M-ESTIMATION OF WAVELET VARIANCE †

Debashis Mondal
University of Chicago, Chicago, USA.
Donald B. Percival
University of Washington, Seattle, USA.

Summary. 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 multiscale 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 introduces a scale-based contamination model and describes robust estimation of the wavelet variance that can 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.

Keywords: Hermite expansion; Multiscale processes; Pockets of open cells; Robust estimation; Time series analysis

1. Introduction

Wavelets decompose a stochastic process (e.g., a time series or a random field) into sub-processes, each one of which is associated with a particular scale. A wavelet variance is the variance of a sub-process at a given scale and quantifies the amount of variation present at that particular scale. Wavelet variances at different scales enable us to perform an analysis of variance; see Percival and Walden (2000) and the references therein. This variance is of interest in numerous geophysical and related applications, where the overall observed process is an ensemble of sub-processes that operate at different scales. In addition, the wavelet variance serves as an exploratory tool to study power law processes (Stoev and Taqqu, 2003), detect inhomogeneity (Whitcher et al., 2002), estimate spectral densities indirectly (Tsakiroglou and Walden, 2002), and 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), 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).

†*Address for correspondence*: Debashis Mondal, Department of Statistics, Univerisity of Chicago, Chicago, IL, 60637-1415, USA. E-mail: mondal@galton.uchicago.com

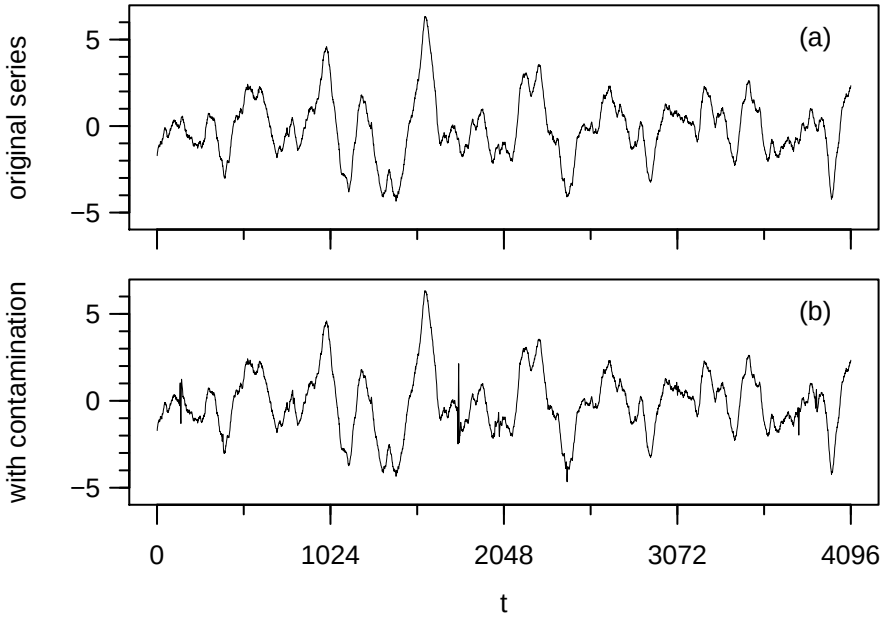

Fig. 1. Simulated ocean shear data. The top plot (a) is of the original simulated series, while the bottom plot (b) shows the series after contamination.

.

In practice, however, data collected in the form of a time series or random field often suffer from contamination, which can occur in various ways. For example, the satellite cloud data we consider in section 8 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. A moderate amount of contamination often has a very adverse effect on conventional estimates of the wavelet variance. In classical statistics, it is common to handle contamination via removing questionable data points and then applying conventional methods; however, locating contamination in a time series or in a random field is often complicated because of dependence in the data. A better approach is to consider a contamination model, and then derive a robust procedure that protects against that model. In such situations, one hopes to measure robustness via influence functions and to form estimates that reduce the effect of bias in an optimum way. Several contamination models have been suggested in the time series literature. These include pure replacement models, additive outliers, level shift models and innovation outliers that hide themselves in the original time series. Chapter 8 of Maronna, Martin and Yohai (2006) provides an excellent account of the theory and methods of robust time series statistics based on these types of contamination models. However, robust nonparametric estimates of second-order statistics based on these contaminated models still present difficult problems. For example, Robinson (1986) observes that, under both pure replacement and additive outlier models, autocovariances of the unobserved uncontaminated process are unrecognizable from those of the observed contaminated process. Robinson's critique extends to wavelet variances, making the estimation and the inference of wavelet

variances for contaminated data a difficult problem. Moreover contamination can be quite complex when a time series or random field is multiscale in nature. Different contamination processes can act on different scales independently, but contamination from one scale can leak into another and be hard to detect. A practical approach is to use the median of the squared wavelet coefficients rather than their mean to estimate the wavelet variance. Stoev et al. (2006) touch on the usefulness of median-type estimators to guard against contamination. In this paper, we develop a full M-estimation theory for the wavelet variance and derive its large sample theory when the underlying process is Gaussian. Our approach treats the wavelet variance as a scale parameter and offers protection mostly against scale-based multiplicative contamination (i.e., additive on the log of squared wavelet coefficients) that acts independently at different scales. We will return to the discussion on robustness in section 9.

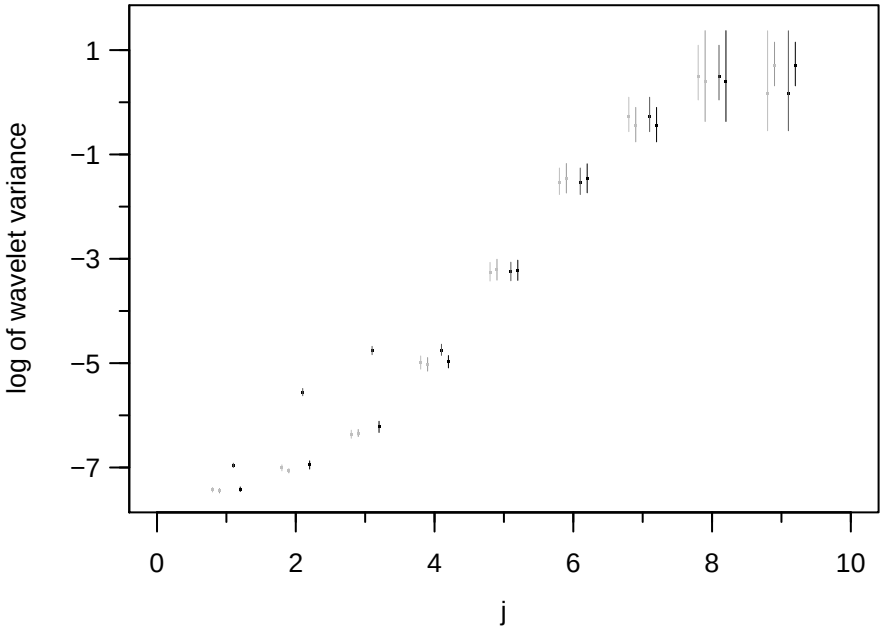

Fig. 2. Confidence intervals (CIs) for the wavelet variance based upon mean- and median-type estimators of the wavelet variance using simulated ocean shear data (Figure 1a) and a contaminated version thereof (Figure 1b). CIs for nine scales $\tau_j = 2^{j-1}$, $j = 1, \ldots, 9$, are displayed. For each scale there are four lines jittered about the level j. From left to right, these show 95% CIs based upon the mean-type estimator with uncontaminated data; median-type with uncontaminated; mean-type with contaminated; and median-type with contaminated.

To illustrate our proposed methodology, let us consider the following example motivated by vertical shear measurements in the ocean (see section 7 for details). Such series are subject to bursts of turbulence isolated in a select number of scales, but burst-free segments are well modeled by a stationary process. Figure 1a shows a simulated Gaussian series of length 4096 from this stationary process. We compute its wavelet coefficients for scales $\tau_j = 2^{j-1}$, $j = 1, \ldots, 9$, and, following Percival (1995), form 95% confidence intervals (CIs) for the wavelet variance based upon averaging squared coefficients for a given scale (left-most lines in Figure 2 in the set of four lines jittered about each of the nine scales).

We form an alternative estimator by taking the median of the squared coefficients and correcting for bias as dictated by our theory (see equation (15)). The second-from-left lines in each group of four show the 95% CIs based upon the robust median-based estimates. We see good agreement with the mean-type estimates at all scales. Next we contaminate the time series by introducing scale-based multiplicative noise, which is intended to simulate bursts of turbulence focused around scale τ_3 (Figure 1b). The second-from-right lines in Figure 2 show the CIs for the wavelet variance based upon this contaminated data and the usual mean-type estimator. Outliers significantly change the wavelet variance estimates at small scales. Any inferences that we might want to draw about the turbulent-free vertical shear process at small scales would be materially influenced by the contamination. Finally the right-most lines in Figure 2 show CIs based upon our robust median-based estimator and the contaminated series. The CIs are close to those for the uncontaminated series, making it possible to draw inferences about the turbulent-free vertical shear process from the contaminated data. Our robust procedure thus works well when observed data are subjected to scale-based multiplicative contamination of an underlying Gaussian process.

The remainder of the paper is organized as follows. Section 2 gives some background on the theoretical wavelet variance and its estimation for both time series and random fields. Section 3 presents our theory for M-estimation of the wavelet variance, with proof of the main result (Theorem 1) deferred to the appendix. Due to use of a logarithmic transform, our raw M-estimators are biased, so section 4 discusses how to correct for this bias. Section 5 considers how to construct CIs for the wavelet variance based upon our M-estimators. Sections 6 and 7 discuss computer experiments that investigate the efficacy of our proposed methodology, and section 8 considers a substantive application to two-dimensional cloud data. Section 9 concludes the paper with some discussion.

2. Wavelet Analysis of Variance

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:

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

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 $i = 1, \ldots, L/2$. Let $H_1(f)$ denote the transfer functions (Fourier transforms) of the filters $\{h_{1,l}\}$. The jth level wavelet filter $\{h_{j,l}\}$ is defined as the inverse discrete Fourier transform (DFT) of

$$H_j(f) = H_1(2^{j-1} f) \prod_{l=0}^{j-2} H_1(\tfrac{1}{2} - 2^l f). \tag{1}$$

The width of this filter is given by $L_j = (2^j - 1)(L - 1) + 1$.

2.2. Wavelet Variance for Time Series

Let $\{X_t, t \in \mathcal{Z}\}$ 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

$$\nu_j^2 = \text{var}\,\{W_{j,t}\}.$$

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

$$\text{var}\,\{X_t\} = \sum_{j=1}^{\infty} \nu_j^2 \tag{2}$$

as an alternative to the classical decomposition

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

The decomposition of 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 ν_j^2 is given by

$$\hat{\nu}_j^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} W_{j,t}^2, \tag{3}$$

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 CIs.

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 four types of coefficients, of which the so-called wavelet-wavelet coefficients are of primary interest to us:

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

The wavelet variance is defined in terms the variance of these coefficients:

$$\nu_{j,j'}^2 = \text{var}\,\{W_{j,j',u,v}\}.$$

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 $\nu^2_{j,j'}$ is well defined if $L \geq d$. The wavelet variance decomposes the process variance since

$$\text{var}\,\{X_{u,v}\} = \sum_{j=1}^{\infty} \sum_{j'=1}^{\infty} \nu^2_{j,j'}. \tag{4}$$

This generalizes the result for time series stated in (2) 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 variance by the unbiased estimator

$$\hat{\nu}^2_{j,j'} = \frac{1}{N_j M_{j'}} \sum_{u=L_j-1}^{N-1} \sum_{v=L_{j'}-1}^{M-1} W^2_{j,j',u,v}, \tag{5}$$

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.

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 $\text{var}\,\{Y_i\} = \text{E}\, Y_i^2 = \nu^2$. Here the index i represents either time t or spatial location (u, v), and hence $\mathcal{L}$ is an integer lattice, either $\mathcal{Z}$ or $\mathcal{Z}^2$. Typically $\{Y_i\}$ is either the wavelet coefficient process $\{W_{j,t}\}$ or the wavelet-wavelet coefficients $\{W_{j,j',u,v}\}$ for any fixed scale (or scale pair). In practice the simple mean-type estimators (3) and (5) are vulnerable to data contamination, so we are interested in developing robust estimators that guard against such contamination, yet still work well when Gaussianity holds. Under our assumptions, ν^2 is a scale parameter. A logarithmic transformation converts it 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

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

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 $\text{E}\, \epsilon_i = 0$ and $\text{var}\,\{\epsilon_i\} = \pi^2/2$.

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*

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

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 differentiable and $\lambda'(x)$ is continuous around μ_0.

The relationship between the solution point μ_0 and the location parameter μ is discussed in section 4.

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 first derivative of λ satisfies the relation

$$\lambda'(x) = -\int f_Q(y+x)\, d\varphi(y).$$

Given the form of F_Q and the fact that φ is of bounded variation, it follows that $\lambda'(x)$ is bounded as well.

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

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

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

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

for $p > 0$ and

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

(see equation (3.9) of Thall, 1979, for another choice). To better understand M-estimation, consider an independent and identically distributed (IID) 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 the 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-IID 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 (proven in the appendix) 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, and $\{Y_i\}$ has a square integrable SDF. 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*

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

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, so the general proof of Theorem 1 in the appendix takes a substantially different approach.

4. 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

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

This bias depends on the choice of the weight function φ and on the distribution function F_Q. We can compute it analytically in some cases. To do so, we first determine the function $\lambda(x)$. Let Z, ϕ and Φ denote the standard Gaussian random variable and its density and distribution functions. 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

$$\lambda_I(x) = 3 - 4\Phi\Big(\mathrm{e}^{\frac{x}{2} - \log \nu}\Big). \tag{6}$$

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

$$\mu_0 = 2 \log \nu + 2 \log\big[\Phi^{-1}(\tfrac{3}{4})\big]. \tag{7}$$

and hence

$$\text{bias}_I = \mu - \mu_0 = \psi(\tfrac{1}{2}) + \log 2 - 2 \log\big[\Phi^{-1}(\tfrac{3}{4})\big]. \tag{8}$$

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

$$\varphi_{II}(x) = 4\,\Phi\Big(\mathrm{e}^{\frac{x}{2}+\frac{1}{2}\psi(\frac{1}{2})+\frac{1}{2}\log 2}\Big) - 3.$$

Therefore we obtain

$$\lambda_{II}(x) = 4\,\mathrm{C}_a\Big(\mathrm{e}^{\log\nu-\frac{x}{2}-\frac{1}{2}\psi(\frac{1}{2})-\frac{1}{2}\log 2}\Big) - 3, \tag{9}$$

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

$$\mu_0 = 2\log\nu - \psi(\tfrac{1}{2}) - \log 2 + 2\log\big[\mathrm{C}_a^{-1}(\tfrac{3}{4})\big]. \tag{10}$$

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.

5. 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 CI for μ_0 and hence for μ and ν^2. Since A_φ is equal to the SDF of the stationary process $\{\varphi(Q_i - \mu_0)\}$ at zero frequency, we use a multitaper spectral approach to estimate it (Serroukh et al., 2000). 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 and design bandwidth parameter $W = 4/N$, 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

$$J_k = \sum_i \beta_{k,i}\varphi(Q_i - T_N), \tag{11}$$

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

$$\tilde{\mu} = \frac{\sum_k J_k\beta_{k,\cdot}}{\sum_k \beta_{k,\cdot}^2}, \text{ where } \beta_{k,\cdot} = \sum_i \beta_{k,i}.$$

We then estimate A_φ by

$$\hat{A}_\varphi = \frac{1}{K}\sum_k (J_k - \tilde{\mu}\beta_{k,\cdot})^2, \tag{12}$$

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 (11). Thus the resulting multitaper 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 bandwidth of the resulting multitaper estimator is approximately $7/N$.

What remains is to compute $\lambda'(\mu_0)$. For $\varphi = \varphi_I$, we use equation (6) and equation (7) to obtain

$$\lambda_I^{'}(\mu_0) = -2\phi\Big(\mathrm{e}^{\frac{\mu_0}{2}-\log\nu}\Big)\mathrm{e}^{\frac{\mu_0}{2}-\log\nu} = -2\phi\Big(\Phi^{-1}(\tfrac{3}{4})\Big)\Phi^{-1}(\tfrac{3}{4}). \tag{13}$$

Similarly for $\varphi = \varphi_{II}$, use of equation (9) gives

$$\lambda_{II}^{'}(\mu_0) = -2\,\mathrm{c}_a\Big(\mathrm{e}^{\log\nu-\frac{\mu_0}{2}-\frac{1}{2}\psi(\frac{1}{2})-\frac{\log 2}{2}}\Big)\mathrm{e}^{\log\nu-\frac{\mu_0}{2}-\frac{1}{2}\psi(\frac{1}{2})-\frac{\log 2}{2}},$$

where c_a is the density function of C_a. By using equation (10), we can simplify the above to

$$\lambda_{II}^{'}(\mu_0) = -2\,\mathrm{c}_a\Big(\mathrm{C}_a^{-1}(\tfrac{3}{4})\Big)\,\mathrm{C}_a^{-1}(\tfrac{3}{4}). \tag{14}$$

This has the same form as that of equation (13) with the Gaussian density function being replaced by the Cauchy. For other choices of φ, namely, $\varphi = \varphi_{III}$ and φ_{IV}, there is no convenient analytic form (although one might surmise that it will have a form similar to equations (13) and (14)), but we can evaluate $\lambda^{'}(\mu_0)$ numerically.

6. Efficiency Study

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

$$\mathrm{e}^{T_N} \sim \text{ Log-normal } \Big(\mu_0, \frac{\sigma^2}{B}\Big), \text{ where } \sigma^2 = \frac{A_\varphi}{[\lambda'(\mu_0)]^2}.$$

Since $\mathrm{E}\,\exp(T_N) \approx \exp(\mu_0 + \sigma^2/(2B))$, an approximately unbiased and robust estimator of ν^2 is given by

$$\tilde{\nu}^2 = \exp\Big(T_N + \text{bias} - \psi(\tfrac{1}{2}) - \log 2 - \frac{\hat{\sigma}^2}{2B}\Big), \text{ where } \hat{\sigma}^2 = \frac{\hat{A}_\varphi}{[\lambda'(\mu_0)]^2}. \tag{15}$$

Assuming that $\hat{A}_\varphi$ is a consistent estimator of A_φ, the ARE is equal to

$$ARE = \lim_{B\to\infty} \frac{\mathrm{var}\,\{\hat{\nu}^2\}}{\mathrm{var}\,\{\tilde{\nu}^2\}} = \frac{A[\lambda'(\mu_0)]^2}{\nu^4 A_\varphi}, \tag{16}$$

where $\hat{\nu}^2 = B^{-1}\sum_i Y_i^2$ is the usual mean-type estimator, which is asymptotically normal with mean ν^2 and variance

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

(Percival, 1995; Mondal, 2007). 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, specializing to the case $\varphi = \varphi_I$, i.e., the median-type estimator.

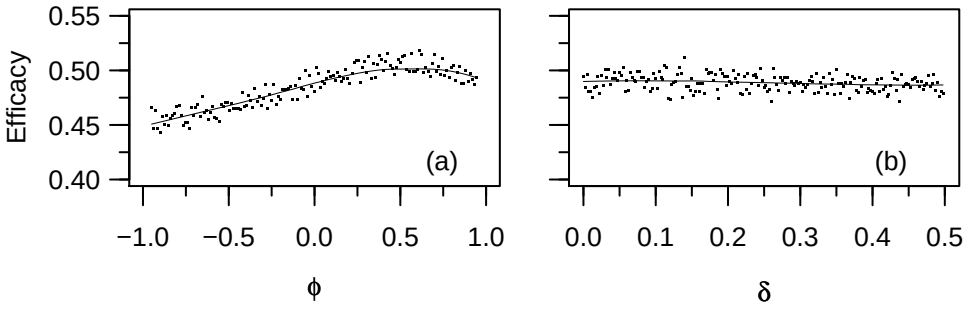

Fig. 3. Approximate asymptotic relative efficiency of the median-type estimator with respect to the mean-type estimator for (a) AR(1) processes with parameter ϕ and (b) FD processes with parameter δ.

In the first example, we simulate 10,000 AR(1) time series of length 1024 for various values of the unit lag correlation ϕ (Kay, 1981). For each series we take $Y_t = W_{2,t}$, the level $j = 2$ Haar wavelet coefficients. We compute exact values of ν^2, μ_0 and $\lambda'(\mu_0)$, so we only need A/A_φ to determine the ARE via (16). This ratio can be approximated by $\text{var}\{\sum Y_t^2\}/\text{var}\{\sum \varphi_I(\log Y_t^2 - \mu_0)\}$. 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 10,000 replications. Figure 3a plots the estimated ARE against $\alpha \in (-1, 1)$. The ARE is about 50% for all ϕ, is smallest when ϕ is close to -1 and attains its peak value at about $\phi = .75$, above which it then declines slightly. As an additional check of our theory, we note that we obtain similar results if we just compute the ratio of the Monte Carlo variances for $\hat{\nu}^2$ and $\tilde{\nu}^2$ as suggested by equation (16).

For the second example, we consider a stationary fractionally differenced (FD) process with long memory parameter $\delta \in (0, \frac{1}{2})$ ($\delta = 0$ corresponding to white noise, and the process becomes more highly correlated as δ approaches 1/2). For selected δ we simulate 10,000 FD time series of length 1024 (Craigmile, 2003) and estimate the ARE for median-type versus mean-type estimators as in the AR(1) example. Figure 3b plots the ARE versus δ. We see that the ARE is close to 50% for all δ.

A simulation experiment with fractional Brownian surfaces (see, e.g., Zhu and Stein, 2002) gives about 60–65% efficiency (details are omitted).

7. A Simulation Study

Here we report the results of a simulation study that demonstrates the efficacy of our theory and that expands upon the example considered in Figures 1 and 2. We consider a Gaussian stationary process that models a burst-free portion of an actual ocean shear series previously considered in Percival and Guttorp (1994), Percival (1995) and Percival and Walden (2000). The SDF for this model is piecewise power law (Percival and Walden, 2000, p. 331). We generate realizations of length $N = 4096$ using the Gaussian spectral synthesis method (Percival and Walden, 2000, section 7.8; the parameter M described there is set to $4N$). We compute the usual (mean-type) unbiased estimator $\hat{\nu}_j^2$ of the wavelet variance for scales $\tau_j = 2^{j-1}$, $j = 1, \ldots, 9$, using equation (3) with $\{h_{j,l}\}$ based upon D(6) wavelet filter,

Table 1. Summary of wavelet variance estimates of uncontaminated simulated ocean shear data

j	mean $\{\hat{\nu}_j^2\}$	$\tilde{\nu}_j^2/\hat{\nu}_j^2$	SD $\{\hat{\nu}_j^2\}$	SD $\{\tilde{\nu}_j^2\}$	relative efficiency
1	0.000585	1.00	0.000015	0.000022	48.1%
2	0.000904	1.00	0.000031	0.000041	57.7%
3	0.001743	1.00	0.000080	0.000102	60.7%
4	0.007150	1.00	0.000550	0.000693	63.0%
5	0.040047	1.00	0.004489	0.005638	63.4%
6	0.214502	1.01	0.034153	0.042587	64.3%
7	0.858346	1.01	0.216022	0.269830	64.1%
8	1.197886	1.03	0.422244	0.547121	59.6%
9	0.641284	1.27	0.367127	0.567690	41.8%

i.e., the Daubechies extremal phase filter of width $L = 6$ (Daubechies, 1992, Table 6.1). We then compute an approximately unbiased median-type robust estimator $\tilde{\nu}_j^2$. To do so, we let T_N be the log of the median of the same squared coefficients used to form $\hat{\nu}_j^2$ and then use it in equation (15), along with the bias term given by equation (8), $\hat{A}_\varphi$ by (12) and $\lambda'(\mu_0)$ by (13). We next corrupt the simulated time series by taking its orthonormal D(6) discrete wavelet transform (DWT) and multiplying selected level $j = 3$ coefficients by log Gaussian white noise $\exp(\epsilon_t)$ with $\mathrm{E}\,\epsilon_t = 0$ and $\mathrm{var}\,\{\epsilon_t\} = 1.5$. There are 512 DWT coefficients at level $j = 3$, of which we select 51 at random for alteration. Additionally, we alter randomly chosen patches of coefficients, where the patchiness is dictated by a realization of a stationary Markov chain η_t with

$$\Pr(\eta_t = 1|\eta_{t-1} = 0) = 0.09, \quad \Pr(\eta_t = 0|\eta_{t-1} = 1) = 0.01$$

and $\Pr(\eta_t = 0) = 0.1$. Any level $j = 3$ coefficient with a corresponding η_t of zero is multiplied by log Gaussian white noise with the same statistical properties as before. The total number of altered coefficients on the average is approximately 97 allowing some coefficients to be altered twice. We take the inverse DWT to create a contaminated version of the original simulated time series (note that, although we have altered just the level $j = 3$ DWT coefficients, the wavelet coefficients that are used to estimate the wavelet variance are based on the over-complete maximal overlap DWT (Percival and Walden, 2000), for which we can expect the contamination to leak out into scales adjacent to τ_3). We then compute the mean- and median-type wavelet variance estimators. Finally we repeat this entire process over again for 1000 different realizations.

Tables 1 and 2 summarize the results of our simulation study. Table 1 concerns just the uncontaminated series. The second column shows the average of the 1000 estimates of $\hat{\nu}_j^2$. The third column shows the ratio of the average of the median-type estimates $\tilde{\nu}_j^2$ to the average of the mean-type estimates $\hat{\nu}_j^2$. With the exception of the largest scale τ_9, these ratios are very close to unity, which indicates that the mean- and median-type estimates match up quite well. The poorer agreement at scale τ_9 can be attributed to a downward bias in the estimator $\hat{A}_\varphi$ due to a sparsity of relevant data at that scale (if we were to replace $\hat{A}_\varphi$ in (15) with its actual value, the modified version of $\tilde{\nu}_9^2$ would agree well with $\hat{\nu}_9^2$). The fourth and fifth columns give the sample standard deviations (SDs) for the 1000 estimates of, respectively, $\hat{\nu}_j^2$ and $\tilde{\nu}_j^2$. The final column gives the estimated relative efficiency of the median-type estimator to the mean-type estimator. It is interesting that the efficiency is

Table 2. Summary of wavelet variance estimates of contaminated ocean shear data

j	$\hat{\nu}^2_{C,j}/\hat{\nu}^2_j$	$\tilde{\nu}^2_{C,j}/\tilde{\nu}^2_j$	SD $\{\hat{\nu}^2_{C,j}\}$	SD $\{\tilde{\nu}^2_{C,j}\}$	RMSE $\{\hat{\nu}^2_{C,j}\}$	RMSE $\{\tilde{\nu}^2_{C,j}\}$
1	12.35	1.05	0.160084	0.000028	0.160142	0.000042
2	62.68	1.12	1.366784	0.000065	1.367238	0.000130
3	72.02	1.13	3.046282	0.000157	3.047273	0.000278
4	4.98	1.07	0.693477	0.000743	0.693714	0.000885
5	1.04	1.01	0.035961	0.005404	0.035981	0.005413
6	1.01	1.00	0.033752	0.043243	0.033762	0.043225
7	1.00	1.01	0.220646	0.274201	0.220536	0.274209
8	0.97	0.99	0.385875	0.504024	0.387926	0.503808
9	1.02	1.32	0.391146	0.668527	0.391120	0.698292

markedly smaller for scale τ_1 than for scales τ_2 to τ_8. The small efficiency for scale τ_9 can again be attributed to lack of sufficient data relevant to this scale (if we increase the sample size to $N = 4 * 4096$, the efficiency increases back up to above 60%).

Table 2 shows results involving the contaminated series, where $\hat{\nu}^2_{C,j}$ and $\tilde{\nu}^2_{C,j}$ denote $\hat{\nu}^2_j$ and $\tilde{\nu}^2_j$ when applied to the contaminated series. The second column shows the ratio of the average of the mean-type estimates $\hat{\nu}^2_{C,j}$ for the contaminated series to a similar average for the uncontaminated series, while the third column shows a similar ratio, but now involving the average of the median-type estimates $\tilde{\nu}^2_{C,j}$ for the contaminated series. Contamination has a very adverse effect on the mean-type wavelet variance estimates at small scales (τ_1 to τ_4), whereas the median-type estimates are far more resistance to contamination. Using the latter, we can draw inferences close to those we made from the uncontaminated data. Note that the contamination has little effect on the mean- or median-type estimates at scales τ_5 and above (the relatively poor performance of $\tilde{\nu}^2_{C,9}$ can again be attributed to bias in the estimator $\hat{A}_\varphi$). The remaining columns in the table show the sample SDs and root mean-square errors (RMSEs) for the 1000 estimates of $\hat{\nu}^2_{C,j}$ and $\tilde{\nu}^2_{C,j}$. The RMSEs suggest that the robust estimator performs better up to scale τ_4, beyond which it is better to switch over to the mean-type estimator.

The time series that we selected from amongst the 1000 series to show in Figure 1 is typical in the sense that its level $j = 3$ actual RMSE for the contaminated mean-type estimate is closest to sample RMSE for the 1000 such estimates. The CIs based upon this series which are displayed in Figure 2 are thus also typical of what we can expect to get. Note that the CIs for ν^2_9 based upon the median-type estimates are markedly smaller than those based upon the mean-type estimates, which is not the case for scales τ_1 to τ_8. This anomaly can again be attributed to bias in the estimator $\hat{A}_\varphi$ due to lack of sufficient data at that scale.

8. Applications to Cloud Data

Figure 4 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 (Bretherton et al., 2004). 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 (Stevens et

al., 2005). The four squares in Figure 4 indicate regions with four different types of clouds. Region (a) contains POCs; (b) consists of uniform stratus clouds and thus has different characteristics than the POC region; (c) has broken clouds; and (d) has clouds that are forming a POC.

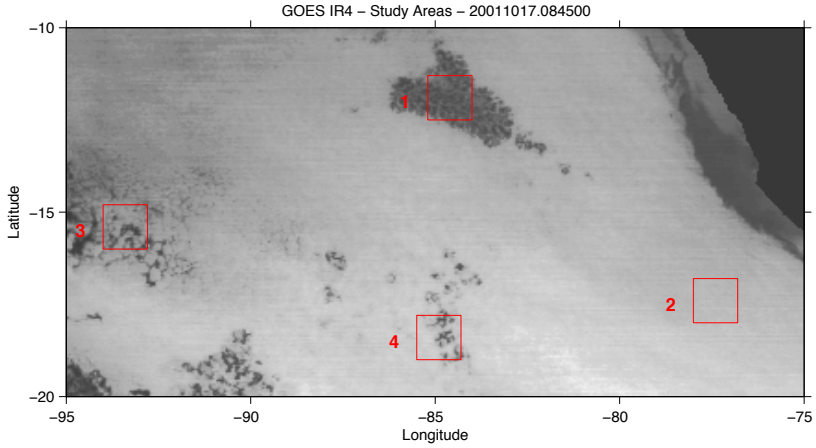

Fig. 4. Image plots of clouds at four different regions

.

Satellite cloud data are often marked by contamination from various sources, but the one of interest to us here is the presence of aberrant cloud types in a region where a single cloud type is dominant. The four regions of focus in Figure 4 all have a dominant cloud type, but they are homogeneous to varying degrees, with regions (a) and (b) being visually more so than (c) and (d). It is thus natural to resort to median-type estimators that are robust and effective in extracting the characteristics of the dominant cloud type.

Figure 5 shows conventional mean-type (gray squares) and robust median-type (black) wavelet variance estimates and their 95% CIs (lines intersecting squares vertically) for four diagonal scales (i.e., $j' = j$) indexed by $j = 1, 2, 3$ and 4. Plot (a) is for the POC region. At the three smallest scales (τ_1 to τ_3) the median-type estimates take somewhat lower values than those for the mean-type estimates, with the largest discrepancy occurring at scale τ_1. While this pattern is consistent with this region being contaminated to some degree by cloud types other than POCs, the fact that the mean- and median-type estimates are comparable to within the sampling variability indicated by the associated overlapping CIs suggests that the POC region is largely unaffected by other cloud types. Plot (b) is for the uniform stratus clouds. Again the mean- and median-type wavelet variance estimates and the associated CIs suggest that this region is largely homogenous; however, unlike the POC region, the CI based upon median-type estimate at scale τ_1 is markedly wider than the one based upon the mean-type estimate, which leads to speculate that the noise characteristics of two regions are different. Plots (c) and (d) show the results for the regions with broken clouds and POC formation. At scales τ_1 and τ_2, the robust estimates of wavelet variances for the broken clouds are almost an order of magnitude lower than the conventional estimates, but also have much larger confidence intervals. Broken clouds are mixtures of various clouds, so the median-type estimates should pick out the characteristics from the dominant cloud type; however, a larger fraction of other cloud types produces bigger confidence intervals.

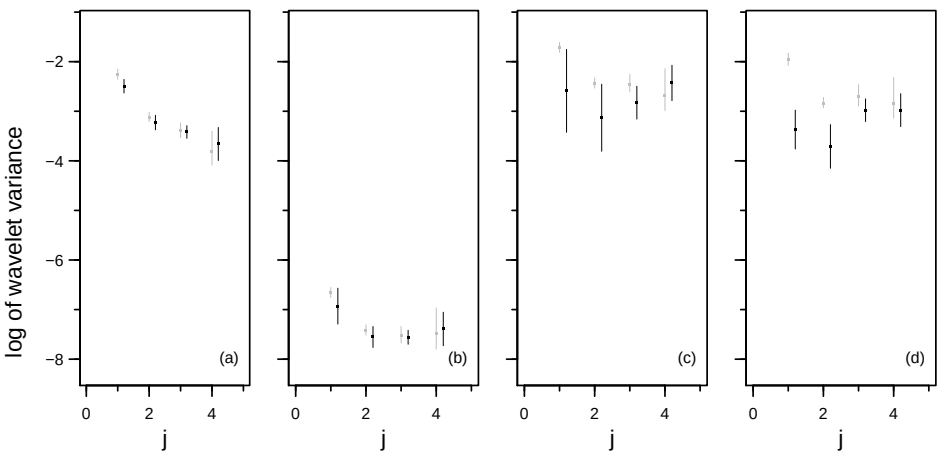

Fig. 5. Log of wavelet variances at diagonal scales indexed by $j = 1$, 2, 3 and 4 for four cloud regions. The gray and black squares show, respectively, the mean-type and robust median-type estimates. The vertical lines bisecting the squares depict 95% CIs. Plot (a) is for the POC region, (b) for uniform clouds, (c) for broken clouds and (d) for a forming POC region

.

The robust estimates of wavelet variance for the region of POC formation are again much smaller than for the conventional estimates, evidently because of the presence of other cloud types; however, unlike the broken clouds region, the CIs based upon the robust procedure are smaller in this region, indicating that the fraction of other clouds present here is smaller than the fraction in the broken clouds region.

One potential use for the wavelet variance in this application is to define features that can be used to classify cloud regions on other images. The wavelet variance curve for a POC region is uniquely identified by a monotonic decrease across scales τ_1 to τ_4, while an overall low level is indicative of uniform clouds. In both cases, the mean- and median-type estimates are comparable. By contrast, these two estimates are markedly different for the broken clouds region and the region of POC formation, with large uncertainty in the median-based estimate being a potential identifier for the broken clouds region. More research is needed to determine if these patterns persist enough across other images to serve as usual features for classification.

9. Discussion

The subject of robustness in statistics has been around for more than a quarter of a century, but the usefulness of robust estimators in practice has been the subject of some controversy; e.g., Stigler (1977) questions the usefulness of the median as an estimator of the location parameter for real data. We have seen in Section 8 that mean- and median-type estimates of the wavelet variance provide different answers at certain scales for inhomogeneous cloud regions. The robust estimates arguably allow us to pick out the characteristics of the dominant cloud type better, thus providing an argument for the usefulness of median-type estimates.

While a number of different contamination models has been entertained for time series data in the literature, we have introduced a new model based upon the idea of scale-based multiplicative contamination. This model is based on the supposition that contamination can occur and affect data at certain scales. Our computer experiments indicate that, even if contamination is present at some small scales, the larger scales will not be influenced much and should have wavelet coefficients that are close to Gaussian. Table 2 shows that the mean-type estimators at larger scales produce smaller mean square error in comparison to robust median-type estimators. A step forward would be to identify the contaminated scales and devise a hybrid scheme whereby we use the robust estimator for scales at which data are affected by contamination, and then switch over to the mean-type estimator at other scales. This strategy is used in certain wavelet shrinkage problems involving non-Gaussian data (Gao, 1997). A question for future research is how to identify the scales at which to switch over in practice.

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).

Finally, we note that, in practical applications, the Huber function $\phi = \phi_{IV}$ is of considerable interest because $\exp(T_N)$ gives rise to the median-type estimator when $-a = b = 0$ with $-a' = b' = 1$ and to the mean-type 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 (16). 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 (16) by a nonparametric (multitaper) estimator for any given h and then solve an optimization problem on a finite grid for a range of values of h.

Acknowledgments

We thank Chris Bretherton, Peter Guttorp and Steve Stigler for discussions. This research was supported by the U.S. National Science Foundation (NSF) under Grant No. DMS 0222115.

Appendix: 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

$$P(x) = \sum_{n=0}^{\infty} c_n P_n(x), \quad \text{where } c_n = \int_{-\infty}^{\infty} P(x)P_n(x)\phi(x)\,dx. \tag{A.17}$$

The function $P(x)$ is said to have Hermite rank ϱ if in the expansion (A.17) 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*

$$\sigma_P^2 = \sum_k s_{P,k} > 0. \tag{A.18}$$

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.18).*

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

$$|\Psi_N(T_N)| \leq |\varphi|/B \text{ a.s.} \tag{A.19}$$

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.19), 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*

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

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

$$\sum a_{2m}^2(2m)! s_{Y,0}^{2m} = \mathrm{E}\left\{1_{x \leq \log Y_i^2 \leq y} - F_Q(x,y)\right\}^2 \leq F_Q(x,y). \tag{A.20}$$

Now

$$\begin{aligned}\mathrm{E}\Big[\sum_{i \in \mathcal{I}} \{1_{x \leq \log Y_i^2 \leq y} - F_Q(x,y)\}\Big]^2 &= \mathrm{var}\Big\{\sum_m a_{2m}(x,y) \sum_i P_{2m}(Y_i)\Big\}^2 \\ &= \sum_m \sum_{m'} a_{2m}(x,y) a_{2m'}(x,y) \sum_i \sum_{i'} \mathrm{cov}\Big(P_{2m}(Y_i), P_{2m'}(Y_{i'})\Big).\end{aligned}$$

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,

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

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.20), we obtain

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

Hence

$$B\mathrm{E}\left\{F_{Q,N}(x,x+y)-F_Q(x,x+y)\right\}^2 = B^{-1}\mathrm{E}\left[\sum_{i\in\mathcal{I}}\{1_{x\le\log Y_i^2\le y}-F_Q(x,y)\}\right]^2$$
$$\le F_Q(x,y)B^{-1}\sum_i\sum_{i'} r^2_{Y,i-i'} \le \text{ constant } F_Q(x,y). \tag{A.21}$$

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

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

Taking expectation we obtain

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

So equation (A.21) yields

$$B\mathrm{E}\left[\int\{\varphi(x-y)-\varphi(x)\}\,d\{F_{Q,N}(x)-F_Q(x)\}\right]^2$$
$$\le\int B\mathrm{E}\left\{F_{Q,N}(x,x+y)-F_Q(x,x+y)\right\}^2 d\varphi(x)$$
$$\le\text{constant}\int F_Q(x,x+y)\,d\varphi(x)$$
$$\le\text{constant}\int y\sup f_Q(z)\,d\varphi(x)\le\text{constant } y,$$

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

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

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 $0 \le y \le B^{-\gamma}$ and φ nondecreasing. We put $y_B = B^{-\gamma}$ and let

$$K = \lfloor \log_2(By_B) \rfloor.$$

We consider a sequence of partitions

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

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

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

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

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

Let

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

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

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

Hence

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

We now apply Cauchy–Schwartz inequality to obtain

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

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

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

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 $k = 0, 1, \ldots, K-1$

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

Hence in view of Lemma 5, we get

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

and similarly

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

Consequently,

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

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

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

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. ■

References

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 Statistical 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.

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.

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

Craigmile, P. F. (2003). Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis* **24**, 505–511.

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

Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis* **18**, 231–251.

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.

Kay, S. M., (1981). Efficient generation of colored noise. *Proceedings of the IEEE* **69**, 480–481.

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.

Maronna, R. A., Martin, D. R. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Chichester, England: John Wiley & Sons, Ltd.

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.

Robinson, P. M. (1986). Discussion: Influence functionals for time series. *The Annals of Statistics* **14**, 832–834.

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

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.

Siigler, S. M. (1977). Do robust estimators work with real data? *Ann. Stats.* **5** , 1055–1098.

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.

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.