International Workshop on
Advanced Mathematical Tools in Metrology
Torino, October 1993

AN INTRODUCTION TO SPECTRAL ANALYSIS AND WAVELETS

Donald B. Percival
Applied Physics Laboratory, HN-10
University of Washington
Seattle, WA 98195
U.S.A.
dbp@apl.washington.edu

Abstract. In metrology experimentalists often measure quantities with respect to an independent variable that is ordered such as time, depth or distance along a line. Time series analysis is the study of the statistical properties of such ordered measurements. Here we briefly introduce two important techniques in time series analysis, namely, spectral analysis and wavelet analysis, both of which can be of value in understanding metrological data.

1. Introduction

The spectral analysis of a time series is one of the oldest and most widely used analysis techniques in the physical sciences. The basic idea behind spectral analysis is to decompose the variance of a time series into a number of components, each one of which can be associated with a particular frequency. Although the basic ideas and concepts behind spectral analysis are quite old (two early references are Stokes, 1879, and Schuster, 1898), the current wide-spread use and interest in the subject arose because of three important events in late 1950s and mid 1960s: the publication of the influential exposition by Blackman and Tukey (1958); the rediscovery of the fast Fourier transform algorithm (Cooley and Tukey, 1965); and the increasing availability of powerful electronic computers which allowed practictioners to carry out the necessary computations. Today spectral analysis is an 'industry standard' in the physical sciences and can be readily done on general purpose computers using software in a variety of forms (several vendors also offer special purpose hardware – known as 'spectrum analyzers' – which can collect time series and perform spectral analysis nearly in real-time).

By contrast, wavelet analysis of a time series is relatively new and not very widespread. The basic idea behind wavelet analysis is to decompose the variance of a time series into a number of components, each one of which can be associated with a particular scale at a particular time. The subject of wavelets is currently a 'hot topic' in mathematics and signal processing and formally dates back to just the early 1980s. In some aspects, however, this subject represents a synthesis of older ideas from many different areas. The current interest in wavelets is motivated by the discovery of both many new elegant mathematical results and efficient algorithms for computing the wavelet transform.

Since both spectral analysis and wavelet analysis of a time series are analysis of variance techniques, there are a number of interesting comparisons and contrasts between the

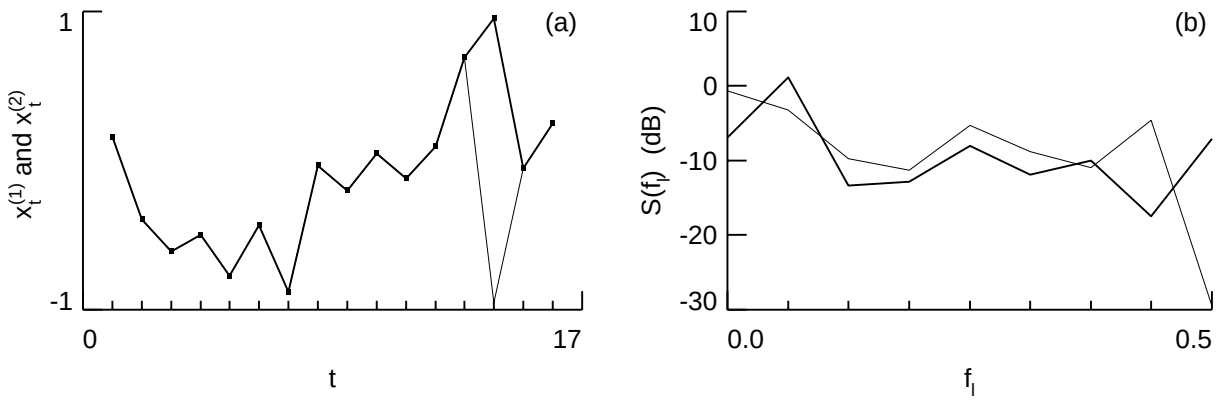

Figure 1. Two time series (plot a) and their discrete energy spectra (plot b). The small solid squares in (a) indicate the 16 values of the time series $\{x_t^{(1)}\}$ plotted versus its time index t for $t = 1, \ldots, 16$. These squares are connected by thick lines merely to aid the eye in following the series. The two thin lines define a second series of 16 values, namely $\{x_t^{(2)}\}$, that differs only from $\{x_t^{(1)}\}$ in that $x_{14}^{(2)} = -x_{14}^{(1)}$. The discrete energy spectra $\{S(f_l)\}$ plotted versus Fourier frequency f_l are shown in plot (b) for $\{x_t^{(1)}\}$ (thick curve) and for $\{x_t^{(2)}\}$ (thin). Both spectra consist of 9 points connected by lines.

two subjects. This paper is intended to be a brief introduction to both subjects with an emphasis on ideas rather than details. More extensive tutorial introductions to spectral analysis can be found in Koopmans (1983), Brillinger (1974), Kay and Marple (1981), Jones (1985), De Marchi and Lo Presti (1993) and Percival (1994), while tutorials on wavelets are given in Strang (1989), Rioul and Vetterli (1991), Section 13.10 of Press *et al.* (1992), Strang (1993), Meyer (1993), Strichartz (1993) and Tewfik *et al.* (1993). There are numerous books on spectral analysis, including Jenkins and Watts (1967), Koopmans (1974), Bloomfield (1976), Priestley (1981), Kay (1988) and Percival and Walden (1993). The book by Daubechies (1992) discusses the mathematical side of wavelets, while Vaidyanathan (1993) presents the subject from a signal processing point of view (also, the March, 1992, issue of the *IEEE Transactions on Information Theory* is devoted to wavelet transforms and multiresolution signal analysis and contains a wealth of interesting material).

2. Orthogonal Transformation of Time Series

Both the spectral analysis and the wavelet analysis of a time series can be viewed in terms of an orthogonal transformation, so let's briefly review the ideas behind such a transformation. Let $x_1, x_2, \ldots, x_N$ represent a time series of N observations (two such series with $N = 16$ values are shown in Figure 1(a)). Let $g_j(\cdot)$ for $j = 1, 2, \ldots, N$ be a set of real-valued functions such that

$$\sum_{t=1}^{N} g_j(t) g_k(t) = \begin{cases} 1, & \text{if } j = k; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

We can then analyze our time series with respect to the basis provided by the $g_j(\cdot)$'s by

computing the transform coefficients χ_j defined as

$$\chi_j = \sum_{t=1}^{N} x_t g_j(t), \;\; j = 1, 2, \ldots, N. \tag{2}$$

Given the transform coefficients, we can then reconstruct ('synthesize') our time series from the χ_j's using

$$x_t = \sum_{j=1}^{N} \chi_j g_j(t). \tag{3}$$

Let us *define* the 'energy' in our time series as $\sum_{t=1}^{N} x_t^2$. Note that x_t^2 is just the contribution to the energy at time t so that a plot of x_t^2 versus t gives us a decomposition of energy with respect to time. A fundamental identity (known as Parseval's theorem) is that

$$\sum_{t=1}^{N} x_t^2 = \sum_{j=1}^{N} \chi_j^2. \tag{4}$$

The above identity states that the energy of our time series is preserved in the transform coefficients χ_j and can be decomposed with respect to the squares of these coefficients. While there are an uncountable number of possible orthogonal transformation for a time series, our ability to make any sense out of the energy decomposition provided by arbitrary χ_j's is limited. The key to a meaningful analysis is to pick a transformation such that the χ_j's have some physical interpretation so that the energy decomposition with respect to these χ_j's tells us something interesting about our time series. Two orthogonal transformations that do so are the Fourier transform (providing the basis for spectral analysis) and the wavelet transform. We discuss the former in the next section and the latter in Section 5.

3. The Fourier Transform

The Fourier transform of a time series is an orthogonal transformation in which the $g_j(\cdot)$'s are 'stretched' and normalized versions of the trigonometric functions $\cos(t)$ and $\sin(t)$. Stretching is accomplished by introducing the notion of angular frequency $2\pi f$ to produce $\cos(2\pi ft)$ and $\sin(2\pi ft)$. In order that the orthogonality conditions of Equation (1) be satisfied, we must choose our frequencies carefully. A choice that works is the set known as the Fourier frequencies, namely, $f_l \equiv l/N$ for $l = 0,, \ldots, N/2$ (for simplicity, we assume that N is an even integer in what follows). We can now define the $g_j(\cdot)$'s as

$$g_j(t) = \begin{cases} \sqrt{2/N}\cos\left([j+1]\pi t/N\right), & j = 1,\, 3,\, \ldots,\, N-3; \\ \sqrt{2/N}\sin\left(j\pi t/N\right), & j = 2,\, 4,\, \ldots,\, N-2; \\ \cos\left(\pi t\right)/\sqrt{N}, & j = N-1; \\ \cos\left(0\cdot t\right)/\sqrt{N} = 1/\sqrt{N}, & j = N. \end{cases}$$

The χ_j's we obtain via (2) can then be used to give a Fourier synthesis of our time series using (3).

As an example, the thick curves in the 16 plots of Figure 2 show $\chi_j g_j(t)$ versus t for the 16 point time series $\{x_t^{(1)}\}$ shown in Figure 1(a) (for plotting purposes, we let t

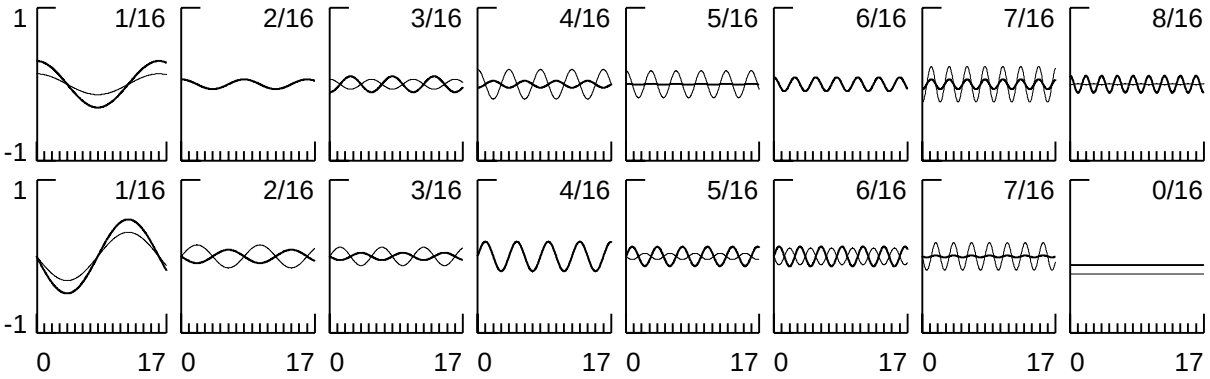

Figure 2. Fourier synthesis of $\{x_t^{(1)}\}$ (thick curves) and $\{x_t^{(2)}\}$ (thin curves).

vary continuously over the range 0 to 17 rather than just take on integer values from 1 to 16). The left-most column of plots shows $\chi_1 g_1(\cdot)$ (top) and $\chi_2 g_2(\cdot)$ (bottom); the next column shows $\chi_3 g_3(\cdot)$ and $\chi_4 g_4(\cdot)$; and so on. The Fourier frequency f_l associated with each $\chi_j g_j(\cdot)$ is written in the upper right-hand corner of each plot. Equation (3) tells us that we can reconstruct ('synthesize') the 16 point time series by adding together the thick curves in the 16 plots at the points $t = 1, \ldots, 16$.

Because of Equation (4), we can decompose the energy in a time series with respect to the $N/2 + 1$ Fourier frequencies in the following way. Let us define a discrete energy spectrum $S(f_l)$ such that

$$S(f_l) = \begin{cases} \chi_{2l-1}^2 + \chi_{2l}^2, & f_l = l/N \text{ for } 1 \le l \le N/2 - 1; \\ \chi_{N-1}^2, & f_l = 1/2 \text{ (i.e., } l = N/2); \\ \chi_N^2, & f_l = 0 \text{ (i.e., } l = 0). \end{cases}$$

Equation (4) now tells us that

$$\sum_{l=0}^{N/2} S(f_l) = \sum_{t=1}^{N} x_t^2; \tag{5}$$

i.e., $S(f_l)$ gives us the contribution to the energy due to sinusoidal components with frequency f_l.

The thick curve in Figure 1(b) shows $S(f_l)$ versus f_l for $\{x_t^{(1)}\}$ (following standard engineering practice, we have plotted $S(f_l)$ on a decibel scale, i.e., $10 \log_{10}(S(f_l))$). This plot tells us, for example, that the lowest frequency term ($f_1 = 1/16$) contributes more to the energy in our time series than any of the other terms. We can readily see this dominance in Figure 2 (the 'synthesis' plot): the amplitudes of the sinusoids associated with f_1 are larger than those for any other frequency. This correspondence is not surprising: the amplitudes are governed by the χ_j's, whereas the $S(f_l)$'s are governed by the χ_j^2's.

In order to demonstrate in Section 5 one key difference between the Fourier and wavelet transforms, let us see what happens if we change a single value in our 16 point time series, namely x_{14}, so that it is now $-x_{14}$ to form the series $\{x_t^{(2)}\}$ (as shown by the light lines in Figure 1(a)). By construction, this time series has the same energy as

$\{x_t^{(1)}\}$. The thin curves in the plots of Figure 2 show the Fourier synthesis of $\{x_t^{(2)}\}$, while the thin curve in Figure 1(b) shows its discrete energy spectrum. Note that, although we have altered just a single point in our time series in an energy preserving fashion, all but three of the χ_j's and all of the $S(f_l)$'s have been altered (some rather substantially). This simple experiment emphasizes the point that Fourier transform coefficients are not well 'localized' in time; i.e., the χ_j's are determined by x_t's spread out over a wide range of indices t so that changes in just a few x_t's can alter all of the coefficients χ_j.

To summarize, the use of sinusoids with a carefully chosen set of frequencies allows us to define a discrete energy spectrum. This spectrum tells us how the energy in a time series is decomposed on a frequency by frequency basis. Large values of the discrete energy spectrum tell us which sinusoids are most significant in synthesizing the time series using a linear combination of sinusoids.

4. The Spectral Estimation Problem

For practical use the simple Fourier theory we outlined in the preceding section is not satisfactory for analyzing a time series because the frequencies are dictated by the sample size N and because we have no way of accounting for the effect of stochastic variations. Properties of the discrete energy spectrum can thus be highly dependent on both the number of observations we have collected and also the particular set of stochastic variations we happened to encounter in collecting our series.

To overcome these defects, considerable effort has been expended over the last half century in extending the theory of the preceding section so that it applies to the class of stationary stochastic process $\{X_t\}$ (by definition, the expected value $E\{X_t\}$ of such a process is a constant, and the covariance $\mathrm{cov}\,\{X_t, X_{t+\tau}\}$ between two members of the process separated by τ units depends on just the lag τ and not the time index t). The theory of stationary processes allows us to rigorously define a spectral density function $S_X(\cdot)$ (sdf) that decomposes ('analyzes') the variance of the process $\{X_t\}$ with respect to frequency; i.e., a fundamental property of the sdf is that

$$\int_{-1/2}^{1/2} S_X(f)\, df = \mathrm{var}\,\{X_t\}. \tag{6}$$

Note the following similarities between the above equation and Equation (5):

[1] The left-hand sides of both equations express the summing of the spectrum or sdf over a set of frequencies (by definition, $S_X(-f) = S_X(f)$ for real-valued processes – the introduction of negative frequencies in the sdf is purely for mathematical convenience).

[2] The right-hand side of Equation (5) is a sum of squares, while a common estimator for $\mathrm{var}\,\{X_t\}$ in the right-hand side of (6) is the sample variance $\sum_{t=1}^N X_t^2/N$, i.e., a normalized sum of squares. If $\sum_{t=1}^N X_t^2$ can be regarded as a measure of 'energy,' then $\sum_{t=1}^N X_t^2/N$ and hence $\mathrm{var}\,\{X_t\}$ can be regarded as a measure of 'power' – the sdf $S_X(\cdot)$ is thus sometimes called the power spectral density function.

With $S_X(\cdot)$ so defined, we can apply the theory of stationary processes under the assumption that our time series $x_1, \ldots, x_N$ is a portion of one realization of the stationary process $\{X_t\}$. The spectral estimation problem is to estimate $S_X(\cdot)$ from the finite time

Figure 3. Three mother wavelets: (a) the Haar wavelet; (b) Daubechies's D_4 wavelet (Strang, 1989); (c) Daubechies's least asymmetric wavelet of order 8 (Daubechies, 1992).

series $\{x_t\}$. The oldest estimator in the literature (dating back to at least Schuster, 1898) is the periodogram, defined (for a process such that $E\{X_t\} = 0$) as

$$\hat{S}_X^{(p)}(f) = \frac{1}{N}\left|\sum_{t=1}^{N} X_t e^{-i2\pi ft}\right|^2, \text{ where } i \equiv \sqrt{-1}$$

(the quantity between the vertical bars is proportional to the discrete Fourier transform of $X_1, \ldots, X_N$ and can be computed efficiently using the fast Fourier transform algorithm). It can be shown that

$$\int_{-1/2}^{1/2} \hat{S}_X^{(p)}(f)\, df = \frac{1}{N}\sum_{t=1}^{N} X_t^2; \tag{7}$$

i.e., just as the integral of the sdf is the process variance (Equation (6)), the integral of the periodogram is the sample variance.

While the obvious appeal of Equation (7) suggests that we can easily solve the spectral estimation problem by using the periodogram, in fact there are substantial statistical reasons for exercising extreme caution with this estimator. The 'tragedy of the periodogram' is twofold:

[1] The expected value of the periodogram $E\{\hat{S}_X^{(p)}(f)\}$ can differ by orders of magnitude from $S_X(f)$ i.e., $\hat{S}_X^{(p)}(f)$ can be a badly biased estimator of $S_X(f)$ (even for sample sizes usually considered as 'large').

[2] The periodogram is not a consistent estimator of the sdf; i.e., its variability does not decrease to zero as the sample size N increases to infinity (as is true for, say, the sample mean as the estimator of the mean of independent and identically distributed normal random variables).

These deficiencies in the periodogram have motivated much of the research in spectral analysis over the last half century. For example, tapering and prewhitening have been introduced to construct spectral estimators with acceptable bias, while smoothing across frequencies, averaging of spectral estimators from different subsequences, and multitapering have been used to obtain consistency. Unfortunately for practictioners, there is no single spectral estimator that is optimum under all circumstances, and there are subtleties and pitfalls galore awaiting the unwary – the reader is advised to carefully study one or more of the references cited in the introduction.

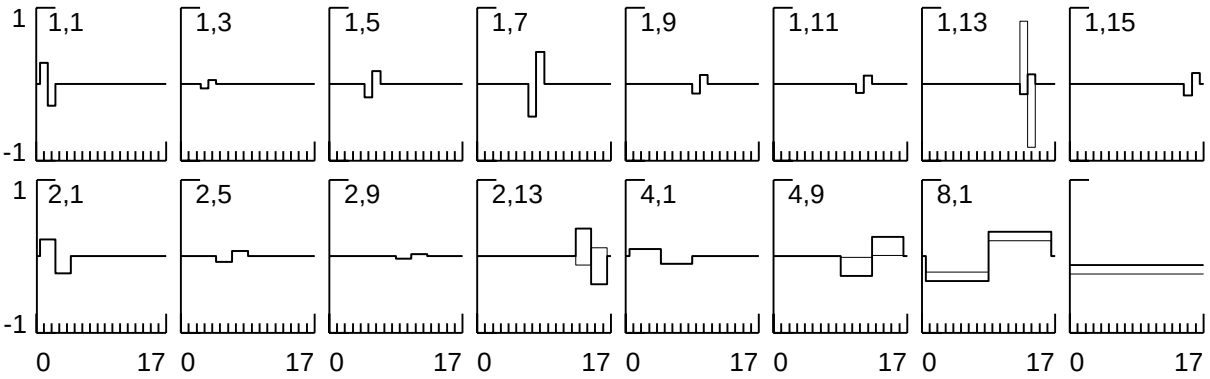

Figure 4. The Haar wavelet synthesis of $\{x_t^{(1)}\}$ (thick curves) and $\{x_t^{(2)}\}$ (thin).

5. The Wavelet Transform

The wavelet transform of a time series is an orthogonal transformation in which the $g_j(\cdot)$'s for $j = 1, \ldots, N-1$ are stretched, shifted and normalized versions of a 'localized' function known as a 'mother wavelet' ($g_N(\cdot)$ is obtained from what is known as the 'scaling function'). Three wavelets are shown in Figure 3, each of which is localized in the sense of being identically zero outside of a finite interval (as we noted previously, sinusoids are *not* localized in this sense). For simplicity, we will concentrate on the Haar wavelet (Figure 3(a)), which we choose to define here as

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

(the scaling function corresponding to the Haar wavelet can be defined as $\varphi(t) = |\psi(t)|$). Stretching is accomplished by introducing the notion of scale $\tau > 0$ to produce $\psi(t/\tau)$, thus defining a function that is nonzero for $0 \le t < 2\tau$. Shifting merely relocates $\psi(t/\tau)$ on the real axis so that it is nonzero for, say, $u \le t < 2\tau + u$; shifting can be accomplished by defining a function of t equal to $\psi([t-u]/\tau)$. In order that the orthogonality conditions of Equation (1) be satisfied, we must choose τ and u carefully. If for simplicity we assume that the sample size is a power of 2 (i.e., $N = 2^K$ for some integer K), then a choice that works is to let $\tau = 2^k$ for $k = 0, \ldots, K-1$ and – within a particular scale – to let $u = 0.5$, $2\tau + 0.5$, $4\tau + 0.5$, $\ldots$, $N - 2\tau + 0.5$. (the 0.5 offset has been included here merely to facilitate plotting). We order the $g_j(\cdot)$'s such that $g_1(\cdot), \ldots, g_{N/2}(\cdot)$ are all proportional to a Haar wavelet of scale $\tau = 1$; $g_{N/2+1}(\cdot), \ldots, g_{3N/4}(\cdot)$ are all proportional to a wavelet of scale $\tau = 2$; and so forth until we reach $g_{N-1}(\cdot)$, the single one of the $g_j(\cdot)$'s that is proportional to a wavelet of scale $2^{K-1} = N/2$. Within each scale, we order the $g_j(\cdot)$'s such that the shift u increases with j. The required normalization for a $g_j(\cdot)$ of scale τ is achieved by dividing the corresponding stretched and shifted $\psi(\cdot)$ by $\sqrt{2\tau}$. Finally, we obtain $g_N(\cdot)$ by setting it proportional to a scaling function of scale $N/2$ so that in fact $g_N(t) = |g_{N-1}(t)| = 1/\sqrt{N}$ (exactly the same as in the Fourier transform). With the $g_j(\cdot)$'s so defined, the χ_j's we obtain via Equation (2) can then be used to give a wavelet synthesis of our time series using Equation (3).

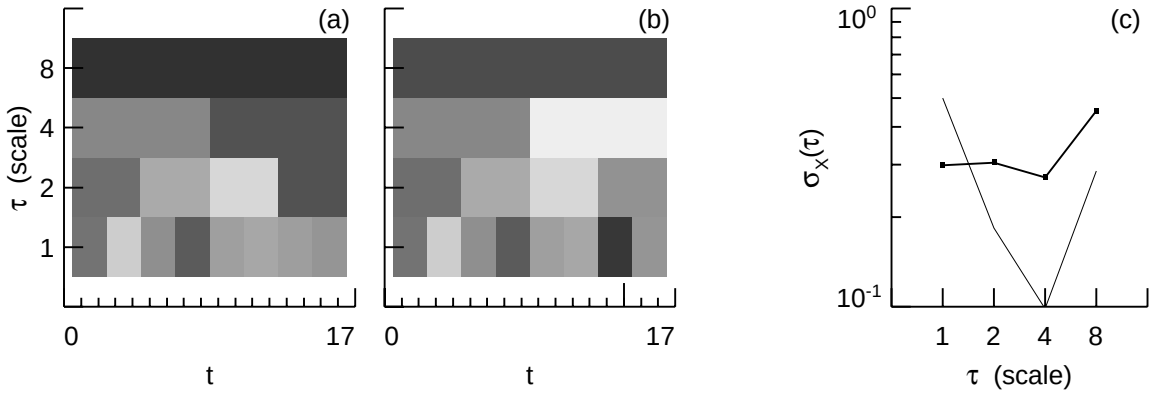

Figure 5. The scalograms for $\{x_t^{(1)}\}$ (plot a) and for $\{x_t^{(2)}\}$ (b) and σ/τ curves (c).

As an example, the thick curves in the 16 plots of Figure 4 show $\chi_j g_j(t)$ versus t for $\{x_t^{(1)}\}$ shown in Figure 1(a). The top row of plots shows, from left to right, $\chi_1 g_1(\cdot)$ to $\chi_8 g_8(\cdot)$ (all of which are of scale one), while the bottom row shows $\chi_9 g_9(\cdot)$ to $\chi_{16} g_{16}(\cdot)$ (the first four of these are of scale two; the next two are of scale four; the next one is of scale eight; and the last one corresponds to the scaling function). The two numbers 'τ, t' in the upper left-hand corner of the plots tells us that the $\chi_j g_j(\cdot)$ in the plot is associated with a wavelet of scale τ that is nonzero starting at time index t. Equation (3) tells us that we can reconstruct $\{x_t^{(1)}\}$ by adding together the thick curves.

Because of Equation (4), we can decompose the energy in a time series with respect to scale τ and time t by forming a scalogram equal to χ_j^2 for $j = 1, \ldots, N-1$. Figure 5(a) shows the scalogram for $\{x_t^{(1)}\}$, in which the values of χ_j^2 have been grey-scaled encoded such that larger values are darker. This plot tells us, for example, that χ_{N-1} (the wavelet coefficient of scale 8) contributes more to the energy in our time series than any of the other terms (cf. the synthesis plots in Figure 4).

Now let us consider what happens to the Haar wavelet transform when we change the sign of x_{14} in $\{x_t^{(1)}\}$ to create $\{x_t^{(2)}\}$. The thin curves in the plots of Figure 4 show how the wavelet synthesis of $\{x_t^{(2)}\}$ differs from that of $\{x_t^{(1)}\}$, while Figure 5(b) shows the scalogram for $\{x_t^{(2)}\}$. The important point to note is that just one coefficient per scale is changed due to the fact that the wavelet basis is localized in time. In comparing the two scalogram, note that only the grey scales of the boxes above $t = 14$ in 5(b) (indicated by the extended tick mark) differ from those in 5(a).

What can a scalogram potentially tell us about a time series? Dark boxes indicate the location of large 'events' of a given scale (consider, for example, the dark box above $t = 14$ for scale one in 5(b) – this is associated with the prominent 'downshoot' we created in $\{x_t^{(2)}\}$). At a fixed time t, going vertically up a scalogram gives a localized decomposition of the time series across different scales (note that the amount of data used in the calculation of a given scale is proportional to that scale). At a fixed scale τ, going horizontally across a scalogram tells us how the time series varies at a certain scale over time. In a certain sense, the scalogram is an attempt to capture some of the

features of both the 'time domain' decomposition of x_t^2 versus t and the 'frequency domain' decomposition of $S(f_l)$ versus f_l (see Section 7). Due to their localized nature, wavelets are a useful tool for studying the evolution of nonstationary series.

To summarize, the use of wavelets with a carefully chosen set of scales and offsets allows us to define a scalogram. This scalogram tells us how the energy in a time series is decomposed on a scale by scale basis at different times. Large values of the scalogram tell us which scales are most significant in synthesizing the time series using a linear combination of wavelets.

6. The Allan Variance

There is an interesting connection between the Haar wavelet coefficients χ_j and the Allan variance, a well-established characterization of the frequency stability of high-performance oscillators (for details, see Allan, 1966; Sullivan *et al.*, 1990; Percival, 1991; and De Marchi and Lo Presti, 1993). To establish this connection, note first that the χ_j's of scale τ are proportional to differences of nonoverlapping averages of length τ. For example, the coefficients of scale one are proportional to $x_1 - x_2$, $x_3 - x_4$, etc.; the coefficients of scale two are proportional to $(x_1 + x_2)/2 - (x_3 + x_4)/2$, $(x_5 + x_6)/2 - (x_7 + x_8)/2$, etc.; and, in general, the coefficients of scale τ are proportional to $(x_1 + x_2 + \cdots + x_\tau)/\tau - (x_{\tau+1} + x_{\tau+2} + \cdots + x_{2\tau})/\tau$, etc. This averaging and differencing scheme has a natural interpretation in the analysis of clock performance, as can be seen from the following argument. A clock is just a device that counts the cycles produced by a frequency standard oscillating at a certain nominal frequency (for example, wall clocks in the U.S. typically keep time based upon the 60 Hz oscillations in the power lines). As a thought experiment, suppose we design a clock that counts the cycles of a perfect oscillator with a 60 Hz frequency. Suppose our clock is correct at 7PM. If we temporarily change the frequency of the oscillator to 90 Hz from 7PM to 7:30PM and then to 30 Hz from 7:30PM to 8PM, the clock will display the correct time at 8PM due to the fact that the *average* frequency of the oscillator from 7PM to 8PM is equal to 60 Hz, the frequency at which the clock was designed to operate; i.e., the oscillator produced the correct number of cycles from 7PM to 8PM (although the cycles certainly did not occur equally spaced in time!).

From our thought experiment we can argue that the performance of a oscillator on a scale of an hour is governed by differences in its hourly average frequencies from one hour to the next – if these differences are small, the oscillator can be used to drive a clock that marks the passage of hours quite well. The Allan variance is a measure of how much the average frequency of an oscillator differs at difference scales (in practice, however, fractional frequency deviations are used in place of actual frequency). To define the Allan variance, let x_t be the average frequency over, say, the hour indexed by t, and let

$$\bar{x}_t(\tau) \equiv \frac{1}{\tau} \sum_{k=0}^{\tau-1} x_{t+k}$$

represent the τ hour average frequency over hours t to $t + \tau - 1$. Then the quantity $\bar{x}_t(\tau) - \bar{x}_{t+\tau}(\tau)$ represents the difference between adjacent τ hour average frequencies and is proportional to the Haar wavelet coefficient of scale τ. If we now replace the x_t's by random variables X_t, the Allan variance is defined as

$$\sigma_X^2(\tau) \equiv E\left\{ \left[\overline{X}_t(\tau) - \overline{X}_{t+\tau}(\tau) \right]^2 \right\} / \, 2$$

(this expectation is independent of the index t if we assume that $\{X_t - X_{t-1}\}$ defines a stationary process). Given a time series $\{x_t\}$ of length N, an estimator for $\sigma_X^2(\tau)$ is

$$\hat{\sigma}_X^2(\tau) = \frac{2}{N} \sum_{\text{all } j \text{ of scale } \tau} \chi_j^2$$

(this corresponds to what is known in the literature as the nonoverlapped estimator of the Allan variance). In terms of the scalogram in which the χ_j^2's are displayed with respect to scales τ and times t, we obtain $\hat{\sigma}_X^2(\tau)$ by summing all the scalogram values of a particular scale τ and then multiplying this sum by $2/N$.

A plot of $\hat{\sigma}_X(\tau)$ (i.e., the square root of the estimated Allan variance) versus scale τ is known in the literature as a 'σ/τ' curve. Figure 5(c) shows such curves for $\{x_t^{(1)}\}$ (small solid squares connected by thick lines) and $\{x_t^{(2)}\}$ (thin curve). Small values on a σ/τ curve indicate scales over which there is relatively little change in the time series (for an oscillator, small values would indicate time scales over which clocks based upon the oscillator could keep good time). Note that the σ/τ curve suffers the same defect previously noted for the discrete energy spectrum: changing a single point in a time series can alter the entire curve. The σ/τ curve is a useful summary of the information in a scalogram under an appropriate assumption of stationarity, but it is of less value for processes that are evolving over time.

To conclude our discussion of the Allan variance, we note that it has been used with considerable success since its introduction 25 years ago. This variance is most likely to be of interest in metrology applications for which the averaging and differencing scheme behind the Haar wavelet makes physical sense and for which the stationarity assumption is appropriate (if the latter is not true, the scalogram is a more appropriate analysis tool). For wavelets other than the Haar, it is possible to define a wavelet variance analogous to the Allan variance. The statistical properties of such wavelet variances are currently under study, but there are already documented examples (in the study of ocean turbulence) in which the Allan variance has been found to be markedly inferior to the other wavelet variances (Percival and Guttorp, 1994).

7. Relating Scale and Frequency

We can relate the concepts of scale and frequency by considering some well-known results from the theory of linear filters. Suppose that our time series $\{x_t\}$ is one portion of a realization of the stationary process $\{X_t\}$ with sdf $S_X(\cdot)$, and suppose that we filter our series to create a new series $\{y_t\}$; i.e.,

$$y_t = \sum_{j=0}^{J-1} a_j x_{t-j}, \quad t = J, \ldots, N,$$

where $a_0, \ldots, a_{J-1}$ are the filter coefficients. The filter output $\{y_t\}$ can be regarded as one portion of a realization of the stationary process $\{Y_t\}$ with sdf $S_Y(\cdot)$. The relationship between the sdf's $S_X(\cdot)$ and $S_Y(\cdot)$ is given by

$$S_Y(f) = |A(f)|^2 S_X(f), \quad \text{where} \quad A(f) \equiv \sum_{j=0}^{J-1} a_j e^{-i2\pi fj}.$$

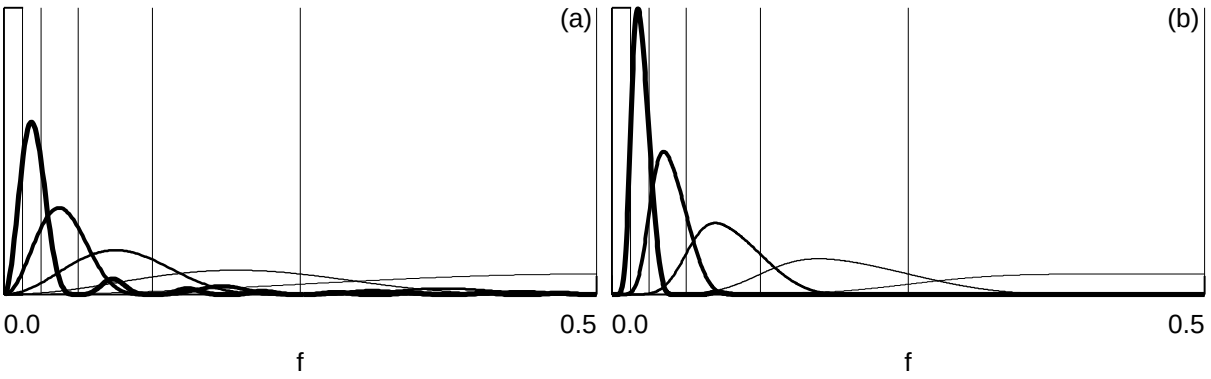

Figure 6. $|A(f)|^2$ versus f for scales 1, 2, 4, 8 and 16 (thin to thick curves) for (a) the Haar wavelet and (b) Daubechies's least asymmetric wavelet of order 8 (see Figure 3(c)). The six thin vertical lines in each plot mark – from left to right – the frequencies 1/64, 1/32, 1/16, 1/8, 1/4 and 1/2.

The function $A(\cdot)$ is known as the transfer function for the filter.

We can obtain the Haar wavelet coefficient of scale τ starting at time t from the output of the following filter of length $J = 2\tau$:

$$\left[(x_1 + x_2 + \cdots + x_\tau) - (x_{\tau+1} + x_{\tau+2} + \cdots + x_{2\tau})\right] / \sqrt{2\tau} \equiv y_{2\tau}.$$

The filter coefficients $\{a_j\}$ are thus

$$a_j = \begin{cases} -1/\sqrt{2\tau}, & 0 \le j \le \tau - 1; \\ 1/\sqrt{2\tau}, & \tau \le j \le 2\tau - 1. \end{cases}$$

Figure 6(a) shows $|A(f)|^2$ for the Haar wavelet filter versus f for scales 1, 2, 4, 8, 16 and 32. This figure shows that the Haar wavelet filter for scale τ can be regarded approximately as a bandpass filter with a passband given by the interval $[1/(4\tau), 1/(2\tau)]$. The notion of scale is thus approximately related to a certain passband of frequencies. Because the various scales τ differ from each other by factors of two, the nominal passbands are nonoverlapping. Thus, if we properly normalize the sum of squares of the output from the wavelet filter of scale τ, we can estimate the integral of the sdf over frequencies $[1/(4\tau), 1/(2\tau)]$. This scheme is exactly the idea behind an old 'quick and dirty' estimator of the sdf called pilot analysis (Blackman and Tukey, 1958, Section 18) and also allows us to translate a σ/τ curve into an estimate of the sdf (and *vice versa*).

Figure 6(b) shows $|A(f)|^2$ versus f for a more complex wavelet than the Haar, namely, Daubechies's least asymmetric wavelet of order 8 (see Figure 3(c)). The advantage of this wavelet over the Haar is the better approximation it offers to a set of nonoverlapping bandpass filters. For certain applications, this advantage is vital (see Percival and Guttorp, 1994), but there are many applications for which the Haar wavelet is entirely adequate.

To summarize, the theory of linear filters allows us to tie the notion of scale to a set of frequencies over a certain passband. With this correspondence, an estimate of the wavelet variance at scale τ can used to estimate the integral of the sdf over the frequencies $[1/(4\tau), 1/(2\tau)]$. Wavelet coefficients at different scales can be obtained by filtering a time series with a set of approximate nonoverlapping bandpass filters.

Acknowledgment

The author would like to thank Drs. Franco Pavese, Patriza Tavella, Andrea De Marchi and other organizers of the workshop for support and for their gracious help and hospitality.

References

Allan, D. W. (1966) Statistics of Atomic Frequency Standards. *Proceedings of the IEEE*, 54, 221–30.

Blackman, R. B. and Tukey, J. W. (1958) *The Measurement of Power Spectra from the Point of View of Communications Engineering.* New York: Dover Publications.

Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction.* New York: John Wiley & Sons.

Brillinger, D. R. (1974) Fourier Analysis of Stationary Processes. *Proceedings of the IEEE*, 62, 1628–43.

Cooley, J. W. and Tukey, J. W. (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19, 297–301.

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

De Marchi, A. and Lo Presti, L. (1993) *Incertezze Di Misura.* Torino: CLUT.

Jenkins, G. M. and Watts, D. G. (1968) *Spectral Analysis and Its Applications.* San Francisco: Holden-Day.

Jones, R. H. (1985) Time Series Analysis – Frequency Domain. In *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, edited by A. H. Murphy and R. W. Katz, Boulder, Colorado: Westview Press, 189–221.

Kay, S. M. (1988) *Modern Spectral Estimation: Theory and Application.* Englewood Cliffs, New Jersey: Prentice-Hall.

Kay, S. M. and Marple, S. L., Jr. (1981) Spectrum Analysis – A Modern Perspective. *Proceedings of the IEEE*, 69, 1380–419.

Koopmans, L. H. (1974) *The Spectral Analysis of Time Series.* New York: Academic Press.

Koopmans, L. H. (1983) A Spectral Analysis Primer. In *Handbook of Statistics, Volume 3: Time Series in the Frequency Domain*, edited by D. R. Brillinger and P. R. Krishnaiah, Amsterdam: North-Holland, 169–183.

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

Percival, D. B. (1991) Characterization of Frequency Stability: Frequency Domain Estimation of Stability Measures. *Proceedings of the IEEE*, 79, 961–72.

Percival, D. B. (1994) Spectral Analysis of Univariate and Bivariate Time Series. In *Statistical Methods for Physical Science*, edited by J. L. Stanford and S. B. Vardeman (a volume in the series *Methods of Experimental Physics*), New York: Academic Press.

Percival, D. B. and Guttorp, P. (1994) Long-Memory Processes, the Allan Variance and Wavelets. In *Wavelets In Geophysics*, edited by E. Foufoula-Georgiou and P. Kumar (a volume in the series *Wavelet Analysis and Its Applications* edited by C. Chui), New York: Academic Press, forthcoming.

Percival, D. B. and Walden, A. T. (1993) *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques.* Cambridge, U. K.: Cambridge University Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1992) *Numerical Recipes: The Art of Scientific Computing* (Second Edition). Cambridge, England: Cambridge University Press.

Priestley, M. B. (1981) *Spectral Analysis and Time Series*. London: Academic Press.

Rioul, O. and Vetterli, M. (1991) Wavelets and Signal Processing. *IEEE Signal Processing Magazine*, 8, 14–38.

Schuster, A. (1898) On the Investigation of Hidden Periodicities with Application to a Supposed 26 Day Period of Meteorological Phenomena. *Terrestrial Magnetism*, 3, 13–41.

Sullivan, D. B., Allan, D. W., Howe, D. A. and Walls, F. L. (1990) *Characterization of Clocks and Oscillators* (NIST Technical Note 1337). Washington, D.C.: U.S. Government Printing Office.

Stokes, G. G. (1879) Note on Searching for Periodicities. *Proceedings of the Royal Society*, 29, 122.

Strang, G. (1989) Wavelets and Dilation Equations: A Brief Introduction. *SIAM Review*, 31, 614–27.

Strang, G. (1993) Wavelet Transforms versus Fourier Transforms. *Bulletin of the American Mathematical Society*, 28, 288–305.

Strichartz, R. S. (1993) How to Make Wavelets. *The American Mathematical Monthly*, 100, 539–56.

Tewfik, A. H., Kim, M. and Deriche, M. (1993) Multiscale Signal Processing Techniques: A Review. In *Handbook of Statistics, Volume 10: Signal Processing*, edited by N. K. Bose and C. R. Rao, Amsterdam: North-Holland, 819–81.

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