This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

A Wavelet Perspective on the Allan Variance

Donald B. Percival

Abstract

The origins of the Allan variance trace back fifty years ago to two seminal papers, one by D.W. Allan (1966) and, and the other by J.A. Barnes (1966). The Allan variance has since then played a leading role in the characterization of high-performance time and frequency standards. Wavelets first arose in the early 1980s in the geophysical literature, and the discrete wavelet transform (DWT) became prominent in the late 1980s in the signal processing literature. Flandrin (1992) briefly documented a connection between the Allan variance and a wavelet transform based upon the Haar wavelet. Percival and Guttorp (1994) noted that one popular estimator of the Allan variance – the maximal overlap estimator – can be interpreted in terms of a version of the DWT now widely referred to as the maximal overlap DWT or MODWT. In particular, when the MODWT is based on the Haar wavelet, the variance of the resulting wavelet coefficients – the wavelet variance – is identical to the Allan variance when the latter is multiplied by one-half. The theory behind the wavelet variance can thus deepen our understanding of the Allan variance. In this paper we review basic wavelet variance theory with an emphasis on the Haar-based wavelet variance and its connection to the Allan variance. We then note that estimation theory for the wavelet variance offers a means of constructing asymptotically correct confidence intervals for the Allan variance without reverting to the common practice of specifying a power-law noise type *a priori*. We also review recent work on specialized estimators of the wavelet variance that are of interest when some observations are missing (gappy data) or in the presence of contamination (rogue observations or outliers). It is a simple matter to adapt these estimators to become estimators of the Allan variance. Finally we note that wavelet variances based upon wavelets other than the Haar offer interesting generalizations of the Allan variance.

I. Introduction

The origins of the Allan variance can be traced back to two seminal papers fifty years ago (Allan, 1966, and Barnes, 1966). The Allan variance was proposed to tackle the problem of quantifying how well high-quality time and frequency standards perform and has been a mainstay in the analysis of clock data since its introduction. The Allan variance has also been used in a wide range of other applications, including rain-attenuated millimeter wave beacon signals from a geostationary satellite (Yoshimura *et al.*, 1983), wave arrival detection in acoustic well logs (Day, 1992) and inertial sensors (Sabatini, 2006, and El-Sheimy *et al.*, 2008). Guerrier *et al.* (2013) also used the Allan variance in the context of inertial sensors and pointed out a number of other interesting applications, including human muscle sympathetic nerve activity (Fadel *et al.*, 2004; Gebber *et al.*, 2006), the rotation of the earth

D.B. Percival is with the Applied Physics Laboratory and the Department of Statistics, University of Washington, Seattle, WA, 98195-5640 USA e-mail: dbp@uw.edu.

Manuscript received May 31, 2015; revised ???.

(Gambis, 2002), cavity attenuated phase shift spectroscopy (Kebabian *et al.*, 2005), sonic anemometerthermometer measurements of temperature and wind speed (Loescher *et al.*, 2005), characterization of the stability of equipment used in radio astronomy (Schieder and Kramer, 2001) and tunable diode-laser absorption spectroscopy (Werle *et al.*, 1993).

Three themes stand out in Allan (1966) and Barnes (1966). The first is the difficulty of using the sample variance as a meaningful measure of clock performance. Clock data in the form of measurements recorded over time often exhibit significant trend-like components (i.e., low-frequency variations). The sample variance is problematic for such data because it can depend strongly on the amount of data available. This fact motivated interest in the mean square successive difference, a precursor of the Allan variance that the celebrated mathematician John von Neumann and co-authors studied in three papers in the 1940s (von Neumann *et al.*, 1941; von Neumann, 1941; von Neumann, 1942). The second theme is the relationship between time domain measures such as the Allan variance and the frequency domain measure known as the spectrum (or spectral density function). A precursor here is the pilot spectrum estimate advocated in Section 18 of Blackman and Tukey's influential 1958 book on spectrum estimation (Blackman and Tukey, 1958). Table III in that section illustrates the computation of pilot estimates, which involves exactly the same computations needed to form one estimator of the Allan variance (and there is a brief nod toward the popular maximal overlap estimator). The final – and most important – theme is the connection between the Allan variance and power-law processes, i.e., processes possessing a spectrum that is proportional to f^{α}, where $f > 0$ is a Fourier frequency. The Allan variance offers a conceptually easy identification of when certain power-law processes might be is a reasonable approximation to the process generating the data. This unique contribution of Allan (1966) and Barnes (1966) has been the focus of much of the subsequent interest in the Allan variance.

The subject of wavelets emerged well after the introduction of the Allan variance. The early 1980s saw the development of the continuos wavelet transform (CWT) mainly in the geophysical literature, while, by the latter part of that decade, the focus of research – particularly in the engineering literature – shifted to the discrete wavelet transform (DWT). There are versions of both the CWT and DWT with intellectual roots in a 1910 paper by the Hungarian mathematician Alfréd Haar (Haar, 1910). Flandrin (1992) briefly pointed out the connection between the Allan variance and Haar wavelet transforms. Percival and Guttorp (1994) noted that one popular estimator of the Allan variance – the maximal overlap estimator – arises from a version of the DWT now widely referred to as the maximal overlap DWT or MODWT. In particular, when the MODWT is based on the Haar wavelet, the variance of the resulting wavelet coefficients – the wavelet variance – is identical to the Allan variance when the latter is multiplied by one-half. The connection between the Allan variance and the Haar wavelet variance is of interest because it ties the former to a well-defined transform, which leads to certain insights not otherwise readily apparent. For example, it transpires that one estimator of the Allan variance is actually a 'slicing and dicing' of the much-maligned sample variance; another insight is that the Allan variance is the simplest possible wavelet variance and that wavelet variances of greater complexity can help assess when the Allan variance is providing a meaningful analysis.

This paper provides a basic introduction to the theory behind wavelets, with an emphasis on the Haar wavelet

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

and its connection to the Allan variance. We start with an overview of the Haar MODWT (Section II). The Haar MODWT leads to a basic analysis of variance of a time series (Section III) and to the definition of a wavelet variance that is proportional to the Allan variance (Section IV). We discuss the statistical theory behind estimators of the Allan and Haar wavelet variances in Section V. While some of the Allan variance estimators are well-known and predate the corresponding wavelet variance estimators, we also review more recent specialized estimators that are less known and have yet to be used in conjunction with the Allan variance. Section VI explores wavelet variances based on wavelets more complex than the Haar. We devote the penultimate section (VII) to two real-world examples illustrating the methodology discussed in previous sections. Section VIII has some concluding remarks.

II. The Haar maximal overlap discrete wavelet transform

In preparation for exploring the relationship between the Allan variance and wavelets, here we formulate the maximal overlap discrete wavelet transform (MODWT), which is one particular version of the wavelet transform. This transform takes a time series $\{X_t : t = 0, \ldots, N-1\}$[1] of length N and produces $J_0 + 1$ new series (each of length N), where typically J_0 is set to an integer less than or equal to $\log_2(N)$. We denote these new series by $\{\widetilde{W}_{1,t}\}$, $\{\widetilde{W}_{2,t}\}$, …, $\{\widetilde{W}_{J_0,t}\}$ and $\{\widetilde{V}_{J_0,t}\}$. The first J_0 of these series are wavelet coefficients and are indexed by $j = 1, 2, \ldots, J_0$, while the final series consists of scaling coefficients and is associated with the index J_0. We create these $J_0 + 1$ series through the use of two filters, a wavelet filter and a complementary scaling filter. The particular filters that lead us to the Allan variance are the Haar wavelet and scaling filters, which are the simplest filters in the Daubechies class of filters (Daubechies, 1992). In what follows, we describe the MODWT in terms of the Haar filters, but key properties of the MODWT continue to hold when we use filters other than the Haar (see Section VI; for more details here and else, we refer the reader to Percival and Walden, 2000, which uses notation consistent with what we adopt in this paper).

The Haar wavelet filter $\{\tilde{h}_{1,l}\}$ and the Haar scaling filter $\{\tilde{g}_{1,l}\}$ are given by

$$\tilde{h}_{1,l} = \begin{cases} 1/2, & l = 0; \\ -1/2 & l = 1; \\ 0 & \text{for all other integers } l; \end{cases} \quad \text{and} \quad \tilde{g}_{1,l} = \begin{cases} 1/2, & l = 0 \text{ or } l = 1; \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Both filters have a width of $L_1 = 2$. These two filters are all we need to define the Haar MODWT of the time series $\{X_t\}$. We begin by using them to create wavelet and scaling coefficients with index $j = 1$:

$$\widetilde{W}_{1,t} = \sum_{l=0}^{L_1-1} \tilde{h}_{1,l} X_{t-l \bmod N} \quad \text{and} \quad \widetilde{V}_{1,t} = \sum_{l=0}^{L_1-1} \tilde{g}_{1,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1, \tag{2}$$

[1] The usual convention in the clock literature is to let X_t stand for a time difference or phase measurement, whereas here X_t is taken to be any measurement recorded at a time indexed by t, as commonly appears in the literature on time series analysis and wavelets. Later on, we will consider fractional frequency deviates as a particular set of measurements recorded over time, and, in violation of the clock literature convention, we will let X_t to refer to these measurements.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

which make use of the 'mod' operator, defined as follows. If $t-l$ is such that $0 \le t-l \le N-1$, then '$t-l \bmod N$' is equal to $t-l$; otherwise, it is equal to $t-l+nN$, where n is the unique integer such that $0 \le t-l+nN \le N-1$. Hence we have

$$\widetilde{W}_{1,0} = \frac{X_0 - X_{N-1}}{2} \text{ while } \widetilde{W}_{1,t} = \frac{X_t - X_{t-1}}{2} \text{ for } t = 1, 2, \ldots, N-1,$$

and

$$\widetilde{V}_{1,0} = \frac{X_0 + X_{N-1}}{2} \text{ while } \widetilde{V}_{1,t} = \frac{X_t + X_{t-1}}{2} \text{ for } t = 1, 2, \ldots, N-1.$$

Note that, for $t \ge 1$, the coefficients $\widetilde{W}_{1,t}$ and $\widetilde{V}_{1,t}$ involve adjacent values of the time series, whereas (assuming $N \ge 3$) the remaining coefficients $\widetilde{W}_{1,0}$ and $\widetilde{V}_{1,0}$ involve the nonadjacent values X_0 and X_{N-1}. We refer to these two anomalous coefficients henceforth as boundary coefficients. Note also that the non-boundary wavelet coefficients are proportional to the first difference of the time series, whereas the non-boundary scaling coefficients are two-point averages.

The set of N wavelets coefficients $\{\widetilde{W}_{1,t}\}$ and the corresponding set of scaling coefficients $\{\widetilde{V}_{1,t}\}$ collectively constitute the level $J_0 = 1$ MODWT of $\{X_t\}$. Typically we are interested in MODWTs of higher levels. The wavelet coefficients indexed by $j = 1$ are one part of a MODWT of level $J_0 \ge 2$, but the scaling coefficients $\{\widetilde{V}_{1,t}\}$ are not. The J_0 remaining parts of the transform are $\{\widetilde{W}_{2,t}\}$, …, $\{\widetilde{W}_{J_0,t}\}$ and $\{\widetilde{V}_{J_0,t}\}$. There are two complementary ways to get these remaining parts. The first way uses higher-level wavelet filters $\{\tilde{h}_{j,l}\}$, $j = 2, \ldots, J_0$, along with a level J_0 scaling filter $\{\tilde{g}_{J_0,l}\}$. Letting $L_j = 2^j$, these filters are given by

$$\tilde{h}_{j,l} = \begin{cases} 1/2^j, & l = 0, \ldots, L_j/2 - 1; \\ -1/2^j, & l = L_j/2, \ldots, L_j - 1; \\ 0 & \text{otherwise;} \end{cases} \text{ and } \tilde{g}_{J_0,l} = \begin{cases} 1/2^{J_0}, & l = 0, \ldots, L_{J_0} - 1; \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

(note that setting $j = J_0 = 1$ in the above yields the Haar filters of Equation (1)). For example, the non-zero elements of $\{\tilde{h}_{j,l}\}$ for $j = 2$ are $\{\tilde{h}_{2,0}, \tilde{h}_{2,1}, \tilde{h}_{2,3}, \tilde{h}_{2,4}\} = \{\frac{1}{4}, \frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}\}$ (see the middle panel of the top row of the forthcoming Figure 2). These filters satisfy three fundamental properties:

$$\sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0, \quad \sum_{l=0}^{L_j-1} \tilde{h}_{j,l}^2 = \tfrac{1}{2^j} \text{ and } \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} \tilde{h}_{j,l+2^j n} = 0 \text{ for all nonzero integers } n. \tag{4}$$

Using these filters, we obtain the required remaining J_0 parts:

$$\widetilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N} \text{ for } j = 2, \ldots, J_0 \text{ and } \widetilde{V}_{J_0,t} = \sum_{l=0}^{L_{J_0}-1} \tilde{g}_{J_0,l} X_{t-l \bmod N}, \tag{5}$$

where $t = 0, 1, \ldots, N-1$. We note that $\{\widetilde{W}_{j,t}\}$ has $L_j - 1$ boundary coefficients (namely, $\widetilde{W}_{j,0}, \ldots, \widetilde{W}_{j,L_j-2}$) whereas $\{\widetilde{V}_{J_0,t}\}$ has $L_{J_0} - 1$ (a technical note: these statements hold if $L_{J_0} < N$, which is almost always the case in practice; if not, then there are $\min(L_j - 1, N)$ boundary wavelet coefficients, and all N scaling coefficients qualify as boundary coefficients). Note that we can express the non-boundary wavelet and scaling coefficients as

$$\widetilde{W}_{j,t} = \frac{1}{L_j} \sum_{l=0}^{L_{j-1}-1} X_{t-l} - \frac{1}{L_j} \sum_{l=L_{j-1}}^{2L_{j-1}-1} X_{t-l} \text{ and } \widetilde{V}_{J_0,t} = \frac{1}{L_{J_0}} \sum_{l=0}^{L_{J_0}-1} X_{t-l}.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

Thus the non-boundary wavelet coefficients are proportional to the difference of adjacent 2^{j-1}-point averages of the time series, whereas the non-boundary scaling coefficients are 2^{J_0}-point averages.

The second way to get exactly the same J_0 remaining parts is a computationally efficient procedure known as the pyramid algorithm. This recursive procedure makes it clear that we really just need the unit level wavelet and scaling filters $\{\tilde{h}_{1,l}\}$ and $\{\tilde{g}_{1,l}\}$ to form the Haar MODWT of arbitrary level J_0. Assume that, for some $j \geq 2$, we have already computed the scaling coefficients $\{\widetilde{V}_{j-1,t}\}$. We can then form the wavelet and scaling coefficients indexed by j via

$$\widetilde{W}_{j,t} = \sum_{l=0}^{L_1-1} \tilde{h}_{1,l} \widetilde{V}_{j-1,t-2^{j-1}l \bmod N} \text{ and } \widetilde{V}_{j,t} = \sum_{l=0}^{L_1-1} \tilde{g}_{1,l} \widetilde{V}_{j-1,t-2^{j-1}l \bmod N}, \quad t = 0, 1, \ldots, N-1. \tag{6}$$

The scheme is to use Equation (2) to get the wavelet and scaling coefficients for $j = 1$ and then to apply the above equations recursively for $j = 2, \ldots, J_0$, thus obtaining the $J_0 + 1$ parts constituting the Haar MODWT of level J_0 (in the process, we also form the series $\{\widetilde{V}_{j,t}\}$, $j = 1, \ldots, J_0 - 1$, but these are not part of the level J_0 MODWT). Note that, if we define $\widetilde{V}_{0,t} = X_t$, then setting $j = 1$ in Equation (6) yields Equation (2). (For pseudo-code describing this algorithm, see pp. 177–8 of Percival and Walden, 2000).

III. Analysis of variance via the Haar MODWT

Let $\mathbf{X}$ be an N-dimensional vector whose elements are $X_0, \ldots, X_{N-1}$, and let $\widetilde{\mathbf{W}}_j$ and $\widetilde{\mathbf{V}}_j$ be similar vectors with elements $\{\widetilde{W}_{j,t}\}$ and $\{\widetilde{V}_{j,t}\}$. Denote

$$\|\mathbf{X}\|^2 = \sum_{t=0}^{N-1} X_t^2$$

for the square of the Euclidean norm of $\mathbf{X}$. A key point about the MODWT is that it is norm preserving, in the sense that

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2. \tag{7}$$

This reduces to $\|\mathbf{X}\|^2 = \|\widetilde{\mathbf{W}}_1\|^2 + \|\widetilde{\mathbf{V}}_1\|^2$ when $J_0 = 1$, which we can readily verify in the Haar case by considering

$$\widetilde{W}_{1,t}^2 + \widetilde{V}_{1,t}^2 = \frac{(X_t + X_{t-1})^2}{4} + \frac{(X_t - X_{t-1})^2}{4} = \frac{X_t^2 + X_{t-1}^2}{2}, \quad t = 1, \ldots, N-1,$$

along with a similar piece involving the boundary coefficients $\widetilde{W}_{1,0}$ and $\widetilde{V}_{1,0}$. Let $\overline{X} = \sum_{t=0}^{N-1} X_t/N$ represent the sample mean of $\mathbf{X}$. We can express the sample variance of our time series as

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{t=0}^{N-1} \left(X_t - \overline{X}\right)^2 \tag{8}$$

$$= \frac{1}{N}\|\mathbf{X}\|^2 - \overline{X}^2 = \frac{1}{N}\left(\sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2\right) - \overline{X}^2 = \sum_{j=1}^{J_0} \hat{\sigma}^2_{\widetilde{W}_j} + \hat{\sigma}^2_{\widetilde{V}_{J_0}}, \tag{9}$$

where

$$\hat{\sigma}^2_{\widetilde{W}_j} = \frac{1}{N}\|\widetilde{\mathbf{W}}_j\|^2 \text{ and } \hat{\sigma}^2_{\widetilde{V}_{J_0}} = \frac{1}{N}\|\widetilde{\mathbf{V}}_{J_0}\|^2 - \overline{X}^2. \tag{10}$$

The following argument says that we can interpret $\hat{\sigma}^2_{\widetilde{W}_j}$ and $\hat{\sigma}^2_{\widetilde{V}_{J_0}}$ as sample variances for $\widetilde{\mathbf{W}}_j$ and $\widetilde{\mathbf{V}}_{J_0}$. The definition of the level j wavelet filter ensures that, if the X_t variables are from a population with a common finite

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

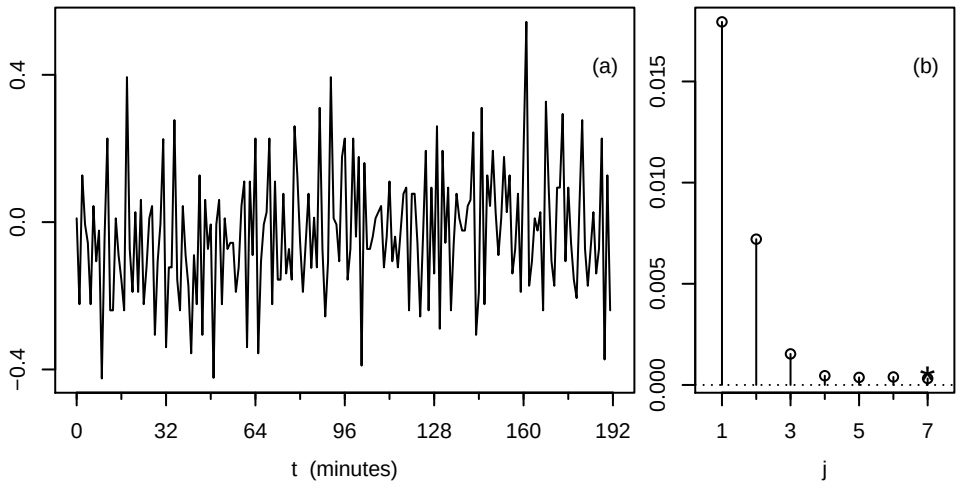

Fig. 1. Time series of rescaled fractional frequency deviates (a) and the Haar empirical wavelet variances (b) at levels $j = 1, \ldots, 7$ (circles) and empirical scaling variance for level 7 (asterisk). Since the sampling time between values in the time series is $\Delta_t = 1$ minute, the jth level wavelet variance of corresponds to changes on a scale of $\tau_j = 2^{j-1}\,\Delta_t$ minutes, while the single scaling variance is associated with averages on a scale of $\lambda_7 = 2^7\,\Delta_t = 128$ minutes The peak at level $j = 1$ in the wavelet variances is thus associated with changes on a scale of $\tau_1 = 1$ minute. (The MODWT behind the calculations shown here used reflection boundary conditions see Section V-B for details.)

mean, then, under mild conditions, the corresponding population mean of the wavelet coefficients is zero. Rather than using the sample mean $\overline{W}_j$ of $\widetilde{\mathbf{W}}_j$ to formulate the sample variance as $\|\widetilde{\mathbf{W}}_j\|^2/N - \overline{W}_j^2$, we can dispense with $\overline{W}_j$ in favor of the population mean of zero. On the other hand, the sample mean of $\widetilde{\mathbf{V}}_{J_0}$ is always $\overline{X}$. To see this in the Haar case, note that

$$\frac{1}{N}\sum_{t=0}^{N-1}\widetilde{V}_{J_0,t} = \frac{1}{N}\sum_{t=0}^{N-1}\frac{1}{L_{J_0}}\sum_{l=0}^{L_{J_0}-1}X_{t-l \bmod N} = \frac{1}{L_{J_0}}\sum_{l=0}^{L_{J_0}-1}\frac{1}{N}\sum_{t=0}^{N-1}X_{t-l \bmod N} = \frac{1}{L_{J_0}}\sum_{l=0}^{L_{J_0}-1}\overline{X} = \overline{X},$$

as claimed. We refer to $\hat{\sigma}^2_{\widetilde{W}_j}$ as the jth level empirical wavelet variance and to $\hat{\sigma}^2_{\widetilde{V}_{J_0}}$ as the level J_0 empirical scaling variance. Equation (9) implies that, as J_0 increases, the scaling variance cannot increase and that $\hat{\sigma}^2_{\widetilde{V}_{J_0}}/\hat{\sigma}^2_X$ represents the proportion of the sample variance that is *not* attributable to wavelet variances at scales indexed by $j = 1, \ldots, J_0$.

In summary the MODWT gives us an analysis of variance (ANOVA) for $\mathbf{X}$, in which we break up the sample variance $\hat{\sigma}^2_X$ into J_0+1 parts. The first J_0 parts are based on the wavelet coefficients are associated with differences in averages over scales $\tau_j = 2^{j-1}$, $j = 1, \ldots, J_0$. The final part is based on the level J_0 scaling coefficients and is associated with averages over a scale of $\lambda_{J_0} = 2^{J_0}$. (Note that τ_j and λ_{J_0} are taken to be standardized unitless scales. If the actual times at which X_{t-1} and X_t were recorded differ by Δ_t, the corresponding physically meaningful scales are $\tau_j\,\Delta_t$ and $\lambda_{J_0}\,\Delta_t$.)

As a simple example of a Haar-based ANOVA, consider Figure 1(a), which shows the first $N = 192$ values of a time series of rescaled fractional frequency deviates described in detail in Section VII-A. Figure 1(b) shows the

empirical wavelet variances $\hat{\sigma}^2_{\widetilde{W}_j}$ (circles) and empirical scaling variance $\hat{\sigma}^2_{\widetilde{V}_{J_0}}$ (asterisk) based upon a level $J_0 = 7$ Haar MODWT. The sum of these eight variances is exactly equal to the sample variance $\hat{\sigma}^2_X \doteq 0.02893$ of the time series. The largest wavelet variance occurs at level $j = 1$, which corresponds to scale $\tau_1 = 1$ minute. The peak at this scale quantifies what a visual inspection picks out, namely, a tendency of the series to fluctuate back and forth from one point to the next (the sample autocorrelation sequence at unit lag is -0.245) The fact that the scaling variance at level $J_0 = 7$ is small relative to the sum of the wavelet variances tells us that the bulk of the variance of the time series can be attributed to changes in averages over scales $\tau_7 = 2^6 = 64$ minutes and smaller. The scale-based ANOVA given by the wavelet variance offers an intuitively sensible explanation of how a time series is structured.

IV. The Haar wavelet variance and the Allan variance

In this section we define the Haar wavelet variance and the Allan variance and indicate their connection. To do so, consider a stochastic process $\{X_t : t \in \mathbb{Z}\}$, i.e., a collection of random variables X_t indexed over the set of all integers $\mathbb{Z}$. We assume that this process is such that its first-order backward difference $X^{(1)}_t = X_t - X_{t-1}$ is a second-order stationary process; i.e., neither the expected value of $X^{(1)}_t$ nor the covariance between $X^{(1)}_{t+\tau}$ and $X^{(1)}_t$ depends upon t, but the latter can depend upon the lag τ. We denote the expected value of X_t and the τ-th member of the autocovariance sequence (ACVS) by

$$\mu^{(1)} = E\{X^{(1)}_t\} \text{ and } s^{(1)}_\tau = \operatorname{cov}\{X^{(1)}_{t+\tau}, X^{(1)}_t\} = E\{(X^{(1)}_{t+\tau} - \mu^{(1)})(X^{(1)}_t - \mu^{(1)})\}.$$

Note that $\{X_t\}$ itself can be either stationary or non-stationary. A simple example of a process that fits into this framework is a white noise process $\{Z_t\}$. By definition, this process is such that $E\{Z_t\}$ is constant, while $\operatorname{cov}\{Z_{t+\tau}, Z_t\} = 0$ for $\tau \neq 0$ and $\operatorname{var}\{Z_t\} = \operatorname{cov}\{Z_t, Z_t\} = \sigma^2_Z$, where $0 < \sigma^2_Z < \infty$. Its first difference is a stationary process with zero mean and with an ACVS given by

$$s^{(1)}_\tau = \begin{cases} 2\sigma^2_Z, & \tau = 0; \\ -\sigma^2_Z, & \tau = \pm 1; \\ 0, & \text{otherwise.} \end{cases}$$

A second example is $X_t = a + bt + Z_t$, which is non-stationary because $E\{X_t\} = a + bt$ depends on t; however, $X^{(1)}_t = b + Z_t - Z_{t-1}$, which defines a process with a constant mean $\mu^{(1)} = b$ and the same ACVS as the first example. For a third example, assume $E\{Z_t\} = 0$, and construct a two-sided random walk process

$$X_t = \begin{cases} \sum_{u=1}^{t} Z_u + C, & t \geq 1; \\ C, & t = 0; \\ -\sum_{u=0}^{|t|-1} Z_{-u} + C, & t \leq -1, \end{cases} \tag{11}$$

where C is a random variable uncorrelated with each Z_t (a common choice is to set $C = 0$; i.e., C has zero mean and zero variance). Here $X^{(1)}_t = Z_t$, so $\mu^{(1)} = 0$, and $s^{(1)}_\tau$ corresponds to the ACVS for a white noise process.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

Other examples of processes whose first-order backward differences are stationary processes include power-law processes with spectra proportional to f^α with $\alpha > -3$, fractional Brownian motion and fractional Gaussian noise (Mandelbrot and van Ness, 1968), autoregressive, integrated moving average (ARIMA(p,d,q) processes of orders $d = 0$ and $d = 1$ (Box *et al.*, 2008) and fractionally integrated versions thereof (ARFIMA(p,δ,q) processes) as long as $\delta < 3/2$ (Beran, 1994).

Given a process $\{X_t\}$ with stationary first-order backward differences, we now define an associated Haar wavelet coefficient process of level j:

$$\overline{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t \in \mathbb{Z}, \tag{12}$$

which makes use of the level j Haar wavelet filter of Equation (3). We can also express this process in terms of the stationary backward differences $\{X_t^{(1)}\}$:

$$\overline{W}_{j,t} = \sum_{l=0}^{L_j-2} \tilde{b}_{j,l}^{(1)} X_{t-l}^{(1)}, \quad t \in \mathbb{Z}, \tag{13}$$

where the filter $\{\tilde{b}_{j,l}^{(1)}\}$ is the cumulative summation of $\{\tilde{h}_{j,l}\}$:

$$\tilde{b}_{j,l}^{(1)} = \begin{cases} \sum_{m=0}^{l} \tilde{h}_{j,m} = \frac{1}{2} - \left|\frac{1}{2} - \frac{l+1}{2^j}\right|, & l = 0, \ldots, L_j - 2; \\ 0, & \text{otherwise} \end{cases}$$

(this is a special case of Lemma 1, Craigmile and Percival, 2005). Since application of a linear filter to a stationary process yields another stationary process as its output, the Haar wavelet coefficient process is stationary. The variance of this process is the jth-level Haar wavelet variance:

$$\nu_X^2(\tau_j) = \text{var}\,\{\overline{W}_{j,t}\}, \tag{14}$$

where, as before, $\tau_j = 2^{j-1}$ is the scale associated with the jth-level wavelet coefficients. If $\{X_t\}$ is a stationary process (the simplest example being white noise), then we have

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \text{var}\,\{X_t\} \tag{15}$$

(for details, see Section 8.1 of Percival and Walden, 2000). The Haar wavelet variance thus affords a scale-based ANOVA for a stationary process that parallels the empirical ANOVA for a time series we described in the previous section. If $\{X_t\}$ is non-stationary (an example being a random walk process), then we have

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \infty, \tag{16}$$

which is in keeping with Equation (15) if we regard the variance of a non-stationary process as being infinite (for the random walk process of Equation (11), if we were to specify C to have infinite variance, then we would have $\text{var}\,\{X_t\} = \infty$ for all t; if C has finite variance, then $\text{var}\,\{X_t\}$ is finite for all t, but time varying). We can use

Equations (12) and (13) to relate $\{\nu_X^2(\tau_j)\}$ to the ACVS $\{s_\tau^{(0)}\}$ for $\{X_t\}$ and/or the ACVS $\{s_\tau^{(1)}\}$ for $\{X_t^{(1)}\}$:

$$\begin{aligned} \nu_X^2(\tau_j) &= \sum_{l=0}^{L_j-d-1} \sum_{m=0}^{L_j-d-1} \tilde{b}_{j,l}^{(d)} \tilde{b}_{j,m}^{(d)} s_{l-m}^{(d)} \\ &= s_0^{(d)} \sum_{l=0}^{L_j-d-1} \left(\tilde{b}_{j,l}^{(d)}\right)^2 + 2 \sum_{\tau=1}^{L_j-d-1} s_\tau^{(d)} \sum_{l=0}^{L_j-d-1-\tau} \tilde{b}_{j,l}^{(d)} \tilde{b}_{j,l+\tau}^{(d)}, \end{aligned} \tag{17}$$

where d can be either 0 or 1 and where we define $\tilde{b}_{j,l}^{(0)} = \tilde{h}_{j,l}$ (an interpretation for d is that it is the number of backward differencing operations needed to reduce $\{X_t\}$ to a stationary process). If $\{X_t\}$ is stationary, then the above expression for $\nu_X^2(\tau_j)$ holds for both $d = 0$ and $d = 1$; on the other hand, if $\{X_t\}$ is non-stationary, then the ACVS $\{s_\tau^{(0)}\}$ doesn't formally exist, so the above only holds for $d = 1$ (our working assumption is that $\{X_t^{(1)}\}$ is stationary, and hence the ACVS $\{s_\tau^{(1)}\}$ always exists). Alternatively we can express the Haar wavelet variance in terms of the semivariogram, which is defined as $\gamma_\tau = \frac{1}{2} \operatorname{var}\{X_\tau - X_0\}$:

$$\nu_X^2(\tau_j) = -\sum_{l=0}^{L_j-1} \sum_{m=0}^{L_j-1} \tilde{h}_{j,l} \tilde{h}_{j,m} \gamma_{l-m}.$$

The wavelet variance became a topic of interest soon after the introduction of the discrete wavelet transform in the late 1980s. The Allan variance was introduced two decades earlier in Allan (1966) specifically as a measure of clock performance (see also Barnes *et al.*, 1971; Rutman, 1978; Percival, 1983, 1991; Greenhall, 1991; Flandrin, 1992; and Percival and Guttorp, 1994). Suppose that, at time $t\,\Delta_t$, we record the difference ϕ_t in the times (or phases) kept by two clocks. Define the τ_1 average fractional frequency deviates as

$$\overline{Y}_t(\tau_1) = \frac{\phi_t - \phi_{t-1}}{\tau_1\,\Delta_t}, \tag{18}$$

which is proportional to the first-order backward difference of the ϕ_t measurements. As an example, suppose that ϕ_t is measured in nanoseconds and that the sampling time between measurements is $\Delta_t = 1$ day $= 8.64 \times 10^{13}$ nanoseconds. Then $\overline{Y}_t(\tau_1)$ is a dimensionless quantity such that a change of 8.64 nanoseconds in ϕ_t from one day to the next corresponds to a fractional frequency deviate of 10^{-13}. A typical clock is a device that counts the number of cycles produced by an oscillator operating at a nominal frequency f_0 (an example being $f_0 = 10$ megahertz). The τ_1 average fractional frequency deviate indicates how much the 'instantaneous' frequency $f(s)$ at time s of the oscillator deviates relative to f_0 when averaged over a time period of $\tau_1\,\Delta_t$:

$$\overline{Y}_t(\tau_1) = \frac{1}{\tau_1\,\Delta_t} \int_{t\,\Delta_t - \tau_1\,\Delta_t}^{t\,\Delta_t} \frac{f(s) - f_0}{f_0}\, ds$$

(conceptually $\overline{Y}_t(\tau_1)$ would not change if f_0 were to modified to be, say, 5 megahertz). Likewise, the τ_j average fractional frequency deviates are defined by

$$\overline{Y}_t(\tau_j) = \frac{1}{\tau_j\,\Delta_t} \int_{t\,\Delta_t - \tau_j\,\Delta_t}^{t\,\Delta_t} \frac{f(s) - f_0}{f_0}\, ds = \frac{1}{\tau_j} \sum_{l=0}^{\tau_j - 1} \overline{Y}_{t-l}(\tau_1) = \frac{\phi_t - \phi_{t-\tau_j}}{\tau_j\,\Delta_t}. \tag{19}$$

The fact that $\overline{Y}_t(\tau_j)$ is an average of a block of the τ_1 deviates is an indication of the leading role that these deviates play.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

If we now regard $\{\overline{Y}_t(\tau_1)\}$ as a stochastic process whose first order backward difference is a stationary process, the Allan variance at scale τ_j for these τ_1 deviates is defined as

$$\sigma^2_{\overline{Y}}(2,\tau_j) = \frac{1}{2}E\left\{\left(\overline{Y}_t(\tau_j) - \overline{Y}_{t-\tau_j}(\tau_j)\right)^2\right\}. \tag{20}$$

The '2' in the notation $\sigma^2_{\overline{Y}}(2,\tau_j)$ merits a comment. In signal processing and time series analysis, the sample variance of, say, $X_0, \ldots, X_{N-1}$ is usually defined as in Equation (8), but a common alternative definition is

$$\frac{1}{N-1}\sum_{t=0}^{N-1}\left(X_t - \overline{X}\right)^2, \tag{21}$$

which involves a division by $N-1$ rather than N (this alternative estimator is sometimes referred to as an unbiased estimator, but in general both it and the sample variance of Equation (8) are biased for finite N – see, e.g., David, 1985). Take $\overline{Y}_{t-\tau_j}(\tau_j)$ and $\overline{Y}_t(\tau_j)$ to be a sample of size $N = 2$. Using the alternative definition, the sample variance becomes

$$\left(\overline{Y}_{t-\tau_j}(\tau_j) - \frac{\overline{Y}_{t-\tau_j}(\tau_j) + \overline{Y}_t(\tau_j)}{2}\right)^2 + \left(\overline{Y}_t(\tau_j) - \frac{\overline{Y}_{t-\tau_j}(\tau_j) + \overline{Y}_t(\tau_j)}{2}\right)^2 = \frac{\left(\overline{Y}_t(\tau_j) - \overline{Y}_{t-\tau_j}(\tau_j)\right)^2}{2}.$$

The expected value of the right-hand side is the Allan variance, which is thus sometimes called the 'two sample' variance, hence explaining the '2' in $\sigma^2_{\overline{Y}}(2,\tau_j)$. If the $\overline{Y}_t(\tau_j)$ deviates are a white noise process with variance, say, $\sigma^2_{\overline{Y}(\tau_j)}$, then $\sigma^2_{\overline{Y}}(2,\tau_j) = \sigma^2_{\overline{Y}(\tau_j)}$; i.e., the Allan variance coincides with the process variance for white noise, which is not the case for any non-white stationary process. (Recall that τ_j is an integer-valued standardized scale that is unitless but can be converted into a physically meaningful scale upon multiplication by the sampling interval Δ_t. Traditionally the Allan variance is regarded as a function of a physical scale, which would dictate using $\sigma^2_{\overline{Y}}(2,\tau_j\,\Delta_t)$ rather than $\sigma^2_{\overline{Y}}(2,\tau_j)$; however, we will stick with the more compact notation throughout.)

For the τ_1 average fractional frequency deviates $\overline{Y}_t(\tau_1)$, the relationship between the Allan variance at scale τ_j and the Haar wavelet variance $\nu^2_{\overline{Y}}(\tau_j)$ at scale τ_j is simple:

$$\sigma^2_{\overline{Y}}(2,\tau_j) = 2\nu^2_{\overline{Y}}(\tau_j). \tag{22}$$

To establish this connection, it suffices to show that

$$\frac{\left(\overline{Y}_t(\tau_j) - \overline{Y}_{t-\tau_j}(\tau_j)\right)^2}{2} = 2\left(\sum_{l=0}^{L_j-1}\tilde{h}_{j,l}\overline{Y}_{t-l}(\tau_1)\right)^2$$

or, equivalently,

$$\overline{Y}_t(\tau_j) - \overline{Y}_{t-\tau_j}(\tau_j) = 2\sum_{l=0}^{L_j-1}\tilde{h}_{j,l}\overline{Y}_{t-l}(\tau_1);$$

however, considering the left-hand side of the above and using Equation (19) gives

$$\overline{Y}_t(\tau_j) - \overline{Y}_{t-\tau_j}(\tau_j) = \frac{1}{2^{j-1}}\sum_{l=0}^{2^{j-1}-1}\overline{Y}_{t-l}(\tau_1) - \frac{1}{2^{j-1}}\sum_{l=0}^{2^{j-1}-1}\overline{Y}_{t-2^{j-1}-l}(\tau_1)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

since $\tau_j = 2^{j-1}$, while considering the right-hand side and using Equation (3) gives

$$\begin{aligned} 2\sum_{l=0}^{L_j-1} \tilde{h}_{j,l}\overline{Y}_{t-l}(\tau_1) &= 2\left(\frac{1}{2^j}\sum_{l=0}^{2^{j-1}-1}\overline{Y}_{t-l}(\tau_1) - \frac{1}{2^j}\sum_{l=2^{j-1}}^{2^j-1}\overline{Y}_{t-l}(\tau_1)\right) \\ &= \frac{1}{2^{j-1}}\sum_{l=0}^{2^{j-1}-1}\overline{Y}_{t-l}(\tau_1) - \frac{1}{2^{j-1}}\sum_{l=0}^{2^{j-1}-1}\overline{Y}_{t-2^{j-1}-l}(\tau_1) \end{aligned}$$

since $L_j = 2^j$, hence establishing the claimed connection.

Why do the Allan and Haar wavelet variances differ by a factor of two? Equality between the two variances can be achieved by redefining either the Haar wavelet variance or the Allan variance. Let's explore the consequences of doing one or the other. Considering first the wavelet variance, agreement with the Allan variance occurs if we redefine the Haar wavelet filters by multiplying $\tilde{h}_l$ by $\sqrt{2}$; however, doing so would destroy the important ANOVA property of the Haar wavelet variance expressed in Equations (15) and (16). On the other hand, the Allan variance is a special case of the sample variance when there are only two samples. Equation (21) is the definition of the sample variance used in Allan (1966). There is a division by $N-1$, which is unity in the two-sample case. The alternative definition of the sample variance given by Equation (8) has a divisor of N, which is two in the two-sample case. Use of this definition leads to a two-sample variance that is identical to the Haar wavelet variance. The discrepancy between the Allan and Haar wavelet variances would thus not have arisen if the two-sample variance had been formulated using the definition for the sample variance prevalent in signal processing and time series analysis.

V. Estimation of the Haar wavelet variance and the Allan variance

We now consider the problem of estimating the Haar wavelet variance and the Allan variance given a realization (observations) from a portion $X_0, \ldots, X_{N-1}$ of a process $\{X_t : t \in \mathbb{Z}\}$ whose first-order backward differences $X_t^{(1)} = X_t - X_{t-1}$ constitute a second-order stationary process. Since Equation (22) says that the Allan variance is twice the Haar wavelet variance, we can readily adjust any estimator of the latter to become an estimator of the former (and vice versa). In this section we review various estimators that have been proposed for wavelet variances in general (both Haar and non-Haar), but we will focus on the Haar case due to its connection with the Allan variance. Some of the Haar wavelet variance estimators correspond to well-known estimators of the Allan variance, but others do not. In addition, the statistical theory behind the Haar wavelet variance estimators provides insights into how to assess Allan variance estimators without having to make certain restrictive assumptions commonly entertained in the literature.

All the estimators of the jth-level Haar wavelet variance $\nu_X^2(\tau_j)$ to be discussed below are based on the level j MODWT wavelet coefficients, which make use solely of $X_0, \ldots, X_{N-1}$:

$$\widetilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N}, \quad t = 0, \ldots, N-1$$

(see Equation (5); as before, $L_j = 2^j = 2\tau_j$, and $\{\tilde{h}_{j,l}\}$ is the jth level Haar wavelet filter of Equation(3)). By

definition the wavelet variance is the variance of

$$\overline{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}.$$

A comparison of $\widetilde{W}_{j,t}$ with $\overline{W}_{j,t}$ shows that these are necessarily equal when $L_j - 1 \leq t \leq N-1$, but not when $0 \leq t < L_j - 1$. A coefficient $\widetilde{W}_{j,t}$ that is equal to $\overline{W}_{j,t}$ is termed a non-boundary coefficient; all others are called boundary coefficients.

A. Unbiased variance estimators

We can formulate an unbiased estimator of $\nu_X^2(\tau_j)$ using the non-boundary wavelet coefficients from the Haar MODWT of $X_0, \ldots, X_{N-1}$. Under the condition that $E\{\overline{W}_{j,t}\} = 0$ and that $M_j = \max\{N - L_j + 1, 0\} > 0$, the unbiased estimator is

$$\hat{\nu}_X^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2. \tag{23}$$

The corresponding estimator of the Allan variance is the well-known maximal overlap estimator and can be expressed as

$$\hat{\sigma}_X^2(2, \tau_j) = 2\hat{\nu}_X^2(\tau_j). \tag{24}$$

The zero mean condition is key here. If $\{X_t\}$ is a stationary process, then necessarily $E\{\overline{W}_{j,t}\} = 0$ no matter what $E\{X_t\}$ is. This result follows from Equation (12):

$$E\{\overline{W}_{j,t}\} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} E\{X_{t-l}\} = 0 \text{ because } \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0.$$

On the other hand, if $\{X_t\}$ is non-stationary, then $E\{\overline{W}_{j,t}\}$ need not be zero unless $E\{X_t^{(1)}\} = 0$, as can be seen from Equation (13):

$$E\{\overline{W}_{j,t}\} = \sum_{l=0}^{L_j-2} \tilde{b}_{j,l}^{(1)} E\{X_{t-l}^{(1)}\} = 2^{j-2} E\{X_t^{(1)}\} \text{ because } \sum_{l=0}^{L_j-2} \tilde{b}_{j,l}^{(1)} = 2^{j-2}.$$

In the context of measuring clock performance, two commonly assumed models are white phase noise and white frequency noise (i.e., $\{\phi_t\}$ and $\{\overline{Y}_t(\tau_1)\}$ are white noise processes). Taking X_t to be $\overline{Y}_t(\tau_1) = (\phi_t - \phi_{t-1})/(\tau_t \, \Delta_t)$, both of these models imply that $\{X_t\}$ is a stationary process so that the zero mean condition on $\overline{W}_{j,t}$ holds. Two other models of interest are flicker frequency noise and random walk frequency noise. In these cases $\{X_t\}$ is a non-stationary process, whereas $\{X_t^{(1)}\}$ is stationary, but now the zero mean condition on $\overline{W}_{j,t}$ depends critically on the additional assumption that $E\{X_t^{(1)}\} = 0$. For all four models, the zero mean condition is problematic in the presence of a linear drift in frequency because, while $\{X_t\}$ is then no longer stationary, the process $\{X_t^{(1)}\}$ still is, but $E\{X_t^{(1)}\}$ cannot be assumed to be zero because of its dependence on the slope of the drift.

How well do $\hat{\nu}_X^2(\tau_j)$ and $\hat{\sigma}_X^2(2, \tau_j)$ estimate $\nu_X^2(\tau_j)$ and $\sigma_X^2(2, \tau_j)$? To properly address this question, we would need to know the distributions for these estimators, which in turn require knowledge of the hard-to-come-by multivariate distribution of $X_0, \ldots, X_{N-1}$. In using the Allan variance to assess clock performance, this difficulty

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

is traditionally overcome by assuming one of five canonical power-law models for $\{\phi_t\}$ or, equivalently, $\{\overline{Y}_t(\tau_1)\}$, namely, white phase, flicker phase, white frequency (also called random walk phase), flicker frequency and random walk frequency noise. This approach in effect completely specifies $\{s_\tau^{(1)}\}$ up to an unknown multiplicative constant. With this assumption, we can produce a, say, 95% confidence interval (CI) for the unknown $\sigma_X^2(2,\tau_j)$ based upon its estimator $\hat{\sigma}_X^2(2,\tau_j)$. One critique of this approach is that the only assumed source of uncertainty is the unknown multiplicative constant – the type of model is assumed to be perfectly known. The approach also implicitly assumes that a single power-law model is in effect over all τ_j, which is rarely the case in practice. The usual work-around is to assume a series of power-law models, with different models being dominant over selected τ_j. Strictly speaking, these composite models are problematic because the procedure used to generate the CIs assumes a single model holds for all τ_j.

An alternative to assuming power-law models is to obtain large-sample distributions for $\hat{\nu}_X^2(\tau_j)$ and $\hat{\sigma}_X^2(2,\tau_j)$ and to use these distributions to provide asymptotically correct CIs for the true variances. These large-sample distributions are tractable if we make some additional assumptions about the level j wavelet coefficient process $\{\overline{W}_{j,t}\}$. In particular, suppose that this process is stationary and Gaussian with an ACVS $\{s_{j,\tau}\}$ that is square summable:

$$A_j = \sum_{\tau=-\infty}^{\infty} s_{j,\tau}^2 < \infty. \tag{25}$$

With these assumptions it can be argued that

$$\frac{M_j^{1/2}(\hat{\nu}_X^2(\tau_j) - \nu_X^2(\tau_j))}{(2A_j)^{1/2}} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1) \text{ and } \frac{M_j^{1/2}(\hat{\sigma}_X^2(2,\tau_j) - \sigma_X^2(2,\tau_j))}{(8A_j)^{1/2}} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1)$$

approximately for large M_j, where '$\stackrel{\mathrm{d}}{=}$' stands for 'is equal in distribution to', and $\mathcal{N}(0,1)$ is a standard Gaussian random variable (Mondal, 2007, proved the above using Theorem 5 of Giraitis and Surgailis, 1985, which is an improvement upon earlier – but more complicated and less general – proofs in Percival, 1983, for the Haar wavelet and in Percival, 1995, for general Daubechies wavelet filters). Based upon this large-sample result, we can claim that the random interval

$$\left[\hat{\nu}_X^2(\tau_j) - \Phi^{-1}(1-p)\left(\frac{2A_j}{M_j}\right)^{1/2},\ \hat{\nu}_X^2(\tau_j) + \Phi^{-1}(1-p)\left(\frac{2A_j}{M_j}\right)^{1/2}\right] \tag{26}$$

constitutes an approximate $100(1-2p)\%$ CI for $\nu_X^2(\tau_j)$, where $\Phi^{-1}(p)$ is the $p \times 100\%$ percentage point for the standard Gaussian distribution. The corresponding CI for $\sigma_X^2(2,\tau_j)$ is obvious: just replace $\hat{\nu}_X^2(\tau_j)$ and $2A_j$ with $\hat{\sigma}_X^2(2,\tau_j)$ and $8A_j$ in the above.

The lower limit of the CI displayed in (26) is not restricted to be nonnegative even though the true wavelet variance is (replacing this limit with zero yields a technically valid CI, but one which Scheffé, 1959, pp. 229–31, criticizes as being potentially misleading). An alternative – but asymptotically equivalent – approach that yields CIs with positive lower limits is to assume that

$$\frac{\eta_j \hat{\nu}_X^2(\tau_j)}{\nu_X^2(\tau_j)} \stackrel{\mathrm{d}}{=} \chi_{\eta_j}^2 \text{ or, equivalently, } \frac{\eta_j \hat{\sigma}_X^2(2,\tau_j)}{\sigma_X^2(2,\tau_j)} \stackrel{\mathrm{d}}{=} \chi_{\eta_j}^2, \tag{27}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

where $\chi^2_{\eta_j}$ is a chi-square random variable with η_j degrees of freedom. We can set η_j using a moment matching scheme. Recalling first that $E\{\chi^2_{\eta_j}\} = \eta_j$ and $\mathrm{var}\,\{\chi^2_{\eta_j}\} = 2\eta_j$ so that

$$\frac{2(E\{c\chi^2_{\eta_j}\})^2}{\mathrm{var}\,\{c\chi^2_{\eta_j}\}} = \eta_j \text{ for any constant } c \ ,$$

we use the facts that $E\{\hat{\nu}^2_X(\tau_j)\} = \nu^2_X(\tau_j)$ and $\mathrm{var}\,\{\hat{\nu}^2_X(\tau_j)\} \approx 2A_j/M_j$ to obtain

$$\eta_j = \frac{2\left(E\{\hat{\nu}^2_X(\tau_j)\}\right)^2}{\mathrm{var}\,\{\hat{\nu}^2_X(\tau_j)\}} = \frac{2\nu^4_X(\tau_j)}{\mathrm{var}\,\{\hat{\nu}^2_X(\tau_j)\}} \approx \frac{M_j\nu^4_X(\tau_j)}{A_j} \text{ or, equivalently, } \eta_j \approx \frac{M_j\sigma^4_X(2,\tau_j)}{4A_j}. \tag{28}$$

The random intervals

$$\left[\frac{\eta_j\hat{\nu}^2_X(\tau_j)}{Q_{\eta_j}(1-p)}, \frac{\eta_j\hat{\nu}^2_X(\tau_j)}{Q_{\eta_j}(p)}\right] \text{ and } \left[\frac{\eta_j\hat{\sigma}^2_X(2,\tau_j)}{Q_{\eta_j}(1-p)}, \frac{\eta_j\hat{\sigma}^2_X(2,\tau_j)}{Q_{\eta_j}(p)}\right] \tag{29}$$

are thus approximate $100(1-2p)\%$ CIs for $\nu^2_X(\tau_j)$ and $\sigma^2_X(2,\tau_j)$, where $Q_{\eta_j}(p)$ is the $p \times 100\%$ percentage point for the $\chi^2_{\eta_j}$ distribution.

We need to know A_j to form the CIs of either Equation (26) or (29). Treating $\widetilde{W}_{j,L_j-1}, \ldots, \widetilde{W}_{j,N-1}$ as a time series whose mean is known to be zero, the standard estimator of $s_{j,\tau}$ becomes

$$\hat{s}_{j,\tau} = \frac{1}{M_j}\sum_{t=L_j-1}^{N-|\tau|-1}\widetilde{W}_{j,t}\widetilde{W}_{j,t+|\tau|}, \ \ 0 \le |\tau| \le M_j - 1,$$

and $\hat{s}_{j,\tau} = 0$ for $|\tau| \ge M_j$. An approximately unbiased estimator of A_j is given by

$$\hat{A}_j = \sum_{\tau=-(M_j-1)}^{M_j-1}\frac{\hat{s}^2_{j,\tau}}{2} = \frac{\hat{s}^2_{j,0}}{2} + \sum_{\tau=1}^{M_j-1}\hat{s}^2_{j,\tau} = \frac{\hat{\nu}^4(\tau_j)}{2} + \sum_{\tau=1}^{M_j-1}\hat{s}^2_{j,\tau} \text{ or, equivalently, } \hat{A}_j = \frac{\hat{\sigma}^4_X(2,\tau_j)}{8} + \sum_{\tau=1}^{M_j-1}\hat{s}^2_{j,\tau} \tag{30}$$

(comparison of $\hat{A}_j$ with the definition of A_j in Equation (25) reveals a counter-intuitive division by two in the above, which can be traced to the moments of a χ^2 distribution – see Percival and Walden, 2000, Section 8.4, for details). We can substitute $\hat{A}_j$ for A_j in Equation (26) to get approximate Gaussian-based CIs. The corresponding $\chi^2_{\eta_j}$-based CIs require plugging $\hat{A}_j$ along with either $\hat{\nu}^4(\tau_j)$ or $\hat{\sigma}^2_X(2,\tau_j)$ into Equation (28) to estimate η_j, which in turn can be used in Equation (29) to produce the desired CIs. Monte Carlo studies indicate that CIs based on $\hat{A}_j$ are reasonably accurate as long as $M_j \ge 128$. If M_j is too small to get a decent estimate of A_j, a fallback is to set η_j to $\max\{M_j/2^j, 1\}$ in Equation (29). When applicable, this approach tends to produce conservative CIs (for details, see Percival and Walden, 2000, Section 8.4).

A key assumption behind the CIs of Equations (26) and (29) is the Gaussianity of $\{\overline{W}_{j,t}\}$. If $\{X_t\}$ is Gaussian, then $\{\overline{W}_{j,t}\}$ must also be such since $\overline{W}_{j,t}$ is a linear combination of X_t, X_{t-1}, …, X_{t-L_j+1}. Even if $\{X_t\}$ is non-Gaussian, Gaussianity might still be a reasonable approximation for $\{\overline{W}_{j,t}\}$: the wavelet coefficient process comes about by applying a linear filter to $\{X_t\}$, and such filtering tends to induce Gaussianity (Mallows, 1967). If Gaussianity is not a reasonable assumption (at least approximately) for $\{\overline{W}_{j,t}\}$, the large-sample distributions for $\hat{\nu}^2_X(\tau_j)$ and $\hat{\sigma}^2_X(2,\tau_j)$ are still tractable under additional assumptions (see Serroukh *et al.*, 2000, for details). It can be argued that

$$\frac{M_j^{1/2}(\hat{\nu}^2_X(\tau_j) - \nu^2_X(\tau_j))}{S^{1/2}_{\overline{W}^2_{j,t}}(0)} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1) \text{ and } \frac{M_j^{1/2}(\hat{\sigma}^2_X(2,\tau_j) - \sigma^2_X(2,\tau_j))}{2S^{1/2}_{\overline{W}^2_{j,t}}(0)} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

approximately for large M_j, where $S_{\overline{W}^2_{j,t}}(0) > 0$ is the spectral density function for the presumed-to-be strictly stationary process $\{\overline{W}^2_{j,t}\}$ (Serroukh *et al.*, 2000). This large-sample result leads to $\chi^2_{\eta_j}$-based CIs of the same form as Equation (29), but with degrees of freedom now given by

$$\eta_j \approx \frac{2M_j\nu^4_X(\tau_j)}{S_{\overline{W}^2_{j,t}}(0)} \text{ or, equivalently, } \eta_j \approx \frac{M_j\sigma^4_X(2,\tau_j)}{2S_{\overline{W}^2_{j,t}}(0)}.$$

In practical applications we need to estimate $S_{\overline{W}^2_{j,t}}(0)$. Serroukh *et al.* (2000) propose a multitaper spectral density estimator of order $K = 5$ based upon four Slepian (discrete prolate spheroidal sequence) data tapers $\{v_{k,t}\}$, $k = 0, \ldots, K-1$, with a design bandwidth of $7/M_j$ (Thomson, 1982; Percival and Walden, 1993). Letting

$$J_k(0) = \sum_{t=L_j-1}^{N-1} v_{k,t}\widetilde{W}^2_{j,t},\ V_k(0) = \sum_{t=L_j-1}^{N-1} v_{k,t} \text{ and } \check{\nu}^2_X(\tau_j) = \frac{\sum_{k=0}^{K-1} J_k(0)V_k(0)}{\sum_{k=0}^{K-1} V^2_k(0)}, \tag{31}$$

this estimator is given by

$$\hat{S}_{\widetilde{W}^2_{j,t}}(0) = \frac{1}{K}\sum_{k=0}^{K-1}\left(J_k(0) - V_k(0)\check{\nu}^2_X(\tau_j)\right)^2. \tag{32}$$

Two comments about $\check{\nu}^2_X(\tau_j)$ are in order. First, its computation can be simplified because $V_k(0) = 0$ for odd k. Second, its notation suggests that it is an estimator of the wavelet variance and that, because of the exponent '2', it is nonnegative. In fact, $E\{\check{\nu}^2_X(\tau_j)\} = \nu^2_X(\tau_j)$ so, similar to $\hat{\nu}^2_X(\tau_j)$, it is an unbiased estimator of $\nu^2_X(\tau_j)$; however, whereas $\hat{\nu}^2_X(\tau_j)$ must be nonnegative, $\check{\nu}^2_X(\tau_j)$ is not constrained to be such.

B. Biased variance estimators

The unbiased estimators $\hat{\nu}^2_X(\tau_j)$ and $\hat{\sigma}^2_X(2,\tau_j)$ make use of just the non-boundary level j Haar MODWT wavelet coefficients. As j increases, the number $M_j = \max\{N - 2^j + 1, 0\}$ of such coefficients decreases to zero, while the number $N - M_j$ of boundary coefficients increases to N. The MODWT-based ANOVA of Equation (9) makes use of both types of coefficients. Is there anything to be gained by using all the wavelet coefficients to estimate $\nu^2_X(\tau_j)$ and $\sigma^2_X(2,\tau_j)$ rather than just the non-boundary coefficients? For the wavelet variance, this estimator takes the form of one piece of the MODWT-based ANOVA, namely,

$$\hat{\sigma}^2_{\widetilde{W}_j} = \frac{1}{N}\|\widetilde{\mathbf{W}}_j\|^2 = \frac{1}{N}\sum_{t=0}^{N-1}\widetilde{W}^2_{j,t} \tag{33}$$

(see Equation (10)); the corresponding estimator for the Allan variance is $2\hat{\sigma}^2_{\widetilde{W}_j}$. Because $E\{\widetilde{W}^2_{j,t}\}$ need not be equal to $\nu^2_X(\tau_j)$ when $t = 0, \ldots, N - M_j - 1$, estimators of $\nu^2_X(\tau_j)$ and $\sigma^2_X(2,\tau_j)$ that are based on $\hat{\sigma}^2_{\widetilde{W}_j}$ are in general biased. For a fixed level j, it can be argued that $E\{\hat{\sigma}^2_{\widetilde{W}_j}\} \to \nu^2_X(\tau_j)$ as $N \to \infty$ if $\{X_t\}$ is a stationary process; i.e., $\hat{\sigma}^2_{\widetilde{W}_j}\}$ is asymptotically unbiased. The same cannot be said if $\{X_t\}$ is a non-stationary process but its first-order backward difference $\{X^{(1)}_t\}$ is stationary. In this case $E\{(X_0 - X_{N-1})^2\} \to \infty$ as $N \to \infty$. This increasingly serious mismatch between the beginning and end of the time series affects the $N - M_j$ boundary coefficients to the point that the bias in $\hat{\sigma}^2_{\widetilde{W}_j}$ does not decrease with increasing N even though the proportion $(2^j - 1)/N$ of boundary coefficients decreases to zero.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

We can create biased estimators of $\nu_X^2(\tau_j)$ and $\sigma_X^2(2,\tau_j)$ that work well both when $\{X_t\}$ is stationary and when $\{X_t\}$ is non-stationary (but $\{X_t^{(1)}\}$ is) by considering the following time series of length $2N$:

$$X_0, X_1, \ldots, X_{N-2}, X_{N-1}, X_{N-1}, X_{N-2}, \ldots, X_1, X_0,$$

which we denote by $X'_0, \ldots, X'_{2N-1}$. The sample variance for this series is the same as for $X_0, \ldots, X_{N-1}$. so ANOVA based on X_t and X'_t are decompositions of the same variance. Denote the MODWT of the X'_t series by

$$\widetilde{W}'_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X'_{t-l \bmod 2N}, \quad t = 0, 1, \ldots, 2N-1,$$

which we also refer to as the MODWT of the X_t series using reflection boundary conditions. Greenhall *et al.* (1999) proposed the Allan variance estimator

$$\overline{\sigma}_X^2(2,\tau_j) = \frac{1}{N} \sum_{t=0}^{2N-1} (\widetilde{W}'_{j,t})^2, \tag{34}$$

which is biased in general. Aldrich (2005) showed that the corresponding biased wavelet variance estimator, namely,

$$\overline{\nu}_X^2(\tau_j) = \frac{1}{2N} \sum_{t=0}^{2N-1} (\widetilde{W}'_{j,t})^2, \tag{35}$$

is asymptotically equivalent to $\hat{\nu}_X^2(\tau_j)$ when $\{X_t\}$ is stationary and also when $\{X_t\}$ is non-stationary but $\{X_t^{(1)}\}$ is stationary. He also considered the mean squared errors of $\overline{\nu}_X^2(\tau_j)$ and $\hat{\nu}_X^2(\tau_j)$ for representative processes through both exact expressions and computer experiments. The biased estimator proved to be superior to the unbiased estimator, particularly in cases where M_j is small relative to N.

C. Variance estimators for gappy series

The unbiased and biased estimators of $\nu_X^2(\tau_j)$ and $\sigma_X^2(2,\tau_j)$ we have considered so far presume that X_0, X_1, …, X_{N-1} represent data sampled over a regular grid of times; i.e., the spacing between the times at which we recorded X_{t-1} and X_t is the same for $t = 1, 2, \ldots, N-1$. For a variety of reasons, this ideal sampling scheme might be compromised, leading to a gappy series; i.e., one for which certain data values are missing. To use a gappy series with the estimators described above, we must fill in the gaps somehow. There are a variety of possible interpolation schemes, ranging from the simple (e.g., linear interpolation) to the sophisticated (e.g., use of expectations and covariances conditioned on what was actually observed). Interpolation is an effective work-around when both the number of gaps and the gap sizes are small, but can be problematic otherwise. Mondal and Percival (2010) propose two specialized wavelet variance estimators to handle gappy time series, neither of which involve interpolation, but both of which make assumptions about how the gaps arise. Since both estimators can be readily adjusted to become Allan variance estimators (see the end of this subsection), it is appropriate to offer the following overview. For details, we refer the reader to Mondal and Percival (2010).

Suppose that $\{\delta_t\}$ is a strictly stationary binary-valued process that models the gap structure; i.e., $\delta_t = 0$ means that X_t is missing, while $\delta_t = 1$ says that it is available. Let β_k^{-1} denote the probability that both $\delta_t = 1$ and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

$\delta_{t+k} = 1$ (i.e., neither X_t nor X_{t+k} is missing). For $0 \le l \le L_j - 1$ and $0 \le m \le L_j - 1$, let

$$\hat{\beta}_{l,m}^{-1} = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \delta_{t-l}\delta_{t-m},$$

which is an estimator of β_{l-m}^{-1} (we assume that $\hat{\beta}_{l,m}^{-1} > 0$, which is restrictive because it might not hold for a time series with too many gaps). The first estimator is a covariance-type estimator given by

$$\hat{u}_X^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{m=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,m}\hat{\beta}_{l,m} X_{t-l}X_{t-m}\delta_{t-l}\delta_{t-m}, \tag{36}$$

while the second is a semivariogram-type estimator given by

$$\hat{v}_X^2(\tau_j) = -\frac{1}{2M_j} \sum_{t=L_j-1}^{N-1} \sum_{l=0}^{L_j-1} \sum_{m=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,m}\hat{\beta}_{l,m} (X_{t-l} - X_{t-m})^2\delta_{t-l}\delta_{t-m}. \tag{37}$$

For a gap-free time series (i.e., $\delta_t = 1$ for all t), both estimators reduce to the unbiased estimator $\hat{\nu}_X^2(\tau_j)$ of Equation (23). In the presence of gaps, the expected value of both estimators is $\nu_X^2(\tau_j)$, but neither estimator is guaranteed to be nonnegative.

An advantage to using $\hat{u}_X^2(\tau_j)$ or $\hat{v}_X^2(\tau_j)$ rather than resorting to interpolation is the availability of large-sample theory that takes into account the uncertainty in estimating $\nu_X^2(\tau_j)$ due to missing data. In the case of the covariance-type estimator $\hat{u}_X^2(\tau_j)$, if $\{X_t\}$ is a stationary process and if certain technical conditions hold, then

$$\frac{M_j^{1/2}(\hat{u}_X^2(\tau_j) - \nu_X^2(\tau_j))}{S_{U_{j,t}^2}^{1/2}(0)} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1)$$

approximately for large M_j, where the numerator involves the spectral density function at zero frequency for the stationary process

$$U_{j,t}^2 = \sum_{l=0}^{L_j-1} \sum_{m=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,m}\beta_{l-m} X_{t-l}X_{t-m}\delta_{t-l}\delta_{t-m}.$$

This random variable has a mean of $\nu_X^2(\tau_j)$ and collapses to $\overline{W}_{j,t}^2$ in the gap-free case, but can be negative. If we replace β_{l-m} in the above with $\hat{\beta}_{l,m}$ to create, say, $\widetilde{U}_{j,t}^2$, $t = L_j - 1, \ldots, N-1$, then we can estimate $S_{U_{j,t}^2}(0)$ using the multitaper approach of Equations (31) and (32) with $J_k(0)$ redefined to be $\sum_t v_{k,t}\widetilde{U}_{j,t}^2$. On the other hand, in the case of the semivariogram-type estimator $\hat{v}_X^2(\tau_j)$, if $\{X_t\}$ is a stationary process or if it is non-stationary but its first-order backward differences $\{X_t^{(1)}\}$ are stationary, then, under suitable technical conditions,

$$\frac{M_j^{1/2}(\hat{v}_X^2(\tau_j) - \nu_X^2(\tau_j))}{S_{V_{j,t}^2}^{1/2}(0)} \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1),$$

which, as before, involves an SDF at zero frequency, but this time for the stationary process defined by

$$V_{j,t}^2 = -\frac{1}{2} \sum_{l=0}^{L_j-1} \sum_{m=0}^{L_j-1} \tilde{h}_{j,l}\tilde{h}_{j,m}\beta_{l-m} (X_{t-l} - X_{t-m})^2\delta_{t-l}\delta_{t-m}.$$

Again, the above has a mean of $\nu_X^2(\tau_j)$, collapses to $\overline{W}_{j,t}^2$ in the gap-free case and can be negative. Upon replacing β_{l-m} with $\hat{\beta}_{l,m}$ to create $\widetilde{V}_{j,t}^2$, $t = L_j - 1, \ldots, N-1$, we can estimate $S_{V_{j,t}^2}(0)$ using the multitaper approach again with $J_k(0)$ redefined in Equation (31) to be $\sum_t v_{k,t}\widetilde{V}_{j,t}^2$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

If $\{X_t\}$ is stationary, we can use either $\hat{u}_X^2(\tau_j)$ or $\hat{v}_X^2(\tau_j)$, but if it is non-stationary with stationary first-order backward differences, then we need to use $\hat{v}_X^2(\tau_j)$. Can we just dispense with $\hat{u}_X^2(\tau_j)$ in favor of $\hat{v}_X^2(\tau_j)$? Interestingly, for certain – but not all – stationary processes, $\hat{u}_X^2(\tau_j)$ is more efficient asymptotically than $\hat{v}_X^2(\tau_j)$ as measured by the ratio $S_{V_{j,t}^2}(0)/S_{U_{j,t}^2}(0)$. Thus the estimator $\hat{u}_X^2(\tau_j)$ is valuable even though it is not as widely applicable as $\hat{v}_X^2(\tau_j)$. A warning about $\hat{u}_X^2(\tau_j)$ is that, in contrast to $\hat{v}_X^2(\tau_j)$ (or $\hat{\nu}_X^2(\tau_j)$ when $\{X_t\}$ is stationary), it can change upon addition of a constant to the observed time series. It is thus important to center a time series by subtracting off its sample mean prior to computing $\hat{u}_X^2(\tau_j)$.

Turning $\hat{u}_X^2(\tau_j)$ and $\hat{v}_X^2(\tau_j)$ into estimators of the Allan variance is easy: just use $2\hat{u}_X^2(\tau_j)$ and $2\hat{v}_X^2(\tau_j)$ (cf. Equation (22)). Because the large-sample variances of $\hat{u}_X^2(\tau_j)$ and $\hat{v}_X^2(\tau_j)$ are, respectively, $S_{U_{j,t}^2}(0)/M_j$ and $S_{V_{j,t}^2}(0)/M_j$, the large-sample variances for the corresponding Allan variance estimators must be $4S_{U_{j,t}^2}(0)/M_j$ and $4S_{V_{j,t}^2}(0)/M_j$.

D. Robust variance estimators

The unbiased and biased estimators of $\nu_X^2(\tau_j)$ are sample means of squared wavelet coefficients. A general problem with any sample mean is its sensitivity to potential contamination from a small number of observations that do not truly reflect the underlying process of interest, but nonetheless have a disproportionately large influence on the sample mean. This problem motivates consideration of robust estimators, which are intended to be affected less by the presence of contamination than the sample mean. As an example of a robust estimator, consider the sample median for a population where the median is equal to mean. The sample mean $\bar{x}$ of, say, 1, 2 and a variable c is such that, as $c \to \infty$, then $\bar{x} \to \infty$ also; by contrast, the sample median stays fixed at 2. A robust alternative to the unbiased estimator $\hat{\nu}_X^2(\tau_j)$ is thus the sample median of $\widetilde{W}_{j,L_j-1}^2, \ldots, \widetilde{W}_{j,N-1}^2$ after an adjustment to take into account the difference between the population mean $\nu_X^2(\tau_j)$ and the population median of $\{\overline{W}_{j,t}^2\}$ (Stoev *et al.*, 2006). Mondal and Percival (2012) develop the statistical theory behind this approach under the assumption that $\{\overline{W}_{j,t}\}$ is a Gaussian process with a spectral density function whose square integrates to a finite value. Here we summarize the key elements of this theory.

Let

$$\text{sign}(x) = \begin{cases} 1, & x \geq 0; \text{ and} \\ -1, & x < 0, \end{cases}$$

and let $\phi(\cdot)$ and $\Phi^{-1}(\cdot)$ represent the probability density function and the inverse of the cumulative distribution function for a standard Gaussian random variable. Let

$$U_t = \text{sign}\left(\log\left(\overline{W}_{j,t}^2\right) - \log\left(\nu_X^2(\tau_j)\right) - 2\,\log\left(\Phi^{-1}(\tfrac{3}{4})\right)\right), \quad t \in \mathbb{Z},$$

which defines a binary-valued stationary process with spectral density function $S_U(\cdot)$. The robust estimator of $\nu_X^2(\tau_j)$ is given by

$$\hat{r}_X^2(\tau_j) = \frac{\text{median}\{\widetilde{W}_{j,t}^2\} \cdot \exp\left(-\hat{S}_U(0)/[2M_jC]\right)}{\left(\Phi^{-1}(\frac{3}{4})\right)^2}, \tag{38}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

where $\hat{S}_U(0)$ is obtained from the right-hand side of Equation (32) by redefining $J_k(0)$ in Equation (31) to be $\sum_t v_{k,t}\,\text{sign}(\log(\widetilde{W}_{j,t}^2) - \text{median}\{\log(\widetilde{W}_{j,t}^2)\})$, while $C = 4\left[\phi(\Phi^{-1}(\frac{3}{4}))\Phi^{-1}(\frac{3}{4})\right]^2$. Then

$$\frac{M_j^{1/2}(\hat{r}_X^2(\tau_j) - \nu_X^2(\tau_j))}{\nu_X^2(\tau_j) S_U^{1/2}(0)/C^{1/2}} \stackrel{\text{d}}{=} \mathcal{N}(0,1),$$

approximately for large M_j. The corresponding robust estimator of the Allan variance is

$$\hat{r}_X^2(2,\tau_j) = 2\hat{r}_X^2(\tau_j). \tag{39}$$

We can use the above robust estimators to form $\chi^2_{\eta_j}$-based CIs analogous to those of Equation (29), for which we can in practice set

$$\eta_j = \frac{2M_jC}{\hat{S}_U(0)}. \tag{40}$$

The median-type estimator $\hat{r}_X^2(\tau_j)$ guards against data contamination but is a less efficient estimator of $\nu_X^2(\tau_j)$ than the unbiased mean-type estimator $\hat{\nu}_X^2(\tau_j)$ when in fact the $\widetilde{W}_{j,t}$'s are free of contamination. Mondal and Percival (2012) found that, for moderate sample sizes, $\hat{r}_X^2(\tau_j)$ has approximately twice the variance of $\hat{\nu}_X^2(\tau_j)$ for a representative selection of stationary processes exhibiting both short- and long-range dependence. Thus, if the $\widetilde{W}_{j,t}$'s are truly Gaussian, we can expect $\hat{r}_X^2(\tau_j)$ to perform markedly poorer than $\hat{\nu}_X^2(\tau_j)$, but the presence of significant contamination can lead to a preference for the latter estimator.

The median-type estimator is a special case of the well-known robust M-estimators pioneered in Huber (1964), Mondal and Percival (2012) provide theory paralleling the above for M-estimators other than the median, to which we refer the reader for details.

VI. Beyond the Haar wavelet

As discussed in Section IV, the Allan variance is associated with the Haar wavelet filter, which has width $L_1 = 2$. This filter is the shortest Daubechies wavelet filter (Daubechies, 1992). All Daubechies filters $\{\tilde{h}_{1,l} : l = 0,\ldots,L_1-1\}$ have an even width L_1. The first column of Figure 2 shows wavelet filters of widths $L_1 = 2$ (the Haar case), 4 and 6 (the filter coefficients $\tilde{h}_{1,l}$ are defined to be zero for $l \geq L_1$ and also for $l < 0$). The displayed filters of widths 4 and 6 are known as the D(4) and D(6) wavelet filters, where 'D' designates a Daubechies extremal phase filter. All Daubechies wavelet filters have three fundamental properties:

$$\sum_{l=0}^{L_1-1} \tilde{h}_{1,l} = 0, \quad \sum_{l=0}^{L_1-1} \tilde{h}_{1,l}^2 = \tfrac{1}{2} \quad \text{and} \quad \sum_{l=0}^{L_1-1} \tilde{h}_{1,l}\tilde{h}_{1,l+2n} = 0 \ \text{ for all nonzero integers } n$$

(cf. Equation (4) with $j = 1$). In other words, the wavelet filter sums to zero; the sum of the squares of its coefficients is $1/2$; and the filter is orthogonal to its even shifts. For each wavelet filter $\{\tilde{h}_{1,l}\}$, there is a corresponding scaling filter given by

$$\tilde{g}_{1,l} = (-1)^{l+1}\tilde{h}_{1,L_1-1-l}, \quad l = 0,\ldots,L_1-1 \tag{41}$$

(note that the Haar scaling filter of Equation (1) is related to the Haar wavelet filter in this manner). The wavelet and scaling filters of levels $j \geq 2$ all depend on just the level $j = 1$ wavelet filter – thus specification of this single

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

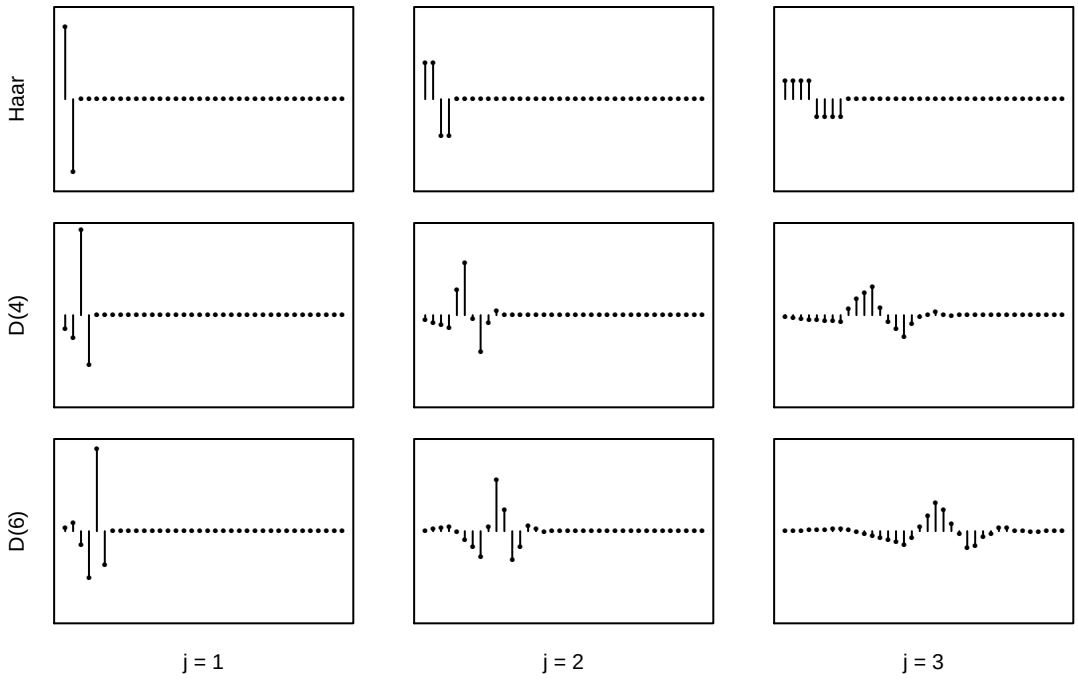

Fig. 2. Daubechies wavelet filters of widths $L_1 = 2$, 4 and 6 (Haar, D(4) and D(6) filters, respectively) for levels $j = 1$, 2 and 3.

filter sets $\{\tilde{h}_{j,l} : l = 0, \ldots, L_j - 1\}$ and $\{\tilde{g}_{j,l} : l = 0, \ldots, L_j - 1\}$ for all levels j, where we now need to make the definition $L_j = (2^j - 1)(L_1 - 1) + 1$ (note that this reduces to the definition $L_j = 2^j$ we used in the Haar case). The three fundamental properties of Equation (4) hold here also. The second and third columns of Figure 2 show $\{\tilde{h}_{2,l}\}$ and $\{\tilde{h}_{3,l}\}$ for the Haar, D(4) and D(6) filters.

We can use the non-Haar wavelet filters to create MODWTs of a time series $\{X_t\}$ that have the same key properties as the Haar MODWT. As with the Haar wavelet, Equation (2) gives us the $j = 1$ level wavelet and scaling coefficients, and we can use either Equation (5) or (6) to get the coefficients for levels $j \geq 2$. The second approach (Equation (6)) uses the pyramid algorithm, which makes it clear that the MODWT only depends on the level $j = 1$ wavelet and scaling filters (and actually just on the wavelet filter, since the scaling filter comes about via Equation (41)). The norm preserving property of Equation (7) still holds, leading to the same decomposition of the sample variance as before (Equations (8), (9) and (10)). We can define a non-Haar wavelet coefficient process of level j using Equation (12) and a corresponding jth-level wavelet variance $\nu_X^2(\tau_j)$ via Equation (14). These wavelet variances offer the same decomposition of $\mathrm{var}\{X_t\}$ as expressed by Equation (15) when $\{X_t\}$ is a stationary process, and the sum of these variances will diverge to infinity as per Equation (16) for certain non-stationary processes.

In the context of the Allan variance, why might non-Haar wavelet variances be of interest? The Allan variance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

is based on the output from the Haar wavelet filter, which, when applied to $\{X_t\}$, yields a process proportional to the first-order backward differences $\{X_t^{(1)}\}$. This fact translates into the Allan variance being well-defined (i.e., finite) when $\{X_t^{(1)}\}$ is stationary (note that, if $\{X_t\}$ is stationary, then $\{X_t^{(1)}\}$ must be also). The squared gain function for a first difference filter is approximately proportional for f^2 at low frequencies. If $S_X(f)$ obeys a power law f^α at low frequencies and is bounded elsewhere, then subjecting $\{X_t\}$ to a first differencing operation yields a process $\{X_t^{(1)}\}$ whose spectrum is approximately proportional to $f^{\alpha+2}$ at low frequencies. For $\{X_t^{(1)}\}$ to be stationary, this spectrum must integrate to a finite value, which happens as long as $\alpha > -3$. Suppose we estimate the Allan variance using, say, the unbiased maximal overlap estimator $\hat{\sigma}_X^2(2, \tau_j)$ of Equation (24). This estimator is predicated not just on the stationarity of $\{X_t^{(1)}\}$ but also on the assumption that the wavelet coefficient process $\{\overline{W}_{j,t}\}$ has mean zero. Because the Haar wavelet filter involves a differencing operation, this zero mean assumption holds when $\{X_t\}$ is stationary, but not necessarily when $\{X_t\}$ is non-stationary but $\{X_t^{(1)}\}$. In terms of $S_X(f)$, the stationarity condition for $\{X_t\}$ is $\alpha > -1$. Of the five canonical power-law noise models ($\alpha = 2$, 1, 0, -1 and -2 corresponding to white phase, flicker phase, white frequency, flicker frequency and random walk frequency noise), two are non-stationary processes, namely, flicker frequency and random walk frequency. For these two processes, even though the Allan variance is well-defined, its estimator requires the additional assumption that $E\{X_t^{(1)}\} = 0$, which might be dicey in certain applications (particularly when frequency drift is a potential problem).

Now consider a wavelet variance based upon the D(4) wavelet filter, which has width $L_1 = 4$. This filter has embedded within it a cascade of two backward differences (in general, a Daubechies wavelet filter of width L_1 involves $L_1/2$ backward differences). The associated squared gain function is approximately proportional to f^4 at low frequencies. If $S_X(f)$ varies as f^α at low frequencies and is bounded otherwise, then subjecting $\{X_t\}$ to a D(4) wavelet filter yields a process whose spectrum is approximately proportional to $f^{\alpha+4}$. The condition needed for the D(4) wavelet variance to be well-defined is $\alpha > -5$. In terms of the unbiased estimator $\hat{\sigma}_X^2(2, \tau_j)$ of the D(4) wavelet variance, the requirement that $\{\overline{W}_{j,t}\}$ has zero mean is $\alpha > -3$. Thus the estimator for the D(4) wavelet variance (or any wavelet variance associated with a filter of width $L_1 \geq 4$) can be used for all five canonical power-law noise models without the need for an additional zero mean assumption. For all wavelet filters (including the Haar), we have

$$\nu_X^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2} S_X(f)\, df \tag{42}$$

to a certain degree; i.e., a level j wavelet filter acts as a band-pass filter. Recalling the relationship $\sigma_X^2(2, \tau_j) = 2\nu_X^2(\tau_j)$ between the Allan variance $\sigma_X^2(2, \tau_j)$ and a Haar-based wavelet variance $\nu_X^2(\tau_j)$, we can define an Allanized version of a general wavelet variance $\nu_X^2(\tau_j)$ as just $2\nu_X^2(\tau_j)$, and the corresponding unbiased estimator would be $2\hat{\nu}_X^2(\tau_j)$. Comparison of Allanized D(4) wavelet variance estimates with Allan variance estimates can help identify when the latter might be suffering from problems associated with an inappropriate assumption that $E\{X_t^{(1)}\} = 0$.

Besides being able to handle a wider range of power-law noise models, an additional reason why a non-Haar wavelet variance $\nu_X^2(\tau_j)$ might be of interest is to provide a tighter connection between the time domain characterization provided by $\nu_X^2(\tau_j)$ and the frequency domain characterization provided by the spectrum $S_X(f)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

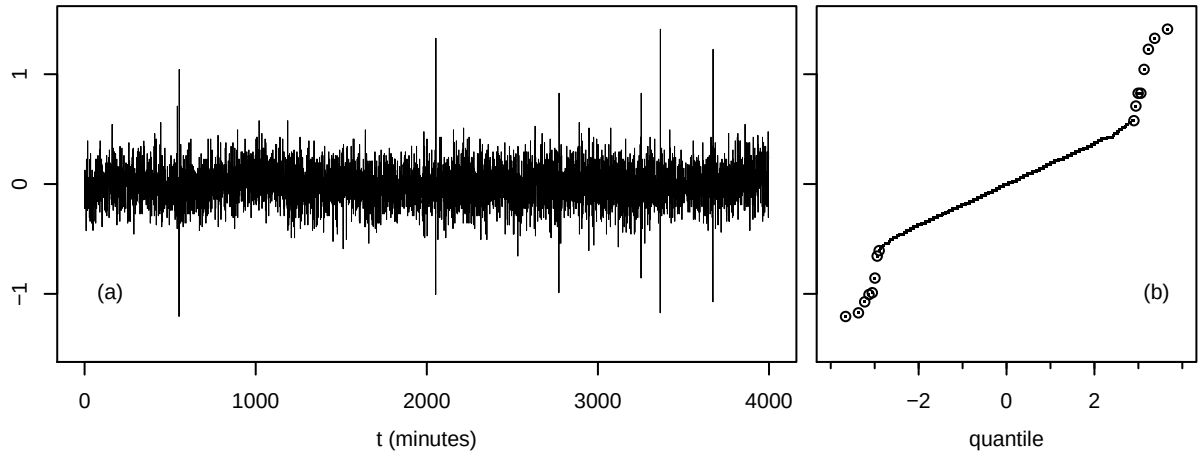

Fig. 3. Rescaled fractional frequency deviates derived from phase differences between two hydrogren masers (a) and empirical quantiles of these deviates versus quantiles from a standard Gaussian distribution (b).

As the width of a Daubechies wavelet filter increases, the approximation of Equation (42) becomes better and better. The Haar wavelet filter in general gives the poorest approximation, which means that, though well-defined, the Allan variance can suffer from undesirable leakage from one pass-band to another (see Section VII-B for an example).

VII. Examples

Here we consider two examples that illustrate some of the points brought up in the previous two sections.

A. Phase measurements between two hydrogen masers

Our first example involves a time series of phase differences $\{\phi_t\}$ between two hydrogen masers (data courtesy of L. Schmidt and D. Matsakis, US Naval Observatory). These phase differences were measured once per minute over a span of 4000 minutes and then converted into fractional frequency deviates $\{\overline{Y}_t(\tau_1)\}$ as per Equation (18) with $\tau_1 \, \Delta_t$ set to one minute. Merely to simplify the labelling of upcoming figures, rather than using $\overline{Y}_t(\tau_1)$, we will rescale it and deal with $X_t = \overline{Y}_t(\tau_1) \times 10^{12}$, which is plotted in Figure 3(a) versus t. The prominent spikes in the figure are due to phase glitches, which are rogue occurrences that do not reflect the underlying time-keeping capabilities of the masers. Figure 3(b) shows a quantile–quantile (Q–Q) plot, in which the empirical quantiles of $\{X_t\}$ are compared against corresponding theoretical quantiles from a standard normal distribution (Chambers *et al.*, 1983). A Q–Q plot such as this one provides an informal test of the null hypothesis that $\{X_t\}$ is Gaussian – an approximate linear relationship between the quantiles would say that there is little evidence against the null hypothesis. Linearity holds for the bulk of the data, but is violated in the upper and lower tails of the distribution (circles show the eight most positive and eight more negative quantiles). The lack of concordance in the tails is due to the phase glitches, which induce X_t values with magnitudes out of keeping with overall Gaussianity. If we use

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

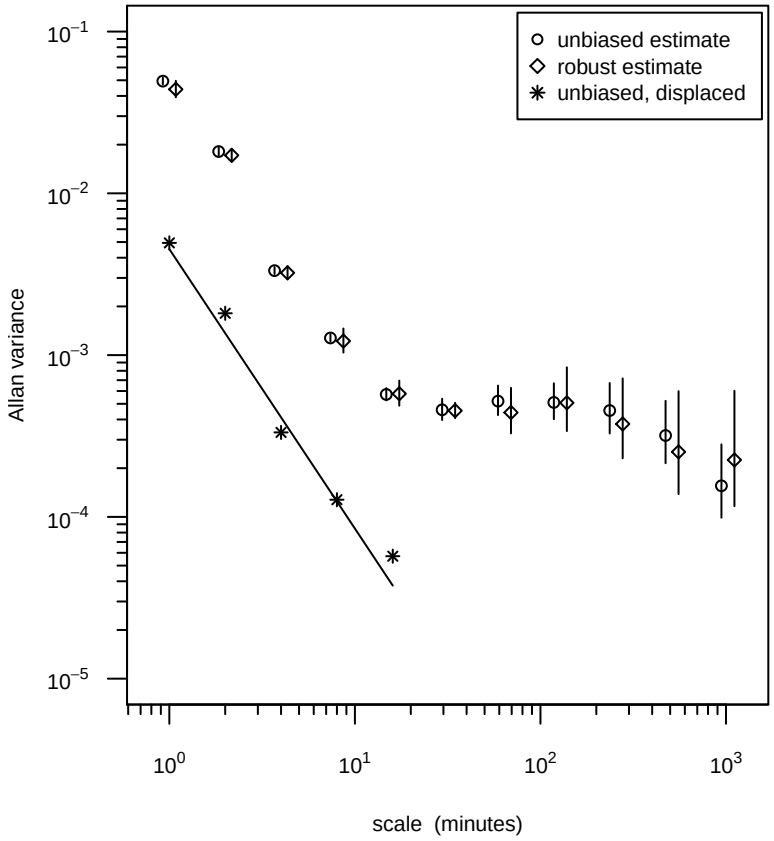

Fig. 4. Unbiased (circles) and robust (diamonds) Allan variance estimates for phase measurements between two hydrogen masers, along with 95% confidence intervals (vertical lines). The asterisks are the unbiased estimates displaced downwards by an order of magnitude, through which is placed a line fitted by least squares.

$\{X_t\}$ to estimate the Allan variance via the usual unbiased or biased estimators, there is cause for concern since these estimators are prone to distortion from discordant measurements; however, a phase glitch manifests itself as a pair of rogue X_t values, one positive, and the other negative, which might tend to cancel each other out in forming all but a few of the MODWT Haar wavelet coefficients. It is thus not unreasonable to use one of the usual estimators of the Allan variance, but it behooves us to check our results against a robust Allan variance estimator.

Figure 4 shows two sets of estimates of the Allan variance $\sigma^2_X(2,\tau_j)$ for $\{X_t\}$ plotted versus $\tau_j\,\Delta_t$, $j = 1,\ldots,11$ (the values for $\tau_j\,\Delta_t$ are offset a bit horizontally so that both sets can be more readily seen). The first set (circles) is based on the unbiased maximal overlap estimator $\hat{\sigma}^2_X(2,\tau_j)$ of Equation (24). There is a vertical line through each circle, which depicts a 95% confidence interval (CI) for the true Allan variance $\sigma^2_X(2,\tau_j)$. These CIs (some of which are so small as to be barely visible) are based on Equation (29), which involves η_j of Equation (28) – this is set by plugging $\hat{\sigma}^4_X(2,\tau_j)$ and $\hat{A}_j$ of Equation (30) into the formula for η_j. The second set (diamonds) is based on the robust Allan variance estimator $\hat{r}^2_X(2,\tau_j)$ of Equation (39). The corresponding CIs are based on Equation (29) with η_j set as per Equation (40) and with replaced $\hat{\sigma}^2_X(2,\tau_j)$ by $\hat{r}^2_X(2,\tau_j)$. The two sets of estimates are generally

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

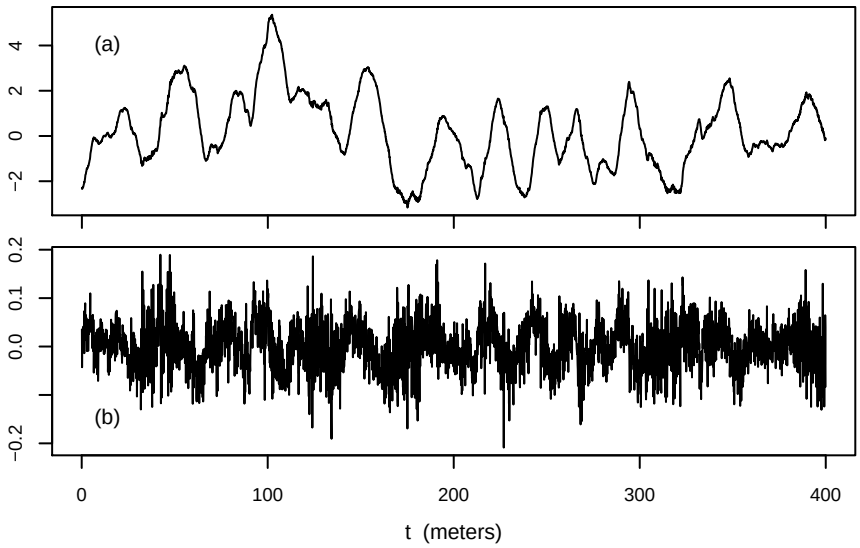

Fig. 5. Vertical ocean shear measurements (a) and their first-order backward differences (b).

in good agreement, but note that the robust estimates tend to be a bit smaller than the unbiased estimates, which is undoubtedly due to the (somewhat small) influence of the rogue X_t values. Also note that the CIs based on the robust estimates tend to be wider than those based on the unbiased estimated, which is a reflection of the fact that robustness comes at the price of increased variability in estimates.

Linear variation of $\sigma^2_X(2, \tau_j)$ versus τ_j Δ_t on a log-log plot is a classic indication of a power-law process. In the present case, linearity is not a viable hypothesis over all eleven scales, but it is arguably so over the first five scales. The asterisks in Figure 4 are the unbiased estimates for these scales relocated downward by an order of magnitude for easier viewing. The line through the asterisks is a least squares fit that takes into account the joint statistical properties of $\hat{\sigma}^2_X(2, \tau_j)$, $j = 1, \ldots, 5$ (for details, see Section 7.1 of Percival and Mondal, 2012). The slope of the line (-1.73) maps onto an estimated power-law exponent for fractional frequency noise of $\hat{\alpha} = 0.73$ since, with X_t taken to be the fractional frequency deviate $\overline{Y}_t(\tau_1)$, a spectrum $S_X(f)$ that is proportional to f^α when $f > 0$ implies that $\sigma^2_X(2, \tau_j)$ is proportional to $\tau_j^{-\alpha-1}$ to a good approximation (least square theory gives a 95% CI for the true exponent as $[0.7, 0.76]$). This estimated exponent falls between two canonical power-law processes, namely, flicker phase ($\alpha = 1$) and white frequency ($\alpha = 0$). In addition the $\hat{\sigma}^2_X(2, \tau_5)$ estimate might just as well be lumped with the τ_6, τ_7 and τ_8 estimates. A least squares fit through these four values has a slope close to zero, which maps onto a flicker frequency process ($\alpha = -1$). The ambiguity in attaching a canonical power law to the τ_5 estimate argues in favor of using Equation (29) to form a CI, which does not depend upon the identification of a power-law model.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

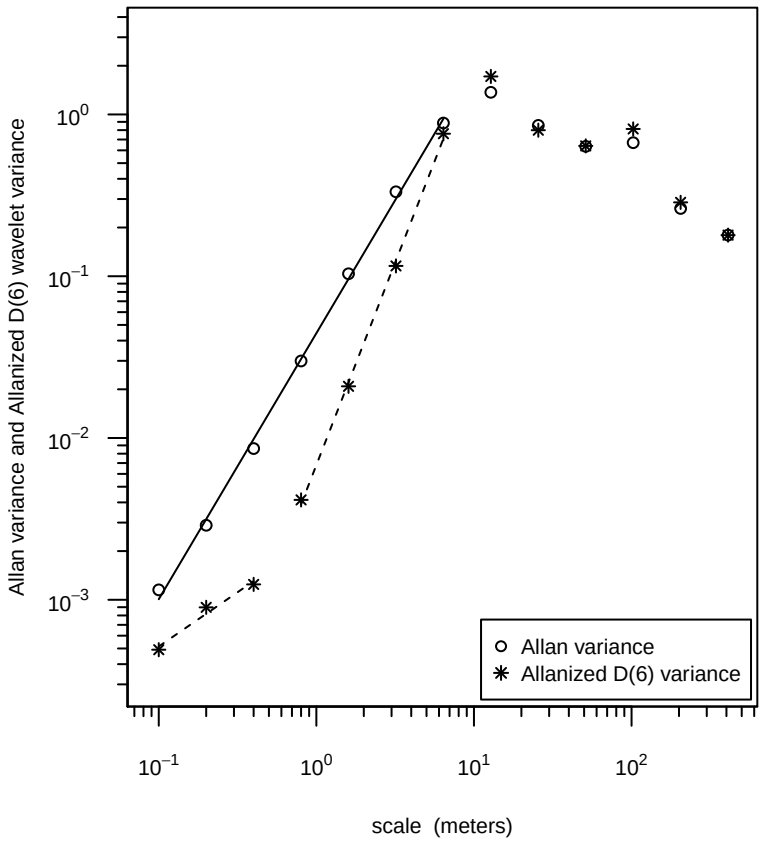

Fig. 6. Allan variance estimates (circles) and Allanized D(6) wavelet variance estimates (asterisks) versus scale for vertical ocean shear measurements (the lines are least squares fits over selected small scales).

B. Measurements of ocean shear

Our second example is a series of recordings that, while collected over time, actually reflect measurements over a grid of equally spaced depths below the surface of the ocean. These data (courtesy of M. Gregg, Applied Physics Laboratory, University of Washington) were collected by a probe dropped over the side of a ship and designed to descend vertically at a constant rate into the ocean. As it descended, the probe measured the horizontal component of water velocity every 0.1 meters, which was then differenced over 10 meter intervals and low-pass filtered to yield a depth series related to vertical shear in the ocean. Figures 5(a) and (b) show a 400 meter segment $\{X_t\}$ of these data and its first-order backward difference $\{X_t^{(1)}\}$. The appearance of $\{X_t^{(1)}\}$ is consistent with that of a realization from a stationary process, and its sample mean is 0.0005. It is thus reasonable to assume that $\{X_t\}$ is a realization of a process whose backward difference is a zero-mean stationary process, suggesting that the Allan variance might be an appropriate tool for studying the variability of ocean shear on a scale-by-scale basis.

The circles in Figure 6 show biased estimates of the Allan variance versus physical scales $\tau_j\,\Delta_t$, $j = 1, \ldots,$ 13, where $\Delta_t = 0.1$ meters (these estimates are given by $2\hat{\sigma}^2_{\widetilde{W}_j}$, where $\hat{\sigma}^2_{\widetilde{W}_j}$ is defined in Equation (33); use of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2015.2495012, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

unbiased maximal overlap estimator $\hat{\sigma}^2_X(2,\tau_j)$ of Equation (24) gives similar results at scales for which there are enough non-boundary coefficients to compute unbiased estimates). The estimates for the first seven scales appear to vary linearly on a log-log plot with a slope close to $5/3$, suggesting that ocean shear obeys a power law with exponent $\alpha = -8/3$ over these scales. Is this a reasonable conclusion to draw? One way to check is to look at wavelet variances based on wavelets of higher order than the Haar (the wavelet behind the Allan variance). The asterisks in Figure 6 show Allanized biased estimates of the D(6) wavelet variance (D(4) biased estimates give qualitatively similar results, as do D(4) and D(6) unbiased estimates). There is good agreement between the Allan variance and the Allanized D(6) wavelet variance at the seven largest scales, but not at the six smallest scales. The D(6) estimates suggest two power laws, rather than a single one: the log-log slopes are approximately $2/3$ and $5/2$, which map into power-law exponents of $\beta = -5/3$ and $\beta = -7/2$. Fully developed Kolmogorov turbulence is a power-law process with an exponent of $-5/3$, which offers a description of the small-scale variability in ocean shear. The dominant variability is at the largest scales and is due to deep jets and internal waves in the ocean. The $-7/2$ power law at intermittent scales is interpretable as a transition region between the small- and large-scale regimes.

The D(6) wavelet variances appear to have a reasonable oceanographic interpretation, which, if correct, says that the Allan variance is misleading even though the conditions needed for the Allan variance to make sense seem to hold (a series whose first difference can reasonably be regarded as a realization of a stationary process with zero mean). What went wrong here? One explanation is that the Allan variance requires $\alpha > -3$ to be well-defined. If indeed there is a transition region associated with a $\alpha = -7/2$ power law, we cannot expect the Allan variance to handle it properly. The Allan variance estimates point to a $\alpha = -8/3$ power law, which does satisfy the $\alpha > -3$ requirement, but this might be merely covering up what is really going on. (Section 8.9 of Percival and Walden, 2000, explores another explanation, namely, that the Allan variance at small scales is suffering from leakage from the dominant large scales – the Haar-based band-pass approximation of Equation (42) is so poor that the small scales are being inordinately influenced by the large scales.)

Finally we note that dividing the Allan and Allanized D(6) variance estimates in Figure (6) by two yields corresponding Haar and D(6) wavelet variance estimates. In both the Haar and D(6) cases, the sum of the thirteen empirical wavelet variances plus the corresponding empirical scaling variance $\hat{\sigma}^2_{\widetilde{V}_{13}}$ is equal to the sample variance $\hat{\sigma}^2_X \doteq 2.67$ for the time series shown in Figure 5(a) (recall Equations (10) and (8)). In the Haar case, $\hat{\sigma}^2_{\widetilde{V}_{13}} \doteq 8.72 \times 10^{-4}$, while $\hat{\sigma}^2_{\widetilde{V}_{13}} \doteq 5.59 \times 10^{-7}$ in the D(6) case. The ratios $\hat{\sigma}^2_{\widetilde{V}_{13}}/\hat{\sigma}^2_X$ are, respectively, 3.27×10^{-4} and 2.09×10^{-7}, which says that all but a small proportion of the sample variance is accounted for by thirteen empirical wavelet variances (either Haar or D(6)). Thus, although the Allan variance was motivated in part by a dissatisfaction with the sample variance, this example demonstrates that an appropriate sum of rescaled Allan variance estimates is – to a very good approximation – equivalent to the sample variance (however, the analysis of variance corresponding to the Allan variance is arguably misleading in comparison to the one given by the Allanized D(6) variance).

VIII. Conclusion

The Allan variance has figured prominently in the assessment of high-performance clocks and frequency standards and in many other applications since its introduction fifty years ago by D.W. Allan and J.A. Barnes. The arrival of the discrete wavelet transform (DWT) two decades later led to a retrospective connection between a wavelet variance based on a Haar DWT and the Allan variance. After an innocuous division by two, the Allan variance constitutes a scale-based analysis of variance for stationary processes and for non-stationary processes whose first-order backward differences are stationary. Standard estimators of the Allan variance are also standard estimators of the wavelet variance, but the latter has a well-developed asymptotic theory that can be applied to the former, obviating the need to assume a (sometimes questionable) power-law noise model. Recent work on estimators of the wavelet variance in specialized situations (gappy data and contaminated data) can be readily adapted to serve as Allan variance estimators. Finally wavelet variances associated with non-Haar wavelet filters can serve as useful checks on the validity of an Allan variance analysis. Viewing the Allan variance through the retrospective lens of wavelets has thus served to deepen our understanding of this important analysis tool.

References

[1] Aldrich, E.M., *Alternative Estimators of Wavelet Variance*, MS dissertation, Department of Statistics, University of Washington, 2005.

[2] Allan, D.W., Statistics of atomic frequency standards, *Proceedings of the IEEE*, **54**, 221–230, 1966.

[3] Barnes, J.A., Atomic timekeeping and the statistics of precision signal generators, *Proceedings of the IEEE*, **54**, 207–220, 1966.

[4] Barnes, J.A., A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot and G.M.R. Winkler, Characterization of frequency stability, *IEEE Transactions on Instrumentation and Measurement*, **20**, 105–120, 1971.

[5] Beran, J., *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.

[6] Blackman, R.B., and J.W. Tukey, *The Measurement of Power Spectra*, Dover Publications, New York, Cambridge University Press, Cambridge, England, 1958.

[7] Box, G.E.P., G.M. Jenkins and G.C. Reinsel, *Time Series Analysis: Forecasting and Control* (Fourth Edition), John Wiley & Sons, Hoboken, New Jersey, 2008.

[8] Chambers, J.M., W.S. Cleveland, B. Kleiner and P.A. Tukey, *Graphical Methods for Data Analysis*, Duxbury Press, Boston, 1983.

[9] Craigmile, P.F., and D.B. Percival, Asymptotic decorrelation of between-scale wavelet coefficients, *IEEE Transactions on Information Theory*, **51**, 1039–1048, 2005.

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

[11] David, H.A., Bias of S^2 Under Dependence, *The American Statistician*, **39**, 201, 1985.

[12] Day, D.D., Wave arrival detection in acoustic well logs via the Wigner distribution and generalized Allan Variance, *Digital Signal Processing*, **2**, 135–145, 1992.

[13] El-Sheimy, N., H. Hou and X. Niu, Analysis and modeling of inertial sensors using Allan variance, *IEEE Transactions on Instrumentation and Measurement*, **57**, 140–149, 2008.

[14] Fadel, P.J., H.S. Orer, S.M. Barman, W. Vongpatanasin, R.G. Victor and G.L. Gebber, Fractal properties of human muscle sympathetic nerve activity, *American Journal of Physiology – Heart and Circulatory Physiology*, **286**, H1076–H1087, 2004.

[15] Flandrin, P., Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Transactions on Information Theory*, **38**, 910–917, 1992.

[16] Gambis, D., Allan Variance in earth rotation time series analysis, *Advances in Space Research*, **30**, 207–212, 2002.

[17] Gebber, G.L. H.S. Orer and S.M. Barman, Fractal noises and motions in time series of presympathetic and sympathetic neural activities, *Journal of Neurophysiology*, **95**, 1176–1184, 2006.

[18] Giraitis, L., and D. Surgailis, CLT and other limit theorems for functionals of Gaussian processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **70**, 191–212, 1985.

[19] Greenhall, C.A., Recipes for degrees of freedom of frequency stability estimators, *IEEE Transactions on Instrumentation and Measurement*, **40**, 994–999, 1991.

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

[21] Guerrier, S., J. Skaloud, Y. Stebler, M.-P. Victoria-Feser, Wavelet-variance-based estimation for composite stochastic processes, *Journal of the American Statistical Association*, **108**, 1021–1031, 2013.

[22] Haar, A., Zur Theorie der Orthogonalen Funktionensysteme, *Mathematische Annalen*, **69**, 331–371, 1910.

[23] Huber, P.J., Robust estimation of a location parameter, *Annals of Mathematical Statistics*, **35**, 73–101, 1964.

[24] Kebabian, P.L. S.C.!Herndon and A. Freedman Detection of nitrogen dioxide by cavity attenuated phase shift spectroscopy, *Analytical Chemistry*, **77**, 724–728, 2005.

[25] Loescher, H.W., T. Ocheltree, B. Tanner, E. Swiatek, B. Dano, J. Wong, G. Zimmerman, J. Campbell, C. Stock, L. Jacobsen, Y. Shiga, J. Kollas, J. Liburdy and B.E. Law, Comparison of temperature and wind statistics in contrasting environments among different sonic anemometer-thermometers, *Agricultural and Forest Meteorology*, **133**, 119–139, 2005.

[26] Mandelbrot, B.B. and J.W. van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Review*, **10**, 422–437, 1968.

[27] Mondal, D., *Wavelet Variance Analysis for Time Series and Random Fields*, PhD dissertation, Department of Statistics, University of Washington, 2007.

[28] Mondal, D., and D.B. Percival, Wavelet variance analysis for gappy time series, *Annals of the Institute of Statistical Mathematics*, **62**, 943–966, 2010.

[29] Mondal, D., and D.B. Percival, M-estimation of wavelet variance, *Annals of the Institute of Statistical Mathematics*, **64**, 27–53, 2012.

[30] Percival, D.B., *The Statistics of Long Memory Processes*, PhD dissertation, Department of Statistics, University of Washington, 1983.

[31] Percival, D.B., Characterization of frequency stability: frequency-domain estimation of stability measures, *Proceedings of the IEEE*, **79**, 961–972, 1991.

[32] Percival, D.B., On estimation of the wavelet variance, *Biometrika*, **82**, 619–631, 1995.

[33] Percival, D.B., and P. Guttorp, Long-memory processes, the Allan variance and wavelets, in *Wavelets in Geophysics*, edited by E. Foufoula-Georgiou and P. Kumar, Academic Press, San Diego, 325–344, 1994.

[34] Percival, D.B., and D. Mondal, A wavelet variance primer, in *Time Series Analysis: Methods and Applications (Handbook of Statistics, Vol. 30)*, edited by T. Subba Rao, S. Subba Rao and C.R. Rao, Elsevier, Amsterdam, 623–657, 2012.

[35] Percival, D.B., and A.T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press, Cambridge, England, 1993.

[36] Percival, D.B., and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, England, 2000.

[37] Rutman, J., Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress, *Proceedings of the IEEE*, **66**, 1048–1075, 1978.

[38] Sabatini, A.M., A wavelet-based bootstrap method applied to inertial sensor stochastic error modelling using the Allan variance, *Measurement Science and Technology*, **17**, 2980–2988, 2006.

[39] Scheffé, H., *The Analysis of Variance*, John Wiley & Sons, New York, 1959.

[40] Schieder, R. and C. Kramer, Optimization of heterodyne observations using Allan variance measurements, *Astronomy and Astrophysics*, **373**, 746–756, 2001.

[41] Serroukh, A., A.T. Walden and D.B. Percival, 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, 2000.

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

[43] Thomson, D.J., Spectrum estimation and harmonic analysis, *Proceedings of the IEEE,* **70**, 1055–1096, 1982.

[44] von Neumann, J., R.H. Kent, H.R. Bellinson and B.I. Hart, The mean square successive difference, *Annals of Mathematical Statistics*, **12**, 153–162, 1941.

[45] von Neumann, J., Distribution of the ratio of the mean square successive difference to the variance, *Annals of Mathematical Statistics*, **12**, 367–395, 1941.

[46] von Neumann, J., A further remark concerning the distribution of the ratio of the mean square successive difference to the variance, *Annals of Mathematical Statistics*, **13**, 86–88, 1942.

[47] Werle, P., R. Miicke and F. Slemr, The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS), *Applied Physics B: Lasers and Optics*, **57**, 131–139, 1993.

[48] Yoshimura, K., T. Shinozuka, Y. Suzuki, H. Mineno, M. Shimada and H. Kuroiwa, Calculation of the Allan variances for the rain-attenuated millimeter wave beacon from a geostationary satellite, and their application to the site-diversity switching, *Electronics and Communications in Japan*, **66-B**, 81–89, 1983.

Donald B. Percival received the B.A. degree in astronomy from the University of Pennsylvania, Philadelphia, PA, USA, in 1968, the M.A. degree in mathematical statistics from George Washington University, Washington, DC, USA, in 1976, and the Ph.D. degree in statistics from the University of Washington, Seattle, WA, USA, in 1982. He was an Astronomer with the Time Service Division, U.S. Naval Observatory, Washington, from 1968 to 1978, where he was involved in the generation of atomic clock time scales and analysis of the frequency stability of high-performance oscillators. Since 1982, he has been with the Applied Physics Laboratory and the Department of Statistics, University of Washington, where he is currently a Senior Principal Mathematician and Professor. He has co-authored (with A.T. Walden) the textbooks *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge, U.K.: Cambridge University Press, 1993) and *Wavelet Methods for Time Series Analysis* (Cambridge, U.K.: Cambridge University Press, 2000). His research interests include spectral analysis, wavelets, and the use of statistical methodology in physical sciences.