6), they are closely connected with wavelet-wavelet variances and lead to an interesting alternative variance decomposition. In particular, we have

$$\nu_{X,g,h}^2(\tau_j, \tau_{j'}) = \sum_{k=j}^{\infty} \nu_{X,h,h}^2(\tau_k, \tau_{j'}), \quad \nu_{X,h,g}^2(\tau_j, \tau_{j'}) = \sum_{k=j'}^{\infty} \nu_{X,h,h}^2(\tau_j, \tau_k),$$

allowing us to rewrite equation (2.6) as

$$\operatorname{var}(X_{u,v}) = \sum_{j=1}^{\infty} \nu_{X,h,h}^2(\tau_j, \tau_j) + \sum_{j=1}^{\infty} \nu_{X,g,h}^2(\tau_{j+1}, \tau_j) + \sum_{j=1}^{\infty} \nu_{X,h,g}^2(\tau_j, \tau_{j+1}). \quad (2.8)$$

The above is a tri-diagonal analysis of variance. The main diagonal involves wavelet-wavelet variances, whereas the first diagonals below and above the main diagonal involve, respectively, the wavelet-scaling and scaling-wavelet variances.

3. M-estimation of wavelet variance

Let $\{Y_i\}$, $i \in \mathcal{L}$, be a zero-mean Gaussian process, and suppose we are interested in estimating $\operatorname{var}(Y_i) = \operatorname{E} Y_i^2 = \nu^2$. Here the index i represents either time t or spatial location (u, v), and hence $\mathcal{L}$ is an integer lattice, either $\mathbb{Z}$ or $\mathbb{Z}^2$. Typically $\{Y_i\}$ is either the wavelet coefficient process $\{W_{j,t}\}$ or the wavelet-wavelet coefficients $\{W_{j,j',u,v}\}$. In practice the simple average-type estimators (2.5) and (2.7) are vulnerable to data contamination, so we are interested in developing robust estimators that can guard against such contamination, yet still work well when Gaussianity holds. Under our assumptions, ν is a scale parameter. A logarithmic transformation converts ν to a location parameter and allows use of M-estimation theory to formulate a robust estimator. Accordingly, let

$$Q_i = \log Y_i^2.$$

Then $\{Q_i\}$ is a stationary process and, using Bartlett and Kendall (1946), we obtain

$$\operatorname{E} Q_i = \log \nu^2 + \psi(\tfrac{1}{2}) + \log 2, \quad \operatorname{var} Q_i = \psi'(\tfrac{1}{2}) = \tfrac{1}{2}\pi^2,$$

where ψ and ψ' are the di- and tri-gamma functions. Let $\mu = \log \nu^2 + \psi(\frac{1}{2}) + \log 2$. Then Q_i can be written as

$$Q_i = \mu + \epsilon_i,$$

where $\mathrm{E}\,\epsilon_i = 0$ and $\operatorname{var} \epsilon_i = \frac{1}{2}\pi^2$.

Assumption 1 *Let* $\varphi(x)$, $x \in \mathbb{R}$, *be a nondecreasing real-valued function of bounded variation with* $\varphi(-\infty) < 0$ *and* $\varphi(\infty) > 0$ *such that*

$$\lambda(x) = \mathrm{E}\,\varphi(Q_i - x)$$

is well defined, strictly decreasing on $\mathbb{R}$ *and has a solution point* μ_0 *such that*

$$\lambda(\mu_0) = 0.$$

Moreover we assume φ *is such that* $\lambda(x)$ *is infinitely differentiable as well.*

The relationship between the solution point μ_0 and the location parameter μ is discussed in Section 4.1.

Because of the Gaussian assumption on $\{Y_i\}$, the marginal distribution F_Q of Q_i is that of the logarithm of a squared Gaussian random variable and hence is infinitely differentiable. Integration by parts allows us to write

$$\lambda(x) = -\int_{-\infty}^{\infty} F_Q(x + y)\,d\varphi(y),$$

and the kth derivative of λ satisfies the relation

$$\lambda^{(k)}(x) = -\int F_Q^{(k)}(y + x)\,d\varphi(y).$$

Given the form of F_Q and the fact that φ is of bounded variation, it follows that $\lambda^{(k)}$ is bounded for each k.

The corresponding M-estimator T_N of the solution point μ_0 based on observations $\{Q_i, i \in \mathcal{I}\}$ is defined by

$$T_N = \operatorname{argmin}\left\{ \left| \sum_{i \in \mathcal{I}} \varphi(Q_i - x) \right| : x \in \mathbb{R} \right\}.$$

The index set $\mathcal{I}$ is equal to $\{0, \ldots, N - 1\}$ for time series and is $\{(u, v) : u, v = 0, \ldots, N - 1\}$ for a random field (thus, when Y_i represents $W_{j,t}$, N stands for M_j). In what follows, let B be the size of $\mathcal{I}$.

Assumption 1 holds for various choices of φ, including

$$\varphi_I(x) = \text{ sign } (x), \ \varphi_{II}(x) = 2\Pr(\epsilon_i \leq x) - 1, \ \varphi_{III}(x) = p \text{ sign } (x)1_{|x|>p} + x1_{|x|\leq p}$$

for $p > 0$ and

$$\varphi_{IV}(x) = \begin{cases} a' & \text{if } x \leq a, \\ e^x - 1 & \text{if } a < x \leq b \text{ and} \\ b' & \text{if } x > b \end{cases}$$

(see equation (3.9) of Thall, 1979, for another choice). To better understand M-estimation, consider an independent and identically distributed (i.i.d.) sample. Then T_N corresponding to φ_I is the same as the maximum likelihood estimator (MLE) for the location parameter when the observations arise from a double exponential distribution. The function φ_{II} corresponds to an MLE under logistic errors, whereas φ_{III} corresponds to an MLE under a distribution whose central part behaves like a Gaussian but whose tail is like a double exponential. Similarly, for φ_{IV}, the estimator is an MLE under a distribution whose central part behaves like a log of a chi-square distribution. The choice φ_I gives rise to median-type estimators, which, when compared to mean-type estimators, are less sensitive to data contamination. The choice φ_{III} yields Huber's φ function for a location parameter, which maps extreme values of $\log Y_i^2$ to either $\pm p$. Similarly φ_{IV} is a Huberized mean of Y_i^2, which replaces extreme values of Y_i^2 with either a' or b'. The median-type estimator φ_I is invariant to monotone transformation of the data and can be regarded as limiting cases of the Huber-type estimators φ_{III} and φ_{IV}.

M-estimation under a non-i.i.d. set up has been considered by a large number of authors; see, for example, Beran (1991) and Koul and Surgailis (1997). Here we follow the work of Koul and Surgailis (1997), which allows a very general class of weight functions φ. The following central limit theorem provides the basis for inference on the solution point μ_0 using the estimator T_N.

Theorem 1 *Assume φ and λ satisfy Assumption 1. Then $B^{\frac{1}{2}}(T_N - \mu_0)$ is asymptotically normal with mean zero and variance given by $A_\varphi / \{\lambda'(\mu_0)\}^2$, where*

$$A_\varphi = \sum_{i \in \mathcal{L}} \text{cov}\{\varphi(Q_i - \mu_0), \varphi(Q_0 - \mu_0)\}.$$

When φ is smooth (twice continuously differentiable), for example, $\varphi = \varphi_{II}$, then $\sum_{i \in \mathcal{I}} \varphi(Q_i - T_N) = 0$. We can then use a Taylor series expansion to deduce that

$$B^{\frac{1}{2}}(T_N - \mu_0) = \frac{B^{-\frac{1}{2}} \sum \varphi(Q_i - \mu_0)}{B^{-1} \sum \varphi'(Q_i - \mu_0) + (T_N - \mu_0) B^{-1} \sum \varphi''(Q_i - T^*)},$$

where T^* takes values between μ_0 and T_N. Consequently the central limit theorem follows from that of $B^{-\frac{1}{2}} \sum \varphi(Q_i - \mu_0)$ and by proving the consistency of T_N. However, when φ is no longer smooth, e.g., $\varphi = \varphi_I$, we can not make use of a Taylor series expansion, and the proof of Theorem 1 differs substantially. We give a general proof of Theorem 1 in the Appendix.

4.1. Correction for Bias

The statistics T_N is consistent for the solution point μ_0, which is not necessarily the same as the location parameter μ. We can obtain a robust estimator $\hat{\mu}$ of μ by adding bias $= \mu - \mu_0$ to the estimator T_N, yielding

$$\hat{\mu} = T_N + \text{bias}.$$

We thus need to compute the bias $\mu - \mu_0$, which depends on the choice of the weight function φ and on the distribution function F_Q. We can compute this bias analytically in some cases. To do so, we first compute the function $\lambda(x)$. Let Z denote the standard Gaussian random variable. Let ϕ and Φ be its density and distribution function. Then

$$\lambda(x) = \mathrm{E}\, \varphi(Q_i - x) = \mathrm{E}\, \varphi(\log Z^2 + \log \nu^2 - x),$$

Thus the choice of $\varphi = \varphi_I$ yields

$$\lambda_I(x) = 3 - 4\Phi\left(\exp(\tfrac{1}{2}x - \log \nu)\right). \tag{4.1}$$

Therefore, $\lambda_I(\mu_0) = 0$ implies

$$\mu_0 = 2\log \nu + 2\log\left[\Phi^{-1}(\tfrac{3}{4})\right]. \tag{4.2}$$

and hence

$$\text{bias}_I = \mu - \mu_0 = \psi(\tfrac{1}{2}) + \log 2 - 2\log\left[\Phi^{-1}(\tfrac{3}{4})\right].$$

Next consider $\varphi = \varphi_{II}$. First we simplify φ_{II} as:

$$\varphi_{II}(x) = 4\,\Phi\left(\exp(\tfrac{1}{2}x + \tfrac{1}{2}\psi(\tfrac{1}{2}) + \tfrac{1}{2}\log 2)\right) - 3.$$

Therefore, we obtain

$$\lambda_{II}(x) = 4\,\mathrm{C}_a\Big(\exp(\log\nu - \tfrac{1}{2}x - \tfrac{1}{2}\psi(\tfrac{1}{2}) - \tfrac{1}{2}\log 2)\Big) - 3, \qquad (4.3)$$

where C_a is the distribution function of a standard Cauchy random variable. Now $\lambda_{II}(\mu_0) = 0$ implies

$$\mu_0 = 2\log\nu - \psi(\tfrac{1}{2}) - \log 2 + 2\log\big[\mathrm{C}_a^{-1}(\tfrac{3}{4})\big]. \qquad (4.4)$$

and hence

$$\mathrm{bias}_{II} = \mu - \mu_0 = -2\log\big[\mathrm{C}_a^{-1}(\tfrac{3}{4})\big].$$

For φ_{III} and φ_{IV}, there are no easy closed forms for $\lambda(x)$; however, we can numerically evaluate the bias correction in these cases.

4.2. Construction of Confidence Intervals

Given a consistent estimator of A_φ and knowledge of $\lambda'(\mu_0)$, we can use Theorem 1 to construct an asymptotically correct confidence interval for μ_0 and hence for μ and ν^2. Since A_φ is equal to the value of the SDF of the stationary process $\{\varphi(Q_i - \mu_0)\}$ at zero frequency, we use a multi-taper spectral approach (Serroukh et al., 2000) to estimate A_φ. Let $\{\gamma_{c,t}, t = 0, \ldots, N - 1\}$ for $c = 0, \ldots, C - 1$ be the first C orthogonal Slepian tapers of length N, where C is an odd integer. When dealing with a time series, let $\mathcal{K}$ be the index set $\{0, \ldots, C-1\}$; otherwise, for a random field, let it be $\{(c, c') : c, c' = 0, \ldots, C - 1\}$. For $k \in \mathcal{K}$, we define

$$J_k = \sum_i \beta_{k,i}\varphi(Q_i - T_N), \qquad (4.5)$$

where $\beta_{k,i} = \gamma_{c,t}$ if we have a time series with $i = t$ or else $\beta_{k,i} = \gamma_{c,u}\gamma_{c',v}$ for a random field with $i = (u, v)$. Define $\beta_{k,.} = \sum_i \beta_{k,i}$ and

$$\tilde{\mu} = \frac{\sum_k J_k\beta_{k,.}}{\sum_k \beta_{k,.}^2}.$$

We then estimate A_φ by

$$\hat{A}_\varphi = \frac{1}{K}\sum_k (J_k - \tilde{\mu}\beta_{k,.})^2,$$

where K is the size of the index set $\mathcal{K}$. Since μ_0 is unknown, we use the consistent estimator T_N in its stead in equation (4.5). Thus the resulting multi-taper

estimate $\hat{A}_\varphi$ is consistent for A_φ if the SDF of the process $\{\varphi(Q_i - x)\}$ at zero frequency is continuous at $x = \mu_0$. The latter holds for a wide range of Gaussian processes $\{Y_i\}$ and for many choices of φ. In particular, a sufficient condition is that the process $\{Y_i\}$ is ergodic, for which T_N is also strongly consistent. Following the recommendation of Serroukh et al. (2000), we choose $C = 5$ so that the Slepian tapers are band-limited to the intervals $[-7/N, 7/N]$.

Next we discuss the computation of $\lambda'(\mu_0)$. For $\varphi = \varphi_I$, we use equation (4.1) and equation (4.2) to obtain

$$\lambda'_I(\mu_0) = -2\phi\Big(\exp(\tfrac{1}{2}\mu_0 - \log \nu)\Big)\exp(\tfrac{1}{2}\mu_0 - \log \nu) = -2\phi\Big(\Phi^{-1}(\tfrac{3}{4})\Big)\Phi^{-1}(\tfrac{3}{4}). \quad (4.6)$$

Similarly for $\varphi = \varphi_{II}$, use of equation (4.3) gives

$$\lambda'_{II}(\mu_0) = -2\,c_a\Big(\exp(\log \nu - \tfrac{1}{2}\mu_0 - \tfrac{1}{2}\psi(\tfrac{1}{2}) - \tfrac{1}{2}\log 2)\Big)\exp(\log \nu - \tfrac{1}{2}\mu_0 - \tfrac{1}{2}\psi(\tfrac{1}{2}) - \tfrac{1}{2}\log 2),$$

where c_a is the density function of C_a. Now by using equation (4.4), we simplify the above as:

$$\lambda'_{II}(\mu_0) = -2\,c_a\Big(C_a^{-1}(\tfrac{3}{4})\Big)C_a^{-1}(\tfrac{3}{4}). \quad (4.7)$$

This has the same form as that of equation (4.6) with the Gaussian density function being replaced by the Cauchy one.

However for other choices of φ, namely, $\varphi = \varphi_{III}$ and φ_{IV}, there is no convenient analytic form (although one may surmise that it will have a form similar to equation 4.7), and we rely on numerical computation to evaluate $\lambda'(\mu_0)$.

5. Efficiency Study

Robust estimators guard against data contamination but are less efficient than estimators designed to be efficient when underlying assumptions are correct. Thus if a time series or random field is truly Gaussian, the median-type estimator will perform poorly compared to the mean-type estimator. It is therefore of interest to study the asymptotic relative efficiency (ARE) of the robust estimators for a range of Gaussian processes. Application of the delta method to the result of Theorem 1 yields

$$B^{\frac{1}{2}}(\exp T_N - \exp \mu_0) \to_d N\Big(0, \exp(2\mu_0)\frac{A_\varphi}{\lambda'(\mu_0)^2}\Big).$$

Hence $\exp\{T_N + \text{bias} - \psi(\tfrac{1}{2}) - \log 2\}$ is an asymptotically unbiased robust esti-

mator of ν^2. When compared to the mean-type estimator, we have

$$ARE = \frac{\mathrm{var}\,\hat{\nu}^2}{\mathrm{var}\left\{\exp T_N + \mathrm{bias} - \psi(\frac{1}{2}) - \log 2\right\}} = \frac{A\{\lambda'(\mu_0)\}^2}{(\nu^2)^2 A_\varphi}, \qquad (5.1)$$

where $\hat{\nu}^2 = B^{-1}\sum_i Y_i^2$ is the usual mean-type estimator and

$$A = 2 \sum_{i=-\infty}^{\infty} \mathrm{cov}^2(Y_0, Y_i).$$

Using a Hermite expansion, we could write A_φ in terms of the ACVS of $\{Y_i\}$, but this expansion is not useful in practice for computing exact AREs. We therefore resort to some simulation studies.

In the first example, we use the Haar wavelet, and we simulate 10000 AR(1) time series of length 1024 for various values of the correlation parameter. For each series we take $Y_t = W_{2,t}$, the level $j = 2$ wavelet coefficients, and use $\varphi = \varphi_I$. We compute exact values of ν^2, μ_0 and $\lambda'(\mu_0)$. We note that A/A_φ can be approximated by $\mathrm{var}\{\sum Y_t^2\}/\mathrm{var}\{\sum \varphi_I(\log Y_t^2 - \mu_0)\}$. Hence for each time series we compute $\sum Y_t$ and $\sum \varphi_I(\log Y_t^2 - \mu_0)$ and then compute their corresponding sample variances using all 10000 replications. Finally we estimate the ARE using (5.1). Figure 5.1 plots the ARE against various values of the correlation parameter. The ARE is about 50% for the AR(1) time series. This efficiency is small when the correlation parameter ϕ is close to -1, attains its peak at about $\phi = .75$, but then starts to decrease again. As an alternative procedure, we also compute a robust estimator of ν^2 based on T_N and the usual estimator $\hat{\nu}^2$ for each simulated series and obtain an estimate of the ARE by computing the ratio of estimates of the Monte Carlo variances. This alternative experiment leads to the same conclusions.

For the second example, we consider X_t to be a stationary fractionally differenced (FD) process with

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

for $k = 1, 2, \ldots$; see, e.g., Granger and Joyeux (1980), Hosking (1981). Here $\delta < \frac{1}{2}$ is the long memory parameter, with $\delta = 0$ corresponding to white noise and δ close to $\frac{1}{2}$ corresponding to a highly correlated process whose ACVS damps down to zero very slowly. We simulate 10000 FD time series of length 1024. We
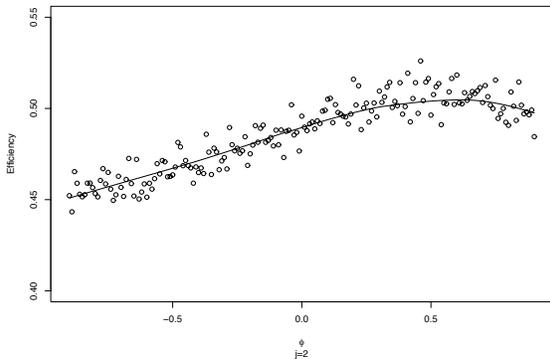
Figure 5.1: Approximate asymptotic relative efficiency of the median-type estimator with respect to the mean-type estimator for AR(1) processes
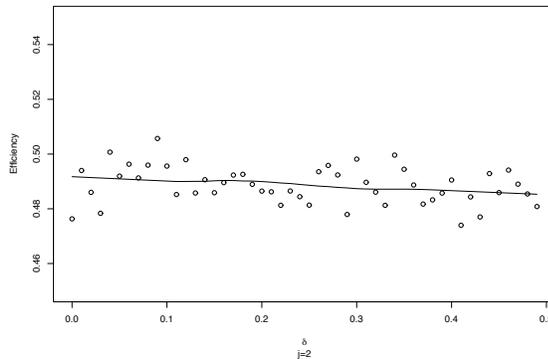
.

Figure 5.2: As in Figure 5.1, but now for FD processes

.

choose the long memory parameter δ in the range 0 to $\frac{1}{2}$ and compute the ARE via Monte Carlo estimates. Here we again use the Haar wavelet and restrict ourselves to the median-type estimator. Figure 5.2 plots the ARE against the long memory parameter δ. We see that the ARE is about 50% for all δ.

In the final example we repeat the same experiment for fractional Brownian surfaces (FBSs). Here $X_{u,v}$ satisfies

$$\tfrac{1}{2}\mathrm{var}\left(X_{u,v} - X_{0,0}\right) = 2(u^2 + v^2)^\theta,$$

where $\theta \in (0,1)$; see e.g., Zhu and Stein (2002). The parameter θ is known as a smoothness or persistence parameter and controls the roughness of the surface; i.e., θ close to unity (zero) corresponds to a smooth (very rough) surface. We simulate 10000 FBSs on 32×32 array using the R package RandomFields (Schlather, 2006) and consider $Y_{u,v} = W_{2,2,u,v}$, i.e., the $(2,2)$ level wavelet-wavelet coefficient process. We compute the ARE using the procedure described above. We observed about 60–65% efficiency for selected $\theta \in [.01, 1]$.

6. Non-Gaussian processes

The application of the wavelet variance to non-Gaussian processes has been considered by, among others, Serroukh et al. (2000) and Vannucci and Liò (2001). Serroukh et al. (2000) obtain asymptotic normality of the conventional estimator (2.5) when the wavelet coefficient process satisfies appropriate mixing conditions, a result that also easily extends to the estimator (2.7); see Mondal (2007). Our interest here is to develop a corresponding robust theory. First we note that, when Y_i is non-Gaussian, $\mathrm{E}\, Y_i^2 = \nu^2$ need not be a scale parameter, so we consider an approach that does not rely upon a logarithmic transformation. Following Chave et al. (1987), we consider a block averaging scheme, where we divide the data into several blocks and consider a conventional estimator of ν^2 within each block. Thus let $\mathcal{V}_0 = \{0, \ldots, N' - 1\}$ and $\mathcal{U} = \{0, \ldots, N - N'\}$ for a time series and $\mathcal{V}_0 = \{(u, v) : u, v = 0, \ldots, N' - 1\}$ and $\mathcal{U} = \{(u, v) : u, v = 0, \ldots, N - N'\}$ for a random field. For $k \in \mathcal{U}$, we consider

$$R_k = \frac{1}{V} \sum_{i \in \mathcal{V}_k} Y_i^2,$$

where V is the size of $\mathcal{V}_0$, and $\mathcal{V}_k = \mathcal{V}_0 + k$ is defined via translation. We then obtain a robust estimator of ν^2 by robustly combining the conventional estimators R_k across $k \in \mathcal{U}$. Under an appropriate mixing condition (see Appendix A.2), R_k is approximately Gaussian with mean ν^2, allowing us to treat ν^2 as a location parameter. Our M-estimator is thus

$$\tilde{T}_N = \ \operatorname{argmin}\ \left\{ \left| \sum_k \tilde{\varphi}(R_k - x) \right| : x \in \mathbb{R} \right\}.$$

Let $\tilde{\lambda}(x) = \mathrm{E}\, \tilde{\varphi}(R_k - x)$ satisfy the condition

$$\tilde{\lambda}(\nu^2) = \tilde{\lambda}'(\nu^2)(x - \nu^2) + o(|x - \nu^2|).$$

Then we have the following theorem.

Theorem 2 *Suppose that $\{R_k\}$ is strongly mixing, i.e., $\alpha_R(m) \to 0$ as $m \to \infty$, where $\alpha_R(m)$ is the mixing condition defined by Equation (A.6) in Appendix A.2. Suppose also that*

$$\sum_{m=1}^{\infty} m^{b-2} \alpha_R(m)^{\frac{q}{b+q}} < \infty \tag{6.1}$$

for some $b \geq 2$ and some $q > 0$. Then $U^{\frac{1}{2}}(\tilde{T}_N - \nu^2)$ is asymptotically normal with mean zero and variance given by $\tilde{A}_\varphi / \{\tilde{\lambda}'(\nu^2)\}^2$, where U be the size of $\mathcal{U}$, and

$$\tilde{A}_{\tilde{\varphi}} = \sum_{k \in \mathcal{L}} \operatorname{cov}\{\tilde{\varphi}(R_k - \nu^2), \tilde{\varphi}(R_0 - \nu^2)\}.$$

We defer the proof to Appendix A.2. Confidence intervals are constructed in a manner similar to what is described in Section 4.2.

7. Applications to Cloud Data

Figure 7.1 shows the pre-processed brightness temperature image of a cloud field over the southeast Pacific Ocean obtained on 17 October 2001 as part of the East Pacific Investigation of Climate (EPIC) field experiment. Strato-cumulus cloud fields in that part of the world tend to be homogenous except for pockets of seemingly cloud-free air. These pockets of open cells (POCs) are distinct from broken clouds, are coupled to the development of marine rainfall and are characterized by low-aerosol air mass. We refer to Bretherton et al. (2004) and Stevens et al. (2005) for details. The four squares in Figure 7.1 indicate regions with four different types of cloud. The first region (the top-most square) is a POC region; the second (right-most) consists of uniform stratus clouds and thus has different characteristics than the POC region; the third (left-most) has broken clouds; and the last (bottom-most) has clouds that are forming a POC.

Satellite cloud data are often marked by contamination in the form of sudden bursts of intensities or speckles and other outliers. In addition, while each of the four regions of focus in Figure 7.1 has a dominant cloud type of interest, each also contains a small fraction of clouds with different characteristics. It is thus natural to resort to median-type estimators that are robust and effective in extracting the characteristics of the dominant clouds.

Figure 7.2 shows conventional mean-type (thin circles) and robust median-type (thick) wavelet variance estimates and their confidence intervals (lines intersecting circles vertically) for four diagonal scales indexed by $j = 1, 2, 3$ and 4. The top row shows, from left to right, the wavelet-wavelet, wavelet-scaling and scaling-wavelet variances and their 95% confidence intervals for the POC region. We see that at the smaller scales ($\tau_1 - \tau_3$) the median-type estimates take lower values than that the mean-type estimates, indicating that there are outliers in
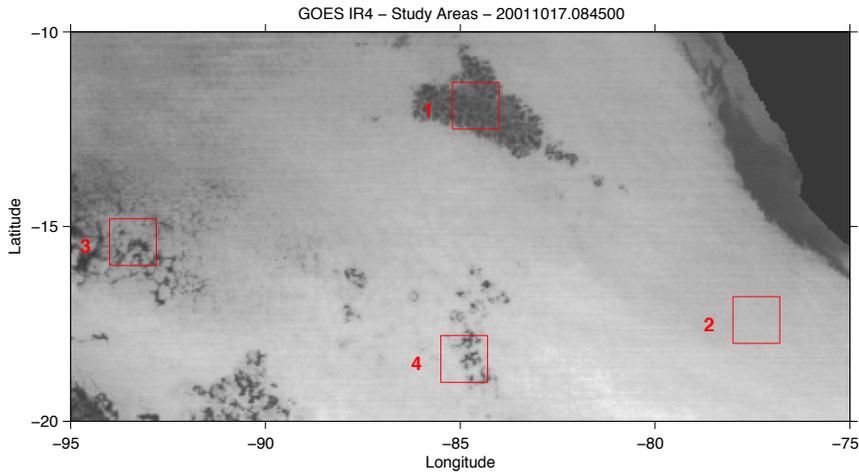
Figure 7.1: Image plots of clouds at four different regions

.

the data. Moreover the median-type estimates produce somewhat smaller confidence intervals. The next row has similar plots for the uniform stratus clouds. We again note that the robust estimates take slightly smaller values. Unlike the POC region, the median-type estimator here has a bigger confidence interval at scale τ_1, indicating that the percentage of noise may be greater in the uniform clouds. The third and bottom rows show the results for the regions with broken clouds and forming POC. The robust estimates for these two regions differ substantially different from the conventional estimates. At scales τ_1 and τ_2, the robust estimates of wavelet-wavelet variances for the broken clouds are much lower than the conventional estimates and have much larger confidence intervals. Broken clouds are mixtures of various clouds and the median-type estimates attempt to pick up the characteristics from the dominant cloud type; however, the bigger fraction of other cloud types produces bigger confidence intervals. The robust estimates of wavelet-wavelet variance for the forming POC region are again much smaller than for the conventional estimates, mainly because of the presence of other clouds. Moreover, unlike the broken clouds region, the robust estimators have smaller confidence intervals in this region, indicating that the fraction of other clouds present here is smaller than the fraction of other clouds present in the broken clouds. We also see that conventional and robust estimates

are markedly different in the broken clouds and forming POC regions. Scales τ_3 and τ_4, rather than scales τ_1 and τ_2, are more useful for differentiating a POC region from other cloud types. Large uncertainty in the estimates of the wavelet variances mark the broken clouds, while big differences between mean-type and median-type estimates at scales τ_1 and τ_2 are indicative of a forming POC region.

8. Discussion

In practical applications, the Huber function $\phi = \phi_{IV}$ is of interest because $\exp(T_N)$ then gives rise to the median (robust) estimator when $-a = b = 0$ with $-a' = b' = 1$ and to the mean-type (conventional) estimator when $-a = b = \infty$. With a' and b' set appropriately, other values of $-a = b = h$ can the thought of as a compromise between the robust and the conventional estimation procedures. It is thus of interest to set h so that the Huberized estimator has a certain asymptotic relative efficiency as given in (5.1). Following Koul and Surgailis (1997), we could obtain an expression for the ARE via the Hermite expansion and then try to find an h that achieves the required efficiency. This would, however, require knowledge of the ACVS of the underlying process. A better strategy is to estimate the ARE in (5.1) by a non-parametric (multi-taper) estimator for any given h and then solve an optimization problem on a finite grid for a range of values of h.

In the presence of outlying observations, conventional estimators of the SDF of a time series or a random field are problematic, so there is a need for robust estimators. The wavelet variance provides a simple and useful estimator of the integral of the SDF over octave bands. In particular, the Blackman and Tukey (1958, Sec. 18) pilot spectra coincide with Haar wavelet variances. Recently Tsakiroglou and Walden (2002) extended the pilot spectra of Blackman and Tukey by utilizing the (maximum overlap) discrete wavelet packet transform. The result is an SDF estimator that is competitive with existing estimators. In the same vein, our proposed methodology can be extended to handle wavelet packet transforms, thus providing a robust SDF estimator in the style of Tsakiroglou and Walden (2002).
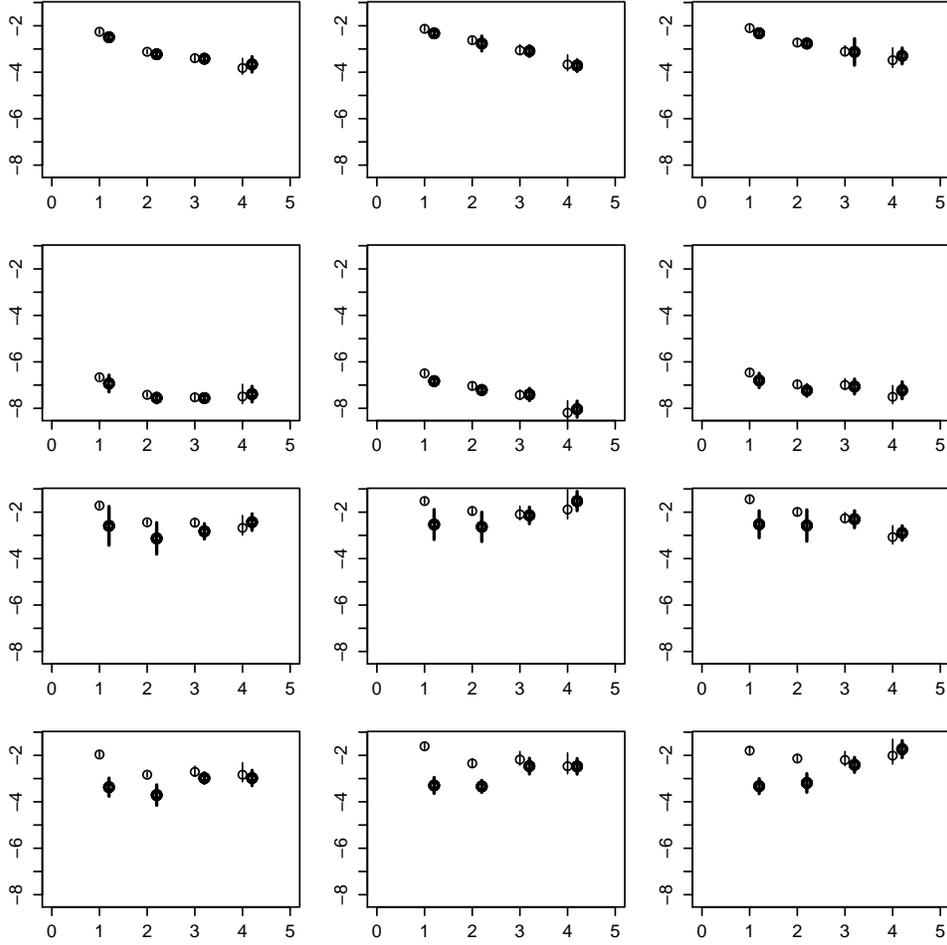
Acknowledgments

Figure 7.2: Log of wavelet variances at diagonal scales indexed by $j = 1, 2\ 3$ and 4 for four cloud regions. The thin and thick circles show, respectively, the mean-type and robust median-type estimates. The vertical lines bisecting the circles depict 95% confidence intervals. The left, middle and right columns are for wavelet-wavelet, wavelet-scale and scale-wavelet variances. The top row is for the POC region; the second row, for uniform clouds; the third row, for broken clouds; and the bottom row, for a forming POC region

.

Appendix: Proofs

A.1. Proof of Theorem 1

We denote by $\{P_n, \ n = 0, 1, \ldots\}$ the sequence of Hermite polynomials. Let ϕ be the density function for a standard Gaussian random variable Z. For a function P with $\int_{-\infty}^{\infty} P(x)\phi(x)\,dx < \infty$, we write the Hermite expansion as

$$P(x) = \sum_{n=0}^{\infty} c_n P_n(x), \ \text{ where } \ c_n = \int_{-\infty}^{\infty} P(x)P_n(x)\phi(x)\,dx. \qquad (A.1)$$

The function $P(x)$ is said to have Hermite rank ϱ if in the expansion (A.1) we have

$$c_0 = c_1 = \cdots = c_{\varrho-1} = 0, \ \ c_\varrho \neq 0.$$

Prior to proving the theorem, we need to prove the following six lemmas.

Lemma 1 *The following functions have Hermite rank two.*

(i) $P(x) = x^2 - 1$
(ii) $P(x) = \log(x^2) - \int \log(x^2)\phi(x)\,dx$
(iii) $P(x) = 1_{(\log x^2 \leq y)} - \int 1_{(\log x^2 \leq y)}\phi(x)\,dx, \ y \in \mathbb{R}$
(iv) $P(x) = \varphi(\log x^2 - y) - \int \varphi(\log x^2 - y)\phi(x)\,dx, \ y \in \mathbb{R}.$

Proof of Lemma 1. Recalling that $P_0(x) = 1$, $P_1(x) = x$ and $P_2(x) = x^2 - 1$, note that each $P(x)$ is an even function such that $\int P(x)\phi(x)\,dx = 0$, implying that $c_0 = c_1 = 0$, whereas $\int P(x)(x^2 - 1)\phi(x)\,dx \neq 0$. ∎

Lemma 2 *If P is any of the functions in Lemma 1 and if $\{Y_i\}$ has a square integrable spectral density, then the autocovariances $s_{P,k}$ of the random process $\{P(Y_i)\}$ satisfy*

$$\sigma_P^2 = \sum_k s_{P,k} > 0. \qquad (A.2)$$

Proof of Lemma 2. Since P is even, we can use the Hermite expansion to write

$$P(Y_i) = \sum c_{2m} P_{2m}(Y_i).$$

The SDF of $\{P(Y_i)\}$ is given by

$$S_P(f) = \sum c_{2m}^2 (2m)! S_Y^{(*2m)}(f)$$

(see Hannan 1970, p. 83), where $S_Y^{(*2m)}$ is $2m$-fold convolution of S_Y. Since $S_Y^{(*2m)}$ is strictly positive at the origin and there exists one m such that $c_{2m} \neq 0$, we see that $\sigma_P^2 = S_P(0)$ is strictly positive, and hence the result follows. ∎

Lemma 3 *If P is any of the functions in Lemma 1 and if $\{Y_i\}$ has a square integrable SDF, then, as $N \to \infty$, $B^{-1} \sum_{i \in \mathcal{I}} P(Y_i)$ converges in distribution to $\sigma_P Z$, where σ_P^2 is as in equation (A.2).*

Proof of Lemma 3. This follows directly from Lemma 1, Lemma 2 and Theorem 2 of Breuer and Major (1983). ∎

Lemma 4 *Let $\Psi_N(x) = B^{-1} \sum \varphi(Q_i - x)$ and T_N be defined as earlier. Then $T_N - \mu_0 = o_P(1)$ and $\Psi_N(T_N) = O_P(B^{-1})$.*

Proof of Lemma 4. Let $F_{Q,N}$ be the empirical distribution function of $\{Q_i, i \in \mathcal{I}\}$. As $Q_i \neq Q_j$, $i \neq j$ a.s, the jumps of the empirical distribution $F_{Q,N}$ are such that $\Delta F_{Q,N}(x) = F_{Q,N}(x) - F_{Q,N}(x-) \leq B^{-1}$ a.s., and therefore $\Delta \Psi_N(x) = O(B^{-1})$ a.s.; indeed, almost surely

$$|\Delta \Psi_N(x)| \leq \int |\Delta F_{Q,N}(y + x)||d\varphi(y)| \leq |\varphi|/B,$$

where $|\varphi|$ is the variation of φ. Now since $\varphi(-\infty) < 0$ and $\varphi(\infty) > 0$, we have that $\Psi_N(-\infty) < 0$ and $\Psi_N(\infty) > 0$. Since $\Psi_N(x)$ is nondecreasing, the graph of Ψ_N crosses the x-axis in a neighborhood of 0 at some point T_N with $\Psi_N(T_N+) \leq 0$ and $\Psi_N(T_N-) \geq 0$ and hence $|\Psi_N(T_N+)| + |\Psi_N(T_N-)| = |\Psi_N(T_N+) - \Psi_N(T_N-)| \leq |\varphi|/B$. Hence, for all N, we have

$$|\Psi_N(T_N)| \leq |\varphi|/B \ \text{ a.s.} \tag{A.3}$$

We now prove consistency of T_N. Let $\epsilon > 0$. Since $\Psi_N(x)$ is nondecreasing, we note that

$$\Pr(T_N < \mu_0 + \epsilon) > \Pr\big(\Psi_N(T_N) < \Psi_N(\mu_0 + \epsilon)\big).$$

By equation (A.3), it then follows that

$$\Pr\big(\Psi_N(T_N) < \Psi_N(\mu_0 + \epsilon)\big) > \Pr\big(|\varphi|/B < \Psi_N(\mu_0 + \epsilon)\big).$$

However, by Lemma 3, $\Psi_N(\mu_0 + \epsilon)$ converges in probability to $\lambda(\mu_0 + \epsilon)$, which is strictly greater than zero. Thus $\Pr\big(|\varphi|/B < \Psi_N(\mu_0 + \epsilon)\big)$ converges to one. Hence

$$\Pr(T_N < \mu_0 + \epsilon) \to 1.$$

A similar argument implies that $\Pr(T_N > \mu_0 - \epsilon)$ converges to one, from which the consistency of T_N follows. ∎

Lemma 5 *For $y \geq 0$ we have*

$$B\mathrm{E}\left[\int \{\varphi(x - y) - \varphi(x)\} \, d\{F_{Q,N}(x) - F_Q(x)\}\right]^2 \leq \ constant \ y.$$

Proof of Lemma 5. Consider the Hermite expansion of $P(Y_i) = 1_{\log Y_i^2 \leq x} - F_Q(x)$, namely,

$$1_{\log Y_i^2 \leq x} - F_Q(x) = \sum_m a_{2m}(x) P_{2m}(Y_i).$$

Lemma 1 says that P has Hermite rank 2. For $x \leq y$ we define $F_Q(x, y) = F_Q(y) - F_Q(x)$, $a_{2m}(x, y) = a_{2m}(y) - a_{2m}(x)$ and $F_{Q,N}(x, y) = F_{Q,N}(y) - F_{Q,N}(y)$. Then we can write

$$1_{x \leq \log Y_i^2 \leq y} - F_Q(x, y) = \sum_m a_{2m}(x, y) P_{2m}(Y_i).$$

It then follows from the orthogonality of the Hermite polynomials that

$$\sum a_{2m}^2 (2m)! s_{Y,0}^{2m} = \mathrm{E}\left\{1_{x \leq \log Y_i^2 \leq y} - F(x, y)\right\}^2 \leq F(x, y). \qquad (\mathrm{A.4})$$

Now

$$\mathrm{E}\left[\sum_{i \in \mathcal{I}} \{1_{x \leq \log Y_i^2 \leq y} - F(x, y)\}\right]^2 = \mathrm{var}\left\{\sum_m a_{2m}(x, y) \sum_i P_{2m}(Y_i)\right\}^2$$

$$= \sum_m \sum_{m'} a_{2m}(x, y) a_{2m'}(x, y) \sum_i \sum_{i'} \mathrm{cov}\Big(P_{2m}(Y_i), P_{2m'}(Y_{i'})\Big).$$

However, it follows from Hannan (1970, p. 117) that $\mathrm{cov}\big(P_{2m}(Y_i), P_{2m'}(Y_{i'})\big) = 0$ for $m \neq m'$ and $\mathrm{cov}\big(P_{2m}(Y_i), P_{2m}(Y_{i'})\big) = s_{Y,i-i'}^{2m}$. Therefore,

$$\mathrm{E}\left[\sum_{i \in \mathcal{I}} \{1_{x \leq \log Y_i^2 \leq y} - F(x, y)\}\right]^2 = \sum_m a_{2m}^2(x, y)(2m)! \sum_i \sum_{i'} s_{Y,i-i'}^{2m}.$$

Let $\{r_{Y,k} = s_{Y,k}/s_{Y,0}, k \in \mathcal{L}\}$ be the autocorrelation sequence of $\{Y_i\}$. Then, by equation (A.4), we obtain

$$\sum a_{2m}^2(x,y)(2m)! \sum \sum s_{Y,i-i'}^{2m} \leq F(x,y) \sum_i \sum_{i'} r_{Y,i-i'}^2.$$

Hence

$$BE\left\{F_{Q,N}(x,x+y) - F_Q(x,x+y)\right\}^2 = B^{-1}E\left[\sum_{i\in\mathcal{I}}\{1_{x\leq\log Y_i^2\leq y} - F(x,y)\}\right]^2$$

$$\leq F(x,y)B^{-1}\sum_i\sum_{i'} r_{Y,i-i'}^2 \leq \text{ constant } F(x,y). \qquad (A.5)$$

The last inequality follows since the SDF of $\{Y_i\}$ is square integrable. Next we note that

$$\int\{\varphi(x-y) - \varphi(x)\}\,d\{F_{Q,N}(x) - F_Q(x)\}$$

$$= \int\left\{F_{Q,N}(x+y) - F_{Q,N}(x) - F_Q(x+y) + F_Q(x)\right\}d\varphi(x)$$

$$= \int\{F_{Q,N}(x,x+y) - F_Q(x,x+y)\}\,d\varphi(x),$$

so that

$$\left[\int\{\varphi(x-y) - \varphi(x)\}\,d\{F_{Q,N}(x) - F_Q(x)\}\right]^2$$

$$\leq \int\left\{F_{Q,N}(x,x+y) - F_Q(x,x+y)\right\}^2 d\varphi(x).$$

Taking expectation we obtain

$$E\left[\int\{\varphi(x-y) - \varphi(x)\}\,d\{F_{Q,N}(x) - F_Q(x)\}\right]^2$$

$$\leq \int E\left\{F_{Q,N}(x,x+y) - F_Q(x,x+y)\right\}^2 d\varphi(x).$$

So equation (A.5) yields

$$BE\left[\int\{\varphi(x-y) - \varphi(x)\}\,d\{F_{Q,N}(x) - F_Q(x)\}\right]^2$$

$$\leq \int BE\left\{F_{Q,N}(x,x+y) - F_Q(x,x+y)\right\}^2 d\varphi(x)$$

$$\leq \text{ constant}\int F_Q(x,x+y)\,d\varphi(x)$$

$$\leq \text{ constant}\int y\sup f_Q(z)\,d\varphi(x) \leq \text{ constant } y,$$

where the last inequality follow since the density function $f_Q(x)$ of $F_Q(x)$ is bounded. This completes the proof. ∎

Lemma 6

$$h_N = B^{\frac{1}{2}} \int \{\varphi(x - T_N) - \varphi(x - \mu_0)\} \, d\{F_{Q,N}(x) - F_Q(x)\} = o_P(1).$$

Proof of Lemma 6. Assume WLOG $\mu_0 = 0$. Then we note that

$$\Pr(|h_N| > \delta) \leq \Pr\left[\sup_{|y| < B^{-\gamma}} B^{\frac{1}{2}} | \int \{\varphi(x - y) - \varphi(x)\} \, d\{F_{Q,N}(x) - F_Q(x)\}| > \delta\right]$$
$$+ \Pr(|T_N| > B^{-\gamma}).$$

The second term is $o_P(1)$ by Lemma 4. The first term follows by mimicking the chaining argument in Lemma 2.2 of Koul and Surgailis (1997). As in Koul and Surgailis (1997), we prove the result for $0 \leq y \leq B^{-\gamma}$ and φ nondecreasing. We put $y_B = B^{-\gamma}$ and let

$$K = \lfloor \log_2(B y_B) \rfloor.$$

We consider a sequence of partitions

$$\{x_{i,k} = y_B i 2^{-k}, \quad 0 \leq i \leq 2^k\}, \quad k = 0, 1, \ldots, K$$

of intervals $[0, y_B]$. For a y in $[0, y_B]$ and a k in $\{0, 1, \ldots, K\}$, we define $i(k, y)$ by

$$x_{i(k,y),k} \leq y < x_{i(k,y)+1,k}.$$

We then obtain a chain by linking 0 to a given point $y \in [0, y_B]$ as

$$0 = x_{i(0,y),0} \leq x_{i(1,y),1} \leq \cdots \leq x_{i(K,y),K} \leq y < x_{i(K,y)+1,K}.$$

Let

$$R_N(y) = B^{\frac{1}{2}} \int \{\varphi(x - y)\} \, d\{F_{Q,N}(x) - F_Q(x)\}$$
$$= B^{\frac{1}{2}} \int \{F_{Q,N}(x + y) - F_Q(x + y)\} \, d\varphi(x),$$

and $R_N(y, z) = R_N(z) - R_N(y)$. We can then use the above chain to write

$$R_N(0, y) = R_N(x_{i(0,y),0}, x_{i(1,y),1}) + R_N(x_{i(1,y),1}, x_{i(2,y),2}) + \cdots$$
$$+ R_N(x_{i(K-1,y),K-1}, x_{i(K,y),K}) + R_N(x_{i(K,y),K}, y).$$

Hence

$$\sup_{y \in [0, y_B]} R_N^2(0, y)$$

$$\leq 2 \Big(\sum_{k=0}^{K-1} \sup_{y \in [0, y_B]} |R_N(x_{i(k-1,y),k-1}, x_{i(k,y),k})| \Big)^2 + 2 \sup_{y \in [0, y_B]} R_N^2(x_{i(K,y),K}, y).$$

We now apply Cauchy–Schwartz inequality to obtain

$$\begin{aligned}
\mathrm{E} \sup_{y \in [0, y_B]} R_N^2(0, y) \;\leq\; & 2K \sum_{k=0}^{K-1} \mathrm{E} \sup_{y \in [0, y_B]} R_N^2(x_{i(k-1,y),k-1}, x_{i(k,y),k}) \\
& + 2\mathrm{E} \sup_{y \in [0, y_B]} R_N^2(x_{i(K,y),K}, y).
\end{aligned}$$

We now give a bound to the last term. We use the monotonicity of $F_{Q,N}$, boundedness of φ, and the fact that F_Q is the distribution of log of a chi-square random variable. We obtain

$$\begin{aligned}
|R_N(x_{i(K,y),K}, y)| \;=\; & B^{\frac{1}{2}} \Big| \int F_{Q,N}(z + x_{i(K,y),K}, z + y) \, d\varphi(z) \\
& \qquad\qquad - \int F_Q(z + x_{i(K,y),K}, z + y) \, d\varphi(z) \Big|,
\end{aligned}$$

which is less than or equal to

$$B^{\frac{1}{2}} \int F_{Q,N}(z + x_{i(K,y),K}, z + x_{i(K,y)+1,K}) \, d\varphi(z) + \text{ constant } B^{\frac{1}{2}} y_B 2^{-K}.$$

The above is also less than or equal to

$$|R_N(x_{i(K,y),K}, x_{i(K,y)+1,K})| + \text{ constant } B^{\frac{1}{2}} y_B 2^{-K}$$

for a different choice of constant.

Next we observe that for $k = 0, 1, \ldots, K-1$

$$\begin{aligned}
\sup_{y \in [0, y_B]} |R_N(x_{i(k,y),k}, x_{i(k,y)+1,k+1})| \\
= \max_{0 \leq i \leq 2^{k+1}-1} \sup_{y \in [x_{j,k+1}, x_{j+1,k+1}]} |R_N(x_{i(k,y),k}, x_{i(k,y)+1,k+1})| \\
\leq \max_{0 \leq i \leq 2^{k+1}-1} |R_N(x_{i,k+1}, x_{i+1,k+1})|
\end{aligned}$$

Hence in view of Lemma 5, we get

$$\mathrm{E} \sup_{y \in [0, y_B]} R_N^2(x_{i(k,y),k}, x_{i(k,y)+1,k+1}) \leq \sum_{i=0}^{2^{k+1}-1} \mathrm{E} \, R_N^2(x_{i,k+1}, x_{i+1,k+1}) \leq \text{constant } y_B,$$

and similarly

$$\mathrm{E}\, R_N^2(x_{i(K,y),K}, x_{i(K,y)+1,K}) \leq \sum_{i=0}^{2^K-1} \mathrm{E}\, R_N^2(x_{i,K}, x_{i+1,K}) \leq \text{ constant } y_B.$$

Consequently,

$$\mathrm{E}\, \sup_{y \in [0, y_B]} R_N^2(0, y) \leq \text{ constant } y_B K^2 + \text{ constant } B y_B^2 2^{-2K}.$$

Now from the definition of K, we obtain $2^{-2K} = B^{-2(1-\gamma)}$, and thus

$$B y_B 2^{-2K} = O(B^{1-2\gamma-2+2\gamma}) = O(B^{-1}), \quad K^2 y_B = O(B^{-\gamma} \log_2^2 B).$$

This completes the proof. ∎

Proof of Theorem 1. By virtue of Lemma 4, we can write

$$
\begin{aligned}
O_P(B^{-1}) = \Psi_N(T_N) &= B^{-1} \sum_{i \in \mathcal{I}} \varphi(Q_i - T_N) - \int \varphi(x - \mu_0)\, dF_Q(x) \\
&= \int \varphi(x - T_N)\, dF_{Q,N}(x) - \int \varphi(x - \mu_0)\, dF_Q(x) \\
&= \int \varphi(x - T_N)\, d\{F_{Q,N}(x) - F_Q(x)\} \\
&\quad + \int \{\varphi(x - T_N) - \varphi(x - \mu_0)\}\, dF_Q(x) \\
&= \rho_N + \lambda(T_N) - \lambda(\mu_0),
\end{aligned}
$$

where $\rho_N = \int \varphi(x - T_N)\, d\{F_{Q,N}(x) - F_Q(x)\}$. This implies

$$\lambda(T_N) - \lambda(\mu_0) = o_P(B^{-\frac{1}{2}}) - \rho_N.$$

We observe that $B^{\frac{1}{2}} \rho_N$ converges to $A_\varphi^{1/2} Z$ because Lemma 3 implies that $B^{\frac{1}{2}} \int \varphi(x - \mu_0)\, d\{F_{Q,N}(x) - F_Q(x)\}$ is asymptotically normal with mean zero and variance A_φ, whereas Lemma 6 implies that

$$B^{\frac{1}{2}} \rho_N = B^{\frac{1}{2}} \int \varphi(x - \mu_0)\, d\{F_{Q,N}(x) - F_Q(x)\} + o_P(1).$$

We use a Taylor series expansion to write

$$\lambda(T_N) - \lambda(\mu_0) = \lambda'(\mu_0)(T_N - \mu_0) + o_P(|(T_N - \mu_0)|).$$

Hence

$$\lambda^{'}(\mu_0)(T_N-\mu_0)+o_P(|(T_N-\mu_0|) = B^{-1}\int \varphi(x-\mu_0)\,d\{F_{Q,N}(x)-F_Q(x)\}+o_P(B^{-\frac{1}{2}}).$$

Taking the absolute value on both sides and using the definition of $\Psi_N(\mu_0)$ and the fact that $\lambda(\mu_0) = 0$, we obtain

$$|T_N - \mu_0||\lambda^{'}(\mu_0) + o_P(1)| \;\leq\; |B^{-1}\Psi_N(\mu_0) + o_P(B^{-\frac{1}{2}})|.$$

The right hand side is $O_P(B^{-\frac{1}{2}})$, and hence $(T_N - \mu_0) = O_P(B^{-\frac{1}{2}})$. Finally

$$B^{\frac{1}{2}}(T_N - \mu_0) = \left[\lambda^{'}(\mu_0)\right]^{-1}B^{\frac{1}{2}}\Psi_N(\mu_0) + o_P(1),$$

completing the proof of Theorem 1. ∎

A.2. Proof of Theorem 2

First we assume $\{Y_i\}$ is strongly mixing, i.e.,

$$\alpha_Y(r) = \sup \alpha_Y(D, D') \to 0, \quad \text{as } r \to \infty, \tag{A.6}$$

where the supremum is taken over all pair of disjoint subsets D and D' such that $\text{dist}(D, D') = \inf_{i\in D, i'\in D'} \|i - i'\| \geq r$ and

$$\alpha_Y(D, D') = \sup_{R\in\Omega_D, R'\in\Omega_{D'}} |\text{Pr}(R\cap R') - \text{Pr}(R)\text{Pr}(R')|.$$

In the above D and D' are two sets of indices and $\Omega_D = \Omega\{Y_i : i \in D\}$ and $\Omega_{D'} = \Omega\{Y_i : i \in D'\}$ are sigma fields generated by the random variables Y_i with subscript elements of D and D' (see Bradley, 2005, or Rosenblatt, 1985, for details). It then follows that $\{R_k\}$ are strongly mixing. We also assume $\text{E}\,|Y_i|^{2+\epsilon}$ and $\text{E}\,|R_k|^{2+\eta}$ are finite for some $\epsilon, \eta > 0$.

The proof follows from the typical application given in Section 4 of Andrews and Pollard (1994). Our M-estimator $\tilde{T}_N$ is chosen to make the random function

$$\tilde{\Psi}_N(x) = B^{-1}\sum_k \tilde{\varphi}(R_k - x)$$

close to zero, in the sense that

$$\tilde{\Psi}_N(\tilde{T}_N) = o_P(U^{-\frac{1}{2}}),$$

which establishes consistency of $\tilde{T}_N$. The true $\tilde{\mu}$ is identified as the root of the corresponding expected value $\tilde{\lambda}(x)$.

Now in order to prove asymptotic normality we ensure that steps in Section 4 of Andrews and Pollard (1994) go through. Thus we need to show that $U^{\frac{1}{2}} \sum_k \tilde{\varphi}(R_k - \tilde{\mu})$ is asymptotically normal with mean zero and variance $\tilde{A}_{\tilde{\varphi}}$ and that stochastic equi-continuity of

$$U^{\frac{1}{2}} \sum_k \left(\tilde{\varphi}(R_k - x) - \tilde{\lambda}(x) \right)$$

holds at $x = \tilde{\mu}$. The asymptotic normality of $U^{\frac{1}{2}} \sum_k \tilde{\varphi}(R_k - \tilde{\mu})$ follows from Theorem 18.5.3 of Ibragimov and Linik (1971) with the use of mixing conditions. The stochastic equi-continuity holds because in our case $\tilde{\varphi}(. - x)$ is monotone and of bounded variation and thus

$$\sup_{N,i} \mathrm{E} \sup_{B(\tilde{\mu}, \delta)} |\tilde{\varphi}(R_k - x') - \tilde{\varphi}(R_k - x)|^2 \le C\delta,$$

where $B(x, \delta)$ is the ball of radius r around x. Thus the bracketing number $\mathcal{N}(\delta, \mathcal{F})$ satisfies

$$\int_0^1 x^{-\gamma/(2+\gamma)} \mathcal{N}(\delta, \mathcal{F})^{1/q} \, dx < \infty,$$

where $\mathcal{F} = \{ \tilde{\varphi}(. - x) : x \in \mathbb{R} \}$ with semi-norm ρ defined by

$$\rho(x, x') = \rho \left(\tilde{\varphi}(. - x), \tilde{\varphi}(. - x') \right).$$

Hence Theorem 2.2 of Andrews and Pollard (1994) applies. Therefore

$$U^{\frac{1}{2}} \sum_k \left(\tilde{\varphi}(R_k - \tilde{T}_N) - \tilde{\lambda}(\tilde{T}_N) \right) = U^{\frac{1}{2}} \sum_k \tilde{\varphi}(R_k - \tilde{\mu}) + o_P(1).$$

Since

$$\tilde{\lambda}(\tilde{T}_N) = \tilde{\lambda}'(\mu)(\tilde{T}_N - \tilde{\mu}) + o_P(|\tilde{T}_N - \tilde{\mu}|),$$

we obtain

$$
\begin{aligned}
o_P(U^{-\frac{1}{2}}) &= \tilde{\Psi}_N(\tilde{T}_N) = \tilde{\lambda}(\tilde{T}_N) + U^{-1} \sum_k \left(\tilde{\varphi}(R_k - \tilde{T}_N) - \tilde{\lambda}(\tilde{T}_N) \right) \\
&= \tilde{\lambda}'(\tilde{\mu})(\tilde{T}_N - \tilde{\mu}) + o_P(|\tilde{T}_N - \tilde{\mu}|) + U^{-1} \sum_k \tilde{\varphi}(R_i - \tilde{\mu}) + o_P(U^{-\frac{1}{2}}).
\end{aligned}
$$

Hence

$$\left| U^{-1} \sum_k \tilde{\varphi}(R_k - \tilde{\mu}) + o_P(U^{-\frac{1}{2}}) \right| \ge |\tilde{\lambda}'(\tilde{\mu}) - o_P(1)||\tilde{T}_N - \tilde{\mu}|.$$

The left hand side is $O_P(U^{-\frac{1}{2}})$, and hence $(\tilde{T}_N - \tilde{\mu}) = O_P(U^{-\frac{1}{2}})$. Finally

$$U^{\frac{1}{2}}(\tilde{T}_N - \tilde{\mu}) = \tilde{\lambda}'(\tilde{\mu})U^{-\frac{1}{2}}\sum_k \tilde{\varphi}(R_k - \tilde{\mu}) + o_P(1),$$

and this completes the proof.   ■

References

Andrews, D. W. K. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* **62**, 119–132.

Bartlett, M. S. and Kendall, D. G. (1946). The statistical analysis of variance-homogeneity and the logarithmic transformation. *Supplement to the Journal of the Royal Statstical Society* **8**, 128–138.

Beran, J. (1991). *M*-estimators of location for data with slowly decaying correlations. *Journal of the American Statistical Association* **86**, 704–708.

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

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys* **2**, 107–144.

Bretherton, C. S., Uttal, T., Fairall, C. W., Yuter, S., Weller, R., Baumgardner, D., Comstock, K., Wood, R. and Raga, G. (2004). The EPIC 2001 stratocumulus study. *Bull. Amer. Meteor. Soc.* **85**, 967–977.

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

Chave, A. D., Thomson, D. J. and Ander, M. E. (1987). On the robust estimation of power spectra, coherences, and transfer functions. *Journal of Geophysical Research* **92**, 633–648.

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

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

Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society B* **43**, 350–363.

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

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

Hannan, E. J. (1970). *Multiple Time Series*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley and Sons Inc.

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

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Koul, H. L. and Surgailis, D. (1997). Asymptotic expansion of M-estimators with long-memory errors. *The Annals of Statistics* **25**, 818–850.

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

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

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

Mondal, D. (2007). *Wavelet variance analysis for time series and random fields*. PhD dissertation, University of Washington.

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

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

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

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge: Cambridge University Press.

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

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser Boston Inc. Boston, MA.

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

Schlather, M. (2006). *The RandomFields Package.* Simulation and analysis of random fields. Cran.r-project.org\doc \packages\RandomFields.pdf. Version 1.3.29.

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

Stevens, B., Vali, G., Comstock, K., Wood, R., van Zanten, M. C., Austin, P. H., Bretherton, C. S. and Lenschow, D. H. (2005). Pockets of open cells (POCs) and drizzle in marine stratocumulus. *Bull. Amer. Meteor. Soc.* **86**, 51–57.

Stoev, S. and Taqqu, M. S. (2003). Wavelet estimation of the Hurst parameter in stable processes. *Processes with Long Range Correlations: Theory and Applications*, eds. D. Rangarajan and M. Ding. Springer Verlag. Lecture notes in Physics no. **621**, 61–87.

Stoev, S., Taqqu, M. S., Park, C., Michailidis, G. and Marron, J. S. (2006). LASS: a tool for the local analysis of self-similarity. *Computational Statistics and Data Analysis* **50**, 2447–2471.

Thall, P. F. (1979). Huber-sense robust *M*-estimation of a scale parameter, with applications to the exponential distribution. *Journal of the American Statistical Association* **74**, 147–152.

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

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

Unser, M. (1995). Texture classification and segmentation using wavelet frame. *IEEE Transactions on Image Processing* **4**, 1549–1560.

Vannucci, M. and Liò, P. (2001). Non-decimated wavelet analysis of biological sequences: applications to protein structure and genomics. *Sankhyā Ser. B* **63**, 218–233.

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

Zhu, Z. and Stein, M. L. (2002). Parameter estimation for fractional Brownian surfaces. *Statistica Sinica* **12**, 863–883.

Department of Statistics, University of Chicago, 5734 S University Ave, Chicago, IL, 60637

E-mail: (mondal@galton.uchicago.edu)

Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195

E-mail: (dbp@apl.washington.edu)