1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

WAVELET ANALYSIS FOR SYNCHRONIZATION AND TIMEKEEPING

D. A. Howe
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

D. B. Percival
Applied Physics Laboratory
University of Washington
HN-10
Seattle, WA 98195

Abstract

We discuss the concept of the wavelet variance as a generalized formalism for representing variations in a time series on a scale by scale basis. In particular, we note that the wavelet variance corresponding to some of the recently discovered wavelets can provide a more accurate conversion between the time and frequency domains than can be accomplished using the Allan variance. This increase in accuracy is due to the fact that these wavelet variances give better protection against leakage than does the Allan variance.

I. Introduction and Summary

The analysis of a time-ordered set of phase measurements $\{x_t\}$ often falls into one of three categories. The first approach treats the time series as a function to be expressed in terms of a set of basis functions defined globally over a finite interval (one example of such basis functions are orthogonal polynomials). The purpose here is to summarize the phase measurements with a few coefficients in order to quantify underlying physical effects such as drift.

The second approach takes first or second-order differences of the time series in order to transform nonstationary variations (due to low frequency components) into stationary variations. The mean square of the second difference of phase measurements at various sampling intervals τ quantifies the incremental variability of the phase and for many clocks is stationary. This is the Allan (or pair) variance as a function of τ.

The third approach is the windowed discrete Fourier transform (DFT) which is used to determine the spectral features underlying the time series. A common quantity derived from this transform is the power spectral density.

The problems with the first approach (polynomial fits) are that the coefficients are sensitive to time shifts over the finite interval; many coefficients may be needed to adequately represent the phase measurements; and, because the coefficients are calculated with respect to global basis functions, local features in the phase measurements can be misrepresented. Potential problems with the second approach include sensitivity to deterministic drifts and leakage due to the fact that the transfer function for the finite impulse response (FIR) filter associated with the Allan variance has substantial sidelobes (this leakage is quite similar to that occurring in the unwindowed DFT). The problems with the third approach (the windowed DFT) are that the results can depend on the choice of the particular window and that, because the windowed DFT is inherently narrowband, it is necessarily also highly variable and hence — without further processing — does not summarize the salient features of broadband processes.

Wavelet analysis attempts to address the potential problems with polynomial fits, the Allan variance and spectral analysis in one unified approach. First, wavelet analysis is based upon the discrete wavelet transform, which provides a "time and scale" representation of time-ordered observations. The time series is still treated as a function on a finite interval, but the wavelet basis functions are hierarchical rather than global so that, in contrast to polynomial fits, localized features (such as a step or discontinuity in phase) can be easily represented. Second, the wavelet transform can be implemented using a variety of basis functions and is narrowband at low fre-

0-7803-1945-1/94 $3.00 © 1994 IEEE

quencies and broadband at high frequencies ("multiresolutional," as it is referred to in the wavelet literature). The analyzing function used in the Allan variance when it is computed using fractional frequency deviates is identical to the Haar wavelet, a common starting point in discussions on wavelets. The variances corresponding to wavelets beyond the Haar wavelet are natural extensions to the Allan variance. These wavelet variances have potential advantages over the Allan variance in terms of leakage and insensitivity to deterministic drifts: because the transfer functions for the FIR filters associated with wavelets beyond the Haar wavelet have substantially reduced sidelobes, these wavelet variances have substantially less leakage; and because the FIR filters for higher order wavelets are based in part on differencing operations, a wavelet variance of order n will be invariant to a polynomial drift of order n. Finally, plots of the square root of the wavelet variance versus averaging time (or scale) yield curves that are analogous to the usual "σ/τ" curve for the Allan variance. Users familiar with the Allan variance can thus readily interpret the wavelet variance. In particular, as is true for the Allan variance, the wavelet variance can be regarded as an octave-band estimate of the spectrum and hence does not suffer from the high variability of the windowed DFT. Because higher-order wavelets provide a better approximation to octave-band filters than does the Haar wavelet, it is easier to translate higher-order wavelet variances into reasonable spectral estimates.

Even though wavelets are a relatively new topic, there is already an enormous literature about them — see [4] and references therein. In what follows, we merely attempt to motivate the use of wavelet analysis for synchronization and timekeeping, with particular emphasis on the problem of leakage. Space precludes a full discussion of many important aspects of wavelet analysis (such as the existence of fast computational algorithms [9], procedures for determining confidence intervals for the wavelet variance [8], and the use of the scalogram as a diagnostic tool for monitoring oscillator stability in real time [7]).

II. Power-Law Noise Processes

It has often been claimed that, of all the physical measures, we can realize frequency or periodicity with the greatest accuracy. What we mean is that some basic periodic (repeating) event is very consistent in its recurrence so that, for example, it is independent of environmental influence. This event can thus be used to define consistent and repeatable intervals of time such as the second. Departures from this consistent recurrence are classed as noise. Perfectly recurrent noiseless events mean that events happen now exactly as happened before, that is, an observation now (the present moment) can be perfectly predicted based upon what has happened before (a past moment). Since reproduction is never quite exact, the degree with which present recurrence duplicates past recurrence indicates how well the events remember or duplicate themselves.

When the phase or time difference between two oscillators or clocks is measured as a function of the nominal time of the clocks, what we are measuring is the relative time deviation of the two clocks. Time deviation is generally modeled by two parts:

[1] a deterministic part quantified by a time offset, frequency offset, and frequency drift, and
[2] a random part quantified by various classes of noise processes.

Historically, power-law (or long memory) noise processes have played a vital role in characterizing the performance of clocks. The statement that the relative phase $\{x_t\}$ between a test clock and reference obeys a power-law noise process means that the power spectral density function $S_x(\cdot)$ for the process modelling $\{x_t\}$ is proportional to f^α for positive Fourier frequencies f. Correctly classifying a clock's power-law behavior, which is equivalent to determining α, is a primary objective of analysis techniques such as spectral analysis. Once the exponent α has been determined, we derive estimates of how a clock's timekeeping ability might evolve [11].

The spectrum of the residual time difference between two clocks or oscillators sometimes contains periodicities (such as from 50 or 60 Hz AC power) and always contains nonperiodic (stochastic) characteristics quantified as power-law processes. In timekeeping metrology, there are five commonly used models of power-law noise processes [11]: white PM ($S_x(f) \propto f^0$, a constant), flicker PM ($S_x(f) \propto f^{-1}$), white FM ($S_x(f) \propto f^{-2}$), flicker FM ($S_x(f) \propto f^{-3}$), and random walk FM ($S_x(f) \propto f^{-4}$). Examples of time series drawn from these five processes are shown in Fig. 1.

III. Narrowband vs. Broadband Processing

A common approach to estimating the power spectrum $S_x(\cdot)$ of phase residuals $\{x_t\}$ uses a digital processor to compute a windowed discrete Fourier transform (DFT) of $\{x_t\}$. Here "windowed DFT" refers to multiplying the phase residuals by a data window $\{h_t\}$ (sometimes called a data taper) to produce a windowed series $\{h_t x_t\}$, to which we then apply the DFT. The purpose of windowing is to reduce a potential bias known as leakage, in which power "leaks" from high power into low power portions of the spectrum, thus causing a significant positive bias in unwindowed spectral estimates. There is a

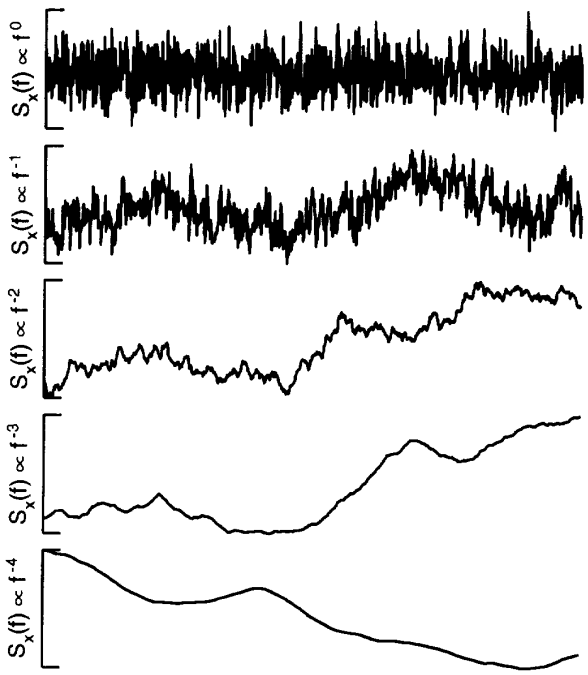

Fig. 1. Examples of time series that are portions of realizations from power-law processes $S_x(f) \propto f^\alpha$, with $\alpha = 0, -1, -2, -3$ and -4 (from top to bottom).

vast literature on the windowed DFT (see [10] for a review). Commercial digital spectrum analyzers typically compute a power spectral density estimate from a sampled varying voltage by converting the voltage to a time series $\{x_t\}$, windowing the series using a user-selected window $\{h_t\}$, and then taking the squared modulus of a properly normalized version of the DFT of $\{h_t x_t\}$. The windowed DFT is inherently narrowband and hence highly variable across frequencies, which makes straightforward interpretation of DFT-based spectral estimates somewhat problematic for the novice (these estimates can also be highly dependent on the choice of a particular window – see [12] for a discussion of this dependence in the case of power-law processes).

Because narrowband processing is not required for broadband processes such as power-law processes, time and frequency standards laboratories have alternatively handled power-law noise processes using the Allan variance [1] or in some cases a modified version of it [2]. These variances can be interpreted as the variance of a process after it has been subjected to an approximate bandpass filter of constant Q (the ratio of the center frequency of the analyzing filter to the width of the filter's pass band is constant [3]). The Allan variance can be used to construct a broadband spectral estimate using well-known conversion schemes [11], [6]. However, while broadband processing produces spectral estimates with inherently less variability than those of narrowband processing, both types of processing are subject to a bias known as leakage. Leakage has long been recognized as an important concern for narrowband spectral estimates (and in fact is the rationale for using windows), but its importance in broadband processing has not received much attention. As we argue below, one rationale for considering the wavelet variance is that higher-order wavelets effectively address the leakage problem.

IV. Wavelets and the "Scale Domain"

Suppose that $x_0, x_1, \ldots, x_{N-1}$ form a sequence $\{x_t\}$ of N time-ordered phase measurements. Let us define

$$\sum_{t=0}^{N-1} |x_t|^2 \equiv \mathcal{E}_x$$

to be the "energy" in our finite set of measurements. We can then trivially regard $|x_t|^2$ as the contribution to the energy $\mathcal{E}_x$ due to the component of $\{x_t\}$ with time index t. We can also regard $\{x_t\}$ as the "time domain" representation of our phase measurements.

Next, consider the discrete Fourier transform (DFT) of $\{x_t\}$, namely,

$$X_k \equiv \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t e^{-i2\pi f_k t}, \qquad k = 0, 1, \ldots, N-1,$$

where X_k is the kth DFT coefficient and is associated with the kth Fourier frequency $f_k \equiv k/N$. Parseval's theorem tells us that

$$\sum_{k=0}^{N-1} |X_k|^2 = \mathcal{E}_x.$$

Hence we can regard $|X_k|^2$ as the contribution to the energy $\mathcal{E}_x$ due to the component of $\{X_k\}$ with frequency index k, and we can regard $\{X_k\}$ as the "frequency domain" representation of our phase measurements. The time and frequency domain representations are equivalent in the sense that we can recover one given the other because of the inverse DFT, namely,

$$x_t = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{i2\pi f_k t}, \qquad t = 0, 1, \ldots, N-1.$$

As is true for the DFT, the discrete wavelet transform (DWT) of $\{x_t\}$ preserves the energy $\mathcal{E}_x$ in a set of

coefficients; however, unlike the DFT, these coefficients are not indexed by frequency, but rather doubly indexed by time shift j and "scale" τ. The DWT is defined in terms of a "mother wavelet" $\psi(\cdot)$ and an associated "scaling function" $\phi(\cdot)$, where $\psi(\cdot)$ can be any member of a large class of functions satisfying certain stringent conditions [4]. Assuming for convenience that $N = 2^p$ for some positive integer p, we define $\psi_{j,\tau}(\cdot)$ as a shifted and scaled version of $\psi(\cdot)$:

$$\psi_{j,\tau}(t) = \frac{1}{\sqrt{2\tau}} \psi\left(\frac{t}{2\tau} - j\tau\right),$$

where $\tau = 1, 2, 4, \ldots, N/2$ indexes a "power of 2" scale, while $j = 0, 2\tau, 4\tau, \ldots, N - 2\tau$ indexes shifts in time commensurate with scale τ. The DWT coefficients are the doubly indexed series $\{d_{j,\tau}\}$ defined by

$$d_{j,\tau} \equiv \sum_t x_t \psi_{j,\tau}(t)$$

along with $c \equiv \sum x_t \phi(t/N)/\sqrt{N}$ (depending upon the precise implementation of the DWT, c is typically proportional to either the average $\sum x_t/N$ of the sequence $\{x_t\}$ or a quantity that converges to the average as N gets large). Parseval's theorem tells us that

$$\sum_\tau \sum_j |d_{j,\tau}|^2 + |c|^2 = \mathcal{E}_x.$$

Hence we can regard $|d_{j,\tau}|^2$ as the contribution to the energy $\mathcal{E}_x$ due to the component of $\{d_{j,\tau}\}$ with time shift index j and scale index τ, and we can regard $\{d_{j,\tau}\}$ as the "scale domain" (or "time/scale" domain) representation of our phase measurements. This scale domain representation is fully equivalent to the time and frequency domain representations because of the inverse DWT, namely,

$$x_t = c + \sum_\tau \sum_j d_{j,\tau} \psi_{j,\tau}(t).$$

As an example of a scale domain representation, let us set our mother wavelet $\psi(\cdot)$ equal to the Haar wavelet $\psi^{(\text{Haar})}(\cdot)$, which we define here as

$$\psi^{(\text{Haar})}(t) = \begin{cases} -1, & 0 \le t < 1/2; \\ 1, & 1/2 \le t < 1; \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding scaling function $\phi(\cdot)$ is given by

$$\phi^{(\text{Haar})}(t) \equiv |\psi^{(\text{Haar})}(t)|$$

(this simple relationship between the mother wavelet and the scaling function is unique to the Haar wavelet). For the Haar wavelet, we find that

$$d_{j,\tau} = \frac{1}{\sqrt{2\tau}} \left(\sum_{l=0}^{\tau-1} x_{(2j+2)\tau-1-l} - \sum_{l=0}^{\tau-1} x_{(2j+1)\tau-1-l} \right)$$
$$= \frac{\sqrt{\tau}}{\sqrt{2}} \left[\overline{x}_{(2j+2)\tau-1}(\tau) - \overline{x}_{(2j+1)\tau-1}(\tau) \right],$$

where $\overline{x}_t(\tau) \equiv \sum_{j=0}^{\tau-1} x_{t-j}/\tau$. Thus, at scale $\tau = 1$, we have $d_{j,1} = (x_{2j+1} - x_{2j})/\sqrt{2}$ for $0 \le j \le N/2 - 1$, while, at the largest scale $\tau = N/2$, we have the single coefficient

$$d_{0,N/2} = (x_{N-1} + \cdots + x_{N/2} - x_{N/2-1} - \cdots - x_0)/(\sqrt{2})^p.$$

Let us now define the wavelet variance for scale τ as

$$\sigma_x^2(\tau) \equiv \text{var}\{d_{j,\tau}\}/\tau.$$

Under the assumption that $E\{d_{j,\tau}\} = 0$ so that the variance of $d_{j,\tau}$ is equal to $E\{d_{j,\tau}^2\}$, an obvious estimator of this wavelet variance is

$$\tilde{\sigma}_x^2(\tau) = \frac{1}{\tau} \times \frac{1}{N/2\tau} \sum_{j=0}^{N/2\tau-1} d_{j,\tau}^2 = \frac{2}{N} \sum_{j=0}^{N/2\tau-1} d_{j,\tau}^2.$$

Specializing now to the case of the Haar wavelet, we find that

$$\tilde{\sigma}_x^2(\tau) = \frac{\tau}{N} \sum_{j=0}^{N/2\tau-1} \left[\overline{x}_{(2j+2)\tau-1}(\tau) - \overline{x}_{(2j+1)\tau-1}(\tau) \right]^2.$$

If the x_t's represented average fractional frequency deviations rather than phase measurements, then the above would be the well-known "nonoverlapped" estimator of the Allan variance. The Allan variance therefore corresponds to a wavelet variance when the Haar wavelet is used with average fractional frequency deviations. When viewed from the perspective of wavelets, the Allan variance is thus not a "time domain" quantity, but rather is a "scale domain" or "time/scale domain" quantity.

V. Determination of Power-Law Noise Types

As a function of time, two-oscillator phase deviations might look like one of the plots of a realization of a pure power-law process shown in Fig. 1. More realistically, these deviations resemble a linear combination of such processes, whose spectrum can be described mathematically as

$$S_x(f) = \sum_\alpha h_\alpha |f|^\alpha,$$

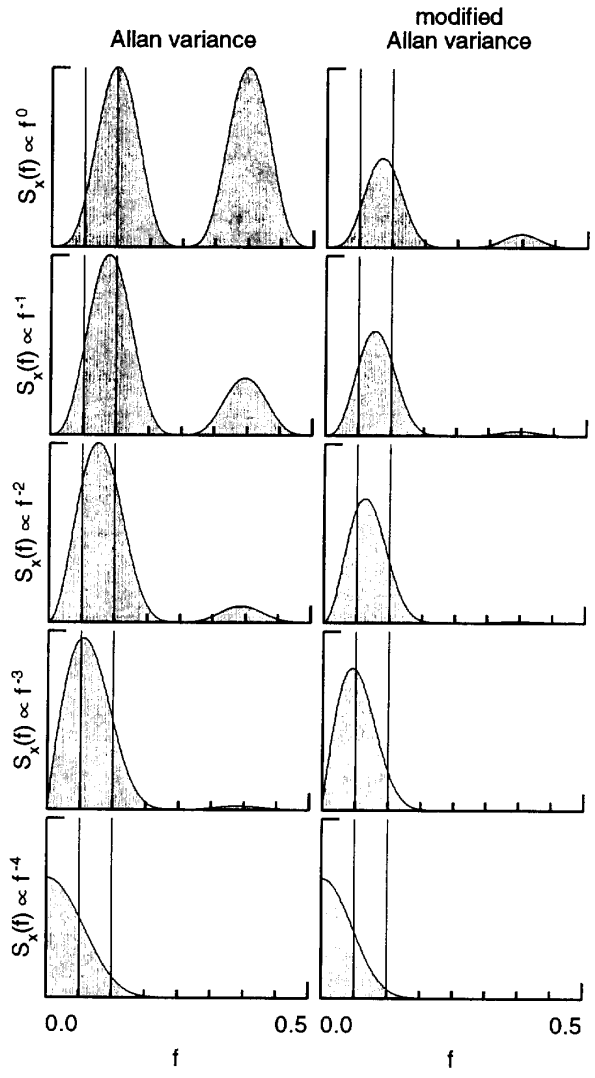

Fig. 2. Modulus squared of the transfer function for the Allan variance (left-hand column) and the modified Allan variance (right-hand column) times power-law spectra $S_x(\cdot)$ proportional to f^0 (top row), f^{-1} (second row), f^{-2} (third row), f^{-3} (fourth row) and f^{-4} (bottom row) for scale $\tau = 4$. The integrals of the shaded areas yield the Allan variance or modified Allan variance for scale $\tau = 4$.

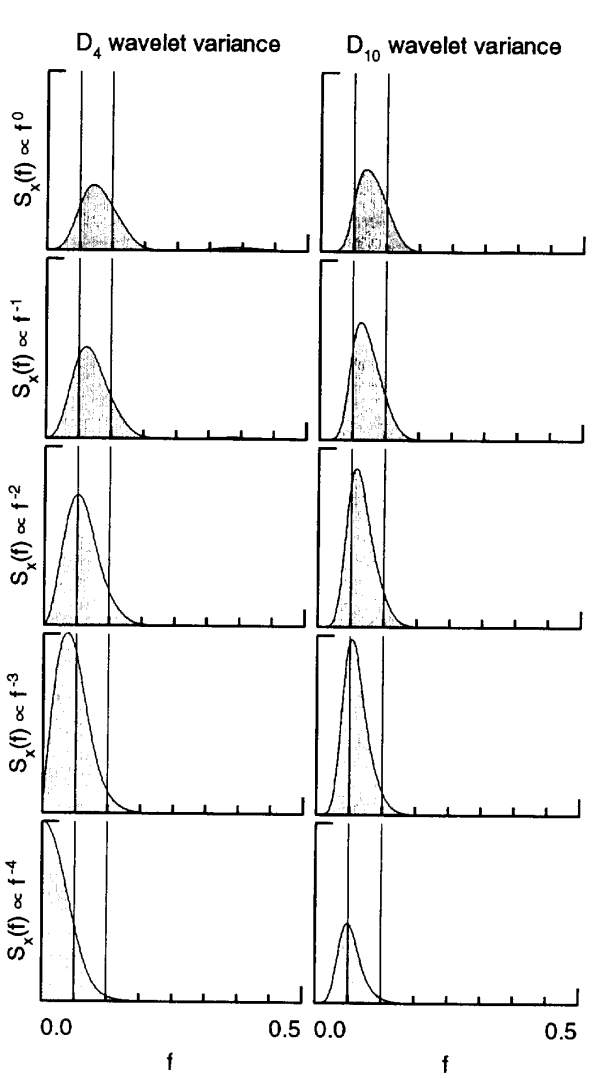

Fig. 3. Modulus squared of the transfer function for the D_4 wavelet variance (left-hand column) and the D_{10} wavelet variance (right-hand column) times the same set of power-law spectra shown in Fig. 2. The integrals of the shaded areas yield the D_4 or D_{10} wavelet variance for scale $\tau = 4$.

where the summation is over a finite number of different α's (usually a subset of $\alpha = 0, -1, -2, -3$ and -4), with h_α determining the relative contribution of the power-law process with exponent α. We refer to a process with the above spectrum as a composite power-law process. For pure power-law processes, there are well-known formulae for converting from the Allan variance to the frequency domain [11]. Here we argue that, for composite power-law processes, this conversion can become problematic for both the usual Allan variance and

the modified Allan variance. Let $\sigma_x^2(\tau)$ represent either of these variances (this notation should not be confused with similar notation for "TVAR" in [3]). We can then write

$$\sigma_x^2(\tau) = \int_{-1/2}^{1/2} \mathcal{F}_\tau(f) S_x(f)\, df,$$

where $\mathcal{F}_\tau(\cdot)$ is the modulus squared of the appropriate transfer function for the filters associated with these variances at scale τ [5]. The shaded areas in the plots of Fig. 2 show the product $\mathcal{F}_\tau(f)S_x(f)$ versus f for the Allan variance (left-hand column) and the modified Allan variance (right-hand column) for five pure power-law spectra and scale $\tau = 4$ (the power-law spectrum is constant in the top row of plots, so this row really shows just $\mathcal{F}_\tau(f)$ versus f). The integral of each shaded area gives $\sigma_x^2(4)$ for the appropriate pure power-law process. In the octave-band interpretation of these variances, either variance at scale τ should roughly reflect the power in the spectrum in the frequency interval $[1/4\tau, 1/2\tau]$. For $\tau = 4$, this interval is $[1/16, 1/8]$ and is delineated on each plot by a pair of thin vertical lines. If the filters associated with these variances were perfect octave-band filters, the shaded area in each plot would be entirely contained between the vertical lines. The amount of the shaded area that lies outside of the vertical lines represents the contribution to the Allan or modified Allan variance attributable to leakage. These plots indicate that there is substantial leakage for the Allan variance, but less so for the modified Allan variance. Leakage is most pronounced in the Allan variance for white PM ($S_x(f) \propto f^0$ in the top left-hand plot), a deficiency that in fact accounts for the development of the modified Allan variance [5]. If we now consider a composite power-law process dominated between the vertical lines by a power-law with a different exponent than the one displayed in the plots of Fig. 2, we can see the potential problem with leakage, namely, that the integral of $\mathcal{F}_\tau(f)S_x(f)$ (the Allan or modified Allan variance) can be influenced mainly by values of f outside of the vertical lines and hence cannot accurately reflect the values of $S_x(f)$ between the vertical lines.

Fig. 3 shows corresponding plots for the wavelet variance using the D_4 (left-hand column) and D_{10} (right-hand column) "extremal phase" wavelets [4]. The D_4 wavelet was chosen because it is "one order up" from the Haar wavelet (and in fact rather closely mimics the behavior of the modified Allan variance), while the D_{10} wavelet is an example of a higher-order wavelet. The main point to notice here is that the D_{10} wavelet variance for scale $\tau = 4$ reflects the spectrum in the passband $[1/16, 1/8]$ to a much better degree than the other variances because the shaded areas are concentrated between the vertical lines to a higher degree for the D_{10} wavelet variance than for the Allan or modified Allan variances.

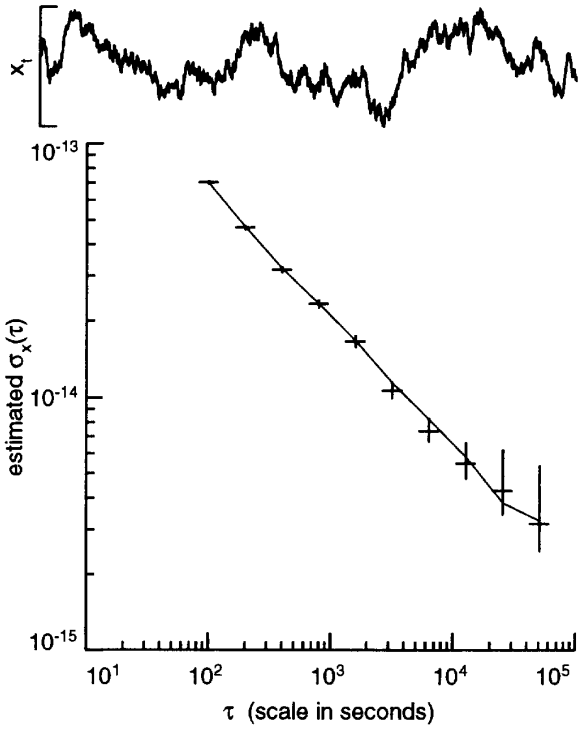

Fig. 4. NIST–7 vs. hydrogen maser phase measurements (top plot) and estimated $\sigma_x(\tau)$ versus τ (bottom plot) for the Allan variance (connected curve) and the D_4 wavelet variance (crosses).

VI. Examples

We present two examples of the limited tests we have made to date using the wavelet variance with phase measurements (see [9] for an example involving geophysical data, for which the D_8 wavelet variance performed considerably better than the Allan variance). The top plot of Fig. 4 shows phase measurements recorded every 100 seconds over a 3.7 day interval comparing NIST–7 to a hydrogen maser. The bottom plot shows the estimated Allan standard deviation (which is just the square root of the Allan variance) versus scale τ (the connected curve) and also the estimated D_4 wavelet standard deviation versus τ (the crosses). The center of each cross indicates the appropriate D_4 estimate, whereas the vertical portion of the cross delineates a "one sigma" (68.3%) confidence interval for the true D_4 wavelet standard deviation [8]. The Allan and D_4 wavelet standard deviations agree fairly well here, although there are two scales

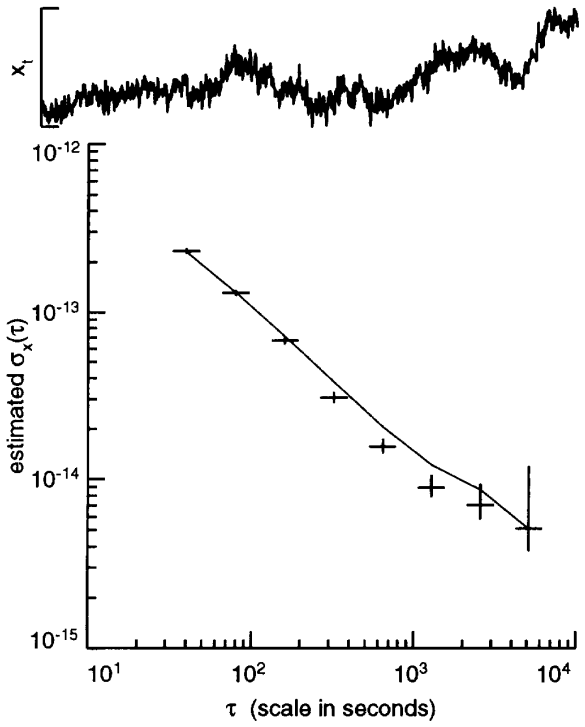

Fig. 5. Time-synchronization phase measurements using the NIST satellite two-way transfer modem configured in an in-cabinet loop test (top plot) and estimated $\sigma_x(\tau)$ versus τ (bottom plot) for the Allan variance (connected curve) and the D_4 wavelet variance (crosses).

(τ = 3200 and 6400 seconds) for which the Allan standard deviation is just inside the "one sigma" confidence limits for the D_4 wavelet standard deviation. Use of the D_4 wavelet here tells us that leakage is not a major problem with the Allan variance for this set of phase measurements.

The top plot of Fig. 5 shows phase measurements recorded every 40 seconds over a half day interval reflecting time-synchronization using the NIST satellite two-way transfer modem configured in an in-cabinet loop test. The bottom plot here shows the same quantities as in the bottom plot of Fig. 4. While the Allan and D_4 standard deviations agree quite well in the smallest three and largest scales shown, there is significant difference in the middle three scales; moreover, the difference is consistent with an interpretation of leakage in the Allan variance (because the Allan variance is higher than the D_4 variance). Use of the higher-order D_6 wavelet yields good agreement with the D_4 wavelet.

References

[1] D.W. Allan, "Statistics of Atomic Frequency Standards," *Proc. IEEE*, vol. 54, pp. 221–230, 1966.

[2] D.W. Allan and J.A. Barnes, "A Modified 'Allan Variance' with Increased Oscillator Characterizaton Ability," *Proc. 35th Ann. Frequency Control Symp.*, pp. 470–475, 1981 (reprinted in [11]).

[3] D.W. Allan, M.A. Weiss, and J.L. Jespersen, "A Frequency-Domain View of Time-Domain Characterizaton of Clocks and Time and Frequency Distribution System," *Proc. 45th Ann. Frequency Control Symp.*, pp. 667–678, 1991.

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

[5] C.A.Greenhall,"A Shortcut for Computing the Modified Allan Variance," *Proc. 46th Ann. Frequency Control Symp.*, pp. 262–264, 1992.

[6] D.B. Percival, "Characterization of Frequency Stability: Frequency-Domain Estimation of Stability Measures," *Proc. IEEE*, vol. 79, pp. 961–972, 1991.

[7] D.B. Percival, "An Introduction to Spectral Analysis and Wavelets," in *Advanced Mathematical Tools in Metrology*, (Edits P. Ciarlini, M. Cox, R. Monaco and F. Pavese). Singapore: World Scientific, pp. 175—186, 1994.

[8] D.B. Percival, "On Estimation of the Wavelet Variance," submitted to *Biometrika*, 1994.

[9] D.B. Percival and P. Guttorp, "Long Memory Processes, the Allan Variance and Wavelets," in *Wavelets in Geophysics*, (Edits E. Foufoula-Georgiou and P. Kumar). Boston: Academic Press, pp. 325–343, 1994.

[10] D.B. Percival and A.T. Walden, *Spectral Analysis for Physical Applications.* Cambridge, U.K.: Cambridge Univ. Press, 1993.

[11] D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls (Edits), *Characterization of Clocks and Oscillators*, NIST Technical Note 1337, 1990 (available from NIST, 325 Broadway, Boulder, CO 80303–3328).

[12] F.L. Walls, D.B. Percival, and W.R. Ireland, "Biases and Variances of Several FFT Spectral Estimators as a Function of Noise Type and Number of Samples," *Proc. 43rd Ann. Frequency Control Symp.*, pp. 336–341, 1989 (reprinted in [11]).